

Översyn av
Undersökningen av lastbilstransporter i
Sverige (UVAV)

Bengt Rosén och Mahnaz Zamani



R&D Report
Statistics Sweden
Research - Methods - Development
1993:2

INLEDNING

TILL

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

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

Föregångare:

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

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

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

Efterföljare:

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

R & D Report 1993:2. Översyn av Undersökningen av lastbilstransporter i Sverige (UVAV) / Bengt Rosén, Mahnaz Zamani.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

Översyn av
Undersökningen av lastbilstransporter i
Sverige (UVAV)

Bengt Rosén och Mahnaz Zamani



R&D Report
Statistics Sweden
Research - Methods - Development
1993:2

Från trycket
Producent
Ansvarig utgivare
Förfrågningar

Januari 1993
Statistiska centralbyrån, utvecklingsavdelningen
Lars Lyberg
Bengt Rosén, 08-783 44 90

© 1993, Statistiska centralbyrån
ISSN 0283-8680

SCB Tryck. Örebro 1993.01 Miljövänligt papper

Översyn av Undersökningen av lastbilstransporter i Sverige (UVAV)

Abstract: The report gives an account of a major revision of the sampling design in the survey "Transports by Swedish Lorries in Sweden", commonly referred to by the acronym UVAV. In particular, it provides a new design which is expected to yield considerable improvement of the survey efficiency. Theoretical and technical issues are treated in a separate appendix. The appendix presents, in addition to technical aspects specific to the revision, an extensive account of UVAV's estimation procedures and their rationale.

INNEHÅLLSFÖRTECKNING

	Sida
1. Inledning	1
2 UVAV-undersökningarna och statistiken från dem	2
2.1 Centrala Bilregistret	
2.2 UVAV-statistikens målstorheter	
2.3 Genomförande av UVAV-undersökningar	
2.4 Kilometerskattregistret	
3 Premisser för UVAVs framtida utformning	3
3.1 Angående transportvariabler	
3.2 Angående redovisningsgrupper	
3.3 Angående redovisningsperiod	
3.4 Angående EG-anpassning av UVAV-statistiken	
3.5 Sammanfattning	
4 Genomförande av översynen	4
4.1 Empiriskt material, beräkningar, mm	
4.2 Belysning av, och beslut i "EG-frågor"	
4.3 Översyn av stratifiering och urvalsallokering	
4.3.1 Utgångspunkter och vägledande principer	
4.3.2 Studerade stratifieringar och allokeringar	
4.3.3 Optimeringsförfarande och resultat	
5 Rekommenderad ny utformning av UVAVs urvalsförfarande	10
5.1 Lastbilspopulation, urvalsram, mm	
5.2 Stratifiering	
5.3 Urvalsallokering	
5.4 Estimation, skattningsprecision	
Appendix. Teoretiska och tekniska aspekter på översynen	
A1 Om UVAVs estimationsproblematik	16
A1.1 Återföring av estimationsproblem för års- och kvartals- aggregat på kvartalsvisa stratumaggregat	
A1.2 Urvals- och observationsförfarandet i en kvartalsundersökning	
A1.3 Estimation baserad på urvalsbeskrivningen "OSU med disjunkta delurval"	
A1.4 Estimation baserad på urvalsbeskrivningen "OSU av lastbilsveckor"	
A2 Estimationsprecision vid alternativa undersökningsutformningar	21
A3 Optimering av stratifiering och urvalsallokering	23
A3.1 Några allmänna aspekter på optimeringsproblematiken	
A3.2 Möjliga stratifieringar i UVAV	
A3.3 Restriktioner avseende tillgängliga undersökningsresurser	
A3.4 Stegen i sökandet efter en optimal utformning	
A4 OSU med disjunkta delurval	25
Referenser	28

Översyn av "Undersökningen av Lastbilstransporter i Sverige" (UVAV)

1. Inledning

SCBs enhet för statistik över Servicenärings, F/SE, utför återkommande undersökningar av varutransporter inom landet, bl a de s k UVAV-undersökningarna, vilka avser transporter med svensk lastbil inom Sverige. Statistiken från UVAV omfattar framför allt skattningar av aggregat för variabler som mäter transportarbete vid lastbilskörning; transporterad kvantitet, transportsträcka, presterade tonkm och fraktintäkt. Aggregaten betraktas med olika gruppindelningar, varvid indelning efter slag av transporterad vara är en av de viktigaste, och aggregering görs över såväl kvartal som (kalender)år.

UVAV startade 1972, och under perioden 1972 - 1987 genomfördes undersökningen varje år. För drygt fem år sedan beslöts, framför allt av besparingsskäl, att UVAV skulle genomföras mindre ofta. Sedan 1987 gäller periodiciteten "vart tredje år". I anslutning till periodicitetsändringen framförde dock huvudanvändare av transportstatistik önskemål om fortsatt årlig statistikredovisning, åtminstone av särskilt intressanta årsaggregat. Mot den bakgrunden utarbetades ett förfarande för årlig framskrivning av vissa av UVAVs årsaggregat, och framskrivningar har gjorts sedan 1988. För framskrivningssyftet togs Kilometer-skattregistret för första gången i bruk för statistikändamål.

I enlighet med treårsperiodiciteten gjordes en UVAV-undersökning under 1990. Nästa UVAV skall genomföras under 1993, och inför den undersökningen bestämdes att göra en översyn av undersökningens utformning, särskilt avseende dess urvals- och estimationsförfarande. Nedan anges huvudskäl för att en översyn bedömdes vara önskvärd.

- Vid arbetet med framskrivningarna vanns fördjupade insikter i den hjälpinformation som står UVAV till buds, och dessa insikter förmodades kunna användas för att effektivisera undersökningen. (1.1)
- Transportstatistiken berörs av EG/EES-anpassning av Sveriges statistik. (1.2)
- UVAVs utformning har legat orörd under förhållandevis lång tid, medan "transportverkligheten" har undergått avsevärda förändringar; lastbilar har blivit genomsnittligt större, transportverksamheter har ändrat struktur och organisation, mm. Detta ger anledning att pröva ändamålsenligheten i UVAVs hittillsvarande utformning. (1.3)

Den tekniska delen av översynen har genomförts av Bengt Rosén och Mahnaz Zamani. För Mahnaz handlade det om att utföra examensarbete på matematikerlinjen/statistikinriktning vid Stockholms universitet. Från F/SE biträdde UVAV-ansvariga, Glenn Sahlberg och Kerstin Forssén, vilka också utgjorde referens- och beslutsgrupp för översynen.

2 UVAV-undersökningarna och statistiken från dem

I det följande beskrivs UVAV i stort, så som senare års undersökningar varit utformade.

2.1 Centrala Bilregistret

Trafiksäkerhetsverkets Centrala Bilregister (CBR) har alltid spelat en central roll för UVAV, som urvalsram, som uppgiftskälla och som främsta källa för hjälpinformation. CBR förtecknar, med **registreringsnummer** som identifierare, alla svenska (icke-militära) motorfordon, och det innehåller information om de enskilda fordonen. För lastbilar är CBR-informationen förhållandevis rikhaltig, och omfattar bl a lastbilens **maxlastvikt** (**MAXLV**, = maximalt tillåtna lastvikt) och **karosserikod**. Den senare ger upplysning om bilens typ, t ex om den är en tankbil, en timmerbil, etc. Vidare ger CBR information om en lastbils **ägarkategori**, som avser om bilen används för **yrkesmässig trafik** (= körningar

för andras räkning mot ersättning) eller **firmatrafik** (= körningar för "egen firmas" räkning), huruvida bilen är (skattemässigt) **aktivregistrerad** eller ej.

2.2 UVAV-statistikens målstorheter

Ett centralt objekt i UVAV-undersökningen är (med hittillsvarande definition);

LASTBIL = lastbil i CBR med MAXLV 2 ton eller mer (med undantag för några få karosserikoder).

De för UVAV-statistiken fundamentala objekten är **KÖRNINGar** enligt nedan;

KÖRNING = körning med eller utan last i Sverige, utförd av **LASTBIL**.

För **KÖRNINGar med last** är nedanstående **körnings/transportvariabler** av huvudintresse;

KVANT = pålastad kvantitet under körningen,
KÖRKMMED = körningens sträcka,
TONKM = genom körningen presterade tonkilometrar,
FRAKTINK = fraktinkomst från körningen.

De nämnda aggregeringarna avser ovanstående variabler. Aggregering över olika delgrupper av körningar är av intresse, och en central gruppindelning görs med hjälp av variabeln

VARUSLAG = det transporterade godsets varuslag.

VARUSLAG klassificeras enligt CTSE (= Commodity Classification for Transport Statistics in Europe), som omfattar cirka 20 huvudgrupper av varor.

Vidare görs indelning med avseende på nedanstående variabel **YMARK**, vilken i första omgången relaterar till **LASTBIL**, men därigenom också till **KÖRNING**;

YMARK = kod för alternativen "yrkesmässig trafik", betecknas **YRKM**, eller firma-
trafik, **FIRM**. Endast kategorierna **YRKM** och **FIRM** föreligger. Bortseende
från ägarkategori betecknas med **OAVSett**.

UVAVs främsta statistiska målstorheter är **aggregat** av typen;

$$AGG(X;V;\ddot{A};P) = \text{summa } X\text{-värden för körningar med last i varugrupp } V \\ \text{utförda av lastbilar med ägarkategori } \ddot{A} \text{ under period } P, \quad (2.1)$$

som erhålls när $X, V, \ddot{A}$ och P varierar enligt nedan;

- X är någon av variablerna **KVANT**, **TONKM**, **KÖRKMMED** eller **FRAKTINK**,
- V är någon av CTSEs varugrupper,
- $\ddot{A}$ är någon av ägarkategorierna **YRKM**, **FIRM** eller **OAVS**,
- P är ett kvartal eller ett år. (2.2)

Förändringstal från ett år till ett annat ges av kvoter mellan motsvarande aggregat.

En ytterligare indelning av transportaggregat, som vissa transportstatistikanvändare är mycket intresserade av, gäller indelning efter körningarnas geografiska ändpunkter, och sådan statistik åberopas som **flödesstatistik**. UVAV redovisar flödesstatistik mellan län, men är återhållsam med mer finindeldad redovisning trots att efterfrågan finns, av huvudskälet att sådan statistik skulle bli mycket otillförlitlig. (Redan på länsnivå är flödesstatistiken klart otillförlitlig för många länspar.)

2.3 Genomförande av UVAV-undersökningar

I stora drag har UVAV-undersökningarna utförts på följande sätt. För varje kvartal i undersökningsåret görs en **kvartalsundersökning**, baserad på ett sannolikhetsurval av

aktivregistrerade LASTBILAR. Som urvalsram används ett utdrag ur CBR, taget cirka två månader före undersökningskvartalet. Lastbilsurvalet dras som ett stratifierat urval där stratifieringen ges av en korsindelning med avseende på landets 24 län och två MAXLV-klasser; 2-8 ton ("lätta" lastbilar) respektive 8 ton eller mer ("tunga" lastbilar). Vid 1990 års UVAV var urvalsstorleken 4000 lastbilar per kvartal, vilken fördelades på stratan med hjälp av en variant av Neyman-allokering. Till varje utvald lastbil tillordnas med slumpens hjälp en **mätvecka** bland de 13 veckorna i undersökningskvartalet. Undersökningskvartalen är definierade så att de omfattar precis 13 veckor.

Insamlingen av observationer går till på följande sätt. Uppgifter om utvalda LASTBILAR inhämtas främst från CBR, men också från CFAR och lastbilens ägare. Uppgifter om de utvalda lastbilarnas KÖRNINGAR, särskilt avseende variablerna KVANT, TONKM, KÖRKM-MED, FRAKTINK och VARUSLAG, insamlas med hjälp av en **körjournal**, som förs av lastbilens förare under mätveckan. Journalen skall omfatta varje körning, med eller utan last, under mätveckan. Urvalet av KÖRNINGAR är ett stratifierat klusterurval.

De insamlade uppgifterna ligger sedan till grund för skattningar av värdena på de statistiska målstorheterna; aggregat, förändringstal, mm. I första omgången görs skattningar för undersökningskvartalen. Efter undersökningsårets slut ger materialet från alla fyra kvartalen underlag för skattning av årsaggregat och annan årsstatistik.

2.4 Kilometerskattregistret

Information från Kilometerskattregistret spelar en instrumentell roll som hjälpinformation för de tidigare nämnda framskrivningarna. Det skulle vara fördelaktigt att kunna utnyttja den hjälpinformationen även för att öka precisionen i skattningarna från reguljära UVAV-undersökningar. Dock, som diskuteras utförligare längre fram har detta tyvärr bedömts vara omöjligt, men trots det sägs några ord om registret här. Kilometerskattregistret (KMR) handhas, liksom CBR, av Trafiksäkerhetsverket. Dess primära uppgift är att ge underlag för drivmedelsbeskattning av dieseloljedrivna fordon. Med jämna mellanrum, ungefär kvartalsvis, skall sådana fordon infinna sig hos myndighet och stämpla av en särskild kilometermätare, och avstämplingarna bokförs i KMR på bilarnas registreringsnummer. Litet förenklat kan man säga att informationen i KMR gör det möjligt att för varje enskild diesel driven lastbil beräkna bilens totala körsträcka under ett kalenderår. Bilarna i UVAVs LASTBILSpopulation, bilar med MAXLV ≥ 2 ton, är till helt övervägande del dieseloljedrivna och de omfattas till nära 100% av KMR.

