

Interne notater

STATISTISK SENTRALBYRÅ

88/20

18. oktober 1988

ESTIMERING AV ANTALL YRKESAKTIVE FORDELT ETTER KJØNN, ALDER OG NÆRING
VED KOMBINASJON AV UTVALG OG REGISTRE. BØMLO KOMMUNE

AV

IB THOMSEN

I n n h o l d

1. Innledning og foreløpige konklusjoner
2. Estimering av antall yrkesaktive fordelt etter kjønn, alder og næring.
3. Yrkesaktive fordelt på næring på områdenivå.
(Tabell 23)
4. Omallokering av utvalget.
5. Effekten av bedre registre

VEDLEGG: Oppfølging av Bømloprosjektet

1. Innledning og foreløpige konklusjoner

Arbeidet som her legges fram er et ledd i metodeutviklingen for å finne hensiktsmessige metoder til kombinasjon av data fra et utvalg og registre for å produsere arbeidsmarkedstall i Folke- og bolig tellingen 1990, FoB 90. Tidligere arbeider innen feltet har gjort spesielle, teoretiske studier av potensielle metoder. Heldal og Spjøtvoll (1985); Rygh Swensen (1986); Heldal (1986); Spjøtvoll og Thomsen (1988); Heldal, Swensen og Thomsen (1987). I disse arbeidene har en lagt vekt på å foreta en grundig teoretisk og praktisk vurdering av de mange metoder som er anbefalt i litteraturen. Det viste seg da et behov for å se på hvordan en ved hjelp av en kombinasjon av noe enklere metoder kan produsere tall med en kvalitet som ikke er vesentlig lavere enn ved bruk av de mer kompliserte metodene studert tidligere. Formålet med dette arbeid er først og fremst å finne frem til slike enkle metoder samt anslå kvaliteten til estimatene. Hensikten med å anslå kvaliteten til tallene er først og fremst å kunne sammenligne forskjellige metoder. For å kunne si noe om kvaliteten i FoB-90 må vi ha mer kjennskap til registersituasjonen i 1990. Først når det foreligger et register likt det vi vil ha i 1990 kan dette spørsmålet besvares fullstendig. I dette notat har vi nærmet oss denne problemstillingen ved å se på konsistensen mellom et AKU-utvalg og et register i 1986. Resultatene av dette eksperimentet er gitt i avsnitt 5, og antyder visse reduksjoner i variansen sammenlignet med situasjoner i 1980.

Det er klart at usikkerheten til tallene avhenger av utvalgsstørrelsen. For kommuner med mellom 6 000 og 10 000 innbyggere kan det derfor bli nødvendig å "oversample" på bekostning av de større kommunene. For å vurdere konsekvensene av en oversampling, har vi presentert usikkerhetsmål både for et 20 % og et 10 % utvalg. Andre konsekvenser knyttet til dette valg er beskrevet i avsnitt 4. Det endelige valget mellom å oversample de mindre kommunene eller ikke bør tas så sent som praktisk mulig for å få best mulig oversikt over registerkvaliteten i 1990. Det er likevel allerede nå flere argumenter som taler for å oversample de mindre kommunene.

I avsnitt 2 gis en kort presentasjon av metodene som er brukt, samt resultatene fra simuleringene av tabell 6 i kommuneheftet. Det er gitt resultater både ved bruk av 10 % og 20 % utvalg.

2. Estimering av antall yrkesaktive, fordelt etter kjønn, alder og næring. (Tabell 6 i kommuneheftet.)

Den estimeringsmetode som her legges fram bruker flere trinn for å komme fram til en fordeling av de yrkesaktive etter næring og alder. I det følgende skal vi beskrive hvert trinn, vurdere resultatene og også foreslå mulige forbedringer.

1. Trinn

I første trinn fordeles alle personer 16-74 år etter yrkesaktivitet og kjønn. Estimeringen utføres ved å etterstratifisere utvalget etter samme kjennemerke i registeret. I tabell 1 er vist gjennomsnittlig resultat etter 15 forsøk, standardavviket samt tallet ifølge Fob-80. Resultatene er gitt for 20 % og 10 % utvalg.

Tabell 1. Estimert antall yrkesaktive etter kjønn

		Yrkesaktive		Ikke yrkesaktive	
		20 %	10 %	20 %	10 %
Menn	Fob	2 581	2 581	784	784
	Est	2 583	2 578	784	789
Standardavvik		41,9	55,8	41,9	55,8
Kvinner	Fob	1 152	1 152		
	Est	1 149	1 144	2 022	2 028
Standardavvik		25,6	38,8	25,6	38,8

Ettersom estimeringsmetoden er forventningsrett er det ikke så rart at gjennomsnittet etter 15 forsøk ligger meget nær folketellingstallene.

