

Undersökningar med flera paneler av stratifierade nätverksurval

Estimationsförfarande vid undertäckande paneler
och överlappande urvalsstratan. Med tillämpning
på HINK92-undersökningen

Bengt Rosén och Håkan Lindén



R&D Report
Statistics Sweden
Research - Methods - Development
1994:2

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 1994:2. Undersökningar med flera paneler av stratifierade nätverksurval. Estimationsförfarande vid undertäckande paneler och överlappande urvalsstratan. Med tillämpning på HINK92-undersökningen / Bengt Rosén; Håkan Lindén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

Undersökningar med flera paneler av stratifierade nätverksurval

Estimationsförfarande vid undertäckande paneler
och överlappande urvalsstratan. Med tillämpning
på HINK92-undersökningen

Bengt Rosén och Håkan Lindén



R&D Report
Statistics Sweden
Research - Methods - Development
1994:2

Från trycket
Producent
Ansvarig utgivare
Förfrågningar

April 1994
Statistiska centralbyrån, utvecklingsavdelningen
Lars Lyberg
Bengt Rosén, tel 08-783 44 90
Håkan Lindén, tel 019-17 68 98

© 1994, Statistiska centralbyrån
ISSN 0283-8680

Printed
Producer
Address

April, 1994
Statistics Sweden,
Department of Research and Development
S-115 83 Stockholm
Lars Lyberg
Bengt Rosén, telephone +46 08-783 44 90
Håkan Lindén, telephone +46 019-17 68 98

Publisher
Inquiries

© 1994, Statistics Sweden
ISSN 0283-8680

Undersökningar med flera paneler av stratifierade nätverksurval

Estimationsförfarande vid undertäckande paneler och överlappande urvalsstratan. Med tillämpning på HINK92-undersökningen

Abstract. The standard sample design in Statistics Sweden's annual sample Survey of Family Incomes (HINK) is as follows. The sample of families for a survey year comprises two panels, each of which is established by first selecting a stratified sample of adults, the sample persons, from the register of the total population. The sample persons' families constitute the family sample, leading to a family sample which is a stratified network sample. Estimation procedures for a HINK survey with this standard design are well established.

The sample to HINK with reference year 1992, HINK92, was selected in a non-standard way. To allow for good evaluation of the effects of the Swedish 90/91 tax reform, it was desirable that HINK92 and HINK89 had a great common sample. Therefore, the two panels of sample persons in HINK89 were re-used in HINK92. This leads to undercoverage problems, since by 1992 some age groups and also immigrants are not properly represented in the HINK89 panels. However, also two additional sample panels were employed. One was selected as a sub-sample from the sample of individuals to the 1992 Household Expenditures Survey. Also this panel raised certain undercoverage problems. The fourth panel was new drawn to HINK92, but it carried the non-standard feature that one of its sampling strata overlapped other strata, thereby raising another non-standard estimation problem. This report presents procedures for overcoming the mentioned obstacles due to undercoverage and overlapping sampling strata.

INNEHÅLLSFÖRTECKNING

	Sid
1 Inledning	1
2 Allmän teori för estimation vid urval av HINK-typ	2
2.1 Familje- och individpopulationerna	2
2.2 Statistiska storheter som relaterar till familjepopulationen	2
2.3 Urval av HINK-typ	3
2.4 Grundläggande relationer för punktestimation av familjestorheter	3
2.4.1 Fall 1. Urvalsstratana är disjunkta	4
2.4.2 Fall 2. Urvalsstratana tillåts vara överlappande	4
2.4.3 Punktestimation när svarsbortfall föreligger	5
2.5 Familjeregistret, och estimation baserat på det	5
2.6 Estimation av statistiska storheter som relaterar till individpopulationen	6
3 Estimation i HINK92	7
3.1 Urvalet	7
3.2 Punktestimation baserad på de olika panelerna	7
3.2.1 Punktestimation baserad på observationerna i Panel 15	7
3.2.2 Punktestimation baserad på observationerna i Panel 16	8
3.2.3 Punktestimation baserad på observationerna i Panel 19	8
3.2.4 Punktestimation baserad på observationerna i Panel 60	9
3.3 Punktestimation baserad på observationerna i samtliga paneler	10
3.3.1 Utgångspunkt	10
3.3.2 Underlag för justeringar	10
3.3.3 Sammansatta skattningar	12
3.4 Beräkning av α^* -vikterna	13
3.4.1 Panelvisa formler för α^* -vikterna	13
3.4.2 Numeriska värden på sammanvägningskoefficienterna	15
3.5 Totalt estimationsförfarande för HINK92	17
Referenser	17

Estimation vid undersökningar med flera paneler av stratifierade nätverksurval, när paneler är undertäckande och stratan överlappande. Med tillämpning på HINK92-undersökningen

1 Inledning

SCB:s Inkomstfördelningsundersökning (HINK) redovisar statistik över inkomster, och i viss mån också förmögenheter och utgifter, för landets familjer och individer. HINK är en urvalsundersökning som genomförs årligen sedan ett par decennier.

Urvalet till undersökningen för ett visst år organiseras i urvalspaneler, vilka upprättas på följande sätt. I första omgången utväljs, eller har redan tidigare utvalts, *paneler av urvalspersoner* genom urval från landets vuxna individer (18 år eller mer), medelst stratifierade urval från registret över totalbefolkningen (RTB). I andra omgången, som utgör första steget i undersökningsårets uppgiftsinsamling, inhämtas uppgift om vilka individer som ingår i urvalspersonernas familjer. Därigenom framkommer undersökningsårets *urvalspaneler av familjer* respektive *individer*. Urvalsförfarande innebär att familjepanelerna genereras som s.k. *stratifierade nätverksurval*. Rosén (1987b) ger en mer utförlig diskussion av den aspekten på HINK. HINK-undersökningar före 1992 har omfattat två paneler av urvalspersoner mot motsvarande paneler av familjer och individer. Den ena nydrogs till undersökningsåret, medan den andra drogs redan ett år tidigare (med ett tilläggsurval av 17-åringar), och användes också i föregående års undersökning.

Den fortsatta uppgiftsinsamlingen gäller samtliga familjemedlemmar i individpanelerna. De uppgifter som inhämtas gäller variabler som hänför sig dels till individer, s.k. individvariabler (t.ex. arbetsinkomst), dels till familjer, s.k. familjevariabler (t.ex. hyreskostnad). Värden på ett antal "nya" familjevariabler härleds sedan genom att addera individvariabelvärdena för medlemmarna i familjerna (t.ex. familjens arbetsinkomst). De primära och härledda variabelvärdena ligger sedan till grund för HINK:s statistik.

Urvalet till HINK med referensår 1992, *HINK92*, gjordes på "ogängse" sätt av följande skäl. För att HINK92 skulle belysa effekterna av 1990/91 års skattereform väl, önskades stort gemensamt urval för HINK92 och HINK89. Detta åstadkoms genom att de två panelerna av urvalspersoner i HINK89, benämnda Panel 15 och 16, användes om igen i HINK92. Sådan "återanvändning" av paneler medför dock undertäckningsproblem beroende på att vissa åldersklasser inte är representerade. Vidare, för att förbättra kopplingen mellan statistik om inkomster och utgifter, drogs en urvalspanel, Panel 60, som ett delurval från undersökningen av Hushållens utgifter 1992 (HUT92). Dessutom drogs en helt ny urvalspanel, Panel 19, på i stort sett gängse sätt, men den hade det ogängse inslaget att den drogs från en urvalsram med överlappande urvalsstratan, vilket inte förekommit i tidigare HINK-undersökningar.

Avstegen från tidigare urvalsförfarande medför att HINK:s gängse estimationsförfarandet måste modifieras för HINK92, och huvudsyftet med denna rapport är att ange hur estimationsförfarandet i HINK92 bör utformas. Dock, även om resonemangen förs i termer som ansluter nära till HINK92, låter sig resultaten i stort sett omgäende generaliseras till följande allmänna situation. Man vill genomföra estimation på basis av uppgifter för ett flerpanelurval, där panelerna dragits med stratifierade nätverksurval, och är behäftade med komplikationer såtillvida att urvalsramen undertäcker populationen eller har överlappande stratan.

2 Allmän teori för estimation vid urval av HINK-typ

Beskrivning av estimationsförfarandet i en gängse HINK-undersökning ges i Rosén (1987), fortsättningsvis återopad med BR87. I det följande ansluts i stort till terminologi och beteckningar enligt BR87, men modifikationerna av urvalsförfarandet i HINK92 medför att något mer omfattande begrepp och beteckningar krävs.

2.1 Familje- och individpopulationerna

Ett referensår, kallat *Rår*, förutsätts vara fastlagt. Låt;

I = populationen av *individer* som var bosatta i landet under (minst halva) *Rår*,

V = delpopulationen av alla *vuxna individer* (= individer med ålder 18+) i I .

I kallas för *individpopulationen*, och bokstaven i används för att beteckna ett genereriskt element i populationen I , eller någon del av den.