3 Premisser för UVAVs framtida utformning

Kriteriet för eventuell ändring av UVAVs utformning efter översynen har varit det gängse. En alternativ utformning är att föredra om den leder till bättre skattningsprecision. Liksom flertalet av SCBs undersökningar, har UVAV **flera syften**. Det förhållandet komplicerar, som bekant, när man skall välja målfunktion i optimeringsstudier. I samråd med UVAV-ansvariga formulerades nedanstående huvudpunkter för viktiga aspekter på framtida UVAV-statistik.

3.1 Angående transportvariabler

KVANT, TONKM, KÖRKMED och FRAKTINK är alla klart intressanta variabler, men KVANT och TONKM hålls som viktigare än de andra, med ungefär samma vikt för båda.

3.2 Angående redovisningsgrupper

Det är av klart intresse att redovisa separat statistik för yrkes- respektive firmatrafik.

CTSEs cirka 20 varugrupper är många, och sedda ur transportsynvinkel mycket olikstora.

Alla tillmäts inte samma vikt. Som viktigast hålls, utöver "total"-gruppen ALLAV de i Tabell 1 preciserade varugrupperna LIVSM, RVIRKE, GRUS, och OLJA. De utgör de transportvolymmässigt fyra största varugrupperna. Vidare är de särskilt intressanta ur den synvinkeln att de återspeglar utvecklingen i olika branscher av ekonomin.

Varugrupp	CTSE-kod	Akronym
Livsmedel	3	LIVSM
Rundvirke (för lekmannen = timmer)	51	RVIRKE
Grus, jord, sten och sand	71	GRUS
Mineraloljor och produkter därav	12	OLJA
Samtliga varuslag	Alla	ALLAV

Tabell 1: Ur transportsynvinkel särskilt viktiga varugrupper.

3.3 Angående redovisningsperiod

Det centrala utflödet från UVAV-undersökningen är årsstatistiken, medan kvartalsstatistiken mer betraktas som "preliminär statistik".

3.4 Angående EG-anpassning av UVAV-statistiken

I enlighet med EES-avtalet, skall Sveriges statistik anpassas till EGs inom ett antal områden, inklusive transportstatistiken. En klar skillnad mellan EGs krav på transportstatistik och hittillsvarande UVAV-statistik är följande. EG föreskriver statistik för transporter med lastbilar med MAXLV $\geq 3,5$ ton, medan UVAVs hittillsvarande statistik gällt transporter med lastbilar med MAXLV ≥ 2 ton. UVAV har avgränsat sin LASTBILspopulation med tanke på statistik för ">2-tons-transporter", och den avgränsningen medger också redovisning enligt EG-normen. UVAV-ansvariga anser att för framtiden är redovisning enligt EG-normen mycket angelägen, medan redovisning enligt gamla normen är mindre angelägen.

Vidare ställer EG långtgående krav på flödesstatistik för transporter; kvartalsvis redovisning med NUTS III-indelning av på- och avlastningsregioner. UVAV har ingen principiell invändning mot kraven, som ju faktiskt harmonierar med redan tidigare framförda användarönskemål, men i skrivelse av 1992-07-16 till Finansdepartementet har UVAV framfört betänkligheter vad gäller att till rimliga kostnader kunna möta kraven med någorlunda tillförlitlig flödesstatistik.

3.5 Sammanfattning

- De viktigaste transportvariablerna är KVANT och TONKM. (3.1)
- Uppdelning efter YRKM och FIRM är viktig. (3.2)
- De viktigaste varugrupperna är LIVSM, RVIRKE, GRUS, OLJA samt ALLAV. (3.3)
- Årsaggregaten är viktigare än kvartalsaggregaten. (3.4)
- Transportstatistik avseende lastbilar med MAXLV $\geq 3,5$ ton är angelägen, statistik för bilar med MAXLV ≥ 2 ton är mindre angelägen. (3.5)
- EG ställer långtgående krav på flödesstatistik för transporter. (3.6)

4 Genomförande av översynen

Gången i översynsarbetet beskrivs översiktligt i detta avsnitt. Mer teoretiska och tekniska delar har förts till Appendix.

4.1 Empiriskt material, beräkningar, mm

Som källa för empirisk information användes materialet från 1990 års UVAV, omfattande mikromaterialen från de fyra kvartalsundersökningarna. Datafilerna transformerades till SAS-dataset, och beräkningarna genomfördes på PC med hjälp av programvaran SAS.

I ett första steg utarbetades program för punkt- och variansskattningar för transportaggregat. För att få en hanterlig optimeringssituation gjordes inskränkning ifråga om varugrupper. Med (3.3) som bakgrund valdes att ta fram skattningar för ALLAV, LIVSM, RVIRKE, GRUS och OLJA. Dessutom för GÖDSEL (CTSE-kod=6), som exempel på en liten varugrupp, vilken dock utelämnas i de flesta av de följande sammanställningarna.

Bergeppet **felmarginal** var centralt, och används med följande innebörd. Låt D stå för standardavvikelse och $\hat{D}$ för skattad standardavvikelse. Den teoretiska felmarginalen för den (approximativt) väntevärdesriktiga estimatoren $\hat{\theta}$ av parametern θ definieras som $2 \cdot D(\hat{\theta})$, och den empiriska felmarginalen som $2 \cdot \hat{D}(\hat{\theta})$. Analogt, den teoretiska relativa felmarginalen är $2 \cdot D(\hat{\theta})/\theta$, och den empiriska relativa felmarginalen är $2 \cdot \hat{D}(\hat{\theta})/\hat{\theta}$. (Dvs, felmarginal = dubbla standardavvikelsen och relativ felmarginal = dubbla variationskoefficienten.) I fortsättningen görs för det mesta inte distinktion mellan teoretiska och empiriska felmarginaler.

4.2 Belysning av, och beslut i "EG-frågor"

Tabell 2 redovisar skattningar av årsaggregat, med felmarginaler, såväl enligt hittillsvarande UVAV-norm som enligt EG-norm, dvs med aggregering över körningar med bilar med MAXLV ≥ 2 ton respektive $\geq 3,5$ ton.

Resultaten var litet överraskande. Lastbilar med MAXLV mellan 2 och 3,5 ton utgör antalsmässigt cirka 15% av LASTBILSbeståndet. Som framgår av Tabell 2, bidrar dessa bilar dock litet till det aggregerade transportarbetet, mellan 0 och några % beroende på variabel och varugrupp. I urvalet till 1990 års UVAV utgjorde lastbilar av ifrågavarande slag cirka 8%. Med (3.5) och det nyss anförda som bakgrund, ansågs att de undersökningsresurser, som hittills ägnats lastbilar med MAXLV 2 - 3,5 ton, utnyttjas mer effektivt om de satsas på bilar med MAXLV $\geq 3,5$ ton. UVAV-ansvariga fattade följande beslut.

Framtida UVAV-undersökningar skall avse lastbilar med MAXLV $\geq 3,5$ ton. (4.1)

Förändring i enlighet med (4.1) kommer naturligtvis att medföra vissa tidsseriebrott, men faktiskt mycket små. Som framgår av Tabell 2, är skillnaderna mellan ">2-ton-aggregat" och ">3,5-ton-aggregat" så gott som försumbara, särskilt om urvalsfelen tas med i bilden.

Härnäst en annan "EG-fråga". EG-samordning avseende statistik påbjuds redan av EES-avtalet. I och med att Sverige aspirerar på att bli medlem i EG, pågår också (frivillig) samordning i andra avseenden. En samordningspunkt gäller drivmedelsbeskattning av dieselfordon. Sveriges nuvarande system med körsträcke-baserad beskattning avviker från EGs, och används bara i Sverige och Norge. När översynen genomfördes hade riksdagsproposition aviserats om att ersätta kilometerskatten med en ny skatt på dieselolja. Om byte av beskattningsform sker, kommer Kilometerskattregistret att läggas ned. Det bedömdes vara för riskabelt att räkna med att kilometerskattregistrets information skulle finnas tillgänglig i framtiden, och följande beslut togs;

UVAVs urvals- och skattningsförfarande skall utformas under premissen att Kilometerskattregistrets information inte finns att tillgå. (4.2)

Om Kilometerskattregistret läggs ner, kommer inte heller de tidigare nämnda framskrivningarna att kunna göras. Sett ur statistiksynvinkel kompenseras detta dock av att

Skattade värden på årsaggregat för 1990, med felmarginaler, avseende transportvariablerna KVANT, TONKM, KÖRKMMED och FRAKTINK. Skattningar redovisas för aggregat över körningar av lastbilar med MAXLV ≥ 2 ton och MAXLV $\geq 3,5$ ton. Sort: "Obestämd".

Varugrupp	Ägar-kategori	MAXLV	KVANT	Fel-marg	TONKM	Fel-marg	KÖR-KMMED	Fel-marg	FRAKT-INK	Fel-marg
ALLAV	OAVS	≥ 2 ton	3881	133	2652	81	1782	42		
		$\geq 3,5$ ton	3855	131	2642	81	1691	41		
	YRKM	≥ 2 ton	3073	122	2300	80	1258	36	2083	52
		$\geq 3,5$ ton	3067	120	2297	80	1231	36	2039	51
	FIRM	≥ 2 ton	808	69	351	26	524	29		
		$\geq 3,5$ ton	788	68	344	26	460	27		
LIVSM	OAVS	≥ 2 ton	344	26	470	37	346	21		
		$\geq 3,5$ ton	337	26	467	37	321	20		
	YRKM	≥ 2 ton	195	21	337	34	173	15	263	21
		$\geq 3,5$ ton	193	21	337	34	170	15	257	21
	FIRM	≥ 2 ton	149	17	133	16	173	15		
		$\geq 3,5$ ton	143	16	131	16	152	14		
RVIRKE	OAVS	≥ 2 ton	345	42	258	29	75	8		
		$\geq 3,5$ ton	345	42	258	29	75	8		
	YRKM	≥ 2 ton	317	40	238	28	69	8	137	16
		$\geq 3,5$ ton	317	40	238	28	69	8	137	16
	FIRM	≥ 2 ton	28	11	20	8	6	2		
		$\geq 3,5$ ton	28	11	20	8	6	2		
GRUS	OAVS	≥ 2 ton	1561	110	232	19	127	8		
		$\geq 3,5$ ton	1560	110	232	19	127	8		
	YRKM	≥ 2 ton	1280	95	206	19	105	8	265	17
		$\geq 3,5$ ton	1280	95	206	19	105	8	265	17
	FIRM	≥ 2 ton	281	21	26	4	22	3		
		$\geq 3,5$ ton	281	21	26	4	22	3		
OLJA	OAVS	≥ 2 ton	186	21	133	16	59	6		
		$\geq 3,5$ ton	186	21	133	16	58	6		
	YRKM	≥ 2 ton	159	19	115	15	49	6	114	13
		$\geq 3,5$ ton	159	19	115	15	49	6	114	13
	FIRM	≥ 2 ton	28	10	18	6	9	2		
		$\geq 3,5$ ton	28	10	18	6	9	2		

Tabell 2: Aggregatskattningar, med felmarginaler, för körningar av lastbilar om MAXLV ≥ 2 och $\geq 3,5$ ton.

EG kräver årliga transportundersökningar. SCB har därför beslutat att UVAV åter skall utföras årligen, troligen f o m 1995. Därmed försvinner behovet av framskrivningar.

När denna rapport skrivs har riksdagsbeslut tagits om att kilometerskatten avskaffas från och med 1 oktober 1993.

4.3 Översyn av stratifiering och urvalsallokering

Det fortsatta översynsarbetet gällde framför UVAVS stratifiering och urvalsallokering, med sikte på förbättrad, och naturligtvis helst optimal, utformning.

4.3.1 Utgångspunkter och vägledande principer

För att kunna bedömma konsekvenserna av eventuella ändringar, behövs en lämplig referenspunkt. Med (4.1) och (3.1)-(3.5) som bakgrund, valdes följande som det centrala referensfallet: De skattningsprecisioner för årsaggregat enligt "EG-norm" som 1990 års UVAV leder till. Dessa redovisas i Tabell 3.

Relativa felmarginaler i % för årsaggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK med aggregering över körningar av lastbilar med MAXLV $\geq 3,5$ ton. Population, stratifiering och urvalsallokering: Enligt UVAV 1990. Urvalsstorlek: 4000 lastbilar per kvartal.					
Varugrupp	Ägarkategori	KVANT	TONKM	KÖRKMMED	FRAKTINK
ALLAV	OAVS	3,40	3,05	2,44	2,49
	YRKM	3,87	3,47	2,90	
	FIRM	8,64	7,64	5,95	
LIVSM	OAVS	7,63	7,91	6,11	8,15
	YRKM	10,90	10,01	8,56	
	FIRM	10,82	12,42	9,06	
RVIRKE	OAVS	12,06	11,38	11,12	11,58
	YRKM	12,65	11,87	11,65	
	FIRM	40,60	41,58	38,42	
GRUS	OAVS	6,89	8,13	6,23	6,35
	YRKM	7,38	8,98	7,12	
	FIRM	19,18	16,03	13,05	
OLJA	OAVS	11,53	12,30	10,64	11,02
	YRKM	12,14	13,16	11,69	
	FIRM	35,06	35,20	26,45	

Tabell 3: Relativa felmarginaler för "EG-årsaggregat" på basis av 1990 års UVAV.

