

Om HINK-undersökningens design- och allokeringsproblematik

Bengt Rosén

SCB

R & D Report
Statistics Sweden
Research - Methods - Development
U/STM - 29

INLEDNING

TILL

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

Föregångare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

Efterföljare:

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

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

R & D Report, U/STM 1987:29. Om HINK-undersökningens design- och allokeringsproblematik / Bengt Rosén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-RnD-USTM-1987-29

Om HINK-undersökningens design- och allokeringsproblematik

Bengt Rosén

SCB

R & D Report
Statistics Sweden
Research - Methods - Development
U/STM - 29

Från trycket Februari 1987
Producent Statistiska centralbyrån, Enheten för statistiska metoder
Ansvarig utgivare Staffan Wahlström
Förfrågningar Bengt Rosén, tel. 08 7834490

© 1987, Statistiska centralbyrån
ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1987

SUMMARY

On the design and sample allocation for the survey on household incomes (HINK)

The HINK survey is conducted yearly and its main objective is to describe income conditions for different groups of households (household being determined by factual cohabitation).

In lack of a register of the units of analysis, i.e. (cohabitation) households, the following "second best" type of sampling design is employed. The national population register is used as frame for drawing a "primary", stratified sample of adults. By interviews, the households of the sampled individuals are identified. Then various income (and expenditure) data are collected for the "entire" households, and used to achieve desired estimates for groups of households.

In an "evaluation" of the HINK survey, carried out by a group from U/STM, it quite soon became clear that HINK's design (=stratification) and sample allocation could be improved a good deal, and at the same time in fact also simplified. Towards this background the report presents, in a fairly non-technical way, various suggestions and material which hopefully should be of value when deciding on more efficient stratification and sample allocation for the HINK survey.

INNEHÅLL

	Sid.
0. OM RELATIONEN MELLAN DETTA OCH TIDIGARE PM	1
I. BAKGRUND FÖR DISKUSSION AV STRATIFIERING OCH URVALSALLOKERING	2
1 Några allmänna och speciella ord om HINK-undersökningen	2
2 En formulering av design- och allokeringsproblematiken	3
3 Hittillsvarande precision i skattningarna för de socioekonomiska grupperna	4
II SÖKANDE EFTER BÄTTRE DESIGN OCH ALLOKERING	6
1 Riktade strata	6
2 Panel-11-stratifieringen	8
3 Estimationsprecision som kan uppnås med en HINK-undersökning som innehåller fyra strata enligt Panel-11-stratifieringen	14
4 Kommentarer till Tabell 10 samt fortsatt diskussion av stratifiering, allokering och skattningsprecision	20
4.1 Diverse kommentarer	20
4.2 Om precisionerna i årsskattningar på basis av en välallokerad HINK-undersökning	22
4.2.1 Vid skattning av genomsnittlig disponibel inkomst för grupper (av hushåll)	22
4.3 Om möjligheter att ytterligare förbättra HINKs stratifiering	25

OM HINK-UNDERSÖKNINGENS DESIGN- OCH ALLOKERINGSPROBLEMATIK

0. OM RELATIONEN MELLAN DETTA OCH TIDIGARE PM

Vid U/STMs "genomlysning" av HINK-undersökningen framkom rätt snart att HINKs precision bör kunna förbättras en hel del genom lämplig omläggning av undersökningens stratifiering och urvalsallokering. HINK-sektionen har framfört önskemålet att om möjligt kunna göra omläggningen redan från och med den urvalsdragning (av Panel 13) som skall genomföras i slutet av innevarande år. Eftersom produktionsomläggningar är förhållandevis tidskrävande processer så hölls 1986-07-11 ett första "beslutssammanträde" för att staka ut huvudkursen för en omläggning.

Som beslutsunderlag till det sammanträdet frantog jag ett PM 1986-07-10 med precis samma titel som det här föreliggande PM:et. I det sammanhanget anmäldes att arbetet bedrivits under viss tidsnöd, och att det därför fanns anledning att gå igenom materialet en gång till. Vidare, under sammanträdet 86-07-11, och även därefter har framkommit önskemål om ytterligare belysning av HINKs stratifierings- och allokeringsproblem samt undersökningens precisionsegenskaper. Syftet med detta PM, relativt PM 1986-07-10, är tvåfaldigt;

- att korrigera vissa felaktigheter som smög sig in i det tidigare PM:et,
- att på vissa punkter fördjupa beslutsunderlaget för omläggningen av stratifiering och urvalsallokering till Panel 13.

Vad gäller de ovan nämnda felaktigheterna i PM 1986-07-10 så hänför de sig framför allt till de tidigare (och nuvarande) Tabellerna 5 och 6 men de har också konsekvenser för de efterföljande variansberäkningarna. (Jag vill dock nämna att felaktigheterna faktiskt inte berodde på felräkningar från min sida, utan på att man gjort "oaviserade" förändringar i grunddataregistret för HINK 84, Panel 11). Hur som helst, föreliggande PM är såväl mer fullständigt som (förhoppningsvis) mer korrekt än PM 1986-07-10, och PM 1986-07-10 utgår därför och ersätts av detta PM.

I. BAKGRUND FÖR DISKUSSION AV STRATIFIERING OCH URVALSALLOKERING

1 Några allmänna och speciella ord om HINK-undersökningen

HINK-undersökningens syfte är att redovisa inkomstförhållanden för olika grupper av hushåll och även för olika grupper av individer. Redovisningen sker företrädesvis på riksnivå. En "biprodukt" av intresse är att HINK också ger information om antal hushåll av olika slag.

HINK är en urvalsundersökning. Det "primära" urvalet görs som ett stratifierat urval av (vuxna) individer. I nästa steg inhämtas önskad information avseende de hushåll som de urvalda individerna tillhör. På basis av de så erhållna observationerna skattas sedan förhållandena i riket.

Nedan ges några tämligen speciella tekniska upplysningar om HINKen, vilka har relevans i den fortsatta diskussionen.

a) I hittillsvarande HINK-undersökningar har man arbetat med bortemot tjugotalet olika strata vid stratifieringen av individpopulationen. I Appendix på sida 32 ges ett utdrag ur HINK-SM-1984 som beskriver stratifiering, urvalsallokering m m för HINK 84.

b) I HINKen använder man två likstora paneler. En urvald individ (med tillhörande hushåll) ligger med i urvalet i två år, under urvalsåret

samt efterföljande år. De två paneler som ingår i HINKen ett visst år, kalla dem Panel p respektive Panel (p+1), väljes (i allt väsentligt) oberoende av varandra. Härigenom kan följande formler för skattning av det aktuella årets värde på en populationskaraktäristiska χ användas, och används också (med aningen av modifikation),

$$\hat{\chi} = \frac{1}{2} (\hat{\chi}_p + \hat{\chi}_{p+1}),$$

$$V(\hat{\chi}) = \frac{1}{4} (V(\hat{\chi}_p) + V(\hat{\chi}_{p+1}))$$

$$\text{Medelfel } (\hat{\chi}) = \frac{1}{2} \sqrt{V(\hat{\chi}_p) + V(\hat{\chi}_{p+1})}.$$

Av skäl som kommer att framgå finner vi det lämpligt att, åtminstone till att börja med, redovisa skattningsprecision per panel istället för per år. Omräkning till årsvärden sker sedan med hjälp av ovanstående formler.

