

Översyn av urvalen i de objektiva skördeuppskattningarna

Erling Andersson



R&D Report
Statistics Sweden
Research-Methods-Development
1989:12

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 1989:12. Översyn av urvalen i de objektiva skördeuppskattningarna / Erling Andersson.

Digitaliserad av Statistiska centralbyrån (SCB) 2016.

Översyn av urvalen i de objektiva skördeuppskattningarna

Erling Andersson



R&D Report
Statistics Sweden
Research-Methods-Development
1989:12

Från trycket Augusti 1989
Producent Statistiska centralbyrån, Utvecklingsavdelningen
Ansvarig utgivare Staffan Wahlström
Förfrågningar Erling Andersson, tel. 08-783 47 90

© 1989, Statistiska centralbyrån
ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1989

INNEHÅLL

	sid
1. Inledning och sammanfattning	1
2. Urvalsmetod för huvudurvalet i spannmål och slåttervall	4
2.1 Allmänt om urvalsmetoder	4
2.2 Några olika metoder för PPS-urval	4
2.3 Val av storleksmått för urvalet av företag	6
2.4 En jämförelse av olika urvalsmetoder	9
2.5 Ny urvalsmetod	14
3. Dimensionering av undersökningarna	15
4. Allokeringen av huvudundersökningen i spannmål och vall	17
5. Återväxtundersökningen i slåttervall	19
6. Undersökningarna i potatis	20
6.1 Urvalsram för huvudurvalen	20
6.2 Urvalsmetod och allokering	22
6.3 Arealskattningar för potatis	24
6.4 Spillundersökningen i potatis	24
7. Urvalen för snabbstatistiken	26
8. Arealskattningarna	28
8.1 Tilläggsurval för skattningen av arealer	28
8.2 LBR-justeringen av arealskattningarna	29
9. Övrigt	32
9.1 Varierande antal provytor	32
9.2 Hektarskördeskattningarna - utnyttjande av arealuppgifter	32
9.3 Skattningar för små områden	34
10. Förslag till uppföljning och vidare utredningar	34
11. Referenser	36
Tabellbilaga	

SCB
A/FU
Erling Andersson

1989-06-06

ÖVERSYN AV URVALEN I DE OBJEKTIVA SKÖRDEUPPSKATTNINGARNA

1. Inledning och sammanfattning.

Från och med 1988 har ansvaret för skördeskadeskyddet överförts från staten till en stiftelse inom LRF. Skördeskadeskyddet har därvid fått en annan utformning än förut. Behovet av statistik från de objektiva skördeuppskattningarna har härigenom minskat, eftersom dessa tidigare i första hand har varit dimensionerade med hänsyn till skördeskadeskyddets behov. Skördeuppskattningarnas omfattning minskar nu till drygt en tredjedel av den tidigare, exempelvis kommer huvudundersökningen i spannmål och vall år 1989 att omfatta 4590 undersökningsenheter mot tidigare ca 12000. Från och med 1989 införs även en ny skördeområdesindelning med 106 skördeområden (SKO) mot tidigare 420. Skördeuppskattningar kommer att ske i 102 SKO.

I samband med en så stor nedskärning finns det anledning att bli intresserad av urvals- och skattningsmetoder gjordes i den sk URVASK-utredningen (slutrapport i PM A/FU 1986:14). Denna ledde till införandet av varierande antal provytor per uttaget fält. Dessförinnan gjordes en större utredning i början på 70-talet (PM AOS 1974:1), vilken resulterade i en omfördelning av resurserna mellan de olika undersökningarna. Framförallt minskade därvid omfattningen av spillundersökningarna medan återväxtbestämningarna i slättervall ökade.

En viktig målsättning i föreliggande utredning har varit att finna en urvalsmetod för huvudundersökningen som möjliggör områdesskattningar med acceptabel precision även för mindre frekventa grödor (vanligen brödsäd). Fram till 1988 har man för detta ändamål dragit ett sk förstärkningsurval. För detta extra urval av gårdar har man enbart undersökt den gröda som i respektive SKO normalt kan förväntas få strax under 15 observationer. En sådan förstärkningsundersökning är emellertid kostsam att genomföra och det vore därför önskvärt att huvudurvalet är effektivare "riktat" mot de mindre frekventa grödorna. För att undersöka möjligheterna till detta har egenskaperna hos ett antal olika urvalsmetoder studerats, se kapitel 2.

I denna PM studeras i första hand metoden för urvalsdragningen av gårdar i huvudundersökningen i spannmål och vall. Urvalets allokering är en annan viktig fråga. Vidare behandlas potatisundersökningen, snabbstatistiken, arealskattningarna, återväxtundersökningen samt olika detaljmoment. En sammanfattning av de framlagda förslagen, med kapitelhänvisningar till den följande texten, är följande:

- Huvudurvalet för spannmål och vall dras även i fortsättningen med PPS-urval, men med nytt storleksmått S. Storleksmättet kommer att vara en funktion av gårdens arealer av de enskilda grödorna (se formel (2.1)) enligt föregående års lantbruksregister (LBR). Hittills har storleksmättet varit lika med gårdens totala åkerareal. Kapitel 2.3.

- Undersökningspopulationen utgörs av företag med minst 5.1 ha åker. Urvalsram är föregående års LBR. Tidigare har även företag mellan 2.1 och 5.0 ha ingått i populationen. Populationsstorleken minskar därmed från ca 101000 företag till ca 85000 (LBR 1988). Den tidigare undre gränsen för areal utnyttjad åker, 0.3 ha, avskaffas.

- Företag inom samma församling hålles samman i urvalsramen för att garantera god geografisk spridning av observationerna inom varje skördeområde. Samtidigt eftersträvas en "maximalt slumpmässig" urvalsram. Därför sorteras företagen slumpmässigt inom församling medan församlingarna sorteras slumpmässigt inom SKO. En startpunkt väljs med slumpens hjälp, varefter man "stegar" fram i urvalsramen med konstant steglängd. Detta innebär att urvalet stratifieras proportionellt på församlingarna inom ett SKO. Tidigare har LBRs sorteringsordning fått gälla och startpunktens läge har delvis bestämts manuellt. Kapitel 2.5.

- Antalet undersökningsenheter (UE) i huvudundersökningen i spannmål och vall minskas till 4590 år 1989. Detta innebär 45 UE per skördeområde i medeltal. De senaste åren har totala antalet UE varit ca 12000. Kapitel 3.

- Principen med variabelt antal UE per skördeområde återinförs, eftersom de nya skördeområdena skiljer sig mycket ifrån om åkerareal (från 4000 ha till 60000 ha). Antalet UE tillåts variera mellan 36 och 54. Antalet bestäms utifrån skördeområdets arealer av spannmålsgrödorna och vall. Kapitel 4.

- I undersökningen av återväxt i vall ingår samtliga UE med utnyttjad vallåterväxt på provyrefältet. Hittills har ett underurval dragits bland gårdarna i huvudurvalet. Undersökningen minskar i omfattning från ca 2200 UE till ca 2000 UE. Den särskilda indelningen i återväxtområden (ÅVO) avskaffas. I fortsättningen beräknas vallåterväxten per skördeområde. Kapitel 5.

- Urvalen för undersökningarna i mat- och fabrikspotatis dras liksom förut med PPS, där urvalssannolikheten är proportionell mot gårdens potatisareal. Metoden med separata urval avseende bevattningsmöjlighet avskaffas 1989. Kapitel 6.2.

- Populationen för potatisundersökningarna omfattar företag med minst 5.1 ha total åkerareal och minst 0.5 ha mat-/fabrikspotatis. Även i detta fall ändras gränsen för total åker från tidigare 2.1 ha medan däremot gränsen för potatisareal blir oförändrad mot tidigare. Undersökningen i fabrikspotatis begränsas till det skodlingsområdet, omfattande hela Skåne och i stort sett all fabrikspotatisodling i Blekinge och Kalmar län. Detta innebär ingen förändring mot tidigare. Urvalsramen utgörs av potatisodlare som uppfyller ovanstående arealvillkor, enligt föregående års LBR. Sortering av urvalsramen och dragning av urvalet sker på motsvarande sätt som för huvudurvalet i spannmål och vall. Kapitel 6.2.

- Antalet UE reduceras till för matpotatis 2100 (tidigare ca 5400) och för fabrikspotatis 300 (tidigare ca 1000). Maximalt antal UE per skördeområde blir 60 för matpotatis och 70 för fabrikspotatis. I båda fallen tillåts antalet variera från 0 och uppåt, beroende på områdets potatisareal. Kapitel 6.2.

- Arealskattningarna för potatis, baserade på uppgifter i utläggningsprotokollet för mat- och fabrikspotatis, beräknas tills vidare på samma sätt som förut. Kapitel 6.3.

- Spillundersökningen i potatis genomförs även 1989 men avses framdeles att utgå, efter införande av standardtal för spillet i potatis. Undersökningens omfattning kommer 1989 att reduceras från tidigare 400 uttagna UE i matpotatis till 225 UE och från 100 till 60 UE i fabrikspotatis. En ny, tillfällig spillområdesindelning för matpotatis införs genom anpassning av den gamla indelningen (17 SPO) till de nya skördeområdena. Redovisningen av spilltal kommer dock liksom tidigare att ske för sammanlagda spillområden, vilka fastställs efter datainsamlingen. Antalet spill-UE per SPO blir proportionellt mot antalet UE i huvudurvalet. För fabrikspotatis bestäms ett enda spilltal för hela undersökningsområdet. Spillurvalen dras med OSU bland gårdarna i huvudurvalen. Kapitel 6.4.

- För snabbstatistiken i spannmål dras oberoende grödvisa underurval med PPS, där S är ett nytt, "grödinriktat" storleksmått (se formel (7.1)). Tidigare har snabburvalen dragits ur huvudurvalet med OSU. Urvalsramen, som utgörs av huvudurvalet, sorteras på liknande sätt som ramen för huvudurvalet. Snabburvalen allokeras per SKO istället för tidigare allokering per spillområde/naturligt jordbruksområde (SPO/NJO). Antalet observationer per gröda i snabbstatistiken förutsätts bli oförändrat 500 (förväntat antal) för alla spannmålsgrödorna utom höstråg som minskas till 300 observationer. Med den nya metoden blir snabburvalen betydligt bättre riktade mot respektive gröda varför antalet uttagna UE kan minskas avsevärt. Antalet uttagna UE per gröda bör kunna vara detsamma varje år. Medelfelsberäkningar införs i snabbstatistiken. Kapitel 7.

- Omfattningen av snabburvalen i potatis har bestämts till oförändrat 500 UE (förväntat antal observationer) i matpotatis och 125 UE i fabrikspotatis (tidigare 200). Liksom i spillundersökningen sker dragningen ur huvudurvalen med OSU. Kapitel 7.

- För att uppnå tillräcklig precision i arealskattningarna kompletteras huvudurvalet med ett tilläggsurval av samma storlek, 4590 UE. För gårdarna i tilläggsurvalet kommer i första hand uppgifter angående arealer att samlas in, urvalet kan även användas för eventuella tilläggsundersökningar. Arealuppgifterna samlas in via lantbruksregistret, med förtursbehandling av inkomna blanketter från urvalsgårdarna. Detta medför att publiceringen av preliminära arealer fördröjs ca tre veckor jämfört med tidigare. Undersökningen av höst-sådda arealer baseras av kostnadsskäl på enbart huvudurvalet, vilket medför begränsade möjligheter till redovisning på länsnivå. Kapitel 8.1.

- Preliminära arealer skattas liksom förut med regressionsmetoden. Beräkningsunderlaget utgörs av huvudurvalet i spannmål och vall samt tilläggsurvalet. En skattning beräknas för vardera urvalet och den slutliga skattningen beräknas som medelvärde av dessa. En studie av eventuella nivådifferenser mellan skattningarna från de båda urvalen bör göras årligen. Kapitel 8.1.

- Arealen för företag mellan 2.1 och 5.0 ha skattas på länsnivå med fjolårets LBR-uppgift. Kapitel 8.1.

- Den s k LBR-justeringen av primärskattningarna av grödarealer från huvudurvalet genomförs endast för betesvall, slåttervall och havre. För övriga grödor avskaffas nivåjusteringen. Primärskattningarna från tilläggsurvalet nivåjusteras inte i något fall. Kapitel 8.2.

- Principen med varierande antal provytor bibehålls och kommer liksom hittills att användas för korn, havre och vall första skörd i hela landet samt för höstvet och höstråg i Skåne. Dessutom införs metoden även för vallåterväxt, där det kommer att bli samma antal provytor som i vall första skörd. Eftersom urvalsmetoden ändras införs nya regler för när 1,2 eller 3 provytor skall läggas ut. Olika regler gäller för vall, fodersäd respektive brödsäd. Tidigare har en regel använts för samtliga grödor. Kapitel 9.1.

- Skattningar av hektarskördar beräknas på två olika sätt, beroende på områdesnivå. Med undantag för riksskattningarna utgörs beräkningsunderlaget av enbart data från huvudurvalet. Hektarskörden skattas med Horvitz-Thompson-estimatoren, se formel (9.1). Riksskattningarna av hektarskördarna beräknas däremot som ett vägt medelvärde av de skattade SKO-hektarskördarna. Vikterna utgörs av SKO-arealskattningarna från huvudurval+tilläggsurval. Totalskördar skattas med produkten av arealskattning och hektarskördeskattning. Kapitel 9.2.

- Undersökningen av kvalitet i spannmål bygger tills vidare på hela huvudurvalet, med undantag för Norrlandslänen.

2. Urvalsmetod för huvudurvalet i spannmål och slättervall.

2.1 Allmänt om urvalsmetoder

Den enklaste metoden för urvalsdragning torde vara OSU-metoden (obundet slumpmässigt urval). I detta fall ges samtliga objekt i populationen lika stor chans att komma med i urvalet. Metoden brukar flitigt användas i läroböcker om urvalsteknik, för att illustrera olika problemställningar. I praktiken är dock dess användningsområde relativt begränsat. I många statistiska undersökningar har man nämligen tillgång till någon form av hjälpinformation om populationsobjekten. Rätt utnyttjad kan denna information ge betydande förbättringar av skattningarnas precision, jämfört med OSU-skattningarna. Ett mycket vanligt sätt att utnyttja hjälpinformationen är att stratifiera urvalet, dvs populationen indelas i grupper, strata. Inom varje stratum dras sedan ett delurval, exempelvis med OSU. En annan metod, som på ett mera "långtgående" sätt använder tillgänglig information, är PPS-metoden. Vid PPS-urval ges varje objekt en urvalssannolikhet som är proportionell mot dess "storlek" (Probability Proportional to Size). Storleksmättet S kan i princip väljas fritt, i det optimala fallet är S proportionellt mot den storhet man vill skatta, vilket naturligtvis är ouppnåeligt i praktiken (dessutom vill man vanligen skatta flera populationsstorheter ur samma urval). För att PPS-metoden skall vara effektiv måste storleksmättet S vara relativt starkt korrelerat med åtminstone de "viktigaste" storheterna (se vidare sid 6).

I de objektiva skördeuppskattningarna har systematiskt PPS-urval (med slumpmässig startpunkt) tillämpats sedan starten 1961. Vid jordbruksundersökningar i Indien började man använda systematiskt PPS-urval redan på 1940-talet. Eftersom det numera finns en mängd olika PPS-metoder "på marknaden" kan det finnas anledning att undersöka om någon annan metod än den nuvarande är mera lämpad för skördeuppskattningarna. Ett femtiotal olika PPS-metoder behandlas t ex i "Sampling with unequal probabilities", av Brewer & Hanif (Springer 1983). Samtliga handlar om dragning utan återläggning. I följande avsnitt beskrivs kortfattat några av dessa.

2.2 Några olika metoder för PPS-urval

Systematiskt PPS-urval

Varje objekt tillordnas en "sträcka" motsvarande objektets storleksmått. Dessa sträckor får tillsammans bilda en linje. Sträckorna kan antingen vara i slumpmässig ordningsföljd eller sorterade på ett visst sätt. Linjens totala längd L delas i n lika stora delar, där n anger stickprovsstorleken. En startpunkt i första delen, mellan 0 och L/n , väljes slumpmässigt, varefter man "stegar fram" med konstant steglängd L/n och väljer ut de n objekt man därvid hamnar på. Vanligen används Horvitz-Thompson-estimatoren i samband med denna urvalsmetod. Skattningen av t ex totalen för storheten y ges av $Y_{HT}^* = \sum y_i / \pi_i$ där π_i anger förväntat antal UE för företag i . Detta är utan tvekan den enklaste metoden för manuell dragning av ett PPS-urval utan återläggning, vilket förklarar dess tidiga tillämpning. Denna urvalsmetod används för närvarande i de objektiva skördeuppskattningarna. En nackdel med metoden är att det är svårt att skatta estimatorns varians.

En variant av ovanstående metod är följande: Tag m stycken subsampel om vardera n/m objekt genom att välja m oberoende slumpmässiga startpunkter och sedan använda steglängden Lm/n . Metoden infördes av Mahalanobis vid jutearealskattningar i Indien, under beteckningen "interpenetrating subsamples". Här kan variansen enkelt skattas, genom att jämföra resultaten från delurvalen. Variansskattningsmetoden kallas oftast "random-group"-metoden och är ganska vanlig när man använder komplicerade estimatorer. Urvalsmetoden är ju mycket enkel, men problemet är att välja m . För att variansskattningen skall vara någorlunda stabil behöver m vara stort, men ju större m är, desto mer avviker proceduren från att vara dragning utan återläggning. Gränsfallet, när $m=n$, inne-

bär ju att dragningen sker med återläggning, vilket försämrar skattningsprecisionen. Det gäller alltså att söka en kompromiss.

RHC-metoden

En av de mera kända PPS-metoderna är RHC-metoden, uppkallad efter Rao, Hartley och Cochran. Om stickprovsstorleken skall vara n delas först urvalsramen slumpmässigt upp i n olika grupper. Antalet objekt per grupp bör vara så lika som möjligt. Storleksmått summeras inom varje grupp, varefter ett objekt dras ur vardera gruppen, med urvalssannolikheter lika med storleksmättet S_i dividerat med grupptotalen S_g . I detta fall används inte Horvitz-Thompson-estimatorn utan den mycket likartade RHC-estimatorn, $Y^*_{RHC} = \sum S_g y_g / s_g$ där y_g respektive s_g anger y -värde resp storleksmått för utvalt objekt i grupp g . Eftersom värdena på S_g kommer att vara olika blir urvalssannolikheterna inte strikt proportionella mot storleksmättet, vilket kan leda till att skattningarna förlorar något i precision. Metoden beskrivs även i Cochran: Sampling techniques 3rd ed (Wiley 1977), kap 9A.11.

