RAPPORTER 81/6

METODER FOR ESTIMERING AV TALL FOR FYLKER VED HJELP AV UTVALGSUNDERSØKELSER

AV
ERLING SIRING OG IB THOMSEN

STATISTISK SENTRALBYRÅ
OSLO

RAPPORTER FRA STATISTISK SENTRALBYRÅ 81/6

METODER FOR ESTIMERING AV TALL FOR FYLKER VED HJELP AV UTVALGSUNDERSØKELSER

AV
ERLING SIRING OG IB THOMSEN

OSLO 1981
ISBN 82-537-1509-9
ISSN 0332-8422

FORORD

I NOU 1980:20 "Om arbeidet med levekårsspørsmål" finner en følgende:

> "Utvalget mener at Byrået igjen bør utrede muligheten av å utarbeide fylkesstatistikk på grunnlag av utvalgsundersøkelser, herunder spørsmålet om å revidere metoden ved utvalgsundersøkelser slik at disse blir representative også på fylkesnivå."

I den foreliggende rapport vurderes disse muligheter, og dessuten presenteres metoder som kan brukes til estimering av tall for fylker, også i de tilfeller hvor utvalgene ikke er representative på fylkesnivå.

Statistisk Sentralbyrå, Oslo, 23. april 1981

Odd Aukrust

INNHOLD

1. INNLEDNING

Innen teorien for utvalgsundersøkelser arbeider en nesten utelukkende ut fra den forutsetning at det skal estimeres tall for hele populasjonen, eller de grupper av populasjonen for hvilke det trukne utvalg er representativt. Etterhvert som det er blitt mer vanlig med større utvalgsundersøkelser har det vært økende interesse for å estimere tall for geografiske avgrensede områder, for hvilke utvalgene strengt tatt ikke er representative.

Et av de første forsøk på å estimere regionale tall for små regioner ble utført av National Center for Health Statistics i USA i 1968. En brukte da en spesiell estimeringsmetode for å estimere gjennomsnittlig antall sykedager og andre mål for helse for hver stat i USA.

I Statistisk Sentralbyrå har en siden 1976 estimert sysselsettingstall for hvert fylke på grunnlag av data fra arbeidskraftundersøkelsene. Estimeringsmetodene er svært like de som ble brukt av National Center for Health Statistics. I de aller siste årene har interessen for regionale tall tatt seg opp, spesielt i USA, og en del nye teknikker har blitt foreslått.

I dette notat skal vi gi en oversikt over de metoder som er foreslått. Dessuten skal vi se på spørsmålet om det er mulig å lage en utvalgsplan som legger forholdene bedre til rette for å kunne estimere fylkestall. På dette punktet har vi ikke vært i stand til å finne arbeider utført andre steder.

I kapittel 2 gis en oversikt over de viktigste problemer knyttet til estimering av fylkestall på grunnlag av data samlet inn ved hjelp av Byråets utvalgsplan. Problemene belyses ved hjelp av sysselsettingstall for Troms og Finnmark.

I kapittel 3 tar vi opp spørsmålet om det er mulig å endre utvalgsplanen med sikte på å kunne gi bedre sysselsettingstall.

I kapittel 4 gis en ikke-teknisk beskrivelse av de estimeringsmetoder som har vært brukt til nå, og i kapittel 5 foretas en sammenligning av metodene når formålet er å estimere sysselsettingstall for Troms og Finnmark.

I kapittel 6 diskuterer vi kort hvordan en kan gå fram når en ønsker å bruke fylke som forklaringsvariabel i en enkel lineær regresjon.

Konklusjonene er samlet i kapittel 7. Framstillingen er gjort så lite matematisk som mulig. I noen tilfeller har det ikke vært mulig helt å unngå bruk av matematisk statistikk. Disse tilfeller er behandlet i vedleggene, og bare konklusjonene er tatt med på de relevante steder i hovedkapitlene.

2. OVERSIKT OVER DE VIKTIGSTE PROBLEMENE KNYTTET TIL ESTIMERING AV FYLKESTALL PÅ GRUNNLAG AV UTVALGSUNDERSØKELSER

2.1. Utvalgsplan og -størrelse

Den mest alvorlige vanskelighet med å estimere fylkestall på grunnlag av utvalg er knyttet til størrelsen på utvalgene. I tabell 1 er gitt antall observasjoner en kan vente å få fra hvert fylke med den utvalgsplan som Byrået bruker i dag.

Tabell 1. Oversikt over forventet utvalgsstørrelse innen fylkene ved forskjellige størrelser på det totale utvalget

Fylke	Total utvalgsstørrelse		
	2 000	5 000	10 000
Østfold	114	285	570
Akershus	180	450	899
Oslo	223	557	1 115
Hedmark	92	229	458
Oppland	88	221	442
Buskerud	105	262	524
Vestfold	91	228	456
Telemark	79	198	396
Aust-Agder	44	110	220
Vest-Agder	67	166	333
Rogaland	148	371	741
Hordaland	191	479	957
Sogn og Fjordane	52	129	258
Møre og Romsdal	116	289	578
Sør-Trøndelag	119	299	597
Nord-Trøndelag	61	154	307
Nordland	120	299	598
Troms	72	179	358
Finnmark	39	96	193

En ser at selv med et utvalg på 10 000 enheter, vil det være få observasjoner fra de minste fylker. På den andre siden kan en av tabell 1 se at utvalgsstørrelsen i de største fylker er relativt stor, og det er naturlig å spørre om det er hensiktsmessig å redusere antall observasjoner i de største fylkene, og foreta en tilsvarende økning av utvalgsstørrelsene i de mindre fylker. En slik utvalgsplan er brukt i Byrået, og vil bli nærmere beskrevet i neste avsnitt.

Det andre problemet knyttet til estimering av fylkestall skyldes at Byråets utvalg trekkes i to trinn, og at en ved konstruksjon av utvalgsplanen ikke la spesiell vekt på ønsket om å kunne gi fylkestall. Problemet lar seg lettest beskrive ved å se på utvalgsplanen for Troms og Finnmark.

I Byråets utvalgsplan er Troms og Finnmark valgt som et super-stratum. D.v.s. en har på forhånd bestemt at utvalgene skal være representative for disse to fylkene under ett. Innenfor dette super-stratum er kommunene stratifisert etter forskjellige variable som størrelse, kommunetype og beliggenhet. Disse strata krysser fylkesgrensene. Kommunene i Finnmark er inneholdt i 4 forskjellige strata, hvorav to også inneholder kommuner fra Troms. Dersom en bruker de vanlige trekkemetoder, vil derfor de observerte tall for Finnmark og Troms være beheftet med en liten skjevhet. Det er vanskelig å si noe generelt om hvor store slike skjevheter er. Ved hjelp av tall fra Folke- og boligtellingen 1970 er det likevel mulig å beregne skjevheten for de variable som var med i denne tellingen. I tabell 2 er gitt en oversikt over skjevhetene til visse variable.

Tabell 2. Skjevheten til tall for Finnmark for et utvalg av variable fra Folke- og boligtellingen 1970

Næring	Andelen sysselsatte. Tall fra Folke- og boligtellingen 1970	Forventet sysselsetting i utvalg trukket etter Byråets utv. plan	Skjevhet
Totalt	.5251	.5252	.0001
Jord- og skogbruk	.0368	.0346	-.0022
Fiske og hvalfangst	.0613	.0556	-.0057
Industri m.v.	.1248	.1260	.0012
Bygg og anlegg	.0505	.0509	.0004
Varehandel	.0539	.0552	.0013
Samferdsel	.0612	.0623	.0011
Tjenesteytende næringer	.1366	.1407	.0041

Som det framgår av tabell 2 varierer skjevheten mye mellom forskjellige variable. På den totale sysselsetting er det praktisk talt ingen skjevhet, mens skjevheten for variabelen "Andel sysselsatte i fiske og hvalfangst", utgjør nesten 10 prosent av det riktige tallet. Det er verdt å merke seg at disse skjevheter ikke påvirkes av størrelsen på utvalget, slik at skjevheten vil øke sin betydning i forhold til variansen når utvalget økes.

2.2. Hvor nøyaktige bør fylkestallene være?

For å vurdere hvordan en best skal løse problemene knyttet til estimering av fylkestall, må en først bestemme seg for et mål for usikkerheten, og deretter si noe om hvor nøyaktige resultatene må være for å være til nytte. Begge disse valg må delvis baseres på skjønn.

Når det gjelder mål for nøyaktigheten, er det vanlig å bruke utvalgsvariansen. For mange av de estimeringsmetoder som er foreslått nedenfor, gjelder det at de har en mindre skjevhet, og det er da vanlig å bruke bruttovariansen som mål for nøyaktighet. Bruttovariansen fås som en sum av variansen og skjevheten kvadrert. Det er da også dette målet som skal brukes i det følgende når flere estimerings- og utvalgsmetoder skal sammenlignes.

Spørsmålet om hvor gode fylkestallene bør være for å kunne publiseres er det naturligvis vanskelig å svare på, men i det følgende skal beskrives et kriterium som tidligere er blitt brukt i Byrået (Laake og Langva (1976)). Bakgrunnen for dette kriteriet er at en ikke ønsker å publisere tall som er sterkt beheftet med utvalgsvarians. En velger derfor ofte å publisere forventningsrette estimatorer dersom

$$\{Z_{1-\varepsilon/2}\sigma\}/F \leq 0{,}4,$$

hvor $Z_{1-\varepsilon/2}$ er $(1-\varepsilon/2)$-fraktilen i den standardiserte normalfordeling, F er det fylkestall en estimerer og σ er standardavviket til estimatoren for fylkestallet. Ved å sette $\varepsilon = 0.05$ får en kriteriet

$$\sigma/F \leq 0.2.$$

D.v.s. at standardavviket ikke bør være større enn 20 prosent av det tall en ønsker å estimere.

Flere av de estimeringsmetoder som skal vurderes i det følgende er ikke forventningsrette, og vi trenger derfor et kriterium som er tilpasset slike situasjoner. I Laake (1976) er det foreslått at et estimat kan publiseres dersom det maksimale konfidensavviket fra det riktige fylkestallet i forhold til estimanden selv er mindre enn 40 prosent. D.v.s.

$$\frac{Z_{1-\varepsilon/2}\sigma + |B|}{F} \leq 0.4,$$

hvor B er skjevheten til estimatoren. Dette kriteriet blir brukt i avsnitt 5.4. for å avgjøre hva som kan publiseres av tall for Troms og Finnmark.

3. UTVALGSPLANER SOM GIR BEDRE MULIGHETER FOR Å KUNNE ESTIMERE FYLKESTALL

3.1. Innledning

Som nevnt tidligere, skyldes noen av de problemene som er knyttet til oppdeling av utvalg etter geografiske kjennemerker at Byråets standard utvalgsplan i øyeblikket ikke legger forholdene spesielt godt til rette for det. Et spørsmål av stor interesse både for levekårsundersøkelsene og andre utvalgsundersøkelser er derfor i hvilken grad det er mulig gjennom utvalgsplanen å legge forholdene bedre til rette for å kunne lage fylkestall på grunnlag av utvalgsundersøkelsene. I dette avsnittet skal vi se på måter en kan gå fram på, og diskutere fordeler og ulemper ved de forskjellige framgangsmåtene.

3.2. Mindre endringer av den någjeldende utvalgsplanen

En enkel endring som ville fjerne de skjevheter som utvalget i dag er beheftet med, ville være å definere hvert fylke som eget stratum, og deretter trekke selvveiende utvalg slik som vi gjør i dag. Ulempene ved slike utvalgsmetoder er først og fremst at en på denne måten ville få store problemer med også å stratifisere kommunene etter størrelse og kommunetype, hvilket vanligvis er viktige bakgrunnsvariable som mange planleggere ønsker å oppdele utvalget etter. Som demonstrert i kapittel 2, er skjevheten på fylkestallene meget små i forhold til utvalgsstørrelsen, og vi mener derfor at gevinsten ved å definere fylket som superstratum ikke står i forhold til det tap en vil få ved ikke å kunne oppdele utvalget etter kommunetype og størrelse på kommunen.

3.3. Stratifisert ikke selvveiende utvalg

Med den någjeldende utvalgsplan er det slik at det antall observasjoner en får fra et bestemt fylke, er tilnærmet proporsjonalt med antall bosatte i fylket. Dette medfører at vi fra små fylker får få observasjoner og omvendt. Dersom det er viktig at fylkestallene hver for seg er gode, ville det ofte være mer hensiktsmessig å trekke like mange observasjoner fra samtlige fylker. Dette kan enkelt gjøres ved å la trekkesannsynligheten variere fra fylke til fylke. I ungdomsundersøkelsen 1980 er en slik utvalgsplan blitt brukt. En stratifiserte fylkene etter størrelse i fem strata, og varierte trekkesannsynlighetene mellom strata. Stratainndelingen, trekkesannsynligheter og forventet antall personer som skulle trekkes fra hvert fylke, er beskrevet i det følgende.

	Tallet på ungdom fra 17 år og til og med 19 år pr. 31. desember 1978	
	Stratum 1	Forventet antall i utvalget
Hordaland	18 480	573
Oslo	16 878	523
Akershus	16 973	526
I alt	52 331	1 622
Trekkesannsynlighet: 0.031		
	Stratum 2	
Rogaland	14 399	670
Møre og Romsdal	11 693	544
Nordland	11 737	546
Sør-Trøndelag	10 840	502
I alt	48 633	2 261
Trekkesannsynlighet: 0.0465		
	Stratum 3	
Østfold	10 168	630
Buskerud	8 989	557
Vestfold	8 712	540
Oppland	8 153	505
Hedmark	8 080	501
I alt	44 102	2 734
Trekkesannsynlighet: 0.062		

	Tallet på ungdom fra 17 år og til og med 19 år pr. 31. desember 1978	
	Stratum 4	Forventet antall i utvalget
Troms	6 916	643
Telemark	6 881	640
Vest-Agder	6 443	599
Nord-Trøndelag	5 931	552
Sogn og Fjordane	5 226	486
I alt	31 397	2 920
Trekkesannsynlighet: 0.093		
	Stratum 5	
Aust-Agder	4 006	497
Finnmark	3 909	485
I alt	7 917	981
Trekkesannsynlighet: 0.124		

Utvalgsstørrelse: 10 518

Tall for hele landet fås ved å gi observasjonene i stratum 1 en vekt på 12, observasjonene i stratum 2 en vekt på 8, observasjonene i stratum 3 en vekt på 6, observasjonene i stratum 4 en vekt på 4 og observasjonene i stratum 5 en vekt på 3.

Som en ser vil denne utvalgsplanen gi tilnærmet like mange observasjoner for samtlige fylker, hvilket medfører at alle fylkestall vil få tilnærmet samme kvalitet, noe som er en fordel når en for eksempel skal sammenligne resultatene mellom fylker. I tillegg kan en også estimere tall for hele landet ved å veie observasjonene med en faktor omvendt proporsjonal med trekkesannsynligheten.

En slik utvalgsplan har uten tvil mange fordeler når ønsket om fylkestall har høy prioritet. På den andre siden er det sikkert også en del ulemper forbundet med den, hvorav de viktigste er:

(i) Variansen til landstallene blir større enn når utvalget blir allokert proporsjonalt med antall bosatte i fylkene. Varianstapet vil variere fra variabel til variabel, i tabell 3 har en ved hjelp av data for folke- og boligtellingen 1970 beregnet varianstapet for noen variable fra denne tellingen.

(ii) At utvalget ikke er selvveiende får en del konsekvenser som vi ennå ikke helt har oversikt over omfanget av. Enkle tabeller kan en kjøre ut ved å veie observasjonene, men når det gjelder analyse som regresjonsanalyse, log-lineær analyse og lignende, er det i dag ingen veletablert praksis når det gjelder å kjøre analyse på veide eller uveide data.