Vuxna individer klassificeras som antingen *sammanboende* eller *ensamstående*, beroende på om de lever "under äktenskapsliknande former med individ av motsatt kön" eller ej. Vid sammanboende kallas "den andre" för *maken/makan*. Äktenskapsliknande sammanboenden där en eller båda kontrahenterna ej är vuxen är så pass ovanliga att de fortsättningsvis bortses från, även om de nog förekommer. Sammanboenderelationen delar upp vuxenpopulationen;

$$V = V^E \cup V^D, \text{ där } V^E \text{ omfattar ensamstående och } V^D \text{ sammanboende vuxna.} \quad (2.1)$$

Relationen grupperar också den vuxna befolkningen i "familjekärnor", vilka består av antingen en ensamstående eller ett par av sammanboende. Motsvarande *familjer* utgörs av familjekärnan samt de barn (= individer under 18 år) för vilka den/de vuxna är vårdnadshavare. (Anm. HINK:s precisa term är inte "familj" utan *familjeenhet*, av skälet att det låter konstigt att säga att en ensamstående 19-åring utan barn utgör en familj. Trots det används i fortsättningen den kortare termen familj.) Populationen av familjer i landet, *familjepopulationen*, betecknas med U , och index d används för en generisk familj. En till (2.1) analog uppdelningen ges av;

$$U = E \cup D, \text{ där } E = \text{envuxenfamiljer och } D = \text{tvåvuxenfamiljer.} \quad (2.2)$$

Följande beteckningar används;

$d(i)$ = den familj som vuxenindivid i tillhör, $i \in V$,

$i(d)$ = den vuxne som ingår i familjen d , $d \in E$,

$i(d;1), i(d;2)$ = paret av vuxna i familjen d , $d \in D$. (Anm. Numreringen 1 och 2 har ingen annan innebörd än att den skiljer på "den ene" och "den andre".)

$m(i)$ = maken/makan till vuxenindivid i , $i \in V^D$.

I första omgången betraktas estimation av statistiska storheter för familjepopulationen.

2.2 Statistiska storheter som relaterar till familjepopulationen

Med en *familjevariabel* $\mathbf{x} = \{x_d; d \in U\}$ avses en tillordning av ett numeriskt värde till varje familj i i familjepopulationen U . *Populationstotalen* för variabeln $\mathbf{x}$, $\tau(\mathbf{x})$, *populationsstorleken*, $\gamma(U)$, och *populationsmedelvärdet* för variabeln $\mathbf{x}$, $\mu(\mathbf{x})$, definieras;

$$\tau(\mathbf{x}) = \sum_{d \in U} x_d, \quad \gamma(U) = \sum_{d \in U} 1, \quad \mu(\mathbf{x}) = \frac{\tau(\mathbf{x})}{\gamma(U)}. \quad (2.3)$$

För en delgrupp G av U , en *redovisningsgrupp*, definieras *grupp- G -totalen* för variabeln $\mathbf{x}$, $\tau(\mathbf{x};G)$, *grupp- G -storleken*, $\gamma(G)$, och *grupp- G -medelvärdet* för variabeln $\mathbf{x}$, $\mu(\mathbf{x};G)$ som;

$$\tau(\mathbf{x}; G) = \sum_{d \in G} x_d, \quad \gamma(G) = \sum_{d \in G} 1, \quad \mu(\mathbf{x}; G) = \frac{\tau(\mathbf{x}; G)}{\gamma(G)}. \quad (2.4)$$

Med användande av följande *grupp-G-indikator*;

$$1_G(d) \text{ (alternativt skrivsätt } 1(d; G)) = 1 \text{ om } d \in G, = 0 \text{ om } d \notin G, \quad (2.5)$$

kan storheterna i (2.4) uttryckas i termer av populationstotaler på följande sätt;

$$\tau(\mathbf{x}; G) = \tau(\mathbf{x} \cdot \mathbf{1}_G), \quad \gamma(G) = \tau(\mathbf{1}_G), \quad \mu(\mathbf{x}; G) = \tau(\mathbf{x} \cdot \mathbf{1}_G) / \tau(\mathbf{1}_G). \quad (2.6)$$

Från (2.6) framgår att populationstotaler är centrala statistiska storheter, och fortsättningsvis diskuteras framför allt estimation av sådana. Först ytterligare några beteckningar.

Låt Λ vara en delmängd av vuxenpopulationen V . Nedan delas familjepopulationen U upp efter hur familjekärnan relaterar till gruppen Λ ;

$$U = E(\Lambda) \cup E(\Lambda^c) \cup D(\Lambda, \Lambda) \cup D(\Lambda, \Lambda^c) \cup D(\Lambda^c, \Lambda^c), \quad (2.7)$$

$$\begin{aligned} E(\Lambda) &= \text{familjer med ensamstående vuxen som tillhör } \Lambda, \\ E(\Lambda^c) &= \text{familjer med ensamstående vuxen som inte tillhör } \Lambda \text{ (utan komplementet } \Lambda^c), \\ D(\Lambda, \Lambda) &= \text{tvåvuxenfamiljer där båda vuxna tillhör } \Lambda, \\ D(\Lambda, \Lambda^c) &= \text{tvåvuxenfamiljer där en vuxen men inte den andre tillhör } \Lambda, \\ D(\Lambda^c, \Lambda^c) &= \text{tvåvuxenfamiljer där ingen vuxen tillhör } \Lambda. \end{aligned} \quad (2.8)$$

2.3 Urval av HINK-typ

Även om urvalsförfarandena för panelerna av urvalspersoner i HINK92 skiljer sig en del från det vid tidigare HINK-undersökningar, faller de inom ramen *urval av HINK-typ* enligt nedan.

Urvalsramen för en panel utgörs av en delmängd Ω av vuxenpopulationen, och ramen är indelad i H stycken stratan, A_h , $h=1,2,\dots,H$. Det förutsätts att stratana tömmer ut Ω ;

$$\Omega = A_1 \cup A_2 \cup \dots \cup A_H, \quad (2.9)$$

men det antas inte att stratana nödvändigtvis är disjunkta. Från stratana dras *OSU-urval* (av vuxna individer). Stratumstorlek och urvalstorlek i stratum A_h betecknas med N_h respektive n_h . På detta sätt framkommer de s.k. *urvalspersonerna*. Urvalspersonernas familjer utgör *familjeurvalet*. Första steget i uppgiftsinsamlingen går ut på att fastställa precis vilka individer som ingår i de utvalda familjerna (barn såväl som vuxna), och dessa utgör *individueurvalet*. Vidare, *urvalsindikatorer* vid dragningen av urvalspersoner betecknas med I_i ;

$$I_i = 1 \text{ om individ } i \text{ blir utvald} = 0 \text{ om individ } i \text{ ej blir utvald, } i \in \Omega. \quad (2.10)$$

2.4 Grundläggande relationer för punktestimation av familjestorheter

I första omgången *antas att svarsbortfall inte föreligger*. Den för fortsättningen centrala estimatorm, $T(\mathbf{x}; \alpha)$, är baserad på observationerna från en familjepanel enligt nedan. Se Avsnitt 3 i BR87 för mer ingående diskussion av den aktuella estimatortypen.

$$T(\mathbf{x}; \alpha) = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in A_h} x_{d(i)} \cdot \alpha_i(h) \cdot I_i, \quad (2.11)$$

med en lämpligt vald uppsättning av s.k. α -vikter, $\alpha = \{\alpha_i(h); i \in V, h=1,2,\dots,H\}$. Härvid sätts $\alpha_i(h)$ till 0 så snart individ i inte ingår i stratum A_h . När stratan inte överlappar varandra är $\alpha_i(h)$ skilt från 0 bara för det enda stratum som individ i hör till, och vi skriver bara $\alpha = \{\alpha_i; i \in V\}$ (och vi tillåter oss att göra det även i andra fall, för att undvika tunga beteckningar).

2.4.1 Fall 1. Urvalsstratana är disjunkta

Den slags resonemang som genomförs mer utförligt i Avsnitt 3 i BR87 leder till följande relation (2.12), där β -koefficienterna specificeras i (2.13). Eftersom bokstaven E redan är förbrukad, betecknas väntevärde med M.

$$M[T(\mathbf{x}; \alpha)] = \sum_{d \in U} \beta_d \cdot x_d, \quad (2.12)$$

$$\begin{aligned} \beta_d &= \alpha_{i(d)} && \text{för en envuxenfamilj } d \in E(\Omega), \\ &= 0 && \text{för en envuxenfamilj } d \in E(\Omega^c), \\ &= \alpha_{i(d;1)} + \alpha_{i(d;2)} && \text{för en tvåvuxenfamilj } d \in D(\Omega, \Omega), \\ &= \alpha_{i(d;1)} && \text{för en tvåvuxenfamilj } d \in D(\Omega, \Omega^c) \text{ med } i(d;1) \in \Omega, \\ &= 0 && \text{för en tvåvuxenfamilj } d \in D(\Omega^c, \Omega^c). \end{aligned} \quad (2.13)$$

Följande *balanseringsvillkor* för α -vikterna införs (jämför Avsnitt 3.1 i BR87);

$$\begin{aligned} \alpha_i &= 1 && \text{om } i \in V^E \cap \Omega, \\ \alpha_i + \alpha_{m(i)} &= 1 && \text{om } i \in V^D \cap \Omega, m(i) \in \Omega, \\ \alpha_i &= 1 && \text{om } i \in V^D \cap \Omega, m(i) \notin \Omega, \\ \alpha_i &= \text{vad som helst} && \text{om } i \notin \Omega. \end{aligned} \quad (2.14)$$

Anmärkning: I en gängse HINK är $\Omega = V$, och villkoret (2.14) är då det "gamla vanliga" balanseringsvillkoret (3.3) i BR87. ♦

Från (2.12) och (2.13) följer att (2.15) gäller när balanseringsvillkoret (2.14) är uppfyllt;

$$M[T(\mathbf{x}; \alpha)] = \sum_{d \in E(\Omega)} x_d + \sum_{d \in D(\Omega, \Omega)} x_d + \sum_{d \in D(\Omega, \Omega^c)} x_d. \quad (2.15)$$

Nedan anges ett alternativt sätt att skriva relationen (2.15) på;

$$M[T(\mathbf{x}; \alpha)] = \tau(\mathbf{x}) - [\tau(\mathbf{x}; E(\Omega^c)) + \tau(\mathbf{x}; D(\Omega^c, \Omega^c))]. \quad (2.16)$$

2.4.2 Fall 2. Urvalsstratana tillåts vara överlappande

Argumenten i detta avsnitt räcker för det allmänna fall där dels Ω är en äkta delmängd av V, dels godtyckligt många stratan lappar över varandra. Formelkomplexiteten växer dock snabbt med antalet överlappande stratan, varför vi nöjer oss med att skriva ned formler för följande enkla fall, vilket räcker för HINK92;

Ett stratum lappar över andra stratan, medan dessa andra stratan icke-överlappar varandra samt tillsammans tömmer ut vuxenpopulationen V. (2.17)

Villkoret (2.17) innebär att vissa vuxna förekommer i två, men i högst två, stratan. Vi ansluter här beteckningsmässigt till förhållandena i Panel 19 i HINK92. Den panelen innehåller urvalspersoner från ett "särstratum", kallat *ungdomsstratumet* och betecknat med A_u , vilket omfattade individer med (1992)åldrarna 18, 19 och 20 år. Individerna i A_u , ingick också i HINKs "reguljära" stratan 01, 02, 03 och 04 (se Tabell 1).