Murthys och Rajs metoder

Denna metod kallas ibland också för successiv sampling utan återläggning, eller "draw-by-draw sampling" (dock finns flera varianter som kunde ges samma beteckning). Första objektet drages med sannolikhet proportionell mot storleksmättet S_i . Objekt nummer två drages på samma sätt utan återläggning, osv. Observera att urvalssannolikheterna därmed förändras i varje dragning, beroende på vilka objekt som redan dragits. Precis som var fallet med RHC-metoden blir urvalssannolikheten alltså inte strikt proportionell mot storleksmättet utan man får en utjämning av storleksskillnaderna och riskerar precisionsförluster, vid skattning enligt Horvitz-Thompson. Detta kan man i viss mån kompensera för genom att använda en annan estimator, "skräddarsydd" för denna urvalsmetod. Två gångbara alternativ har föreslagits av Raj respektive Murthy. Rajs skattningsmetod går ut på att efter varje enskild dragning bilda en delskattning och som slutlig skattning bilda medelvärdet av alla delskattningarna. Detta är samma princip som för systematisk subsampling enligt ovan och därmed kan variansen skattas enligt random-group-metoden. Formeln för Murthys estimator är mera komplicerad, se Cochrans bok, kap 9A.9.

Lahiris "hel-sampel"-metod

Med denna metod dras hela stickprovet, med n objekt, i samma dragning, enligt följande: Först summeras storleken av de n största (enligt tillämpat storleksmått) objekten i stratumet/populationen, vilket ger värdet MAXSUM. Sedan dras ett OSU-stickprov om n objekt, vilkas storleksmått summeras till URVSUM. Därefter dras ett rektangelfördelat slumpstal X mellan 0 och MAXSUM. Om $X > URVSUM$ förkastas hela stickprovet, i annat fall accepteras det. Proceduren upprepas tills ett stickprov godtas. Metoden innebär att ett stickprov accepteras med en sannolikhet i proportion mot stickprovets aggregerade storleksmått, $\sum S_i (=URVSUM)$. Detta betyder att den vanliga kvotskattningen $Y^*_R = (\sum Y_i / \sum S_i) \cdot S_{tot}$ är unbiased (S_{tot} = summan av alla storleksmått i stratumet/populationen). Systemet med omdragningar kan verka en aning omständligt, men å andra sidan är det mycket enkelt att dra OSU-urval.

Andra metoder

Det finns ytterligare många andra PPS-urvalsmetoder, några av dessa har emellertid ett visst krav på storleksmåttens fördelning. Närmare bestämt måste $n_{sko} S_i / S_{sko} < 1$, för att använda lämpliga beteckningar (här står n_{sko} för antalet uttagna i det skördeområde företag i tillhör, medan S_{sko} anger skördeområdets aggregerade storleksmått). I vårt fall är detta inte alltid uppfyllt eftersom vissa företag är mycket stora, med varje rimligt storleksmått mätt. Därmed faller vissa intressanta urvalsprocedurer bort, t ex Brewer's (se Cochran, kap 9A.8). Problemet skulle kanske kunna kringgåas genom att behandla de

största gårdarna separat, något som emellertid säkert medför andra besvärligheter. Ytterligare andra metoder är i praktiken begränsade till urvalsstorleken $n=2$ (med $n=2$ är det fortfarande möjligt att göra väntevärdesriktiga variansskattningar). I vissa fall behöver detta inte innebära någon allvarlig begränsning, det kan nämligen vara fördelaktigt att göra en finstratifiering av större, ursprungliga strata. För vårt vidkommande torde detta emellertid vara mindre lämpligt, då oftast endast de högfrekventa grödorna skulle få två observationer i ett stratum.

Nuvarande metod i de objektiva skördeuppskattningarna

Hittills har man använt systematiskt PPS-urval enligt ovan, med storleksmättet S =gårdens totala åkerareal. Urvalsramen utgörs numera av föregående års lantbruksregister, LBR. Från början användes dåvarande RLFs register som ram. Detta innehöll uppgift om gårdarnas totalareal men däremot inga arealer för enskilda grödor. Därmed var man i princip "tvungen" att välja storleksmättet total åkerareal, ett i och för sig högst naturligt val under alla omständigheter. Urvalsramen brukar inte sorteras i någon särskild ordning utan lantbruksregistrets ordningsföljd får gälla. Detta ska leda till att urvalsenheterna blir geografiskt utspridda, från början var LBR nämligen sorterat byvis inom församlingarna, något som emellertid gäller i lägre grad numera. Startpunkten för urvalet väljs relativt fjolårets startpunkt. Detta innebär i viss mån ett manuellt val och görs för att minska risken för att få många "omval", jämfört med förra årets urval.

Innan vi jämför de olika urvalsmetoderna kommer valet av storleksmätt att diskuteras i följande avsnitt.

2.3 Val av storleksmätt för urvalet av företag

Hittills har främst PPS-metoder behandlats men även stratifierade urval kan komma ifråga för de objektiva skördeuppskattningarna (utöver den geografiska stratifieringen som sker över skördeområden). En lämplig stratifieringsvariabel är då gårdarnas storlek, dvs i vilket fall behövs ett mått på gårdsstorleken. I avsnitt 2.1 nämndes att ett storleksmätt bör vara starkt korrelerat med de storheter som anses viktigast, för att leda till bra skattningar av dessa. Detta gäller även vid stratifierat urval. Totala åkerarealen har relativt hög korrelation med arealen av de större grödorna, i första hand vall och korn. Skattningarna av hektarskördar och arealer för dessa grödor har vanligen bättre precision än de för övriga grödor i samma område. Detta beror naturligtvis också på att antalet observationer oftast är störst för just vall och korn. Målsättningen i vår översyn av urvalsmetoden har varit att försöka "hjälpa upp" skattningarna för de mindre frekventa grödorna, till en relativt sett bättre precision. Samtidigt får naturligtvis inte skattningsprecisionen för de högfrekventa grödorna försämrats alltför mycket, helst inte alls. Detta är onekligen en något vagt formulerad målsättning men det ligger i sakens natur. Liksom de flesta statistiska undersökningar har de objektiva skördeuppskattningarna flera syften. Hektarskördar och total-skördar skall skattas för flera grödor, arealskattningar skall beräknas för ett ännu större antal grödor. Dessutom genomförs beräkningarna för flera regionala nivåer, från riksnivå ner till församlingar. Olika statistikanvändare har skilda krav, vilka inte samtidigt kan tillmötesgå på ett optimalt sätt. Till de objektiva skördeuppskattningarna är också ofta kopplade olika uppdragsundersökningar, vilka kan ha delvis andra krav på mål och metoder. Det finns visserligen olika tänkbara kriterier för att söka en "optimal strategi" (urvalsteknik, storleksmätt, estimationsmetod) för undersökningen i sin helhet. Någon analytisk lösning av en sådan problemställning är emellertid inte att tänka på. Vi har istället valt att utifrån olika simuleringskörningar försöka göra en rimlig bedömning om vilken strategi som verkar leda till ett "bra" resultat, inom den givna kostnadsramen.

En tidigare studie av en alternativ urvalsmodell för skördeuppskattningarna i spannmål och vall har gjorts i ett projektarbete vid Umeå Universitet (Lena Nyström: Skördeuppskattning då urvalssannolikheten beror på både totalareal och grödans areal, Umeå 1984). Den metod som studeras är PPS-urval för varje enskild gröda g (spannmål och vall) med storleksmättet $S(g) = \text{total åkerareal} + k(g) \cdot \text{grödareal}(g)$. I rapporten genomförs numeriska studier av skattningsprecisionen som funktion av värdet på $k(g)$. Resultaten visar att modellen kan ge vissa precisionsvinster, i första hand för slåttervall. En stor nackdel med den beskrivna modellen är emellertid att urvalen dras grädvis, utan samordning. Besöken av de uttagna gårdarna är kostnadskrävande, varför det är fördelaktigt att kunna undersöka så många grödor som möjligt på varje uttagen UE. Av denna anledning var den tidigare genomförda förstärkningsundersökningen, där enskilda grödor undersöktes, relativt kostsam.

Rapporten av Nyström gav upphov till idén om ett storleksmått som förenar möjligheterna att ta hänsyn till enskilda grödarealer och att dra ett samordnat urval, nämligen följande:

$$S_i = \sum_j k_j (A_{ij}/A_j) \quad (2.1)$$

k_j : viktfaktor för gröda j

A_{ij} : areal av gröda j på gård i enligt urvalsram

A_j : total areal av gröda j inom det skördeområde som gård i tillhör, dvs $A_j = \sum_i A_{ij}$

Med formel (2.1) tilldelas varje gård ett enda storleksmått så att ett samordnat urval blir möjligt. Formeln är flexibel på flera sätt. Vid summeringen över j kan man välja vilka grödor som skall ingå. (Lämpligen ingår alltid också en "gröda" vars areal utgör summan av resterande grödors areal. Detta är nödvändigt för att garantera att alla gårdar får en urvalssannolikhet större än noll). Vidare kan viktfaktorerna k_j väljas fritt. Formeln kan varieras beroende på vilket skördeområde det gäller.

Tanken bakom formel (2.1) är att urvalet riktas mera mot de i respektive område mindre frekventa grödorna för att ge fler observationer av dessa. Man kan säga att urvalet dras med "inbyggd förstärkning" för de lågfrekventa grödorna. Detta sker tack vare att gårdarnas arealer av de i skördeområdet mindre vanliga grödorna (av spannmål och vall) blir något "uppförstorade", genom att dividera med A_j . Till exempel kommer en m^2 höstvet i allmänhet att ha större chans att bli utvald som provyta än en m^2 vall, eftersom höstvetarealen oftast är mindre än vallarealen i ett skördeområde. Ett enkelt exempel: A har totalt 20 ha höstvet och 20 ha vall, medan B i samma skördeområde har enbart vall, 40 ha. Skördeområdets totala vallareal antas vara 5000 ha medan höstvet odlas på 1000 ha. Storleksmättet total åkerareal ger lika urvalssannolikhet åt A och B medan formel (2.1) leder till att A får 3 gånger större sannolikhet än B, vid lika viktfactorer (=1) för höstvet och vall: $S_A = 20/1000 + 20/5000 = 120/5000$, medan $S_B = 40/5000$.

Eftersom vall är den vanligaste grödan får, i likhet med exemplet ovan, gårdar med enbart vall betydligt lägre urvalssannolikhet än med hittillsvarande storleksmått. Detta är en önskvärd effekt. Rena vallgårdar är för det mesta små och brukarna är ofta svåra att nå, eftersom de har sin utkomst på annat håll. Om sådana UE är alltför vanligt förekommande torde detta ha en negativ inverkan på bl a provtagarnas motivation och tilltro till undersökningarna.

En nödvändig förutsättning för att storleksmättet (2.1) skall ha önskad verkan är att gårdarnas arealfördelning är någorlunda stabil över tiden. Om nämligen många odlare ena året odlar fodersäd/vall för att nästa år övergå till brödsäd (eller tvärtom) skulle effekten kunna bli motsatt den önskade. Detta är emellertid av naturliga skäl inte fallet. Vidare är det viktigt att skilja mellan lågfrekventa och obetydliga grödor, så att inte enstaka småodlingar leder till orimligt höga urvalssannolikheter. Detta kan undvikas med hjälp av viktfaktorerna k_j , genom att beräkna dessa på följande sätt:

$$k_j = m_j \cdot \min(100 \cdot A_j / (A_{\text{tot}} \cdot p_{\text{sko}}), 1) \cdot \min(A_j / q_{\text{sko}}, 1) \quad (2.2)$$

m_j = vikt faktorns maximala värde

$A_j = \sum_i A_{ij}$ = areal av gröda j i skördeområdet

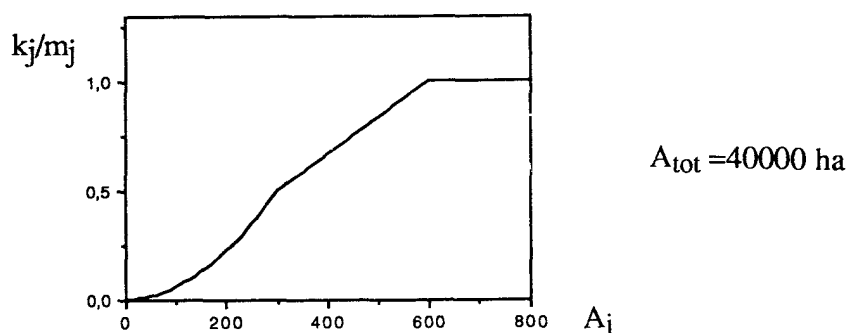
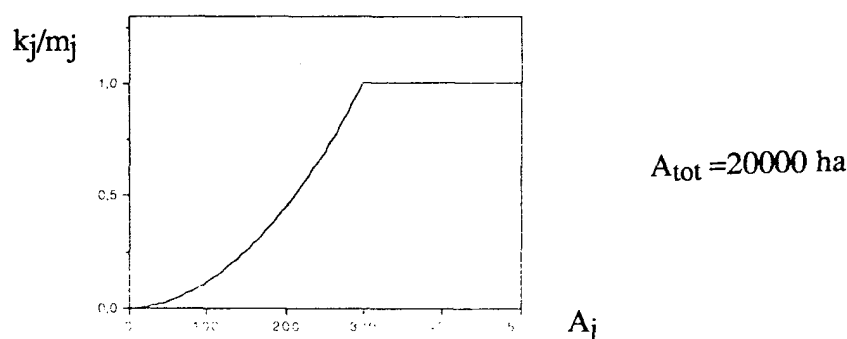
$A_{\text{tot}} = \sum_j A_j$ = total åkerareal i skördeområdet

$\min(a,b)$ ger minsta värdet av a och b

Innebörden av (2.2) är följande. Arealen av gröda j, A_j , måste dels utgöra minst p_{sko} procent av skördeområdets totala åkerareal, dels uppgå till minst q_{sko} hektar, för att grödan skall ges full vikt. Om minst ett av dessa villkor inte är uppfyllt reduceras vikten (båda faktorerna är en linjärt avtagande funktion av A_j när respektive gränsvärde har underskridits). Det relativa gränsvärdet, p_{sko} , bör sättas i relation till antalet utläggningsgrödor som odlas i skördeområdet. Ju färre grödor, desto högre sätts p_{sko} . Vissa skördeområden har mycket liten total åkerareal, därför behövs även det absoluta gränsvärdet, q_{sko} . Detta kan rimligen sättas lika för alla områden. Faktorn m_j , slutligen, gör att grödorna kan tilldelas olika full vikt.

I figur 2.1 visas hur k_j beror av A_j när A_{tot} dels är 20000 ha, dels 40000 ha. Värdet på p_{sko} är 1.5%, medan q_{sko} antas vara 300 ha.

Figur 2.1



Med storleksmättet (2.1) kan man alltså i viss mån styra urvalet och omfördela antalet observationer mellan grödorna. Därmed är man åtminstone en bit på väg mot målet att uppnå bättre balans mellan medelfelen för de olika grödorna. I nästa avsnitt jämförs de i avsnitt 2.2 nämnda urvalsmetoderna, vid användning av storleksmättet (2.1).

2.4 En jämförelse av olika urvalsmetoder

När vi jämför olika urvalsstrategier finns det ett antal faktorer att ta hänsyn till:

1. Skattningarnas effektivitet (vid givet antal observationer)
2. Fördelningen av antalet UE/gröda
3. Är variansskattningarna enkla eller mera komplicerade (omöjliga) ?
4. Är variansskattningarna stabila ? Väntevärdesriktiga ?
5. Är urvalsprincipen enkel eller komplicerad ?

Rangordningen bör vara ungefär som ovan, dock hänger faktorerna delvis samman. Antalet UE per gröda inverkar ju på precisionen i de enskilda skattningarna. Vidare lönar det sig naturligtvis inte att använda en enkel variansskattning om denna har mycket dålig precision.

Beträffande estimatorernas effektivitet finns rapporter över empiriska studier av bl a de procedurer som beskrevs i avsnitt 2.2 (Rao & Bayless, JASA 64 (1969), sid 540-559 och Bayless & Rao, JASA 65 (1970), sid 1645-1668). Om vi till en början bortser från systematiskt PPS-urval, visar resultaten att Murthys metod i allmänhet ger effektivare skattningar än de övriga. Murthys estimator är för övrigt en vidareutveckling av Rajs estimator och är även teoretiskt bevisat effektivare än denna. Denna förbättring kan vara olika stor, men är speciellt markant om urvalsfraktionen är stor eller om det finns betydande storleksskillnader mellan objekten i populationen, vilket inte är ovanligt i vårt fall (varje skördeområde får betraktas som en population). Rajs estimator är i sin tur något effektivare än RHC-estimatorn. Lahiris kvotskattning ger ojämna resultat. För vissa populationer kan den vara avsevärt effektivare än de övriga, speciellt när ett fåtal objekt är betydligt större än resten. Detta har visserligen endast visats för små stickprovsstorlekar, $n=2-4$, men gör ju ändå att metoden verkar intressant för oss, eftersom vissa gårdar är mycket stora. Emellertid kan Lahiris metod ge en mycket sämre skattning för andra populationer. Effektiviteten i Horvitz-Thompson-estimatorn vid systematiskt urval torde i allmänhet vara i nivå med Murthys metod, vid slumpmässig ordningsföljd i urvalsramen. En ordnad ram kan däremot påverka effektiviteten åt båda hållen.

Längre fram i detta avsnitt redovisas resultaten av simuleringsstudier av precisionen i skattningar av hektarskördar och arealer.

Genom simulerade urvalsdragningar har fördelningen av antal UE/gröda undersökts. Antalen i tabell 2.1 gäller urval för en preliminär SKO-indelning med 103 skördeområden, med 50 UE per SKO, dvs totalt 5150 UE. Som urvalsram har använts LBR 1985. Genom matchning av urvalen mot LBR 1986 har antalet UE som odlar respektive gröda kunnat bestämmas. Storleksmättet enligt formel (2.1) har använts i samtliga fall, med undantag för urvalsmetod=nuvarande. Denna innebär enligt ovan systematiskt PPS-urval med storleksmättet total åkerareal.

Tabell 2.1 Antal observationer/gröda vid simulerade urval

Urvalsmetod	Höstvete	Vårvete	Höstråg	Korn	Havre	Vall
Systematisk PPS	1887	1406	835	4305	3579	3610
Syst PPS, subsampl	1817	1335	884	4247	3589	3655
RHC	1736	1243	709	4244	3528	3677
Murthy/Raj	1854	1286	803	4323	3487	3653
Lahiri	1112	628	430	3179	2596	3579
Nuvarande	1611	1064	597	3994	3248	3799

Jämfört med nuvarande urvalsmetod ger ett likaledes systematiskt PPS-urval med användning av det nya storleksmåttet (2.1) en relativ ökning av totala antalet observationer i brödsäd med ca 25%. Största ökningen får höstråg med ca 30-40%. Dessutom ökar antalet observationer i korn och havre med 5-10%. Endast vall får ett minskat antal UE, ca 3-5%. En omfördelning i önskad riktning har alltså skett. Dessutom har antalet "billiga" UE, dvs sådana som odlar flera grödor, ökat kraftigt.