Tabell 3. Sammenligning av varianser til landstall når en henholdsvis bruker en utvalgsplan som i Ungdomsundersøkelsen 1980, og en utvalgsplan der en allokerer proporsjonalt med antall bosatte i strataene fra Ungdomsundersøkelsen 1980

Næring	Varianser til landstall ved en utvalgsplan som i Ungdomsundersøkelsen 1980	Varianser til landstall ved en utvalgsplan med prop.allokering	Variansøkning ved ikke prop.allokering regnet i prosent
Totalt	$\frac{1}{n} \cdot 3.080 \cdot 10^{-1}$*)	$\frac{1}{n} \cdot 2.495 \cdot 10^{-1}$	23.4
Jord- og skogbruk	$\frac{1}{n} \cdot 4.771 \cdot 10^{-2}$	$\frac{1}{n} \cdot 4.668 \cdot 10^{-2}$	2.2
Fiske og hvalfangst	$\frac{1}{n} \cdot 9.112 \cdot 10^{-3}$	$\frac{1}{n} \cdot 9.364 \cdot 10^{-3}$	-2.7
Industri m.v.	$\frac{1}{n} \cdot 1.525 \cdot 10^{-1}$	$\frac{1}{n} \cdot 1.236 \cdot 10^{-1}$	23.4
Bygg og anlegg	$\frac{1}{n} \cdot 5.089 \cdot 10^{-2}$	$\frac{1}{n} \cdot 4.290 \cdot 10^{-2}$	18.6
Varehandel	$\frac{1}{n} \cdot 8.646 \cdot 10^{-2}$	$\frac{1}{n} \cdot 6.405 \cdot 10^{-2}$	35.0
Samferdsel	$\frac{1}{n} \cdot 6.529 \cdot 10^{-2}$	$\frac{1}{n} \cdot 5.158 \cdot 10^{-2}$	26.6
Tjenesteytende næringer	$\frac{1}{n} \cdot 1.542 \cdot 10^{-1}$	$\frac{1}{n} \cdot 1.159 \cdot 10^{-1}$	33.0

*) n = total utvalgsstørrelse

3.4. Metoder for supplering av utvalgsplanen etter behov

En annen måte å nærme seg problemet på består i å finne måter en kan supplere utvalgsplanen på, slik at utvalgene blir representative innen hvert fylke. I Finnmark, for eksempel, er de primære utvalgsområder som er trukket ikke representative for fylket. En kan spørre om det er mulig å trekke en eller flere tilleggskommuner i Finnmark slik at vi får et selvveiende, representativt utvalg for Finnmark, og samtidig bruke intervjukorpset i Finnmark så fornuftig som mulig. Svaret på dette spørsmålet er at det er mulig å trekke tilleggskommuner slik at utvalget for hvert fylke blir representativt. I vedlegg 1 er metoden og visse egenskaper ved den beskrevet i detalj. Her skal vi bare nevne et par fordeler og ulemper ved metoden:

(i) Metoden er meget generell, og kan brukes i mange tilfeller hvor en ønsker representative utvalg for spesielle grupper i befolkningen, og samtidig få maksimal utnyttelse av den intervjuerstab som Byrået disponerer. For eksempel hvis en ønsker et utvalg som er selvveiende for alle fiskerikommuner i landet, kan en bruke en tilsvarende metode og trekke en eller flere tilleggskommuner. Metoden kan også brukes ved utskifting av kommuner i Byråets utvalgsplan, noe som er til stor nytte ved en eventuell justering av utvalgsplanen etter folke- og boligtellingen 1980.

(ii) Med den begrensede etterspørsel som eksisterer etter fylkestall i dag, er det ikke å vente at Byrået kan ansette faste intervjuere i eventuelle tilleggskommuner. En må derfor regne med at innsamlingskostnadene i tilleggskommunene vil bli noe høyere enn for kommuner som er med i Byråets utvalgsplan.

(iii) I kapittel 5 skal det vises at gevinsten ved å bruke supplering av utvalgsplanen med sikte på å estimere fylkestall er meget moderat for de fleste variabler med de utvalgsstørrelser som det er realistisk å regne med i forbindelse med levekårsundersøkelsen. Årsaken til dette er at den skjevhet i fylkestallene som vi nevnte i avsnitt 3.2. er liten i forhold til usikkerhetene som skyldes størrelsen på utvalget.

(iv) Metoden er så vidt vi vet ikke tidligere blitt brukt slik at det er nødvendig å avvente en del erfaringer med den. Det er planer om å bruke suppleringsmetoden i forbindelse med trekking av utvalg av fiskere. De kommuner som i dag finnes i Byråets utvalgsplan kan ikke sies å være representative for fiskerikommuner i Norge. En har derfor planer om å trekke et suppleringsutvalg av fiskerikommuner i tillegg til de fiskerikommuner som i dag er med i Byråets utvalgsplan, og konsentrere intervjuingen innen disse kommuner. Hvis denne undersøkelsen blir gjennomført vil erfaringene med bruk av metoden bli grundig vurdert for senere bruk.

3.5. Foreløpige konklusjoner når det gjelder å bruke spesielle utvalgsmetoder for å kunne gi fylkestall

Som konklusjon på det arbeid som er beskrevet i dette kapittel, er det rimelig foreløpig å anta at det med de utvalgsstørrelser som er realistiske for levekårsundersøkelsene, er lite å hente når det gjelder å trekke utvalgene på en måte som legger forholdene bedre til rette for å kunne gi fylkestall. Når det derimot gjelder å lage spesielle levekårsundersøkelser innen geografisk konsentrerte grupper av befolkningen, ser det ut som om metoden med å supplere Byråets någjeldende utvalgsplan virker meget tiltalende.

4. OVERSIKT OVER DE VIKTIGSTE ESTIMERINGSMETODER SOM BRUKES I FORBINDELSE MED ESTIMERING AV FYLKESTALL

4.1. Innledning

Som nevnt i innledningen er det gjennom de siste 10 årene gjort en del forsøk på å estimere fylkestall på grunnlag av utvalgsdata. Levy (1979), Purcell and Kish (1979), Laake (1976, 1977, 1978). Stort sett kan metodene inndeles i tre typer:

(i) Direkte estimering

(ii) Estimering på grunnlag av en modell

(iii) Kombinasjoner av (i) og (ii).

I dette kapittel skal gis en kort beskrivelse av metodene og vi skal knytte noen kommentarer til hver av dem. Det viser seg at det i praktiske situasjoner ofte er meget vanskelig å gi generelle svar på hvilke metoder som er best, og i neste kapittel skal vi derfor foreta noen numeriske beregninger på data fra Folke- og boligtellingen 1970.

4.2. Inndeling av estimeringsmetodene

(i) Direkte estimering

Den enkleste metode består av å estimere et fylkestall på grunnlag av de observasjoner en har fra fylket. Som nevnt ovenfor vil en med en utvalgsplan som Byråets måtte vente at denne estimeringsmetode fører til skjeve resultater, og vil i tillegg ha stor varians for de mindre fylker på grunn av utvalgsstørrelsen. Skjevheten kan fjernes ved å trekke en eller flere tilleggskommuner slik som nevnt i avsnitt 3.3. Utvalgsstørrelsen i de små fylker kan økes noe på bekostning av de større fylker ved å lage en ikke-selvveiende utvalgsplan som vist i avsnitt 3.2. Likevel er det klart at med de utvalgsstørrelser det er realistisk å regne med i levekårsundersøkelsene vil denne estimeringsmetode, selv med de modifikasjoner av utvalgsplanen som er nevnt, ikke gi fylkestall av tilfredsstillende kvalitet. Nedenfor skal vi foreslå en annen estimator, hvor direkte estimering inngår sammen med en annen estimator.

(ii) Estimering på grunnlag av en modell

En stor klasse av estimeringsmetoder som er blitt utprøvd, er basert på antakelser om at den variabel en ønsker å estimere på fylkesnivå, X, er avhengig av en eller flere variable $Z_1, \ldots, Z_K$, som en har informasjon om fra andre kilder. La f.eks. følgende sammenheng X og Z_1 være alminnelig akseptert:

$$X_i = \beta Z_{1i} + \alpha + U_i, \quad i = 1, 2, \ldots, N.$$

U_i er tilfeldig feil med forventning 0, og N er antall individer. La oss dessuten anta at Z_1 er kjent for alle individer i landet, mens en har målinger for variabel X bare for et utvalg. I utvalget estimeres først α og β. La estimatene være $\hat{\alpha}$ og $\hat{\beta}$. For å estimere tall for et bestemt fylke, finner en summen av alle verdiene for variabel Z innen fylket, og multipliserer summen med $\hat{\beta}$ og legger til $N_1\hat{\alpha}$, hvor N_1 er antall personer i fylket. Estimatoren for summen av X-verdiene i fylket, F_X, blir da

$$\hat{F}_X = N_1\hat{\alpha} + \hat{\beta} \sum_{i \in F} Z_{1i},$$

hvor $\sum_{i \in F}$ betyr summen over alle individer i fylket.

Estimatoren bygger altså delvis på en sammenheng mellom variablene, og delvis på at en har kjennskap til en eller flere av disse variablene fra andre kilder, f.eks. et register eller en totaltelling. Vi kaller den heretter for regresjonsestimatoren.

Nå er det klart at estimatoren F_X sammenlignet med en direkte estimator er god når modellen for sammenhengen er god, derimot kan estimatoren være litt av en katastrofe dersom den forutsatte modell er dårlig og kvaliteten til de verdier for Z en har adgang til er dårlig.

I praksis har en det problemet at en ofte ikke kan si noe om hvor god modellen er, og om en innfører skjevheter ved å bruke den. Dessuten er det sterkt begrenset hvor mange variable en har informasjon om i registrene. Nytten av å bruke slike metoder har derfor variert mye fra felt til felt. En av de mest heldige anvendelser er estimering av endringer i befolkningen innen counties i USA i perioden 1960 til 1970. En tok her utgangspunkt i Folketellingen 1960 og framskrev befolkningen ved hjelp av fødsels- og døds- og flytterater estimert i arbeidskraftundersøkelsene. Avvikene mellom de estimerte tall og resultatene fra Folketellingen 1970 var i alminnelighet ganske moderate.

Mulighetene for å lage estimatorer basert på modeller er naturligvis like store som mulighetene for å lage modeller, og et praktisk problem i forbindelse med levekårsundersøkelsene er at for noen variable vil en type metode være god, mens en for andre variable må lage andre metoder. Det er rimelig å tro at en først vil ha full oversikt over slike metoder når en har foretatt grundig analyse av flere undersøkelser.

Imidlertid er det en modellbasert metode som etter vår mening vil kunne brukes i mange tilfeller i forbindelse med levekårsundersøkelsene. Metoden kalles syntetisk estimering, og har vært brukt i forbindelse med mange undersøkelser i USA de siste 12 årene. I 1968 publiserte National Center for Health Statistics tall for antall sykedager og andre mål for helse for hver stat i USA. Siden 1976 har Statistisk Sentralbyrå publisert fylkestall fra arbeidskraftundersøkelsene. I begge disse tilfeller ble det brukt syntetisk estimering.

La oss gi en kort beskrivelse av hva som menes med syntetisk estimering:

Anta at en for en stor region (som kan være hele landet) har forventningsrette estimatorer for forekomsten av et fenomen i flere aldersgrupper. (Prinsippet blir det samme om en bruker andre grupperingsvariable enn alder.) La disse estimatorer være $\hat{P}_1$, $\hat{P}_2$, ..., $\hat{P}_L$. Den syntetiske estimator for forekomsten av fenomenet innen et bestemt fylke fås nå ved å veie sammen estimatene $\hat{P}_1$, $\hat{P}_2$, ..., $\hat{P}_L$. Vekten foran $\hat{P}_i$ er den relative andel av personene i fylket som tilhører aldersklasse i. Estimatoren er da

$$F = \sum_{i=1}^{L} \hat{P}_i W_i,$$

hvor W_i er den relative andel av befolkningen i aldersklasse i innen fylket. Det er klart at denne estimatoren er god dersom forekomsten av fenomenet er høyt korellert med alder. På den andre siden er den mindre god dersom forekomsten av fenomenet varierer mye fra fylke til fylke innen aldersgruppene.

Statistiske egenskaper til den syntetiske estimatoren, slik som varians, forventning og bruttovarians, og metoder for å estimere disse parametrene har blitt drøftet av Gonzalez og Waksberg (1978) og av Levy og French (1977). Varians og forventningsskjevhet, B_F, til den syntetiske estimatoren er gitt ved:

$$\text{var}(F) = \sum_{i=1}^{L} W_i^2 \, \text{var}(\hat{P}_i) + 2 \sum_{i<j} W_i W_j \, \text{cov}\,(\hat{P}_i, \hat{P}_j)$$

$$B_F = \sum_{i=1}^{L} W_i (P_i - P_{iF}),$$

der P_i, $i = 1, \ldots, L$ er den virkelige forekomsten i aldersklasse i i den "store" regionen og P_{iF}, $i = 1, \ldots, L$ er den tilsvarende forekomsten i fylket.

I de fleste tilfeller der en bruker en syntetisk estimator er den "store" regionen så stor at $\hat{P}_i$, ..., $\hat{P}_L$ alle blir basert på mange observasjoner. Den syntetiske estimatoren har derfor som regel liten samplingvarians.

Av formelen for skjevheten til den syntetiske estimatoren ser vi at skjevheten er en veiet sum av differansene mellom forekomstene i den "store" regionen og fylket innen forskjellige aldersklasser. Denne skjevheten kan være av betydelig størrelsesorden hvis forekomsten av fenomenet vi skal estimere, varierer mye fra fylke til fylke innen aldersgruppene. Det faktum at skjevheten kan være stor, uten at en i en praktisk situasjon kan vite noe om det, er svakheten til den syntetiske estimatoren.

(iii) Metoder basert på en kombinasjon av metodene (i) og (ii)

Det er naturlig å spørre om en ikke kan lage en estimator, som er en lineærkombinasjon av en direkte estimator og en syntetisk estimator. Kvaliteten til den direkte estimator $\hat{F}_d$ er som før nevnt meget avhengig av hvor mange observasjoner en har fra det aktuelle fylket, mens kvaliteten til den syntetiske estimatoren, $\hat{F}_s$, avhenger nesten utelukkende av hvor god modellen er. I store fylker med mange observasjoner kan det derfor tenkes at den direkte estimatoren er god, mens den syntetiske estimatoren er best i mindre fylker. Det er derfor naturlig å se på en lineærkombinasjon av de to. La

$$\hat{F}_k = C\hat{F}_d + (1-C)\hat{F}_s,$$

hvor C bestemmes slik at $\hat{F}_k$ får minst mulig bruttovarians. Vet et "optimalt" valg av C er alltid $\hat{F}_k$ minst like så god som den beste av estimatorene $\hat{F}_d$ og $\hat{F}_s$. Den optimale verdi av C, Cop, er slik at $\hat{F}_k$ vil legge størst vekt på den av estimatorene $\hat{F}_d$ og $\hat{F}_s$ som har minst bruttovarians. Hvis f.eks. forventningsskjevheten til $\hat{F}_s$ er liten, vil Cop ligge nær 0, siden variansen til $\hat{F}_s$ er langt mindre enn variansen til $\hat{F}_d$ ($\hat{F}_s$ tenkes å være basert på et stort antall observasjoner). I slike tilfeller vil altså $\hat{F}_k$ praktisk talt være identisk med $\hat{F}_s$.

I Schaible, Broch and Schnack (1977) er det vist at når bruttovariansen til $\hat{F}_s$ og $\hat{F}_d$ er like store, kan en oppnå å få redusert bruttovariansen med opptil 50 prosent ved å bruke $\hat{F}_k$ i stedet for F_s eller F_d. Gevinsten ved å bruke $\hat{F}_k$ i stedet for den beste av estimatorene $\hat{F}_s$ og $\hat{F}_d$ avtar når forskjellen i bruttovarians mellom $\hat{F}_s$ og $\hat{F}_d$ øker.

Cop er en funksjon av bruttovariansene til $\hat{F}_d$ og $\hat{F}_s$, som må estimeres når en bare har utvalgsdata til disposisjon. I en praktisk situasjon kan en derfor ikke regne med å finne den optimale verdi for C. I Schaible (1978) er det imidlertid vist at kvaliteten til $\hat{F}_k$ er robust for forskjellige valg av C, og i Schaible, Brock and Schnack (1977) er det presentert empiriske studier som tyder på at $\hat{F}_k$ som regel vil være bedre enn både $\hat{F}_d$ og $\hat{F}_s$, selv om en bruker en dårlig estimator for Cop.

I neste kapittel skal $\hat{F}_k$ sammenlignes med $\hat{F}_d$ og $\hat{F}_s$ ved forskjellige utvalgsplaner.