- c) HINKen har hittills omfattat (cirka) 10 000 observationer per år, 5 000 per panel.
- d) I HINKens skattningssammanhang är olika estimatorer tänkbara, särskilt ett par som tidigare har gått under benämningarna (bl a) Estimator 1 och Estimator 2. Dessa kommer här att kallas för den 1/2-viktade respektive den (urvalssannolikhets)proportionalviktade estimatorn.

2 En formulering av design- och allokeringsproblematiken

När man vill föra en diskussion om huruvida HINKen är bra eller dåligt designad och allokerad, och hur undersökningen eventuellt skulle kunna förbättras så möter man problemet att HINK-undersökningen har flera syften, särskilt föreligger flera intressanta indelningar i redovisningsgrupper. Man kan inte designa och allokerar optimalt för alla tänkbara uppsättningar av redovisningsgrupper, hur som helst så kan man

inte diskutera alla tänkbara indelningar samtidigt. För att få någorlunda klarhet och enkelhet i resonemangen skall vi därför, åtminstone i första omgången, hålla oss till en viss bestämd redovisningssituation. Anledningen till att just denna situation väljs är att den, efter diskussioner med olika "HINK-intressenter", framstår som den mest angelägna bland de möjligheter HINKen erbjuder. Den aktuella redovisningssituationen är följande.

För hushåll (till skillnad från individer), indelade efter socioekonomisk gruppering vill man skatta genomsnittlig disponibel inkomst för de olika grupperna. Av intresse är även antalet hushåll i de olika redovisningsgrupperna.

Vid socioekonomisk indelning arbetar vi med nedanstående grupper.

Nr	Redovisningsgruppens benämning
1	Ej facklärd arbetare
2	Facklärd arbetare
3	Lägre tjänstemän
4	Tjänstemän på mellannivå
5	Högre tjänstemän
6	Företagare
7	Jordbrukare
8	Övriga (pensionärer, ekon inaktiva och oklassificerade)

I fortsättningen kommer vi i första omgången att hålla oss till den ovan specificerade situationen. Det problem vi ställer oss är att välja design och allokering så att man får så god som möjligt precision i de aktuella skattningarna.

3	Hittillsvarande precision i skattningarna för de socioekonomiska grupperna
---	--

I HINK-genomlysningsarbetet har framtagits diverse empiriskt material från HINK82, HINK83 och HINK84, vilket ligger till grund för nedanstående Tabell 1. Vid medelvärdesbildning över olika år har gjorts "infla-

tionsuppräkning" med 9 % och 8 % mellan åren 1982 och 1983 respektive mellan 1983 och 1984. Av skäl som kommer att framgå separatredovisas Panel 11 (vilken ingår i HINK84).

Tabell 1

Redo- vis- nings- grupp	Varians för medelvärdesskattning (per panel) vid proportionalvik- tad estimator. Sort: 1984 års Kkr.		Varians för medelvär- desskattning (per pa- nel) vid ett (hypote- tiskt) OSU-urval ur individpopulationen. Sort: 1984 års Kkr. Genomsnitt över Panelerna 8-11
	Genomsnitt över Panelerna 8(82), 9(82), 9(83), 10(83) och 10(84)	I Panel 11	
1	4.33	4.38	1.66
2	6.97	9.52	2.59
3	11.38	10.43	4.78
4	9.84	16.83	4.69
5	22.90	18.20	14.56 (11.84)*
6	12.72	15.77	18.91
7	22.86 (11.37)*	3.89	26.30 (21.07)*
8	0.88	1.08	0.49

* Inom parentesen står värde som erhålles om man tillåter sig att rensa bort "outliervärde".

Efter att ha tänkt över materialet i Tabell 1 gör man tämligen omgående följande reflexion. De skattningsprecisioner som erhöles i HINK82-84 ser inte alldeles upplyftande ut. Den dåvarande kombinationen av design + allokering gav endast tämligen måttliga förbättringar i förhållande till ett OSU-urval för redovisningsgrupperna 6 och 7 medan skattningsprecisionen för övriga redovisningsgrupper är klart sämre än vad ett OSU-urval skulle ha gett. Detta måste tas som ett starkt indicium på att HINK82-84 var i första hand klart dåligt allokerad, och i andra hand eventuellt även dåligt designad.

II SÖKANDE EFTER BÄTTRE DESIGN OCH ALLOKERING

Ingen vet idag riktigt vilka överväganden som en gång i världen gjordes när HINKens design och allokering utformades (och som i sina huvuddrag består ännu). Det som dock kan konstateras (från t ex Tabell 1) är att (åtminstone) urvalsallokeringen är klart dålig. Mot denna bakgrund finns inte anledning att rota ner sig i detaljer i den nuvarande stratifieringen, utan det är naturligare att försöka bygga upp en ändamålsenlig stratifiering "från scratch".

1 Riktade strata

Från Tabell 1 framgår att ett OSU-urval (och även hittillsvarande HINK-urval) ger mycket olika precision i skattningarna av medelvärdena i de olika redovisningsgrupperna; stora varianser fås för grupperna 6 och 7 (företagare respektive jordbrukare) samt liten varians för grupp 8 (pensionärer m fl). Detta förhållande är framför allt en återspeglning av att 6 och 7 är små redovisningsgrupper medan 8 är stor. Det finns goda skäl för att eftersträva bättre balans mellan skattningsprecisionerna för de olika redovisningsgrupperna (åtminstone under förutsättningen att den aktuella gruppindelningen betraktas som viktig). En naturlig, och god idé är då att använda det som kan kallas riktade strata, och som uppenbarligen är en av de idéer som finns med bakom HINKens nuvarande uppsättning av strata.

I Rosén (1986) redogörs dels för några allmänna tankegångar i samband med riktade strata dels ges diverse mer tekniska resultat angående riktade strata i sammanhang av HINK-typ. Nedan redovisas några huvudpunkter.

- Ett stratum A i (vuxen)individentpopulationen säges vara riktat mot redovisningsgruppen G (av hushåll) om en individ ur A med någorlunda god säkerhet ingår i ett hushåll i G. Riktandet kan göras med olika grad av "träffsäkerhet". Bom av typ I säges föreligga om ett hushåll i G ej kan nås från A, dvs ingen av de (en eller två) vuxna i hushållet tillhör stratum A. (Bom-I-hushållen utgör "undertäckning" vid

riktandet.) Bom av typ II säges föreligga när en individ i A ingår i ett hushåll utanför G. (Bom-II-hushållen kan uppfattas som "övertäckning" vid riktandet.)

Riktade stratan kan åstadkomma bättre balans i estimationsprecisionerna mellan "små" och "stora" redovisningsgrupper genom att man samlar högt (= med högre urvalsfraktion än den genomsnittliga) i stratan riktade mot små redovisningsgrupper och samlar lågt (= med lägre urvalsfraktion än den genomsnittliga) i stratan riktade mot stora redovisningsgrupper.

För att riktade strata skall ge åsyftad effekt för små redovisningsgrupper är det väsentligt att de kan riktas så att antalet bommar av typ I blir lågt (= man får liten "undertäckning" av de redovisningsgrupper man riktar mot). Det är också av värde, men avsevärt mindre väsentligt, att antalet bommar av typ II ("övertäckningen") blir litet.