I fall med överlappande stratan måste β -koefficienterna i (2.12) modifieras jämfört med (2.13). Den modifiering som behöver göras under (2.17) presenteras i (2.20), som fordrar några ytterligare beteckningar. I fortsättningen kommer endast sådana D-mängder (enligt (2.8)) att vara av intresse, som lägger samma villkor på båda makarna, dvs D-mängder av typen $D(\Lambda, \Lambda)$.

Den kortare beteckningen $D(\Lambda)$ används fortsättningsvis istället för $D(\Lambda, \Lambda)$. (2.18)

Sätt;

$$\begin{aligned}
V(21+) &= \text{den del av vuxenpopulationen } V \text{ som utgörs av individer med ålder } 21+ \text{ år,} \\
V(21-) &= \text{den del av } V \text{ som utgörs av individer med ålder } 18, 19 \text{ eller } 20 \text{ år,} \\
E(21+) &= \text{envuxenfamiljer där den vuxne är } 21+ \text{ år,} \\
E(21-) &= \text{envuxenfamiljer där den vuxne är } 18, 19 \text{ eller } 20 \text{ år,} \\
D(21+) &= \text{tvåvuxenfamiljer där båda vuxna är } 21+ \text{ år,} \\
D(21-) &= \text{tvåvuxenfamiljer där båda vuxna är } 18, 19 \text{ eller } 20 \text{ år.}
\end{aligned} \tag{2.19}$$

Variationer av beteckningarna i (2.19) används också, t.ex. $V^D(21+)$, $E(75+)$ och $D([21-] \cup \mathcal{S})$, med tron och förhoppningen att beteckningarna är självförklarande.

Resonemanget i Avsnitt 3 i BR87 leder till följande. Under (2.17) antar β -koefficienterna i (2.12) värden enligt nedan, där u står för undomsstratumet och r för ett reguljärt stratum (01, 02, 03 eller 04 i Tabell 1);

$$\begin{aligned}
\beta_d &= \alpha_{i(d)}(r) && \text{om } d \text{ är en envuxenfamilj med } d(i) \in E(21+), \\
&= \alpha_{i(d)}(u) + \alpha_{i(d)}(r) && \text{om } d \text{ är en envuxenfamilj med } d(i) \in E(21-), \\
&= \alpha_{i(d;1)}(r) + \alpha_{i(d;2)}(r) && \text{om } d \text{ är en tvåvuxenfamilj } (d \in D) \text{ med } d(i) \in D(21+), \\
&= \alpha_{i(d;1)}(u) + \alpha_{i(d;1)}(r) + \alpha_{i(d;2)}(r) && \text{om } d \in D \text{ med } i(d;1) \in V(21-) \text{ och } i(d;2) \in V(21+), \\
&= \alpha_{i(d;1)}(r) + \alpha_{i(d;2)}(u) + \alpha_{i(d;2)}(r) && \text{om } d \in D \text{ med } i(d;1) \in V(21+) \text{ och } i(d;2) \in V(21-), \\
&= \alpha_{i(d;1)}(u) + \alpha_{i(d;1)}(r) + \alpha_{i(d;2)}(u) + \alpha_{i(d;2)}(r) && \text{om } d \in D \text{ med } d(i) \in D(21-).
\end{aligned} \tag{2.20}$$

Följande *modifierade balanseringsvillkor* för α -vikterna införs;

$$\begin{aligned}
\alpha_i &= 1 && \text{om } i \in V^E(21+), \\
\alpha_i(u) + \alpha_i(r) &= 1 && \text{om } i \in V^E(21-), \\
\alpha_i + \alpha_{m(i)} &= 1 && \text{om } i \in V^D(21+), m(i) \in V^D(21+), \\
\alpha_i(u) + \alpha_i(r) + \alpha_{m(i)}(r) &= 1 && \text{om } i \in V^D(21-), m(i) \in V^D(21+), \\
\alpha_i(r) + \alpha_{m(i)}(u) + \alpha_{m(i)}(r) &= 1 && \text{om } i \in V^D(21+), m(i) \in V^D(21-), \\
\alpha_i(u) + \alpha_i(r) + \alpha_{m(i)}(u) + \alpha_{m(i)}(r) &= 1 && \text{om } i \in V^D(21-), m(i) \in V^D(21-).
\end{aligned} \tag{2.21}$$

Från (2.12) och (2.20) följer att nedanstående relation är uppfylld under (2.21);

$$M[T(x; \alpha)] = \tau(x). \tag{2.22}$$

2.4.3 Punktestimation när svarsbortfall föreligger

I Appendixen 1 och 2 i BR87 diskuteras hur estimationsförfarandet bör justeras vid bortfall. Den metod som rekommenderas där, och även här, är rak uppräkningsmetod enligt nedan. Hur bortfallsjusteringen görs, är dock inte väsentligt för huvudproblemet i denna rapport.

Rak uppräkning: Formeln (2.11) används med I_i som "urval-och-svarande-indikator", och

$$\text{faktorn } \frac{N_h}{n_h} \text{ byts mot } \frac{N_h}{n(s)_h}, \tag{2.23}$$

där

$$n(s)_h = \text{antal svarande familjer i stratum } A_h, \text{ med räkning av såväl "intressanta" svarande som övertäckningssvarande.} \tag{2.24}$$

2.5 Familjeregistret, och estimation baserat på det

Efter genomförd uppgiftsinsamling upprättar HINK det s.k. *familjeregistret*, betecknat $\mathcal{R}_F$, vilket beskrivs i Avsnitt 2 i BR87. Det omfattar en post för varje intressant (till skillnad från övertäckning) svarande familj med värden på familjevariabler.

Med bortfallsjustering enligt ovan kan estimatorn i (2.11), baserad på familjevariabelvärdena för en panel, skrivas på följande sätt (där estimators beteckning inte ändras trots bortfallsjusteringen);

$$T(\mathbf{x}; \alpha)_p = \sum_{d \in \text{Panel } p \text{ i } \mathcal{R}_F} X_d \cdot W_d, \quad (2.25)$$

där X_d är $\mathbf{x}$ -värdet för familj d och W_d är *uppräkningsstalet* (även kallat *familjevikten*) som beräknas enligt nedan, där $h(d)$ = det stratum som familjens *urvalsperson* $u(d)$ kommer från;

$$W_d = \frac{N_{h(d)}}{n(s)_{h(d)}} \cdot \alpha_{u(d)}(h). \quad (2.26)$$

Motsvarande estimator för en familjegrupptotal är, där 1_G är gruppindikatorn i (2.5);

$$\hat{\tau}(\mathbf{x}; G)_p = \sum_{d \in \text{Panel } p \text{ i } \mathcal{R}_F} X_d \cdot 1_G(d) \cdot W_d. \quad (2.27)$$

För framtida bruk skrivs den sista estimators även på den form som svarar mot (2.11);

$$T(\mathbf{x}; G; \alpha)_p = \sum_{h=1}^H \frac{N_h}{n(s)_h} \sum_{i \in A_h} x_{d(i)} \cdot \alpha_i(h) \cdot 1_G(d(i)) \cdot I_i. \quad (2.28)$$

I Avsnitt 4 i BR87 beskrivs hur formeln (2.28) ger underlag för variansskattningar med hjälp av programvaran SMED.

2.6 Estimation av statistiska storheter som relaterar till individpopulationen

Individpopulationen I infördes i Avsnitt 2.1. Följande begrepp och beteckningar är analoga till dem i Avsnitt 2.2. Med en *individvariabel* $\mathbf{x} = \{x_i; i \in I\}$ avses en tillordning av ett numeriskt värde till varje individ i populationen I . *Populationstotalen* för variabeln $\mathbf{x}$, $\tau(\mathbf{x})$, *populationsstorleken*, $\gamma(I)$, och *populationsmedelvärdet* för variabeln $\mathbf{x}$, $\mu(\mathbf{x})$, definieras;

$$\tau(\mathbf{x}) = \sum_{i \in I} x_i, \quad \gamma(I) = \sum_{i \in I} 1, \quad \mu(\mathbf{x}) = \frac{\tau(\mathbf{x})}{\gamma(I)}. \quad (2.29)$$

För en delgrupp G av populationen I definieras *grupp- G -totalen* för variabeln $\mathbf{x}$, $\tau(\mathbf{x}; G)$, *grupp- G -storleken*, $\gamma(G)$, och *grupp- G -medelvärdet* för variabeln $\mathbf{x}$, $\mu(\mathbf{x}; G)$, enligt (2.4). Relationerna i (2.6) gäller med *grupp- G -indikatorn*;

$$1_G(i) \text{ (alternativt skrivsätt } 1(i; G)) = 1 \text{ om individ } i \in G, = 0 \text{ om individ } i \notin G, \quad (2.30)$$