Härnäst diskuteras den bild av UVAVs skattningsprecision som framkommer ur Tabell 3. Liksom för flertalet av SCBS undersökningar, gäller att UVAV dels har flera syften (variabler, redovisningsgrupper, målstorheter) dels är utsatt för användarönskemål om "längre gående uppdelning av statistiken". Det går då inte att ge generella omdömen av typen att skattningsprecisionen är "tillfredsställande" eller "icke tillfredsställande". Det finns hela tiden ytterligare statistikönskemål, som inte kan tillgodoses på grund av att den efterfrågade statistiken skulle bli för otillförlitlig, vilket exemplifieras i det följande.

De relativa felmarginalerna i Tabell 3, avser årsaggregat för de stora varugrupperna. För mindre varugrupper är dock precisionsbilden en annan, och då naturligtvis sämre, vilket illustreras i Tabell 4, som gäller varugruppen GÖDSEL.

Relativa felmarginaler i % för årsaggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK med aggregering över körningar av lastbilar med MAXLV $\geq 3,5$ ton. Population, stratifiering och urvalsallokering: Enligt UVAV 1990. Urvalsstorlek: 4000 lastbilar per kvartal.					
Varugrupp	Ägarkategori	KVANT	TONKM	KÖRKMMED	FRAKTINK
GÖDSEL	OAVS	48,1	32,3	30,3	33,7
	YRKM	50,2	33,0	31,4	
	FIRM	77,9	88,8	74,1	

Tabell 4: Relativa felmarginaler för årsaggregat, på basis av 1990 års UVAV.

Analogt, för kvartalsaggregat är precisionsbilden en annan än i Tabell 3, vilket framgår av Tabell 5. Eftersom urvalen i de fyra kvartalen var lika stora gäller (4.3) nedan, vilket också illustreras av innehållen i Tabellerna 5 och 3.

Felmarginaler, absoluta och relativa, för kvartalsaggregat är ungefär dubbelt så stora som för motsvarande årsaggregat. (4.3)

Relativa felmarginaler i % för kvartal-1-aggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK med aggregering över körningar av lastbilar med MAXLV $\geq 3,5$ ton. Population, stratifiering och urvalsallokering: Enligt UVAV 1990. Urvalsstorlek: 4000 lastbilar per kvartal.					
Varu-grupp	Ägarkategori	KVANT	TONKM	KÖRKMMED	FRAKTINK
ALLAV	OAVS	6,8	5,9	4,5	5,0
	YRKM	8,1	6,6	5,8	
	FIRM	15,2	15,9	10,1	
LIVSM	OAVS	14,3	15,3	11,6	16,0
	YRKM	20,3	18,7	16,5	
	FIRM	20,5	27,4	17,0	
RVIRKE	OAVS	19,5	19,0	18,7	19,1
	YRKM	20,3	19,7	19,5	
	FIRM	71,4	69,1	65,0	
GRUS	OAVS	13,2	16,0	12,4	13,0
	YRKM	14,8	17,7	14,3	
	FIRM	31,7	36,1	26,3	
OLJA	OAVS	24,7	25,9	22,8	24,1
	YRKM	26,8	27,8	24,9	
	FIRM	62,8	72,8	58,7	

Tabell 5: Relativa felmarginaler för kvartalsaggregat på basis av 1990 års UVAV.

Vidare, för förändringstal förstoras felmarginaler med ungefär faktorn $\sqrt{2}$. Som tidigare nämnts, föreligger också krav/önskemål om långt nedbruten flödesstatistik, vilket ytterligare skulle fresta på de relativa felmarginalerna.

Tabell 3 tas därför inte som utgångspunkt för någon absolut bedömning av UVAVs skattningsprecision, utan används i första omgången bara för att peka på vissa mindre tillfredsställande drag i UVAVs precisionsbild. Först formuleras ett par allmänna principer för bedömning av en urvalsundersöknings skattningsprecision.

När variabler betraktas som olika viktiga, bör man sträva efter större skattningsprecision, ju viktigare en variabeln anses vara. (4.4)

Man bör eftersträva att skattningar för redovisningsgrupper i samma indelningskategori får någorlunda lika god skattningsprecision. (4.5)

Nedan kommenteras de skattningsprecisioner som redovisas i Tabell 3.

Aggregaten för KVANT och TONKM har större relativa felmarginaler än de för KÖRKMMED och FRAKTINK, vilket mot bakgrund av (3.1) och (4.4) är mindre tillfredsställande. (4.6)

Skattningsprecisionen för FIRM är markant sämre än för YRKM, vilket mot bakgrund av (3.2) och (4.5) är mindre tillfredsställande. (4.7)

Skattningarna för RVIRKE och OLJA är klart sämre än för LIVSM och GRUS, vilket mot bakgrund av (3.3) och (4.5) är mindre tillfredsställande. (4.8)

Obalanserna (4.6) - (4.8) utgjorde, tillsammans med det allmänna önskemålet om generell förbättring av skattningsprecisionen, utgångspunkter för sökandet efter bättre stratifiering och urvalsallokering. Innan vi går närmare in på det sökandet, formuleras ett par allmänna stratifieringsprinciper.

Stratifiering är fördelaktig om man uppnår "stratumvis homogenisering" med avseende på viktiga undersökningsvariabler. (4.9)

Stratifiering efter viktiga redovisningsgrupper är fördelaktig, eftersom en sådan ger möjlighet att "styra urvalet redovisningsgruppsvis". (4.10)

4.3.2 Studerade stratifieringar och allokeringar

Hittills har UVAV stratifierats genom en korsindelning LÄN×MAXLV. Indelningen efter LÄN homogeniserar med avseende på regionala variationer. Dessutom befrämjar "styrd regional utspridning" av lastbilsurvalet möjligheten till någorlunda god länsvis flödestatistik. Variabeln MAXLV är markant korrelerad med samtliga körningsvariabler; KVANT, TONKM, KÖRKMMED och FRAKTINK. Lämplig stratifiering efter MAXLV leder därför till homogeniseringar, som medför variansreduktion för i stort sett alla aggregat.

Nedan anges de stratifieringsaspekter som beaktades vid översynen.

Stratifiering efter LÄN: Tidigare stratifiering ifrågasattes inte.

Stratifiering efter ägarkategori: Stratifiering efter YMARK studerades, och det fanns två goda skäl att tro att sådan stratifiering skulle vara fördelaktig, varav det ena var följande. Trots att FIRM-bilarna antalsmässigt utgör cirka 50% av LASTBILSbeståndet, producerar de endast 15 à 20% av transportarbetet. Mot den bakgrunden kunde man förvänta att stratumindelning efter YMARK, skulle leda till mer homogena stratan, vilket i sin tur skulle medföra förbättrad skattningsprecision för OAVS-aggregat. Det andra skälet, ges av (3.2), (4.7) och (4.10).

Särskild stratifiering för RVIRKE och OLJA: Här kom en av lärdomarna från framskrivningsarbetet till pass. Det otillfredsställande förhållande som pekas på i (4.8) bottnar i att transporter av RVIRKE och OLJA utförs av specialiserade, "högproduktiva" lastbilar, vilka dock är förhållandevis få och därför svagt representerade i urvalet. De aktuella lastbilarna kan identifieras med hjälp av den karosserikodvariabel (KARKOD) som finns tillgänglig i CBR. Till transporter av RVIRKE korresponderar s k BANKEBILAR (KARKOD=61) och till transporter av OLJA korresponderar TANKBILAR (KARKOD=40, 41, 42, 45). BANKE- och TANK-bilarna är dock så få, att de inte medger uppdelning efter LÄN. Därför betraktades "riksstratan" för dessa lastbilstyper.

Stratifiering med hjälp av MAXLV: I hittillsvarande UVAV-stratifiering har indelning gjorts i två MAXLV-klasser; bilar mellan 2 och 8 ton samt bilar med MAXLV ≥ 8 ton. Varför just två klasser valdes, och varför gränsen mellan "lätta" och "tunga" bilar sattes till 8 ton, går dessvärre inte att få veta, eftersom ingen dokumentation finns bevarad från senaste stratifieringsomläggningen. Därför prövades MAXLV-stratifieringen förutsättningslöst.

4.3.3 Optimeringsförfarande och resultat

Nedan ges en kortfattad redogörelse av gången i sökandet efter en optimal, eller åtminstone någorlunda optimal, stratifiering. Ytterligare beskrivning ges i Appendix.

Vi etablerade formler och datorprogram för att med hjälp av 1990 års UVAV-material

skatta relativa felmarginaler för de centrala aggregaten (= de i Tabell 3) under en allmän undersökningsutformning ifråga om stratifiering och urvalsallokering. Därigenom kunde olika tänkbara utformningar jämföras med avseende på resulterande skattningsprecision. Efter omfattande sökande, med relativa felmarginaler som bedömningskriterium, framstod en viss stratifiering med tillhörande urvalsallokering som bästa möjliga inom den ram för tänkbara utformningar som anammats. Nedan redovisas de "huvudparametrar" som varierades under sökandet efter optimum, och de val som slutligen framstod som de bästa. Den resulterande urvalsutformningen anges detaljerat i nästa avsnitt.

YMARK-stratan: Infördes, och visade sig vara så fördelaktiga som förmodats.

MAXLV-stratan: Såväl antalet indelningar efter MAXLV som **stratumgränserna** varierades, med resultatet att man i storstadslänen lämpligen använder tre MAXLV-klasser, men nöjer sig med två i övriga län. Rekommenderade stratumgränserna anges i nästa avsnitt.

Stratan av BANKEBILar och TANKBILar: Infördes, och visade sig fördelaktiga.

Val av variabel för bestämning av urvalsallokering: När intresset gäller statistiska storheter för flera olika variabler, är det inte självklart vilken variabel man skall ta fasta på när man söker efter (nästan)optimal urvalsallokering. Det val som visade sig leda till den bästa "totala precisionsbilden", var att ta 50-50%-medelvärde av Neyman-allokeringar med avseende på KVANT och TONKM.

Bestämning av de s k urvalsproportionalerna (se Avsnitt 5): När man, som här, med hjälp av ett föreliggande observationsmaterial gör "retrospektiv" analys av vilken urvalsutformning som skulle ha varit optimal, föreligger risk för "glädjekalkyl" genom överanpassning till de föreliggande observationerna. Förhoppningsvis har stordelen av sådant "glädjekalkylerande" elimineras genom att de s k variationsproportionalerna beräknades som medelvärden över 1990 års fyra kvartal. Detta förfarande leder, rimligen, också till bättre framtida allokering än vad en "överanpassning" till 1990 års observationer skulle göra. Vidare prövades, men förkastades, huruvida variationsproportionalerna borde beräknas utan hänsyn till länsindelning.

Fördelning av urvalsstorlek mellan YRKM-stratan och FIRM-stratan: Varierades med resultat enligt nästa avsnitt.

Fördelning av årsurvalsstorleken på kvartal: Varierades, med resultatet att det bästa är att hålla fast vid hittillsvarande ordning med samma urvalsstorlek i varje kvartal.

Minsta urvalsstorlek i ett stratum: Varierades, med resultat enligt nästa avsnitt.

5 Rekommenderad ny utformning av UVAVs urvalsförfarande

5.1 Lastbilspopulation, urvalsram, mm

UVAVs lastbilspopulation avgränsas fortsättningsvis på följande sätt;

LASTBIL = lastbil i CBR med maxlastvikt $\geq 3,5$ ton, och med någon av karoseri-koderna: 00-21, 27-47, 49-70, 72, 73, 77, 79, 80, 83-94, 96-98.

Som tidigare genomförs en årsgång av UVAV medelst separata urvalsundersökningar för de fyra undersökningskvartalen, som vart och ett består av 13 (mät)veckor. LASTBILsurvalen för olika kvartal dras oberoende av varandra. Vid urvalsdragningarna utgörs urvalsramen av en ur Centrala Bilregistret (CBR) framselektad urvalsram med alla aktivregistrerade LASTBILar. Urvalsramarna upprättas så nära före undersökningskvartalet som möjligt, vilket f n innebär cirka två månader före. Urvalen dras som **stratifierade urval** med OSU-urval inom stratan, med stratifiering enligt nedan.

5.2 Stratifiering

Huvudstrukturen i stratifieringen av LASTBILarna i urvalsramen ges av en korsindelning av typen LÄN × YMARK-klass × MAXLV-klass. Från den görs undantag såtillvida att BANKEBILar och TANKBILar utgör särskilda "riksstratan", för vilka varken LÄNsindelning eller MAXLV-indelning görs. Specifikation av stratumindelningen ges av följande steg.

Steg 1: LASTBILarna i urvalsramen delas upp efter **karosserikod** (KARKOD) i nedanstående tre huvudgrupper;

BANKEBIL = LASTBIL med KARKOD 61,

TANKBIL = LASTBIL med KARKOD 40, 41, 42 eller 45,

FLERTBIL = LASTBIL med KARKOD som inte är 61, 40, 41, 42 eller 45.

