

Ann Marit Kleive og Leiv Solheim

**Kalibrering av vektene i
utvalgsundersøkelsen**

Erfaringer fra utvalgene til inntekts-
og formuesundersøkelsene
1991 og 1992



Innhold

1. Innledning	3
2. Generelt om populasjon, utvalg og vektorer for 1991 og 1992	3
2.1. Utvalg	3
2.2. Vektorer	4
2.3. Oppretting av skjevheter	6
2.4. Sammenkobling av IF og IFN	7
2.5. Kalibrering	8
3. IF/IFN 1991	9
3.1. Utvalgene	9
3.2. Aldersjustering av IF-utvalget	10
3.3. Kontroll av IFN-utvalget og næringsdelen av IF-utvalget	13
3.4. Sammenkobling av IF og IFN	16
3.5. Kalibreringsvariable	19
3.6. Diskusjon	21
4. IF/IFN 1992	22
4.1 Utvalg	22
4.2. Aldersjustering av IF-utvalget	23
4.3. Kontroll av IFN-utvalget og næringsdelen av IF-utvalget	25
4.4. Sammenkobling av IF og IFN	26
4.5. Kalibreringsvariable	27
4.6. Fordelingen i utvalget	29
4.7. Konklusjoner	31
5. Diskusjon	31
5.1. Noen kritiske merknader til utvalgene i IF91 og IF92	31
5.2. Arbeidsoppgaver fremover - forslag til forbedringer	32
6. Referanser	33

1. Innledning

Inntekts- og formuesundersøkelsene ble frem til 1984 gjennomført av Statistisk sentralbyrå hvert tredje år. Etter dette er undersøkelsen blitt årlig. En fullstendig dokumentasjon for IF-undersøkelsen finnes i Notat 94/8 av Grete Sparby. Dette notatet ble skrevet til undersøkelsen i 1990, men har relevans også for de senere undersøkelsene. For å ha et grunnlag for å evaluere skattereformen, ble en tilsvarende inntekts- og formuesundersøkelse gjennomført også for selvstendig næringsdrivende i 1991 og 1992.

For 1990 undersøkelsen ble de estimerte totalene beregnet ved vektor basert på en kombinasjon av etterstratifiserte inverse trekksannsynligheter i det enkelte utvalget, etterfulgt av en veiet summering av vektene. Denne metoden er beskrevet av Leiv Solheim i vedlegg 11 i Notat 94/8. I 1989 begynte Johan Heldal et arbeid med å forbedre vektene slik at de estimerte verdiene ut fra utvalget stemte bedre overens med totalene i ligningsregisteret. Dette arbeidet resulterte i en metode for kalibrering av vektene. Metoden er beskrevet i Notat 92/3 *A method for Calibration of Weights in Sample Surveys* av Johan Heldal. Denne teknikken ble bl.a. benyttet i et arbeid med bruk av utvalgsundersøkelser i nasjonalregnskapet. En artikkel om dette arbeidet er publisert av Heldal og Røstadsand i ØA 2/94.

Den kalibreringsteknikken som er beskrevet i dette notatet ble første gang benyttet i inntekts- og formuesundersøkelsene i 1991. Som det skal vises senere i notatet resulterte dette i totaler som lå nærmere ligningsregisterets verdier enn det som var tilfelle ved kun å etterstratifisere etter alder. Samme teknikk ble benyttet i 1992, men med litt andre valg av kalibreringsvariable. I 1992-undersøkelsene ble fordelingen i utvalget undersøkt nærmere, og det viste seg at kalibrering kan skape en kunstig skjevhet i utvalget i forhold til registerverdiene. Det er derfor i det siste punktet i notatet skissert en del forslag til forbedringer som vil bli studert nærmere senere.

2. Generelt om populasjon, utvalg og vektor for 1991 og 1992

2.1. Utvalg

Utvalgene som danner grunnlaget for inntekts- og formuesundersøkelsene er trukket etter forskjellige utvalgsplaner for å ta hensyn til forskjellige krav til det datagrunnlaget en har behov for.

Utvalget til den ordinære inntekts- og formuesundersøkelsen (IF) består av tre delutvalg:

- levetårsutvalget**
- forbruksutvalget paneldel**
- forbruksutvalget ny del.**

Levetårsutvalget og den nye delen av forbruksutvalget er trukket som et selvveiende personutvalg der husholdningsmedlemmene over 15 år inkluderes i utvalget. Paneldelen i forbruksutvalget er for 1991 trukket på familienivå, mens den for 1992 er trukket på personnivå. Begge disse utvalgene er selvveiende.

Spesialutvalget til inntekts- og formuesundersøkelsen for næringsdrivende (IFN) ble trukket blant de som var registrert som selvstendig næringsdrivende, definert ved å ha næringsinntekt i ligningsregisteret for 1990. Grunnlaget for IFN-utvalget er 8.3 % (eller hver 12. person) av befolkningen over 15 år som var med i FOB90. Denne populasjonen ble koblet sammen med ligningsregisteret fra 1990 for å finne fram til de personene som hadde næringsinntekt i 1990. IFN-utvalget ble trukket fra denne populasjonen og skal derfor være et representativt utvalg blant personer med næringsinntekt i 1990.