Av større interesse er det å se på standardavvikene. Det er viktig, men ikke særlig overraskende, at disse er små i forhold til estimatene. Det er liten tvil om at disse standardavvik kan reduseres ytterligere, men vi ser ingen grunn til å forfølge dette spørsmålet her.

Et spørsmål av en viss interesse er hvor mye standardavvikene reduseres ved å bruke etterstratifisering i stedet for en enkel oppblåsing av utvalgstillene. Forholdene mellom standardavvikene til disse to estimeringsmetoder har vi beregnet til å være 1,9 og 2,4 for menn og kvinner henholdsvis.

Alt i alt må en kunne si at trinn 1 er en enkel og grei estimering, som gir en ikke ubetydelig effekt ved kombinasjon med registeret.

2. Trinn

I annet trinn fordeles de yrkesaktive etter næring (9 grupper). Som i trinn 1 brukes enkel etterstratifisering etter næring i registeret. Vi får nå 9 etterstrata, men ellers er teknikken som i første trinn. Resultatene er vist i tabell 2, hvor gjennomsnittene av 15 forsøk er gitt sammen med Fob-tallene og standardavvikene. Resultatene med et 20% og 10 % utvalg er vist i tabell 2A og 2B, henholdsvis. Som i tabell 1 er forskjellene mellom Fob-tallene og gjennomsnittet etter 15 forsøk meget små, hvilket igjen skyldes at estimeringsmetoden er forventningsrett. Det er imidlertid den store forskjell at standardavvikene er mye større i tabell 2A og 2B enn i tabell 1. Gevinsten ved å bruke etterstratifisering er mye mindre for næringsfordelingen. På de to siste linjer i tabell 2A er vist standardavvikene til de etterstratifiserte tallene og til de vanlig oppblåste tallene. Det sees at gevinsten ved etterstratifisering er sterkt redusert i forhold til hva som er tilfellet i tabell 1.

Tidligere resultater tyder på at en ved hjelp av mer kompliserte metoder kan forbedre resultatene i tabell 2A og 2B noe. (Heldal (1986)) foreløpige mener vi likevel at den enkle metoden er å foretrekke.

3. Trinn

I tredje trinn fordeles de yrkesaktive etter kjønn og alder. I dette trinnet brukes en noe annen metode enn i trinn 1 og 2. Metoden har flere navn, men kalles ofte for "raking". Den består av å "justere" tall i en tabell, slik at de stemmer med gitte marginaler. Vi skal kort beskrive hvordan metoden virker:

La oss ta ut en linje i tabellen

Observerte tall				Kjent total
n_1	n_2		n_j	K

I første omgang estimeres antall personer i hver celle ved en enkel oppblåsing. Ettersom totalen, K , er kjent, regner en ut forskjellen mellom den kjente totalen og summen av de oppblåste tallene.

$$(K - \sum n_i).$$

Denne forskjellen fordeles nå på cellene, proporsjonalt med størrelsen. Det justerte tall i celle i blir derved

$$m_i = n_i + \frac{n_i}{\sum n_i} (K - \sum n_i).$$

I en tabell som består av flere linjer utføres denne justering for hver linje. Hvis den andre marginalen også er kjent, utføres deretter en lignende justering for hver kolonne. Etterpå undersøkes det om den første marginalen stemmer, hvis ikke, justeres enda en gang. Denne prosessen fortsetter inntil begge marginalene stemmer. Denne enkle metoden er brukt for å fordele de yrkesaktive etter alder. Resultatene er vist i tabell 3(A) og 3(B). I tabellene er vist gjennomsnittlig

resultat etter 15 forsøk, Fob-resultatene samt standardavvik. Forskjellene mellom de estimerte tallene og tellingstallene er så små, at metoden synes å ha en meget liten skjevhet. Dessuten synes det som om metoden gir vesentlig mindre standardavvik enn enkel oppblåsing. (Tallene er bare vist for menn, men helt tilsvarende resultater gjelder kvinner.) Dette trinnet synes altså å fungere godt, men vi har likevel forsøkt å forbedre det.