HINK upprättar också ett *individregister* $\mathcal{R}_I$ med en post för varje familjemedlem (inklusive barn) i de familjer som ingår i familjeregistret $\mathcal{R}_F$. Varje post påförs ett *uppräkningsstal* (även kallat *individvikt*) IW_i , nämligen familjens uppräkningsstal;

$$IW_i = W_{d(i)}, \text{ med } W_d \text{ enligt (2.26).} \quad (2.31)$$

Observationsmaterialet i individregistret används för att skatta statistiska storheter som hänför sig till individpopulationen. Ur teoretisk synvinkel är det enklast att se skattning av individgruppstotaler som specialfall av skattning av familjepopulationstotaler på följande sätt (vilket behandlas mer ingående i Avsnitt 3.2 i BR87). Låt $\mathbf{x} = \{x_i; i \in I\}$ vara en individvariabel och G en individgrupp. Definiera familjevariabeln $\mathbf{z} = \{z_d; d \in U\}$ genom;

$$z_d = \sum_{i \in d, i \in G} x_i = \sum_{i \in d} x_i \cdot 1_G(i). \quad (2.32)$$

Då gäller att individgruppstotalen $\tau(\mathbf{x}; G) =$ familjepopulationstotalen $\tau(\mathbf{z})$. Estimationen baseras sedan på naturligt sätt på denna relation. I termer av observationerna i individregistret fås

följande estimationsformel för en individgruppstotal, där X_i är det för individ i observerade värdet på individvariabel x ;

$$\hat{\tau}(x;G)_p = \sum_{i \in \text{Panel } p \text{ i } \mathcal{G}_i} X_i \cdot \mathbf{1}_G(i) \cdot IW_i. \quad (2.33)$$

I Avsnitt 4 i BR87 beskrivs hur variansskattningar för skattade värden på statistiska storheter för grupper av individer kan genomföras med hjälp av programvaran SMED.

3 Estimation i HINK92

3.1 Urvalet

Tabell 1 redovisar paneler, urvalsstratan och urvalsstorlekar i HINK92. Stratumen 01, 02, 03 och 04 är undersökningens "reguljära" stratan (se t ex BR87).

Stratum	Urvalsstorlek			
	Panel 15	Panel 16	Panel 19	Panel 60
01 ("jordbrukare")	361	382	72	48
02 ("företagare")	631	460	139	142
03 ("pensionärer")	153	327	40	70
04 ("de många")	3338	3583	341	1240
Höginkomsttagare		954		
Ungdomar (18, 19, 20 år)			399	
(Dåvarande) 17-åringar	90			
Summa	4573	5706	991	1500

Tabell 1. Paneler, urvalsstratan och urvalsstorlekar i HINK92.

3.2 Punktestimation baserad på de olika panelerna

I det följande behandlas estimation först panel för panel, och därefter sammantaget.

3.2.1 Punktestimation baserad på observationerna i Panel 15

Panel 15 drogs i slutet av 1988, ursprungligen till HINK88. Urvalsram var (dåvarande) vuxen-individer i RTB[88] samt ett tilläggstratum med (dåvarande) 17-åringar. Vid återanvändningen i HINK92 skall panelen betraktas som ett urval från vuxendelen av RTB[92], vilken betecknas V . För enkelhets skull antas att V också omfattar alla individer som fanns i 1988 års urvalsram, inklusive sådana som avlidit, emigrerat, o.dyl., med uppgift om individens "RTB-öde". Den ursprungliga urvalsramen är bara en delmängd Ω av V , och dess undertäckning av RTB[92] består av följande huvuddelar;

$V(21-)$: Vuxna i RTB[92] som var för unga för att omfattas av urvalet till Panel 15. (3.1)

$\mathfrak{S}$: (1992-)vuxna som inte var bosatta i landet 1988, men ingick i RTB[92].

Dessa individer kallas fortsättningsvis *panel-15-invandrare*. (3.2)

Urvalsramen Ω kan skrivas (där $\setminus$ står för mängddifferens);

$$\Omega = V \setminus (V(21-) \cup \mathfrak{S}). \quad (3.3)$$

Med Ω enligt (3.3) antar balanseringsvillkoret (2.14) nedanstående form (3.4) - (3.6), där speciella val gjorts för $\alpha:n$ som kan väljas godtyckligt. Notera att urvalsramen för Panel 15 inte innehöll överlappande stratan.

$$\alpha_i = 1 \quad \text{om } i \text{ är ensamstående,} \quad (3.4)$$

$$\alpha_i + \alpha_{m(i)} = 1 \quad \text{om } i \text{ är sammanboende och } i \text{ och } m(i) \text{ är 21+ och ingen är invandrare,} \quad (3.5)$$

$$\alpha_i = 1 \quad \text{om } i \text{ är sammanboende och } m(i) \text{ är 18-20 år eller/och invandrare.} \quad (3.6)$$

Villkoret (3.5) ger utrymme för alternativa α -val. Det val som används i gängse HINKar kallas *gifta-proportionalviktning* (se (A2.6) i BR87). Sådan viktning rekommenderas även här, när den är tillämplig, och den åberopas som "*vanlig HINK-vikt*", liksom andra α -val som överensstämmer med dem i gängse HINK-förfarande. Därigenom kommer man till följande α -vikter;

$$\alpha_i = 1 (= \text{"vanlig HINK-vikt"}) \quad i \in V^E, \quad (3.7)$$

$$\alpha_i = \text{"vanlig HINK-vikt"} \quad i \in V^D(21+) \cap \mathcal{S}^c, \quad m(i) \in V(21+) \cap \mathcal{S}^c \quad (3.8)$$

$$\alpha_i = 1 \quad i \in V^D(21+) \cap \mathcal{S}^c, \quad m(i) \in V(21-) \cup \mathcal{S}. \quad (3.9)$$

Låt $T(\mathbf{x};\alpha)_{15}$ och $T(\mathbf{x};G;\alpha)_{15}$ vara Panel-15-estimatorer enligt (2.25) och (2.28) med ovanstående α -val. Från (2.16) och (2.6) följer att nedanstående relationer (3.10) och (3.11) gäller (för beteckningar se även (2.18));

$$M[T(\mathbf{x};\alpha)_{15}] = \tau(\mathbf{x}) - \{\tau(\mathbf{x}; E([21-] \cup \mathcal{S})) + \tau(\mathbf{x}; D([21-] \cup \mathcal{S}))\}. \quad (3.10)$$

$$M[T(\mathbf{x};G;\alpha)_{15}] = \tau(\mathbf{x};G) - \{\tau(\mathbf{x}; E([21-] \cup \mathcal{S}) \cap G) + \tau(\mathbf{x}; D([21-] \cup \mathcal{S}))\}. \quad (3.11)$$

3.2.2 Punktestimation baserad på observationerna i Panel 16

Det "ogängse" höginkomsttagarstratumet i Panel 16 var ett gängse stratum i så måtto att det inte överlappade andra stratan. Det som sägs i Avsnitt 3.2.1 är därför tillämpligt också för Panel 16, bara en mindre modifikation görs. Panel 16 drogs med RTB[89] som urvalsram, och *panel-16-invandrare* definieras därför på följande sätt ;

$$\mathcal{S}^* = (1992-)\text{vuxna som inte var bosatta i landet 1989, men ingick i RTB[92]}. \quad (3.12)$$

Med α som i (3.7) - (3.9) med $\mathcal{S}$ bytt mot $\mathcal{S}^*$ fås att följande relationer gäller för Panel-16-estimatorer $T(\mathbf{x};\alpha)_{16}$ och $T(\mathbf{x};G;\alpha)_{16}$ enligt (2.25) och (2.28);

$$M[T(\mathbf{x};\alpha)_{16}] = \tau(\mathbf{x}) - \{\tau(\mathbf{x}; E([21-] \cup \mathcal{S}^*)) + \tau(\mathbf{x}; D([21-] \cup \mathcal{S}^*))\}. \quad (3.13)$$

$$M[T(\mathbf{x};G;\alpha)_{16}] = \tau(\mathbf{x};G) - \{\tau(\mathbf{x}; E([21-] \cup \mathcal{S}^*) \cap G) + \tau(\mathbf{x}; D([21-] \cup \mathcal{S}^*) \cap G)\}. \quad (3.14)$$

3.2.3 Punktestimation baserad på observationerna i Panel 19

Panel 19 är gängse i så måtto att den drogs vid "rätt tidpunkt", dvs med de vuxna i RTB[92] som urvalsram. Den innehåller dock det irreguljära inslaget att ungdomsstratumet (med 18-20-åringar) överlappar andra stratan. Det som sägs i Avsnitt 2.4.2 är tillämpligt på Panel 19.