5. SAMMENLIGNING AV KVALITETEN TIL FORSKJELLIGE ESTIMERINGS- OG UTVALGSMETODER

5.1. Innledning

I kapitlene 3 og 4 har vi presentert noen metoder som kan brukes for å skaffe bedre fylkestall på grunnlag av utvalgsdata, samt nevnt visse fordeler og ulemper ved de forskjellige metodene. I dette kapittel skal vi supplere disse betrakninger ved å utprøve noen av de foreslåtte framgangsmåtene på to datasett. Som det ene datasett har vi valgt Folke- og boligtellingen 1970, og som det andre Helseundersøkelsen 1975. Ved hjelp av metodene diskutert i kapitlene 2 og 4, skal vi estimere tall i de to fylkene Troms og Finnmark.

Ved bruk av data fra Helseundersøkelsen kan en bare studere enkelte sider ved de forskjellige metodene, f.eks. er det ikke mulig å estimere skjevheten til den syntetiske estimatoren.

Årsaken til at vi har valgt å bruke data fra folketellingen, er at dette er en totaltelling, hvilket gjør det mulig å studere de forskjellige metodene i alle detaljer. For en gitt utvalgsstørrelse kan en altså utføre nøyaktige beregninger for skjevheter og varianser til de forskjellige estimatorer. Videre kan en finne den verdi av C som minimerer bruttovariansen til $\hat{F}_k$ (jfr. kap. 4.2. (iii)). På tross av at en bare kan studere et meget begrenset sett av variable ved hjelp av folketellingen, mener vi at resultatene for dette begrensede sett av variable er nyttige når det gjelder å vurdere de forskjellige metodene for andre variable.

5.2. Sammenligning av forskjellige metoder når en ønsker å estimere noen sysselsettingstall for Troms og Finnmark

Vi skal nå tenke oss at målet er å estimere de fire parametrene andelen som er sysselsatt, andelen som er sysselsatt innen jord- og skogbruk, andelen ansatte innen fiske og hvalfangst og andelen ansatte innen varehandel, for hvert av de to fylkene Troms og Finnmark. De estimeringsmetodene som skal vurderes er direkte estimering, syntetisk estimering og en lineær-kombinasjon av disse to. Samtlige estimatorer skal vurderes under to forskjellige utvalgsplaner, nemlig den någjeldende utvalgsplan, og den utvalgsplanen som er beskrevet i avsnitt 3.4. og vedlegg 1, i det følgende kalt ny utvalgsplan. Den någjeldende utvalgsplanen blir i dette avsnittet kalt gammel utvalgsplan. Dessuten skal vi anta at de eneste variable vi har opplysninger om i registeret er alder og kjønn til samtlige personer.

I Laake og Longva (1976) er det gitt resultater som viser at for samtlige fire variablers vedkommende, er den syntetiske estimatoren en får ved å bruke alder og kjønn som grupperingsvariable, ikke bedre enn den estimatoren en får når en ser bort fra alders- og kjønnsgrupperingen. Som representant for de syntetiske estimatorer velger vi oss derfor ganske enkelt andelen sysselsatte i en større region som kan være fylkesparet Troms/Finnmark, Nord-Norge eller hele landet. Dette gjelder samtlige fire variable studert i dette avsnittet. De tre estimatorer som derfor skal vurderes er:

(i) Andelen observert i vedkommende fylke (Den direkte estimator).
(ii) Andelen observert i en større region enn vedkommende fylke. (Den syntetiske estimator).
(iii) En lineærkombinasjon av (i) og (ii).

Nedenfor skal disse tre estimatorer vurderes for ny og gammel utvalgsplan.

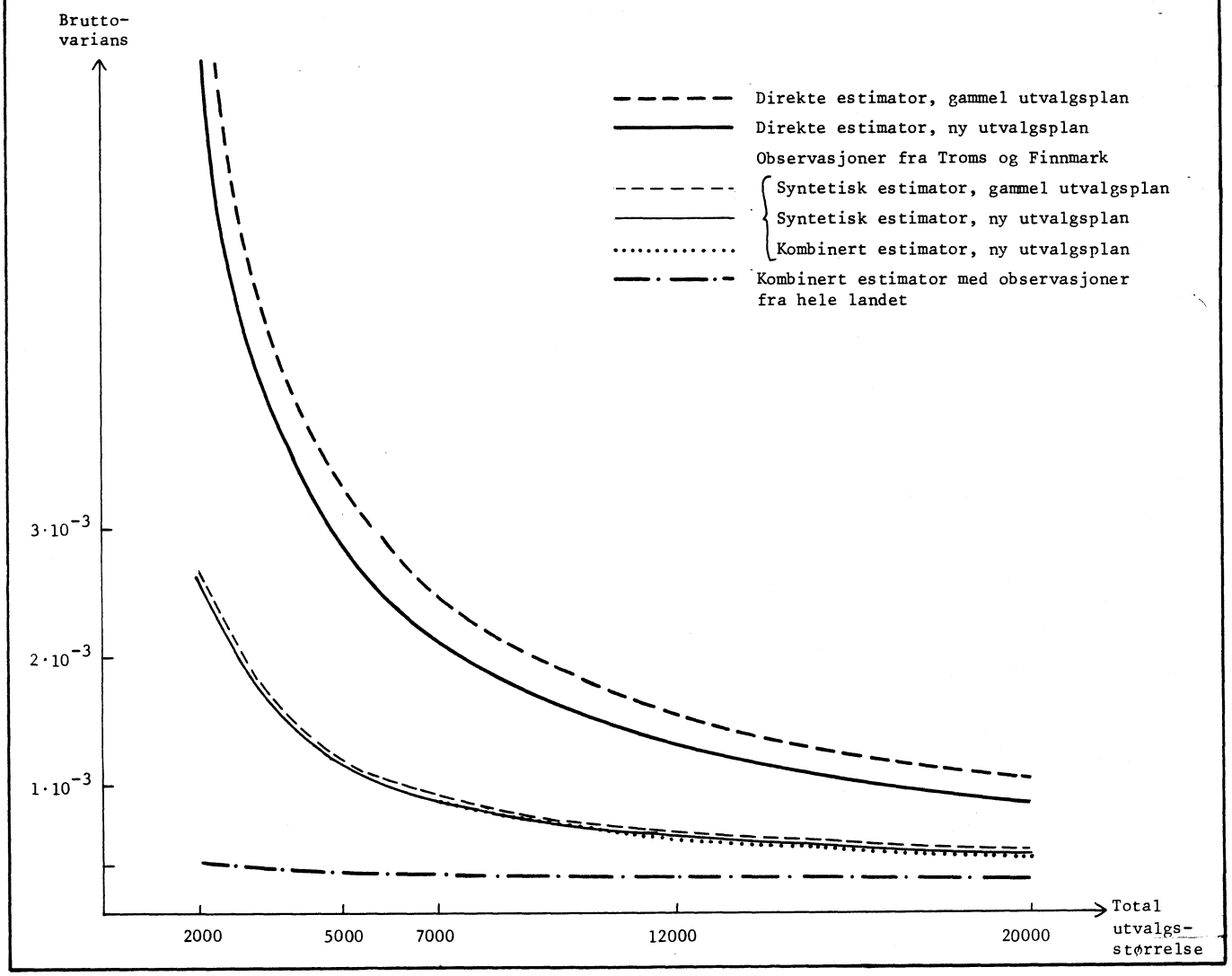
5.2.1. Bruttovarianser for forskjellige estimatorer for andelen sysselsatte i Troms og Finnmark

I fig. 5.1. og 5.2. er bruttovariansen vist som funksjon av total utvalgsstørrelse for noen forskjellige framgangsmåter. Både for Troms og Finnmark er den direkte estimator dårligst uansett hvilken utvalgsplan som brukes. Den beste estimator er i begge fylker å lage en lineærkombinasjon av andelen sysselsatte i hele landet og andelen sysselsatte i det aktuelle fylket. I figurene kan en lese at det gir stor gevinst å bruke denne estimatoren, som vi heretter kaller den kombinerte estimatoren, framfor den direkte estimator.

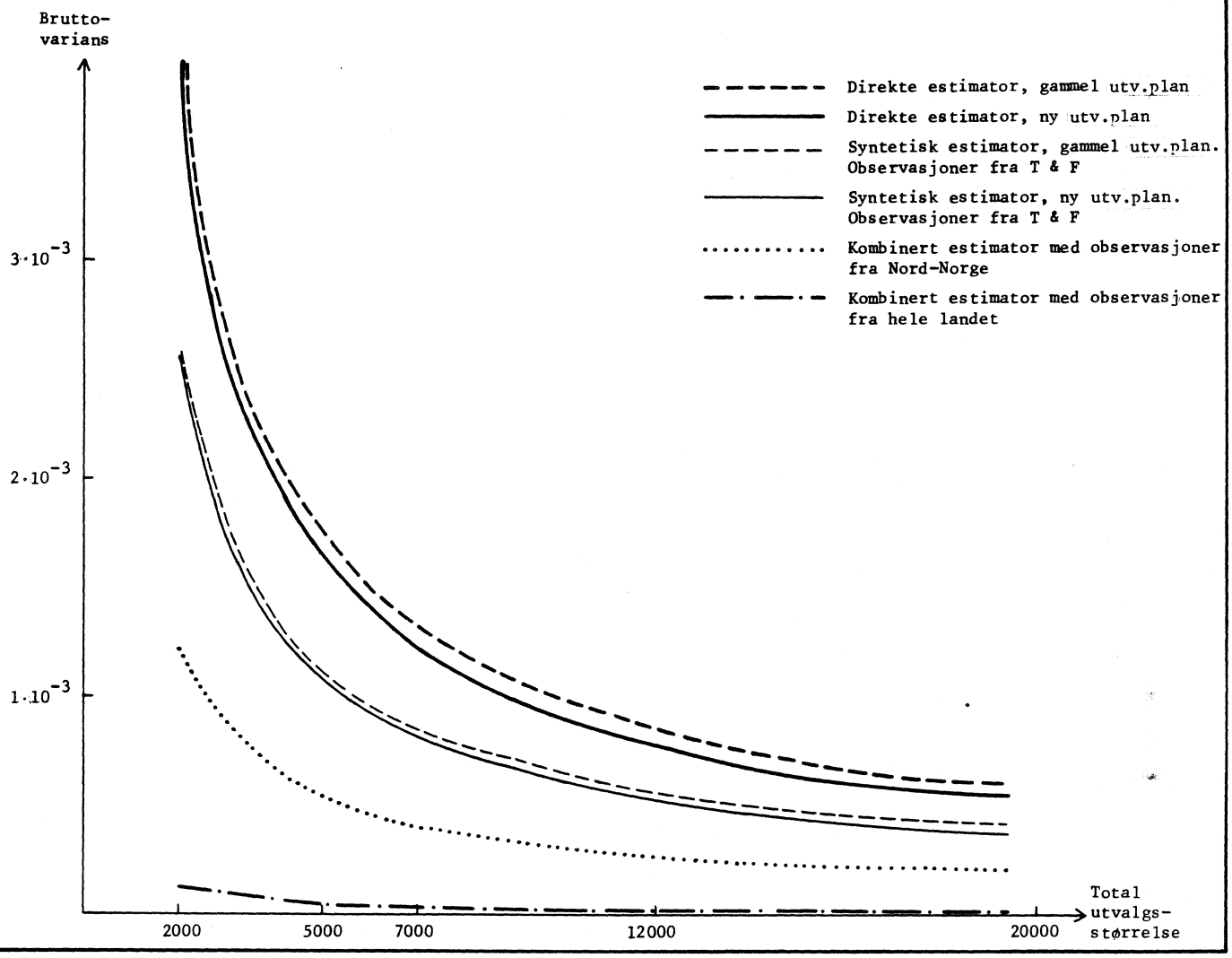
Når det gjelder forholdet mellom den kombinerte og den syntetiske estimator, er det så liten forskjell mellom disse to estimatorene, at kurvene for bruttovariansene til de to estimatorene nesten ville falle helt sammen dersom de begge ble tegnet inn på figurene. Dette kommer av at den syntetiske estimatoren er så mye bedre enn den direkte estimatoren, at den kombinerte estimatoren legger nesten all vekt på den syntetiske estimatoren. Dette har igjen sammenheng med at andelen sysselsatte i Troms og Finnmark bare skiller seg lite fra andelen sysselsatte i hele landet.

For den syntetiske estimator, og følgelig også for den kombinerte estimator betyr det praktisk talt ingenting om en bruker ny eller gammel utvalgsplan. Dette kommer av at skjevheten til den syntetiske estimatoren er den samme ved de to utvalgsplanene, og at det betyr lite for variansen om en bruker ny eller gammel utvalgsplan når en skal estimere tall for hele landet.

Figur 5.1. Bruttovariansen til forskjellige estimatorer for andelen sysselsatt i Finnmark som funksjon av total utvalgsstørrelse



Figur 5.2. Bruttovariansen til forskjellige estimatorer for andelen sysselsatt i Troms som funksjon av total utvalgsstørrelse



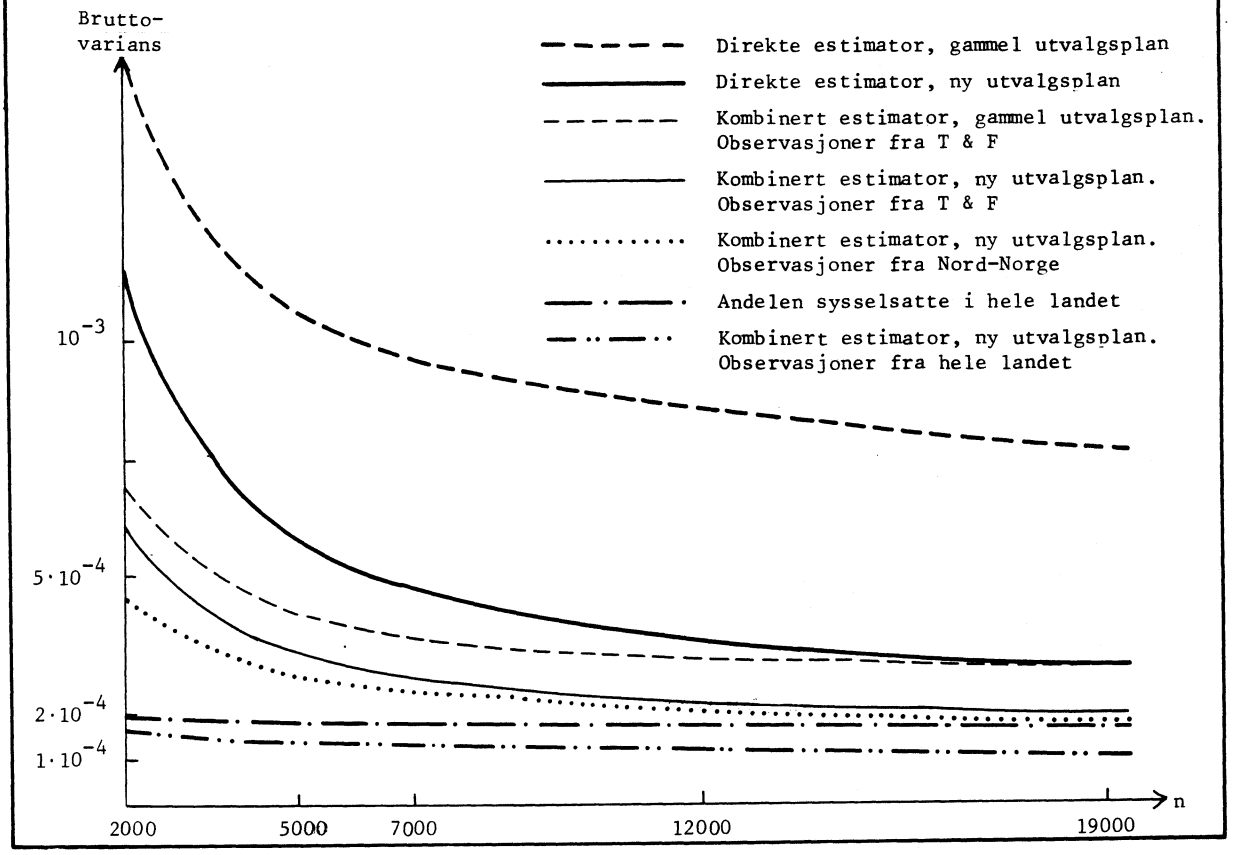
5.2.2. Bruttovarianser for forskjellige estimatorer for andelen sysselsatte i jord- og skogbruk i Troms og Finnmark

I fig. 5.3. og 5.4. er bruttovariansen som funksjon av total utvalgsstørrelse vist for forskjellige estimatorer og utvalgsplaner. Figurene viser at ved begge utvalgsplaner kan en få redusert bruttovariansen mye ved å bruke en kombinert estimator framfor den direkte estimatoren. For Troms vedkommende ser en at den kombinerte estimatoren med observasjoner fra Nord-Norge er best ved store utvalgsstørrelser, og at den kombinerte estimatoren med observasjoner fra hele landet er best for utvalgsstørrelser mindre enn 10 000. Når det gjelder Finnmark er den kombinerte estimatoren med observasjoner fra hele landet best for alle utvalgsstørrelser som vi har studert.