IFN-populasjonen kan deles i tre stratum ut fra sammenhengen mellom ligningsregisteret og opplysningene fra FOB90:

- A: Har kun næringsinntekt og er FOB90-selvstendig.
- B: Har både nærings- og lønnsinntekt og er IKKE FOB90-selvstendig.
- C: Har kun næringsinntekt og er IKKE FOB90-selvstendig.

Stratum A kan videre deles inn i 5 undergrupper: primærnærings, håndverk, handel, industri og frie yrker på grunnlag av næringsopplysningene i FOB90. Husholdningsmedlemene ble inkludert i utvalget ved å ta utgangspunkt i husholdningene som ble etablert på grunnlag av FOB90, registrerte endringer i det sentrale personregisteret og informasjon fra selvangivelsene.

Delutvalgene til IF-undersøkelsen og utvalget til IFN-undersøkelsen kobles til ligningsregisteret det enkelte året for å finne verdier for aktuelle inntekts- og formuesvariable.

2.2. Vekter

Som utgangspunkt for videre beregninger benyttes inverse trekk sannsynligheter som vekter. For de av delutvalgene i IF som er personutvalg, er vektene omvendt proporsjonale med antall personer over 15 år i husholdningen. Det er kun personer som er 16 år og eldre som er trekkbare til de opprinnelige personutvalgene. For IFN vil det si at kun de som er 17 år og eldre er trekkbare i 1991, og de som er 18 år og eldre er trekkbare i 1992. Vektene til utvalg som er trukket på grunnlag av familiene er proporsjonale med antall familier i husholdningene.

Grunnlaget for IFN-utvalget var først å trekke ut et 8.3 % personutvalg (hver 12. person) blant alle personer født senest i 1974. Disse hadde alle som nevnt tidligere vært med i FOB90. Dessuten skulle IO ha næringsinntekt i ligningsregisteret for 1990. Dette kriteriet krever imidlertid at IO i tillegg til å være med i FOB skal være selvstendig næringsdrivende, definert ved at IO hadde næringsinntekt i 1990. Det betyr at en husholdning kunne komme med i dette utvalget dersom den(de) personen(e) som ble trukket ut til FOB samtidig var selvstendig næringsdrivende. For å finne trekk sannsynligheten til husholdningen må trekk sannsynligheten til de andre selvstendige i husholdningen finnes. Trekk sannsynligheten til IO-ene avhenger av hvilket stratum de tilhører. For å beregne trekk sannsynligheten til de andre selvstendige i husholdningen må det utvikles en metode for å plassere disse i et riktig stratum. Siden disse ikke var med i FOB90 må det finnes en annen kilde som erstatter denne. Disse personene ble klassifisert etter forholdet mellom næringsinntekt og lønnsinntekt i 1990 (pensjonsgivende inntekt høy og lav sats).

- I. Dersom IO tilhører stratumgruppe A plasseres en annen næringsdrivende i husholdningen i stratumgruppe A dersom
næringsinntekt90 > lønnsinntekt90
ellers i stratumgruppe B.
- II. Dersom IO tilhører stratumgruppe B plasseres en annen næringsdrivende i husholdningen i stratumgruppe B dersom
næringsinntekt90 < lønnsinntekt90
ellers i stratumgruppe A.
- III. Dersom IO tilhører stratumgruppe C plasseres andre næringsdrivende i stratumgruppe C dersom disse **KUN har næringsinntekt**

ellers i stratumgruppe A eller stratumgruppe B avhengig av om næringsinntekten er større enn lønnsinntekten eller ikke.

Tabellen nedenfor viser et eksempel på anvendelse av denne metoden på dataene fra 1991-undersøkelsen.

Tabell 2.1. Sammenhengen mellom faktisk stratumtilhørighet og beregnet stratum-tilhørighet for husholdningsmedlemmene i IFN91 på grunnlag av I-III for IO-ene.

Beregnet Faktisk	1	2	3	4	5	6	7	Sum
1	481	0	0	0	0	19	0	500
2	0	458	0	0	0	42	0	500
3	0	0	460	0	0	40	0	500
4	0	0	0	454	0	45	0	499
5	0	0	0	0	795	55	0	850
6	142*	41*	41*	15*	92*	912	0	1245
7	0	0	0	0	0	0	500	500

*Fordelingen på strata 1-5 er beregnet ut fra fordelingen til FOB-populasjonen.

Som tabell 2.1. viser, så er graden av korrekt plassering ganske bra. Dette eksempelet er basert på data fra 1990 og 1991, men metoden er også benyttet ved senere undersøkelser.