Balanseringsvillkoret (2.21) för α -vikterna ger som vanligt utrymme för viss valfrihet, och i (3.15) - (3.23) anges möjliga val. Därvid har i första hand eftersträfvats att α -vikterna skall kunna beräknas någorlunda enkelt, men även efter att vikterna skall vara förhållandevis effektiva vad gäller att uppnå liten estimatorvarians. För effektivitetsöverbäganden hänvisas till Avsnitt 6 i Rosén (1987b). För de individer i som förekommer i två stratan, i ungdomsstratumet A_u och ett annat, betecknas vikterna med $\alpha_i(u)$ respektive $\alpha_i(r)$, för övriga individer med bara α_i . A_4 betecknar stratum 04 i Tabell 1 ("de många").

$$\alpha_i = 1 (= \text{"vanlig HINK-vikt"}) \quad i \in V^E(21+), \quad (3.15)$$

$$\alpha_i(u) = \frac{\frac{n(A_u)}{N(A_u)}}{\frac{n(A_u)}{N(A_u)} + \frac{n(A_4)}{N(A_4)}} \quad i \in V^E(21-), \quad (3.16)$$

$$\alpha_i(r) = \frac{\frac{n(A_4)}{N(A_4)}}{\frac{n(A_u)}{N(A_u)} + \frac{n(A_4)}{N(A_4)}} \quad i \in V^E(21-), \quad (3.17)$$

$$\alpha_i = \text{"vanlig HINK-vikt"} \quad i \in V^D(21+), \quad m(i) \in V(21+), \quad (3.18)$$

$$\alpha_i(u) = \frac{\frac{n(A_u)}{N(A_u)}}{\frac{n(A_u)}{N(A_u)} + 2 \cdot \frac{n(A_4)}{N(A_4)}} \quad i \in V^D(21-), \quad m(i) \in V(21+), \quad (3.19)$$

$$\alpha_i(r) = \frac{\frac{n(A_4)}{N(A_4)}}{\frac{n(A_u)}{N(A_u)} + 2 \cdot \frac{n(A_4)}{N(A_4)}} \quad i \in V^D(21-), \quad m(i) \in V(21+), \quad (3.20)$$

$$\alpha_i = \frac{\frac{n(A_4)}{N(A_4)}}{\frac{n(A_u)}{N(A_u)} + 2 \cdot \frac{n(A_4)}{N(A_4)}} \quad i \in V^D(21+), \quad m(i) \in V(21-), \quad (3.21)$$

$$\alpha_i(u) = \frac{1}{2} \cdot \frac{\frac{n(A_u)}{N(A_u)}}{\frac{n(A_u)}{N(A_u)} + \frac{n(A_4)}{N(A_4)}} \quad i \in V^D(21-), \quad m(i) \in V(21-), \quad (3.22)$$

$$\alpha_i(r) = \frac{1}{2} \cdot \frac{\frac{n(A_4)}{N(A_4)}}{\frac{n(A_u)}{N(A_u)} + \frac{n(A_4)}{N(A_4)}} \quad i \in V^D(21-), \quad m(i) \in V(21-). \quad (3.23)$$

Låt $T(x; \alpha)_{16}$ och $T(x; G; \alpha)_{16}$ vara Panel-16-estimatorer enligt (2.25) och (2.28) med α -val enligt (3.15) - (3.23). Från (2.22) och (2.6) fås att följande relationer gäller;

$$M[T(x; \alpha)_{19}] = \tau(x), \quad (3.24)$$

$$M[T(x; G; \alpha)_{19}] = \tau(x; G). \quad (3.25)$$

3.2.4 Punktestimation baserad på observationerna i Panel 60

Panel 60 är ett delurval från HUT92 som drogs på följande sätt. Det totala urvalet till HUT92 omfattade cirka 6000 individer dragna som ett systematiskt urval, med urvalsramen ordnade efter ålder, från ålderssegmentet 0 - 74 år i RTB[92]. HUT-urvalet stratifierades enligt HINKs standardstratan (= stratan 01, 02, 03 och 04 i Tabell 1), och OSU-urval drogs från stratana med urvalsstorlekar enligt Tabell 1. Enligt principen att inferens bör göras betingat med avseende på ancillära storheter, betraktas dessa delurval i estimationssteget som självvägande urval med givna urvalsstorlekar från de fyra stratana 01, 02, 03 respektive åldersgruppen 65 - 74 år. Antal individer i det senare stratamet var 794 985. Vidare, vid variansskattningar förfars som om delurvalen vore OSU-urval, vilket innebär viss (men obetydlig) approximation med tanke på att HUT-urvalet drogs som ett systematiskt urval. (Approximationen drar åt hållet att estimatorvarianser överskattas.) Med dett nyssnämnda som bakgrund utförs estimationen från Panel-60-observationerna under premisen att Panel-60-urvalet har probabilistiska egenskaper som ett urval av HINK-typ. Panel 60 är dock förenad med samma irregularitet som Panelerna 15

och 16, nämligen att urvalspersonerna är dragna från en delmängd av RTB[92]. För Panel 60 är den aktuella delmängden;

$$\Omega = V(75-) = \text{individer med (1992-)ålder 18 - 74 år.} \quad (3.26)$$

Rättfram modifikation av det som sägs i Avsnitt 3.2.1 ger följande. Med α -vikter enligt (3.27) - (3.29) gäller (3.30) och (3.31).

$$\alpha_i = 1 (= \text{"vanlig HINK-vikt"}), \quad \text{om } i \in V^E, \quad (3.27)$$

$$\alpha_i = \text{"vanlig HINK-vikt"} \quad \text{om } i \in V^D(75-), \quad m(i) \in V(75-), \quad (3.28)$$

$$\alpha_i = 1 \quad \text{om } i \in V^D(75-), \quad m(i) \in V(75+), \quad (3.29)$$

$$M[T(x; \alpha)_{60}] = \tau(x) - \{\tau(x; E(75+)) + \tau(x; D(75+))\}, \quad (3.30)$$

$$M[T(x; G; \alpha)_{60}] = \tau(x; G) - \{\tau(x; E(75+) \cap G) + \tau(x; D(75+) \cap G)\}. \quad (3.31)$$

3.3 Punktestimation baserad på observationerna i samtliga paneler

3.3.1 Utgångspunkt

Vid estimation på basis av HINK92-materialet skall naturligtvis samtliga observationer användas. Detta görs genom att på lämpligt sätt väga samman skattningar baserade på observationerna i de olika panelerna.

Låt γ_{15} , γ_{16} , γ_{19} och γ_{60} vara tal som satisfierar;

$$\gamma_{15} + \gamma_{16} + \gamma_{19} + \gamma_{60} = 1, \quad (3.32)$$

och betrakta nedanstående statistika, där T definieras enligt (2.25);

$$\gamma_{15} \cdot T(x; \alpha)_{15} + \gamma_{16} \cdot T(x; \alpha)_{16} + \gamma_{19} \cdot T(x; \alpha)_{19} + \gamma_{60} \cdot T(x; \alpha)_{60}. \quad (3.33)$$

Formlerna (3.10), (3.13), (3.24) och (3.30) ger;

$$\begin{aligned} M[\gamma_{15} \cdot T(x; \alpha)_{15} + \gamma_{16} \cdot T(x; \alpha)_{16} + \gamma_{19} \cdot T(x; \alpha)_{19} + \gamma_{60} \cdot T(x; \alpha)_{60}] = \\ = \tau(x) - \gamma_{15} \cdot [\tau(x; E([21-] \cup \mathfrak{S})) + \tau(x; D([21-] \cup \mathfrak{S}))] - \gamma_{16} \cdot [\tau(x; E([21-] \cup \mathfrak{S}^*)) + \\ + \tau(x; D([21-] \cup \mathfrak{S}^*))] - \gamma_{60} \cdot [\tau(x; E(75+) + \tau(x; D(75+))]. \end{aligned} \quad (3.34)$$

Från (3.34) framgår att statistikan (3.33) inte ger väntevärdesriktig estimation av populationstotalen $\tau(x)$. Skevheten beror av nedanstående storheter;

$$\tau(x; E([21-] \cup \mathfrak{S})) \text{ och } \tau(x; D([21-] \cup \mathfrak{S})), \quad (3.35)$$

$$\tau(x; E([21-] \cup \mathfrak{S}^*)) \text{ och } \tau(x; D([21-] \cup \mathfrak{S}^*)), \quad (3.36)$$

$$\tau(x; E(75+)) \text{ och } \tau(x; D(75+)). \quad (3.37)$$

Estimatoren i (3.33) kan dock justeras genom att storheterna i (3.35) - (3.37) skattas, och "elimineras" ur (3.33).

3.3.2 Underlag för justeringar

Justeringsskattningar med hjälp av Panel 19

Från (3.25) fås nedanstående relationer, vilka utsäger att T-estimatoren på vänstra sidan ger väntevärdesriktig estimation av storheten på högra sidan. B betecknar en godtycklig delmängd av familjepopulationen, som kommer att specialiseras längre fram.

$$M[T(x; E([21-] \cup \mathfrak{S}) \cap B; \alpha)_{19}] = \tau(x; E([21-] \cup \mathfrak{S}) \cap B), \quad (3.38)$$

$$M[T(x; D([21-] \cup \mathfrak{S}) \cap B; \alpha)_{19}] = \tau(x; D([21-] \cup \mathfrak{S}) \cap B), \quad (3.39)$$

$$M[T(x; E([21-] \cup \mathfrak{S}^*) \cap B; \alpha)_{19}] = \tau(x; E([21-] \cup \mathfrak{S}^*) \cap B), \quad (3.40)$$

$$M[T(x; D([21-] \cup \mathfrak{S}^*) \cap B; \alpha)_{19}] = \tau(x; D([21-] \cup \mathfrak{S}^*) \cap B), \quad (3.41)$$

$$M[T(\mathbf{x};E(75+);\alpha)_{19}] = \tau(\mathbf{x};E(75+)), \quad (3.42)$$

$$M[T(\mathbf{x};D(75+);\alpha)_{19}] = \tau(\mathbf{x};D(75+)). \quad (3.43)$$

Justeringskattningar med hjälp av Panel 60

Från (3.31) fås nedanstående relationer, som utsäger att T-estimatorn på vänstra sidan ger väntevärdesriktig estimation av storheten på högra sidan.

$$M[T(\mathbf{x};E([21-]\cup\mathfrak{I});\alpha)_{60}] = \tau(\mathbf{x};E([21-]\cup\mathfrak{I})) - \tau(\mathbf{x};E([75+]\cap\mathfrak{I})), \quad (3.44)$$

$$M[T(\mathbf{x};D([21-]\cup\mathfrak{I});\alpha)_{60}] = \tau(\mathbf{x};D([21-]\cup\mathfrak{I})) - \tau(\mathbf{x};D([75+]\cap\mathfrak{I})). \quad (3.45)$$

Genom att bortse från bidrag till totaler som kommer från envuxenfamiljer med invandrare som är 75+ år samt från tvåvuxenfamiljer där båda vuxna är invandrare och 75+ år, fås följande enklare, men något approximativa relationer (3.46) - (3.49). I de relationerna görs också restriktion till en godtycklig delmängd B av familjepopulationen.