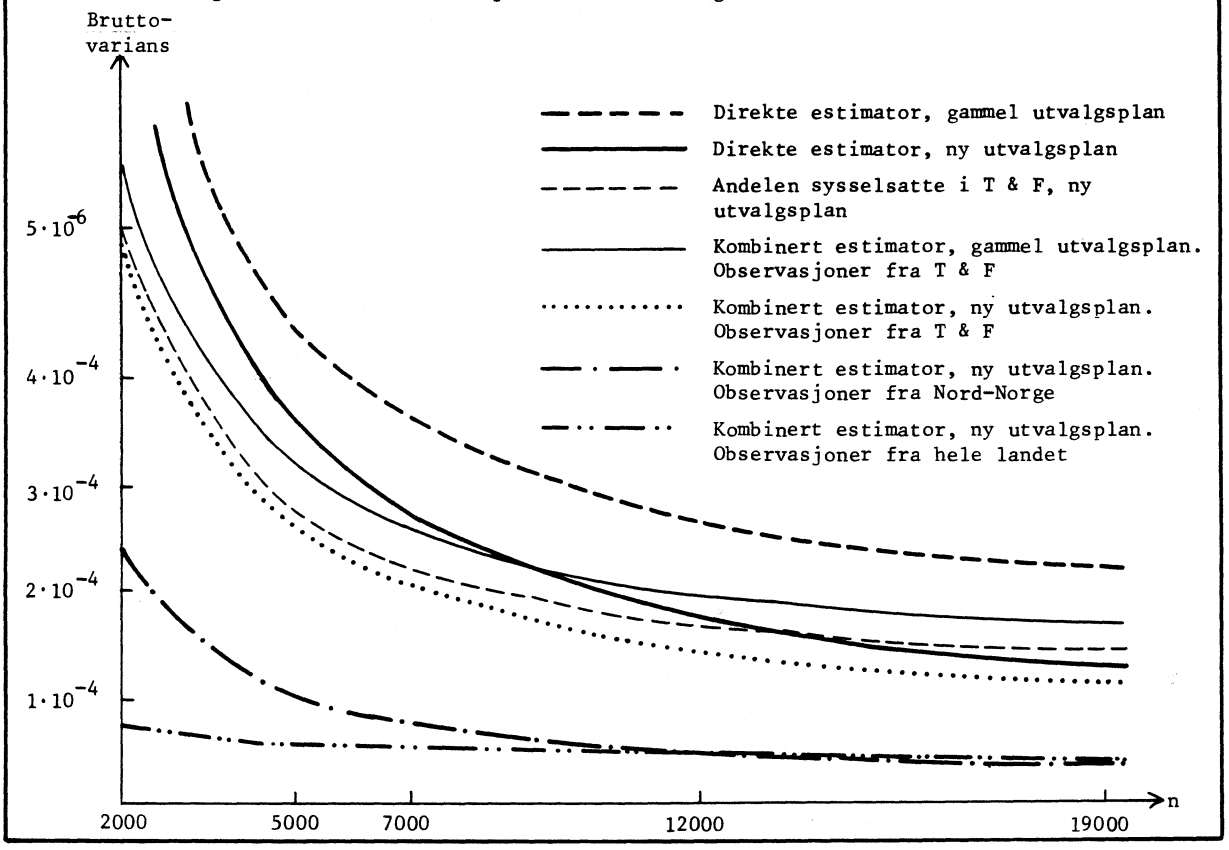
Ved estimering av andelen sysselsatte i jord- og skogbruk spiller utvalgsmetoden en noe større rolle enn tilfellet var for variabelen "andel sysselsatte". Dette skyldes at skjevheten til den direkte estimatoren ved gammel utvalgsplan er forholdsvis stor som vist i tabell 2, og at strataene av utvalgsområder blir mer homogene m.h.t. studievariabelen ved ny utvalgsplan enn ved gammel utvalgsplan. Det er likevel viktig å merke seg at utvalgsplanen betyr langt mindre når en bruker en kombinert estimator enn når en bruker den direkte estimatoren.

Når det gjelder forholdet mellom den syntetiske og den kombinerte estimatoren, er det ved estimering av andelen sysselsatte i jord- og skogbruk en del å tjene på å bruke den kombinerte estimatoren framfor den syntetiske estimatoren.

Figur 5.3. Bruttovariansen til forskjellige estimatorer for andelen sysselsatt i jord- og skogbruk i Finnmark som funksjon av total utvalgsstørrelse



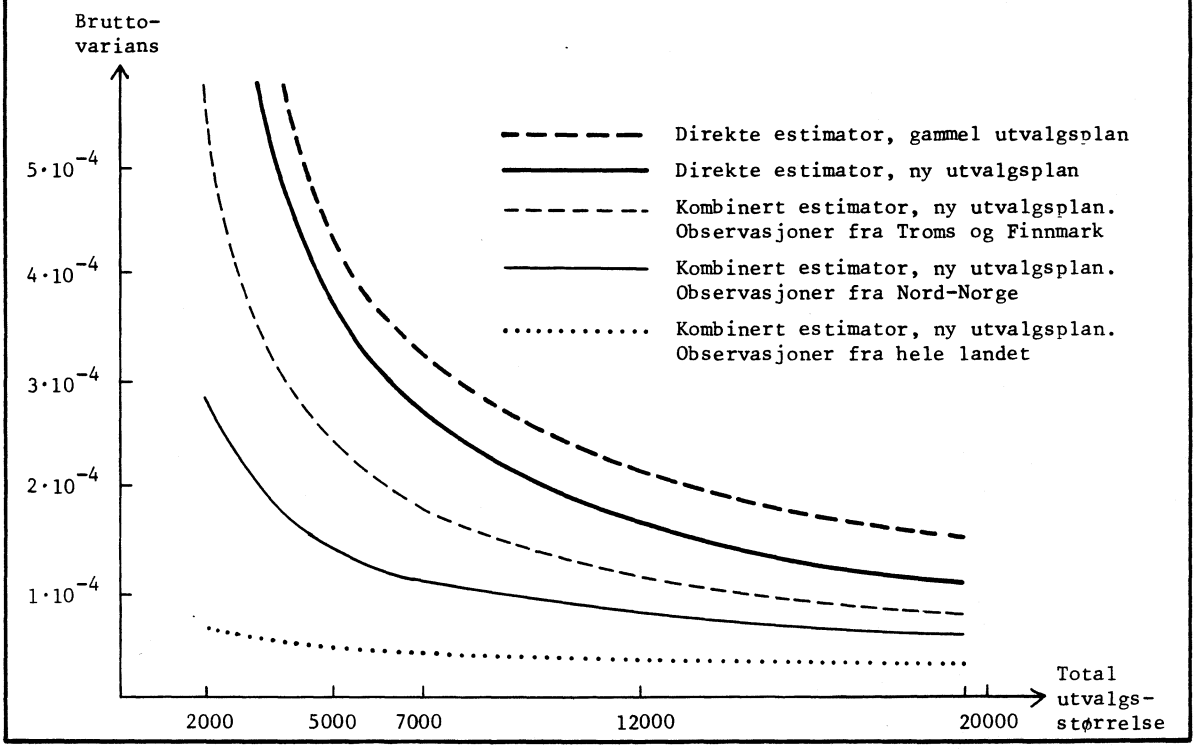
Figur 5.4. Bruttovariansen til forskjellige estimatorer for andelen sysselsatte i jord- og skogbruk i Troms som funksjon av total utvalgsstørrelse



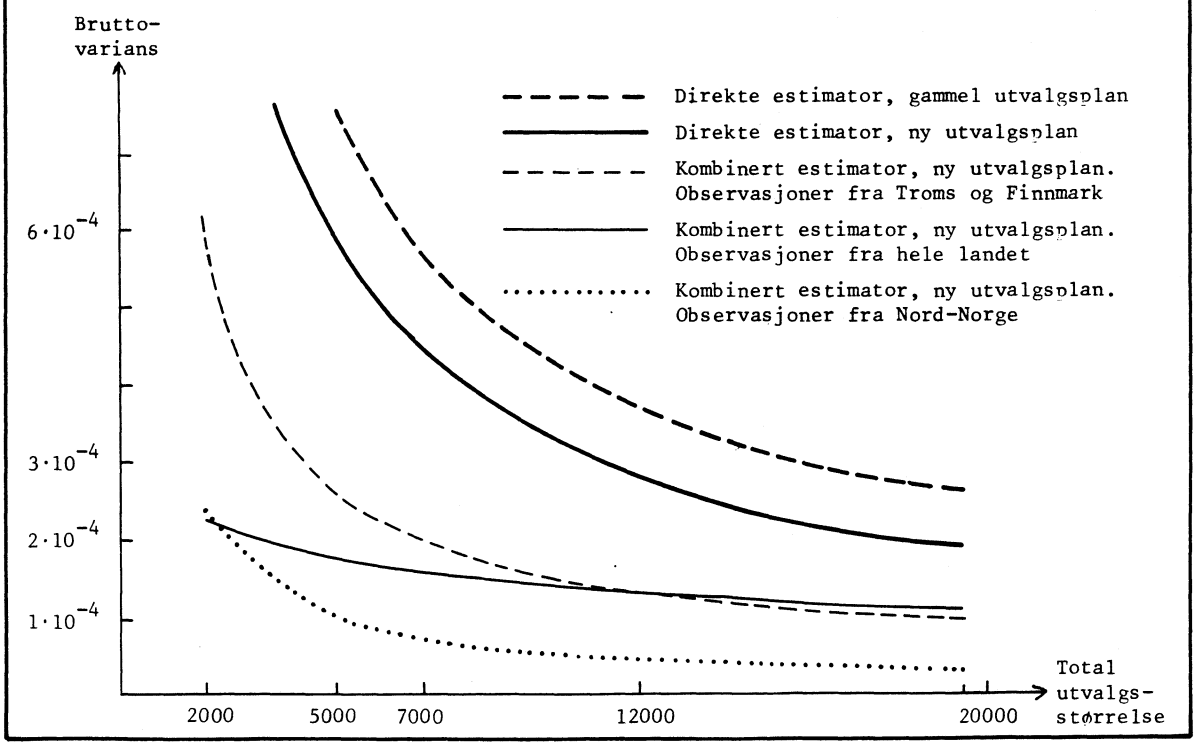
5.2.3. Bruttovarianser for forskjellige estimatorer for andelen sysselsatte i varehandel i Troms og Finnmark

På fig. 5.5. og 5.6. er bruttovariansen som funksjon av total utvalgsstørrelse vist for forskjellige estimatorer og utvalgsplaner. Igjen er den direkte estimatoren dårligst uansett utvalgsplan. For Troms vedkommende er den kombinerte estimatoren med observasjoner fra hele landet best, mens for Finnmarks vedkommende er den kombinerte estimatoren med observasjoner fra Nord-Norge best. Ved bruk av de kombinerte estimatorer gjelder det også her at det spiller liten rolle om en bruker ny eller gammel utvalgsplan.

Figur 5.5. Bruttovariansen til forskjellige estimatorer for andelen sysselsatte i varehandel i Troms som funksjon av total utvalgsstørrelse



Figur 5.6. Bruttovariansen til forskjellige estimatorer for andelen sysselsatt i varehandel i Finnmark som funksjon av total utvalgsstørrelse



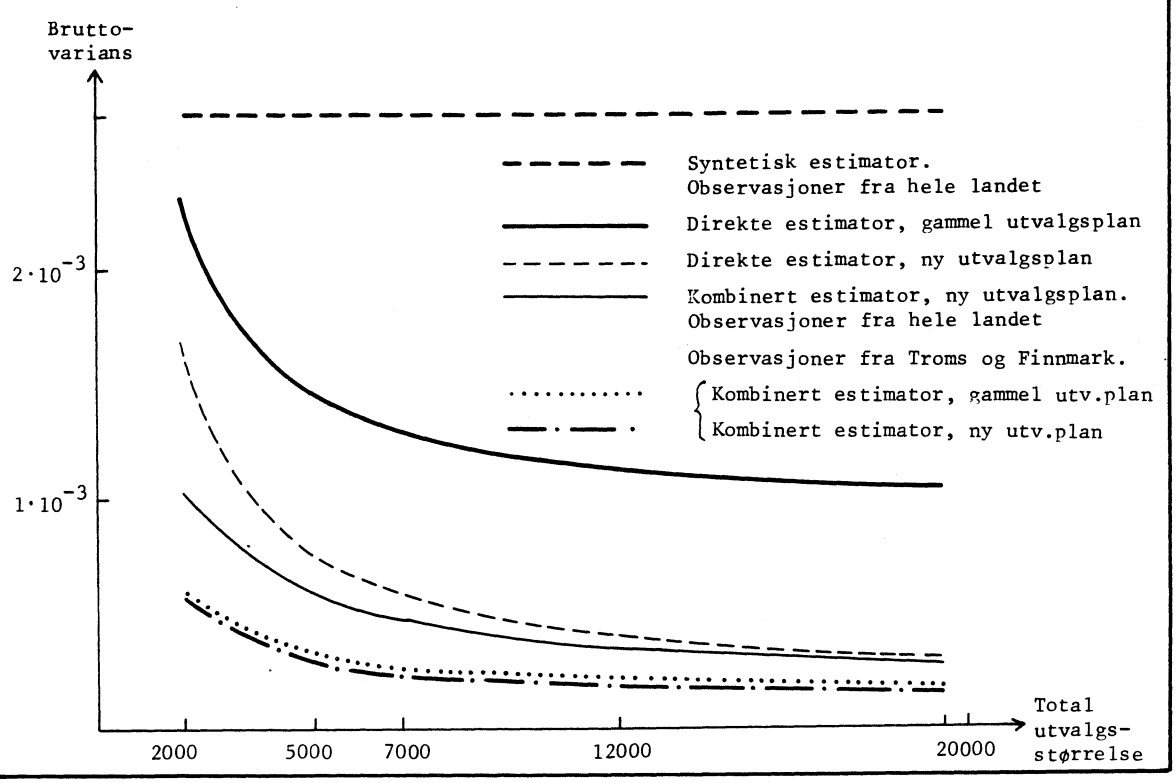
5.2.4. Bruttovarianser til forskjellige estimatorer for andelen sysselsatte innen fiske- og hvalfangst i Troms og Finnmark

På fig. 5.7. og 5.8. er bruttovariansen som funksjon av total utvalgsstørrelse vist for forskjellige estimatorer og utvalgsplaner. Variabelen "andelen sysselsatte innen fiske og hvalfangst" skiller seg ut fra de andre variablene vi har studert, ved at de syntetiske estimatorene med observasjoner fra hele landet er svært dårlige. Dette skyldes at andelen fiskere i Troms og Finnmark avviker mye fra andelen fiskere i hele landet. Til tross for dette kan en i figur 5.7. se at ved små utvalgsstørrelser kan en faktisk få redusert bruttovariansen en del ved å bruke den kombinerte estimatoren med observasjoner fra hele landet i stedet for å bruke den direkte estimatoren.