Usikkerhetene i de justerte tallene er naturligvis avhengig av usikkerhetene til de estimer vi starter med. Det kan derfor være fristende å forsøke å sette inn noen syntetiske estimer og deretter justere disse til de stemmer med marginalene. Vi har forsøkt med flere syntetiske estimer, men har funnet det fornuftig å bruke tall for hele fylket som startverdi, og deretter justere dem til de gitte marginalene. Resultatet av å bruke denne metoden er vist i tabell 3(AA) og 3(BB). Her ser en at det faktisk er et avvik mellom gjennomsnittet av 15 forsøk og tellingsresultatene. Metoden har en klar skjevhet. Til gjengjeld er standardavviket blitt ytterligere redusert.

Resultatene i tabell 3(A) og (B) gir eksempler på vesentlige trekk ved de metoder vi anvender når resultater "bringes ned" til små områder. For å gjøre det bedre enn vanlig oppblåsing, må vi bruke estimer som har en viss skjevhet. Hvis denne skjevhet blir for stor må estimatoren forkastes uansett reduksjonen i standardavvik. På den andre siden er det ofte mulig å finne estimer som har en liten skjevhet, men som gir vesentlig mindre standardavvik.

Som en konklusjon til resultatene kan en si at metodene til fordeling av yrkesaktive etter kjønn og alder synes å virke tilfredsstillende.

4. Trinn

I trinn 1 til 3 har vi estimert marginalene til tabell 6 i kommuneheftet. For å estimere simultanfordelingen bruker vi raking på samme

måten som i trinn 3. Også her har vi forsøkt med flere startverdier. Vi har valgt å bruke tall for hele fylket som startverdi, og justert disse til marginalene stemmer. Resultatene er gitt i tabell 4A og 4B. Her ser en at noen av tallene har til dels stor skjevhet, som estimeres ved å ta forskjellen mellom folke- og bolig tellingstallet (Fob) og gjennomsnittet av 15 forsøk (Est). I tabell 4A og 4B er noen av estimatene svært usikre. De tilfeller, hvor summen av standardavviket og skjevheten utgjør mer enn 40 % av estimatet, er merket med "*" ved det tilsvarende FoB-tallet. I tabell 4A er 20 tall merket, mens 29 tall er merket i tabell 4B. De fleste usikre tall er små.

Det er av en viss interesse at bruttovariansene i tabell 4A i gjennomsnitt er ca 1/7 av den variansen en ville ha fått ved enkel oppblåsing, mens det tilsvarende tall i tabell 4B er ca. 1/10.

3. Yrkesaktive fordelt på næring på områdenivå

I kommuneheftene er de yrkesaktive fordelt på næring innen hver tellingskrets. Dette klarer vi ikke med de metoder som her er presentert. Derimot klarer vi å få en tilfredsstillende fordeling på områdenivå. Vi bruker igjen raking. Resultatene er vist i tabell 5, hvor registertallene er brukt som startverdi. Resultatene med et 10 % og et 20 % utvalg er så like at vi bare presenterer resultatene av et 20 % utvalg. Når vi er på områdenivå, betyr utvalgets størrelse svært lite. Som i tabell 4A og 4B er tall med stor usikkerhet merket med "*".

4. Omallokering av utvalget.

I avsnitt 2 og 3 ble det presentert standardavvik for både et 10 % og et 20 % utvalg. Dette ble gjort for å vurdere effekten av å øke utvalget i Bømlo kommune. I dette avsnittet skal det kort vises hvordan en kan tenke seg å øke utvalget for Bømlo og samtidig få et selveiende utvalg for hele Hordaland fylke på "ca. 10 %".

Vi skal ikke her ta stilling til hvilket alternativ som bør velges. Dette valget utsettes til vi har bedre oversikt over registerkvaliteten vi kan forvente oss i 1990.

I tabell 6 er folkemengden vist for kommuner i Hordaland med mer enn 6 000 innbyggere.

Tabell 6. Folkemengden og utvalgsstørrelsen ved to alternativer.
Utvalgsplaner i kommuner i Hordaland.

Stratum nr.	Total folkemengde	Selvveiende utvalg* 10 %	Ikke selvveiende utvalg*	
			Trekkeprosent	Utvalgsstørrelse
1 Bergen	208 915	20 892	8	16 713
1 Askøy	18 401	1 840	8	1 472
2 Voss	14 038	1 404	10	1 404
2 Stord	14 026	1 403	10	1 403
2 Fjell	13 303	1 330	10	1 330
2 Kvinnherad	13 155	1 316	10	1 316
2 Os	12 013	1 201	10	1 201
2 Lindås	11 343	1 134	10	1 134
3 Bømlo	9 531	953	20	1 906
3 Kvam	8 704	870	20	1 741
3 Odda	8 466	847	20	1 693
3 Østerøy	6 914	691	20	1 383
Total	338 809	33 881		32 696

* I FoB-90 og avsnitt 2 er utvalgene redusert med ca. 23 % for å utelate alle personer under 16 år.