Subsampling-metoden vid systematiskt PPS-urval ger ingen större skillnad i antalsfördelningen jämfört med vanlig systematisk dragning, vilket heller inte är att vänta. Vad som emellertid inte framgår av tabellen är att subsampling leder till betydligt fler dubletter (företag som är uttagna mer än en gång). Detta beror på att startpunkterna för två eller flera subsampl i samma SKO ibland hamnar nära varandra. Vid dragning utan subsampling blir ett femtiotal företag uttagna mer än en gång, vilket får anses vara i sin ordning eftersom dessa är betydligt större än övriga. Med subsampling kan däremot även de minsta gårdarna bli dubbeluttagna. Vid dragning av fem subsampl per skördeområde (för att ge stabila variansskattningar torde detta vara i minsta laget) ökar antalet dubletter 5-6 ggr jämfört med dragning utan subsampling. Ju större antal subsampl per SKO desto fler dubletter kan förväntas, samtidigt som skattningsprecisionen avtar. Detta innebär en så allvarlig försämring att vanligt systematiskt PPS-urval är att föredra framför dragning med subsampling, trots att subsampling förenklar variansberäkningarna enligt avsnitt 2.2. Av övriga metoder ger Murthy/Rajs metod bäst grödfördelning för vårt syfte, dock något sämre än systematiskt urval. Med RHC-metoden blir fördelningen ytterligare något sämre. Lahiris metod verkar inte alls användbar i vårt fall eftersom samtliga grödor utom vall får betydligt färre observationer än med nuvarande urvalsmetod, vilket ju är i motsats till den önskade effekten.

Några urval har även dragits med en "tvåårig" urvalsram, under användning av storleksmåttet (2.1). Företagens grödarealer beräknades därvid som medelvärde av LBR-arealerna under de två närmast föregående åren. I ramen medtogs endast företag som fanns med i det sista årets LBR. Vitsen med detta skulle vara en gardering mot att förändringar i grödfördelningen på de enskilda företagen kan motverka den önskade omfördelningseffekten. Metoden ger emellertid något sämre fördelning av observationerna än vid användning av "ettårig" urvalsram, vilket måste betyda att korrelationen mellan de enskilda företagens grödfördelning är starkare för intilliggande år än med två års mellanrum. Därmed vinner man knappast något på att utnyttja flera års LBR-uppgifter i urvalsramen.

Angående skattning av varianser är skillnaden relativt stor mellan metoderna. RHC-metoden, systematisk PPS-subsampling, Rajs samt Lahiris metoder leder samtliga till mycket enkla variansestimater. Samtliga är väntevärdesriktiga, men beträffande variansestimaternas stabilitet är RHC-metoden och Rajs metod klart bäst. Murthys estimator har en mycket enkel varians estimator om $n=2$, men vid våra stickprovsstorlekar ($n=36-54$) är beräkningen av variansen i det närmaste ogenomförbar. Några approximationsformler är heller inte bekanta.

Vid systematiskt PPS-urval används Horvitz-Thompson-estimatorn, vars varians vanligen skattas med Sen-Yates-Grundy's estimator,

$$v^*_{SYG} = \sum \sum (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} (y_i / \pi_i - y_j / \pi_j)^2. \quad (2.3)$$

I formeln ingår storheterna π_{ij} (=förväntat antal par av UE från företagen i och j), vilka är svåra att bestämma. Dessutom är i många fall $\pi_{ij}=0$ vilket gör v^*_{SYG} meningslös att använda. Detta gäller i synnerhet om urvalsramen är ordnad i någon mening. Hittills har i de objektiva skördeuppskattningarna LBRs ordningsföljd fått gälla, vilket ger en ordnad urvalsram. Om urvalsramen är i slumpmässig ordningsföljd kan man använda approximativa skattningsformler för att bestämma π_{ij} , men beräkningarna är relativt komplicerade och kan resultera i negativa variansskattningar.

En approximativ variansskattning för Horvitz-Thompson-estimatoren bygger på ett modellantagande enligt följande (se t ex Brewer & Hanif, kap 1.8). Låt Y beteckna den storhet som skall skattas medan S anger det storleksmått som använts vid urvalsdragningen. Ett linjärt samband antas föreligga: $Y_i = \beta S_i + e_i$, där e_i är okorrelerade stokastiska variabler med $E(e_i)=0$, $Var(e_i)=v_i$. Eftersom man bör eftersträva stark korrelation mellan Y och S är detta ett rimligt antagande. Vidare antas att "felvariansen" uppfyller $v_i = v S_i^{2\mu}$, där $0.5 \leq \mu \leq 1$. Då kan följande approximativa variansformel härledas:

$$Var^*(Y^*_{HT}) = (n/n-1)[1 - (\sum \pi_i^{2\mu}) / (\sum \pi_i^{2\mu-1})] \sum (y_i/\pi_i - Y^*_{HT}/n)^2 \quad (2.4)$$

Formeln kan lätt utvidgas till att även gälla flerstegsurval. Uttrycket (2.4) innehåller ju inga π_{ij} , men istället parametern μ . Denna kan antingen skattas eller bestämmas med en kvalificerad gissning. För de flesta populationer som studerats under denna modell brukar gälla $0.6 \leq \mu \leq 0.9$ varför $\mu=0.75$ torde vara en god gissning. Tar man bort uttrycket inom klammer övergår estimatoren (2.4) i variansskattningen för dragning med återläggning. Uttrycket inom klammer kan således betraktas som en korrektionsfaktor för ändlig population. Om speciellt $\mu=0.5$ erhålles den välbekanta korrektionsfaktorn $(1-n/N)$.

I de objektiva skördeuppskattningarna har man valt att skatta variansen som vid dragning med återläggning och får därmed en viss överskattning. Denna överskattning har dock visat sig vara måttlig och har bedömts sakna betydelse. Detta borde isåfall i ännu högre grad vara fallet efter nedskärningen av skördeuppskattningarna, då urvalsfraktionen blir ännu mindre.

Den sista jämförelsepunkten gäller urvalsmetodernas enkelhet. I detta avseende är de systematiska procedurerna oöverträffade, åtminstone vid manuell dragning. I datorernas tidevarv kan man emellertid utan någon större ansträngning vara mera sofistikerad vid urvalsdragningen. Därför är det i dagens läge ingen nämnvärd skillnad på denna punkt mellan de studerade metoderna.

En sammanfattning av ovanstående diskussion ges i tabell 2.2, med "betyg" (högsta betyg är 5) för de olika metoderna avseende de fem jämförelsepunkterna. Huruvida betygen är helt rättvisande kan diskuteras men tabellen ger åtminstone en viss överblick. (Observera att faktorerna är rangordnade från vänster till höger i tabellen).

Tabell 2.2 "Betyg" för de studerade urvalsmetoderna

Urvalsmetod	Y* effektiv?	Fördelning UE/gröda	Var* enkel?	Var* stabil?	Urval enkelt?
Systematisk PPS	5	5	1	1	5
Syst PPS, subsampl	4	5	5	2	5
RHC	3	3	5	5	5
Murthy	5	4	1	4	4
Raj	4	4	5	4	4
Lahiri	3	1	5	3	4

Mot bakgrund av tabell 2.2 kan vi i fortsättningen bortse ifrån Lahiris metod. Tidigare har redan nämnts att systematisk PPS med subsampling bedöms vara sämre än vanlig systematisk PPS-dragning. Av tidsskäl uteslöts i detta skede även RHC-metoden från vidare

jämförelser även om metoden hade varit intressant att studera närmare. Med Murthys metod skulle problemet med skattningen av variansen inte vara löst. När det dessutom finns någorlunda approximativa variansskattningar att ta till vid systematiskt PPS-urval (betyget 1 för "Var* enkel?" är därför kanske alltför lågt) kan inte Murthys metod anses utgöra ett bättre alternativ. Därmed återstår endast systematisk PPS och Rajs metod. Dessutom kvarstår stratifierat urval (stratifiering efter storlek) som ett alternativ även om metoden inte finns med i jämförelsen ovan.

För att närmare undersöka skattningsprecisionen vid olika urvalsförfaranden har skattningar av grödarealer studerats genom simulerade urval med systematisk PPS, Rajs metod samt stratifiering. Allokeringen var 50 UE per skördeområde i en preliminär indelning med 103 SKO. Arealen per SKO skattades för spannmålsgrödorna och vall och jämfördes sedan med de "sanna" arealerna enligt LBR. Vid skattningen användes Horvitz-Thompson-estimatoren, Rajs estimator respektive kvotestimatoren, beroende på urvalsmetod. Regressionsskattningar med hjälp av föregående års LBR-uppgifter beräknades ej i detta fall.

Vid stratifierat urval uppdelades företagen i varje SKO på fem strata. I stratum 0 lades alla företag som med säkerhet blir dragna vid systematiskt PPS-urval, för vilka gäller $n_{sko}S_j/S_{sko} > 1$. För dessa företag skedde dragningen så att förväntat antal UE blev samma som vid systematiskt PPS-urval. Övriga företag fördelades så att de 40% minsta företagen i varje SKO bildade stratum 1, nästa 30% utgjorde stratum 2, nästa 20% stratum 3 medan de 10% största (bortsett från de "säkra") bildade stratum 4. Antalet UE per stratum 1-4 inom samma SKO skilde sig åt med högst en enhet.

Den optimala allokeringen enligt Neymans formel kan bestämmas för varje gröda för sig. Stratumindelningen enligt ovan innebär en grov kompromiss mellan Neymanallokeringarna för de olika grödorna.

För varje metod användes både LBR 1984 och LBR 1985 som urvalsram. Utifrån skattningarna för skördeområden beräknades variabeln $100 \cdot (\text{skattad grödareal} - \text{sann grödareal}) / \text{sann grödareal}$, dvs skattningens procentuella avvikelse från det sanna värdet. Efter varje urval beräknades grädvis denna variabels standardavvikelse över skördeområdena. Denna standardavvikelse är ett mått på respektive skattnings genomsnittliga avvikelse, sett över hela riket. Observera dock att "svarsfelet" är 0 i dessa beräkningar eftersom skattningarna har skett utifrån LBR-arealerna istället för arealuppgifter från besöksintervjuer. Detta saknar dock betydelse i sammanhanget.

Vid systematiskt PPS-urval användes både nuvarande storleksmättet total åkerareal och formel (2.1). För övriga metoder användes enbart (2.1). Vid beräkningen av (2.1) testades även olika värden på viktfaktorerna k_j , vilket snart gav ett stort antal olika kombinationer. Sammanfattningsvis kan sägas att maximala viktfaktorerna, m_j , inte bör vara alltför olika. I annat fall minskar skattningsprecisionen till följd av en för stor "snedvridning" av urvalet och en ökning av antalet företag som uttas mer än en gång. Lämpliga värden på m_j har bedömts vara $m_j=1$ för samtliga spannmålsgrödor, $m_j=0.5$ för vall i län 1-21, $m_j=1$ för vall i län 22-25 samt $m_j=0.3$ för övrig areal. Beträffande parametrarna p_{sko} och q_{sko} verkar det rimligt att använda $p_{sko}=1.5-3\%$, beroende på antal utläggningsgrödor i skördeområdet, samt $q_{sko}=300$ ha för alla SKO. Dessa parametervärden har använts vid urvalen som ligger bakom siffrorna i tabell 2.3. I tabellen redovisas standardavvikelser beräknade enligt ovan, baserade på minst fyra urval för vardera metoden. (Siffrorna för de olika grödorna är ej jämförbara sinsemellan då dessa bygger på olika antal skördeområden, eftersom endast skördeområden med en viss, minsta sann areal av respektive gröda inkluderats).

Tabell 2.3 Mått på SKO-arealskattningars medelavvikelse från sant värde (%)
Simulerade urval med olika urvalsmetoder

<u>Gröda</u>	<u>Nuvarande</u>	<u>Syst PPS</u>	<u>Rajs metod</u>	<u>Stratifierat urval</u>
Höstvete	16	15	17	15
Vårvete	21	18	20	22
Höstråg	31	25	30	30
Korn	12	12	12	13
Havre	14	14	13	15
Vall	16	18	18	18

Vid jämförelse av resultaten vid systematiskt PPS-urval enligt nuvarande metod samt med storleksmåttet (2.1) framträder samma bild som tidigare: Storleksmåttet (2.1) leder till förbättrade skattningar för brödsäd, ungefär oförändrat resultat för fodersäd samt något sämre resultat för vall, jämfört med storleksmåttet total åkerareal. Såväl Rajs metod som stratifierat urval ger i allmänhet något sämre resultat än systematiskt PPS-urval. Modifieringar av den ovan beskrivna stratumindelningen, genom annan fördelning i stratum 1-4, ledde inte till någon förbättring.

För att erhålla ytterligare jämförelseunderlag har även simulerade hektarskördeskattningar beräknats. I urvalsramen har i detta fall ingått enbart företagen i huvudurvalet för år 1985 respektive 1986, dvs företag med uppgifter om provyteskördar. Därmed bygger resultaten på enbart verkliga skördedata. Alternativt hade man kunnat simulera skördesiffror för resterande företag och därmed kunnat utnyttja hela LBR som urvalsram, men detta skulle göra tolkningen av resultaten osäkrare. Samma skördeområdesindelning användes som vid de simulerade arealskattningarna, dvs en preliminär indelning på 103 skördeområden. "Sanna" SKO-hektarskördar beräknades utifrån skattade totalskördar och arealuppgifter (enligt intervju) för företagen i urvalsramen. Urvalen drogs med 10-15 UE/SKO, där antalet UE varierade beroende på arealen brödsäd. Totalt drogs i varje urval ca 1300 UE, motsvarande en urvalsfraktion på ca 12%. Detta är visserligen ungefär dubbelt så hög urvalsfraktion som planeras i huvudurvalet, men var nödvändigt för att ge någorlunda säkra resultat. Även i detta fall studerades skattningarnas procentuella avvikelse från sant värde och beräknades denna variabls standardavvikelse över skördeområden. Medelvärden av 15-20 simuleringar per metod/år redovisas i tabell 2.4.

Tabell 2.4 Mått på SKO-hektarskördeskattningars medelavvikelse från sant värde (%)
Simulerade urval med olika urvalsmetoder

<u>Gröda</u>	<u>Nuvarande</u>	<u>Syst PPS</u>	<u>Stratifierat urval</u>
1985-86			
Höstvete	11.9	11.5	14.0
Vårvete	12.4	10.8	12.0
Höstråg	15.5	13.7	14.0
Korn	10.7	10.8	11.3
Havre	13.1	12.6	13.1
Vall	11.8	13.3	13.4
1984-85			
Höstvete	12.7	12.4	
Vårvete	12.2	10.8	
Höstråg	17.1	15.9	
Korn	10.8	10.9	
Havre	13.2	12.2	
Vall	12.2	13.8	

Liksom i tabell 2.3 bygger siffrorna för de olika grödorna på olika antal SKO varför resultaten inte är jämförbara mellan grödor. En jämförelse av metoderna ger ett liknande utfall som arealskattningarna ovan. För 1984-85 drogs inget stratifierat urval eftersom metoden inte längre ansågs utgöra ett tillräckligt bra alternativ. Vid dragning med Rajs metod uppstod problem angående behandlingen av de största företagen. Med systematiskt PPS-urval kan ju dessa utgöra mer än en UE. Det visade sig emellertid svårt att ha kvar denna möjlighet med Rajs procedur, då man vid estimationen tar hänsyn till de dragna objektens ordningsföljd. Försöken att lösa detta gav ibland uppenbart felaktiga skattningar, varför inga resultat presenteras för Rajs metod (arealskattningarna enligt Rajs metod i tabell 2.3 bygger däremot på högst en UE per företag). I detta läge ansågs inte heller Rajs metod tillräckligt konkurrenskraftig. Därför återstår enbart systematiskt PPS-urval som alternativ.

I ett senare skede av utredningen har ytterligare simuleringar av arealskattningar skett, för den fastlagda nya SKO-indelningen och med föreslagen allokering. Enbart systematiskt PPS-urval har studerats, med storleksmått=total åkerareal respektive enligt formel (2.1). Med LBR 1987 som urvalsram har skattningarna jämförts med arealerna i LBR 1988, på samma sätt som i tabell 2.3. Regressionsmetoden har ej använts. Denna skulle medföra förbättringar för framförallt de högfrekventa grödorna, jämfört med tabell 2.5. I tabellen redovisas medelvärden från fem simuleringar med vardera metoden, på SKO- samt länsnivå. Områden med mindre än 300 ha av respektive gröda har ej ingått i beräkningarna.

Tabell 2.5 Mått på arealskattningars medelavvikelse från sant värde (%)
Systematiskt PPS-urval med storleksmått=total åkerareal resp formel (2.1)

Gröda	SKOarealer		Länsarealer	
	Total åker	Formel (2.1)	Total åker	Formel (2.1)
Höstvete	13.8	12.7	9.2	8.4
Vårvete	21.9	19.4	14.8	12.2
Höstråg	22.3	21.1	23.4	20.4
Korn	10.7	10.6	5.1	5.9
Havre	13.3	13.9	7.5	8.9
Vall	11.0	14.4	4.8	8.1

Resultaten bekräftar tidigare slutsatser.

2.5 Ny urvalsmetod

Mot bakgrund av de resultat som redovisats i föregående avsnitt kommer huvudurvalet i spannmål och vall att dras med systematisk PPS-metod där storleksmålet beräknas enligt formel (2.1). Viktfaktorerna k_j bestäms därvid enligt formel (2.2) med följande parametervärden:

$m_j=1$ för spannmål i samtliga områden
0.5 för vall i SKO 0111-2122
1 för vall i SKO 2211-2521
0.3 för övrig areal i samtliga områden

$p_{sko}=1.5\%$ för SKO 0111-2122
2.0% för SKO 2211-2512
3.0% för SKO 2521

$q_{sko}=300$ ha i samtliga områden

De två första siffrorna i skördeområdenas beteckningar är lika med läns-koden för det län där skördeområdet har sin största åkerareal (enligt LBR 1987). Den tredje siffran anger antalet län inom vilket skördeområdet har åkerareal. En detaljerad redogörelse för områdesindelningen ges i SCB:s Meddelande i samordningsfrågor 1989:1.

De olika värdena på p_{sko} förklaras av att antalet utläggningsgrödor varierar mellan skördeområdena. Brödsäd odlas inte i Norrland (SKO 2211-2521), i SKO 2521 odlas heller ingen havre. Parametervärdena har efter simuleringskörningarna bedömts vara rimliga men kan naturligtvis efterhand justeras när så anses lämpligt, för att uppnå bättre fördelning inom varje enskilt SKO.