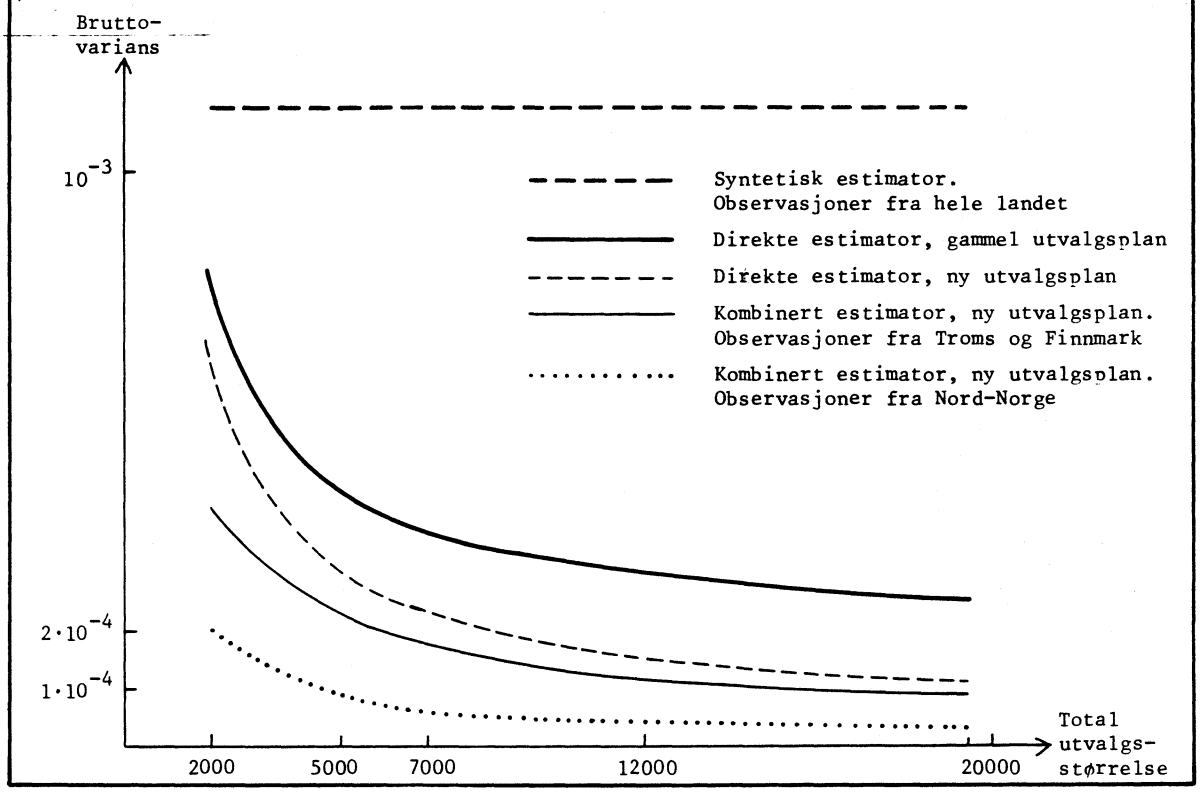
Figurene viser at når det gjelder de direkte estimatorene, betyr det her en del om en bruker ny eller gammel utvalgsplan. Dette er også tilfellet for de kombinerte estimatorene med observasjoner fra hele landet. Årsaken til dette er at en ved bruk av disse estimatorene her vil legge stor vekt på de direkte estimatorene og liten vekt på de dårlige syntetiske estimatorene.

Den beste estimator for andelen fiskere i Troms er den kombinerte estimatoren med observasjoner fra Nord-Norge. Den beste estimator for andelen fiskere i Finnmark er for små utvalgsstørrelser (mindre enn 5-6000) den kombinerte estimatoren med observasjoner fra Nord-Norge, og for store utvalgsstørrelser den kombinerte estimatoren med observasjoner fra fylkesparet Troms/Finnmark. Ved bruk av disse estimatorene betye det svært lite om en bruker ny eller gammel utvalgsplan.

Figur 5.7. Bruttovariansen til estimatorer for andelen sysselsatt med fiske og hvalfangst i Finnmark som funksjon av den totale utvalgsstørrelse



Figur 5.8. Bruttovariansen til forskjellige estimatorer for andelen sysselsatt med fiske og hvalfangst i Troms som funksjon av den totale utvalgsstørrelse



5.2.5. Noen foreløpige konklusjoner på grunnlag av sammenligningene gjort ved hjelp av data fra Folke- og boligtellingen 1970

For variablene som er studert ovenfor, er den kombinerte estimatoren totalt sett helt klart bedre enn den direkte og den syntetiske estimatoren når det gjelder å estimere tall for Troms og Finnmark. Dette kombinert med andre resultater vi har fått, men ikke publisert her, samt resultater som er publisert i Levy (1979) og i Schaible, Brock and Schnack (1977) tyder på at det er denne typen estimatorer en bør satse på i framtiden. Et annet viktig resultat er at det ser ut som om det er av liten betydning om en bruker ny eller gammel utvalgsplan når en bruker en kombinert estimator.

For de fleste variable vi har sett på er det den kombinerte estimatoren som er en lineærkombinasjon av observasjoner fra Troms eller Finnmark og observasjoner fra hele landet, som er best. Når det gjelder estimering av andelen fiskere i Troms og Finnmark er det den kombinerte estimatoren som er en lineærkombinasjon av observasjoner fra det enkelte fylket og observasjoner fra Nord- Norge som er best. Ut fra dette kan en konkludere med at det som regel lønner seg å bruke observasjoner fra hele landet når en skal bruke den kombinerte estimatoren for å estimere fylkestall. Hvis en imidlertid vet at en landsdel skiller seg vesentlig ut fra resten av landet m.h.t. studievariabelen, bør en bare benytte seg av observasjoner fra landsdelen når en skal estimere fylkestall i denne landsdelen.

I det neste avsnittet skal vi estimere andelen personer med kroniske lidelser i Troms og Finnmark ved hjelp av forskjellige estimeringsmetoder, og demonstrere noen av de problemer som er knyttet til bruk av de syntetiske og kombinerte estimatorene. Data er hentet fra Helseundersøkelsen 1975.

5.3. Et eksempel på bruk av den kombinerte estimatoren på data fra Helseundersøkelsen 1975

I de sammenligninger som er gjort ovenfor har det vært mulig å konstruere "optimale" kombinerte estimatorer, og variansene og skjevhetene ble regnet ut eksakt. Begge disse fordeler forsvinner når metodene skal brukes på utvalgsdata. Her må konstanten i den kombinerte estimatoren estimeres, og det er i alminnelighet ikke mulig å estimere skjevheten og dermed bruttovariansen til estimatoren. På bakgrunn av resultatene ovenfor er det likevel hensiktsmessig å bruke den kombinerte estimatoren når en ønsker å estimere fylkestall på grunnlag av data fra Helseundersøkelsen 1975.

I tabell 4 er gitt estimater for andelen personer med kroniske lidelser innen alle fylker. For denne variabelen viste det seg at alder er meget viktig, og det er derfor i tabell 4 brukt en syntetisk estimator, som bruker aldersfordelingene i de forskjellige fylkene slik som nevnt i avsnitt 4.2. pkt. (ii). Det som karakteriserer resultatene i tabell 4 er at variasjonen i fylkestallene estimert ved hjelp av den kombinerte estimatoren, er vesentlig mindre enn variasjonen en får ved å bruke den direkte estimatoren. Dette skyldes antakeligvis to forhold.

(i) I den kombinerte estimatoren er resultatet "trukket inn mot" landsgjennomsnittet.

(ii) Den tilfeldige variasjon til den kombinerte estimatoren er mindre enn den tilfeldige variasjon til den direkte estimatoren.

Måten konstanten i den kombinerte estimatoren er estimert på, samt noen problemer knyttet til denne estimeringen er beskrevet i vedlegg 3.

Tabell 4. Estimater for andelen personer med kroniske lidelser

Fylke	Direkte estimator	Syntetisk[1] estimator I	Kombinert[2] estimator I	Aldersjustering: Syntetisk[3] estimator II	Aldersjustering: Kombinert[4] estimator II
Østfold	.4229	.4133	.4139	.4170	.4173
Akershus	.3935	.4133	.4092	.3886	.3891
Hedmark	.4062	.4133	.4130	.4282	.4267
Oppland	.4389	.4133	.4164	.4233	.4245
Buskerud	.3513	.4133	.3918	.4189	.3930
Vestfold	.4079	.4133	.4130	.4139	.4136
Telemark	.4115	.4133	.4132	.4232	.4227
Aust-Agder	.3421	.4133	.4120	.4154	.4140
Vest-Agder	.4322	.4133	.4141	.4011	.4030
Rogaland	.3991	.4133	.4115	.3926	.3932
Hordaland	.4024	.4133	.4119	.4022	.4022
Sogn og Fjordane ..	.3969	.4133	.4128	.4144	.4139
Møre og Romsdal ...	.3793	.4133	.4067	.4035	.4003
Sør-Trøndelag	.4685	.4133	.4357	.4080	.4350
Nord-Trøndelag	.4318	.4133	.4139	.4065	.4075
Nordland	.4149	.4133	.4134	.4037	.4044
Troms	.3639	.4133	.4059	.3884	.3868
Finnmark	.5419	.4133	.4309	.3723	.4079
Oslo	.4436	.4133	.4270	.4464	.4461

1) Syntetisk estimator I = Landsgjennomsnitt.

2) Kombinert estimator I = En lineærkombinasjon av den direkte estimatoren og syntetisk estimator I.

3) Syntetisk estimator II = Landsgjennomsnittet justert for aldersfordeling i fylket.

4) Kombinert estimator II = En lineærkombinasjon av den direkte estimatoren og syntetisk estimator II.

5.4. Publisering av sysselsettingstall for Troms og Finnmark

I dette avsnittet skal vi se på hva som kan publiseres av tall for Troms og Finnmark etter publiseringskriteriet beskrevet i avsnitt 2.2. Som i avsnitt 5.2. tenker vi oss at målet er å estimere andelen sysselsatte i forskjellige næringer. Igjen anvendes data fra Folke- og boligtellingen 1970 for å foreta skjevhets- og variansberegninger til estimeringsmetodene beskrevet i kapitlene 3 og 4.

Tabell 5 og 6 viser hva som kan publiseres av sysselsettingstall for henholdsvis Finnmark og Troms ved forskjellige estimeringsmetoder og utvalgsstørrelser.

Tabell 5. Estimater for sysselsettingstall for Finnmark som kan publiseres. ("ja" betyr at tall kan publiseres)

Estimator	Total utvalgsstørrelse	Sysselsetting							
		To talt	Jord- og skogbruk	Fiske og hvalfangst	Industri m.v.	Bygg og anlegg	Varehandel	Samferdsel	Tjenesteytende næringer
Direkte estimator, ny utvalgsplan	2 000	ja							
	5 000	ja							
	20 000	ja			ja				ja
Syntetisk estimator, observasjoner fra Troms og Finnmark	2 000	ja							
	5 000	ja							ja
	20 000	ja						ja	ja
Syntetisk estimator, observasjoner fra hele landet	2 000	ja			ja	ja		ja	ja
	5 000	ja			ja	ja		ja	ja
	20 000	ja			ja	ja		ja	ja
Kombinert estimator, observasjoner fra Nord-Norge	2 000	ja							ja
	5 000	ja				ja	ja		ja
	20 000	ja			ja	ja	ja	ja	ja
Kombinert estimator, observasjoner fra hele landet	2 000	ja			ja	ja		ja	ja
	5 000	ja			ja	ja		ja	ja
	20 000	ja			ja	ja		ja	ja

Tabell 6. Estimater for sysselsettingstall for Troms som kan publiseres. ("ja" betyr at tall kan publiseres)

Estimator	Total utvalgsstørrelse	Sysselsetting							
		To talt	Jord- og skogbruk	Fiske og hvalfangst	Industri m.v.	Bygg og anlegg	Varehandel	Samferdsel	Tjenesteytende næringer
Direkte estimator, ny utvalgsplan	2 000	ja							
	5 000	ja							
	20 ooo	ja	ja		ja	ja	ja	ja	ja
Syntetiske estimator, observasjoner fra Troms og Finnmark	2 000	ja							
	5 000	ja							ja
	20 000	ja				ja	ja	ja	ja
Syntetisk estimator, observasjoner fra hele landet	2 000	ja	ja			ja	ja	ja	ja
	5 000	ja	ja			ja	ja	ja	ja
	20 000	ja	ja			ja	ja	ja	ja
Kombinert estimator, observasjoner fra Nord- Norge	2 000	ja							
	5 000	ja	ja					ja	ja
	20 000	ja	ja	ja	ja	ja	ja	ja	ja
Kombinert estimator, observasjoner fra hele landet	2 000	ja	ja			ja	ja	ja	ja
	5 000	ja	ja			ja	ja	ja	ja
	20 000	ja	ja		ja	ja	ja	ja	ja

Grunnen til at en i tabellene ikke skiller mellom ny og gammel utvalgsplan for de syntetiske og kombinerte estimatorene, er at det her ikke blir noen forskjell i hva som kan publiseres av tall om en bruker ny eller gammel utvalgsplan.

Tabellene viser at ved utvalgsstørrelser på 2-5000 personer fra hele landet er det bare tall for andelen "sysselsatte totalt" som kan publiseres ved direkte estimering i begge fylkene. Ved bruk av en syntetisk eller kombinert estimator der en anvender observasjoner fra hele landet, ser en at en kan publisere tall for 6 av 8 kjennetegn i Troms og 5 av 8 kjennetegn i Finnmark ved de samme utvalgsstørrelser.

Når en kommer opp i utvalgsstørrelser på 15-20 000 personer ser det ut til at en kan publisere flest tall for kjennetegn i de to fylkene når en bruker en kombinert estimator med observasjoner fra Nord-Norge.

6. FYLKE SOM FORKLARINGSVARIABEL

6.1. Innledning

I avsnittene foran har vi forsøkt å sammenligne forskjellige metoder som kan brukes når en ønsker å estimere fylkestall så godt som mulig. Et annet problem som ofte dukker opp er: Har det noen betydning å bo i et bestemt fylke? De to problemene henger nøye sammen, men det er også viktige forskjeller.

Når en er interessert i hvor hyppig et fenomen opptrer i et fylke, bryr en seg ofte mindre om hvorfor dette er tilfelle. Hvis en f.eks. skal allokere midler proporsjonalt med forekomsten av et fenomen, trenger en bare et kvantitativt uttrykk for utbredelsen av fenomenet i hvert fylke. I andre tilfeller er en mer opptatt av å finne ut om variabelen bosted har en egen effekt på forekomsten av et fenomen, etter at effektene av andre variable er tatt bort.

6.2. To eksempler på å bruke fylke som forklaringsvariabel

Ved Sosialforskningsinstituttet i Stockholm har en på en enkel måte nærmet seg spørsmålet om variabelen fylke (len) har en egen effekt etter at effektene av noen andre variable er tatt bort. Vi skal gi en kort beskrivelse av metoden, og kommentere den i lys av de resultater vi fant ovenfor.

For Norrbotten len har en presentert gjennomsnitt for en lang rekke levekårsvariable. Disse gjennomsnittene er basert på et utvalg bestående av 202 personer i alderen 15-75 år. Som estimator har en brukt det vi tidligere har kalt direkte estimering. Utvalget er trukket i ett trinn, slik at en unngår den skjevhet vi har i et fylkesutvalg fordi vi trekker utvalget i to trinn. For de fleste variables vedkommende mener vi likevel at denne skjevheten er meget liten i forhold til den usikkerhet som skyldes utvalgets størrelse. Vi skal derfor se helt bort fra denne skjevheten.

På bakgrunn av de resultater vi kom fram til i kapitlene foran er det naturlig å spørre om en ikke med fordel kunne ha brukt en annen estimator, f.eks. en kombinert estimator, som for de fleste variables vedkommende ville ha vesentlig mindre varians enn den direkte estimator. Svaret på dette spørsmålet er at dersom en ønsker det "beste" tall for Norrbotten len, kunne en med fordel bruke en kombinert estimator, men dersom en ønsker å undersøke om variablen len har noen effekt på levekårsvariablene etter at effekten av andre variable er tatt bort, bør en ta utgangspunkt i den direkte estimatoren. I medborgerrapporten fra Institutet för social forskning i Stockholm, er det dette siste spørsmålet en er opptatt av. I tillegg til å gi den direkte estimator for lenet har en også publisert to standardiserte gjennomsnitt, som er tenkt å skulle estimere gjennomsnittet for Norrbotten len når effektene av variablene alder og yrke er tatt bort. En tester deretter om det er signifikant forskjell mellom det standardiserte gjennomsnittet og det vanlige gjennomsnittet. Bortsett fra visse svakheter ved metoden er denne framgangsmåten svært lik det en ville ha fått ved å lage en regresjonsanalyse og teste om variablen len har noen signifikant effekt, etter at effektene av variablene alder og yrke er tatt bort. I neste avsnitt skal vi foreta en lignende test på data fra Helseundersøkelsen 1975. Vi skal bruke regresjonsanalyse i stedet for standardisering.