- För att kunna utnyttja den fulla potentialen i strata som man verkligen lyckats rikta någorlunda väl mot förhållandevis små redovisningsgrupper, är det väsentligt att använda proportionalviktade estimerer.

I det följande skall vi i första omgången intressera oss för strata som riktas dels mot de små redovisningsgrupperna 6 och 7 dvs "företagare" och "jordbrukare" samt mot den stora redovisningsgruppen 8 dvs företrädesvis "pensionärer". Nedan införs numreringar och benämningar av de aktuella tre riktade stratana i individpopulationen samt "reststratumet".

Tabell 2

<u>Stratum:</u>	<u>Nr</u>	<u>Benämning</u>
	1	LBR-Jordbrukare
	2	Förmodade företagare (och jordbrukare)
	3	Ålderspensionärer
	4	"Löntagare"

2 Panel-11-stratifieringen

I HINK har man successivt gjort diverse (oftast smärre) ändringar i stratumdefinitionerna. Från och med Panell 11 gjordes en förhållandevis omfattande ändring, en ändring som berör just definitionerna av de strata som nämns i Tabell 2. I den nuvarande HINK-utformningen utgör stratifieringen enligt Tabell 2 dock endast ett första steg på vägen mot den fullständiga "fin"-stratifieringen (vilken som sagt omfattar bortemot tjugotalet strata). I det fortsatta resonemanget skall vi emellertid, åtminstone till att börja med, bortse från denna efterföljande fin-stratifiering. Med Panel-11-stratifieringen menar vi i fortsättningen den stratifiering enligt Tabell 2 som erhålles med hjälp av nedanstående stratumdefinitioner.

Tabell 3 Panel-11-stratifieringen

<u>Stratum</u>	<u>Definition</u>
1	Individer som finns med i Lantbruksregistret (LBR).
2	Individer utanför Stratum 1, vilka enligt SCBTB (för "två år innan") satisfierar kriteriet att inkomst av rörelse eller av jordbruk > 50 % av arbetsinkomst.
3	Individer utanför Stratana 1 och 2 som är $\geq$ 65 år.
4	Resten (bland de vuxna individerna).

I nedanstående Tabell 4 anges hur individpopulationen vid Panel-11-stratifiering delar upp sig antalsmässigt på de fyra stratana i Tabell 2.

Tabell 4

<u>Stratum</u>	<u>Antal individer</u>
1	88 000
2	293 000
3	1 334 000
4	<u>4 785 000</u>
	6 500 000

Från nedanstående Tabell 5 framgår att Panel-11-stratifieringen var lyckosam såtillvida att antalet bommar av typ I reduceras avsevärt för redovisningsgruppen "jordbrukare" och träffförhållandena förbättras även för redovisningsgruppen "företagare". Vi finner därför inte anledning att gå in på detaljer i tidigare definitioner av stratana i Tabell 2. De förbättrade träffförhållandena vad gäller "jordbrukare" återspeglar sig också i Tabell 1, där variansen för jordbrukargruppen sjunker avsevärt i Panel 11. (Siffran skall dock tagas med en nypa salt, eftersom det handlar om ett skattat värde som är behäftat med åtskillig slump-osäkerhet.)

I Tabellerna 5 och 6 används följande beteckningar,

g_h = antalet envuxen-hushåll i den aktuella redovisningsgruppen, där den vuxne tillhör stratum h , $h=1,2,3,4$

g_{hl} = antalet tvåvuxen-hushåll i den aktuella redovisningsgruppen, där den ene vuxne tillhör stratum h och den andre vuxne stratum l , $1 \leq h \leq l \leq 4$.

Det bör nämnas att för materialet bakom Tabellerna 5-8 är den faktiska stratumtillhörighetsklassificeringen litet "oegentlig" för sammanboende i de fall då hushållets verkliga sammansättning skiljer sig från dess "RTB-sammansättning". (I sådana fall ges moatjén samma stratumtillhörighet som urvalsindividen.) Från denna "komplikation" bortses i fortsättningen.

Tabell 5 Skattade värden för hur landets jordbrukar- och företagarkushåll fördelar sig på stratumtillhörighetsgrupper vid stratifiering enligt Tabell 2 med Panel-11- resp tidigare stratumdefinitioner.

	Redovisningsgrupp				
	Jordbrukare			Företagare	
	Stratumdefinitioner enligt			Stratumdefinitioner enligt	
	Panel-11	Tidigare		Panel-11	Tidigare
g ₁	17 700	11 800		100	550
g ₂	8 600	750		25 000	25 000
g ₃	0	6 000		0	3 000
g ₄	150	5 000		10 400	17 000
g ₁₁	4 850	20 000		400	650
g ₁₂	18 300	500		600	500
g ₁₃	2 400	1 400		0	100
g ₁₄	16 800	16 000		1 200	800
g ₂₂	1 550	500		25 000	17 000
g ₂₃	2 150	100		2 050	1 400
g ₂₄	2 950	500		53 800	45 000
g ₃₃	0	4 000		0	1 700
g ₃₄	0	2 200		3 200	1 950
g ₄₄	2 250	5 000		23 900	39 000
Summa	77 700	73 750		145 650	153 650

Tabell 6 Skattade värden för hur hushållen i de olika redovisningsgrupperna fördelar sig på stratum-tillhörighetsklasser vid stratifiering enligt Tabell 2 med Panel-11-stratifiering.

	Redovisningsgrupp							
	1	2	3	4	5	6	7	8
g ₁	2 700	150	0	0	150	100	17 700	3 100
g ₂	3 300	1 100	900	1 100	450	25 000	8 600	37 200
g ₃	0	0	3 200	0	3 200	0	0	667 700
g ₄	475 200	253 500	212 900	154 300	65 000	10 400	150	596 400
g ₁₁	150	600	0	100	100	400	4 850	850
g ₁₂	350	450	0	0	0	600	18 300	600
g ₁₃	0	0	0	0	0	0	2 400	700
g ₁₄	3 100	1 750	150	650	1 400	1 200	16 800	1 200
g ₂₂	1 200	900	1 000	1 250	350	25 000	1 550	5 100
g ₂₃	0	0	0	0	200	2 050	2 150	22 900
g ₂₄	7 500	4 900	4 350	5 400	4 400	53 800	2 950	14 900
g ₃₃	0	0	1 600	0	1 500	0	0	259 900
g ₃₄	15 800	0	4 000	4 000	3 400	3 200	0	84 800
g ₄₄	296 200	259 100	165 200	260 100	193 000	23 900	2 250	139 900
Summa	805 500	522 450	393 300	426 900	273 150	145 650	77 700	1 835 250

Som tidigare nämnts så innehåller Panel-11-stratifieringen tre riktade strata, nämligen

Stratum 1 som är riktat mot den "lilla" redovisningsgruppen "jordbrukarhushåll",

Stratum 2 som är riktat mot i första hand den "lilla" redovisningsgruppen "företagarhushåll" men även mot (den "lilla") redovisningsgruppen "jordbrukarhushåll",

Stratum 3 som är riktat mot den "stora" gruppen "ålderspensionärshushåll".

Med tanke på vad som sades tidigare är det av intresse att se vilka under- och övertäckningar av målgrupperna som erhålles vid dessa riktanden. Nedanstående Tabeller 7 och 8 ger belysning av frågorna. Tabell 7 erhålles genom att summera ihop på lämpligt sätt i Tabell 6.