Treksanssynligheten til husholdningen er summen av treksanssynlighetene til de selvstendig næringsdrivende i husholdningen og som kunne vært trukket ut i FOB90, d.v.s. de som er over 15 år. Settes $p_{IO,j}$ lik treksanssynligheten for de enkelte næringsdrivende i husholdningen til IO medregnet IO, blir treksanssynligheten til husholdningen, $p_{H(IO)}$, da lik

$$(2.1) \quad p_{H(IO)} = \sum_j p_{IO,j}$$

Vekten til husholdningen, $w_{H(IO)}$, blir videre lik

$$(2.2) \quad w_{H(IO)} = 1/p_{H(IO)}$$

De selvstendig næringsdrivende i en husholdning kan deles inn i en eller flere bedrifter. En bedrift i husholdningen trekkes ut hvis og bare hvis en selvstendig næringsdrivende IO i husholdningen er trukket ut. Det betyr at bedriftene i husholdningen får samme treksanssynlighet som husholdningen, altså blir vekten til bedriftene i husholdningen lik $w_{H(IO)}$.

Utvalget er trukket på grunnlag av opplysninger om næringsinntekt i 1990. Siden populasjonen er fra 1990 og undersøkelsene skjer i 1991 og 1992 medfører det at de som er nyetablerte som selvstendig næringsdrivende etter 1990 ikke hadde mulighet til å bli trukket ut til utvalget. I tillegg kan det skje ulike forandringer med personene i det utvidede utvalget. Disse endringene er det viktig å skille mellom.

- i IO hadde næringsinntekt i 90, men ikke i 91.
- ii Husholdningen til IO hadde næringsinntekt i 90, men ikke i 91.
- iii IO har næringsinntekt i 90 og 91, men i forskjellig næring.

iv Antallet "bedrifter" i husholdningen og sammensettingen av bedriftene er forandret.

I det første tilfellet dreier det seg om en reell avgang for IO fra hele populasjonen, men nødvendigvis behøver ikke husholdningen forsvinne fra næringshusholdningene. I det andre tilfellet er det en reell avgang blant de næringsdrivende husholdningene. I det tredje tilfellet har IO skiftet næring. Det fjerde tilfellet dekker veldig mange muligheter. På grunn av disse endringene kan det bli inkonsistens mellom populasjon og utvalg. Ligningsverdiene for de enkelte ligningsvariablene er ikke kjent for de ulike næringstypene slik at det ikke kan korrigeres for antall innen hvert strata. Antall næringsdrivende innen hvert strata er heller ikke kjent fra noe register. Derfor er det ikke mulig å korrigere slik at det er konsistens mellom utvalget og registeret. Derimot benyttes som kalibreringsvariable ligningsverdier for næringsinntekt jord/skog/fiske og annen næring. Kalibreres det med hensyn på disse variablene vil vektene korrigeres slik at det er konsistens mellom utvalg og register for strata A og B.

2.3. Oppretting av skjevheter

Det er en klar sammenheng mellom alder og inntekt. Undersøkelser viser at den gjennomsnittlige inntekten for dagens populasjon øker med alder inntil 50 år for deretter å avta. Eksempler på dette er gitt i figur 7.5.3c og 7.5.3d i NOU 1993:17 Levekår i Norge. Endel trygdeytelser som barnetrygd og alderspensjon avhenger også sterkt av alderen. Dette betyr at det kan oppstå problemer i estimeringen av ligningstotalene dersom aldersfordelingen er skjev. For modellanalyser ved bruk av skattemodellen LOTTE er det også vesentlig at aldersfordelingen er riktig i forhold til befolkningen.

For å rette opp skjevhetene i aldersfordelingen i IF-utvalgene og dermed også unngå de analytiske problemene som kan oppstå, er utvalgene etterstratifisert ved kalibrering mot alderstotalene fra befolkningsregisteret. For å ta hensyn til ytelser som varierer med alder, regler for hvem som skal levere selvangivelse og behov for å begrense antall aldersgrupper ut fra mer metodiske hensyn, er materialet delt inn etter følgende aldersgruppering

Intervall	Kommentar
0-6 år	
7-12 år	
13-15 år	Leverer egen selvangivelse, barnetrygd inntil fylte 16 år.
16-19 år	Skoleungdom
20-24 år	Deler intervallet 20-29 år i to deler p.g.a. stor forskjell i
25-29 år	yrikesaktiviteten i disse to gruppene.
30-39 år	
40-49 år	
50-59 år	
60-66 år	
67- år	Alderspensjonister

Disse intervallinndelingene ble benyttet i inntekts- og formuesundersøkelsen for 1992. Undersøkelsen i 1991 hadde en finere inndeling for aldersgruppen 30-60 år, men denne forskjellen i inndeling betyr lite. Aldersgruppene fra 1992 vil bli benyttet i senere inntekts- og formuesundersøkelser.

Det er rimelig å tro at den skjevheten som finnes i IF-utvalget, også er et problem i IFN-utvalget. Siden populasjonen av personer som tilhører husholdninger med næringsinntekt ikke er kjent i noe register, er det ikke mulig å rette opp aldersfordelingen i IFN-utvalget på grunnlag av registerverdier. En mulighet hadde vært å benytte opplysninger om husholdningene i FOB90 som grunnlag.