Steg 2: Grupperna BANKEBILar, TANKBILar och FLERTBILar delas upp efter **yrkeskategori** i undergrupperna **YRKESBILar** (YMARK=1) och **FIRMABILar** (YMARK=0).

För BANKEBILar och TANKBILar utgör de så erhållna fyra grupperna urvalsstratan. För FLERTBIL fortsätter uppdelningen.

Steg 3. Fortsatt uppdelning av FLERTBILar: FLERTBILarna delas upp efter landets alla 24 **län** (LÄN). Inom varje LÄN×YMARK-grupp indelas efter **maxlastvikt** (MAXLV). Den indelningen görs olika för å ena sidan **storstadslänen**, Stockholm, Malmö och Göteborg län, och å andra sidan för **länen utanför storstäderna**.

I länen utanför storstäderna (= länskod skild från 01, 12 och 14) görs indelningen efter MAXLV enligt nedan:

(FLERTBIL; YRKM, LÄTT) = YRKESBILar med 3,5 ton ≤ MAXLV < 11,5 ton,

(FLERTBIL; YRKM, TUNG) = YRKESBILar med 11,5 ton ≤ MAXLV,

(FLERTBIL; FIRM, LÄTT) = FIRMABILar med 3,5 ton ≤ MAXLV < 10,5 ton,

(FLERTBIL; FIRM, TUNG) = FIRMABILar med 10,5 ton ≤ MAXLV.

I storstadslänen (= länskod 01, 12 eller 14) görs indelningen enligt nedan:

(FLERTBIL; YRKM, LÄTT) = YRKESBILar med 3,5 ton ≤ MAXLV < 7,5 ton,

(FLERTBIL; YRKM, MEDEL) = YRKESBILar med 7,5 ton ≤ MAXLV < 12,5 ton,

(FLERTBIL; YRKM, TUNG) = YRKESBILar med 12,5 ton ≤ MAXLV,

(FLERTBIL; FIRM, LÄTT) = FIRMABILar med 3,5 ton ≤ MAXLV < 8,0 ton,

(FLERTBIL; FIRM, MEDEL) = FIRMABILar med 8,0 ton ≤ MAXLV < 13,0 ton,

(FLERTBIL; FIRM, TUNG) = FIRMABILar med 13,0 ton ≤ MAXLV.

De så bildade lastbilsgrupperna utgör urvalsstratan.

Ovan införda indelning av LASTBILarna i urvalsramen leder till **totalt 106 urvalsstratan**. FLERTBILarna utanför storstäderna delas upp i 21 (län) × 2 (yrkesm/firma) × 2 (MAXLV-klasser) = 84 stratan, och FLERTBILarna i storstadslänen delas upp i 3 (län) × 2 (yrkesm/firma) × 3 (MAXLV-klasser) = 18 stratan. Därtill kommer 4 stratan från BANKEBILar och TANKBILar uppdelade i vardera två YMARK-klasser.

5.3 Urvalsallokering

Det förutsätts att en **total urvalsstorlek** för LASTBILar för det aktuella undersökningsåret är fastställd. Också urvalsallokeringen genomförs i steg.

Steg 1: Årsurvalsstorleken fördelas lika på de fyra undersökningskvartalen.

Steg 2: Kvartalsurvalsstorleken fördelas på grupper av urvalsstratan enligt Tabell 6.

	YRKESBILar	FIRMABILar
FLERTBILar	56,0 %	30,0 %
BANKEBILar	5,0 %	1,7 %
TANKBILar	4,0 %	3,3 %

Tabell 6: Fördelning av urvalsstorleken på stratumgrupper.

Därmed är urvalsallokeringen för BANKEBILarna och TANKBILarna gjord.

Kommentar: Vid kvartalsurvalsstorlek 4000 medför allokering enligt Tabell 6 följande: 268 BANKEBILar utväljs; varav 200 (75%) är YRKESBILar och 68 (25 %) FIRMABILar. 292 TANKBILar utväljs, varav 160 (55%) är YRKESBILar och 132 (45 %) FIRMABILar. ■

Steg 3. Urvalsallokering för FLERTBILar: Urvalsallokeringen sker separat för stratumgrupperna (FLERTBIL; YRKESBIL) och (FLERTBIL; FIRMABIL). Inom var och en av dessa används "restrifierad Neyman-allokering" enligt nedan.

Restrifierad Neyman-allokering:

En population med H disjunkta stratan, numrerade $h = 1, 2, \dots, H$, föreligger. Låt

N_h = antalet objekt i stratum h , $h = 1, 2, \dots, H$,

κ_h = ett angivet positivt tal, kallat **variationsproportional**, $h = 1, 2, \dots, H$,

n = en föreskriven (total) urvalsstorlek.

Med (vanlig) **Neyman-allokering** avses att den totala urvalsstorleken fördelas på stratan enligt formeln;

$$n^*_h = n \cdot (N_h \cdot \kappa_h) / \sum_r (N_r \cdot \kappa_r), \quad h = 1, 2, \dots, H.$$

Restrifierad Neyman-allokering med minsta urvalsstorlek q erhålls på följande sätt. Först görs en Neyman-allokering. För alla stratan i vilka $n^*_h \leq q$ sätts urvalsstorleken till q . Sedan Neyman-allokeras den återstående urvalsstorleken på resterande stratan, dvs de stratan som i första omgången fick urvalsstorlek $> q$. Om något stratum i andra omgången får en urvalsstorlek som är mindre än q , höjs den upp till q .

Kommentar: Det kan inträffa att urvalsstorlekar i andra omgången blir mindre än q , men det händer förhållandevis sällan, och när det sker blir urvalsstorleken normalt endast obetydligt mindre än q . Den slutliga, "upphöjda" totala urvalsstorleken kan därför bli något större än den föreskrivna. För UVAV kommer det att handla om 0 till 5 "extra" lastbilar i ett kvartalsurval av storlek 4000. ■

Vid tillämpning av en restrifierad Neyman-allokering måste man specificera stratumstorlekar, variationsproportionaler och minsta urvalsstorlek.

Stratumstorlekarna erhålls genom **räkning i urvalsramen**.

Minsta urvalsstorleken i ett stratum sätts till $q = 13$ vid kvartalsurvalsstorlek 4000. Detta värde reduceras i samma proportion som urvalsstorleken eventuellt reduceras. (Anm. Valet $q = 12$ är aningen bättre än $q = 13$, men det senare valet är att föredra med tanke på att ett kvartal omfattar 13 mätveckor.)

Variationsproportionaler anges i Tabell 7 nedan.

LÄN	Variationsproportionaler för YRKESBILAR				Variationsproportionaler för FIRMABILAR		
	LÄTTA lastbilar	MEDEL lastbilar	TUNGA lastbilar		LÄTTA lastbilar	MEDEL lastbilar	TUNGA lastbilar
01	15	55	195		20	35	110
03	30		175		25		80
04	70		165		15		100
05	40		180		25		155
06	85		165		20		155
07	45		185		20		145
08	55		145		20		130
09	50		170		70		115
10	85		195		30		100
11	70		185		15		45
12	30	105	175		15	45	125
13	50		200		20		120
14	15	85	175		10	30	105
15	85		160		15		80
16	85		180		30		95
17	40		180		20		85
18	65		170		20		60
19	35		160		25		200
20	90		250		20		90
21	70		215		30		95
22	45		220		15		95
23	15		200		10		90
24	40		230		15		130
25	50		210		15		205

Tabell 7: Variationsproportionaler för FLERTBIL-stratana.

5.4 Estimation, skattningsprecision

Med den ovan föreslagna urvalsutformningen utförs estimationen på basis av erhållna observationer i allt väsentligt som tidigare. Närmare beskrivning ges i Appendix.

Tabell 9 redovisar (skattade värden för) de relativa felmarginaler som skulle ha erhållits i 1990 års UVAV om den ovan föreslagna urvalsutformningen använts. Värdena i den tabellen skall också uppfattas som prognos för den skattningsprecision som kommer att uppnås vid 1993 års UVAV om förslaget följs, och urvalsstorleken är som 1990, dvs 4000 lastbilar per kvartal. Prognosen vilar på antagandet att "strukturella förhållanden" för lastbilstransporter är ungefär desamma under 1993 som de var 1990. Tabell 9 redovisar också de (estimator)variansreduktioner som uppnås med den föreslagna utformningen jämfört med 1990 års utformning. Variansreduktionerna har beräknats enligt formeln;

Variansreduktion =

$$= 1 - (\text{rel. felmarginal med ny utformning} / \text{rel. felmarginal enligt Tabell 3})^2.$$

Variansreduktion relaterar till "utökning av urvalsstorleken med bibehållen urvalsutformning", enligt vad som specificeras i Tabell 8.

Variansreduktion	2%	5%	10%	15%	20%	25%	30%
Ekvivalent ökning av urvalsstorleken	2%	5%	11%	18%	25%	33%	43%
Variansreduktion	35%	40%	45%	50%	60%	70%	80%
Ekvivalent ökning av urvalsstorleken	54%	67%	82%	100%	150%	233%	400%

Tabell 8: Relation mellan variansreduktion och utökning av urvalsstorleken med oförändrad urvalsutformning.

Relativa felmarginaler i % för årsaggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK. Den redovisade variansreduktionen (var.red) är räknad relativt skattningsprecisionen i Tabell 3. Lastbilspopulation: Bilar med 3,5 tons MAXLV eller mer. Stratifiering: Enligt Avsnitt 5. Urvalsallokering: Enligt Avsnitt 5. Urvalsstorlek: 4000 lastbilar per kvartal.									
Varu-grupp	Ägarkategori	KVANT	Var.-red	TONKM	Var.-red	KÖRKMMED	Var.-red	FRAKTINK	Var.-red
ALLAV	OAVS	2,85	30%	2,54	31%	2,18	20%	1,88	43%
	YRKM	3,24	30%	2,83	33%	2,42	30%		
	FIRMA	6,02	51%	5,08	56%	4,73	37%		
LIVSM	OAVS	6,57	26%	6,71	28%	5,87	8%	7,68	11%
	YRKM	9,52	24%	8,69	24%	7,99	13%		
	FIRMA	8,55	38%	8,66	51%	8,66	9%		
RVIRKE	OAVS	7,12	65%	6,34	69%	6,10	70%	6,38	70%
	YRKM	7,51	65%	6,69	68%	6,45	69%		
	FIRMA	21,82	71%	18,24	81%	17,27	80%		
GRUS	OAVS	5,87	27%	6,93	27%	5,62	19%	5,63	21%
	YRKM	6,49	23%	7,63	28%	6,26	23%		
	FIRMA	13,80	48%	12,94	35%	12,63	6%		
OLJA	OAVS	7,91	53%	7,29	65%	6,60	61%	7,79	50%
	YRKM	8,87	47%	7,97	63%	7,26	61%		
	FIRMA	15,80	80%	17,59	75%	15,79	64%		

Tabell 9: Relativa felmarginaler och variansreduktioner med ny urvalsutformning.

Nedan formuleras huvudkonklusioner från Tabell 9.

- Den föreslagna urvalsutformningen leder till avsevärd förbättring av skattningsprecisionen, mer totalt sett variansreduktioner av storleksordningen 30-40%.
- För varugrupperna RVIRKE och OLJA är variansreduktionerna särskilt markanta, av storleksordningen 60-70%. Den obalans som pekas på i (4.8) elimineras.
- Skattningsprecisionerna för YRKM och FIRM är mer lika än tidigare, vilket mildrar obalansen enligt (4.7). De skulle kunna föras ännu närmare varandra, särskilt för RVIRKE och OLJA, men det skulle ske till priset av klar försämring av skattningsprecisionerna för OAVS-aggregat. (Ett sådant försämringpris har redan betalats.)
- Skattningsprecision har "flyttats" från KÖRKMMED-aggregat till aggregat avseende KVANT och TONKM, vilket är respons på invändningen (4.6).

Den skattningsprecision som erhålls med den föreslagna urvalsutformningen, beror naturligtvis på vilken urvalsstorlek som används. Tabellerna 10 och 11 ger prognoser för de relativa felmarginalerna om urvalsstorleken minskas med 10% respektive 20%.

Relativa felmarginaler i % för årsaggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK. Den redovisade variansreduktionen (var.red) är räknad relativt skattningsprecisionen i Tabell 3. Lastbilspopulation: Bilar med 3,5 tons MAXLV eller mer. Stratifiering: Enligt Avsnitt 5. Urvalsallokering: Enligt Avsnitt 5. Urvalsstorlek: 3600 lastbilar per kvartal.									
Varu-grupp	Ägar-kategori	KVANT	Var.-red	TONKM	Var.-red	KÖRKMMED	Var.-red	FRAKTINK	Var.-red
ALLAV	OAVS	3,01	22%	2,68	23%	2,30	11%	1,99	36%
	YRKM	3,41	22%	2,98	26%	2,56	22%		
	FIRMA	6,35	46%	5,35	51%	5,00	29%		
LIVSM	OAVS	6,93	18%	7,08	20%	6,20	-3%	8,10	1%
	YRKM	10,04	15%	9,16	16%	8,43	3%		
	FIRMA	9,01	31%	9,12	46%	9,15	-2%		
RVIRKE	OAVS	7,50	61%	6,67	66%	6,42	67%	6,72	66%
	YRKM	7,91	61%	7,05	65%	6,80	66%		
	FIRMA	22,94	68%	19,15	79%	18,14	78%		
GRUS	OAVS	6,19	19%	7,30	19%	5,93	9%	5,94	12%
	YRKM	6,84	14%	8,05	20%	6,60	14%		
	FIRMA	14,54	43%	13,65	27%	13,36	-5%		
OLJA	OAVS	8,34	48%	7,69	61%	6,96	57%	8,22	44%
	YRKM	9,35	41%	8,40	59%	7,66	57%		
	FIRMA	16,68	77%	18,55	72%	16,61	61%		

Tabell 10: Relativa felmarginaler och variansreduktion vid urval om 3600 lastbilar.