Tabell 7 Skattade värden för hur hushållen i grupperna "företagarhushåll", "jordbrukarhushåll" samt "resterande hushåll" fördelar sig på stratumtillhörighetsklasser enligt Panel-11-stratifieringen

	Företagarhushåll	Jordbrukarhushåll	Resterande hushåll
g ₁	100	17 700	6 100
g ₂	25 000	8 600	44 000
g ₃	0	0	674 100
g ₄	10 400	150	1 757 300
g ₁₁	400	4 850	1 800
g ₁₂	600	18 300	1 400
g ₁₃	0	2 400	700
g ₁₄	1 200	16 800	8 300
g ₂₂	25 000	1 550	9 800
g ₂₃	2 050	2 150	23 100
g ₂₄	53 800	2 950	41 400
g ₃₃	0	0	263 000
g ₃₄	3 200	0	112 000
g ₄₄	23 900	2 250	1 313 500
Summa	145 650	77 700	4 256 500

Som tidigare antytts så består Redovisningsgrupp 8 av "restpotten" relativt Redovisningsgrupperna 1-7. (Skälet till detta är "tekniskt".) Förutom att nedanstående Tabell 8 ger en bild av vilka hushåll individer i Stratum 3 leder till, så redovisar den också en finare uppdelning av Redovisningsgruppen 8, vilket kommer att vara av värde längre fram.

Tabell 8 Vissa hushålls fördelning på vissa stratum tillhörighetsklasser enligt Panel-11-stratifieringen.

	Redovisningsgrupp 8				Redovisningsgrupperna 1-7
	Ålderspensionärshushåll	Studerandehushåll	Arbetslösa, förtidspensionär	Oklassificerade hushåll	
g ₃	668 000	0	0	0	6 450
Övriga g _h	10 000	160 000	260 000	207 000	1 230 000
g ₃₃	260 000	0	0	0	0
g ₃₁	700	0	0	0	2 400
g ₃₂	23 000	0	0	0	4 400
g ₃₄	85 000	0	5 000	0	30 500
Övriga g _{hλ}	16 000	4 500	62 000	75 000	1 415 700
Totalt	1 062 700	164 500	327 000	282 000	2 689 450

3 Estimationsprecision som kan uppnås med en HINK-undersökning
som innehåller fyra strata enligt Panel-11-stratifieringen.

Med hjälp av de allmänna formlerna i Rosén (1986) samt det framtagna empiriska materialet angående dels hur stratumtillhörigheterna (med Panel-11-stratifiering) sönderlägger hushållspopulationen dels medelvärden och varianser i de olika stratumtillhörighetsgrupperna kan man härleda nedanstående "teoretisk-empiriska" formler för hur precisionerna i medelvärdesskattningarna (för disponibel inkomst) för de olika socioekonomiska redovisningsgrupperna beror av vilka stickprovsstorlekar som väljs. (Detaljerna i beräkningarna redovisas annorstädes.)

Formlerna förutsätter att proportionalviktade estimatorer används.

Vidare används beteckningarna

n_h = urvalsstorlek i stratum h , $h = 1, 2, 3, 4$.

$\hat{\mu}_u$ = skattningen (baserad på en urvalspanel) av medelvärdet i redovisningsgrupp u , $u = 1, 2, \dots, 8$.

Formlerna i Tabell 9 är något förenklade versioner av mer generella formler. De förenklade versionerna fungerar dock fullt tillfredsställande över det variationsområde för stickprovsstorlekarna som vi är intresserade av i det här sammanhanget.

Inför kommande diskussion för vi också in gruppen "ålderspensionärshushåll" i bilden.

Tabell 9

Redovisningsgrupp av hushåll	Variansen $V(\hat{\mu}_u)$ som funktion av urvalsstorlekarna i de fyra stratana. Sort: 1984 års Kkr.
Ej facklärd arbetare	$\frac{6000}{n_4}$
Facklärd arbetare	$\frac{10200}{n_4}$
Lägre tjänstemän	$\frac{15880}{n_4} + \frac{55}{n_3}$
Mellan- tjänstemän	$\frac{17070}{n_4} + \frac{22}{n_2}$
Högre tjänstemän	$\frac{35200}{n_4} + \frac{300}{n_3} + \frac{45}{n_2 + 230} + \frac{12}{n_2} + \frac{5}{n_1}$
Företagare	$\frac{19500}{n_4} + \frac{1900}{n_2} + \frac{2460}{n_2 + 230} + \frac{20}{n_1 + 70}$
Jordbrukare	$\frac{420}{n_1} + \frac{430}{n_1 + 70} + \frac{730}{n_2} + \frac{250}{n_2 + 230} + \frac{470}{n_1 + 0.3n_2} + \frac{1700}{n_4}$
Övriga	$\frac{950}{n_4} + \frac{260}{n_3} + \frac{9}{n_2}$
Ålderspensionärer	$\frac{750}{n_4} + \frac{780}{n_3} + \frac{20}{n_2}$

I följande Tabell 10 redovisas varianserna för medelvärdesskattningar för de i Tabell 9 nämnda redovisningsgrupperna för ett antal olika urvalsallokeringar som satisfierar villkoret att totala urvalet är 5 000.

Som bakgrund till Tabell 10 behöver följande överväganden redovisas.

- Gruppen "ålderspensionärshushåll" är naturligtvis av intresse i olika sammanhang. För denna grupp kan dock god information om inkomstförhållanden erhållas från andra undersökningar än HINKen, bl a från den totalräknade inkomststatistiken. Detta talar för att HINK inte behöver ge bättre skattningsprecision för gruppen "ålderspensionärshushåll" än för andra (någorlunda stora) socioekonomiska grupper av hushåll.

Mot den bakgrunden har urvalsstorleken i Stratum 3, i de fall där den "står till förfogande", satts förhållandevis lågt, jämfört med tidigare i alla fall, företrädesvis till $n_3 = 200$.

- En helt annan grupp som fått ökad aktualitet under senare år är (den kanske något diffusa) gruppen "ungdomar". Det finns därför anledning att sträva efter förbättrad skattningsprecision för (förhållandevis små) redovisningsgrupper som karaktäriseras av någon "ungdomsegenskap". Det naturliga medlet för ett sådant syfte vore att införa ett "ungdomsstratum" i individpopulationen och att sampla över genomsnittet i detta stratum; t ex.

Stratum 5: Individer i åldersintervallet $[18,26]$ år.

En föreslagen allokeringsprincip för detta stratum är följande. Den "besparing" som gjorts på pensionärerna spenderas på ungdomarna. (Men andra principer är naturligtvis tänkbara.) Nyss nämnda princip leder till att Stratum 5 tilldelas $400-200=200$ fler observationer än vad det skulle få enligt sin "Stratum-4-tilldelning".

Även för denna fem-stratum-situation har formlerna i Tabell 9 använts för att beräkna estimatorvarianserna för de socioekonomiska redovisningsgrupperna. Dessa leder då till aningar av överskattningar, vilka dock torde vara helt försumbara för de mest "kritiska" redovisningsgrupperna, "högre tjänstemän", "företagare" samt "jordbrukare".