Hittills har LBRs sorteringsordning fått gälla i urvalsramen, vilket ger en ordnad urvalsram. Slumpmässighet i urvalsramens ordningsföljd är emellertid av flera skäl att föredra. Risker minskar för att företag skall komma med i urvalet flera år i följd samtidigt som variansberäkningen blir mer tillförlitlig. Dock skall nämnas att en ordnad urvalsram kan ge väsentliga förbättringar av skattningsprecisionen, men även motsatsen kan vara fallet. Den geografiska spridningen är emellertid viktig att ta hänsyn till, eftersom det kan föreligga betydande regionala skillnader inom skördeområdena enskilda år. Därför bör ordningen inte vara fullständigt slumpmässig, utan företag inom samma församling (-del) hålls samman. Därigenom blir urvalet proportionellt stratifierat på församlingarna i skördeområdet. Inom församling sker dock slumpmässig sortering av företagen liksom församlingarna sorteras slumpmässigt inom SKO. Den nya metoden innebär alltså att man övergår från ordnat systematiskt PPS-urval till slumpmässigt systematiskt PPS-urval.

Vid beräkning av varianser utgår man även i fortsättningen från formeln för dragning med återläggning:

$$\text{Var}^*(Y^*) = (n/n-1) \cdot \sum (y_i/\pi_i - Y^*/n)^2 \quad (2.5)$$

Detta har tidigare ansetts vara en acceptabel approximation även om man får en viss överskattning av variansen. Eftersom urvalsfraktionen i de olika undersökningarna nu blir betydligt lägre än förut får överskattningen ännu mindre betydelse i fortsättningen.

3. Dimensionering av undersökningarna

För att bestämma omfattningen av de olika undersökningarna i 1989 års skördeuppskattningar har kostnaderna för fältarbetet uppskattats. Beräkningarna grundar sig på resultaten av en kostnadsstudie från 1981-82, baserad på analys av uppgifter i provtagarnas veckorapporter 1981. I denna beräknades genomsnittlig arbetstid och reslängd per UE/fält/provtagare för olika arbetsmoment i fältarbetet. (Resultaten finns redovisade i bilaga 4 till "Utredning om de objektiva skördeuppskattningarnas omfattning" från lantbruksekonomiska samlingsnämndens sektorsgrupp (LES), juni 1982). Under antagandet att tidsåtgång och reslängder för de enskilda arbetsmomenten har varit oförändrade sedan dess, har en "framskriven" totalkostnad för fältarbetet åren 1982-87 kunnat beräknas. Dessa skattningar av fältarbetskostnaderna har sedan jämförts med de faktiska kostnaderna. Därvid framkom att de framskrivna resultaten ger en överskattning av de faktiska kostnaderna enligt en stigande trend på 1.45% per år. Denna siffra tolkas som den genomsnittliga årliga effektivitetsförbättringen i fältverksamheten. Den ökade effektiviteten kan främst förklaras av olika rationaliseringsåtgärder, exempelvis införandet av böjliga markeringskäppar och användningen av den ekonomiska kartan. För närmare detaljer angående beräkningarna hänvisas till PM A/FU 1988:28.

Efter nedskärningen av skördeuppskattningarnas omfattning kommer reslängden och därmed också restiden per arbetsmoment att öka, eftersom det genomsnittliga avståndet mellan provyttegårdarna ökar. Förändringens storlek är mycket svår att uppskatta, men ur

ovannämnda LES-rapport kan vissa uppgifter hämtas. I rapporten anges medelkörhastigheten till 40 km/tim vid dåvarande fulla omfattning av skördeuppskattningarna. Vid halverad omfattning antogs reslängden öka med 75% samtidigt som medelhastigheten ökade till 48 km/tim. Motsvarande värden vid nedskärning till en tredjedel antogs vara 100% respektive 50 km/tim.

Antalet UE för huvudundersökningen i spannmål och vall 1989 har efter beräkningar på olika alternativa omfattningar fastlagts till 4590 (45 UE/SKO i genomsnitt). Detta motsvarar drygt en tredjedel av tidigare omfattning. Låt oss följa ovanstående antagande om förändringen av reslängden vid nedskärning till en tredjedel, dvs för arbetsmomenten besöksintervju+utläggning av provytor, provtagning i vall samt provtagning i spannmål antas reslängden öka med 100% samtidigt som medelkörhastigheten blir 50 km/tim. Genom provurval av 4590 UE har antalet provytor uppskattats till 3200 i vall och 10600 i spannmål. Antalet observationer i återväxtundersökningen i slättervall uppskattas till 2000. Detta innebär endast 10% nedskärning gentemot tidigare. Antag att reslängden därmed ökar med 10% medan medelkörhastigheten stiger till 41 km/tim. Huvudundersökningen i potatis skall 1989 omfatta totalt 2400 UE, motsvarande en nedskärning till en dryg tredjedel. Reslängden antas därför öka med 100% och medelkörhastigheten sätts till 50 km/tim. Slutligen sker en minskning av spillundersökningen i potatis till drygt hälften av tidigare omfattning, varför vi antar att den genomsnittliga körhastigheten blir 48 km/tim medan reslängden ökar med 75%. I tabell 3.1 anges uppskattad tidsåtgång och reslängd för de olika arbetsmomenten enligt undersökningen 1981 samt för 1989 års undersökningar i enlighet med ovanstående antaganden. Observera dock att i tabellen inte har tagits hänsyn till den förbättrade effektiviteten sedan 1981 utan siffrorna för 1989 gäller under förutsättning av oförändrad effektivitet.

Tabell 3.1 Uppskattad genomsnittlig arbetstid (timmar) och reslängd (km) per enhet i de objektiva skördeuppskattningarna

		<u>1981</u>		<u>1989</u>	
	<u>Enhet</u>	<u>Tid</u>	<u>Reslängd</u>	<u>Tid</u>	<u>Reslängd</u>
Intervju+utläggning	UE	4.40	41.4	5.02	82.8
Provtagning vall	fält	1.86	30.7	2.32	61.4
Provtagning spannmål	fält	1.46	16.8	1.71	33.6
Återväxt utsätn+prov	fält	4.40	67.4	4.52	74.1
Potatis utläggn+prov	UE	3.14	35.3	3.67	70.6
Potatis spill	UE	1.85	26.2	2.15	45.8

Vid skattningen av totala fältkostnaden efter nedskärningen av skördeuppskattningarna har siffror angående provtagarnas löner och reseersättningar för 1987 använts. Skattningen kan jämföras med budgeterad fältarbetskostnad (efter nedskärning) för 1987/88. Dels fanns samtliga dessa uppgifter tillgängliga, dessutom är det osäkert om den skattade rationaliseringseffekten under 1981-87 kan antas gälla även under efterföljande år. Kostnaden per provtagartimme (lön, semesterersättning och lönekostnadspålägg) var 1987 85.41 kr, medan kilometerersättningen första halvåret var 1.59 kr samt under andra halvåret 1.67 kr. Total budget för fältarbetet var 8511 tkr i 1987 års priser. Den årliga rationaliseringsvinsten skattades enligt PM A/FU 1988:28 till 1.45% per år, dvs från 1981 till 1987 skulle rationaliseringseffekten vara 8.7%. En skattning av arbetstid och reslängd per arbetsmoment för 1987 (vid nedskuren omfattning), med hänsyn till effektivitetsförbättringen, erhålls därför genom att dividera siffrorna i de två högra kolumnerna i tabell 3.1 med faktorn 1.087. Med hjälp av angivna lönekostnader och reseersättningar kan man sedan skatta 1987 års "styckepriser" för de enskilda arbetsmomenten vid nedskuren omfattning, se tabell 3.2. Eftersom besöksintervju och provyteutläggning sker under våren har för denna post reseersättningen satts till 1.59 kr/km, för övrigt har kilometerersättningen antagits vara 1.67 kr. Till fältarbetet hör utöver uppgifterna i tabell 3.1 även "diverse uppgifter", innefattande undersökningar av höstsådda arealer, vallåterväxtens tillvaratagande, obärgade arealer,

insamling av kvalitetsprover m m. Med skördeuppskattningarnas faktiska omfattning under 1987 blev kostnaden för dessa uppgifter 1093 tkr, enligt beräkningarna i PM A/FU 1988:28. Efter föreslagen nedskärning beräknas kostnaden bli en dryg tredjedel och uppskattas i 1987 års priser till 350 tkr. Kostnaden för porto, elström och lagerlokaler var under åren 1985 och 1986 ca 500 tkr/år. Denna utgift antas i våra beräkningar vara 300 tkr. Vidare utgår kursarvoden och fasta ersättningar till provtagarna. Dessa uppgick under 1985 till sammanlagt 1201 tkr och under 1986 till 1246 tkr. Antalet provtagare minskar till knappt hälften jämfört med tidigare, varför denna utgiftspost antas bli 600 tkr i 1987 års priser. Därmed kan den totala fältkostnaden skattas, enligt tabell 3.2.

Tabell 3.2 Kostnader för fältarbetet efter omläggningen, 1987 års priser

	<u>Antal</u>	<u>Kostnad/enhet.kr</u>	<u>Total kostnad.tkr</u>
Intervju+utläggning	4590	515.56	2366
Provtagning vall	3200	276.66	885
Provtagning spannmål	10600	186.14	1973
Återväxt utsättn+provt	2000	469.32	939
Potatis utlägg+provt	2400	396.79	952
Potatis spill	285	239.39	68
Porto, elström, lagerlokaler			300
Diverse uppgifter			350
Kurser, fasta ersättningar			<u>600</u>
			8433

Marginalen till budgeterade 8511 tkr kan således uppskattas till 78 tkr. Utrymmet för andra, intermittenta undersökningar är därmed mycket knappt. Detta kan dock eventuellt vara något större ifall det har skett en effektivitetsförbättring även mellan 1987 och 1989. Med tanke på beräkningarnas osäkerhet bör man dock avvakta 1989 års faktiska utfall för att bättre kunna bestämma undersökningarnas fortsatta omfattning.

4. Allokeringen av huvudundersökningen i spannmål och vall

Under åren 1966-1980 varierade antalet UE per skördeområde i förhållande till områdenas sk jämförelsetal. Detta beräknades som summan av områdets åkerareal, uttryckt i ha, och antalet företag multiplicerat med 15. Från 1968 fördelades 30-39 UE per skördeområde. Tilldelningen var dock i de flesta fallen 30 UE, varför medelantalet blev knappt 31 UE per SKO. Efterhand, under trycket av besparingskraven, har antalsfördelningen utjämnats genom att "skära av topparna" och under senare år har majoriteten av områdena tilldelats samma antal. Sedan 1985 har antalet varit 28 UE, enstaka skördeområden har dock fått upp till 30 UE.

I den nya skördeområdesindelning som gäller från 1989 är det emellertid mycket stora skillnader i områdenas totala åkerareal. Arealen för företag större än 5 ha varierar från knappt 4000 ha till ca 60000 ha. Eftersom skördeområdena därmed kommer att få mycket olika vikt vid skattningar på högre regionala nivåer är det inte längre rimligt att ge samma antal UE till alla områden. Dessutom ökar skördeuppskattningarnas flexibilitet genom att tillämpa variabelt antal UE, eftersom man vid framtida anpassning till kostnadsanslagen har möjlighet att styra den regionala omfattningen av undersökningarna.

Som redan nämnts har omfattningen av 1989 års undersökning bestämts till genomsnittligen 45 UE per skördeområde. Lämpligen tilldelas varje SKO ett fast minimiantal samt ett tilläggsantal, som är en funktion av en eller flera variabler. En rimlig variation av antalet UE har bedömts vara mellan 80% och 120% av genomsnittsantalet, dvs 36-54 UE per skördeområde. Tilläggsantalet, utöver 36 UE, kan tänkas styras av olika variabler,

exempelvis skördeområdets totala åkerareal, areal spannmål+vall, areal spannmål, areal brödsäd, skördevariationer etc. Vi har valt att låta en funktion av skördeområdets arealer av utläggningsgrödorna (spannmål och vall) styra tilläggsantalet. Denna funktion har i princip samma form som formel 2.1, för beräkningen av företagets storleksmått. I det här fallet beräknas istället ett storleksmått, S_{sko} , för skördeområdet:

$$S_{sko} = \sum A_j / R_j \quad (4.1)$$

A_j = SKO-areal av gröda j (j =höstvet, vårvete, höstråg, korn, havre, slåttervall)

R_j = riksareal av gröda j

Antalet UE i ett skördeområde beräknas sedan enligt följande linjära funktion av storleksmättet (4.1):

$$(\text{Antal UE})_{sko} = 36 + \text{INT}(18 \cdot \min(S_{sko}/c, 1)) \quad (4.2)$$

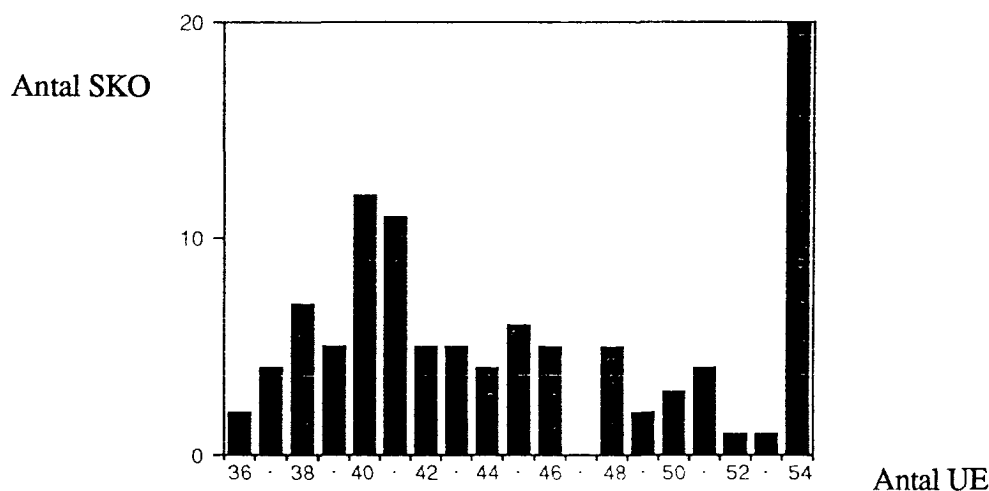
INT (=integer) ger heltalsvärdet av uttrycket inom parentes

$\min(a, b)$ ger minsta värdet av a och b , dvs alla SKO där storleksmättet S_{sko} är större än konstanten c får maximala antalet UE (=54)

c är en konstant som styr totala antalet UE. Konstanten måste bestämmas på nytt varje år för att ge exakt önskat antal (se programdokumentation).

Figur 4.1 visar hur fördelningen blev vid allokeringen för 1989 års undersökning.

Figur 4.1



Formel (4.2) kan naturligtvis även användas för andra storleksmått, exempelvis S_{sko} = skördeområdets totala åkerareal.

Fördelen med storleksmättet (4.1) jämfört med t ex total åkerareal är att bättre hänsyn tas till om ett skördeområde är av betydelse för någon av de mera lågfrekventa grödorna. Områden där flera utläggningsgrödor odlas ges ett ökat antal UE, vilket är önskvärt med tanke på att flera grödor "konkurrerar" om de tilldelade observationerna och inte alla grödor odlas på samtliga uttagna företag. I områden där färre grödor odlas är det däremot mera sannolikt att en uttagen gård har de flesta grödorna. Storleksmättet 4.1 leder till ett större genomsnittligt antal provyttefält per UE och därmed lägre kostnad per fält jämfört med om man skulle allokera i förhållande till total åkerareal. (Formel 4.1 förstärker alltså den ökning av antalet flergrodsodlare i urvalet som storleksmättet 2.1 leder till). Å andra sidan blir naturligtvis totalkostnaden högre vid fast antal UE.

Vid valet av allokeringsprincip gäller det bl a att ta hänsyn till de två delvis motstridiga kraven på bra skattningar för både skördeområden och riket. Om skattningar på SKO-nivå anses viktigare bör möjligen även skördevariationerna få påverka allokeringen, för att utjämna skillnader i SKO-skattningarnas medelfel. Detta skulle kunna ske genom att i storleksmåttet för skördeområdena även väga in ett mått på områdenas skördevariationer. Heterogena områden, med stora skördevariationer, skulle på så vis kunna bli "större" än homogena områden med lika stora arealer. Denna ansats har tagits upp tidigare, bl a i URVASK-utredningen där den dock förkastades. Inte heller här har frågan studerats närmare, eftersom de heterogena områdena anses ha fått tillräcklig "kompensation" genom den nya skördeområdesindelningen. Antalet skördeområden har nämligen fördelats mellan länen inte bara utgående från åkerareal och antal företag i länen som i den tidigare indelningen, utan även med viss hänsyn tagen till skördevariationerna i länen. Län med högre skördevariationer har därmed fått relativt något fler skördeområden än homogena län, jämfört med tidigare indelning. Eftersom varje SKO tilldelas minst 36 UE blir därmed skördeområdena i heterogena län i viss mån kompenserade för de högre skördevariationerna.

I tabell 1 i tabellbilagan redovisas allokeringen för 1989 års undersökning samt skördeområdenas totala åkerareal enligt LBR 1988.

5. Återväxtundersökningen i slåttervall

Återväxten i vall har på flera håll ökat i betydelse i förhållande till första skörden. I den enkätundersökning A/FU genomförde under våren 1987, angående de olika undersökningarnas fortsatta omfattning, ansåg därför flera statistikanvändare att återväxtundersökningen bör få en relativt större omfattning än tidigare. Ett vanligt förekommande svarsalternativ var att undersökningen bör anpassas efter återväxtens betydelse i förhållande till första skörden.

För att studera återväxtens betydelse beräknades medelvärdet per SKO (i den gamla indelningen) av återväxtens procentandel av total vallskörd för åren 1961-86 samt för endast de senaste tio åren av perioden, 1977-86. Medelvärdet för den senare delen av perioden var nästan genomgående högre, bortsett från i fjällområdet, och visar klart att återväxten med tiden har ökat i betydelse. För majoriteten av skördeområdena har återväxten svarat för mellan 30 och 40% av den totala vallskörden under perioden 1977-86. Variationerna i tioårsmedelvärdena är dock stora, från 5% till 51%. Så stora skillnader borde rimligtvis få påverka urvalsallokeringen i återväxtundersökningen.

Hittills har urvalet för undersökningen av vallåterväxten dragits som ett underurval (ca 2200 UE) bland gårdarna i huvudundersökningen av spannmål och vall. Allokeringen har skett genom att till varje återväxtområde (ÅVO) ge ett fast antal observationer, samt ett tillskott i proportion till vallarealen. Därefter har antalet UE i varje ÅVO fördelats bland ingående skördeområden i förhållande till deras vallareal. Eftersom den nya SKO-indelningen omfattar endast drygt 100 områden finns det inte längre någon rimlig anledning att ha kvar indelningen i 59 återväxtområden. Denna avskaffas därför och även återväxten undersöks framdeles per skördeområde.