6.2.1. Regresjonsanalyse med fylker som dummy-variable

Som eksempler på hvordan en kan undersøke om fylke har noen effekt på en variabel etter at effektene av andre variable er "tatt bort" har vi utført to regresjonsanalyser.

I den ene har vi forsøkt å undersøke om bosted (fylke) har noen effekt på husholdningsinntekt. Dette har vi gjort ved å utføre regresjonsanalyse i flere trinn med husholdningsinntekt som avhengig variabel. I første trinn hadde vi "fylke" som uavhengig variabel (eller rettere sagt én variabel for hvert fylke). I andre trinn hadde vi alder som uavhengig variabel, mens i tredje trinn hadde vi både alder og "fylke" med som uavhengige variable.

Resultatene fra første trinn av analysen tyder på at "fylke" har en signifikant effekt på husholdningsinntekten når "fylke" er eneste forklaringsvariabel. Fra andre og tredje trinn fant vi at tilpassingen av predikerte inntektsbeløp til observerte inntektsbeløp var signifikant bedre når vi hadde med både alder og "fylke" som forklaringsvariable, enn når vi bare hadde med alder som forklaringsvariabel. Dette tyder på at "fylke" også har en signigikant effekt på husholdningsinntekt etter at effekten av variasjonen i aldersfordelingen fra fylke til fylke er "tatt bort".

I den andre regresjonsanalysen hadde vi "antall kroniske lidelser" som avhengig variabel, mens alder og fylke igjen var uavhengige variable. Etter å ha benyttet oss av samme framgangsmåte som beskrevet ovenfor, fant vi at "fylke" igjen hadde en signifikant effekt på den avhengige variablen, både når vi "tok bort" effekten av alder og når vi hadde "fylke" alene som forklaringsvariabel.

Resultatene fra begge regresjonsanalysene bør tas med "en klype salt" av flere grunner. Hensiktene med å ta med analysene er:

(i) Å antyde hvordan en kan gå fram når en ønsker å undersøke om "fylke" har en effekt.

(ii) Presentere et alternativ til bruken av standardisering.

(iii) Gjøre oppmerksom på at en med Byråets någjeldende utvalgsplan kan undersøke effekten av "fylke".

7. KONKLUSJONER

Den viktigste konklusjon en kan trekke av resultatene ovenfor er at en med utvalgsstørrelser under 10 000 bare oppnår små fordeler ved å legge om utvalgsplanen med sikte på å kunne gi bedre fylkestall. Det finnes sikkert variable for hvilke denne konklusjon ikke er riktig, men for langt de fleste gjelder den.

Hvis en ønsker å lage spesielle undersøkelser blant geografisk konsentrerte befolkningsgrupper, som f.eks. fiskere, vil det ofte være nødvendig å supplere den någjeldende utvalgsplan, og det ser da ut som om suppleringsmetoden beskrevet i avsnitt 4.4. er meget velegnet. Da metoden så vidt vi vet ikke er brukt tidligere, er det viktig å få dokumentert anvendelser av den.

Dersom en med et vanlig utvalg skal estimere fylkestall ser det ut som om den kombinerte estimatoren er den beste. Også denne metoden er relativt ny, og ikke alle problemer knyttet til bruken av den er endelig løst. Dersom det blir nødvendig med omfattende bruk av kombinerte estimatorer, må metoden studeres mer grundig enn det har vært mulig å gjøre her.

Dersom en ønsker å bruke fylke som forklaringsvariabel i en lineær regresjonsanalyse, kan dette etter vår mening best gjøres ved å innføre det nødvendige antall dummy-variable, og se bort fra utvalgsplanen.

Denne siste konklusjonen forutsetter at utvalgene er selvveiende. Dersom en ikke har selvveiende utvalg, finnes det foreløpig ingen veletablerte regler for hvordan en regresjonsanalyse bør utføres. Holt et.al. (1980).

For tiden har en ikke helt oversikt over behovet for fylkestall. Når dette behovet blir bedre kartlagt, vil det i Byrået bli drøftet i hvilken utstrekning en skal publisere tall for fylker.

8. REFERANSER

Des Raj: "Sampling Theory".

Gonzalez, M. E. and Waksberg, J. (1973): "Estimation of the error of synthetic estimates". International Association of Survey Statisticians, Vienna, Austria 1973.

Holt D., Smith, T. M. E. and, P. D. Winther (1980): "Regression analysis of data from Complex surveys". J. R. Statist. Soc. A (1980). 143, Part 4, pp 474-487.

Keyfitz, N. (1951): "Sampling with probabilities proportional to size: Adjustment for changes in the probabilities". Am.stat. Ass. Journal, march -51.

Laake, P. (1977): "A prediction approach to sub-domain estimation in infinite populations". Statistisk Sentralbyrå.

Laake, P. (1977): "An evaluation of synthetic estimates of employment". Scandinavian Journal of Statistics 5, 57-60.

Laake, P. and Longva, H. K. (1976): "Estimering av sysselsetting i geografiske regioner, om estimatorenes skjevhet, varians og bruttovarians". Statistisk Sentralbyrå. Artikler 88.

Levy and French (1977): "Synthetic Estimation of State Health Characteristics Based on the Health Interview Survey". Vital and Health Statistics: Series 2, NO. 75, DHEW Publication (PHS) 78-1349. Washington: U.S. Government Printing Office, 1977.

Levy (1979): "Monograph based on papers presented at a workshop conducted by Response Analysis", Princeton, New Jersey, under NIDA Contract No. 271-77-3425.

Purcell, N. J. and Kish, L. (1980): "Postcensal Estimates for Local Areas (or Domains)". Int. Stat.Review, 48 (1980) 3-18.

Schaible, W. L., Brock, D. B. and Schnack, G. A.: "An empirical comparison of the simple inflation, synthetic and composite estimators for small area statistics". Proceedings of the American Statistical Association, Social Statistics Section: 1017-1021, 1977.

TREKKING AV ET REPRESENTATIVT, SELVVEIENDE UTVALG FOR FINNMARK MED MAKSIMAL UTNYTTELSE AV DE TIDLIGERE UTTRUKNE UTVALGSOMRÅDENE

Byråets generelle utvalgsplan er konstruert slik at større utvalg trekkes i to trinn. I Norge er det ca. 440 kommuner som er delt inn i ca. 100 strata. I første trinn trekkes fra hvert stratum et primært utvalgsområde som består av en eller flere kommuner. I annet trinn blir så det endelige utvalget trukket fra de uttrukne utvalgsområdene.

Trekking på første trinn foretas ikke for hver ny utvalgsundersøkelse. En trekker ut utvalgsområder for noen år om gangen, og bygger opp en fast intervjuerstab i disse områdene.

Byråets utvalgsplan er ikke konstruert med tanke på å kunne produsere tall for hvert fylke. I noen tilfeller krysser stratagrensene fylkesgrensene slik at vi ved trekking av utvalg ikke får representative delutvalg for fylkene.

Vi skal nå presentere en metode som kan brukes for å lage et representativt, selvveiende utvalg for Finnmark med maksimal utnyttelse av de tidligere uttrukne utvalgsområdene. Metoden forutsetter at vi beholder den samme stratainndelingen av utvalgsområdene som tidligere.

Kommunene i Finnmark er inndelt i 4 forskjellige strata. I to av disse strataene er det også med kommuner fra Troms. De tidligere uttrukne utvalgsområdene ble trukket ut med sannsynligheter proporsjonale med folketallene. I vår nye utvalgsplan ønsker vi av to grunner å forandre på trekkesannsynlighetene på 1. trinn. Den ene er at folketallene i kommunene har endret seg noe de siste årene. Den andre er at vi ønsker at kommuner fra Troms skal ha en sannsynlighet på 0 for å komme med i vårt utvalg for Finnmark. Samtidig som vi ønsker å endre trekkesannsynlighetene innenfor de forskjellige strataene, ønsker vi at flest mulig av de tidligere uttrukne utvalgsområdene skal bli med i vårt nye utvalg. Med en framgangsmåte som er kalt Keyfitz-metoden kan vi få begge disse ønskene oppfylt. Keyfitz (1951).

Beskrivelse av Keyfitz-metoden.

Anta at vi har et stratum bestående av kommunene A, B, C og D, og at vi tidligere har trukket ut en kommune med trekkesannsynligheter lik α, β, γ og δ for de 4 kommunene. Anta videre at vi ønsker at disse sannsynlighetene skal bli endret til henholdsvis a, b, c og d ($\alpha+\beta+\gamma+\delta=a+b+c+d=1$), og at $a>\alpha$, $b>\beta$, $c<\gamma$ og $d<\delta$. Metoden som vi skal presentere kan brukes for et generelt antall trekkeenheter.

Det første vi gjør er å dele trekkeenhetene inn i to grupper ettersom de nye trekkesannsynlighetene er større eller mindre enn de opprinnelige. Vi kan kalle de to gruppene gruppe 1 og gruppe 2 henholdsvis. I dette tilfellet utgjør A og B gruppe 1, og C og D gruppe 2.

For å få oppfylt våre tidligere nevnte ønsker skal en etter Keyfitz-metoden nå gå fram på følgende måte:

Hvis kommunen som tidligere er trukket ut, tilhører gruppe 1, beholder vi denne kommunen i vårt nye utvalg.

Hvis kommunen som tidligere er trukket ut tilhører gruppe 2 (C eller D), må vi foreta et valg mellom de to begivenhetene "behold trekkeenheten" og "bytt trekkeenheten". Den siste begivenheten gir vi en sannsynlighet på $\frac{\gamma-c}{\gamma}$ eller $\frac{\delta-d}{\delta}$ ettersom det er C eller D som er trukket ut tidligere. Hvis utfallet av valget blir "behold trekkeenheten", foretar vi ingen forandring. Hvis utfallet av valget blir "bytt trekkeenheten", trekker vi en enhet fra gruppe 1 med sannsynlighet proporsjonal til differansen mellom ny og gammel trekkesannsynlighet. Den betingede trekkesannsynligheten for at A skal bli trukket, gitt at vi skal bytte trekkeenhet, lar vi f.eks. være $\frac{a-\alpha}{a-\alpha+b-\beta}$.

Skisse av bevis for egenskapene til Keyfitz-metoden.

Det er lett å verifisere at hver enhet har den riktige trekkesannsynligheten. Vi nøyer oss med å vise at A har sannsynligheten a for å bli valgt. A kan komme med i utvalget på 3 måter. Den ene er at A er blitt trukket ut tidligere. Sannsynligheten for dette er α. De 2 andre måtene er at henholdsvis C og D er blitt trukket ut tidligere, og at vi bytter kommune, og at vi til slutt trekker A. Disse tre hendelsene er disjunkte, og sannsynligheten for at A skal bli valgt er:

$$\alpha+\gamma\,\frac{\gamma-c}{\gamma}\cdot\frac{a-\alpha}{a-\alpha+b-\beta}+\delta\frac{\delta-d}{\delta}\cdot\frac{a-\alpha}{a-\alpha+b-\beta}$$

$$=\alpha+\frac{(a-\alpha)\,(\gamma-c+\delta-d)}{a-\alpha+b-\beta}=a$$

siden $\gamma-c+\delta-d = a-\alpha+b-\beta$.

Det er lett å generalisere dette beviset til å gjelde en vilkårlig trekkeenhet i en populasjon med vilkårlig mange trekkeenheter.

Vi skal nå vise at sannsynligheten for å bytte ut den tidligere uttrukne trekkeenheten blir minimert. En endring kan intreffe ved at C eller D er blitt trukket ut tidligere og at begivenheten "bytt trekkeenhet" inntreffer.
Sannsynligheten for dette er $\gamma\frac{\gamma-c}{\gamma}+\delta\frac{\delta-d}{\delta}=\gamma-c+\delta-d$.

Vi ser at sannsynligheten for endring er lik summen av differansene mellom de gamle og de nye trekkesannsynlighetene for alle enhetene i gruppe 2.

Anta at C ble trukket ut i første trekning, og at det er mulig å anvende en sannsynlighet p for "bytt trekkeenhet" som er mindre enn $\frac{\gamma-c}{\gamma}$. Sannsynligheten for at C først skal bli trukket ut, og at vi siden skal beholde C blir da $\gamma(1-p)$. Denne sannsynligheten kan ikke være større enn c. Men dette strider mot forutsetning om at $p<\frac{\gamma-c}{\gamma}$ som gir ulikheten $\gamma(1-p) > c$.

På samme måte kan en bevise at $\frac{\delta-d}{\delta}$ er den minste sannsynligheten som kan anvendes for endring hvis D er den tidligere uttrukne enheten. Hvis $\frac{\gamma-c}{\gamma}$ og $\frac{\delta-d}{\delta}$ er de minste sannsynligheter som kan bli brukt for utbytting av C og D henholdsvis, kan det ikke anvendes noen mindre sannsynlighet for endring enn $\gamma-c+\delta-d$.

Anvendelse av Keyfitz-metoden ved trekking av kommuner fra Finnmark.

Som nevnt tidligere er kommunene i Finnmark med i 4 forskjellige strata. Under er disse 4 strataene listet opp med folketallene 1. januar 1979, gamle og nye (ønskede) trekkesannsynligheter.

Navn på utvalgsområdene	Folketallene 1. jan. 1979	Opprinnelig trekkesanns. p'	Ny (ønsket) trekkesanns. p"
	Stratum 1		
I alt	16 957	1.0000	1.0000
Bjarkøy & Ibestad (Troms)	uinteressant	0.0962	0
Tranøy & Torsken & Berg (Troms)	"	0.1359	0
Karlsøy & Lyngen (Troms)	"	0.1809	0
Skjervøy & Kvænangen (Troms)	"	0.1412	0
Loppa & Hasvik	3 765	0.0987	0.2220
Sørøysund & Måsøy	4 857	0.1271	0.2864
Lebesby & Gamvik	3 705	0.0997	0.2185
Berlevåg & Båtsfjord	4 630	0.1203	0.2730

Navn på utvalgsområdene	Folketallene 1. jan. 1979	Opprinnelig trekkesanns. p'	Ny (ønsket) trekkesanns. p"
	Stratum 2		
I alt	22 124	1.0000	1.0000
Hammerfest	7 459	0.3352	0.3371
Vardø	3 769	0.1769	0.1704
Vadsø	6 054	0.2603	0.2736
Nordkapp	4 842	0.2277	0.2189
	Stratum 3		
I alt	23 546	1.0000	1.0000
Alta	12 998	0.5252	0.5520
Sør-Varanger	10 548	0.4748	0.4480
	Stratum 4		
I alt	16 104	1.0000	1.0000
Balsfjord (Troms)	uinteressant	0.2146	0
Storfjord & Kåfjord (Troms)	"	0.1539	0
Nordreisa (Troms)	"	0.1347	0
Kautokeino & Karasjok	5 478	0.1679	0.3402
Kvalsund & Porsanger	6 238	0.1944	0.3874
Tana & Nesseby	4 388	0.1344	0.2725
Totalt i Finnmark	78 731		

Når vi heretter omtaler et bestemt utvalgsområde, vil vi la p' betegne opprinnelig og p" ny trekkesannsynlighet for dette utvalgsområdet.

Vi skal nå anvende Keyfitz-metoden på hvert av de 4 strataene.

i) Stratum 1

Først deler vi utvalgsområdene i stratum 1 inn i to grupper ettersom de nye trekkesannsynlighetene er større eller mindre enn de opprinnelige. Gruppen av utvalgsområder som har fått økt sine trekkesannsynligheter, kaller vi gruppe 1, og gruppen av utvalgsområder som har fått redusert sine trekkesannsynligheter, kaller vi gruppe 2. Vi ser at utvalgsområdene i Finnmark utgjør gruppe 1, og utvalgsområdene i Troms utgjør gruppe 2.

Opprinnelig ble Bjarkøy & Ibestad i gruppe 2 valgt ut. I følge Keyfitz's prosedyre skal vi da gi Bjarkøy & Ibestad en sannsynlighet på $\frac{p'-p"}{p"} = 1$ for å bli byttet ut med et utvalgsområde i gruppe 1. Hvilket av utvalgsområdene i gruppe 1 som skal erstatte Bjarkøy & Ibestad, skal avgjøres ved en trekning, der hvert av utvalgsområdene i gruppe 1 gis en trekkesannsynlighet som er proporsjonal med differansen mellom ny og "gammel" trekkesannsynlighet (p"-p').