TABELL 10

Fall		a	b	c	d	e	f	g
Allokering		n ₁ = 70 n ₂ = 225 n ₃ = 1025 n ₄ = 3680	n ₁ = 550 n ₂ = 900 n ₃ = 400 n ₄ = 3150	n ₁ = 55 n ₂ = 100 n ₃ = 410 n ₄ = 4435	n ₁ = 50 n ₂ = 1500 n ₃ = 200 n ₄ = 3250	n ₁ = 1450 n ₂ = 1500 n ₃ = 200 n ₄ = 1850	n ₁ = 200 n ₂ = 800 n ₃ = 200 n ₄ = 3600	n ₁ = 200 n ₂ = 700 n ₃ = 200 n ₄ = 3700
Kommentarer till allokeringen		Proportio- nalalloke- ring dvs väsentligen ekvivalent med OSU	Nuvarande allokering före allo- kering på "finstra- ta"	Optimalallo- kering för redovis- ningsgruppen högre tjänstemän	Optimalallo- kering för företagare under premis- sen n ₃ = 200	Optimalallo- kering för jordbrukare under premis- sen n ₃ = 200	Några tänkbara allokeringar, varvid 200 observationer "avsatts till ungdomar".	
Varianser för skatt- ningar av genom- snittlig disponibel inkomst. Per panel, Sort: 1984 års Kkr	Redovisningsgrupp							
	Ej facklärda arbetare	1.63	1.90	1.35	1.85	3.24	1.67	1.62
	Facklärda arbetare	2.77	3.24	2.30	3.14	5.51	2.83	2.76
	Lägre tjänstemän	4.37	5.18	3.71	5.16	8.86	4.69	4.57
	Mellan-tjänstemän	4.74	5.44	4.07	5.27	9.24	4.77	4.64
	Högre tjänstemän	10.08	11.99	9.02	12.46	20.56	11.36	11.10
	Företagare	19.29	10.51	31.01	8.86	13.24	10.25	10.70
	Jordbrukare	16.75	3.60	25.05	14.08	2.37	6.39	6.61
	Övriga	0.55	0.96	-	-	-	-	-
	Ålderspensionärer	-	-	2.27	4.14	4.32	4.13	4.13

TABELL 10

Fall		h	i	j	k	l	m	n
Allokering		n ₁ = 200 n ₂ = 600 n ₃ = 200 n ₄ = 3800	n ₁ = 200 n ₂ = 500 n ₃ = 200 n ₄ = 3900	n ₁ = 300 n ₂ = 800 n ₃ = 200 n ₄ = 3500	n ₁ = 300 n ₂ = 700 n ₃ = 200 n ₄ = 3600	n ₁ = 300 n ₂ = 600 n ₃ = 200 n ₄ = 3700	n ₁ = 300 n ₂ = 500 n ₃ = 200 n ₄ = 3800	n ₁ = 400 n ₂ = 700 n ₃ = 200 n ₄ = 3500
Kommentarer till allokeringen		Några tänkbara allokeringar, varvid 200 observationer "avsatts till ungdomar".						
Varianser för skatt- ningar av genom- snittlig disponibel inkomst. Per panel, Sort: 1984 års Kkr	Redovisningsgrupp							
	Ej facklärda arbetare	1.58	1.54	1.71	1.67	1.62	1.58	1.71
	Facklärda arbetare	2.68	2.62	2.91	2.83	2.76	2.68	2.91
	Lägre tjänstemän	4.45	4.35	4.81	4.69	4.57	4.45	4.81
	Mellan-tjänstemän	4.53	4.42	4.90	4.77	4.65	4.54	4.91
	Högre tjänstemän	10.86	10.64	11.63	11.36	11.10	10.87	11.64
	Företagare	11.34	12.24	10.39	10.83	11.45	12.36	10.97
	Jordbrukare	6.89	7.27	5.07	5.27	5.52	5.86	4.53
	Övriga	-	-	-	-	-	-	-
	Ålderspensionärer	4.13	4.13	4.14	4.14	4.14	4.14	4.14

TABELL 10

Fall		o	p	q	r	s	t	u	v
Allokering		n ₁ = 400 n ₂ = 600 n ₃ = 200 n ₄ = 3600	n ₁ = 400 n ₂ = 500 n ₃ = 200 n ₄ = 3700	n ₁ = 400 n ₂ = 400 n ₃ = 200 n ₄ = 3800	n ₁ = 300 n ₂ = 700 n ₃ = 250 n ₄ = 3550	n ₁ = 300 n ₂ = 700 n ₃ = 200 n ₄ = 3600	n ₁ = 300 n ₂ = 700 n ₃ = 150 n ₄ = 3650	n ₁ = 300 n ₂ = 700 n ₃ = 100 n ₄ = 3700	n ₁ = 400 n ₂ = 700 n ₃ = 200 n ₄ = 3 700
Kommentarer till allokeringen		Några tänkbara allokeringar, varvid 200 observationer "avsatts till ungdomar".							Trolig allokering för Panel 13
Varianser för skatt- ningar av genom- snittlig disponibel inkomst. Per panel, Sort: 1984 års Kkr	Redovisningsgrupp								
	Ej facklärd arbetare	1.67	1.62	1.58	1.69	1.67	1.64	1.62	1.62
	Facklärd arbetare	2.83	2.76	2.68	2.87	2.83	2.79	2.76	2.76
	Lägre tjänstemän	4.69	4.57	4.45	4.69	4.69	4.72	4.84	4.57
	Mellan-tjänstemän	4.78	4.66	4.55	4.84	4.77	4.71	4.64	4.64
	Högre tjänstemän	11.36	11.11	10.88	11.20	11.36	11.73	12.60	11.09
	Företagare	11.59	12.48	13.83	10.91	10.83	10.76	10.68	10.67
	Jordbrukare	4.77	5.08	5.54	5.27	5.27	5.26	5.25	4.51
	Övriga	-	-	-	-	-	-	-	-
	Ålderspensionärer	4.14	4.14	4.15	3.36	4.14	5.43	8.03	4.13

4. Kommentarer till Tabell 10 samt fortsatt diskussion av stratifiering, allokering och skattningsprecision.

4.1 Diverse kommentarer

- A. Vi börjar med några kommentarer i anslutning till en jämförelse mellan Tabell 10 och Tabell 1.
- a) Vid ett första ögonkast på Tabell 1 får man kanske en känsla av att HINK lika gärna kunde använda ett OSU-urval som ett stratifierat urval. Dock, som tidigare påpekats är den mer adekvata förståhandedsslutsatsen av Tabell 1 att HINKen troligen varit olyckligt allokerad. Man torde väl heller icke tveka om att så har varit fallet, när man jämför vad som uppnås under "OSU-allokeringen" a) i Tabell 10 med vad som kan uppnås vid välvalda strata och god urvalsallokering, se lämpligt vald allokering bland f) - v).
- b) Tabell 10 ger också information om var de "stora olyckorna" i hittillsvarande allokering inträffat.

Som tidigare nämnts så har hittillsvarande HINK-stratifiering gått till så (åtminstone i stora drag) att man först gjort en Panel-11-stratifiering med dess fyra stratan, och därefter finuppdelat dessa fyra stratan i totalt 16 stratan. Urvalsfraktionerna har sedan varierats högst avsevärt mellan de 16 stratana.