2.4. Sammenkobling av IF og IFN

I statistikkårene 91 og 92 inneholder utvalget til IF-undersøkelsen selvstendig næringsdrivende, mens IFN vil inneholde husholdninger uten selvstendig næringsdrivende. Næringshusholdninger (NH) defineres som husholdninger med minst en næringsinntekt, og lønnsusholdninger (LH) defineres som husholdninger med kun lønnsinntekt. Dermed kan utvalgene karakteriseres ved følgende skjema:

IF: LH90 NH90
 IFN: NH90

De ulike typene statistikk produseres ved hjelp av data og enheter som definert nedenfor.

IFN-STATISTIKK: Data = IF/NH og IFN/NH
 Enhhet = Bedriftene i husholdningen
 = Foretaket næringsvirksomhet

IF-STATISTIKK: Data = IF og IFN
 Enhhet = Personer
 = Husholdninger

Lotte-modeller: Data = IF og IFN
 Enhhet = Personer
 = Husholdninger

Når utvalgene for IF og IFN skal settes sammen må det beregnes felles vekter. Det må da tas hensyn til om personer/husholdninger kun kan være med i IF (LH90) eller i både IF og IFN (NH90). Deretter må vektene korrigeres slik at det tas hensyn til antall i hver gruppe. For å forenkle situasjonen kan det lønne seg å sette opp en skjematisk oversikt over de to utvalgene.

Undersøkelse	Populasjon/ Utvalg	Familier uten nær.innt.		Familier med nær. innt.	
		Antall	Vekter	Antall	Vekter
IF	Utvalg	$n_{L,IF}$	w_{IF}	$n_{S,IF}$	w_{IF}
	Populasjon	$N_{L,IF}$	$N_{S,IF}$		
IFN	Utvalg			$n_{S,IFN}$	w_{IFN}
	Populasjon			$N_{S,IFN}$	

der w_{IF} og w_{IFN} beregnes ut fra om personen tilhører en husholdning med næringsinntekt eller ikke.

IFN-utvalget er trukket på grunnlag av populasjonen i 1990. Det var kun personer med næringsinntekt i 1990 som var trekkbare. IF-utvalget er trukket direkte fra populasjonen det aktuelle året. Det betyr at vektene for hvert av de to utvalgene beregnes på bakgrunn av to forskjellige populasjoner, men sammenvektingen av utvalgene må skje på grunnlag av data fra samme populasjon.

IF-utvalget i 91 besto av leveårsutvalget som er et personutvalg og utvalget til forbruksundersøkelsen. Det sistnevnte utvalget ble som nevnt delt i to deler, et familieutvalg (paneldelen) og et personutvalg (de nye i forbruksutvalget i 1991). Det betyr at IO-ene i levekår og utenfor panelet i forbruk er et selvveiende personutvalg, mens panelet i forbruk er et selvveiende

familieutvalg. I tillegg inkluderes alle personene i husholdningene til IO-ene. IF-utvalget i 1992 er et rent personutvalg, og slik kommer det til å bli i de fremtidige IF-undersøkelsene. Det utvidede personutvalget får ut fra ulike sammensetninger følgende egenskaper:

Det utvidete personutvalget i **levetår og utenfor panelet i forbruk** har trekksannsynligheter som blir proporsjonale med antall trekkbare personer i husholdningen. Husholdningene får også trekksannsynligheter som er proporsjonale med antall personer i husholdningene. Det betyr at personvektene og husholdningsvektene blir omvendt proporsjonale med antall personer i utvalget.

Det utvidete personutvalget i **paneldelen til forbruk** er et selvveiende personutvalg siden det opprinnelige familieutvalget er selvveiende familieutvalg. Altså blir personvektene proporsjonale med antall familier i husholdningen.

For undersøkelser der hele utvalget er personutvalg (som for levetår og utenfor paneldelen i forbruk 1991) blir trekksannsynligheten proporsjonal med antall personer i husholdningen.

For utvalget til IFN er hovedsaken at personvektene og den tilhørende husholdningsvekten er tilnærmet omvendt proporsjonale med antall personer i husholdningen med næringsinntekt i 1990. Mer presis beskrivelse av dette er gitt tidligere.

Både for IF- og IFN-undersøkelsen skulle det ideelt sett vært skilt mellom husholdninger helt uten næringsinntekt og husholdninger med næringsinntekt. Dette er mulig i utvalget, men ikke i populasjonen siden husholdning ikke er et registerkjenne-tegn.

Definer nå I_{IF} slik at I_{IF} settes lik 1 dersom et individ er med i IF og null ellers, og tilsvarende settes I_{IFN} lik 1 dersom et individ er med i IFN og null ellers. Ny vekt til en person i en næringshusholdning i det kombinerte utvalget kan da beregnes på følgende vis:

$$(2.3) \quad w_s = (n_{s,IF} / n_s) * I_{IF} * w_{IF} + (n_{s,IFN} / n_s) * I_{IFN} * w_{IFN}$$

der $n_L = n_{L,IF} + n_{L,IFN}$ og $n_L = n_{L,IF} + n_{L,IFN}$.

Vektfaktoren til det sammensatte utvalget blir beregnet på grunnlag av (2.3) avhengig av om personen tilhører en husholdning med næringsinntekt eller ikke. Hvis personen tilhører en ren lønnstakerhusholdning har det ingen innvirkning på vekten.