Relativa felmarginaler i % för årsaggregat avseende KVANT, TONKM, KÖRKMMED och FRAKTINK. Den redovisade variansreduktionen (var.red) är räknad relativt skattningsprecisionen i Tabell 3. Lastbilspopulation: Bilar med 3,5 tons MAXLV eller mer. Stratifiering: Enligt Avsnitt 5. Urvalsallokering: Enligt Avsnitt 5. Urvalsstorlek: 3200 lastbilar per kvartal.									
Varu-grupp	Ägar-kategori	KVANT	Var.-red	TONKM	Var.-red	KÖRKMMED	Var.-red	FRAKTINK	Var.-red
ALLAV	OAVS	3,19	12%	2,85	13%	2,44	0%	2,11	28%
	YRKM	3,62	13%	3,16	17%	2,71	13%		
	FIRMA	6,72	40%	5,69	45%	5,29	21%		
LIVSM	OAVS	7,35	7%	7,51	10%	6,57	-16%	8,59	-11%
	YRKM	10,65	5%	9,72	6%	8,94	-9%		
	FIRMA	9,57	22%	9,71	39%	9,69	-14%		
RVIRKE	OAVS	7,97	56%	7,09	61%	6,82	62%	7,14	62%
	YRKM	8,40	56%	7,49	60%	7,22	62%		
	FIRMA	24,47	64%	20,42	76%	19,34	75%		
GRUS	OAVS	6,57	9%	7,75	9%	6,29	-2%	6,29	2%
	YRKM	7,26	3%	8,54	10%	7,01	3%		
	FIRMA	15,39	36%	14,47	19%	14,17	-18%		
OLJA	OAVS	8,84	41%	8,15	56%	7,37	52%	8,71	38%
	YRKM	9,92	33%	8,90	54%	8,12	52%		
	FIRMA	17,67	75%	19,67	69%	17,55	56%		

Tabell 11: Relativa felmarginaler och variansreduktion vid urval om 3200 lastbilar.

Appendix. Teoretiska och tekniska aspekter på översynen

I detta appendix redovisas teoretiska och tekniska delar av de överväganden som ledde fram till den i Avsnitt 5 föreslagna nya utformningen av UVAVs urvalsförfarande. I Avsnitt A1 diskuteras UVAVs estimationsproblematik, i Avsnitt A2 den metod som användes för att uppskatta estimationsprecisionen vid alternativa urvalsutformningar och i Avsnitt A3 indikeras huvuddragen i sökandet efter en optimal urvalsutformning. Slutligen, i Avsnitt A4 behandlas ett speciellt urvalsförfarande av relevans för UVAV, kallat "OSU med disjunkta delurval".

A1 Om UVAVs estimationsproblematik

Som framgår av Avsnitt 4, i centrum för intresset i översynen stod precisionen för skattningar av aggregat av typen (2.1), då särskilt årsaggregat såsom varande den för UVAV viktigaste typen av statistisk storhet. Huvudkaraktäristikerna för bedömning av en tänkbar urvalsutformning var aggregatskattningars relativa felmarginal, vars bedömning fordrar uppskattning av estimatorvarianser. Delar av UVAVs punkt- och variansskattningsproblem är av litet ovanligt slag, vilket vi snart kommer in på, men först görs några mer standardmässiga reduktioner av estimationsproblematiken.

A1.1 Återföring av estimationsproblem för års- och kvartalsaggregat på kvartalsvisa stratumaggregat

En UVAV-undersökning utförs genom fyra kvartalsundersökningar, vilka sedan ger underlag för årsstatistik. En grundläggande relation för steget från kvartal till år är (A1.1) nedan, där beteckningarna är enligt (2.1) och där k löper över Års fyra kvartal;

$$AGG(X;V;\ddot{A};\ddot{A}_r) = \sum_{k=1}^4 AGG(X;V;\ddot{A};k). \quad (A1.1)$$

Vidare, i UVAVs hittillsvarande såväl som framtida urvalsutformning spelar stratifiering en central roll, och med stratum avses "lastbilsstratum". Allmänt används bokstaven h , $h=1,2,\dots,H$, för att etikettera lastbilsstratana, och ett index h markerar att den indicerade storheten hänför sig till stratum no h . Följande relation gäller;

$$AGG(X;V;\ddot{A};k) = \sum_{h=1}^H AGG(X;V;\ddot{A};k)_h, \quad k=1,2,3,4. \quad (A1.2)$$

Den sedvanliga "hatten" ^ används för "approximativt väntevärdesriktig" skattning (= med försumbar skevhet). (A1.1) och (A1.2) ger bakgrund för nedanstående relationer;

$$\hat{AGG}(X;V;\ddot{A};\ddot{A}_r) = \sum_{k=1}^4 \hat{AGG}(X;V;\ddot{A};k) = \sum_{k=1}^4 \left\{ \sum_{h=1}^H \hat{AGG}(X;V;\ddot{A};k)_h \right\}. \quad (A1.3)$$

Kvartalsundersökningars urval dras oberoende av varandra, och i en kvartalsundersökning dras urvalen från olika stratan oberoende av varandra. Dessa oberoendeförhållanden rättfärdigar nedanstående relation (A1.4), där "var" betecknar varians:

$$\text{var}[\hat{AGG}(X;V;\ddot{A};\ddot{A}_r)] = \sum_{k=1}^4 \text{var}[\hat{AGG}(X;V;\ddot{A};k)] = \sum_{k=1}^4 \sum_{h=1}^H \text{var}[\hat{AGG}(X;V;\ddot{A};k)_h], \quad (A1.4)$$

vilken i sin tur ger bakgrund för nedanstående relationer för skattade varianser;

$$\hat{\text{var}}[\hat{\text{AGG}}(X;V;\ddot{A};\ddot{A}r)] = \sum_{k=1}^4 \hat{\text{var}}[\hat{\text{AGG}}(X;V;\ddot{A};k)] = \sum_{k=1}^4 \sum_{h=1}^H \hat{\text{var}}[\hat{\text{AGG}}(X;V;\ddot{A};k)_h] . \quad (\text{A1.5})$$

En huvudaspekt på (A1.3) och (A1.5) är följande. Om man har punkt- och variansskattningar för kvartalsvisa stratumaggregat, kan man genom att applicera (A1.3) och (A1.5) utvidga till skattningar för kvartals- och årsaggregat. Vi koncentrerar oss därför i fortsättningen på skattning av kvartalsvisa stratumaggregat.

A1.2 Urvals- och observationsförfarandet i en kvartalsundersökning

Nedan ges en beskrivning av urvals- och observationsförfarandet i en kvartalsundersökning, som är litet mer teknisk än den i Avsnitt 2.3. Mer utförliga beskrivningar finns bl a i Appendix 2 i Carlsson, Forssén & Rosén (1989) samt i Rosén & Sundgren (1991). Alla kvartalsundersökningar genomförs på principiellt samma sätt, enligt det följande.

Urvals- och observationsförfarandet i en kvartalsundersökning:

Urvalsramen för LASTBILar utgörs av Centrala Bilregistrets (CBR) förteckning över aktivregistrerade lastbilar med MAXLV över en bestämd gräns. Urvalselementen är lastbilsveckor. En **lastbilsvecka** specificeras av ett par (i,a), där i står för en LASTBIL och a för en vecka. Den totala populationen av lastbilsveckor erhålls genom att i löper över LASTBILsramen och a över undersökningskvartalets 13 veckor.

LASTBILspopulationen stratifieras enligt någon fastlagd regel. **Urvalet av lastbilsveckor** genereras genom att man först drar ett OSU av LASTBILar inom varje stratum, och sedan lottar en av kvartalets veckor till varje utvald LASTBIL. Därvid har man på förhand bestämt det antal LASTBILar som skall utväljas till var och en av kvartalets 13 veckor. Detta sker genom att LASTBILar som utväljs i ett stratum i urvalsordning fördelas på mätveckor enligt mönstret: 1,2,...,13,1,2,...,13,1,2... osv. Om lastbilsveckan (i,a) utväljs, **observeras** LASTBIL i under **mätveckan** a, via journalföring av samtliga körningar under veckan. Med urvalsteoretisk terminologi innebär detta att observationer avseende körningar samlas in för ett **stratifierat klusterurval**, där klustren består av lastbilarnas körningar under en vecka.

Kvartalet betraktas som fixerat, och k utesluts ofta ur beteckningarna. U betecknar populationen av LASTBILar och $\{U_h\}$ dess stratan. Index j etiketterar de körningar som en lastbil utför under kvartalet. För en körningsvariabel X definieras variabeln $x = \{x_{ij}\}$ på populationen av körningar under kvartalet enligt nedan;

$$x_{ij} = X\text{-värdet för körning } j \text{ av LASTBIL } i, \text{ jebil } i\text{:s körningar, } i \in U. \quad (\text{A1.6})$$

I vissa resonemang behövs också "tidsmärkning" av variabler, och beteckningen $x(t)_{ij}$ utsäger då att körningen utförs vid tid t.

De intressanta redovisningsgrupperna av körningar specificeras av villkor på varuslaget v och/eller ägarkategorin $\ddot{A}$. Motsvarande **redovisningsgruppindikator** betecknas $1(V;\ddot{A})$, där $1(V;\ddot{A})_{ij} = 1$ om körning j av LASTBIL i tillhör gruppen $(V;\ddot{A})$, och $= 0$ annars.

Variablerna $y(X;V;\ddot{A};a)$, vilka är klusteraggregat, definieras i (A1.7) nedan. I ord innebär de: $y(X;V;\ddot{A};a)_i$ = över vecka a aggregerade X-värden för körningar i redovisningsgruppen $(V;\ddot{A})$, som utförs av LASTBIL i.

$$y(X;V;\ddot{A};a)_i = \sum_{j \text{ för vilka } t \in a} x(t)_{ij} \cdot 1(V;\ddot{A})_{ij}, \quad i \in U, \quad a = 1, 2, \dots, 13. \quad (\text{A1.7})$$

Motsvarande kvartalsvisa stratumaggregat är;

$$AGG(X;V;\ddot{A};k)_h = \sum_{a=1}^{13} \sum_{i \in U_h} y(X;V;\ddot{A};a)_i, \quad h=1,2,\dots,H, \quad k=1,2,3,4. \quad (A1.8)$$

Det är statistiska storheter av typen (A1.8) som det gäller att ge punkt- och variansskattningar för, och som sedan via (A1.3) och (A1.5) leder till skattningar för kvartal och år. Vilka skattningsformler som därvid bör användas beror på den "formalbeskrivning" som väljs för urvalet av lastbilsveckor. Här finns åtminstone två möjligheter, vilka diskuteras i det följande. Vi börjar med en "korrekt" beskrivning, vars skattningsformler dock ej används, av skäl som ges längre fram. Därefter diskuteras en approximativ beskrivning av urvalsförfarandet, vilken leder till det skattningsförfarande som faktiskt används i UVAV, inklusive i den här översynen.

A1.3 Estimation baserad på urvalsbeskrivningen "OSU med disjunkta delurval"

I detta avsnitt spelar begreppet "OSU med disjunkta delurval" en central roll. Det urvalsförfarandet definieras och studeras i Avsnitt 4, som läsaren fortsättningsvis förutsätts vara bekant med. Till att börja med bortses från svarsbortfall och ramtäckningsbrister. Följande utsagor skall vara uppenbara vid eftertanke.

Urvalet av lastbilsveckor till en kvartalsundersökning är ett OSU med 13 disjunkta delurval, ett för var och en av kvartalets mätveckor. (A1.9)

Den i (A1.7) definierade variabeln $y(X;V;\ddot{A};a)$ observeras endast på delurvalet för mätvecka a , $a=1,2,\dots,13$. (A1.10)

Vi inför ytterligare några beteckningar. Låt

N_h = antal lastbilar i stratum U_h , (A1.11)

$n(a)_h$ = antal utvalda lastbilar med mätvecka a i stratum U_h . (A1.12)

Observerade variabelvärden betecknas med den stora bokstav som svarar mot populationsvariabelns lilla bokstav, och index v löper över observerade lastbilar inom ett par (h,a) av ett stratum och en mätvecka. Speciellt står $\{Y(X;V;\ddot{A};a)_{hav}; v=1,2,\dots,n(a)_h, a=1,2,\dots,A\}$ för den uppsättning värden på variablerna $y(X;V;\ddot{A};a)$ enligt (A1.7), som genereras av observerade "(h,a)-bilar".