I 3. og 4. kolonne i tabell 6 er det vist hvordan en kan øke utvalgsprosenten blant kommuner med under 10 000 innbyggere ved å redusere utvalgsprosenten fra 10 til 8 i kommuner med mer enn 15 000 innbyggere. Det største, selvveiende utvalg en da kan trekke for Hordaland fylke blir 8 %, hvilket bør kunne sies å være "ca. 10 %".

Det følger også av totaltallene i tabell 6 at en kan øke utvalget noe i kommuner med et inbyggerantall mellom 10 000 og 15 000. Dette vil føre til en jevnere fordeling av utvalget mellom kommunene. Dette er i seg selv en fordel, fordi statistikken for de fleste kommuner da får tilnærmet samme kvalitet.

5. Effekten av bedre register.

Flere av de viktige marginalene estimert i avsnitt 2 bruker bare etterstratifisering som kombinasjonsmetode. I disse tilfeller er det relativt enkelt å anslå effekten av et bedre register. Det store problem er derfor å gi anslag for hvor gode registrene vil være i 1990. For å gå litt nærmere inn på denne problemstillingen skal vi i dette avsnittet se på den registerkvaliteten som finnes i dag, og anslå hvilken effekt det har på standardavviket til tallene gitt i tabell 2.

Sammenhengen mellom registerkvaliteten og usikkerheten på etterstratifiserte tall kan beskrives på følgende måte:

La enhver person tilhøre en av to kategorier, A og \bar{A} . Etter kobling med registeret kan følgende tabell da konstrueres:

		Register		
		A	\bar{A}	
Utvalg	A	M_{11}	M_{12}	$M_{1.}$
	\bar{A}	M_{21}	M_{22}	$M_{2.}$
		$M_{.1}$	$M_{.2}$	

M_{11} er antall personer i utvalget som tilhører kategori A både i registeret og utvalget.

M_{21} er antall personer som tilhører kategori A i utvalget, men A i registeret.

M_{12} er antall personer som tilhører A i utvalget, men \bar{A} i registeret

M_{22} er antall personer som tilhører kategori \bar{A} både i utvalget og og registeret.

I tillegg til disse tallene kjenner en

N_1 , antall personer som tilhører kategori A i registeret,

N_2 , antall personer som tilhører kategori \bar{A} i registeret.

Den etterstratifiserte estimatoren for antall personer i kategori A ifølge utvalget blir da

$$E = \frac{M_{11}}{M_{\cdot 1}} N_1 + \frac{M_{12}}{M_{\cdot 2}} N_2.$$

Dessuten er variansen til E tilnærmet lik

$$\text{var}(E) = \text{var}(\bar{X})(I - K),$$

Hvor $\text{var}(\bar{X})$ er variansen til det oppblåste utvalgsgjennomsnittet, og

$$K = \left(\frac{M_{11}}{M_{\cdot 1}} - \frac{M_{21}}{M_{\cdot 2}} \right)^2$$

Reduksjonene en oppnår ved etterstratifisering er altså en funksjon av K.

Dersom $M_{11} = M_1$ og $M_{21} = 0$ er det perfekt konsistens mellom register og utvalg, og variansen til det etterstratifiserte estimator blir null.

Fordi M_{21} ofte er liten, skal vi bruke tilnærmingen $K = \frac{M_{11}}{M_1}$.

Vi ønsker nå å si litt om hvordan K har utviklet seg over tid. Med de data vi har nå kunne dette gjøres ved å koble AKU-utvalg og A/A-register både i 1980 og 1986. Vi har imidlertid bare gjort denne koblingen i 1986. For 1980 har vi brukt K-verdien for Bømlo kommune ved kobling mellom tilleggsutvalget og FoB-80. Sammenligningen mellom de to K-verdiene i tabell 7 blir derfor noe vanskelig, men større endringer i registerkvaliteten burde vise seg.

Tabell 7. K-verdier i 1980 og 1986 etter næring.