Under allokering b) i Tabell 10 återfinns de estimatorvarianser man skulle erhålla i HINK-undersökningen med nuvarande allokering på "grovstratana" 1-4 (enligt Tabell 2) om man låte bli att göra den efterföljande finstratifieringen och "finallokeringen". De två vänstra kolumnerna i Tabell 1 beskriver vad man hittills har uppnått (åtminstone för de socioekonomiska grupperna) efter finstratifiering och "finallokering". Den konklusion man leds till är att den hittillsvarande "finallokeringen" är den del av allokeringen som varit klart olycklig.

c) OSU-värdena enligt a) i Tabell 10 är beräknade som specialfall av de allmänna variansformlerna i Tabell 9, vilka framtagits genom en tämligen komplicerad kombination av teori och empiri. OSU-värdena i Tabell 1 har beräknats avsevärt mera "rakt på". Man kan konstatera att OSU-värdena i Tabellerna 1 och 10 faktiskt visar god överensstämmelse. Detta är i alla fall ett indicium på att formlerna i Tabell 9 verkligen återspeglar vad som händer i "verkligheten".

B. Att en omallokering av HINKen leder till förbättrad skattningsprecision framgår av Tabell 10. Dessutom uppstår viss kostnadsminskning av följande skäl.

Kostnaderna för "deklarationsöverföring" är större för företagar- och jordbrukardeklarationer än för vanliga löntagardeklarationer. Med kostnaden för överföring av en persons löntagardeklaration satt till 1 gäller följande.

Tabell 11

Genomsnittlig deklarations-
överföringskostnad för
det urvalda hushållet

För urvalsperson från Stratum	1	2.7
" -	2	2.5
" -	3	1.5
" -	4	1.6

Tidigare "grovallokering" ($n_1 = 550$, $n_2 = 900$, $n_3 = 400$, $n_4 = 3150$) leder till en överföringskostnad om 9375 medan en ny "välallokering" (t ex $n_1 = 400$, $n_2 = 700$, $n_3 = 200$, $n_4 = 3700$) ger en överföringskostnad om 9050, dvs en besparing i överföringskostnaderna om cirka 4 %.

4.2 Om precisionerna i årsskattningar på basis av en välallokerad HINK-undersökning

Syftet här att ge en bild av vilka skattningsprecisioner som kan förväntas för årsskattningar baserade på en välallokerad HINK-undersökning. En fyrstratifierad och en femstratifierad (av ovan diskuterad typ) HINK är så pass lika varandra att det som sägs i det följande gäller i allt väsentligt för dem bägge.

I fortsättningen av detta avsnitt arbetar vi under följande förutsättningar.

- (i) HINK-undersökningens primärurval består av 10 000 observationer per år.
- (ii) Som "välallokering" väljes allokeringen v) i Tabell 10.

4.2.1 Vid skattning av genomsnittlig disponibel inkomst för grupper (av hushåll).

För de socioekonomiska grupperna har vi god information via materialet i Tabell 10, som under (i) och (ii) ovan ger nedanstående resultat

Tabell 12

Redovisningsgrupp (av hushåll)	Ungefärlig längd för 95 % konfidensgränser för årsskatt- ningen av genomsnittlig dispo- nibel inkomst. Sort: 1984 års Kkr.
Ej facklärda arbetare	± 1.8
Facklärda arbetare	± 2.3
Lägre tjänstemän	± 3.0
Mellantjänstemän	± 3.0
Högre tjänstemän	± 4.7
Företagare	± 4.6
Jordbrukare	± 3.0
Ålderspensionärer	± 2.9

För andra redovisningsgrupper än de socioekonomiska kan vi inte uttala oss med samma noggrannhet. Nedan ges dock en tumregel som ger någorlunda korrekt storleksordning på medelfelet för skattningar av genomsnittlig disponibel inkomst under tämligen allmänna förutsättningar.

Låt G vara en redovisningsgrupp (av hushåll) och sätt

$\hat{\mu}_G$ = HINKS årsskattning av genomsnittlig disponibel inkomst
(i 1984 års Kkr) för hushållen i G ,

$g(G)$ = antal hushåll i G .

Inför följande villkor,

- (iii) G hämtar endast en mindre del av sitt innehåll från företag-, jordbrukar- och ålderspensionärshushåll,
- (iv) inkomstförhållandena i G är inte "extrema" (vare sig uppåt eller nedåt).

Följande tumregel ges mer utförlig motivering annorstädes.

TUMREGEL. Under (i) - (iv) ges standardavvikelsen för $\hat{\mu}_G$, $\sigma(\hat{\mu}_G)$, med tämligen god approximation av nedanstående formel,

$$\sigma(\hat{\mu}_G) \approx \begin{cases} \frac{500}{\sqrt{g(G)}} & \text{om } G \text{ bara omfattar en-vuxenhushåll} \\ \frac{630}{\sqrt{g(G)}} & \text{om } G \text{ bara omfattar två-vuxenhushåll} \\ \frac{950}{\sqrt{g(G)}} & \text{om } G \text{ innehåller en blandning av en- och två-vuxenhushåll.} \end{cases}$$

Tumregeln illustreras numeriskt nedan.

Tabell 13

Antal hushåll i G	Tumregelns värde på $\sigma(\hat{\mu}_G)$ när		
	G består av bara envuxenhushåll	G består av bara tvåvuxenhushåll	G är en blandning av en- och tvåvuxenhushåll
50 000	2.2	2.8	4.2
100 000	1.6	2.0	3.0
200 000	1.1	1.4	2.1
300 000	0.9	1.2	1.7
500 000	0.7	0.9	1.3
800 000	0.6	0.7	1.1
1 000 000	0.5	0.6	1.0

4.2.2 Vid skattning av antalet hushåll i en redovisningsgrupp

TUMREGEL: Låt förutsättningarna (i) - (iii) vara uppfyllda. Låt $g_E(G)$ beteckna antalet envuxenhushåll i G, $\hat{g}(G)$ skattningen av $g(G)$ och $\sigma(\hat{g}(G))$ motsvarande standardavvikelse. Då gäller, faktiskt med god approximation,

$$\sigma(\hat{g}(G)) \approx \sqrt{g(G)} \cdot 18 \sqrt{1 + \frac{g_E(G)}{g(G)} - 4.2 \cdot g(G) \cdot 10^{-7}}$$

Nedan ges numerisk belysning av tumregeln.

Tabell 14

Antal hushåll i G (= g(G))	Tumregelns värde på $\sigma(\hat{g}(G))$ när		
	G består av bara en-vuxenhushåll	G består av bara två-vuxenhushåll	G består av lika många envuxen- och tvåvuxenhushåll
50 000	5 660	3 980	4 890
100 000	7 960	5 570	6 870
200 000	11 140	7 700	9 580
300 000	13 500	9 220	11 560
500 000	17 030	11 310	14 460
800 000	20 770	13 120	17 370
1 000 000	22 630	13 710	18 700

4.3 Om möjligheter att ytterligare förbättra HINKs stratifiering

Även om Panel-11-stratifieringen är bra så är det naturligtvis möjligt att den skulle kunna byggas ut/modifieras till en ännu bättre stratifiering. Nedan diskuteras några tänkbara förbättringsmöjligheter.

- a) Högre tjänstemän. När man med hjälp av de riktade Stratana 1 och 2 lyckats sänka estimatorvarianserna för "företagare" och "jordbrukare" så blir det gruppen "högre tjänstemän" som får den största estimatorvariansen. Denna stora varians beror dels på att gruppen är förhållandevis liten dels på att variabiliteten inom gruppen är förhållandevis stor (vilket bl a återspeglar att de högre tjänstemännens löner rör sig över ett förhållandevis brett spektrum).