En tänkbar metod för allokeringen av återväxtundersökningen på SKO-nivå, med hänsyn till återväxtens betydelse, vore följande: Beräkna för varje skördeområde medelvärdet över de tio senaste åren av den årliga kvoten mellan återväxtskörd (=utnyttjandegrad·skörd på utnyttjad areal) och första skörd av slåttervall. Beteckna detta medelvärde med $\bar{a}_{sko}$. Låt oss utgå från målsättningen att variansen (kg/ha) för hektarskördeskattningen i ett SKO skall vara lika stor för återväxten och första skörden. De två delarna ger då lika stort bidrag till variansen i skattningen av totala vallskörden i området, om vi bortser från korrelationstermen. (Beräkningar för åren 1961-86 visar att korrelationen mellan första skörd och återväxt vanligen ligger nära noll). Beteckna variansen för

skattningarna av första skörd och återväxt i ett SKO med Var_f respektive Var_a . Antalet observationer som ligger till grund för skattningarna betecknas n_f respektive n_a . Varianterna antas vara omvänt proportionella mot antalet observationer. Antag vidare att förhållandet mellan de relativa skördevariationerna kan skrivas $Vrel_a = k \cdot Vrel_f$. Vi gör här det förenklade antagandet att faktorn k är densamma överallt. Återväxten har större variationer än första skörden, varför $k > 1$. Faktorn k kan skattas eller ges ett schablonvärde. Bortse från skillnader i mätfel, även om en annan mätmetod används när återväxten betas. Vidare antas felet i skattningen av utnyttjandegraden vara försumbart i sammanhanget. Detta skulle leda till det ungefärliga sambandet

$$Var_a = Var_f \cdot k \cdot \hat{a}_{sko} \cdot n_f / n_a \quad (5.1)$$

För att uppnå $Var_a = Var_f$ bör man således välja $n_a = n_f \cdot k \cdot \hat{a}_{sko}$ vid allokeringen av återväxtundersökningen. Antalet uttagna företag som behövs för att erhålla n_a observationer är n_a / p_{utn} , där p_{utn} anger utnyttjandegraden i skördeområdet och kan skattas ur tidigare material. Detta antal måste i förekommande fall sänkas för att inte överstiga antalet uttagna i huvudundersökningen. Eftersom antalet observationer i första skörd inte är känt vid urvalsdragningen får n_f skattas med det förväntade antalet observationer i området. Återväxtundersökningens omfattning blir på detta vis direkt beroende av storleken på huvudundersökningen. Med 4590 UE i den senare uppskattas antalet observationer i återväxten med ovanstående metod till ca 1500 jämfört med 3200 i första skörd.

Ovanstående metod kan ibland förväntas leda till att samtliga UE i huvudundersökningen tas ut även för återväxtundersökningen. Att låta detta gälla överallt, dvs på samtliga vallgårdar i huvudundersökningen provta även återväxten, skulle innebära en avsevärd förenkling mot att använda ett underurval. Denna metod kommer därför att användas vid 1989 års skördeuppskattningar. Antalet observationer i återväxt förväntas då bli ca 2000, dvs omfattningen minskar endast obetydligt från tidigare 2200 UE. Om framtida besparingskrav framtvingar en återgång till underurval för återväxtundersökningen kan den ovan beskrivna allokeringemetoden bli aktuell.

6. Undersökningarna i potatis

6.1 Urvalsram för huvudurvalen

Före 1973 utgjordes urvalet för huvudundersökningarna i potatis av samma urval som för huvudundersökningen i spannmål och vall. Till följd av potatisodlarnas tilltagande specialisering ökade emellertid behovet av ett bättre "riktat" potatisurval. När lantbruksregistret infördes 1968 blev det möjligt att identifiera potatisodlarna i urvalsramen. Under 1969-72 kompletterades huvudurvalet med ett tilläggsurval för större potatisodlingar. Från och med 1973 har urvalet för potatisundersökningen dragits helt separat. Sedan 1977 har dessutom skilda urval dragits för bevattnade och obevattnade odlingar.

Urvalsramen har sedan 1973 utgjorts av företag med minst 2 ha total åker och minst 0.5 ha mat-/fabrikspotatis enligt föregående års LBR. Målpopulationen är de företag som uppfyller de nämnda arealvillkoren undersökningsåret. Urvalen dras med systematiskt PPS, där företagens potatisareal utgör storleksmått. Överensstämmelsen mellan urvalsram och målpopulation är emellertid klart sämre än i undersökningen i spannmål och vall, relativt sett. Detta kommer till uttryck både som under- och övertäckning. Enligt tidigare beräkningar (PM A/FU 1984:15) är undertäckningen för matpotatis nästan 10% av arealen, jämfört med knappt 1% för spannmål och vall. Undertäckningen kan medföra bias i skattningarna.

Tabell 6.1 visar hur antalet uttagna UE i matpotatis av olika orsaker reduceras innan provtagningen genomförs. Observera att kolumnen "övertäckning" anger antalet under-

sökta UE där provyteutläggning ej har skett. I detta antal ingår även undersökningsenheter med minst 0.5 ha matpotatis, men mindre än 0.5 ha utläggningsbar matpotatisareal. För arealskattningarna är dessa av intresse, t ex kan det finnas företag med stor areal tidigpotatis. Antalet UE som utgör övertäckning avseende bestämningen av både skörd och areal är därför något mindre än i tabellen.

Tabell 6.1 Antal UE, undersökningen i matpotatis

År	Uttagna	Vägrare, delade	Under- sökta	"Över- täckning"	Antal med grödan	Bortfall, obärgat	Antal obs ¹⁾
1985	5451	242	5209	602	4607	217	4390
1984	5447	262	5185	526	4659	320	4339
1983	5445	146	5299	496	4803	228	4575
1982	5468	182	5286	513	4773	111	4662
1981	5442	167	5275	484	4791	166	4625

¹⁾ Antal UE med skördeuppgift

Övertäckningen uppgår för matpotatis till ca 10% av antalet uttagna enligt tabell 6.1. För fabrikspotatis är övertäckningen ca 5-6%. Detta innebär dåligt utnyttjade resurser eftersom företagen i övertäckningen har besökts för intervju i utläggningsskedet, men sedan inte bidrar med några provyterresultat.

En mindre del av undertäckningen utgörs av företag som nybildats under året. Dessa kan naturligtvis inte täckas in med en urvalsram som bygger på tidigare års register. Övertäckningen kan å andra sidan delvis förklaras med att uttagna företag har lagts ned. Större delen av både under- och övertäckningen måste emellertid bero på tillfälliga förändringar i populationen av potatisodlare: Flera företag "vandrar" från eller till målpopulationen mellan åren. Möjligen gäller detta i särskilt hög grad företag med små potatisarealer. Ett sätt att minska undertäckningen skulle därför kunna vara att ta med även företag med 0.1-0.4 ha potatis i urvalsramen. För fabrikspotatis visar sig detta inte innebära någon vinst, medan för matpotatis undertäckningen skulle minska en del. Samtidigt ökar emellertid övertäckningen med ca 15000-20000 företag, vilket gör metoden oanvändbar.

En bättre lösning på täckningsproblemet kan istället vara att ha en "flerårig" urvalsram: Genom matchning på ID-nr kan arealuppgifter från de x senaste lantbruksregistren ligga till grund för urvalsdragningen. Medelarealen av mat- respektive fabrikspotatis beräknas över de x åren för varje ID-nr. Endast de företag, vars genomsnittliga potatisareal är minst 0.5 ha och som har mer än 5 ha total åker i det senaste lantbruksregistret, skall ingå i respektive urvalsram. På så vis kommer säkert alla stabila potatisodlare med i ramen, men även "intermittenta" odlare av större potatisarealer. Samtidigt försvinner småodlare som enstaka år har strax över 0.5 ha potatis. Därmed finns möjligheten att både under- och övertäckningen minskar. Metoden har testats genom att jämföra resultaten för x=1, 2 och 3, med potatisodlarna (≥ 0.5 ha) enligt LBR 1987 som målpopulation. Urvalsramen har alltså bildats ur LBR 1986, respektive LBR 1985 och LBR 1986 samt LBR från åren 1984-86. Odlarna har uppdelats efter bevattningsmöjlighet. Tabellerna 6.2 och 6.3 anger antal odlare samt potatisareal i under- respektive övertäckningsgrupperna.

Tabell 6.2 Undertäckning 1987, antal odlare samt areal, med x års urvalsram

Gröda	x=1		x=2		x=3	
	Antal	Areal	Antal	Areal	Antal	Areal
Matpotatis utan bev	546	1123	582	1078	619	1095
Matpotatis med bev	204	1027	207	834	209	809
Fabrikspot utan bev	84	231	76	197	77	195
Fabrikspot med bev	80	353	75	314	75	308

Tabell 6.3 Övertäckning 1987, antal odlare samt areal, med x års urvalsram

Gröda	x=1		x=2		x=3	
	Antal	Areal	Antal	Areal	Antal	Areal
Matpotatis utan bev	557	1181	674	1326	792	1493
Matpotatis med bev	218	1053	315	1152	379	1230
Fabrikspot utan bev	119	342	170	364	236	427
Fabrikspot med bev	52	356	89	397	122	408

Av tabell 6.2 framgår att undertäckningsarealen minskar något om man använder flerårig urvalsram, men blir ändå ganska stor. Detta betyder att de under 1987 nytillkomna odlarna i stor utsträckning varit utan potatisareal under de tre närmast föregående åren. Övertäckningen visar sig enligt tabell 6.3 öka vid flerårig urvalsram. Ökningen är större än reduktionen av undertäckningen. Orsaken till denna ökning torde i första hand vara den fortlöpande nedgången av antalet potatisodlare.

Resultaten visar att det knappast lönar sig att bilda urvalsramen med den alternativa metoden, utan den hittillsvarande lösningen innebär en vettig kompromiss beträffande under- och övertäckning och kommer därför att användas även i fortsättningen. Om undertäckningen leder till bias i arealskattningarna jämfört med LBR kan skattningarna modifieras för att reducera felet (se kapitel 8.2).

Genom att förändra undersökningarnas allokering kan antagligen täckningsproblemet reduceras till en del. Vid tidigare allokering har skördeområden med 1.1 ha potatis enligt urvalsramen tilldelats två UE. Det är troligt att täckningsfelen är större i skördeområden med så små potatisarealer. Problemet minskar genom den nya indelningen i större skördeområden, men det är ändå viktigt att vid allokeringen studera arealstatistiken i de nya områdena (se avsnitt 6.2).

Efter önskemål från det tidigare skördeskadeskyddet har sedan 1977 separata urval dragits avseende potatisodling med och utan bevattningsmöjlighet. Det nya skördeskadeskyddet har inget behov av en uppdelning, därför upphör arealuppdelningen i LBR avseende bevattning från och med 1989. Metoden med separata urval kan därför inte längre användas i de objektiva skördeuppskattningarna.

6.2 Urvalsmetod och allokering

Beslutet att i LBR upphöra med arealuppdelningen avseende bevattningsmöjlighet i potatis fattades under slutskedet av föreliggande utredning. Då hade redan frågan om urvalsmetod för potatisundersökningen behandlats, utgående från att arealuppdelningen skulle finnas kvar. Vissa av de förslag som hade framlagts kan visserligen inte användas när arealuppdelningen försvinner, men kan ändå vara värda att nämna.

Motiveringen för separata urval avseende bevattningsmöjlighet är att man därigenom har möjlighet att påverka skattningsprecisionen i respektive bevattningsgrupp. Målsättningen är att uppnå godtagbar precision för båda grupperna. Potatisarealerna i de nya skördeområdena framgår av tabell 2 i tabellbilagan. I de flesta områden med större potatisodling är arealen obevattnad gröda mindre än den bevattnade. Obevattnad potatis har i allmänhet högre skördevariationer än bevattnad potatis. Dessa båda faktorer gör att i första hand skattningsprecisionen för obevattnad gröda kan behöva förbättras. Detta har skett genom att dra ett genomsnittligen "tätare" urval i obevattnad potatis. Emellertid är detta möjligt utan att dela upp potatisodlarna i två bevattningspopulationer, vilket tidigare har lett till vissa beräkningsmässiga komplikationer: en UE kan vara uttagen i urvalet för gårdar med bevattning, men provytetältet råkar vara icke-bevattningsbart och vice versa. Istället beräknas för gårdarna i potatisurvalsramen ett storleksmått i analogi med formel (2.1):

$$S_i = k_{bev}(A_{bev,i}/A_{bev,sko}) + k_{obev}(A_{obev,i}/A_{obev,sko}) \quad (6.1)$$

k_{bev} : viktfaktor för bevattnad potatis

$A_{bev,i}$: areal av bevattnad potatis på gård i enligt urvalsram

$A_{bev,sko}$: total areal av bevattnad potatis inom det skördeområde som gård i tillhör

Analoga beteckningar för obevattnad potatis.

Ett storleksmått beräknas för vardera mat- och fabrikspotatisurvalet.

Ett lämpligt val av viktfactorer är, enligt simuleringskörningar

$$k_{bev} = \min((100 \cdot A_{bev,sko} / 40 \cdot (A_{bev,sko} + A_{obev,sko})), 1) \quad (6.2)$$

Faktorn k_{obev} beräknas analogt.

Formel (6.2) innebär att vikt faktorn för bevattnad (obevattnad) potatis är 1 om bevattnad (obevattnad) areal utgör minst 40% av skördeområdets totala potatisareal, i annat fall avtar vikt faktorn mot noll (jfr figur 2.1).

Dragning av PPS-urval, där storleksmättet beräknas enligt formlerna (6.1) och (6.2), ger en rimlig resursfördelning mellan bevattnad och obevattnad potatis. Antalet observationer i den mindre gruppen i respektive område förstärks, utan att använda separata urval.

Ovannämnda urvalsmetod kan inte användas eftersom bevattningsklassificeringen tas bort i LBR. Istället kommer från och med 1989 urvalen för potatisundersökningarna att dras med PPS, där S =gårdens totala potatisareal enligt föregående års LBR. Därmed går det inte att påverka fördelningen av antalet observationer mellan bevattningsgrupperna i varje enskilt SKO. Fortfarande finns dock naturligtvis möjlighet att skatta arealer och skördar med uppdelning på bevattning, genom bevattningsklassificering av provytfälten.

Huvudundersökningen i spannmål och vall reduceras till ca 39% av tidigare omfattning. Matpotatisundersökningen kan enligt vår bedömning skäras ned lika mycket, också med hänsyn till svaren i den tidigare nämnda enkätundersökningen. Undersökningen i matpotatis omfattar därvid 2100 UE år 1989. Undersökningen i fabrikspotatis bedöms kunna reduceras ytterligare något, till 30% dvs 300 UE.

Som framgår av tabell 2 i tabellbilagan varierar potatisarealen kraftigt mellan skördeområdena. Två områden har betydligt större areal matpotatis än övriga, nämligen SKO 1222 respektive SKO 1321, med vardera ca 10% av rikets totala areal. Fabrikspotatisodlingen är koncentrerad till skördeområdena 1011 och 1111, vilka tillsammans har ca 60% av fabrikspotatisarealen i riket. De stora skillnaderna i potatisareal måste naturligtvis avspeglas i allokeringen. För potatisundersökningarna tillåts därför antalet UE per SKO variera från 0 upp till ett maximiantal. Maximitilldelningen för undersökningen i spannmål och vall är satt till 54 UE. Eftersom matpotatis i allmänhet har något högre skördevariationer än de stora spannmålsgrödorna och slättervall, har maximiantalet för matpotatis satts till 60 UE. Fabrikspotatisodlingens starka koncentration till två skördeområden föranleder en ännu högre gräns och har satts till 70 UE. Denna kan synas onödigt hög men har följande förklaring: För skattningar på riksnivå är det optimalt att låta antalet UE per SKO vara proportionellt mot produkten av areal och skördevariation. De två "stora" skördeområdena, 1011 och 1111, skulle isåfall få ca 90 UE respektive ca 120 UE. För att öka säkerheten i SKO-skattningarna i övriga skördeområden bör emellertid resurserna omfördelas till förmån för dessa områden. Maximiantalet 70 UE bedöms vara en lämplig avvägning.

Antalet UE per skördeområde beräknas i analogi med formel (4.2), t ex för matpotatis:

$$(\text{Antal UE})_{sko} = \text{INT}(60 \cdot \min(S_{sko}/c, 1)) \quad (6.3)$$

S_{sko} = skördeområdets totala matpotatisareal enligt urvalsramen
konstanten c bestäms för att ge önskat totalt antal UE

Undersökningen i fabrikspotatis begränsas liksom tidigare till grödans huvudsakliga odlingsområde. Detta omfattar hela Skåne (samtliga SKOn med de två första siffrorna = 11 eller 12) samt områdena 0731, 0811 och 1011. Därmed täcker man även in 99.8% av fabrikspotatisarealen i Blekinge län. Utanför undersökningsområdet finns ca 6% av rikets fabrikspotatisareal, till stor del bestående av utsädesodlingar.

I tabellbilagans tabell 3 redovisas allokeringen för 1989 års undersökningar i potatis.

6.3 Arealbeskattningar för potatis

Utöver skattningarna baserade på huvudurvalet för spannmål och vall beräknas även skattningar av potatisarealer utgående från uppgifterna i utlägningsprotokollen för mat- och fabrikspotatis. För dessa skattningar anses regressionsmetoden inte ge någon vinst eftersom urvalet dras med PPS mot potatisarealen. Man använder istället den vanliga Horvitz-Thompson-estimatoren (se avsnitt 2.2). Denna ger en primärskattning med vilken förändringskvoten i förhållande till fjolårets primärskattning kan beräknas. Förändringskvoten multipliceras med fjolårets LBR-areal. Matpotatisodlingar mindre än 0.5 ha ingår dock ej i beräkningarna enligt ovan. För dessa antas arealen vara oförändrad jämfört med föregående års LBR. Total matpotatisareal skattas länsvis genom summering av skattningarna för odlingar med minst 0.5 ha och fjolårets LBR-areal för mindre odlingar (se även PM A/FU 1988:17).

En alternativ, enklare skattningsmetod är att använda den ojusterade primärskattningen, dvs att bortse från de sk förändringskvoterna. Dock bör småodlingar av matpotatis behandlas som förut och arealen på dessa skattas med fjolårets LBR-areal. Resultatet av en jämförelse mellan denna alternativa metod och den hittillsvarande metoden redovisas i tabell 6.4. I tabellen anges differensen mellan riksarealskattningarna och LBR-arealen samt summan av länsarealskattningarnas absolutdifferenser gentemot LBR.

Tabell 6.4 Potatisarealskattningar med två olika metoder ("Nuv"=hittillsvarande metod; "Alt"=alternativ metod). Differenser mot LBR.

	1987		1986	
	Nuv	Alt	Nuv	Alt
Matpotatis, riksareal	0	-670	+670	-170
Matpotatis, länsarealer	1470	1490	1310	620
Fabrikspotatis, riksareal	-120	-190	+180	-80
Fabrikspotatis, länsarealer	170	190	190	220

Det verkar inte vara någon större skillnad mellan metoderna. Dock kan man lägga märke till att den alternativa metoden ger en underskattning av riksarealen, till följd av undertäckningen. Före 1986 finns inget underlag att tillgå, varför det är svårt att avgöra vilken metod som är att föredra. Tills vidare skattas därför potatisarealerna på samma sätt som förut.