År	Jordbr. etc.	Ind.+ bergv.	Kraft	Bygg, anlegg	Vareh.	Transp.	Fi- nans	Tj. yting
1980	0,37	0,86	0,80	0,66	0,73	0,42	0,70	0,71
1986	0,59	0,80	0,82	0,65	0,68	0,78	0,71	0,78

En ser at for næringene jordbruk etc. og transport er K mye større i 1986 enn i 1980, for de øvrige næringer er endringene små. For jordbruk vil standardavviket ved et 20 % utvalg bli redusert til ca. 25, mens det tilsvarende tall for transport bli ca. 23.

En betydelig del av inkonsistensene mellom AKU og registeret skyldes at sysselsatte personer i AKU ikke finnes i registeret, samt uoppgitt næring for selvstendige. Ytterligere reduksjoner i usikkerhetene avhenger av i hvilken grad en ved hjelp av informasjon fra andre registre kan avgjøre om disse personene virkelig er sysselsatte eller ikke og tildele dem en næringskode.

Tabell 2A. YRKESAKTIVE FORDELT ETTER KJØNN OG NÆRING. 20 % UTVALG

		Uopp- gitt	Jord- bruk	Ind.- Berg	Kraft	Bygg og anlegg	Vareh.	Transp.	Fi- nans	Tj. yting
Menn	Reg	26	246	1068	17	219	106	265	36	345
	Fob	34	353	773	19	270	121	590	37	384
	Est	39	349	776	19	271	121	588	37	382
	St.avv.	9.2	31.5	42.8	5.8	18.3	14.0	28.9	12.2	16.8
Kvinner	Reg	13	27	120	3	15	217	59	88	560
	Fob	26	66	103	1	25	234	67	30	600
	Est	28	67	109	0	24	234	71	31	585
	St.avv.	8.4	22.9	12.1	0.8	10.1	13.8	13.2	5.9	21.2
Totalt	Fob	60	419	876	20	295	355	657	67	984
	Est	67	416	885	19	295	355	659	68	967
	St.avv.	12.5	38.9	44.5	6.0	20.9	19.7	31.8	13.6	27.0
	Standardavv. som vanlig oppbåsing	16.2	39.4	55.1	8.7	33.6	36.7	48.6	16.3	57.8

Tabell 2B. YRKESAKTIVE FORDELT ETTER KJØNN OG NÆRING. 10 % UTVALG

		Uopp- gitt	Jord- bruk	Ind.- Berg	Kraft	Bygg og anlegg	Vareh.	Transp.	Fi- nans	Tj. yting
Menn	Fob	34	353	773	19	270	121	590	37	384
	Est	42	335	774	21	272	128	583	35	387
	St.avv.	15.2	35.2	66.9	7.8	30.8	21.3	35.4	19.7	20.1
Kvinner	Fob	26	66	103	1	25	234	67	30	600
	Est	22	58	103	0	17	240	68	31	605
	St.avv.	18.0	30.0	17.1	0	13.7	23.0	18.7	8.4	30.2
Totalt	Fob	60	419	876	20	295	355	657	67	984
	Est	64	393	877	21	289	368	651	66	992
	St.avv.	23.6	46.2	69.1	7.8	33.7	33.1	40.0	21.4	36.3

TABELL 3A. YRKESAKTIVE FORDELT PÅ ALDER. 20 % UTVALG

		16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-69	70-+
Menn	Fob	237	371	350	633	366	356	193	34	41
	Est	239	369	349	633	365	362	188	35	42
	St.avv.	17.1	6.4	7.7	10.5	12.4	11.5	14.4	8.7	12.4
Standardavv. ved oppblåsing		29.7	37.7	36.0	47.5	37.5	37.8	28.4		
Kvinner	Fob	130	166	159	230	172	200	83	12	0
	Est	130	160	164	225	183	196	79	13	0
	St.avv.	14.0	11.7	22.3	22.7	20.0	22.9	14.5	8.0	0
Totalt	Fob	367	537	512	863	538	556	276	46	41
	Est	369	529	513	858	548	558	267	48	42

TABELL 3AA. YRKESAKTIVE FORDELT PÅ ALDER. 20 % UTVALG

		16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-69	70-+
Menn	Fob	237	371	350	633	366	356	193	34	41
	Est	230	356	342	622	371	380	209	37	34
	St.avv.	7.4	4.1	3.8	5.2	3.8	6.0	8.0	3.4	3.6
Standardavv. ved oppblåsing		29.7	37.7	36.0	47.5	37.5	37.8	28.4		
Kvinner	Fob	130	166	159	230	172	200	83	12	0
	Est	117	152	146	248	200	193	78	10	6
	St.avv.	2.6	3.3	5.3	4.8	5.2	5.5	3.1	0.6	0.4
Totalt	Fob	367	537	512	863	538	556	276	46	41
	Est	347	508	488	870	571	573	287	47	40