Från formlerna i Tabell 9 framgår att det enda sättet att åstadkomma en någorlunda rejäl sänkning av denna varians (utöver en ökning av totala stickprovsstorleken) vore att införa ett stratum som riktar sig mot just högre tjänstemän. Nedan anges två tänkbara möjligheter.

- (i) Med hjälp av existerande utbildningsregister bör man kunna rikta in förhållandevis väl mot de högre tjänstemännen. Notera att i stort sett alla med högskoleexamen förs till den socio-ekonomiska gruppen "högre tjänstemän".

Denna väg tror jag på, om man anser det vara angeläget att förbättra skattningsprecisionen för just "högre tjänstemän".

- (ii) En annan möjlighet att rikta in sig mot de högre tjänstemännen ges kanske av ett "höglönestratum" i individpopulationen (vilket rätt lätt borde kunna åstadkommas med hjälp av SCBTB).

Själv är jag inte optimistisk om att särskilt mycket står att vinna längs den vägen. Man borde dock kanske göra lämpliga "empiriska" studier som kan ge klarhet i frågan.

- b) Ytterligare redovisningsgrupper av särskilt intresse. Hittills har diskussionen om riktade strata framför allt gällt de socioekonomiska grupperna, men vi har även fört in gruppen "ungdomar" i bilden. Det är naturligtvis möjligt att andra gruppindelningar innehåller små redovisningsgrupper som är så pass angelägna att de förtjänar ett riktat stratum. Dock, man måste naturligtvis vara på sin vakt så att man inte "stratifierar sönder" HINKen (igen).
- c) Definitionen av Stratum 2. En minskning av antalet typ-I-bommar från Stratum 2 (vid riktandet mot redovisningsgrupperna "företagare" och "jordbrukare") skulle kunna ge ytterligare förbättrad effekt åt Panel-II-stratifieringen, om inte minskningen av typ-I-bommar (= "undertäckningen") sker på bekostnad av en för stor ökning av antalet typ-II-bommar (= "övertäckningen") (se Rosén (1986)). Mot den bakgrunden kan man fråga sig om parametern "50 %" i definitionen av Stratum 2 i Tabell 3 är optimal?

För att belysa den nyss nämnda frågeställningen har följande "stratifierings-omstratifieringssituation" betraktats.

Låt vuxenindividpopulationen vara fyr-stratifierad på "sedvanligt" sätt i Stratan 1, 2, 3 och 4. Vi "omstratifierar" populationen, återigen i fyra stratan, vilka betecknas A_1 , $A_2(x)$, $A_3(x)$ och $A_4(x)$, $0 \leq x \leq 1$, på följande sätt.

Stratum A_1 består av (precis) de individer som ingår i Stratum 1.

Stratum $A_2(x)$ består av de individer utanför A_1 , för vilka gäller att

$$\max (F, J) > x \cdot (L + F + J),$$

där

L = individens löneinkomst "året innan",

F = individens företagarinkomst "året innan",

J = individens jordbrukarinkomst "året innan".

Kommentar: Helst skulle man i ovanstående definitioner av L , F och J ha velat använda "två år innan", eftersom det är sådana inkomster som används i SC8TB när HINK-Stratum 2 upprättas (se Tabell 3). Anledningen till att året-innan-inkomster valts här, är att den informationen finns lätt tillgänglig för andra-års-observationerna i gamla HINK-paneler, medan två år gamla inkomstdata vore avsevärt mer komplicerat att ta fram. Med tanke på ändringarna i stratumdefinitionerna så är den i det här sammanhanget mest intressanta panelen Panel 11, men den är tyvärr ännu bara första-års-observerad, varför vi valt att arbeta med Panel 10. •

Stratum $A_3(x)$ består av de individer utanför A_1 och $A_2(x)$ som tillhör Stratum 3.

Stratum $A_4(x)$ består av resterande individer, dvs individerna utanför A_1 , $A_2(x)$ och $A_3(x)$.

Fortsättningsvis betecknar vi tillhörighet till $A_2(x)$, $A_3(x)$ och $A_4(x)$ med 2^* , 3^* respektive 4^* .

På basis av observationerna i Panel 10 erhålles följande skattningar av antalet hushåll i olika stratumtillhörighetsklasser.

Tabell 15

	x	Företagar- hushåll	Jordbrukar- hushåll	Andra hushåll än företagar- och jord- brukarhushåll
g_2^*	50%	30 400	600	3 900
	25%	36 700	5 800	15 300
	10%	41 200	13 600	33 900
	0%	46 600	14 300	81 300
g_2		23 400	700	12 900
g_4^*	50%	15 300	6 100	1 636 600
	25%	9 000	900	1 625 200
	10%	8 700	900	1 623 600
	0%	3 300	200	1 591 700
g_4		23 400	6 000	1 626 700
$g_1+g_3^*$	50%	4 600	17 200	687 700
	25%	4 600	17 200	687 700
	10%	400	9 400	670 800
	0%	400	9 400	655 200
g_1+g_3		4 600	17 200	687 700
$g_2^*g_2^*$	50%	25 300	300	3 900
	25%	29 600	300	4 800
	10%	31 500	300	5 700
	0%	35 200	400	17 800
g_{22}		16 900	200	7 300
$g_2^*g_4^*$	50%	49 900	400	18 900
	25%	46 700	800	28 700
	10%	50 600	1 300	41 700
	0%	47 900	4 100	77 100
g_{24}		43 300	600	26 800

Tabell 15

	x	Företagar- hushåll	Jordbrukar- hushåll	Andra hushåll än företagar- och jord- brukarhushåll
$g_{2*1} +$ g_{2*3}^*	50%	2 000	2 400	700
	25%	2 000	3 600	1 300
	10%	3 000	9 000	17 400
	0%	3 000	9 900	21 800
$g_{21} + g_{23}$		2 100	500	800
g_{4*4}^*	50%	27 300	3 000	1 345 600
	25%	26 300	2 700	1 334 800
	10%	23 000	2 600	1 326 900
	0%	22 000	1 400	1 281 400
g_{44}		42 400	2 900	1 334 200
g_{3*3}^{*+} g_{3*4}^*	50%	3 700	7 500	397 300
	25%	3 700	7 500	397 000
	10%	200	3 400	376 000
	0%	200	1 800	370 600
$g_{33} + g_{34}$		3 400	7 600	396 900

Vi ger oss nu in på att med hjälp av materialet i Tabell 15 försöka belysa frågeställningen angående lämplig %-sats.

Som tidigare sagts så fungerar Panel-11-stratifieringen väl vad gäller riktandet mot jordbrukarhushåll med hjälp av Stratana 1 och 2. Det är därför egentligen bara träffförhållandena avseende företagarhushåll som vi är intresserade av här. Att vi ändå gett jordbrukarhushållen sin egen kolumn i Tabell 15 beror på att "övertäckning" (relativt företagarhushåll) från Stratum 2 som leder till jordbrukarhushåll är att betrakta som något positivt medan "övertäckning" som leder till andra

hushåll än jordbrukarhushåll är att räkna som negativt. Dock, med tanke på att vi här arbetar med Panel-10-material är "träffbilden" i kolumnen "jordbrukarhushåll" missvisande för vad som uppnås med Panel-11-stratifieringen.