Från (A1.9) och (A1.10) följer att formeln (A4.8) kan tillämpas stratumvis för punkt-skattning av AGG-storheter enligt (A1.8), med följande korrespondenser mellan beteckningarna ovan och de i Avsnitt A4: $y(a)=y(X;V;\ddot{A};a)$, $N=N_h$, $n(a)=n(a)_h$. Vidare, för UVAV-undersökningen bedöms villkoret (A4.10) vara uppfyllt med god approximation, varav följer att även (A4.12) kan tillämpas. Man kommer då till;

$$\hat{AGG}(X;V;\ddot{A};k)_h = \sum_{a=1}^{13} \frac{N_h}{n(a)_h} \cdot \sum_{v=1}^{n(a)_h} Y(X;V;\ddot{A};a)_{hav}, \quad (A1.13)$$

$$\hat{\text{var}}[\hat{AGG}(X;V;\ddot{A};k)_h] = N_h^2 \cdot \sum_{a=1}^{13} \left(\frac{1}{n(a)_h} - \frac{1}{N_h} \right) \cdot S^2[Y(X;V;\ddot{A};a)]_h. \quad (A1.14)$$

(A1.13) och (A1.14) anger punkt- och variansskattning för kvartalsvisa stratumaggregat, och de ger i kombination med (A1.3) och (A1.5), åtminstone i princip, lösning på det "totala" estimationsproblemet för transportaggregat. Dock, som redan har antytts, är skattningsformlerna ifråga förenade med problem, vilka hänger ihop med det nedanstående.

Användning av (A1.13) fordrar att $n(a)_h \geq 1$ för alla (h,a) , (A1.15)

Användning av (A1.14) fordrar att $n(a)_h \geq 2$ för alla (h,a) . (A1.16)

För UVAV är villkoret i (A1.15) uppfyllt för alla (h,a) om $n(a)_h$ uppfattas som antal **utvalda** lastbilsveckor. I praktiken föreligger dock ett visst mått av svarsbortfall, och då tolkas $n(a)_h$ som antal **svarande** lastbilar. Med UVAVs urvalsstorlekar och bortfallsfrekvens uppträder då ett flertal 0:or bland $n(a)_h$. Detta vållar problem i (A1.13), och problemen blir än större i (A1.14). I UVAV är (A1.16) inte uppfyllt för alla (h,a) ens om $n(a)_h$ uppfattas som antal utvalda lastbilar, ty för vissa (h,a) är urvalsstorleken = 1. Med svarsbortfall kommer ytterligare 1:or och, som sagt även 0:or, in bland $n(a)_h$ (=antal svarande). Slutsatsen blir att (A1.13) och (A1.14) inte är praktiskt användbara i UVAV.

A1.4 Estimation baserad på urvalsbeskrivningen "OSU av lastbilsveckor"

En väg runt ovan diskuterade svårighet, ges av att basera estimationen på följande alternativa beskrivning av den probabilistiska strukturen i urvalet av lastbilsveckor;

Urvalet av lastbilsveckor till en kvartalsundersökning är ett OSU från populationen av alla lastbilsveckor under kvartalet. (A1.17)

Om beskrivningen (A1.17) kan omgående sägas att den inte är exakt riktig, utan endast approximativ, vilket inses av t ex följande. Om (A1.17) gällde exakt, skulle en och samma lastbil kunna väljas ut till flera mätveckor, vilket emellertid är omöjligt med det urvalsförfarande som används. Dock, även om beskrivningen (A1.17) inte är helt korrekt, kan den mycket väl leda till tillfredsställande estimation, sett ur praktisk synvinkel.

Härnäst skall vi skriva ned de skattningsformler som (A1.17) leder till. Notera följande;

Totala antalet lastbilsveckor för bilar i stratum $U_h = 13 \cdot N_h$. (A1.18)

Under (A1.17) är värdet på mätveckan inte längre intressant vid skattningen, och a släpps därför i flertalet beteckningar. Låt

$n_h =$ antal utvalda lastbilar i stratum $h = n(1)_h + n(2)_h + \dots + n(13)_h$. (A1.19)

Vidare, låt $\{Y(X;V;\ddot{A})_{h\nu}; \nu = 1,2,\dots,n_h\}$ beteckna den uppsättning av $y(X;V;\ddot{A})$ -värden enligt (A1.7), som ges av de under kvartalet observerade lastbilarna från stratum U_h . Notera att ν nu löper över alla för kvartalet utvalda lastbilar i stratum U_h , och inte som tidigare bara över lastbilarna med samma mätvecka.

Till att börja med bortses från svarsbortfall. Stratumvis tillämpning av välkända OSU-formler för skattning ger, för $h = 1,2,\dots,H$;

$$\hat{AGG}(X;V;\ddot{A};k)_h = \frac{13 \cdot N_h}{n_h} \cdot \sum_{\nu=1}^{n_h} Y(X;V;\ddot{A})_{h\nu} , \quad (A1.20)$$

$$\text{var}[\hat{AGG}(X;V;\ddot{A};k)_h] = (13 \cdot N_h)^2 \cdot \left(\frac{1}{n_h} - \frac{1}{13 \cdot N_h} \right) \cdot \sigma^2[y(X;V;\ddot{A})]_h , \quad (A1.21)$$

$$\hat{\text{var}}[\hat{AGG}(X;V;\ddot{A};k)_h] = (13 \cdot N_h)^2 \cdot \left(\frac{1}{n_h} - \frac{1}{13 \cdot N_h} \right) \cdot S^2[Y(X;V;\ddot{A})]_h , \quad (A1.22)$$

där σ_h^2 och $\dot{S}_h^2$ är enligt nedan;

$$\sigma^2(y)_h = \frac{1}{13 \cdot N_h - 1} \cdot \sum_{i \in U_h} (y_i - \mu(y)_h)^2, \quad S^2(Y)_h = \frac{1}{n_h - 1} \cdot \sum_{\nu=1}^{n_h} (Y_{h\nu} - \bar{Y}_h)^2. \quad (A1.23)$$

Formlerna (A1.20) och (A1.22) tillsammans med (A1.3) och (A1.5) anger huvudstrukturen i det skattningsförfarande som tillämpas i UVAV. De i praktiken använda formelnerna är dock något annorlunda, vilket beror på att man också måste göra justeringar för svarsbortfall och undertäckning, och dessa justeringar diskuteras nedan.

Populationen av LASTBILAR förändras förhållandevis snabbt. Fortlöpande försvinner bilar ur den (genom skrotning, avställning, mm), och bilar tillkommer (som "genuint nya", genom aktivering av tidigare avställda bilar, mm). "Försvunna" bilar utgör **övertäckning**, och "tillkomna" bilar utgör **undertäckning**. Dessa populationsförändringar måste beaktas på lämpligt sätt för att inte skattningsскеvhет skall uppstå, och samma sak gäller för svarsbortfallet (vilket är cirka 10%). Populationsförändringarna spelar roll av två huvudskäl, dels därför att urvalet dras en tid före (cirka två månader) undersökningskvartalet dels därför att mätveckorna sprider sig i tiden (över ett kvartal). UVAVs urvalsförfarande har den struktur som brukar åberopas som "urval i rum och tid". I Rosén (1990a) behandlas justeringsproblem för den slags urval, och där redovisas det justeringsförfarande som används i UVAV, inklusive i den här översynen. Förfarandet beskrivs kortfattat i det följande, och den som önskar utförligare beskrivning hänvisas till nyssnämnda rapport.

Vad gäller svar från utvalda lastbilar, föreligger tre huvudmöjligheter: (i) Svar inkommer, och omfattar ifylld körjournal. (ii) Svar inkommer, från vilket framgår att lastbilen tillhör övertäckningen. (iii) Inget svar inkommer, dvs svarsbortfall föreligger. Låt;

$$m_h = \text{antal (körjournal)svarande lastbilar i stratum } U_h, \quad h=1,2,\dots,H. \quad (A1.24)$$

$$m_h^* = \text{antal svarslämnande lastbilar i stratum } U_h \text{ (= såväl de körjournal-svarande som de övertäckningssvarande)}, \quad h=1,2,\dots,H. \quad (A1.25)$$

I denna typ av situation gäller följande: Om undertäckningen bedöms medföra endast försumbar skattningsскеvhет och om svarsbortfallet kompenseras medelst s k "rak uppräkning", skattas ett aggregat genom att motsvarande urvalssumma multipliceras med uppräkningsfaktorn $13 \cdot N_h / m_h^*$. I UVAV är dock undertäckningssкеvhетen inte försumbar, och med den urvalsstruktur som föreligger uppnås adekvata justeringar för undertäckning och bortfall genom att istället använda uppräkningsfaktorn $13 \cdot N_h^* / m_h$, där

$$N_h^* = \text{"kvartalsmittvägd" populationsstorlek (se Rosén (1990a))}. \quad (A1.26)$$

Nedan anges resulterande skattningsformler;

$$\hat{AGG}(X;V;\ddot{A};k)_h = \frac{13 \cdot N_h^*}{m_h} \cdot \sum_{\nu=1}^{m_h} Y(X;V;\ddot{A})_{h\nu}, \quad (A1.27)$$

$$\hat{\text{var}}[\hat{AGG}(X;V;\ddot{A};k)_h] = (13 \cdot N_h^*)^2 \cdot \left(\frac{1}{m_h} - \frac{1}{13 \cdot N_h^*} \right) \cdot S^2[Y(X;V;\ddot{A})]_h, \quad (A1.28)$$

$$S^2[Y(X;V;\ddot{A})]_h = \frac{1}{m_h - 1} \cdot \sum_{\nu=1}^{m_h} [Y(X;V;\ddot{A})_{h\nu} - \bar{Y}(X;V;\ddot{A})_h]^2. \quad (A1.29)$$

För framtida bruk anges den till (A1.28) svarande "teoretiska" variansformeln;

$$\text{var}[\hat{\text{AGG}}(X;V;\ddot{A};k)_h] = (13 \cdot N_h^*)^2 \cdot \left(\frac{1}{m_h} - \frac{1}{13 \cdot N_h^*} \right) \cdot \sigma^2[y(X;V;\ddot{A})]_h, \quad (\text{A1.30})$$

De skattningsformler som används i UVAV, är de som utgår från (A1.27) - (A1.29), och sedan "utvidgas" via (A1.3) och (A1.5).

Nyssnämnda formeluppsättningen skiljer sig från den "korrekta" som utgår från (A1.13) och (A1.14), men som ju inte kan användas i UVAV av tidigare nämnda skäl. Att (A1.27) - (A1.30) verkligen ger praktiskt tillfredsställande skattningar och beskrivning av estimatorvariansen, är dock grundat mer på bedömning än på empirisk prövning. På grund av problemen i anslutning till (A1.15) och (A1.16) är det inte självklart hur en empirisk prövning av approximationsgodheten i (A1.30) skulle genomföras. Tiden medgav inte heller någon mer ingående utredning av frågan, även om en sådan hade varit önskvärd. Bedömningen är dock att variansformeln (A1.30) gäller med god nog approximation för syftena i översynen, där det viktigaste var att kunna göra relativa jämförelser av skattningsprecisionen vid olika urvalsutformningar. Dessutom, vi känner inget praktiskt användbart alternativt estimationsförfarande.

A2 Estimationsprecision vid alternativa undersökningsutformningar

Vid översynen gällde det framför allt att skaffa uppfattning om den skattningsprecision som skulle erhållas i framtida UVAV-undersökningar med alternativa, och förhoppningsvis bättre, urvalsutformningar. Empirisk information om förhållandena i UVAVs lastbils- och körningspopulationer hämtades ur materialet från UVAV90 (= 1990 års UVAV), där urval och observation var enligt Avsnitt A1.2. Vid beräkningar på det empiriska materialet var det av intresse att samtidigt betrakta (åtminstone) två olika stratifieringar och urvalsallokeringar, dels den stratifiering och allokering som användes i UVAV90 dels den alternativa stratifiering och allokering som man ville vinna information om. Följande beteckningar används, och de hänför sig till kvartalsurval;

U_h betecknar ett stratum i UVAV90:s stratifiering, $N^*(U_h)$ motsvarande (kvartalsmittvägda) stratumstorlek, och m_h antalet svarande LASTBILAR i statumet, $h = 1, 2, \dots, H$. (A2.1)

W_ℓ betecknar ett stratum i en alternativ stratifiering, $N^*(W_\ell)$ motsvarande (kvartalsmittvägda) stratumstorlek, och r_ℓ antal svarande LASTBILAR i stratumet, $\ell = 1, 2, \dots, L$. (A2.2)

Det förutsätts att urvals- och observationsförfarandet vid en alternativ undersökningsutformning (A2.2) också omfattas av beskrivningen i Avsnitt A1.2. Genom att "byta bokstäver" i (A1.27) och (A1.30) erhålls nedanstående formler för skattning av ett kvartalsaggregat och tillhörande teoretiska estimatorvarians;

$$\hat{\text{AGG}}(X;V;\ddot{A};k) = \sum_{\ell=1}^L \frac{13 \cdot N^*(W_\ell)}{r_\ell} \cdot \sum_{\nu=1}^{r_\ell} Y(X;V;\ddot{A})_{h\nu}, \quad (\text{A2.3})$$

$$\text{var}[\hat{\text{AGG}}(X;V;\ddot{A};k)] = \sum_{\ell=1}^L 169 \cdot N^*(W_\ell)^2 \cdot \left(\frac{1}{r_\ell} - \frac{1}{13 \cdot N^*(W_\ell)} \right) \cdot \sigma^2[y(X;V;\ddot{A})]_\ell. \quad (\text{A2.4})$$

Den för översynen mest intressanta formeln var (A2.4), vilken anger estimatorvariansen för ett kvartalsaggregat under den alternativa undersökningsutformningen. Den formeln

kunde dock inte tillämpas alldeles omgående, därför att $N^*(W_t)$ och $\sigma^2[y(X;V;\ddot{A})]_t$ inte var kända. En aspekt på de storheterna, är att intresset framför allt gäller deras framtida värden (=vid UVAV93), och framtiden är det ju svårt att uttala sig om. Det problemet "löstes" på det "vanliga" sättet, genom att anta att den bästa prognos man kan ge för framtida UVAV-populationer, åtminstone för 1993 års, är att tro att de ser ut ungefär som den senast tillgängliga, dvs 1990 års. De aktuella storheterna $N^*(W_t)$ och $\sigma^2[y(X;V;\ddot{A})]_t$ var dock inte kända heller för 1990, men kunde skattas enligt nedan.