TABELL 3B. YRKESAKTIVE FORDELT PÅ ALDER. 10 % UTVALG

		16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-69	70-+
Menn	Fob	237	371	350	633	366	356	193	34	41
	Est	238	368	253	634	353	358	199	30	46
	St.avv.	21.4	14.3	15.7	15.0	22.1	26.6	29.8	16.3	17.0
Kvinner	Fob	130	166	159	230	172	200	83	12	0
	Est	132	164	150	240	163	202	80	12	0
	St.avv.	28.1	19.3	29.8	37.5	22.1	46.1	28.9	9.4	0

TABELL 3BB. YRKESAKTIVE FORDELT PÅ ALDER. 10 % UTVALG

		16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-69	70-+
Menn	Fob	237	371	350	633	366	356	193	34	41
	Est	230	356	342	621	371	378	210	37	34
	St.avv.	12.0	4.5	4.0	7.1	4.6	8.2	11.5	4.3	4.7
Kvinner	Fob	130	166	159	230	172	200	83	12	0
	Est	116	151	145	248	200	191	77	10	6
	St.avv.	7.3	7.4	9.6	18.4	11.6	15.1	6.4	1.1	0.7

Tabell 5. YRKESAKTIVE FORDELT PÅ NÆRING OG OMRÅDE

Område	Uoppgitt			Jordbruk etc.			Ind. Bergv.			Kraft			Bygg			Vareh.			Transp.			Finans		
	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta	Fob	Est	Sta
S. Bømlø	10	14	4	94	109	10	69	82	4	0	0	0	40*	28	2	45	52	5	162	150	7	8	9	3
Finnås	4*	12	3	44*	18	2	71	76	3	1	1	0	35	34	3	40	31	3	94	96	4	4	4	1
Moster	6*	12	3	39	46	5	126	144	5	1	0	0	46	51	4	55	50	4	181*	127	5	19	17	4
Rubbesta	18*	9	2	59	70	7	273	230	8	3	4	1	41	45	6	61	63	6	67	110	5	17	14	4
Bremnes	22*	11	2	183	170	17	336	321	13	15	16	3	133	138	12	153	171	15	151	182	8	19	23	7

Off-priv.
Tjenestyt.

Fob Est Sta

106	103	4
103	104	3
149	136	5
218	220	7
407	413	14

* Summen av standardavviket og
skjevheten er større enn 40 %
av estimatet.

Tabell 4 (A) Yrkesaktive fordelt etter alder og næring 10 % - utvalg

NÆRING	ALDER																										
	16-19 år			20-24 år			25-29 år			30-39 år			40-49 år			50-59 år			60-66 år			67-69 år			70 +		
	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.
Uoppgitt	7	9	3	6*	8	3	4*	8	3	8*	12	4	15*	9	4	13*	9	4	6*	7	3	1*	2	1	0*	2	1
Jordbr.etc.	48	46	6	35	29	6	43*	29	6	85	67	9	68	61	6	63	74	11	43	54	7	13	14	2	21	20	3
Industri-Bergv.	74	90	9	94	112	9	104	106	8	208	196	12	151	132	9	160	149	12	75	77	9	6*	10	2	4	4	1
Kraft	1*	2	1	2*	2	1	2*	2	1	5	5	2	3	3	1	3*	4	2	2*	3	1	1*	0	0	1*	0	0
Bygg og Anlegg	32	32	5	48	43	5	58	46	6	66	70	9	27	39	6	35	36	6	25	19	3	2	2	1	2*	1	0
Varehandel	75	64	8	56	48	5	46	38	5	62	79	9	52	57	5	46	50	6	12*	25	3	3	4	1	3	3	1
Transport	62*	42	5	85	88	7	82	100	7	193	172	10	102	105	7	89	98	7	34	37	4	5	5	1	5*	2	1
Finans	5*	4	2	3*	8	3	18*	12	5	19	21	8	9	10	4	5*	7	3	5*	3	1	2*	0	0	1*	0	0
Tj. yting	63	59	6	208	169	9	152	146	12	217	246	17	111	154	12	142	142	13	74	63	6	13*	8	2	4	5	1