Det önskemål vi ställer på ett bra Stratum $A_2(x)$ är framför allt att det skall ge liten "undertäckning" (typ-I-bommar) för företagarrhushåll, dvs litet bidrag från stratumtillhörighetsklasserna (4^*) , (3^*) , (4^*4^*) och (4^*3^*) men även liten "övertäckning" (typ-II-bommar), dvs stratumtillhörighetsklasserna (2^*) , (2^*2^*) , (2^*1) och (2^*3^*) skall helst leda till få hushåll andra än företagarr- och jordbrukarrhushåll.

Vi går nu till Tabell 15 och börjar med följande observation. Träffbilden med "50%-stratan" är genomgående bättre än med de ursprungliga stratana. Detta är dock bara en återspeglning av att en "prognos" (om vem som är företagare) blir bättre när den baseras på information från året innan än på information från två år innan. Detta är väl helt i linje med vad man kan förvänta sig. Det skulle naturligtvis vara fördelaktigt om HINKs Stratum-2-bildning kunde baseras på året-innan-uppgifter, men än så länge finns dessa inte tillgängliga vid tidpunkten för HINKs urvalsdragning.

Härnäst betraktar vi hur träffbilden för Stratum $A_2(x)$ ändras med x . Först en sammanställning av hur storleken av Stratum $A_2(x)$ varierar med x .

Tabell 16

x	Antal individer i stratum $A_2(x)$
50 %	168 200
25 %	210 300
10 %	286 700
0%	412 800

Från Tabellerna 15 och 16 framgår följande. Med minskande %-sats i definitionen av $A_2(x)$ så minskar naturligtvis "undertäckningen", men faktiskt tämligen odramatiskt, åtminstone vad gäller g_{44}^{**} . Dock, med minskande %-sats sker en förhållandevis snabb ökning av storleken av $A_2(x)$ och därmed också av övertäckningen.

Den slutsats vi vill dra av det hela är följande. Det ovan betraktade materialet kan inte uppfattas som någon alarmsignal om att den använda siffran 50% skulle vara olyckligt vald. Det finns därför inte underbyggd anledning att på denna punkt göra förändring till Panel 13. En något klarare bild av förhållandena bör kunna erhållas när andra-års-observationerna på Panel 11 finns tillgängliga. Eventuellt kan även hela denna %-problematik bli inaktuell i framtiden genom att man kommer att kunna skapa ett mer effektivt Stratum 2 med hjälp av den utvidgning av företagarregistret (CFR) som håller på att genomföras.

Urvalet i HINK-84

Appendix. Utdrag ur (vilket som helst av de tre) HINK-SM84.

Panel 10 (Återkommande från HINK-83)

Stratum	Familjeinkomst	Antal vuxna individer	
		Population	Urval
1 Pensionärer	- 51 500	652 186	175
2 "	51 501 -	748 802	225
3 Rörelseidkare	- 61 000 ¹	111 152	500
4 "	61 001 ¹ -	61 837	400
14 Jordbrukare	Brukningsareal ² < 20 ha	59 856	225
15 "	" 20 - 100 ha	42 475	300
16 "	" > 100 ha	3 549	25
5 Övriga gifta med barn	- 51 500	47 506	400
6 "-	51 501 - 155 000	961 576	400
7 "-	155 001 -	383 249	375
8 Övriga gifta utan barn	- 51 500	46 910	350
9 "-	51 501 - 155 000	754 457	250
10 "-	155 001 -	341 094	275
11 Övriga ej gifta med barn	- 95 000	305 008	500
12 Övriga ej gifta utan barn	- 95 000	1 794 422	350
"-	95 001 -	155 649	250
99 Nyttillkomna 18-åringar		124 000	50

1) Individinkomst.

2) Individer som tillhör familjeenhet enligt folkbokföringen där någon medlem återfinns i LBR och inkomst av jordbruksfastighet är minst hälften av arbetsinkomsten.

Urvalet i HINK-84

Panel 11 (Nydragen till HINK -84)

Stratum	Familjeinkomst	Antal vuxna individer	
		Population	Urval
14 Jordbrukare		23 537	40
15 "	2	50 393	330
16 "		13 843	181
1 Pensionärer	- 55 000	615 944	175
2 "	55 001 -	717 736	225
3 Rörelseidkare	- 65 000 ¹	204 346	500
4 "	65 001 -	88 375	400
5 Övriga gifta med barn	- 55 000	47 501	400
6 "-	55 001 - 166 000	905 693	400
7 "-	166 001 -	390 978	375
8 Övriga gifta utan barn	- 55 000	46 066	350
9 "-	55 001 - 166 000	741 404	250
10 "-	166 001 -	353 232	275
11 Övriga ej gifta med barn	- 101 000	312 821	500
12 Övriga ej gifta utan barn	- 101 000	1 818 775	350
13 "-	101 000 -	168 520	250

1) Individinkomst.

2) Urvalet av jordbrukare utgör en delmängd av urvalet i A/LEs lantbruksenkät. Stratumtillhörighet avgörs efter antal standardtimmar.

Tidigare nummer av Promemorior från U/STM:

NR

- 1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)
- 2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)
- 3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)
- 4Handledning i AID-analys. Anders Norberg (1980-10-22)
- 5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)
- 6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)
- 7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)
- 8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)
- 9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)
- 10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)
- 11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)
- 12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)
- 13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)
- 14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)
- 15 Granskning och evalvering av surveymodeller, tiden före 1960. Gösta Forsman (1985-05-30)
- 16 Variance estimators of the Gini coefficient - simple random sampling. Arne Sandström, Jan Wretman och Bertil Waldén (Memo, Februari 1985)
- 17 Variance estimators of the Gini coefficient - probability sampling. Arne Sandström, Jan Wretman och Bertil Waldén (1985-07-05)
- 18 Reconciling tables and margins using least-squares. Harry Lütjohann (1985-08-01)

- 19 Ersättnings och uppgiftslämnarbördans betydelse för kvaliteten i undersökningarna om hushållens utgifter. Håkan L. Lindström (1985-11-29)
- 20 A general view of estimation for two phases of selection. Carl-Erik Särndal och Bengt Swensson (1985-12-05)
- 21 On the use of automated coding at Statistics Sweden. Lars Lyberg (1986-01-16)
- 22 Quality Control of Coding Operations at Statistics Sweden. Lars Lyberg (1986-03-20)
- 23 A General View of Nonresponse Bias in Some Sample Surveys of the Swedish Population. Håkan L Lindström (1986-05-16)
- 24 Nonresponse rates in 1970 - 1985 in surveys of Individuals and Households. Håkan L. Lindström och Pat Dean (1986-06-06)
- 25 Two Evaluation Studies of Small Area Estimation Methods: The Case of Estimating Population Characteristics in Swedish Municipalities for the Intercensal Period. Sixten Lundström (1986-10-14)
- 26 A Survey Practitioner's Notion of Nonresponse. Richard Platek (1986-10-20)
- 27 Factors to be Considered in Developing a Reinterview Program and Interviewer Debriefings at SCB. Dawn D. Nelson (1986-10-24)
- 28 Utredning kring bortfallet i samband med den s k Metropolit-debatten. Sixten Lundström (1986-11-10)

Kvarvarande exemplar av ovanstående promemorior kan rekvireras från
Elseliv Lindfors, U/STM, SCB, 115 81 Stockholm, eller per telefon
08 7834178