2.5. Kalibrering

Når det er justert for den skjevheten som skyldes aldersfordelingen i utvalget, resulterer det som det skal vises senere i bedre aggregerte totalverdier enn når inverse trekksannsynligheter blir benyttet som vekter. Imidlertid gjenstår det mindre avvik i forhold til totaltallene fra ligningsregisteret. Disse kan rettes opp ved å kalibrere mot noen av totaltallene. Kalibrering er en metode for korrigeringsvektene basert på en multiplere regresjonsmodell. Metoden justerer vekten slik at de oppblåste totalene for noen av ligningsvariablene tvinges til å stemme eksakt med totalene fra ligningsregisteret. Ved gunstige valg av kalibreringsvariable vil også totalene for de andre ligningsvariablene ligge nær de eksakte verdiene.

Ved valg av kalibreringsvariable må det velges variable som ikke er innbyrdes høyt korrelerte, og det må tas hensyn til at metoden kan være ømfintlig overfor antall kalibreringsvariable. Samtidig må

kalibreringsvariablene være representative for de andre variablene. Kalibreringsmetoden som er benyttet i disse undersøkelsene er beskrevet i Heldal (1992).

3. IF/IFN 1991

3.1. Utvalgene

Utvalget til den ordinære inntekts- og formuesundersøkelsen i 1991 (IF91) deles i tre delutvalg. Det dominerende delutvalget er levekårsutvalget, trukket som et personutvalg der husholdningsmedlemmene innkluderes i utvalget. Videre er forbruksutvalget med i IF91. Dette er det nødvendig å dele i to deler siden panelet i forbruk er et familieutvalg mens de nye i forbruk er et personutvalg etter samme lest som levekårsutvalget. I det videre arbeidet analyseres først disse tre delutvalgene hver for seg.

Til inntekts- og formuesundersøkelsen for næringsdrivende (IFN91) er utvalget trukket blant de som var selvstendig næringsdrivende i 1990.

Tabell 3.1. Antall husholdninger og personer i de ulike utvalgene for IF-undersøkelsen.

Utvalg	Husholdninger	Personer
IF, i alt	3.489	9.303
Levekår	2.509	6.867
Forbruk90	493	1.227
Forbruk91	485	1.209
IFN	4.598	15.055
Totalt	8.075	24.328

I analysen fram til et forslag til vekter har hovedsaken vært å rette opp den skjeve aldersfordelingen i utvalgene. Deretter er vektene kalibrert mot inntektsvariabler, trygdegrunnlaget og statsskatt på formuen.

Levekårsutvalget og panel 1991 i forbruksutvalget er et personutvalg, mens panel 1990 i forbruksutvalget er et familieutvalg. Som utgangspunkt er de inverse trekkesannsynlighetene benyttet som vekter.

Utvalg	Vekt
Levekårsutvalget	Omvendt proporsjonal med antall personer over 15 år i husholdningen.
Forbruksutvalget, panel 1990	Omvendt proporsjonal med antall familier i husholdningen.
Forbruksutvalget, panel 1991	Omvendt proporsjonal med antall personer over 15 år i husholdningen.

Tabell 3.2. Aldersfordeling i utvalgene basert på trekksannsynlighetene

Alder	Levekår	Forbruk		Befolkningen 01.01.92
		Panel 90	Panel 91	
00 - 06	433.586	482.123	437.658	398.008
07 - 12	312.956	332.129	349.893	311.434
13 - 15	152.861	172.761	172.606	161.165
16 - 19	251.512	217.290	143.468	250.124
20 - 24	347.880	287.265	215.736	336.926
25 - 29	309.636	346.191	384.564	328.371
30 - 34	307.079	385.698	366.276	317.121
35 - 39	299.654	361.592	302.499	310.003
40 - 44	327.467	281.238	296.648	310.976
45 - 49	286.806	281.908	239.024	270.157
50 - 54	197.741	252.445	215.787	202.215
55 - 59	185.906	199.880	182.854	182.471
60 - 66	280.894	282.243	290.797	273.517
67 +	606.665	553.102	774.679	570.000 (614.248)*
Totalt	4.300.644	4.435.866	4.372.487	4.222.486 (4 273 634)

*Blant personer over 67 år er de som er på institusjon trukket fra. Anslått befolkning over 67 år utenom institusjon er 570 000.

Tabell 3.2. viser at den estimerte aldersfordelingen i utvalgene er skjev i forhold til befolkningsregisteret når de oppblåste verdiene baserer seg på trekksannsynlighetene. En korrigeringsfaktor for aldersfordelingen er derfor nødvendig.

Først er utvalgene korrigert ved å rette opp aldersfordelingen direkte på personnivå. Dette fører imidlertid til at personer i samme husholdning får forskjellig vekt, noe som ikke er mulig i dagens versjon av skattemodellen LOTTE. Disse korrigeringsfaktorene må derfor ses på som en kontroll når kalibreringsmetoden (Heldal, 1992) benyttes for å rette opp aldersfordelingen. I tillegg benyttes andre kalibreringsvariable til å justere vektene, men med forutsetning om at vektene til alle personer i samme husholdning er like.