$$M[T(\mathbf{x};E([21-]\cup\mathfrak{I})\cap B;\alpha)_{60}] \approx \tau(\mathbf{x};E([21-]\cup\mathfrak{I})\cap B), \quad (3.46)$$

$$M[T(\mathbf{x};D([21-]\cup\mathfrak{I})\cap B;\alpha)_{60}] \approx \tau(\mathbf{x};D([21-]\cup\mathfrak{I})\cap B). \quad (3.47)$$

Helt analogt fås följande relationer;

$$M[T(\mathbf{x};E([21-]\cup\mathfrak{I}^*)\cap B;\alpha)_{60}] \approx \tau(\mathbf{x};E([21-]\cup\mathfrak{I}^*)\cap B), \quad (3.48)$$

$$M[T(\mathbf{x};D([21-]\cup\mathfrak{I}^*)\cap B;\alpha)_{60}] \approx \tau(\mathbf{x};D([21-]\cup\mathfrak{I}^*)\cap B). \quad (3.49)$$

Justeringskattningar med hjälp av Panelerna 15 och 16

Med argument som är analogt till dem för Panel 60 fås följande. Om man bortser från bidrag till totaler som kommer från envuxenfamiljer med invandrare som är 75+ år samt från tvåvuxenfamiljer där båda vuxna är invandrare som är 75+ år samt från tvåvuxenfamiljer där den ena kontrahenten är 21- år och den andra 75+, gäller följande approximativa relationer;

$$M[T(\mathbf{x};E(75+);\alpha)_{15}] \approx \tau(\mathbf{x};E(75+)), \quad (3.50)$$

$$M[T(\mathbf{x};D(75+);\alpha)_{15}] \approx \tau(\mathbf{x};D(75+)), \quad (3.51)$$

$$M[T(\mathbf{x};E(75+);\alpha)_{16}] \approx \tau(\mathbf{x};E(75+)), \quad (3.52)$$

$$M[T(\mathbf{x};D(75+);\alpha)_{16}] \approx \tau(\mathbf{x};D(75+)). \quad (3.53)$$

Kombinerade justeringskattningar

Låt λ_{15} , λ_{16} och λ_{19} vara tal som satisfierar

$$\lambda_{15} + \lambda_{16} + \lambda_{19} = 1. \quad (3.54)$$

Från (3.42), (3.43), (3.50) - (3.53) och (3.54) fås relationerna (3.55) och (3.56) nedan, vilka ger bakgrund för att justera för skevhetsskomponenterna i (3.37);

$$M[\lambda_{15} \cdot T(\mathbf{x};E(75+);\alpha)_{15} + \lambda_{16} \cdot T(\mathbf{x};E(75+);\alpha)_{16} + \lambda_{19} \cdot T(\mathbf{x};E(75+);\alpha)_{19}] \approx \tau(\mathbf{x};E(75+)), \quad (3.55)$$

$$M[\lambda_{15} \cdot T(\mathbf{x};D(75+);\alpha)_{15} + \lambda_{16} \cdot T(\mathbf{x};D(75+);\alpha)_{16} + \lambda_{19} \cdot T(\mathbf{x};D(75+);\alpha)_{19}] \approx \tau(\mathbf{x};D(75+)). \quad (3.56)$$

För att justera för skevhetsskomponenterna i (3.35) och (3.36) kombineras skattningar från Panelerna 19 och 60. Härvid förfars på sätt som kan förefalla litet komplicerat, med uppdelning efter delmängden Q av familjepopulationen enligt (3.57) nedan. Skälet för uppdelningen efter Q är att Panel 19 valdes med koncentration på urvalspersoner i åldersgrupperna 18 - 20 år, medan urvalspersonerna i Panel 60 sprider jämt över alla åldrar upp till 75 år. Därför bör man fästa olika relativ vikt vid Panelerna 19 och 60 beroende på om en skattning gäller "unga familjer" eller ej. Denna viktningsfråga behandlas vidare i Avsnitt 3.4.2.

$$Q = \text{familjer (med en såväl som två vuxna) där minst en vuxen är 18 - 20 år.} \quad (3.57)$$

Följande typ av uppdelning av en grupp-G-total gäller för en allmän delmängder B av familjepopulationen (B^c står för komplementet till B);

$$\tau(\mathbf{x};G) = \tau(\mathbf{x};G \cap B) + \tau(\mathbf{x};G \cap B^c). \quad (3.58)$$

Låt κ_{19} och κ_{60} samt ρ_{19} och ρ_{60} vara tal som satisfierar;

$$\kappa_{19} + \kappa_{60} = 1, \quad \rho_{19} + \rho_{60} = 1. \quad (3.59)$$

Från (3.38) - (3.41) och (3.46) - (3.49) med $B = Q$, samt (3.58) fås nedanstående formler, som kommer till användning litet längre fram;

$$M[\kappa_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q; \alpha)_{60} + \rho_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{60}] \approx \tau(\mathbf{x};E([21-] \cup \mathcal{S})), \quad (3.60)$$

$$M[\kappa_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q; \alpha)_{60} + \rho_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{60}] \approx \tau(\mathbf{x};D([21-] \cup \mathcal{S})), \quad (3.61)$$

$$M[\kappa_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{60} + \rho_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{60}] \approx \tau(\mathbf{x};E([21-] \cup \mathcal{S}^*)), \quad (3.62)$$

$$M[\kappa_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{60} + \rho_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{60}] \approx \tau(\mathbf{x};D([21-] \cup \mathcal{S}^*)). \quad (3.63)$$

3.3.3 Sammansatta skattningar

Kombination av (3.34), (3.55), (3.56) samt (3.60) - (3.63) ger följande (approximativt) väntevärdesriktiga estimator för en populationstotal $\tau(\mathbf{x})$;

$$\begin{aligned} \hat{\tau}(\mathbf{x}) = & \gamma_{15} \cdot T(\mathbf{x}; \alpha)_{15} + \gamma_{16} \cdot T(\mathbf{x}; \alpha)_{16} + \gamma_{19} \cdot T(\mathbf{x}; \alpha)_{19} + \gamma_{60} \cdot T(\mathbf{x}; \alpha)_{60} + \\ & + \gamma_{15} \cdot [\kappa_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q; \alpha)_{60} + \\ & + \rho_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{60} + \\ & + \kappa_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q; \alpha)_{60} + \\ & + \rho_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}) \cap Q^c; \alpha)_{60}] + \\ & + \gamma_{16} \cdot [\kappa_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{60} + \\ & + \rho_{19} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};E([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{60} + \\ & + \kappa_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{19} + \kappa_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q; \alpha)_{60} + \\ & + \rho_{19} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{19} + \rho_{60} \cdot T(\mathbf{x};D([21-] \cup \mathcal{S}^*) \cap Q^c; \alpha)_{60}] + \\ & + \gamma_{60} \cdot [\lambda_{15} \cdot T(\mathbf{x};E(75+); \alpha)_{15} + \lambda_{16} \cdot T(\mathbf{x};E(75+); \alpha)_{16} + \lambda_{19} \cdot T(\mathbf{x};E(75+); \alpha)_{19} + \\ & + \lambda_{15} \cdot T(\mathbf{x};D(75+); \alpha)_{15} + \lambda_{16} \cdot T(\mathbf{x};D(75+); \alpha)_{16} + \lambda_{19} \cdot T(\mathbf{x};D(75+); \alpha)_{19}]. \end{aligned} \quad (3.64)$$

Formeln (3.64) anvisar väntevärdesriktig estimation av en populationstotal, men den är ganska formidabel. Nästa uppgift är att uttrycka (3.64) enklare. Från (2.28) framgår att (3.64) kan skrivas på följande sätt, med α^* -vikter enligt (3.66) nedan;

$$\hat{\tau}(\mathbf{x}) = \sum_{p=15,16,19,60} \sum_{h=1}^{H_p} \frac{N_{ph}}{n(s)_{ph}} \sum_{i \in A_{ph}} x_{d(i)} \cdot \alpha_i^*(h;p) \cdot I_i. \quad (3.65)$$

I (3.66) utelämnas stratumindex h , och $\alpha(p)_i$ står för α_i - vikten i Panel p om urvalsperson i ingår i den panelen, och för 0 annars. Vidare används indikatorbeteckningen i i (2.5).

$$\begin{aligned}
\alpha_i^*(p) = & \gamma_{15} \cdot \alpha(15)_i + \gamma_{16} \cdot \alpha(16)_i + \gamma_{19} \cdot \alpha(19)_i + \gamma_{60} \cdot \alpha(60)_i + \\
& + \gamma_{15} \cdot [\kappa_{19} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i + \\
& + \kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i] + \\
& + \gamma_{16} \cdot [\kappa_{19} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); E([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i + \\
& + \kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i] + \\
& + \gamma_{60} \cdot [\lambda_{15} \cdot \mathbf{1}(d(i); E(75+)) \cdot \alpha(15)_i + \lambda_{16} \cdot \mathbf{1}(d(i); E(75+)) \cdot \alpha(16)_i + \lambda_{19} \cdot \mathbf{1}(d(i); E(75+)) \cdot \alpha(19)_i + \\
& + \lambda_{15} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(15)_i + \lambda_{16} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(16)_i + \lambda_{19} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(19)_i]. \quad (3.66)
\end{aligned}$$

I syfte att göra formeln (3.66) mer överblickbar, samt även att för att skapa en enkel struktur för beräkning av α^* -vikterna, införs följande konvention;

Även för familjer med ensamstående vuxen betraktas vuxendelen som ett par, varvid paret erhålls genom upprepning av den vuxne: För $i \in V^E$ sätts $i(d;1) = i(d;2) = i$. (3.67)

Under (3.67) gäller att en tidigare indikator av typen $\mathbf{1}(d(i); E(\Lambda))$ kan skrivas $\mathbf{1}(d(i); D(\Lambda))$. Vidare gäller att innebörden av indikatorerna $\mathbf{1}(d(i); Q)$ och $\mathbf{1}(d(i); Q^c)$ inte ändras när (3.67) tillämpas. Som konsekvens av det nyss sagda kan (3.66) skrivas på följande enklare sätt under konventionen (3.67);

$$\begin{aligned}
\alpha_i^*(p) = & \gamma_{15} \cdot \alpha(15)_i + \gamma_{16} \cdot \alpha(16)_i + \gamma_{19} \cdot \alpha(19)_i + \gamma_{60} \cdot \alpha(60)_i + \\
& + \gamma_{15} \cdot [\kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i] + \\
& + \gamma_{16} \cdot [\kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(19)_i + \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) \cdot \alpha(60)_i + \\
& + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(19)_i + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \cdot \alpha(60)_i] + \\
& + \gamma_{60} \cdot [\lambda_{15} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(15)_i + \lambda_{16} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(16)_i + \lambda_{19} \cdot \mathbf{1}(d(i); D(75+)) \cdot \alpha(19)_i]. \quad (3.68)
\end{aligned}$$

(3.68) är ett kompakt sätt att skriva α^* -vikterna på. I nästa avsnitt uttrycks (3.68) panelvis, och blir då lättare att överblicka.