Utgå från följande välkända formel för en stratumvarians;

$$\sigma^2[Y(X;V;\ddot{A})]_t = \frac{1}{N^*(W_t) - 1} \cdot \left[\sum_{i \in W_t} y(X;V;\ddot{A})_i^2 - \frac{1}{N^*(W_t)} \cdot \left(\sum_{i \in W_t} y(X;V;\ddot{A})_i \right)^2 \right]. \quad (A2.5)$$

Använd (A1.27) och (A1.3) för att skatta komponenterna i (A2.5) enligt nedanstående mönster, där $\Delta(W_t)_{h\nu}$ indikerar om lastbil no ν i U_h hör till stratum W_t eller ej;

$$N^*(W_t) = \sum_{i \in W_t} 1 \quad \text{skattas med} \quad \sum_{h=1}^H \frac{13 \cdot N^*(U_h)}{m_h} \cdot \sum_{\nu=1}^{m_h} \Delta(W_t)_{h\nu}, \quad (A2.6)$$

$$\sum_{i \in W_t} y_i \quad \text{skattas med} \quad \sum_{h=1}^H \frac{13 \cdot N^*(U_h)}{m_h} \cdot \sum_{\nu=1}^{m_h} Y_\nu \cdot \Delta(W_t)_{h\nu}, \quad (A2.7)$$

$$\sum_{i \in W_t} y_i^2 \quad \text{skattas med} \quad \sum_{h=1}^H \frac{13 \cdot N^*(U_h)}{m_h} \cdot \sum_{\nu=1}^{m_h} Y_\nu^2 \cdot \Delta(W_t)_{h\nu}. \quad (A2.8)$$

Insättning i (A2.5) av skattningar enligt (A2.6) - (A2.8) ger estimat av $\sigma^2[y(X;V;\ddot{A})]_t$. Dessa används sedan i (A2.4) tillsammans med skattningar av $N^*(W_t)$ enligt (A2.6), och ger då skattning av estimatorvarianser för kvartalsaggregat. Genom att applicera (A1.5) erhålls skattningar av estimatorvarianser för årsaggregat.

För att skatta relativa felmarginaler behövs också skattningar av aggregaten själva, dvs av $AGG(X;V;\ddot{A};k)$. Sådana framtas genom att tillämpa formeln (A1.27). (Man kommer till samma sktningar, om man går "omvägen" att utgå från (A2.3) och tillämpar tankegångarna i (A2.6) - (A2.8).)

Relativa felmarginaler (RF) för aggregatskattningar beräknas sedan enligt (A2.9), där P som tidigare står för endera av ett kvartal eller hela året;

$$RF[\hat{AGG}(X;V;\ddot{A};P)] = \frac{2 \cdot \sqrt{\hat{\text{var}}[\hat{AGG}(X;V;\ddot{A};P)]}}{\hat{AGG}(X;V;\ddot{A};P)}. \quad (A2.9)$$

A3 Optimering av stratifiering och urvalsallokering

A3.1 Några allmänna aspekter på optimeringsproblematiken

Huvuduppgiften var att finna en för UVAV optimal, eller åtminstone "nästan optimal", urvalsutformning under givna restriktioner på undersökningsresurserna. En **urvalsutformning** specificeras därvid av nedanstående delregler;

- En **stratifieringsregel**, att användas i samtliga kvartalsundersökningar,
 - En **regel för fördelning av en given årsurvalsstorlek** på de fyra kvartalen,
 - En **regel för fördelning en given kvartalsurvalsstorlek** på aktuella urvalsstratan.
- Termen **urvalsallokeringen** syftar fortsättningsvis främst på denna senare regel.

Som framgår av Avsnitt 4, var huvudkaraktäristikorna vid bedömning av en urvalsutformning de relativa felmarginalerna för de aggregat som pekats ut som särskilt viktiga, och som framgår av t ex Tabell 3. Förhållandet att man är intresserad av skattningsprecisionen för flera olika aggregat (med olika variabler och olika redovisningsgrupper) vållar komplikation vid optimering. Ur teoretisk synvinkel vore det mest renodlade, att på ett adekvat sätt väga ihop felmarginalerna till en formaliserad "**total skattningsprecision**" som sedan minimeras. Det är dock mycket vanskligt att konstruera ett sådant totalkriterium. Inget försök gjordes, utan hopvägningen utfördes "med ögonmåt", varvid (4.6) - (4.8) särskilt beaktades. De i föregående avsnitt formulerade beräkningsalgoritmerna ger förhållandevis enkla beräkningar för uppskattning av relativa felmarginaler för **en** bestämd urvalsutformning. Detta innebär dock inte att optimeringsproblemet är löst. Det blir komplext genom att det föreligger så många tänkbara urvalsutformningar, att genomräkning av dem alla vore praktiskt uteslutet.

Mångfalden beror bl a på följande. När man söker efter en optimal stratifiering, har varje övervägd stratifiering "rätt" att bli bedömd på basis av en urvalsallokering som är optimal för just den stratifieringen. Detta innebär att för varje stratifiering som man vill betrakta, föreligger ett optimeringsproblem med avseende på urvalsallokeringen. I vissa enkla fall kan man ge en analytisk lösning på allokeringproblemet, och paradexemplet är Neyman-allokeringen. Dess optimalitet hänför sig dock till en speciell situation, nämligen när man vill uppnå bästa möjliga skattningsprecision för ett aggregat för **en** bestämd variabel och **en** bestämd redovisningsgrupp. För en urvalsundersökning kan man ju dock bara använda **en** urvalsallokering. Är den bra för ett aggregat behöver den inte vara bra för ett annat, utan det gäller att på lämpligt sätt jämkas ihop önskemålen. Vid den för UVAV aktuella typen av flersyftessituation finns inte någon analytisk lösning på problemet om optimal urvalsallokering. Den möjlighet som står till buds är "numerisk optimering", dvs att man numeriskt beräknar skattningsprecisioner för olika urvalsallokeringar, och genom jämförelse av dessa finner den allokering man bedömer vara bäst. På analogt sätt kan man sedan numeriskt jämföra olika stratifieringar.

I och med att det var uteslutet att göra numeriska beräkningar för alla föreliggande möjligheter, gällde det att arbeta efter någon lämplig sökstrategi, enligt vilken man sökte efter successivt bättre utformningar. Sökandet utfördes genom ett växelspel mellan förändringar med "stora steg" i riktning mot optimum, åtminstone förhoppningsvis, och efterföljande "lokal" numerisk optimering. Den lokala optimeringen gav sedan utgångspunkt för ett nästa "stora steg" osv. Det typiska exemplen på ett "stort steg" var införande av nya stratumtyper, medan "lokal" optimering typiskt gällde saker som "finjustering" av antal stratan, stratumgränser och urvalsallokering. Färdriktningar för de "stora stegen" gavs dels av allmänna principer/tumregler av typen (4.9) och (4.10) dels av F/SES på erfarenhet grundade förmodanden om vad som borde vara bra för UVAV.

En beräkningsbesparande princip vid lokal optimering av urvalsallokeringen var följande. Börja med att beräkna Neyman-allokeringar för några av de särskilt viktiga aggregatskattningarna, och inskränk sedan fortsatt sökande till konvexa höljet av dessa Neyman-allokeringar. Där finns åtminstone "nästan-optimum". Även om principen ifråga är något heuristisk, så fungerade den bra.

A3.2 Möjliga stratifieringar i UVAV

En första fråga som inställer sig när man vill söka efter en god stratifiering, är vilka stratifieringar som överhuvud är möjliga. Eftersom urvalsstratifiering måste ske redan vid urvalsdragningen, bestäms de möjliga stratifieringarna av den information som finns tillgänglig i urvalsramen, eller som någorlunda enkelt skulle kunna göras tillgänglig där. När Kilometerskattregistret betraktas som inaktuellt att använda, är UVAVs enda källa för substantiell hjälpinformation CBR, dvs urvalsramen själv. Nedan ges en förteckning av de (huvud)variabler som finns i CBR, och så gott som alla har redan berörts;

REGNR = bilens registreringsnummer,
YMARK = 1 anger att bilen används för "yrkesmessig trafik", = 0 att ej,
KARKOD = karosserikod,
TOTVIKT = maximalt tillåten totalvikt för fordonet,
TJVIKT = fordonets egen vikt,
MAXLV = TOTVIKT - TJVIKT = maximalt tillåten lastvikt,
LÄN = bilägarens hemmalän,
KOMMUN = bilägarens hemmakommun.

A3.3 Restriktioner avseende tillgängliga undersökningsresurser

För att de framräknade relativa felmarginalerna för olika urvalsutformningar skulle bli jämförbara, gällde det att konstanthålla någon lämplig aspekt på de utformningar som skulle jämföras. Under stordelen av sökandet hölls följande storheter fixerade: **De totala antalen svarande lastbilar för vart och ett av de fyra kvartalen.** Därvid användes de antal svarande som förelåg i UVAV90, vilka var enligt nedan;

För 1:a kvartalet: $m=3308$, För 2:a kvartalet: $m=3312$,
För 3:e kvartalet: $m=3212$, För 4:e kvartalet: $m=3383$.

I optimeringens slutskede betraktades även möjligheten att fritt variera urvalsstorlekarna mellan kvartalen, och då konstanthölls det **totala antalet svarande lastbilar under ett år.**

Specifikation av en urvalsutformning kan naturligtvis inte ges i termer av de antal svarande lastbilar som skall erhållas, utan måste formuleras i termer av de antal lastbilar som skall utväljas. Steget från "antal svarande" till "antal utvalda" lastbilar togs med hjälp av antagandet att den i UVAV90 konstaterade totala bortfallsfrekvensen gäller även i framtiden, och gäller likformigt över totala lastbilspopulationen.

A3.4 Stegen i sökandet efter en optimal utformning

I nedanstående Tabell 12 redovisas huvudgången i optimeringssökandet, med hjälp av de begrepp "stora steg" och "lokal optimering" som diskuteras ovan.

"Stora steg"	"Lokal optimering" avsåg
Startpunkten var UVAV90s populationsavgränsning och stratifiering.	Urvalsallokeringen på stratan.
Stratifieringen utökades med uppdelning efter YMARK.	Urvalsallokeringen på stratan, bl a ifråga om fördelningen av urvalet mellan stratumgrupperna YRKESBILar och FIRMABILar.
Ändrad populationsavgränsning till "EG-population" (MAXLV $\geq 3,5$ ton).	Urvalsallokeringen på stratan, bl a ifråga om fördelningen av urvalet mellan stratumgrupperna YRKESBILar och FIRMABILar.
Stratifieringen utökades med särskilda stratan för BANKEBILar och TANKBILar.	Urvalsallokeringen på stratan, bl a ifråga om fördelningen dels mellan stratumgrupperna YRKESBILar och FIRMABILar dels mellan grupperna BANKE-, TANK- och FLERTBILar.
Stratifieringen gavs ökad flexibilitet ifråga om indelning efter MAXLV.	Antal MAXLV-stratan och stratumgränser för dessa. Urvalsallokeringen på stratan, bl a ifråga om fördelningen dels mellan stratumgrupperna YRKESBILar och FIRMABILar dels mellan grupperna "BANKE-, TANK- och FLERTBILar. Beräkning av urvalsproportionaler. Allmän urvalsfördelning mellan kvartal. Minsta urvalsstorlek i stratum.

Tabell 12: Huvudstruktur i det genomförda sökandet efter en optimal utformning.

Det skulle föra för långt, och kräva för mycket utrymme, att här redovisa den mängd av numeriska resultat som framtoqs i sökandet efter optimum. Framför allt handlade det om resultat-tabeller med utformning enligt Tabell 3, för kvartal såväl som för år, och totalt framtoqs hundratals sådana. Dessa finns samlade i ett arkiv, som också omfattar de SAS-program som användes för beräkningarna. Pärmen ifråga finns hos F/SE, och den som skulle vilja studera mer detaljerade aspekter av optimeringen kan göra det där.