6.4 Spillundersökningen i potatis

Spillområdesindelningen

En ny indelning i 106 skördeområden (tidigare 420) gäller från och med 1989. Detta ger anledning att även revidera indelningen i spillområden (SPO). Inom de närmaste åren kommer troligen att införas standardtal för spillkorrektionen i potatis, liksom tidigare har skett för slättervall och spannmål, varvid spillundersökningen läggs ner. Därmed bör SPO-indelningen kunna utföras utan alltför mycket detaljarbete.

En enkel metod för bildandet av nya spillområden är att anpassa de tidigare 17 spillområdena till de nya skördeområdena genom att föra varje SKO till det SPO (enligt gamla indelningen) där skördeområdet har sin största åkerareal. Resultande SPO-indelning redovisas i tabell 4 i tabellbilagan. Denna indelning kommer att gälla 1989. I tabell 6.5 anges total åkerareal, areal matpotatis, areal fabrikspotatis (i undersökningsområdet) samt förändringen gentemot tidigare indelning, för de nya spillområdena. Förändringen uttrycks i total åkerareal "ut" (=åkerareal som försvinner ur det gamla spillområdet med samma nummer) respektive areal "in". Förändringen i potatisarealer torde vara i proportion till angivna förändringar.

Tabell 6.5 Arealstatistik enligt LBR 1987 för ny spillområdesindelning

SPO	Total åker	Matpotatis	Fabrikspotatis	Förändringar	
				Areal ut	Areal in
1	542 000	9 841	8 561	2 000	20 000
2	204 000	1 139	461	15 000	11 000
3	162 000	1 057	184	54 000	3 000
4	150 000	567	44	27 000	24 000
5	395 000	2 579		5 000	12 000
6	346 000	745		11 000	0
7	108 000	592		3 000	24 000
8	396 000	3 888		11 000	17 000
9	137 000	264		600	20 000
10	45 000	250		6 000	5 000
11	16 000	276		3 000	3 000
12	79 000	531		4 000	13 000
13	85 000	342		15 000	6 000
14	21 000	91		12 000	5 000
15	41 000	163		3 000	6 000
16	82 000	768		2 000	0
17	45 000	704		0	2 000

Förändringarna jämfört med tidigare indelning blir i vissa fall relativt stora, vilket man framdeles bör tänka på vid beräkningen av standardtal för spillkorrektionen. Område nummer 1 har ju klart störst potatisarealer, medan potatisarealerna i flera andra spillområden är mycket små. Antalet spillområden är därför kanske onödigt stort, men reduceras inte för att möjliggöra jämförelser bakåt i tiden. Någon redovisning av spilltal blir emellertid inte möjlig utan måste ske för sammanslagna områden för att uppnå tillräcklig precision. Detta har också varit nödvändigt med den gamla indelningen på senare år, efter nedskärningar av spillundersökningen. Eftersom odlingen av fabrikspotatis är starkt koncentrerad till SPO 1 kommer endast ett spilltal att bestämmas för hela undersökningsområdet. Hur sammanslagningen skall ske för matpotatis bestäms lämpligen efter datainsamlingen.

Allokering

En lämplig allokeringsprincip är att låta antalet UE i spillundersökningen vara proportionellt mot antalet UE i respektive huvudurval. Spilltalen anses nämligen inte ha något "egenvärde" som motiverar en ökad tilldelning till spillområden med mindre potatisareal, för att förbättra möjligheten till redovisning av spilltal. Spillundersökningen i potatis har på senare år omfattat 400 uttagna UE i matpotatis samt 100 UE i fabrikspotatis. En viss nedskärning har bedömts vara rimlig och 1989 års undersökning omfattar 225 UE i matpotatis respektive 60 UE i fabrikspotatis. Urvalen dras med OSU bland gårdarna i huvudurvalen genom dragning av "var n:te" UE, efter slumpmässig sortering inom SKO och mellan SKO. Därmed blir allokeringen proportionell på SKO-nivå enligt ovan, men ej exakt proportionell på SPO-nivå vilket dock är av mindre betydelse. Antal uttagna UE per SPO i 1989 års huvudurval och spillurval framgår av tabell 6.6. I tabellen redovisas även antalet observationer i 1987 års spillundersökning.

Tabell 6.6 Antal uttagna UE 1989 samt antal observationer i 1987 års spillundersökning, för ny spillområdesindelning

SPO	Hu.unders	Matpotatis		Hu.unders	Fabrikspotatis	
		Spill	Spill 1987		Spill	Spill 1987
1	609	66	123	265	51	56
2	146	15	32	26	6	16
3	121	13	17	8	2	10
4	64	7	11	1	1	9
5	200	21	35			
6	88	7	9			
7	75	11	6			
8	412	44	49			
9	26	4	3			
10	23	3	6			
11	32	3	12			
12	61	7	13			
13	42	4	12			
14	12	1	3			
15	19	1	2			
16	93	10	13			
17	77	8	13			

7. Urvalen för snabbstatistiken

Sedan 1974 har varje år beräknats snabbstatistik över preliminära hektarskördar och total-skördar på riksnivå för spannmål och potatis. Beräkningarna utförs på uppdrag av Svensk Spannmålshandel respektive Sveriges Potatisintressenter. Resultaten används i första hand vid prisförhandlingar på hösten. Statistiken grundas på provtagning i de objektiva skördeuppskattningarna på ett underurval av provyttegårdar i de ordinarie huvudurvalen. Omfattningen av underurvalen (förväntat antal observationer) har varit 500 UE i spannmålsgrödorna (län 01-21) och matpotatis samt 200 UE i fabrikspotatis. Uppdraget kvarstår även i fortsättningen, men mot bakgrund av nedskärningen av skördeuppskattningarna har bestämts att urvalen för snabbstatistiken minskas ned för höstråg och fabrikspotatis, till 300 UE respektive 125 UE. För övriga grödor förutsätts oförändrad omfattning. Redovisningen av snabbstatistik för fabrikspotatis sker en vecka senare än förut.

Tidigare har allokeringen av underurvalen för snabbstatistiken skett genom att fördela totalt önskat antal observationer på först SPO, därefter på anpassade naturliga jordbruksområden och slutligen på SKO-nivå. Varje SPO har tilldelats ett fast antal samt ett tillägg i proportion till grödarealen. Allokeringen för undersökningen i potatis har gjorts på liknande sätt, men med län som enda mellannivå. För närmare detaljer, se PM A/FU 1979:32 respektive 1978:41.

Det främsta syftet med snabbstatistiken är att erhålla bra riksskattningar. Därför borde istället övervägas att dra urvalen direkt för hela riket utan hänsyn till några mellannivåer. En annan intressant skillnad gentemot huvudurvalet (i spannmål) är att underurvalen för snabbstatistiken dras grödis, dvs urvalen behöver inte samordnas. Detta kan man dra nytta av genom att rikta underurvalen mot respektive gröda, så att viktiga odlare av grödan med stor sannolikhet kommer med och för att hålla antalet uttagna UE så lågt som möjligt. För potatisundersökningen dras enligt föregående kapitel separata huvudurval för mat- och fabrikspotatis, med PPS mot företagens potatisareal. Därigenom kan man inte "rikta"

underurvalen för snabbstatistiken ännu mera. Dessa bör därför dras med OSU, på samma sätt som urvalen för spillundersökningen (se avsnitt 6.4).

För spannmålsgrödorna kan däremot riktade underurval åstadkommas genom att för var och en av grödorna beräkna ett nytt storleksmått för varje UE i huvudurvalet för spannmål och vall och därefter dra underurvalen med systematiskt PPS-urval. Tidigare har dragningen av underurvalen för snabbstatistiken skett med OSU bland företagen i huvudurvalet. Den nya metoden innebär att uppräkningsstalen (se formel (9.1)) förändras jämfört med huvudurvalet, vilket tidigare inte har varit fallet. De nya uppräkningsstalen måste bestämmas för var och en av spannmålsgrödorna.

För beräkningen av storleksmättet för respektive urval ansätts formelerna

$$S_{hv} = \text{areal höstvetes} + k_1 \cdot (\text{areal vårvetes} + \text{areal höstrågs}) + k_2 \cdot \text{övrig åkerareal} \quad (7.1)$$

$$S_{vv} = \text{areal vårvetes} + k_1 \cdot (\text{areal höstvetes} + \text{areal höstrågs}) + k_2 \cdot \text{övrig åkerareal}$$

$$S_{hr} = \text{areal höstrågs} + k_1 \cdot (\text{areal höstvetes} + \text{areal vårvetes}) + k_2 \cdot \text{övrig åkerareal}$$

$$S_{ko} = \text{areal korn} + k_1 \cdot \text{areal havre} + k_2 \cdot \text{övrig åkerareal}$$

$$S_{ha} = \text{areal havre} + k_1 \cdot \text{areal korn} + k_2 \cdot \text{övrig åkerareal}$$

Tanken är att storleksmättet i möjligaste mån skall vara lika med grödarealen under undersökningsåret. Arealen av en spannmålsgröda på ett företag antas kunna "förklaras" av företagets grödarealer föregående år enligt högerleden i (7.1). Sambandet med arealer av grödor inom samma grödgrupp (brödsäd/fodersäd) torde vara högre än med övriga grödarealer. Därför ingår den andra termen i summan. Även den sista termen måste vara med för att ge alla enheter en urvalssannolikhet större än noll. Konstanterna k_1 och k_2 har bestämts ur några årgångar av LBR genom regressionsberäkningar enligt modellen (7.1), där S_{hv} ersatts av areal höstvetes enligt påföljande års LBR och analogt för övriga spannmålsgrödor. Det visar sig att för höstråg är sambandet med föregående års areal av övrig brödsäd så svagt att k_1 bör sättas lika med k_2 . Regressionsberäkningarna gav följande faktorer:

Höstvetes:	$k_1 = 0.27$	$k_2 = 0.16$
Vårvetes:	$k_1 = 0.21$	$k_2 = 0.14$
Höstråg:	$k_1 = 0.10$	$k_2 = 0.10$
Korn:	$k_1 = 0.19$	$k_2 = 0.11$
Havre:	$k_1 = 0.12$	$k_2 = 0.07$

Simulerade urval visar att antalet uttagna enheter i underurvalen kan minskas betydligt med denna metod, jämfört med tidigare. Vidare bör antalet uttagna kunna hållas konstant över åren. Tabell 7.1 visar antalet uttagna UE för snabbstatistiken i spannmål under de senaste tre åren samt antalet för 1989, vilket förväntas ge önskat antal observationer (observera dock att antalet observationer i höstråg skall minska från 500 till 300).

Tabell 7.1 Antal uttagna UE för snabbstatistik

<u>Gröda</u>	<u>1986</u>	<u>1987</u>	<u>1988</u>	<u>1989</u>
Höstvetes	1175	1223	1273	640
Vårvetes	2887	2329	2218	880
Höstråg	2359	2769	3129	800
Korn	654	647	647	550
Havre	753	735	749	590

För potatis kan naturligtvis ingen större minskning uppnås. Antalet uttagna för 1989 är 600 UE i matpotatis samt 140 i fabrikspotatis.

Enligt simuleringskörningar kommer med den nya metoden 60-70% av brödsädesarealen i huvudurvalet med i snabbstatistiken, medan andelen fodersäd är ca 30-40%.

Som nämndes ovan borde underurvalen dras direkt för hela riket, med tanke på att snabbstatistikens främsta användning är skattningarna på riksnivå. Urvalsramen skulle då kunna sorteras med slumpmässig ordningsföljd inom församlingar varefter församlingarna sorteras slumpmässigt sinsemellan. Emellertid beräknas de ordinarie hektarskörde-skattningarna på riksnivå genom sammanvägning av SKO-skattningar (se avsnitt 9.2). Skattningarna i snabbstatistiken erhålls med samma beräkningsmetod. Urvalen för snabbstatistiken allokteras därför per skördeområde. Allokeringen sker i proportion till skördeområdets totala storleksmått enligt formel (7.1) för respektive gröda. Snabbskattningar på SKO-nivå kan sedan viktas samman till riksskattningar. Medelfelsberäkningar kan då ske på samma sätt som för de ordinarie skattningarna av rikshektarskörden. Hittills har inga löpande beräkningar gjorts av medelfelen i snabbstatistiken.

Ett intressant alternativ beträffande ordningen i urvalsramen för snabbstatistiken vore att ordna skördeområdena efter stigande normskörd av respektive gröda. Det är känt att en systematiskt ordnad urvalsram kan ge betydande precisionsförbättringar om man lyckas ordna ramen efter en känd variabel som är (någorlunda) starkt korrelerad med den variabel man vill skatta. Ett naturligt val av sorteringsvariabel är i vårt fall normskörden. Problemet med en ordnad urvalsram är emellertid att det ofta är svårt att beräkna medelfel. Ansatsen borde trots detta vara värd att studera närmare.

8. Arealskattningarna

8.1 Tilläggsurval för skattningen av arealer

Enligt LES-rapporten angående skördeuppskattningarnas omfattning (se avsnitt 3) är de preliminära arealskattningarna, med publicering i juli, av så stor betydelse att insamlingen av arealuppgifter bör göras i samma omfattning som tidigare vid en nedskärning av skördeuppskattningarna. Samma slutsats dras även beträffande undersökningen av höstsådda arealer, vars resultat publiceras i november. I A/FUs enkät 1987 till statistik-användarna frågades bl a om kraven på säkerheten i arealskattningarna. Åtminstone de främsta användarna av arealskattningarna svarade i enlighet med LES-rapportens slutsatser. Mot denna bakgrund har beslutats att huvudurvalet i spannmål och vall kompletteras med ett tilläggsurval av samma storlek, dvs 4590 UE. För de företag som ingår i tilläggsurvalet skall i första hand uppgifter om arealer insamlas, men urvalet kan vid behov även användas vid olika tilläggsundersökningar. Totalt får man därmed uppgifter från 9180 UE till grund för arealskattningarna, vilket visserligen innebär en nedskärning jämfört med tidigare, dock med endast drygt 20%.

En fördel med att låta tilläggsurvalet vara lika stort som huvudurvalet är att urvalsdragningen blir enkel, genom att man kan "stega fram" i urvalsramen med halverad steglängd och låta varannan UE ingå i respektive urval. Detta ger också maximalt negativ samordning mellan urvalen, vilket här är önskvärt. Trots detta kommer ett antal dubletter till företagen i huvudurvalet att finnas med i tilläggsurvalet. Det kan tyckas vara tillräckligt om arealuppgifter endast inhämtas en gång. Av orsaker som framgår nedan bör man emellertid inhämta uppgifter i båda urvalen för företag som förekommer dubbelt. Beträffande insamlingsförfarandet för uppgifterna i tilläggsurvalet har två alternativ diskuterats, nämligen genom telefonintervjuer eller via lantbruksregistret. Kostnaden för insamling per telefon torde bli alltför hög med tanke på den snäva kostnadsramen (se avsnitt 3). Dessutom ökar uppgiftslämnarbelastningen på jordbrukarna, vilket innebär risk för större bortfall i såväl LBR som i FUs undersökningar. Den enda rimliga metoden är därför att samla in uppgifterna från LBR (se PM A/FU 1989:7). Trots behovet tvingas vi därmed att låta arealskattningarna för höstsådda grödor enbart grundas på uppgifter från företagen i huvudurvalet.

Insamlingen av uppgifterna för tilläggsurvalet kommer i första hand att ske genom förtursbehandling av inkomna LBR-blanketter från urvalsgårdarna. Vid A/LEs ordinarie andra påminnelsetidpunkt, fyra veckor efter uppgiftsdatum, uppskattas bortfallet till ca 700 UE av de uttagna i tilläggsurvalet. En särskild bortfallsuppföljning behövs därför för att snabbare reducera bortfallet. Detta kommer att ske genom telefonpåminnelse från SCB.

En omfattande simuleringsstudie av precisionen i arealskattningarna vid 4500 UE respektive 9000 UE har genomförts, för jämförelse med hittillsvarande precision i skördeuppskattningarna. Det är emellertid svårt att få en rättvisande bild av skillnaderna, eftersom simuleringsresultaten bygger på LBR-arealer och därmed naturligtvis gynnas vid en jämförelse med de "sanna" LBR-arealerna. Arealuppgifterna ifrån A/FUs besöksintervjuer är naturligen behäftade med vissa "svarsfel", dvs uppgifterna stämmer inte överens med arealerna som anges i LBR (se även avsnitt 8.2). Resultaten av jämförelsen redovisas därför inte här. Bedömningen är att man tills vidare bör avvakta med att eventuellt reducera arealundersökningen ytterligare och först studera effekten av tilläggsurvalet.

Eftersom de två olika insamlingsmetoderna kan leda till olika svarsfel är det viktigt att hålla isär de båda urvalen vid skattningen (se vidare avsnitt 8.2). För varje region/gröda beräknas därför en arealskattning ur vardera urvalet, med separata skattningar av regressionsparametrarna. Den slutliga arealskattningen beräknas som medelvärdet av skattningarna från de två urvalen.

Arealskattningarna skall även framdeles gälla samtliga företag större än 2.0 ha. Arealen på de minsta företagen förändras endast måttligt från år till år. Skattningen av grödarealer på företag mellan 2.1 och 5.0 ha kommer därför att ske genom att länsvis använda föregående års LBR-arealer.

8.2 LBR-justeringen av arealskattningarna

Arealskattningarna utgående från uppgifterna i utläggningsprotokollen, de sk primärskattningarna, brukar nivåkorrigeras för att hamna "rätt" i förhållande till LBR. Korrektionen sker genom att beräkna förändringskvoten jämfört med föregående års primärskattning, vilken multipliceras med fjolårets LBR-areal (se PM A/FU 1986:9). Anledningen till justeringen av primärskattningarna är att brukarnas uppgifter till lantbruksregistret i viss mån skiljer sig från arealerna som uppges vid besöksintervjun. Sedan regressionsmetoden infördes i arealskattningarna 1986 har emellertid nivåfelet i primärskattningarna inte närmare studerats. Inklusiv testmaterial föreligger nu underlag för att undersöka nivåfelet under åren 1983-88.

I tabell 5 i tabellbilagan jämförs arealskattningarna på riksnivå med LBRs siffror, dels för primärskattningar, dels för skattningarna med kvotjustering vilka publicerats i Statistiskt Meddelande med preliminär statistik angående åkerarealens användning (kallas i fortsättningen "areal-SM"). I båda fallen har regressionsmetoden använts. I tabellens högra kolumn anges medelvärdet av de absoluta differenserna under de sex studerade åren 1983-88.