3.4 Beräkning av α^* -vikterna

3.4.1 Panelvisa formler för α^* -vikterna

α^* -vikter i Panel 15

Här förutsätts index i löpa över Panel 15, och konventionen (3.67) tillämpas. Formeln (3.68) innebär då att nedanstående relation är uppfylld, där vi för enkelhets skull skriver α_i och α_i^* istället för $\alpha(15)_i$ och $\alpha_i^*(15)$;

$$\alpha_i^* = \alpha_i \cdot [\gamma_{15} + \gamma_{60} \cdot \lambda_{15} \cdot \mathbf{1}(d(i); D(75+))]. \quad (3.69)$$

Beräkning av α_i^* -vikter enligt (3.69) kan uttryckas med nedanstående tvåstegsförfarande.

Steg 1: Beräkna α_i enligt (3.7) - (3.9).

Steg 2: Med tillämpande av konventionen (3.67), beräkna α_i^* enligt Tabell 2.

Formel för α_i^*	Används för
$\alpha_i \cdot \gamma_{15}$	$d(i) \notin D(75+)$
$\alpha_i \cdot (\gamma_{15} + \gamma_{60} \cdot \lambda_{15})$	$d(i) \in D(75+)$

Tabell 2. α^* -vikter i Panel 15.

α^* -vikter i Panel 16

Här förutsätts index i löpa över Panel 16, och konventionen (3.67) tillämpas. Formeln (3.68) innebär då nedanstående relation, med skrivsättet α_i och α_i^* istället för $\alpha(16)_i$ och $\alpha_i^*(16)$;

$$\alpha_i^* = \alpha_i \cdot [\gamma_{16} + \gamma_{60} \cdot \lambda_{16} \cdot \mathbf{1}(d(i); D(75+))]. \quad (3.70)$$

Beräkning av α_i^* -vikter enligt (3.70) kan uttryckas med nedanstående tvåstegsförfarande.

Steg 1: Beräkna α_i enligt (3.7) - (3.9) med $\mathfrak{S}$ bytt mot $\mathfrak{S}^*$.

Steg 2: Med tillämpande av konventionen (3.67), beräkna α_i^* enligt Tabell 3.

Formel för α_i^*	Används för
$\alpha_i \cdot \gamma_{16}$	$d(i) \notin D(75+)$
$\alpha_i \cdot (\gamma_{16} + \gamma_{60} \cdot \lambda_{16})$	$d(i) \in D(75+)$

Tabell 3. α^* -vikter i Panel 16.

α^* -vikter i Panel 60

Här förutsätts index i löpa över Panel 60, och konventionen (3.67) tillämpas. Formeln (3.68) innebär då nedanstående relation, med skrivsättet α_i och α_i^* istället för $\alpha(60)_i$ och $\alpha_i^*(60)$;

$$\begin{aligned} \alpha_i^* = \alpha_i \cdot [& \gamma_{60} + \gamma_{15} \cdot \{ \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q) + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}) \cdot \mathbf{1}(d(i); Q^c) \} + \\ & + \gamma_{16} \cdot \{ \kappa_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q) + \rho_{60} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*) \cdot \mathbf{1}(d(i); Q^c) \}. \end{aligned} \quad (3.71)$$

Beräkning av α_i^* -vikter enligt (3.71) kan uttryckas med nedanstående tvåstegsförfarande.

Steg 1: Beräkna α_i enligt (3.27) - (3.29).

Steg 2: Med tillämpande av konventionen (3.67), beräkna α_i^* enligt Tabell 4. Vi påminner om definitionen av mängden Q (se (3.57));

Q = familjer där minst en av $i(d;1)$ och $i(d;2)$ har åldern 18, 19 eller 20 år.

Formel för α_i^*	Används för
$\alpha_i \cdot \gamma_{60}$	$d(i) \notin D([21-] \cup \mathfrak{S})$
$\alpha_i \cdot (\gamma_{60} + \gamma_{15} \cdot \kappa_{60})$	$d(i) \in D([21-] \cup \mathfrak{S}), d(i) \notin D([21-] \cup \mathfrak{S}^*), d(i) \in Q$
$\alpha_i \cdot (\gamma_{60} + \gamma_{15} \cdot \rho_{60})$	$d(i) \in D([21-] \cup \mathfrak{S}), d(i) \notin D([21-] \cup \mathfrak{S}^*), d(i) \notin Q$
$\alpha_i \cdot (\gamma_{60} + \gamma_{15} \cdot \kappa_{60} + \gamma_{16} \cdot \kappa_{60})$	$d(i) \in D([21-] \cup \mathfrak{S}^*), d(i) \in Q$
$\alpha_i \cdot (\gamma_{60} + \gamma_{15} \cdot \rho_{60} + \gamma_{16} \cdot \rho_{60})$	$d(i) \in D([21-] \cup \mathfrak{S}^*), d(i) \notin Q$

Tabell 4. Beräkning av α^* -vikter i Panel 60.

Rättfärdigande av steget från (3.71) till Tabell 4: Man förväntar kanske att villkoret i övre högra rutan i tabellen skall ha utseendet: $d(i) \notin D([21-] \cup \mathfrak{S})$, $d(i) \notin D([21-] \cup \mathfrak{S}^*)$. Notera dock att från definitionerna (3.2) och (3.12) följer att $D([21-] \cup \mathfrak{S}) \supset D([21-] \cup \mathfrak{S}^*)$, vilket medför att det angivna villkoret är ekvivalent med det nyssnämnda. Den nämnda inklusionsrelationen förklarar också varför möjligheterna $\alpha_i \cdot (\gamma_{60} + \gamma_{16} \cdot \kappa_{60})$ och $\alpha_i \cdot (\gamma_{60} + \gamma_{16} \cdot \rho_{60})$ inte är aktuella i tabellen

α^* -vikter i Panel 19

Här förutsätts index i löpa över Panel 19, och konventionen (3.67) tillämpas. Formeln (3.68) innebär då nedanstående relation, med skrivsättet α_i och α_i^* istället för $\alpha(19)_i$ och $\alpha_i^*(19)$;

$$\alpha_i^* = \alpha_i \cdot [\gamma_{19} + \gamma_{15} \cdot \{\kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S})) \cdot \mathbf{1}(d(i); Q) + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S})) \cdot \mathbf{1}(d(i); Q^c)\} + \\ + \gamma_{16} \cdot \{\kappa_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*)) \cdot \mathbf{1}(d(i); Q) + \rho_{19} \cdot \mathbf{1}(d(i); D([21-] \cup \mathfrak{S}^*)) \cdot \mathbf{1}(d(i); Q^c)\} + \\ + \gamma_{60} \cdot \lambda_{19} \cdot \mathbf{1}(d(i); D(75+))]. \quad (3.72)$$

Beräkning av α_i^* -vikter enligt (3.72) kan uttryckas med nedanstående tvåstegsförfarande.

Steg 1: Beräkna α_i enligt (3.15) - (3.23).

Steg 2: Med tillämpande av konventionen (3.67), beräkna α_i^* enligt Tabell 5. Vi påminner om definitionen av mängden Q (se (3.57));

Q = familjer där minst en av $i(d;1)$ och $i(d;2)$ har åldern 18, 19 eller 20 år.

Formel för α_i^*	Används för
$\alpha_i \cdot \gamma_{19}$	$d(i) \notin D(75+)$, $d(i) \notin D([21-] \cup \mathfrak{S})$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \kappa_{19})$	$d(i) \notin D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S})$, $d(i) \notin D([21-] \cup \mathfrak{S}^*)$, $d(i) \in Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \rho_{19})$	$d(i) \notin D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S})$, $d(i) \notin D([21-] \cup \mathfrak{S}^*)$, $d(i) \notin Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \kappa_{19} + \gamma_{16} \cdot \kappa_{19})$	$d(i) \notin D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S}^*)$, $d(i) \in Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \rho_{19} + \gamma_{16} \cdot \rho_{19})$	$d(i) \notin D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S}^*)$, $d(i) \notin Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{60} \cdot \lambda_{19})$	$d(i) \in D(75+)$, $d(i) \notin D([21-] \cup \mathfrak{S})$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \kappa_{19} + \gamma_{60} \cdot \lambda_{19})$	$d(i) \in D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S})$, $d(i) \notin D([21-] \cup \mathfrak{S}^*)$, $d(i) \in Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \rho_{19} + \gamma_{60} \cdot \lambda_{19})$	$d(i) \in D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S})$, $d(i) \notin D([21-] \cup \mathfrak{S}^*)$, $d(i) \notin Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \kappa_{19} + \gamma_{16} \cdot \kappa_{19} + \gamma_{60} \cdot \lambda_{19})$	$d(i) \in D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S}^*)$, $d(i) \in Q$
$\alpha_i \cdot (\gamma_{19} + \gamma_{15} \cdot \rho_{19} + \gamma_{16} \cdot \rho_{19} + \gamma_{60} \cdot \lambda_{19})$	$d(i) \in D(75+)$, $d(i) \in D([21-] \cup \mathfrak{S}^*)$, $d(i) \notin Q$

Tabell 5. α^* -vikter i Panel 19.