* Her er summen av standardavviket og skjevheten større enn 40 prosent estimert verdi

Tabell 4 (B) Yrkesaktive fordelt etter alder og næring 20 % - utvalg

NÆRING	ALDER																										
	16-19 år			20-24 år			25-29 år			30-39 år			40-49 år			50-59 år			60-66 år			67-69 år			70 +		
	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.	Fob.	Est.	Std.
Uoppgitt	7	9	2	6	8	2	4*	8	2	8	12	2	15*	9	2	13*	9	2	6	7	1	1*	2	1	0*	3	1
Jordbr.etc.	48	46	6	35	30	4	43	32	4	85	70	10	68	64	9	63	80	9	43	57	7	13	15	2	21	20	3
Industri-Bergv.	74	90	6	94	114	5	104	107	7	208	198	10	151	136	7	160	149	7	75	77	5	6*	10	1	4	4	1
Kraft	1*	2	1	2	2	1	2	2	1	5	4	1	3	3	1	3	4	1	2	2	1	1*	0	0	1*	0	0
Bygg og Anlegg	32	32	3	48	44	4	58	47	5	66	72	6	27	41	4	35	37	4	25	20	2	2	2	1	2*	1	1
Varehandel	75	62	4	56	47	3	46	37	3	62	76	4	52	54	4	46	48	3	12*	24	2	3	4	1	3	3	1
Transport	62*	44	4	85	88	5	82	100	6	193	175	10	102	105	6	89	101	7	34	38	4	5	5	1	5*	3	1
Finans	5*	3	1	3*	9	2	18*	12	3	19	22	4	9	10	2	5	7	1	5*	3	1	2*	0	0	1*	1	1
Tj. yting	63	58	2	208	167	5	152	144	4	217	239	7	111	149	6	142	138	6	74	59	4	13*	8	1	4	5	1

* Her er summen av standardavviket og skjevheten større enn 40 prosent estimert verdi

Referanser

- Heldal, J. og Spjøtvoll, E. (1985), Statistical calibration of categorical variables with application to combination of register and surveys. International Statistical review.
- Heldal, J. (1986), Combine estimation for small domains using regression estimators.
- Heldal, J., Swensen, A.R. og Thomsen I, (1987), Census statistics through combined use of surveys and registers? Statistical Journal of the United Nations ECE 5.
- Swensen, A. R. (1986), Double sampling schemes and shrinkage estimators.
- Spjøtvoll, E. og Thomsen, I. (1987), Application of some empirical Bayes methods to small area statistics. Invited paper, 46th session of ISI

OPPFØLGING AV BØMLOPROSJEKTET

1. Bakgrunn og formål med Bømloprosjektet.

I 1986 og 1987 arbeidet Gruppe for metoder med å sette seg inn i og utvikle metoder for kombinasjon av utvalg og registre. Det viste seg da at det var mange mulige kandidater til bruk ved FoB-90. Vi studerte og modifiserte flere metoder både ut fra teoretiske og praktiske synsvinkler. Dette omfattende arbeid er dokumentert i en rekke notater, hvorav flere er publisert i internasjonale tidsskrifter. På grunnlag av dette arbeidet identifiserte vi følgende viktige problemstillinger, som måtte løses:

- i) I FoB-90, må vi finne fram til forenklede versjoner av de studerte metodene.
- ii) Det blir nødvendig å kombinere flere enkle metoder for å oppnå full gevinst ved kobling med registrene.
- iii) Det finnes i dag ingen veletablerte metoder for beregning av usikkerhetene til flere av de aktuelle metoder. I det videre arbeid må vi derfor bruke simulering, og senere ta opp mulighetene for å utvikle metoder for varians- og skjevhetsberegninger.

Bømloprosjektet har som viktigste formål å belyse problemene (i) og (ii) ovenfor. Prosjektet er altså først og fremst et metodeprosjekt, som tar sikte på å finne frem til relativt enkle kombinasjonsmetoder som kan bli aktuelle ved FoB-90. Vi valgte å legge metodeprosjektet så nær den aktuelle produksjonsprosessen som mulig. Dette har ført til at resultatene også kaster lys over en del viktige spørsmål, som f.eks.

- (a) Hvor nøyaktig vil tallene bli i FoB-90?
- (b) Hvilken nøyaktighet kreves?