3.2. Aldersjustering av IF-utvalget

For å ta hensyn til aldersfordelingen er personutvalget etterstratifisert etter alder. Dette skaper et lite problem siden personene i en husholdning ikke får samme trekksannsynlighet. Dersom en skal produsere statistikk med husholdning som enhet og vil ha konsistens med personstatistikken må det enten etterstratifiseres slik at personene i en husholdning får samme vekt, eller en må telle husholdningene ved personvektene ved å telle hver person som $1/h$ er h er antall personer i husholdningen. I praksis vil dette ikke skape noen problemer siden aldersfordelingen rettes opp på husholdningsnivå ved kalibreringsmetoden til Heldal.

Effekten av å rette opp den skjeve aldersfordelingen kan best illustreres ved å sammenligne de oppblåste totalverdiene for ligningsvariablene ved de to vektmetodene: inverse vekter og aldersjusterte vekter. For hver av metodene må det beregnes en felles vekt når de tre delutvalgene i IF skal settes sammen til et utvalg.

Ialt er det 9 303 personer i de tre utvalgene tilsammen. Av disse er det 6 867 personer i levekårsutvalget, 1 227 personer i panel 90 og 1 209 personer i panel 91. La w_L , w_{90} og w_{91} betegne disse vektene (enten inverse trekksannsynligheter eller aldersjusterte) og I_L , I_{90} og I_{91} betegne indikatorfunksjoner som er lik 1 hvis personene er med i det aktuelle utvalget og 0 ellers. Da kan vekten for hver person i det samlede utvalget finnes ved

$$(3.1) \quad w_{IF} = \frac{(6867 * I_L * w_L + 1227 * I_{90} * w_{90} + 1209 * I_{91} * w_{91})}{9303}$$

Dette er basert på metoden beskrevet i vedlegg 11 i Notat 94/8. Benyttes vektene gitt ved (3.1) til å beregne totaler for ligningsvariablene, gir det tall som vist i tabellen nedenfor.

Tabell 3.3. Oppblåste totalbeløp fra IF-utvalget basert på inverse trekksannsynligheter og aldersjusterte vektor sammenlignet med ligningsregisterets totaltall.

Variabel	Inverse	aldersj vektor	Lignings- registeret
Pensjonsgiv. innt. lav sats	338.380	334.160	332.137
Pensjonsgiv. innt. jord/skog/fiske	10.098	9.809	10.295
Pensjonsgiv. innt. høy sats	20.173	19.790	19.437
Trygdegrunnlag	86.905	81.764	83.475
Grunnlag for beregning av toppskatt	453.451	443.615	443.448
Formueskatt kommune	2.739	2.602	2.606
Formueskatt stat	1.580	1.495	1.661
Inntektsskatt kommune	60.086	58.768	58.173
Inntektsskatt stat	9.389	9.131	9.037
Fellesskatt	16.701	16.296	16.308
Toppskatt stat	4.947	4.829	4.638
Skatt på betydelig aksjesalg	0.510	0.484	0.508
Forsinkelsesavgift	0.124	0.121	0.023
Tilleggsskatt	0.351	0.346	0.273
Trygdeavgift lav sats	1.384	1.301	1.326
Trygdeavgift mellom sats	26.875	26.526	26.433
Trygdeavgift høy sats	2.371	2.325	2.328
Skattebegrensning § 78	1.419	1.341	1.477
Forsørgerfradrag	2.060	2.037	1.951
Nettoinntekt stat	372.876	363.991	365.323

Av tabell 3.3 fremgår det at de aldersjusterte tallene er nærmere ligningsregisterets tall enn ved bruk av inverse trekksannsynligheter. Det understreker viktigheten av å ha en korrekt aldersfordeling i utvalget i forhold til populasjonen. Det bør også nevnes at det er klare forskjeller mellom de tre delutvalgene i IF. Spesielt viste det seg at oppblåste tall basert kun på IO-ene i levekårsutvalget ga totaltall som lå veldig nær ligningsregisterets tall.

Tilslutt bør det nevnes at ligningsregisterets bakgrunnspopulasjon ikke er helt identisk med den populasjonen utvalget er trukket fra. I ligningsregisteret er personer på institusjon med, mens utvalget kun omfatter personer utenfor institusjon.

På tilsvarende vis kan antallet bak de forskjellige summene sammenlignes. Resultatene av det er gitt i tabell 3.4.

Tabell 3.4. Oppblåste antall basert på inverse trekksannsynligheter og aldersjusterte vektor sammenlignet med ligningsregisterets totaltall.