Målsättningen bör vara att endast använda kvotjusteringen när den ger en märkbar förbättring av riksskattningarna, dvs reduktion av bias. Metoden med kvotjustering av primärskattningarna medför nämligen att medelfelen måste räknas upp med faktorn roten ur 2. Felet i en skattning uttrycks ibland med medelkvadratfelet (mean squared error, MSE), som definieras enligt formeln $MSE = \text{bias}^2 + \text{medelfel}^2$. Låt oss definiera "bias" enligt följande: Den metod som har minst värde i högra kolumnen i tabell 5 har ingen bias, medan bias för den andra metoden är lika med skillnaden mellan talen i högra kolumnen. Exempelvis skulle den kvotjusterade riksarealskattningen för råg vara unbiased med denna definition, medan den ojusterade primärskattningen har en bias på 600 ha. Sedan skulle följande kriterium kunna användas: Använd kvotjustering i de fall bias enligt ovan-

stående definition reduceras så mycket att MSE minskar, trots det ökade medelfelet. Ett mått på medelfelet vid kvotjustering skulle kunna vara medelvärde av angivna medelfel i areal-SM under 1983-88. Medelfelet vid skattning utan kvotjustering beräknas sedan genom division med roten ur 2. De grödor som kan komma ifråga för kvotjustering med detta kriterium är de som har lägre "bias" vid kvotjustering enligt tabell 5, dvs råg, havre, blandsäd, grönfoder, slåttervall, betesvall, matpotatis och övriga växtslag. I tabell 8.1 anges för dessa grödor de olika "felen" enligt ovan.

Tabell 8.1 "MSE" i riksarealskattningar med och utan kvotjustering

Gröda	Med kvotjustering			Utan kvotjustering		
	"Bias"	"Medelfel"	"MSE"	"Bias"	"Medelfel"	"MSE"
Råg	0	1300	1300 ²	600	900	1100 ²
Havre	0	3600	3600 ²	4400	2600	5100 ²
Blandsäd	0	1700	1700 ²	700	1200	1400 ²
Grönfoder	0	1100	1100 ²	500	800	900 ²
Slåttervall	0	4200	4200 ²	6500	2900	7100 ²
Betesvall	0	2300	2300 ²	26900	1600	26900 ²
Matpotatis	0	900	900 ²	900	600	1100 ²
Övriga växtslag	0	1300	1300 ²	200	900	900 ²

Vid strikt tillämpning av nämnda kriterium skulle kvotjustering utföras för havre, slåttervall, betesvall och matpotatis. För matpotatis är dock skillnaden i MSE liten.

Som komplement till tabellerna 5 och 8.1 kan det vara av intresse att jämföra precisionen i länsarealskattningarna, med och utan kvotjustering. Se tabell 8.2 (matpotatis och övriga växtslag ingår ej eftersom inga separata länsskattningar publiceras för dessa i areal-SM).

Tabell 8.2 Summan av absolutdifferenser i länsarealskattningar jämfört med LBR

Gröda	Med kvotjustering				Utan kvotjustering			
	1988	1987	1986	Medelv	1988	1987	1986	Medelv
Råg	4100	3200	4200	3800	2800	3000	3600	3100
Havre	13000	12700	12600	12800	13900	12500	14100	13500
Blandsäd	7400	4900	4700	5700	5700	4800	5200	5200
Grönfoder	11200	6500	3800	7200	9300	9200	5300	7900
Slåttervall	20800	17200	16700	18200	21700	18000	20400	20000
Betesvall	13300	10800	11600	11900	33400	32000	34000	33100

Ser man till länsskattningarna är det tydligen bara för betesvall som kvotjusteringen ger någon förbättring att tala om. Om man emellertid enbart kvotjusterar betesvall blir skattningen av total åker genomgående för låg. Enligt ovan är slåttervall och havre närmast "i tur" för nivåjustering. Om enbart dessa tre grödor hade nivåjusterats under åren 1983-88 skulle nivåfelet i total åkerareal varit -12800, -6100, +1000, -2100, +3400 samt för år 1988 -13300 ha. Siffrorna kan jämföras med tabell 5 i bilagan.

Överskattningen av betesvallarealen i skördeuppskattningarnas primärskattningar borde till stor del kunna förklaras av att man i LBR har angett någon likartad gröda istället för betesvall: slåttervall, ej utnyttjad vall, grönfoder, träda eller annan obrukad åker. (Även "icke-åkermark" i form av betesmark kan tänkas förväxlas med betesvall). Primärskattningarna av slåttervall är enligt tabell 5 systematiskt underskattade. För övriga

ovan uppräknade grödor är detta emellertid inte fallet, möjligen med undantag för grönfoder. En alternativ metod för justering av primärskattningarna vore därför följande: Skattningen för betesvall kvotjusteras på samma sätt som förut. Den areal som därmed "blir över" fördelas sedan till slåttervall och grönfoder, för justering av primärskattningarna för dessa grödor (endast i undantagsfall justeras betesvallarealen uppåt, varvid slåttervall och grönfoder skulle justeras nedåt). Ur tabell 5 kan underskattningen i slåttervall uppskattas till ca 50% av överskattningen i betesvall, medan motsvarande värde för grönfoder är ca 10%. Primärskattningen i slåttervall och grönfoder skulle alltså kunna korrigeras med -50% respektive -10% av justeringen i betesvall. Resterande 40% kan man eventuellt avstå från att fördela, vilket emellertid leder till en systematisk underskattning av total åkerareal. Eftersom primärskattningen för havre ger en underskattning av arealen kunde man istället korrigera denna med -40% av justeringen i betesvall, varvid primärskattningen av total åkerareal förblir oförändrad. Emellertid kan denna metod knappast ges stöd av att det beträffande havre skulle ske en "förväxling" med betesvall, på liknande sätt som för slåttervall och grönfoder. Istället skulle primärskattningarna för havre, liksom för betesvall, kunna justeras på samma sätt som förut, medan skattningarna för slåttervall och grönfoder korrigeras enligt ovan. Därmed blir primärskattningen av total åkerareal i stort sett oförändrad.

Ovannämnda "omfördelningsmetod" för justering av primärskattningarna samt ytterligare några varianter har studerats närmare i jämförelse med den hittillsvarande metoden. Den alternativa metoden leder inte till någon nämnvärd förbättring av skattningarna och ger även problem vid beräkning av medelfel. Nuvarande metod anses enklare och är därför att föredra.

Av tabellerna 5 och 8.2 framgår att kvotjusteringen av primärskattningarna för betesvall leder till avsevärt förbättrade arealskattningar. Betesvallarealerna enligt LBR visar sig vara ovanligt stabila från år till år, jämfört med andra grödor. Förändringarna i länsarealerna är sällan större än 200 ha. Därför kan det vara intressant att undersöka om fjolårets LBR-areal är en ännu bättre skattning av arealen betesvall än den kvotjusterade primärskattningen i skördeuppskattningarna. För att ta hänsyn till den långsamma minskningen av den totala betesvallarealen kan man avrunda länsarealerna nedåt till närmaste 100-tal ha. Resultatet med denna skattningsmetod för åren 1984-88 framgår av tabell 8.3. Inom parentes anges motsvarande värden vid skattning med kvotjustering, enligt areal-SM.

Tabell 8.3 Differens (ha) gentemot LBR i riksarealskattningar samt absolutdifferenser i länsarealskattningar. Betesvall.

	<u>1988</u>	<u>1987</u>	<u>1986</u>	<u>1985</u>	<u>1984</u>
Riksareal	-8000(+2000) +3600(-3500)	0(+2500)	+600(+2900)	+800(+9700)	
Länsarealer	8000(13300)	4800(10800)	3700(11600)	3100(10700)	3700(11900)

Den alternativa metoden visar sig klart bättre än den nuvarande, särskilt beträffande läns-skattningarna. Riksarealen underskattas dock med 8000 ha år 1988, då arealen betesvall ökade i alla län utom ett, till följd av Omställning 90. Sådana förändringar kan naturligtvis aldrig fångas upp med denna skattningsmetod. Därför är bedömningen att den nuvarande skattningsmetoden för areal betesvall, med kvotjustering av primärskattningarna, bör användas även i fortsättningen.

Mot bakgrund av ovanstående kvotjusteras 1989 primärskattningarna för betesvall, slåttervall och havre. Kvotjusteringen skall ske på samma sätt som tidigare år. För övriga grödor skattas arealen med primärskattningen. Detta gäller skattningarna som bygger på huvudurvalet. Eftersom arealuppgifterna i tilläggsurvalet blir desamma som i LBR, skall arealen för samtliga grödor i detta fall skattas med primärskattningen. Preliminärskattningen av total åkerareal anses ha litet intresse och kommer därför att utgå ur redovisningen.

9. Övrigt

9.1 Varierande antal provytor

Metoden med varierande antal provytor per UE i vissa grödor infördes 1986 och syftar till att öka precisionen i hektarskördeskattningarna. Hektarskörden skattas genom att vikta samman provytemedelskördarna för samtliga UE i området (se även avsnitt 9.2). Skattningsprecisionen kan förbättras genom att ta fler provytor på de företag som "väger tyngst" i skattningen, medan företag med mindre viktfactor får färre provytor. Viktfaktorn utgörs av produkten av det s k uppräkningsstalet och grödarealen på företaget. Uppräkningsstalet är känt när urvalet har dragits, vilket däremot grödarealen naturligtvis inte är. Emellertid kan man studera fördelningen av viktfactorerna under tidigare år. Företagen grupperas efter viktfactor, vilken normeras i förhållande till en "medelviktfactor" (motsvarar vad som tidigare har benämnts medelgrödandel). Denna bör beräknas ur material från ett antal år bakåt, för att gardera mot slumpmässiga svängningar i arealfördelningarna. Därefter kan uppskattas ett undre gränsvärde för relativa viktfactorer, under vilket en tredjedel av företagen hamnar. På samma sätt bestäms ett övre gränsvärde för att avgränsa den övre tredjedelen i fördelningen. (Målsättningen är att medelantalet provytor per UE på riksnivå skall vara 2 för samtliga grödor.) Detta sker för varje skördeområde och gröda, varefter medelvärdet av gränsvärdena beräknas per gröda.

När de slutliga gränsvärdena på detta vis har fastställs kan arealgränserna för 1,2 eller 3 provytor enkelt bestämmas:
undre gränsvärdet · medelviktfactorn ÷ uppräkningsstalet ger arealgränsen mellan 1 och 2 provytor, etc.

Under 1988 var gränsvärdena 0.67 respektive 1.16 för samtliga grödor. För 1989 har gränsvärdena beräknats till 0.55 respektive 1.10 i höstvet och höstråg (gäller endast SKO 1111-1124 samt 1211-1222, i övriga områden utläggs genomgående två provytor per UE), 0.59 respektive 1.17 i korn och havre samt 0.42 respektive 1.21 i slättervall. Den något större spridningen i gränsvärdena jämfört med tidigare återspeglar att variationen i uppräkningsstalen har ökat något.

Skattningen av hektarskörden av vallåterväxt beräknas med samma viktfactorer som för första skörd av vall. Eftersom återväxten ökar i betydelse är det naturligtvis viktigt att även denna skattning har bästa möjliga precision. Därför införs 1989 varierande provyteantal även för bestämningen av vallåterväxten. Antalet provytor blir detsamma som i vall första skörd.

9.2 Hektarskördeskattningarna - utnyttjande av arealuppgifter

Hektarskörden H_k för en gröda i skördeområde k skattas vanligen enligt formeln (Horvitz-Thompson)

$$H_k^* = [\sum_i n_k^{-1} \cdot u_{ki} \cdot a_i \cdot h_i] / [\sum_i n_k^{-1} \cdot u_{ki} \cdot a_i] = T_k^* / A_k^* \quad (9.1)$$

u_{ki} = uppräkningsstal för företag i , SKO k

n_k = antal uttagna UE (korrigerat för bortfall), SKO k

a_i = areal av grödan på företag i

h_i = provytemedelskörd på företag i

T_k^* = skattad totalskörd av grödan, SKO k

A_k^* = skattad areal av grödan, SKO k

När man förutom provyttegårdarna har dragit ett tilläggsurval för insamling av enbart arealuppgifter, skulle man istället för A_k^* kunna använda en annan arealskattning A_k^{**} . Medan A_k^* enbart bygger på arealinformation från provyttegårdarna antas A_k^{**} vara grundad på det totala arealurvalet och därmed ha bättre precision än A_k^* . Trots detta kan det vara bättre att skatta H_k med hjälp av A_k^* . Anledningen är att T_k^* och A_k^* är starkt korrelerade, korrelationskoefficienten är vanligen minst 0.8 enligt tidigare studier. Vid skattning av H_k med $H_k^{**} = T_k^*/A_k^{**}$ är det troligt att vinsten med den bättre precisionen i A_k^{**} inte kan väga upp precisionsförlusten som följer av den lägre korrelationen mellan A_k^{**} och T_k^* . De senare skattningarna bygger ju till hälften på olika urvalsunderlag. Formeln för H_k^{**} kan skrivas om enligt följande: $H_k^{**} = (T_k^*/A_k^*) \cdot (A_k^*/A_k^{**}) = H_k^* \cdot (A_k^*/A_k^{**})$. Detta innebär att H_k^{**} erhålles genom nivåjustering av H_k^* med hjälp av kvoten mellan två arealskattningar. Båda hektarskördeskattningarna kan betraktas som ett viktat medelvärde av provytemedelskördarna på urvalsenheterna. För H_k^{**} är viktsumman (A_k^*/A_k^{**}) , medan för H_k^* vikternas summa är 1, vilket onekligen är mera naturligt. Skattningen H_k^{**} bör därför inte användas.

För skördeområden och på lägre nivåer, i första hand församlingar, skattas hektarskörden med H_k^* , i vissa fall dock med andra värden på u_{ki} och n_k (se avsnitt 9.3). För större områden finns åtminstone två alternativ för skattning av hektarskörden. En skattning i analogi med H_k^* är

$H_{\text{omr}}^* = T_{\text{omr}}^*/A_{\text{omr}}^*$, där skattningarna av T_{omr}^* och A_{omr}^* fås i enlighet med (9.1), genom summering över alla företag i området.

En annan variant ges av

$H_{\text{omr}}^{**} = [\sum_j A_j^{**} \cdot H_j^*] / \sum_j A_j^{**}$, där summeringen sker över alla "byggstenar" j , t ex skördeområden, som området består av.

Den främsta fördelen med H_{omr}^* är att området kan vara uppbyggt "hur som helst", utan att några särskilda beräkningar behöver genomföras för ingående SKO-bitar. Skattningen kan enkelt beräknas för godtyckliga områden, så länge dessa kan identifieras via det sk församlingskartoteket, där områdestillhörigheten anges för varje församling (del av). Emellertid tas ingen hänsyn till arealinformationen från tilläggsurvalet.

I uttrycket för H_{omr}^{**} antas H_j^* vara beräknad enligt (9.1). Vidare antas A_j^{**} bygga på hela arealunderlaget, dvs både huvudurval och tilläggsurval. (Genom att sätta $A_j^{**} = A_j^*$ ser man att H_{omr}^* är ett specialfall av H_{omr}^{**}). H_{omr}^{**} är ett viktat medelvärde av byggstenarnas hektarskördeskattningar. Frågan är hur stor vinsten blir av att använda viktfaaktorena A_j^{**} istället för A_j^* , som i H_{omr}^* . Enligt Gauss' approximationsformel gäller

$$\text{Var}(H_{\text{omr}}^{**}) \approx \sum_j \text{Var}(H_j^*) \cdot p_j^2 + \sum_j \text{Var}(A_j^{**}) \cdot \{(H_j \cdot \sum A_i - \sum A_i \cdot H_j) / (\sum A_i)^2\}^2 \quad (9.2)$$

där $p_j = A_j / \sum A_i$

om vi antar att A_j^{**} och H_j^* är okorrelerade.

Summan till vänster i (9.2) påverkas inte av valet av A_j^{**} , därför behöver vi endast undersöka vad som händer med den högra summan. Man ser att om alla H_j skulle vara lika stora blir högra summan noll. Då har felet i arealskattningen ingen betydelse. Ju större variation mellan H_j :na i området, desto större betydelse får arealskattningsfelet för precisionen i H_{omr}^{**} . Några sifferexempel:

a. Skattning av hektarskörd för korn på riksnivå 1987, utgående från skattningar på riksområden.

Variationen i H_j är stor, från 1700 kg/ha i RO3 till 4200 kg/ha i RO1. $\text{Var}(H_{\text{omr}}^{**})$ kan skattas genom användning av de medelfel som publicerats i SM för 1987 års skörd. Antag

att $\text{Var}(A_j^*) = 2 \cdot \text{Var}(A_j^{**})$, dvs urvalsstorleken skiljer sig med en faktor 2 (bortse från effekten av regression). Då kan medelfelet i H_{rikt}^* uppskattas till 1.03 gånger medelfelet i H_{rikt}^{**} . Annorlunda uttryckt innebär H^{**} att antalet UE kan reduceras med 6% (1.03^2) med bibehållen precision.

b. Skattning av hektarskörd i PO 06 enligt PO18-indelningen. Området motsvarar ungefär länen 1, 3, 4 och 19. Variationen i H_j är måttlig, från 4840 kg/ha till 5320 kg/ha (1987). Förhållandet mellan medelfelen i H_{PO6}^* och H_{PO6}^{**} uppskattas till 1.008, motsvarande en skillnad på ca 1.6% i antal UE.

Exemplen ger en ungefärlig uppfattning om storleken på precisionsvinsten vid användning av H_{omr}^{**} istället för H_{omr}^* . Normalt blir vinsten störst vid skattning på riksnivå, eftersom variationen i H_j är störst i detta fall.

Från 1989 kommer hektarskördarna att skattas med H_{omr}^* för län, produktionsområden och andra större områden, t ex SLR-områden. För rikshektarskördarna kommer skattningarna att ske enligt formeln för H_{omr}^{**} , med skördeområdena som byggstenar. Vid skattning av preliminära totalskördar och arealer kommer självfallet A^{**} att användas för alla områden. I den definitiva statistiken utnyttjas LBR-uppgifter.

9.3 Skattningar för små områden

Även vid skattning av hektarskördar i områden som är betydligt mindre än SKO, exempelvis församlingar, kan formel (9.1) användas. Församlingsskattningar behövs för den historiska databasen, för att möjliggöra omräkningar av hektarskördar vid eventuella framtida ändringar av områdesindelningarna. Man kan annars ifrågasätta om sådana skattningar blir meningsfulla, då underlaget i de flesta fall kommer att utgöras av en eller två observationer. Vid en observation skattas hektarskörden rimligen med provytemedelskörden för urvalsenheten ifråga, vilket även (9.1) ger. När antalet observationer är två eller fler erhålls ett viktat medelvärde av provytemedelskördarna. Precisionen i dessa skattningar blir naturligen låg. En alternativ skattningsmetod som bör övervägas på församlingsnivå är att istället använda det oviktade medelvärdet (se även avsnitt 10).