Det er viktig å merke seg at disse to spørsmålene ikke kan besvares endelig ved hjelp av Bømloprosjektet, hvilket heller ikke var planlagt. Spørsmål (b) krever en egen besvarelse som er helt uavhengig av metoden som brukes, og også omfatter tabellene som ikke er fremkommet ved kombinasjon. For å besvare spørsmål (b) må en ha kjennskap til registerkvaliteten i registrene i 1990.

Etter min mening må Bømloprosjektet oppfattes som avsluttet fordi vi nå har identifisert et sett med enkle kombinasjonsmetoder, hvor tapet i forhold til mer kompliserte metoder er kjent så langt som mulig. Vi har bedre metoder, men foreløpig anbefaler vi de enkle.

2. Oppfølging av Bømlo-prosjektet.

Vedlagt følger en liste over 10 prosjekter med forslag om ansvarlig kontor og ferdigdato. Nedenfor følger en del kommentarer til de enkelte prosjekter.

Prosjekt 1.

Notatet fra Bømloprosjektet skrives om. Det legges vekt på å avgrense hensikten med prosjektet samt se på effektene av et bedre register om omallokering av utvalget.

Prosjekt 2.

Gruppe for metoder utarbeider et utkast til produktbeskrivelse.

Prosjekt 3.

Gruppe for metoder og Folketellingskontoret må utarbeide et notat om ønsket nøyaktighet i tallene, inklusiv de tall som bare estimeres på grunnlag av utvalget alene. Det er vanskelig å tenke seg at dette notat skal komme frem til en offisiell norm i første omgang, men det bør lage utkast til problemstillinger samt antyde hva som bør gjøres i det tilfeller hvor tallene er for unøyaktige.

Prosjekt 4.

De metoder som til nå er utviklet skal overføres til Folketellingskontoret. Dette gjøres best ved å produsere tall for hele fylket. Dette er en ren produksjonsprosess, og Metodegruppa bør derfor spille rollen som konsulenter, mens Folketellingskontoret blir prosjektansvarlig.

Prosjekt 5.

Når et bedre register er etablert, bør effektene på variansene utredes. Dette kan gjøres når 6. ktr. har etablert et bedre register i begynnelsen av 1989.

Prosjekt 6.

Dette bør settes igang snarest som et samarbeidsprosjekt mellom Gruppe for metoder og Folketellingskontoret.

Prosjekt 7.

Den kombinerte metoden er god hvis registeret ikke er for godt eller for dårligt. Når et nytt register foreligger bør det undersøkes om registeret er så godt, at vi bør ta sikte på å rette opp registeret, og ikke bare bruke det som støtte under estimeringen.

Prosjekt 8.

Dersom vi skal estimere usikkerheten til de estimerte tallene, må vi utvikle nye metoder. Dette blir et "tungt" metodeprosjekt, og bør ikke settes i gang for vi er helt sikre på hvilke metoder som vil bli brukt i 1990.

Prosjekt 9.

I Bømloprosjektet måtte vi gi opp å bruke registeropplysninger om timetall. Det er derfor viktig å få satt i gang arbeidet med å forbedre kvaliteten til denne variabelen.

Prosjekt 10.

Ved presentasjonen av tall fra FoB-90 er det viktig å gi så mange opplysninger som mulig om usikkerhetene til tallene. På den andre siden vil det gå en del tid før vi får tilstrekkelig innsikt til å gi klare råd på dette feltet. Prosjektet bør derfor ikke settes i gang før et stykke tid inn i 1989.

OM METODEPROSJEKTER TIL FoB-90

PROSJEKT	ANSVARLIG KONTOR/FERDIGDATO		
	MET. GR.	FoB-KTR.	6. KTR.
1. OMSKRIVING AV NOTAT FRA BØMLOPROSJEKTET	15.6.88		
2. PRODUKTBEKRIVELSE	15.6.88		
3. NOTAT OM ØNSKET NØYAKTIGHET	1.8.88	1.8.88	1.8.88
4. PRODUKSJON AV FYLKESTALL		?	
5. NOTAT OM EFFEKTENE AV BEDRE REGISTRE	1.6.89		1.6.89
6. BESTEMMELSE AV TREKKENHET	1.9.88	1.9.88	
7. ALTERNATIV METODE FOR KOMBINASJON	1.6.89		1.6.89
8. VARIANSBEREGNINGER	1.6.89		
9. BEREGNING AV TIME- VERKSTALL	?		?
10. PRESENTASJON AV USIKKERHETENE I FoB-90	1.9.89	1.9.89	