Vi har beregnet disse trekkesannsynlighetene. De er følgende:

Utvalgsområde	Trekkesannsynlighet
I alt	1.0000
Loppa & Hasvik	0.2225
Sørøysund & Måsøy	0.2874
Lebesby & Gamvik	0.2144
Berlevåg & Båtsfjord	0.2755

ii) Stratum 2

I stratum 2 ble Vardø trukket ut opprinnelig. Vi ser at Vardø har fått redusert trekkesannsynligheten. I følge Keyfitz-metoden må vi da gi Vardø en sannsynlighet på $\frac{p'-p''}{p'} = 0.0367$ for å bli erstattet av et av de utvalgsområdene der trekkesannsynligheten har økt, dvs. Hammerfest eller Vadsø.

Hvis utfallet blir at Vardø skal byttes ut, skal vi foreta et valg mellom Hammerfest og Vadsø med følgende trekkesannsynligheter:

Utvalgsområde	Trekkesannsynlighet
I alt	1.0000
Hammerfest	0.1250
Vadsø	0.8750

iii) Stratum 3

I stratum 3 ble Sør-Varanger trukket ut opprinnelig. Vi ser at Sør-Varanger har fått redusert trekkesannsynligheten ($p'' < p'$). I følge Keyfitz-metoden må vi da foreta et valg mellom å beholde Sør-Varanger eller å erstatte Sør-Varanger med Alta.

Begivenheten å beholde Sør-Varanger gis en sannsynlighet på $\frac{p''}{p'} = 0.9436$.
Begivenheten å erstatte Sør-Varanger med Alta gis en sannsynlighet på $\frac{p'-p''}{p'} = 0.0564$.

iv) Stratum 4

I stratum 4 ble Tana & Nesseby trukket ut opprinnelig. Vi ser at dette utvalgsområdet har fått økt sin trekkesannsynlighet ($p'' > p'$). I følge Keyfitz-metoden skal da Tana & Nesseby fortsette å representere stratum 4.

Trekking av et selvveiende utvalg i Finnmark

Trekking av et selvveiende utvalg i Finnmark skal foregå i to trinn. På første trinn trekker en et utvalgsområde fra hvert stratum etter Keyfitz-metoden, og på annet trinn trekker en så det endelige utvalget fra de uttrukne utvalgsområdene. Anta at vi ønsker at hver person i Finnmark skal ha en trekkesannsynlighet lik f, og at vi har trukket ut utvalgsområde nr. i til å representere stratum nr. j. La p''_{ij} betegne trekkesannsynligheten på første trinn (p"-verdien) for dette utvalgsområdet. Hvis vi lar trekkesannsynligheten på annet trinn være f_{ij}, må f_{ij} oppfylle følgende betingelse:

$$f = f_{ij} \cdot p''_{ij}.$$

Forventet prosentandel av et utvalg som må trekkes utenfor de tidligere uttrukne utvalgsområdene

Når en følger den ovenfor beskrevne fremgangsmåten for trekking av et utvalg i Finnmark, vil forventet prosentandel av utvalget som må trekkes utenfor de tidligere uttrukne utvalgsområdene, være 24.2.

UTLEDNING AV SKJEVHET OG VARIANS TIL FYLKESTALLENE

Innhold

1. Innledning, definisjoner

Vi skal i dette vedlegget gjøre rede for hvordan vi har beregnet skjevheter og varianser til estimatorene som er beskrevet i kapittel 4, for fylkestall fra Finnmark.

I vedlegget er det tale om ny og gammel utvalgsplan. Med gammel utvalgsplan menes Byråets någjeldende utvalgsplan, mens med ny utvalgsplan menes den någjeldende utvalgsplanen supplert med tilleggskommuner som beskrevet i vedlegg 1.

Utvalgsområdene i Finnmark tilhører 4 strata, her nummerert fra 1 til 4. Ved gammel utvalgsplan består stratum 1 og 4 av utvalgsområder fra både Troms og Finnmark, mens ved ny utvalgsplan består stratum 1 og 4 av utvalgsområder bare fra Finnmark.

Vi skal nå definere alle symboler som er brukt i vedlegget.

Definisjoner

N_{ij} = Antall personer i populasjonen i utvalgsområde nr. i i stratum j.

$N_{.j}$ = Antall personer i populasjonen i stratum j.

$N_{.jF}$ = Antall personer i stratum j i Finnmark.

$n_{.j}$ = Antall personer i utvalget fra stratum j.

n_F = Antall personer i utvalget fra Finnmark.

n_R = Antall personer i utvalget fra regionen R.

P_{ij} = Andel personer i populasjonen i utvalgsområde nr. i i stratum j med et bestemt kjennemerke.

$P_{.j}$ = Andel personer i populasjonen i stratum j med det bestemte kjennemerket. $P_{.j} = \frac{1}{N_{.j}} \sum N_{ij} P_{ij}$

P_F = Andel personer i populasjonen i Finnmark med det bestemte kjennemerket.

P_R = Andel personer i populasjonen i regionen R, som kan være Troms og Finnmark, Nord-Norge eller hele landet, med det bestemte kjennemerket.

$\bar{Y}_{.j}$ = Andel personer i utvalget fra stratum j med det bestemte kjennemerket.

$\bar{Y}_F$ = Andel personer i utvalget fra Finnmark med det bestemte kjennemerket.

$\bar{Y}_R$ = Andel personer i utvalget fra regionen R med det bestemte kjennemerket.

Vi forutsetter her at det trekkes utvalg som er selvveiende, og at sannsynligheten for at utvalgsområde nr. i fra stratum j skal bli trukket i 1. trinn er proporsjonal med N_{ij}. Disse to forutsetningene medfører at n_j ikke varierer med utfallet av trekningen på 1. trinn, og at n_j er proporsjonal med $N_{.j}$.

Vi antar at det er P_F som vi er interessert i å estimere.

2. Den direkte estimator

2.1. Ny utvalgsplan

Den direkte estimatoren er identisk med $\bar{Y}_F$ som er hyppigheten av et bestemt kjennemerke blant observasjonene fra Finnmark.

Vi antar først at vi har trukket et utvalg ved ny utvalgsplan.

$\bar{Y}_F$ kan da skrives på formen:

$$\bar{Y}_F = \frac{1}{n_F} \sum_{j=1}^{4} n_{\cdot j} \bar{Y}_{\cdot j}$$

Forventningen til $\bar{Y}_F$ blir:

$$E\bar{Y}_F = \frac{1}{n_F} \sum_{j=1}^{4} n_{\cdot j} P_{\cdot j} = \frac{1}{N_F} \sum_{j=1}^{4} N_{\cdot j} P_{\cdot j} = P_F,$$

dvs. $\bar{y}_F$ er forventningsrett.

Variansen til $\bar{Y}_F$ blir:

$$\text{var } \bar{Y}_F = \frac{1}{n_F^2} \sum_{j=1}^{4} n_{\cdot j}^2 \text{ var } \bar{Y}_{\cdot j},$$

der $\text{var } \bar{Y}_{\cdot j} = \sum_i \frac{N_{ij}}{N_{\cdot j}} (P_{ij} - P_{\cdot j})^2 + \sum_i \frac{N_{ij}}{N_{\cdot j}} \frac{N_{ij} - n_j}{N_{ij} - 1} \frac{P_{ij}(1-P_{ij})}{n_j}$ (Des Raj) (1)

2.2. Gammel utvalgsplan

Anta nå at vi har trukket et utvalg ved gammel utvalgsplan. Vi innfører de to hjelpevariablene I_1 og I_4, der

$$I_1 = \begin{cases} 1 \text{ hvis det blir trukket et utvalgsområde fra Finnmark i stratum 1 ved 1. trinns trekking.} \\ 0 \text{ ellers.} \end{cases}$$

$$I_4 = \begin{cases} 1 \text{ hvis det blir trukket et utvalgsområde fra Finnmark i stratum 4 ved 1. trinns trekking} \\ 0 \text{ ellers.} \end{cases}$$

$$P_r (I_1 = 1) = \frac{N_{\cdot 1F}}{N_{\cdot 1}} = \Pi_1$$

$$P_r (I_4 = 1) = \frac{N_{\cdot 4F}}{N_{\cdot 4}} = \Pi_4$$

$\bar{Y}_F$ kan nå skrives på formen:

$$\bar{Y}_F = I_1 I_4 \bar{Y}_F + I_1 (1-I_4) \bar{Y}_F + (1-I_1) I_4 \bar{Y}_F + (1-I_1)(1-I_4) \bar{Y}_F$$

Forventningen til $\bar{Y}_F$ blir:

$$E\bar{Y}_F = E\,E\,(\bar{Y}_F | I_1, I_4) = \underline{\Pi_1 \Pi_4 E\bar{X}_1 + \Pi_1 (1-\Pi_4) E\bar{X}_2 + (1-\Pi_1)\Pi_4 E\bar{X}_3 + (1-\Pi_1)(1-\Pi_4) E\bar{X}_4},$$

der $E\bar{X}_1 = E\,(\bar{Y}_F | I_1=1, I_4=1)$, $E\bar{X}_2 = E\,(\bar{Y}_F | I_1=1, I_4=0)$, $E\bar{X}_3 = E\,(Y_F | I_1=0, I_4=1)$ og

$E\bar{X}_4 = E\,(Y_F | I_1=0, I_4=0)$

Variansen til $\bar{Y}_F$ kan skrives på følgende form:

$$\text{Var } \bar{Y}_F = \text{Var E } (\bar{Y}_F|I_1,I_4) + \text{E Var } (\bar{Y}_F|I_1,I_4) \qquad (2)$$

Vi skal se på de to leddene i (2) hver for seg:

i) $$\text{E Var } (\bar{Y}_F|I_1,I_4) = \underline{\pi_1 \pi_4 \text{ Var } \bar{X}_1 + \pi_1 (1-\pi_4) \text{ Var } \bar{X}_2 + (1-\pi_1) \pi_4 \text{ Var } \bar{X}_3 + (1-\pi_1)(1-\pi_4) \text{ Var } \bar{X}_4},$$

der $\text{Var } \bar{X}_1 = \text{Var } (\bar{Y}_F|I_1=1,I_4=1)$, $\text{Var } \bar{X}_2 = \text{Var } (\bar{Y}_F|I_1=1,I_4=0)$ $\text{Var } \bar{X}_3 = \text{Var } (Y_F|I_1=0,I_4=1)$ og

$$\text{Var } \bar{X}_4 = \text{Var } (\bar{Y}_F|I_1=0,I_4=0)$$

ii) $$\text{Var E } (\bar{Y}_F|I_1,I_4) = \text{Var } (I_1I_4 (E\bar{X}_1-E\bar{X}_2-E\bar{X}_3+E\bar{X}_4) + I_1 (E\bar{X}_2-E\bar{X}_4) + I_4 (E\bar{X}_3-E\bar{X}_4) + E\bar{X}_4)$$

$$= \underline{(E\bar{X}_1-E\bar{X}_2-E\bar{X}_3+E\bar{X}_4)^2 \pi_1\pi_4 (1-\pi_1\pi_4) + (E\bar{X}_2-E\bar{X}_4)^2 \pi_1 (1-\pi_1)}$$

$$+ \underline{(E\bar{X}_3-E\bar{X}_4)^2 \pi_4 (1-\pi_4) + 2 (E\bar{X}_1-E\bar{X}_2-E\bar{X}_3+E\bar{X}_4)(E\bar{X}_2-E\bar{X}_4)}$$

$$\underline{\pi_1\pi_4 (1-\pi_1) + 2 (E\bar{X}_1-E\bar{X}_2-E\bar{X}_3+E\bar{X}_4)(E\bar{X}_3-E\bar{X}_4) \pi_1\pi_4 (1-\pi_4)}$$

3. Den syntetiske estimatoren

Den syntetiske estimatoren som vi har brukt, er rett og slett den observerte andelen, $\bar{Y}_R$, med et bestemt kjennetegn i regionen R, som vi vekselvis har latt være fylkesparet Troms/Finnmark, Nord-Norge og hele landet.

Forventningen til $\bar{Y}_R$ blir den samme ved ny og gammel utvalgsplan:

$$\text{E } \bar{Y}_R = P_R$$

Forventningsskjevheten til $\bar{Y}_R$ blir altså $P_R - P_F$.

Variansen til $\bar{Y}_R$ blir:

$$\text{Var } \bar{Y}_R = \frac{1}{n_R^2} \sum_{j \in R} n_j^2 \text{ Var } \bar{Y}_{\cdot j}, \qquad (3)$$

der $\text{Var } \bar{Y}_{\cdot j}$ er som i (1), og summeringen er over alle strataene i R.

(3) gjelder ved både ny og gammel utvalgsplan. Ved ny utvalgsplan får en imidlertid noen flere strata å summere over enn ved gammel utvalgsplan.

I de tilfellene der vi har latt R være Nord-Norge eller hele landet, har vi for variansen til $\bar{Y}_R$ brukt tilnærmingen:

$$\text{Var } \bar{Y}_R = \underline{\frac{P_R (1-P_R)}{n_R}}$$

4. Den kombinerte estimatoren

Den kombinerte estimatoren er som følger:

$$F_K = c\,\bar{Y}_R + (1-c)\,\bar{Y}_F$$

c er en konstant som velges slik at bruttovariansen til F_K blir minimert, dvs.

$$c = \frac{BV(\bar{Y}_F) - E(\bar{Y}_F - P_F)(\bar{Y}_R - P_R)}{BV(\bar{Y}_F) + BV(\bar{Y}_R) - 2E(\bar{Y}_F - P_F)(\bar{Y}_R - P_R)}$$

der BV er forkortelse for bruttovarians.

Forventningen til F_K blir:

$$EF_K = \underline{c\,P_R + (1-c)P_F}$$

Forventningsskjevheten til F_K blir da:

$$EF_K - P_F = \underline{c\,(P_R - P_F)}$$

Vi ser at skjevheten til F_K blir mindre enn skjevheten til $\bar{Y}_R$ når $c \in (0,1)$. Vanligvis ligger c i dette intervallet.

Variansen til F_K blir:

$$\text{Var } F_K = \underline{c^2 \text{ Var } \bar{Y}_R + (1-c)^2 \text{ Var } \bar{Y}_F + 2c(1-c) \text{ cov } (\bar{Y}_R, \bar{Y}_F)}$$

Vi skal nå vise hvordan vi har gått fram for å finne kovariansen mellom $\bar{Y}_R$ og $\bar{Y}_F$. $\bar{Y}_R$ kan skrives på formen $\frac{n_F}{n_R}\bar{Y}_F + \bar{z}$, der $\bar{z}$ er en stokastisk variabel slik at $\bar{z}$ og $\bar{Y}_F$ er uavhengige.

Ved ny utvalgsplan blir kovariansen mellom $\bar{Y}_R$ og $\bar{Y}_F$.

$$\text{cov } (\bar{Y}_R, \bar{Y}_F) = \frac{n_F}{n_R} \text{ var } \bar{Y}_F$$

Ved gammel utvalgsplan er det atskillig vanskeligere å finne kovariansen mellom $\bar{Y}_R$ og $\bar{Y}_F$. Vi trenger følgende hjelpesetning.

Hjelpesetning: La X, Y og I være stokastiske variable. Da gjelder følgende:

$$\underline{\text{cov } (X,Y) = E \text{ cov } (X,Y|I) + \text{cov } [E(X|I), E(Y|I)]}$$

Bevis for hjelpesetning:

cov (X,Y) kan skrives på følgende form:

$$\text{cov } (X,Y) = \tfrac{1}{2}\{\text{Var } (X+Y) - \text{Var } X - \text{Var } Y\}$$

$$= \tfrac{1}{2}\{\text{Var } E((X+Y)|I) + E \text{ Var } ((X+Y)|I)$$

$$- E \text{ Var } (X|I) - \text{Var } E(X|I) - E \text{ Var } (Y|I) - \text{Var } E(Y|I)\} \qquad (4)$$

Vi trenger følgende to biresultater:

i) $E((X+Y)|I) = \underline{E(X|I) + E(Y|I)}$

ii) $\text{Var } ((X+Y)|I) = \underline{\text{Var } (X|I) + \text{Var } (Y|I) + 2 \text{ cov } (X,Y|I)}$

i) følger av at $E((X+Y)|F) = \int\int (x+y) f(x,y|I)\,dxdy$

$$= \int_x xf(x|I)\,dx + \int_y yf(y|I)\,dy = \underline{E(X|I) + E(Y|I)}$$

Det er lett å vise ii) hvis en skriver $\text{Var } ((X+Y)|I)$ som $E[(X+Y)^2|I] - [E(X+Y)|I]^2$, løser opp parantesene og bruker (i).