För små områden som består av flera församlingar bör formel (9.1) korrigeras såtillvida att andra värden på u_k och n_k (k betecknar här församling) bör användas för de ingående församlingarna. Uppräkningstalen u_k beräknas lämpligen genom att dividera totala storleksmättet i församlingen med storleksmättet för företag i och inte, som normalt sker, utgående från skördeområdets totala storleksmätt. Samtidigt ändras n_k till att ange totalt antal UE med observationer i församling k . Detta kan innebära en hel del extra arbete men ger i gengäld ökad skattningsprecision i områden med litet antal observationer. Den viktigaste eventuella tillämpningen torde vara vid skattning på kommunnivå.

10. Förslag till uppföljning och vidare utredningar

I detta avsnitt anges kortfattat vad som närmast kan vara angeläget att studera vidare.

- Uppföljning av utfallet av arealskattningarna. För dessa skattningar har man ju den ovanliga möjligheten att i efterhand kunna jämföra med "facit", dvs LBR. Speciellt viktigt blir det att studera utfallet av arealskattningarna från tilläggsurvalet. Dessutom bör man kontinuerligt studera behovet av nivåjustering i

primärskattningarna.

- b. En systematisk översyn av samtliga medelfelsberäkningar i skördeuppskattningarna.
- c. Undersök möjligheten att skatta hektarskördar genom lika viktning av provytemedelskördarna. Denna fråga har tagits upp flera gånger förut, men någon omfattande numerisk studie har aldrig genomförts. Metoden ger lägre varianser, men samtidigt är det stor risk att man får systematiska fel. Det nya storleksmåttet ger en viss ökning av variationen i uppräkningsstalen, jämfört med tidigare (se avsnitt 9.1). Därmed förloras något av den precisionsökning man får i de lågfrekventa grödorna, till följd av fler observationer. Detta gör ansatsen extra intressant.
- d. Skattning av hektarskördar på församlingsnivå genom lika viktning. Denna fråga kan eventuellt studeras för sig, om punkt c anses innebära en alltför stor arbetsinsats. I det här fallet bör skattningsmetoden utvidgas med en metod för korrektion som skall ge konsistens med skattningarna på högre nivåer.
- e. Beräkning av normskördar med uppdelning på företagsstorlek. Förslaget har väckts från några lantbruksnämnder, med tanke på nämndernas normkalkyler.
- f. Uppföljning av utfallet av snabbstatistiken i jämförelse med den preliminära skördestatistiken. Eventuell justering av konstanterna i formel (7.1).
- g. Studera utfallet vid användning av individuella korrektionsfaktorer för bl a obärgad areal i skattningarna. För närvarande beräknas korrektionstalen på områdesbasis. Korrektion på UE-nivå ger beräkningsmässiga fördelar, varför man bör undersöka hur skattningsprecisionen påverkas.
- h. Studera möjligheten att använda ordnad (efter normskörd) urvalsram för snabbstatistiken.
- i. Uppföljning av resultaten för den i kapitel 6.3 angivna alternativa metoden för arealskattningar i potatis.

11. Referenser

1. Bayless, D.L. & Rao, J.N.K., An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling ($n=3$ or 4), J. Amer. Statist. Assoc. **65**, 1645-1667 (1970).
2. Brewer, K.R.W. & Hanif, M., Sampling with unequal probabilities (Springer 1983).
3. Cochran, W.G., Sampling Techniques third edition (Wiley 1977).
4. Nyström, L., Skördeuppskattning då urvalssannolikheten beror på både totalareal och grödans areal (Umeå universitet, matematisk-statistiska institutionen 1984).
5. PM AOS 1974:1, Översyn av urvals- och skattningsmetoder vid de objektiva skördeuppskattningarna.
6. PM A/FU 1978:41, Beräkning av urvalen för olika delundersökningar i de objektiva skördeuppskattningarna.
7. PM A/FU 1979:32, Maskinell fördelning av underurval.
8. PM A/FU 1984:15, Om variansen för hektarskördeskattningarna i matpotatis.
9. PM A/FU 1986:9, Regler för justering av förändringskvoter.
10. PM A/FU 1986:14, Utredning av urvals- och skattningsmetoderna i de objektiva skördeuppskattningarna (URVASK).
11. PM A/FU 1988:17, Potatisarealer 1988.
12. PM A/FU 1988:28, Beräkningar av kostnadsutvecklingen i fältarbetet.
13. PM A/FU 1989:7, Arealskattningarna 1989.
14. Rao, J.N.K. & Bayless, D.L., An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum, J. Amer. Statist. Assoc. **64**, 540-549 (1969).
15. Utredning om de objektiva skördeuppskattningarnas omfattning. Lantbruksekonomiska samarbetsnämndens sektorsgrupp 1982.
16. Områdesindelningar i lantbruksstatistiken 1989. Meddelanden i samordningsfrågor 1989:1. SCB.

TABELLBILAGA

Tabell 1 Allokering av 1989 års huvudundersökning i spannmål och vall
Total åkerareal per skördeområde enligt LBR 1988

<u>SKO</u>	<u>Antal UE</u>	<u>Total åkerareal</u>	<u>SKO</u>	<u>Antal UE</u>	<u>Total åkerareal</u>
0111	50	37767	1411	40	18094
0112	46	32627	1412	43	26829
0311	54	60424	1421	41	18422
0312	43	26459	1511	45	39180
0321	54	55680	1512	46	33237
0322	39	11988	1521	41	18907
0411	50	38953	1522	40	20963
0421	51	39979	1611	39	10587
0422	50	27592	1612	45	29578
0431	49	40060	1613	45	27048
0511	40	13249	1614	46	26400
0512	41	17893	1615	48	26604
0513	54	47782	1616	52	36942
0514	54	51365	1617	40	14499
0515	54	48439	1621	54	44829
0521	45	25490	1622	54	55367
0611	40	16722	1623	54	35462
0612	43	25649	1711	42	24898
0621	41	22946	1712	43	31087
0622	44	34511	1713	40	21272
0711	41	21239	1721	38	13838
0731	38	14045	1722	42	22865
0811	54	39295	1723	37	7694
0812	45	26117	1724	37	7019
0813	41	14771	1811	42	18533
0814	40	18086	1812	54	51472
0821	48	35510	1813	39	14395
0831	41	21569	1821	48	30361
0911	44	18826	1911	54	46258
0912	54	45048	1912	43	23964
0913	42	18643	1921	54	46572
1011	48	28726	1922	44	26418
1111	54	30743	2011	46	43894
1112	46	23074	2012	39	15247
1121	54	42663	2111	44	35244
1122	49	31357	2121	42	26508
1123	54	35710	2122	37	6929
1124	41	17852	2211	39	15896
1131	38	15722	2212	40	23703
1211	54	46316	2221	41	22917
1212	51	29344	2311	41	25352
1213	51	25437	2312	38	11359
1214	54	36512	2331	37	5195
1215	54	55368	2411	40	22114
1216	54	48410	2412	38	10098
1221	41	19868	2413	40	20591
1222	48	24554	2414	40	20300
1311	51	39509	2415	36	4749
1321	53	41681	2511	40	23634
1322	38	12359	2512	38	15002
1331	45	29819	2521	36	3670

Tabell 2 Potatisarealer i de nya skördeområdena enligt LBR 1987
Gäller företag med minst 5.1 ha åker och minst 0.5 ha mat-/fabrikspotatis

SKO	Areal matpotatis, ha		Areal fabrikspotatis, ha	
	Med bev	Utan bev	Med bev	Utan bev
0111	24	30	0	0
0112	43	6	0	0
0311	51	59	0	0
0312	53	118	0	0
0321	132	69	0	0
0322	17	13	0	0
0411	29	71	0	0
0421	31	30	0	0
0422	62	27	0	0
0431	27	46	0	0
0511	13	6	0	0
0512	7	6	0	0
0513	745	250	0	0
0514	160	82	0	0
0515	24	34	0	0
0521	114	56	0	0
0611	13	7	0	0
0612	80	35	0	0
0621	176	54	9	13
0622	37	14	2	2
0711	3	30	7	21
0731	4	3	18	26
0811	143	79	261	200
0812	79	185	2	2
0813	264	60	8	0
0814	23	18	0	0
0821	5	0	0	0
0831	10	11	18	9
0911	30	129	0	0
0912	126	48	0	0
0913	39	235	0	0
1011	88	92	1452	1267
1111	762	51	2687	734
1112	94	39	430	510
1121	288	36	308	122
1122	13	43	71	141
1123	366	191	2	0
1124	215	51	50	30
1131	1	4	0	1
1211	147	5	5	0
1212	110	10	47	8
1213	258	71	320	125
1214	299	52	3	3
1215	498	580	106	216
1216	75	78	4	0
1221	26	33	19	85
1222	1974	408	1	0
1311	394	528	12	5
1321	1621	671	50	41
1322	69	23	0	0
1331	9	38	0	0
1411	5	37	0	0

<u>Tabell 2</u> <u>SKO</u>	(forts) <u>Areal matpotatis, ha</u>		<u>Areal fabrikspotatis, ha</u>	
	<u>Med bev</u>	<u>Utan bev</u>	<u>Med bev</u>	<u>Utan bev</u>
1412	14	29	0	0
1421	3	90	1	3
1511	44	35	0	0
1512	58	43	0	0
1521	17	12	0	0
1522	5	5	0	0
1611	270	36	23	14
1612	278	210	21	25
1613	88	69	6	0
1614	175	90	37	19
1615	424	141	20	28
1616	91	74	10	2
1617	78	20	0	1
1621	342	271	0	12
1622	161	244	0	13
1623	278	368	14	21
1711	239	26	0	0
1712	0	1	0	0
1713	34	28	0	0
1721	11	4	0	0
1722	142	77	0	0
1723	39	36	0	0
1724	81	91	0	0
1811	52	80	0	4
1812	382	505	0	82
1813	61	48	0	0
1821	28	35	0	0
1911	1	20	0	0
1912	0	9	0	0
1921	1	38	0	0
1922	42	30	0	0
2011	120	200	0	0
2012	135	141	0	0
2019	0	0	0	0
2111	41	90	0	0
2121	121	72	0	0
2122	15	4	0	0
2211	20	47	0	0
2212	18	51	0	0
2221	67	75	0	0
2311	30	78	0	0
2312	18	28	0	0
2319	1	9	0	0
2331	3	22	0	0
2411	115	220	0	0
2412	0	8	0	2
2413	119	113	0	0
2414	28	72	2	1
2415	42	49	0	0
2419	0	1	0	0
2511	205	222	0	0
2512	72	193	0	0
2519	0	1	0	0
2521	0	12	0	0

Tabell 3 Allokering av 1989 års huvudundersökningar i potatis

<u>SKO</u>	<u>Antal UE</u> <u>matpotatis</u>	<u>Antal UE</u> <u>fabrikspotatis</u>	<u>SKO</u>	<u>Antal UE</u> <u>matpotatis</u>	<u>Antal UE</u> <u>fabrikspotatis</u>
0111	7		1411	4	
0112	5		1412	4	
0311	13		1421	7	
0312	20		1511	8	
0321	22		1512	10	
0322	3		1521	4	
0411	9		1522	2	
0421	6		1611	35	
0422	8		1612	60	
0431	12		1613	12	
0511	2		1614	36	
0512	1		1615	60	
0513	60		1616	19	
0514	29		1617	13	
0515	7		1621	60	
0521	23		1622	39	
0611	3		1623	60	
0612	13		1711	33	
0621	25		1712	0	
0622	6		1713	3	
0711	4		1721	1	
0731	0	1	1722	23	
0811	30	26	1723	13	
0812	32		1724	19	
0813	38		1811	16	
0814	4		1812	60	
0821	0		1813	9	
0831	1		1821	8	
0911	19		1911	3	
0912	29		1912	1	
0913	32		1921	4	
1011	25	70	1922	11	
1111	60	70	2011	38	
1112	21	49	2012	32	
1121	51	24	2111	16	
1122	7	10	2121	21	
1123	60	0	2122	2	
1124	39	3	2211	9	
1131	0	0	2212	7	
1211	23	0	2221	19	
1212	15	2	2311	14	
1213	44	24	2312	5	
1214	44	0	2331	3	
1215	60	15	2411	41	
1216	19	0	2412	1	
1221	4	5	2413	27	
1222	60	1	2414	13	
1311	60		2415	11	
1321	60		2511	52	
1322	7		2512	24	
1331	5		2521	1	

Tabell 4 Ny SPO-indelning, anpassad till de nya skördeområdena

SPO

1	SKO 1011, 1111, 1112, 1121, 1122, 1123, 1211, 1212, 1213, 1214, 1215, 1216, 1222, 1311, 1321
2	0811, 0812, 0814, 0821, 0911, 0912, 0913
3	0611, 0621, 1124, 1131, 1221, 1322, 1511, 1611
4	0511, 0612, 0622, 0711, 0731, 0813, 0831
5	0411, 0421, 0422, 0431, 0512, 0513, 0514, 0515, 1812, 1821
6	0111, 0112, 0311, 0312, 0321, 0322, 1911, 1921, 1922
7	0521, 1617, 1723, 1811, 1813, 1912
8	1512, 1612, 1613, 1614, 1615, 1616, 1621, 1622, 1623, 1711, 1712, 1722
9	1331, 1411, 1412, 1421, 1521, 1522
10	1713, 1721, 1724
11	2012
12	2011, 2121, 2122
13	2111, 2212, 2221
14	2211, 2331
15	2019, 2311, 2312, 2319
16	2411, 2412, 2413, 2414, 2415, 2419
17	2511, 2512, 2519, 2521

Tabell 5 Differenser i riksarealer mellan primärskattning och LBR. Inom parentes anges differenserna mellan arealerna enligt areal-SM (LBR-justerade) och LBR.

<u>Gröda</u>	<u>1988</u>	<u>1987</u>	<u>1986</u>	<u>1985</u>	<u>1984</u>	<u>1983</u>	Medelvärde av differensernas absolutbelopp
Höstvete	+1900(-1200)	+3300(+400)	+1700(+1600)	0(-2900)	+1400(-1800)	+1700(-2600)	1700(1800)
Vårvete	+700(-200)	+1900(+3100)	-1300(-3300)	+1800(+2800)	-2200(-3700)	-1000(-800)	1500(2300)
Råg	-500(+1400)	-2000(-1700)	-1400(+700)	-2300(-1400)	-2100(-100)	-1400(-700)	1600(1000)
Korn	-1300(-1000)	+1600(+7400)	-6700(-9200)	+3100(+5500)	-4300(-2500)	-4300(+2100)	3600(4600)
Havre	-4300(+200)	-1300(+9000)	-11300(+100)	-10500(-3200)	-13100(+5300)	-13200(-9900)	9000(4600)
Blandsäd	+1100(+3500)	+700(+1700)	-300(-2000)	+4100(+3900)	+6300(+400)	+5400(+2500)	3000(2300)
Baljväxter	-700(+2400)	-400(+1300)	-1000(+1500)	-1600(+1100)	-3700(-400)	-2500(-3200)	1700(1700)
Höstraps	-2400(-1300)	-2300(-2400)	+200(+800)	-300(-1000)	-300(-300)	+400(+200)	1000(1000)
Vårrops	+700(+3000)	-1600(+600)	+700(+1100)	+1000(+4400)	-200(-1300)	-300(+6300)	800(2800)
Höstrybs	0(-1100)	+700(+1200)	-300(-900)	-300(-600)	-100(-100)	-400(+600)	300(800)
Vårrys	+2200(+2600)	0(-400)	+900(+1300)	-100(-2100)	+2000(+4300)	-300(-2100)	900(2100)
Grönfoder	+4300(+6200)	-1300(+2800)	-2600(-1000)	-2200(+2700)	-5000(-800)	-2500(+1600)	3000(2500)
Slåttervall	-21500(-11300)	-8900(+6500)	-16700(-1600)	-16100(+1100)	-15100(-12800)	-2500(+8500)	13500(7000)
Betesvall	+30700(+2000)	+27200(-3500)	+33300(+2500)	+31200(+2900)	+33200(+9700)	+28700(-2200)	30700(3800)
Matpotatis	-1300(+200)	-1500(+400)	-1600(+900)	-2000(-400)	-700(+800)	-1900(+1000)	1500(600)
Fabrikspotatis	+300(-200)	+500(+200)	+500(-100)	+600(+300)	+100(-400)	+800(+1700)	500(500)
Övriga växtslag	+1200(+200)	+1200(-300)	+1500(+700)	+1700(-200)	+1200(-1800)	+2200(+4400)	1500(1300)
Träda	-4100(-5400)	-4300(-12900)	+3900(+6400)	-2000(-3600)	-1600(-1800)	+8000(+6000)	4000(6000)
Vall, ej utnyttjad	-7900(-6700)	0(+1900)	-300(-1600)	+1100(+500)	+800(-100)	-500(0)	1800(1800)
Annan obr. åker	-200(+5000)	-5400(-5600)	+1700(+2700)	-2000(-2100)	-400(+9000)	-13000(-8600)	3800(5500)
Total åker	+700(-1600)	+8400(+9700)	+2200(+1500)	+4800(+7700)	-3300(+900)	+3800(+4400)	3900(4300)

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

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

Reports published earlier during 1989 are:

- | | |
|-------------------|---|
| 1989:1
(grön) | Går det att mäta produktivitetsutvecklingen för SCB?
(Rune Sandström) |
| 1989:2
(grön) | Slutrapporter från U-avdelningens översyn av HINK och KPI (flera författare) |
| 1989:3
(grön) | A Cohort Model for Analyzing and Projecting Fertility by Birth Order (Sten Martinelle) |
| 1989:4
(beige) | Abstracts I - Sammanfattningar av metodrapporter från SCB |
| 1989:5
(gul) | On the use of Semantic Models for specifying Information Needs (Erik Malmberg) |
| 1989:6
(grön) | On Testing for Symmetry in Business Cycles (Anders Westlund och Sven Öhlén) |
| 1989:7
(grön) | Design and quality of the Swedish Family Expenditure Survey (Håkan L Lindström, Hans Lindkvist och Hans Näsholm) |
| 1989:8
(grön) | Om utnyttjande av urvalsdesignen vid regressionsanalys av surveydata (Lennart Nordberg) |
| 1989:9
(grön) | Variations in the Age-Pattern of Fertility in Sweden Around 1986 (Michael Hartmann) |
| 1989:10
(grön) | An Application of Generalized Precision Functions in the 1985 Swedish Family Expenditure Survey (Håkan L Lindström and Peter Lundquist) |
| 1989:11
(gul) | Statistics production in the 90's - decentralization without chaos (Bo Sundgren) |

Kvarvarande BEIGE och GRÖNA exemplar av ovanstående promemorior kan rekvireras från Elisabet Klingberg, U/STM, SCB, 115 81 Stockholm, eller per telefon 08-783 41 78.

Dito GULA exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 Stockholm, eller per telefon 08-783 41 47.