Variabel	Inverse	aldersj vektor	Ligning register
Pensjonsgiv. innt. lav sats	2.373.738	2.362.264	2.379.559
Pensjonsg. innt. jord/skog/fiske	145.697	142.146	141.591
Pensjonsgiv. innt. høy sats	179.654	176.180	180.572
Trygdegrunnlag	1.111.516	1.055.773	1.140.339
Grunnlag for beregning av toppskatt	3.224.496	3.163.938	3.274.144
Formueskatt stat	788.084	753.243	745.000
Inntektsskatt kommune	2.922.600	2.851.378	2.935.000
Inntektsskatt stat	1.273.961	1.246.672	1.259.000
Fellesskatt	2.958.712	2.886.515	2.985.790
Toppskatt stat	651.431	636.766	631.104
Skatt på betydelig aksjesalg	1.413	1.273	997
Forsinkelsesavgift	15.924	15.951	25.476
Tilleggs-skatt	25.791	25.429	23.046
Trygdeavgift lav sats	1.056.587	998.320	1.067.248
Trygdeavgift mellom sats	2.182.829	2.166.515	2.169.987
Trygdeavgift høy sats	166.057	162.982	167.682
Skattebegrensning § 78	407.570	384.183	450.261
Forsørgerfradrag	1.014.555	1.004.337	977.058
Nettoinntekt stat	3.206.953	3.144.550	3.243.176

Tabell 3.4 viser et annet bilde enn verdi-tallene i tabell 3.3. Antallet beregnet ved inverse trekksannsynligheter ligger nærmere ligningsregisterets tall enn antallet beregnet ved hjelp av aldersjusterte vektorer. Viktigere er det imidlertid at tallene basert på de aldersjusterte vektene ligger lavere enn ligningsregisterets. Dette kan delvis forklares med at i ligningsregisteret finnes det personer på institusjon, og disse kan ikke bli inkludert i utvalget.

For å studere fordelingen er de samme analysene gjennomført for nettoinntekt, toppskattegrunnlag og antall fordelt på skatteklasse. Resultatene er vist i tabellene nedenfor.

Tabell 3.5. Antall skattytere fordelt på klasser

Skatteklasse	Invers trekks.	Aldersjusterte	Ligningsregisteret
1	2.831.279	2.776.677	3.065.297
2	358.685	351.321	398.697

Tabell 3.6. Nettoinntekt og toppskattegrunnlag fordelt på skattegrunnlag og inntektsintervaller

Klasse	Intervall	Inverse trekks.		Aldersjust.		Ligningsreg.	
		Ant.	Beløp	Ant.	Beløp	Ant.	Beløp
1							
Nettoinntekt							
	00 - 65	2.831	160.5	2.777	156.5	3.065	159.6
	65 - 130	1.966	97.4	1.920	95.2	1.924	94.6
	130 - 164	1.026	27.8	1.005	27.2	995	26.3
	164 - 207	609	18.3	594	17.8	570	17.1
	207 -	280	23.6	272	22.9	263	21.9
Toppskattgr.lag							
	00 - 207	2.859	352.1	2.805	344.5	3.065	346.2
	207 -	556	44.0	543	42.9	533	40.6
2							
Nettoinntekt							
	00 - 81	359	22.5	351	21.9	399	22.9
	81 - 162	208	12.6	204	12.3	219	12.9
	162 - 191	101	2.4	100	2.4	98	2.4
	191 - 216	67	1.4	66	1.3	67	1.4
	216 -	44	6.1	44	5.9	49	6.1
Toppskattgr.lag							
	00 - 249	365	49.8	359	48.9	399	49.3
	249 -	59	7.4	59	7.3	58	7.3

Tabellene 3.5 og 3.6 viser en forholdsvis klar tendens. I klasse 1 er det relativt sett flere og større beløp i utvalget enn i ligningsregisteret, mens i klasse 2 er fordelingen i utvalget omtrent som i ligningsregisteret. Det som slår sterkest ut er imidlertid et for lavt estimert antall i laveste intervall i utvalget sammenlignet med registeret.

3.3. Kontroll av IFN-utvalget og næringsdelen av IF-utvalget

For å følge opp skattereformens virkning for næringsdrivende ble det som tidligere beskrevet trukket et eget utvalg blant selvstendig næringsdrivende i 1990. Utvalgsplanen for dette utvalget er forklart i kapittel 2.1. I IF-utvalget finnes også næringsdrivende. For å studere disse to gruppene av næringsdrivende nærmere, er totaltallene basert på det spesielle næringsutvalget (IFN), næringsdelen av IF-utvalgene og en registerpopulasjon basert på ligningsregisteret, sammenlignet. Fordi en justering med hensyn på aldersfordelingen viste seg å ha positiv effekt på totalverdiene, er de aldersjusterte vektene benyttet i dette avsnittet.

For å forstå tabellene kan det være nødvendig med en nærmere beskrivelse av utvalgene og populasjonen. IFN-utvalget er trukket blant personer som hadde næringsinntekt i 1990. Det vil si at det skal representere en populasjon av husholdninger der det var næringsinntekt i 1990. Næringsdelen av IF-utvalgene er også definert ved at husholdningen hadde næringsinntekt i 1990. Den populasjonen som er etablert fra ligningsregisteret består av alle familier med næringsinntekt i 1990. Det betyr at denne populasjonen er noe mindre enn det næringsutvalgene representerer. Den faktiske forskjellen er at populasjonen mangler husholdninger bestående av flere familier og der minst en av familiene IKKE har næringsinntekt i 1990. Det er derfor rimelig å forvente at næringsinntektene (pensjonsgivende inntekt jord, skog og fiske og pensjonsgivende inntekt høy sats) skulle være like i de to utvalgene og i populasjonen, mens lønnsinntekten (pensjonsgivende inntekt lav sats) og alle andre størrelser som avhenger av lønnsinntekten, skulle bli høyere i utvalgene enn i populasjonen.