A4 OSU med disjunkta delurval

Resultaten i detta avsnitt finns i Rosén (1990b), men de tas med även här, bl a därför att de bör finnas tillgängliga på någon "lättillgänglig" plats.

Vi börjar med några begrepp och beteckningar, och använder då mer allmänna beteckningsmönster än tidigare. Bl a används E, V och C för att beteckna de sannolikhetssteoretiska operationerna väntevärde, varians respektive kovarians. U står för en **population**, vars objekt är etiketterade med de N första heltalen, $U = \{1, 2, \dots, N\}$. En **variabel** på U är en tillordning av tal till objekten i U; $\mathbf{x} = (x_1, x_2, \dots, x_N)$. (Populations)**medelvärde** för $\mathbf{x}$ är;

$$\mu(\mathbf{x}) = \frac{1}{N} \cdot \sum_{i \in U} x_i. \quad (\text{A4.1})$$

Den till variablerna $\mathbf{x}$ och $\mathbf{y}$ svarande (populations)**kovariansen** definieras

$$\sigma(\mathbf{x}, \mathbf{y}) = \frac{1}{N-1} \cdot \sum_{i \in U} [x_i - \mu(\mathbf{x})] \cdot [y_i - \mu(\mathbf{y})], \quad (\text{A4.2})$$

och den till variabeln $\mathbf{x}$ svarande (populations)**variansen** är $\sigma^2(\mathbf{x}) = \sigma(\mathbf{x}, \mathbf{x})$.

Nedan specificeras ett förfarande för sannolikhetsurval från U, vilket kan konkretiseras på litet olika sätt, varav det följande är ett.

OSU med disjunkta delurval: A och $n(1), n(2), \dots, n(A)$ är givna heltal, sådana att $n(1) + n(2) + \dots + n(A) \leq N$. Ett totalt urval bestående av A stycken "märkta" delurval dras från U på följande sätt. Först utväljs ett OSU Q_1 om $n(1)$ objekt. Från de resterande $N - n(1)$ objekten dras sedan ett OSU Q_2 om $n(2)$ objekt, från de då resterande $N - (n(1) + n(2))$ objekten dras ett OSU Q_3 av storlek $n(3)$, osv tills A stycken delurval dragits. (Dragningsförfarandet omfattar alltså, förutom dragning av ett totalt urval, att utvalda objekt "märks" med det delurval de tillhör.) (A4.3)

Det totala urvalet är utan replikationer och med fix urvalsstorlek, nämligen ett OSU-urval av storlek $n = n(1) + n(2) + \dots + n(A)$. Vidare gäller att det (oordnade) "märkta" urvalets probabilistiska egenskaper faktiskt inte beror av i vilken ordning man låter de olika delurvalen komma in i dragningsprocessen. Det senare illustreras bl a av de inklusions-sannolikheter som anges längre fram. Sätt;

$T(x; a) =$ summan av x -värdena i Q_a , dvs **delurvalssumman** i det a :te delurvalet. (A4.4)

Följande lemma är grundläggande för den fortsatta diskussionen. Vi skjuter dock på beviset ett tag, och ger först ett par tillämpningar av lemmat.

LEMMA A4.1: Låt $z(1), z(2), \dots, z(A)$ vara variabler på populationen U , från vilken ett OSU med A disjunkta delurval dras. T betecknar nedanstående **totala urvalssumma**;

$$T = \sum_{a=1}^A T(z(a); a). \quad (A4.5)$$

Variansen för T ges då av (A4.6), med μ , σ och σ^2 enligt (A4.1) och (A4.2);

$$V\left[\sum_{a=1}^A T\right] = \sum_{a=1}^A n(a) \cdot \left(1 - \frac{n(a)}{N}\right) \cdot \sigma^2[z(a)] - \frac{1}{N} \cdot \sum_{a \neq b} n(a) \cdot n(b) \cdot \sigma[z(a), z(b)]. \quad (A4.6)$$

LEMMA A4.2: Låt $y(1), y(2), \dots, y(A)$ vara variabler på U , och låt intresset gälla nedanstående populationstotal;

$$\tau = \sum_{a=1}^A \sum_{i \in U} y(a)_i. \quad (A4.7)$$

Ett OSU med A disjunkta delurval dras från U , och för objekten i det a :te urvalet observeras variabeln $y(a)$, $a = 1, 2, \dots, A$. Väntevärdesriktig estimation av τ ges då av nedanstående estimator, i vilken $I(a)$ står för urvalsindikatorn enligt (A4.13);

$$\hat{\tau} = \sum_{a=1}^A \frac{N}{n(a)} \sum_{i \in U} y(a)_i \cdot I(a)_i. \quad (A4.8)$$

Estimatorns varians är;

$$V(\hat{\tau}) = N^2 \cdot \sum_{a=1}^A \left(\frac{1}{n(a)} - \frac{1}{N} \right) \cdot \sigma^2[y(a)] - N \cdot \sum_{a \neq b} \sigma[y(a), y(b)]. \quad (A4.9)$$

Rättfärdigande: Väntevärdesriktigheten följer från välkända resultat. Variansformeln (A4.9) fås genom att tillämpa Lemma A4.1 med $z(a) = y(a) \cdot N/n(a)$, $a = 1, 2, \dots, A$. ■

Om man vill använda (A4.9) som utgångspunkt för att skatta variansen ifråga, stöter man i praktiken vanligen på följande komplikation. Skattning av $\sigma[y(a), y(b)]$ fordrar att man för ett antal objekt i U har observerat såväl $y(a)$ som $y(b)$. Vid ett OSU med disjunkta delurval, där man observerar endast en variabel per delurval (vilket antas i Lemma A4.2),

har man dock inte tillgång till sådana simultana observationer. Man kan emellertid komma runt svårigheten, om man bedömer att nedanstående villkor (A4.10) är uppfyllt.

LEMMA A4.3: Låt förutsättningarna vara som i Lemma A4.2. Antag att

$$\frac{1}{N} \cdot \sum_{a \neq b} \sigma[y(a), y(b)] \text{ är försumbar jämförd med } \sum_{a=1}^A \left(\frac{1}{n(a)} - \frac{1}{N} \right) \cdot \sigma^2[y(a)]. \quad (\text{A4.10})$$

Då gäller följande approximativa variansformel;

$$V(\hat{\tau}) \approx N^2 \cdot \sum_{a=1}^A \left(\frac{1}{n(a)} - \frac{1}{N} \right) \cdot \sigma^2[y(a)], \quad (\text{A4.11})$$

som leder till följande approximativt väntevärdesriktiga varians estimator;

$$\hat{V}(\hat{\tau}) = N^2 \cdot \sum_{a=1}^A \left(\frac{1}{n(a)} - \frac{1}{N} \right) \cdot S^2(Y(a); a). \quad (\text{A4.12})$$

där $S^2(y(a); a)$ betecknar urvalsvariansen i det a:te delurvalet.

Rättfärdigande: (A4.11) följer från (A4.9) och (A4.10). (A4.12) fås genom att kombinera (A4.11) med förhållandet att en urvalsvariens ger väntevärdesriktig skattning av motsvarande populationsvariens. ■

Härnäst ges det tidigare uppskjutna beviset av Lemma A4.1.

Bevis av Lemma A4.1: Inför följande (del)urvalsindikatorer

$$\begin{aligned} I(a)_i &= 1 \text{ om objekt } i \text{ ingår i det } a\text{:te delurvalet } Q_a, \\ &= 0 \text{ annars,} \end{aligned} \quad a = 1, 2, \dots, A. \quad (\text{A4.13})$$

Variabeln T i (A4.5) kan då skrivas enligt (A4.14) nedan, som i kombination med välbekanta räkneregler för varianser och kovarianser leder till formeln (A4.15).

$$T = \sum_{a=1}^A \sum_{i \in U} z(a)_i \cdot I(a)_i, \quad (\text{A4.14})$$

$$V(T) = \sum_{a,b=1}^A \sum_{i,j \in U} z(a)_i \cdot z(b)_j \cdot C[I(a)_i, I(b)_j]. \quad (\text{A4.15})$$

Inklusionssannolikheter av första och andra ordningen betecknas enligt nedan;

$$\pi(a)_i = P(I(a)_i = 1) = E[I(a)_i], \quad i = 1, 2, \dots, N, \quad a = 1, 2, \dots, A, \quad (\text{A4.16})$$

$$\pi(a, b)_{ij} = P(I(a)_i = I(b)_j = 1) = E[I(a)_i \cdot I(b)_j], \quad i, j = 1, 2, \dots, N, \quad a, b = 1, 2, \dots, A. \quad (\text{A4.17})$$

Dvs $\pi(a)_i$ är sannolikheten att objekt i ingår i Q_a , och $\pi(a, b)_{ij}$ är sannolikheten att objekten i och j (vilka får sammanfalla) ingår i Q_a respektive i Q_b (vilka också får sammanfalla). Enligt välbekanta räkneregler gäller då;

$$C[I(a)_i, I(b)_j] = \pi(a, b)_{ij} - \pi(a)_i \cdot \pi(b)_j. \quad (\text{A4.18})$$

Härledning av nedanstående uttryck på $\pi(a)_i$ och $\pi(a, b)_{ij}$ är rättframma, och lämnas åt läsaren. De utsäger bl a att inklusionssannolikheterna av första och andra ordningen är oberoende av i vilken ordning delurvalen kommer in i urvalsprocessen.

$$\pi(a)_i = n(a)/N, \quad a = 1, 2, \dots, A, \quad (\text{A4.19})$$

$$\begin{aligned}
\pi(a,b)_{ij} &= n(a)/N, & a=b, \quad i=j, \\
&= n(a) \cdot (n(a)-1)/N(N-1), & a=b, \quad i \neq j, \\
&= 0, & a \neq b, \quad i=j, \\
&= n(a) \cdot n(b)/N(N-1), & a \neq b, \quad i \neq j.
\end{aligned} \tag{A4.20}$$

(A4.18) - (A4.20) ger efter enkla räkningar, vilka återigen lämnas till läsaren:

$$\begin{aligned}
C[I(a)_i, I(b)_j] &= n(a) \cdot (N - n(a))/N^2, & a=b, \quad i=j, \\
&= -n(a) \cdot (N - n(a))/N^2 \cdot (N-1), & a=b, \quad i \neq j, \\
&= -n(a) \cdot n(b)/N^2, & a \neq b, \quad i=j, \\
&= n(a) \cdot n(b)/N^2 \cdot (N-1). & a \neq b, \quad i \neq j.
\end{aligned} \tag{A4.21}$$

Från (A4.21) fås medelst rättfram algebra, som lämnas åt läsaren;

$$\begin{aligned}
\sum_{i,j \in U} z(a)_i \cdot z(b)_j \cdot C[I(a)_i, I(b)_j] &= \\
&= \frac{n(a) \cdot (N - n(a))}{N \cdot (N - 1)} \cdot \left[\sum_{i \in U} z(a)_i^2 - \frac{1}{N} \cdot \left\{ \sum_{i \in U} z(a)_i \right\}^2 \right] = \\
&= n(a) \cdot \left(1 - \frac{n(a)}{N} \right) \cdot \sigma^2[z(a)], \quad a = b,
\end{aligned} \tag{A4.22}$$

$$\begin{aligned}
&= -\frac{n(a) \cdot n(b)}{N \cdot (N - 1)} \cdot \left[\sum_{i \in U} z(a)_i \cdot z(b)_i - \frac{1}{N} \cdot \left\{ \sum_{i \in U} z(a)_i \right\} \cdot \left\{ \sum_{i \in U} z(b)_i \right\} \right] = \\
&= -\frac{n(a) \cdot n(b)}{N} \cdot \sigma[z(a), z(b)], \quad a \neq b.
\end{aligned} \tag{A4.23}$$

(A4.15), (A4.22) och (A4.23) ger nu (A4.6), och lemmat är därmed bevisat.

Referenser

Carlsson, F., Forssén, K. & Rosén, B: (1989): Om framskrivning av statistik avseende varutransporter med lastbil ("UVAV-statistiken"). SCB R&D Report 1989:23.

Rosén, B. (1990a): Om justering för undertäckning vid undersökningar med urval i "rum och tid". SCB, R&D Report 1990:3.

Rosén, B. (1990b). On the Design of the Zimbabwe Bus Transportation Survey. (Appendix 2 i G. Isaksson, A. Nemaikonde & N. Taruvinga: On the Implementation of Transport Statistics in Zimbabwe: The 1990 Transport Census. Statistics Sweden/ICO ZIMSTAT 1990:7.

Rosén, B. & Sundgren, B. (1991-08-15). Exempel på dokumentation enligt SCBDOK (utvidgat DOK-system). "UVAV 1987".

Sahlberg G. (1992-07-16): Angående Arbetsmarknadsdepartementets PM 92-06-16 om NUTS-indelning av Sverige och effekterna på SCBs lastbilstransportstatistik (UVAV).

SCB: F/SE: T 30 SM 9103 Varutransporter med lastbil och järnväg under 1990.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie, som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna.

R & D Reports Statistics Sweden are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers.

Reports published 1993 (or will soon appear*):

1992:1* Kvalitetsfonden 1991/92 - Projekt, handläggning och utvärdering
(grön) (Roland Friberg)

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