Steget från (3.72) till Tabell 5 rättfärdigas med samma argument som för Panel 60.

3.4.2 Numeriska värden på sammanvägningskoefficienterna

För att α^* -vikterna enligt Tabellerna 2 - 5 skall kunna beräknas, måste man först lägga fast numeriska värden på koefficienterna γ , λ , κ och ρ under bivillkoren (3.32), (3.54) och (3.59). Uppgiften för dessa koefficienter är att väga samman estimatorer i syfte att få estimatorer med bättre precision än var och en av delestimatorerna (se (3.33), (3.55), (3.56) och (3.60) - (3.63)). Precisionsoptimal sammanvägning fås om delestimatorerna vägs omvänt proportionellt mot

sina varianser. Dessa är dock okända, och de varierar också beroende på vilken statistisk storhet som skattas. Vi stöder oss därför på det vanligtvis föreliggande förhållandet att estimatorvarianser är, åtminstone i stora drag, omvänt proportionella mot antalet observationer de är baserade på, och väljer värden på koefficienterna enligt den principen. Det bör understrykas att om nyssnämnda förutsättningar om estimatorvarianserna fallerar, så uppstår inte skevhet i en sammanvägd estimator, endast att den inte får så god precision som den skulle kunnat få.

Centrala storheter för fastläggning av γ , λ , κ och ρ är alltså antalet observationer i olika paneler och grupper. I Tabell 6 redovisas sådana, med uppdelning efter urvalspersonens ålder i de två grupperna 18 - 20 år respektive 21+ år.

Urvalspersonens ålder	Antal observerade familjer				
	Panel 15	Panel 16	Panel 19	Panel 60	Alla
18, 19 och 20 år	0	0	372	83	455
21+ år	4 533	5 654	565	1 436	12 188
Totalt	4 533	5 654	937	1 519	12 643

Tabell 6. Antal observerade hushåll i HINK92.

Ovan förda resonemang om variansers omvända proportionalitet mot antal observationer tillsammans med värdena i Tabell 6 ger bakgrund för nedanstående val av sammanvägningskoefficienterna γ , λ , κ och ρ . Kom ihåg att mängden Q infördes i syfte att dela upp familjepopulationen i "unga familjer" (minst en av makarna är 18 - 20 år) respektive "övriga familjer".

$$(\gamma_{15}, \gamma_{16}, \gamma_{19}, \gamma_{60}) \text{ väljs proportionella mot } (4533, 5654, 937, 1519), \quad (3.73)$$

$$(\lambda_{15}, \lambda_{16}, \lambda_{19}) \text{ väljs proportionella mot } (4533, 5654, 565), \quad (3.74)$$

$$(\kappa_{19}, \kappa_{60}) \text{ väljs proportionella mot } (372, 83) \text{ och } (\rho_{19}, \rho_{60}) \text{ mot } (565, 1436). \quad (3.75)$$

(3.73) - (3.75) och villkoren (3.32), (3.54) och (3.57) leder till nedanstående värden på γ , λ , κ och ρ . Värdena visar dock några (mycket små) aritmetiska avvikelser från (3.73) - (3.75), beroende på att de bestämdes med hjälp av preliminära värden i Tabell 6, men värdena i (3.76) - (3.78) är de som användes.

$$\gamma_{15} = 0.36, \gamma_{16} = 0.44, \gamma_{19} = 0.08, \gamma_{60} = 0.12, \quad (3.76)$$

$$\lambda_{15} = 0.42, \lambda_{16} = 0.53, \lambda_{19} = 0.05, \quad (3.77)$$

$$\kappa_{19} = 0.84, \kappa_{60} = 0.16, \text{ och } \rho_{19} = 0.28, \rho_{60} = 0.72. \quad (3.78)$$

En slutlig modifikation

Det finns flera goda skäl att, om möjligt, kalibrera en estimator mot en eller flera kända totaler. Sådan kalibrering gjordes för HINK92 på följande sätt. Med koefficienter γ , λ , κ och ρ enligt ovan fanns underlag för att beräkna α^* -vikter enligt Avsnitt 3.4.1. Därmed blev estimatoren (3.65) operationell liksom även dess analog för estimation av individtotaler, vilken specificeras i (3.82) i nästa avsnitt. Med ovanstående värden på γ , λ , κ och ρ blev det skattade värdet på "HINK92-befolkningen" 8 464 000. På basis av uppgifter från HINK91, HINK92 samt befolkningstal för 1990-92 gjordes bedömningen att den "sanna" HINK92-befolkningen var cirka 8 616 000. Mot den bakgrunden utfördes kalibrering så att alla koefficienter γ , λ , κ och ρ multiplicerades med faktorn $8\,616\,000/8\,464\,000 = 1,018$. I fortsättningen förutsätts denna faktor vara inkluderad i koefficienterna γ , λ , κ och ρ .

3.5 Totalt estimationsförfarande för HINK92

Därmed har härletts förfaranden för punkt- och variansskattning på basis av observationsmaterialet i HINK92, såväl för statistiska storheter som relaterar till familjepopulationen som för storheter som relaterar till individpopulationen. Nedan sammanfattas de centrala stegen i estimationen.

1. För varje panel beräknas α -vikter enligt Avsnitten 3.2.1 - 3.2.4.
2. Sammanvägningskoefficienter $(\gamma_{15}, \gamma_{16}, \gamma_{19}, \gamma_{60}), (\lambda_{15}, \lambda_{16}, \lambda_{19}), (\kappa_{19}, \kappa_{60})$ och (ρ_{19}, ρ_{60}) beräknas enligt Avsnitt 3.4.2.
3. För varje panel beräknas α^* -vikter enligt Avsnitt 3.4.1.
4. Familjevikter beräknas enligt nedan (se Avsnitt 2.5);

$$W_d = \frac{N_{h(d)}}{n(s)_{h(d)}} \cdot \alpha_{u(d)}^*(h). \quad (3.79)$$

5. Individvikter beräknas enligt nedan, vilket innebär att individvikten sätts lika med motsvarande familjevikt (se Avsnitt 2.6);

$$IW_i = W_{d(i)}. \quad (3.80)$$

Därefter kan HINKs sedvanliga skattningsförfaranden användas, där $\mathfrak{R}_F$ och $\mathfrak{R}_I$ betecknar familjeregistret respektive individregistret;

$$6. \hat{\tau}(x; G) = \sum_{d \in \mathfrak{R}_F} X_d \cdot 1_G(d) \cdot W_d \quad \text{skattar en familjegrupptotal } \tau(x; G). \quad (3.81)$$

$$7. \hat{\tau}(x; G) = \sum_{i \in \mathfrak{R}_I} X_i \cdot 1_G(i) \cdot IW_i \quad \text{skattar en individgrupptotal } \tau(x; G). \quad (3.82)$$

Tillsammans med det som sägs om variansskattningar i BR87 följer också det nedanstående;

8. Med vikter och estimatorer enligt ovan kan HINKs sedvanliga variansskattningsförfaranden användas.

Referenser

Rosén, B. (1987a). Om HINK-undersökningens estimationsförfarande. R&D Report 30, U/STM, SCB.

Rosén, B. (1987b). Some Optimality Problems When Estimating Household Data on the Basis of a "Primary" Stratified Sample of Individuals. R&D Report 36, U/STM, SCB.

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.

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.

Reports published 1993-94:

- | | |
|------------------|---|
| 1993:1
(grön) | Kvalitetsfonden 1991/92 - Projekt, handläggning och utvärdering
(Roland Friberg) |
| 1993:2
(grön) | Översyn av Undersökningen av lastbilstransporter i Sverige (UVAV)
(Bengt Rosén, Mahnaz Zamani) |
| 1993:3
(grön) | Bortfallsbarometern nr 8 (Mats Bergdahl, Pär Brundell, Jan Hörngren,
Håkan Lindén, Peter Lundquist, Monica Rennermalm) |
| 1993:4
(gul) | Guidelines on the Design and Implementation of Statistical Meta-
information Systems (Bo Sundgren) |
| 1993:5
(grön) | Kvalitetsrapporten 1993 - Utveckling av kvaliteten för SCBs
statistikproduktion (Jan Eklöf, Per Nilsson) |
| 1993:6
(grön) | Metoder att få konsistens mellan olika skattningar och samtidigt
öka precisionen (Sixten Lundström) |
| 1993:7
(grön) | Variance Estimation in the Swedish Consumer Price Index
(Jörgen Dalén, Esbjörn Ohlsson) |
| 1993:8
(grön) | Quantifying Errors in the Swedish Consumer Price Index
(Jörgen Dalén) |
| 1993:9
(gul) | Ökad tillgänglighet för statistiska data - en studie av några gränssnitt
mot statistiska databaser (Mattias Lindberg) |
| 1994:1
(gul) | Ett verksgemensamt datalager vid SCB? (Bo Sundgren) |

Tidigare utgivna **R&D Reports** kan beställas genom Ingvar Andersson, SCB, U/LEDN, 115 81 STOCKHOLM (telefon 08-783 41 47, telefax 08-783 45 99).

R&D Reports listed above as well as issues from 1988-92 can - in case they are still in stock - be ordered from Statistics Sweden, att. Ingvar Andersson U/LEDN, S-115 81 STOCKHOLM, SWEDEN (telephone nr +46 08 783 41 47, telefax nr +46 08 783 45 99).