Ved innsetting av i) og ii) i (4) finner en at:

$$\text{cov } (X,Y) = E \text{ cov } (X,Y|I) + \text{cov } E[(X|I), E(Y|I)] \text{ QED.}$$

På samme måte som for $\bar{Y}_F$ kan vi skrive $\bar{Y}_R$ på følgende form:

$$\bar{Y}_R = I_1 I_4 \bar{Y}_R + I_1 (1-I_4) \bar{Y}_R + (1-I_1) I_4 \bar{Y}_R + (1-I_1)(1-I_4) \bar{Y}_R,$$

der I_1 og I_4 er som definert tidligere. Ved bruk av hjelpesetningen finner vi at

$$\text{cov}(\bar{Y}_F, \bar{Y}_R) = E \text{ cov}(\bar{Y}_F, \bar{Y}_R | I_1, I_4) + \text{cov}[E(\bar{Y}_F|I_1, I_4), E(\bar{Y}_R|I_1, I_4)]$$

For å finne de to leddene i summen over går en fram på samme måte som i avsnitt 2.2 i vedlegget. Vi finner:

i) $E \text{ cov}(\bar{Y}_F, \bar{Y}_R | I_1, I_4)$

$$= \pi_1\pi_4 \frac{1^{\Sigma^4 n_j}}{n_R} \text{Var } \bar{X}_1 + \pi_1 (1-\pi_4) \frac{n_1+n_2+n_3}{n_R} \text{Var } \bar{X}_2$$

$$+ (1-\pi_1) \pi_4 \frac{n_2+n_3+n_4}{n_R} \text{Var } \bar{X}_3 + (1-\pi_1)(1-\pi_4) \frac{n_2+n_3}{n_R} \text{Var } \bar{X}_4$$

ii) $\text{cov}[E(\bar{Y}_F|I_1,I_4), E(\bar{Y}_R|I_1,I_4)]$

$$= (E\bar{X}_1 - E\bar{X}_2 - E\bar{X}_3 + E\bar{X}_4)(E\bar{z}_2 - E\bar{z}_4) \pi_1\pi_4 (1-\pi_1)$$

$$+ (E\bar{X}_1 - E\bar{X}_2 - E\bar{X}_3 + E\bar{X}_4)(E\bar{z}_3 - E\bar{z}_4) \pi_1\pi_4 (1-\pi_4)$$

$$+ (E\bar{X}_2 - E\bar{X}_4)(E\bar{z}_2 - E\bar{z}_4) \pi_1 (1-\pi_1) + (E\bar{X}_3 - E\bar{X}_4)(E\bar{z}_3 - E\bar{z}_4)$$

$$\pi_4(1-\pi_4),$$

der

$$E\bar{z}_1 = E(\bar{Y}_R | I_1 = 1, I_4 = 1), \quad E\bar{z}_2 = E(\bar{Y}_R | I_1 = 1, I_4 = 0),$$

$$E\bar{z}_3 = E(\bar{Y}_R | I_1 = 0, I_4 = 1) \text{ og } E\bar{z}_4 = E(\bar{Y}_R | I_1 = 0, I_4 = 0)$$

Ellers er alle symbolene i i) og ii) som definert tidligere.

ESTIMERING AV KONSTANTEN I DEN KOMBINERTE ESTIMATOREN

Den kombinerte estimatoren er beskrevet i kap. 4.2. (iii). I kap. 5.2. kunne vi velge konstanten til den kombinerte estimatoren ("C"-en i kap. 4.2. (iii)) eksakt slik at bruttovariansen til den kombinerte estimatoren ble minimert. Når en skal lage estimater på grunnlag av utvalgsdata, har en ikke denne muligheten siden konstanten til den kombinerte estimatoren må estimeres. Vi skal her vise hvordan vi estimerte denne konstanten i kap. 5.3.

La p betegne den andelen personer med et bestemt kjennemerke i et fylke som vi er interessert i å estimere, og la F_F og F_L betegne de observerte frekvenser i henholdsvis fylket og hele landet. Den kombinerte estimatoren blir da:

$$\hat{p} = cF_L + (1-c)\, F_F$$

La c* være den verdi av c som minimerer bruttovariansen til $\hat{p}$. Det er ikke så vanskelig å vise at c* er:

$$C^* = \frac{BV(F_F) - E\,(F_F-p)\,(F_L-p)}{BV(F_F) + BV\,(F_L) - 2\,E\,(F_F-p)\,(F_L-p)},$$

der BV er forkortelse for bruttovarians.

Det er vanskelig å lage noen god estimator for skjevheten til F_L, og følgelig er det også vanskelig å lage noen god estimator for c*.

Estimatoren for c* som vi brukte i kap. 5.3. var:

$$\hat{C}^* = \frac{\frac{1}{n_F} F_F\,(1-F_F)(1-\frac{n_F}{n_L})}{\frac{1}{n_F} F_F\,(1-F_F)\,(1-\frac{2n_F}{n_L}) + \frac{1}{n_L} F_L\,(1-F_L) + \frac{n_F}{n_L}(F_F-F_L)^2}$$

der n_F og n_L er antall observasjoner fra henholdsvis fylket og hele landet.

Vi skal prøve å komme med en begrunnelse for $\hat{c}^*$:

I de fleste tilfeller er $EF_F \approx p$, og vi har derfor gjort tilnærmelsen $EF_F = p$. Under forutsetning av at denne likheten gjelder, har vi at $E\,(F_F-p)\,(F_L-p) = \text{cov}\,(F_F, F_L)$

Det er lett å vise at $\text{cov}\,(F_F,F_L) \approx \frac{n_F}{n_L}\,\text{var}\,F_F$. (Det er eksakt likhet hvis antall observasjoner fra fylket ikke er en stokastisk variabel) c* blir da tilnærmet:

$$C^* \approx \frac{(1-\frac{n_F}{n_L})\text{var}\,F_F}{\frac{(1-2n_F)}{n_L}\text{var}\,F_F + \text{var}\,F_L + (E\,F_L-p)^2}$$

Vi ser at $\hat{c}^*$ er framkommet ved at vi har estimert var F_F med $\frac{F_F(1-F_F)}{n_F}$, var F_L med $\frac{F_L(1-F_L)}{n_L}$ og $(EF_L-p)^2$ med $\frac{n_F}{n_L}(F_F-F_L)^2$ og satt dette inn i formelen for c*.

Vi skal prøve å begrunne hvorfor vi har estimert $(EF_L-p)^2$ med $\frac{n_F}{n}(F_F-F_L)^2$ En tilnærmet forventningsrett estimator for EF_L-p er F_L-F_F. En tanke var da å bruke $(F_L-F_F)^2$ som estimator for $(EF_L-p)^2$. Men siden $E\,(F_L-F_F)^2 = (1-\frac{2n_F}{n_L})\text{var}\,F_F + \text{var}\,F_L + (EF_L-p)^2$ = hele nevneren i c*, ville $(EF_L-p)^2$ som regel bli kraftig overestimert ved bruk av den nevnte estimatoren.

Vi kom da på den tanken å multiplisere $(F_F-F_L)^2$ med $\frac{n_F}{n_L}$. Dette har den store fordel at det ikke blir lagt så stor vekt på differansen mellom F_F og F_L når antall observasjoner i et fylke er lite i forhold til antall observasjoner fra hele landet.

Siden $E\ (F_L-F_F)^2$ = nevneren i c*, kunne en mulighet være å estimere c* med

$$c^{**} = \frac{(1-\frac{n_F}{n_L})\ \frac{1}{n_F}\ F_F\ (1-F_F)}{(F_F-F_L)^2}$$

Men siden vi vet at c* nesten alltid ligger mellom 0 og 1, ser vi straks at c** ville være en umulig estimator.

Bruttovariansen til p er robust overfor avvik i c fra c*. Av denne grunn er det ikke usannsynlig at den kombinerte estimatoren med c* estimert ved $\hat{c}^*$ i de fleste tilfeller er en bedre estimator enn både F_L og F_F. Før en tar i bruk en kombinert estimator, er det likevel nødvendig å se nærmere på estimeringen av c*.

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422

Trykt 1980

Nr. 80/1 Svein Longva, Lorents Lorentsen and Øystein Olsen: Energy in a Multi-Sectoral Growth Model Energi i en flersektors vekstmodell Sidetall 29 Pris kr 9,00 ISBN 82-537-1082-8

- 80/2 Viggo Jean-Hansen: Totalregnskap for fiske- og fangstnæringen 1975 - 1978 Sidetall 33 Pris kr 9,00 ISBN 82-537-1080-1

- 80/3 Erik Biørn og Hans Erik Fosby: Kvartalsserier for brukerpriser på realkapital i norske produksjonssektorer Sidetall 60 Pris kr 11,00 ISBN 82-537-1087-9

- 80/4 Erik Biørn and Eilev S. Jansen: Consumer Demand in Norwegian Households 1973 - 1977 A Data Base for Micro-Econometrics Sidetall 130 Pris kr 13,00 ISBN 82-537-1086-0

- 80/5 Ole K. Hovland: Skattemodellen LOTTE Testing av framskrivingsmetoder Sidetall 30 Pris kr 9,00 ISBN 82-537-1088-7

- 80/6 Fylkesvise elektrisitetsprognoser for 1985 og 1990 En metodestudie Sidetall 56 Pris kr 11,00 ISBN 82-537-1091-7

- 80/7 Analyse av utviklingen i elektrisitetsforbruket 1978 og første halvår 1979 Sidetall 22 Pris kr 7,00 ISBN 82-537-1129-8

- 80/8 Øyvind Lone: Hovedklassifiseringa i arealregnskapet Sidetall 50 Pris kr 9,00 ISBN 82-537-1104-2

- 80/9 Tor Bjerkedal: Yrke og fødsel En undersøkelse over betydningen av kvinners yrkesaktivitet for opptreden av fosterskader Occupation and Outcome of Pregnancy Sidetall 93 Pris kr 11,00 ISBN 82-537-1111-5

- 80/10 Otto Carlson: Statistikk fra det økonomiske og medisinske informasjonssystem Alminnelige somatiske sykehus 1978 Sidetall 65 Pris kr 11,00 ISBN 82-537-1119-0

- 80/11 John Dagsvik: A Dynamic Model for Qualitative Choice Behaviour Implications for the Analysis of Labour Force Participation when the Total Supply of Labour is Latent En dynamisk teori for kvalitativ valghandling Implikasjoner for analyse av yrkesdeltaking når det totale tilbud av arbeid er latent Sidetall 25 Pris kr 9,00 ISBN 82-537-1152-2

- 80/12 Torgeir Melien: Ressursregnskap for jern Sidetall 56 Pris kr 9,00 ISBN 82-537-1138-7

- 80/13 Eystein Glattre og Ellen Blix: En vurdering av dødsårsaksstatistikken Feil på dødsmeldingene Evaluation of the Cause-of-Death-Statistics Sidetall 73 Pris kr 11,00 ISBN 82-537-1136-0

- 80/14 Petter Frenger: Import Share Functions in Input - Output Analysis Importandelsfunksjoner i krysløpsmodeller Sidetall 41 Pris kr 9,00 ISBN 82-537-1143-3

- 80/15 Den statistiske behandlingen av oljevirksomheten Sidetall 56 Pris kr 11,00 ISBN 82-537-1150-6

- 80/16 Adne Cappelen, Eva Ivås og Paal Sand: MODIS IV Detaljerte virkningstabeller for 1978 Sidetall 261 Pris kr 15,00 ISBN 82-537-1142-5

- 80/18 Susan Lingsom: Dagbøker med og uten faste tidsintervaller: En sammenlikning basert på prøveundersøkelse om tidsnytting 1979 Open and Fixed Interval Time Diaries: A Comparison Based on a Pilot Study on Time Use 1979 Sidetall 31 Pris kr 9,00 ISBN 82-537-1158-1

- 80/19 Sigurd Høst og Trygve Solheim: Radio- og fjernsynsundersøkelsen januar - februar 1980 Sidetall 101 Pris kr 13,00 ISBN 82-537-1155-7

- 80/20 Skatter og overføringer til private Historisk oversikt over satser mv. Årene 1969 - 1980 Sidetall 72 Pris kr 11,00 ISBN 82-537-1151-4

- 80/21 Olav Bjerkholt og Øystein Olsen: Optimal kapasitet og fastkraftpotensial i et vannkraftsystem Sidetall 36 Pris kr 9,00 ISBN 82-537-1154-9

- 80/22 Rolf Aaberge: Eksakte metodar for analyse av to-vegstabellar Sidetall 80 Pris kr 11,00 ISBN 82-537-1161-1

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) (forts.) - ISSN 0332-8422

Trykt 1980 (forts.)

Nr. 80/23 P. Frenger, E.S. Jansen and M. Reymert: Tariffs in a World Trade Model An Analysis of Changing Competitiveness due to Tariff Reductions in the 1960's and 1970's Sidetall 47 Pris kr 9,00 ISBN 82-537-1163-8

- 80/24 Jan Mønnesland: Bestandsuavhengige giftermålsrater Sidetall 50 Pris kr 11,00 ISBN 82-537-1167-0

- 80/25 Kari Lotsberg: Virkninger for norsk økonomi av endringer i samhandel Norge - utviklingslandene Sidetall 67 Pris kr 11,00 ISBN 82-537-1170-0

- 80/26 Lasse Fridstrøm: Lineære og log-lineære modeller for kvalitative avhengige variable Linear and Log-Linear Qualitative Response Models Sidetall 122 Pris kr 13,00 ISBN 82-537-1184-0

- 80/27 Aktuelle skattetall 1980 Current Tax Data Sidetall 43 Pris kr 9,00 ISBN 82-537-1194-8

- 80/28 Forbruksundersøkelse blant soldater 1979 Sidetall 45 Pris kr 11,00 ISBN 82-537-1199-9

- 80/29 Konsumprisindeksen Sidetall 61 Pris kr 9,00 ISBN 82-537-1203-0

- 80/30 Totalregnskap for fiske- og fangstnæringen 1976 - 1979 Sidetall 37 Pris kr 9,00 ISBN 82-537-1205-7

- 80/31 P. A. Garnåsjordet, Ø. Lone and H. V. Sæbø: Two Notes on Land Use Statistics Sidetall 47 Pris kr 9,00 ISBN 82-537-1214-6

- 80/32 Knut Ø. Sørensen: Glatting av flytterater i Statistisk Sentralbyrås befolkningsframskrivinger Sidetall 26 Pris kr 7,00 ISBN 82-537-1216-2

Trykt 1981

Nr. 81/2 Tiril Vogt: Referansearkiv for naturressurs- og forurensningsdata 2. utgave Sidetall 424 Pris kr 20,00 ISBN 82-537-1233-2

- 81/3 Nils Håvard Lund: Byggekostnadsindeks for boliger Sidetall 127 Pris kr 15,00 ISBN 82-537-1232-4

- 81/4 Anne Lise Ellingsæter: Intervjuernes erfaringer fra arbeidskraftundersøkelsene Rapport fra 99 intervjuere Field Work Experiences with the Labour Force Sample Survey Reports from 99 Interviewers Sidetall 40 Pris kr 10,00 ISBN 82-537-1234-0

- 81/5 Bjørn Kjensli: Strukturundersøkelse for bygg og anlegg Vann- og kloakkanlegg Sidetall 62 Pris kr 15,00 ISBN 82-537-1235-9

- 81/6 Erling Siring og Ib Thomsen: Metoder for estimering av tall for fylker ved hjelp av utvalgsundersøkelser Sidetall 42 Pris kr 10,00 ISBN 82-537-1509-9

- 81/8 Morten Reymert: En analyse av faktorinnsatsen i Norges utenrikshandel med utviklingsland og industriland Sidetall 55 Pris kr 15,00 ISBN 82-537-1506-4

Pris kr 10,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.

ISBN 82-537-1509-9
ISSN 0332-8422