Også for nærings-utvalgene kan det være interessant å vurdere aldersfordelingen i utvalgene mot fordelingen i populasjonen. Antall innen hver aldersgruppe i de to utvalgene og i populasjonen er presentert i tabellen nedenfor.

Tabell 3.7. Fordeling av personer etter alder og kjønn i næringshusholdningene i 1990, basert på næringsutvalgene, og i populasjonen av næringsfamilier i 1990.

Alder/kjønn	IFN	IF (næring)	Populasjonen næringsfamiliene 1990
00 - 06	60.008	64.316	60.655
07 - 12	64.623	64.038	65.213
13 - 15	39.581	33.707	35.670
16 - 19	61.431	57.603	56.062
20 - 24	66.089	62.877	58.362
25 - 29	53.921	51.945	43.820
30 - 34	54.823	45.185	50.826
35 - 39	61.934	58.550	60.095
40 - 44	68.249	62.197	66.623
45 - 49	64.964	62.846	63.384
50 - 54	47.846	51.555	46.660
55 - 59	41.316	40.134	38.743
60 - 66	54.702	48.100	48.923
67 -	64.663	47.486	52.418
Menn	431.962	409.589	411.538
Kvinner	377.262	340.949	335.916
Totalt	804.229	750.538	747.454

Tabell 3.7 viser at de utvalgsbaserte husholdningspopulasjonene er større enn familiepopulasjonen fra ligningsregisteret. Forskjellen mellom IFN og IF er vanskelig å forklare. Basert kun på levekår blir totalt antall estimert til 769.000 personer, mens de to forbruksutvalgene kun har henholdsvis 717.000 og 679.000 personer i næringshusholdningene. Det er viktig å huske på at populasjonen er fra 1990, mens tallene i utvalgene er fra 1991.

Tilsvarende sammenligning kan gjøres for ligningsvariablene.

Tabell 3.8. Oppblåste totalbeløp for ligningsvariable basert på IFN-utvalget, IF-næringsdelen og en populasjon av næringsfamilier i 1990. Aldersjusterte vektorer.

	IFN	IF (næring)	Ligningsreg
Pensjonsgiv. innt. lav sats	44.136	41.720	39.362
Pensjonsgiv. innt. jord/skog/fiske	9.743	9.745	9.983
Pensjonsgiv. innt. høy sats	17.498	18.345	18.163
Trygdegrunnlag	9.351	7.842	8.122
Grunnlag for beregning av toppskatt	79.756	76.484	74.733
Formueskatt kommune	0.760	0.833	1.222*
Inntektsskatt kommune	11.333	11.358	10.828
Inntektsskatt stat	2.081	2.200	2.071
Fellesskatt	3.098	3.064	2.949
Toppskatt stat	0.941	0.930	0.958
Skatt på betydelig aksjesalg	0.017	0.000	0.022
Forsinkelsesavgift	0.073	0.036	0.009
Tilleggsskatt	0.789	0.182	0.122
Trygdeavgift lav sats	0.147	0.123	0.128
Trygdeavgift mellom sats	4.126	3.918	3.773
Trygdeavgift høy sats	2.005	2.174	2.177
Skattebegrensning § 78	0.156	0.659	0.102
Forsørgerfradrag	0.401	0.398	0.382
Nettoinntekt stat	68.809	67.140	64.760

*Dette gjelder formueskatt samlet til kommune og stat

Tabellen viser det som ble påpekt, nemlig at næringsinntekten er omtrent på samme nivå i populasjonen og i de to utvalgene. For resten av variablene er tallene spesielt for IFN høyere enn for populasjonen. Det skal også bemerkes at ved kun å benytte levekårsdelen i fra IF vil de estimerte tallene ligge mye nærmere IFN-tallene.

Tabell 3.9. Oppglåste antall for ligningsvariable basert på IFN-utvalget, IF-næringsdelen og en populasjon av næringsfamilier i 1990

	IFN	IF(næring)	Ligningsreg
Pensjonsgiv. innt. lav sats	425.101	404.700	385.847
Pensjonsgiv. innt. jord/skog/fiske	130.861	136.767	128.607
Pensjonsgiv. innt. høy sats	152.309	148.733	153.585
Trygdegrunnlag	145.208	111.628	121.682
Grunnlag for beregning av toppskatt	610.211	566.040	556.843
Formueskatt stat	175.985	166.780	197.140*
Inntektskatt kommune	526.634	488.065	481.476
Inntektsskatt stat	232.126	231.589	222.686
Fellesskatt	533.457	491.152	486.004
Toppskatt stat	110.831	118.035	103.715
Skatt på betydelig aksjesalg	57	0	195
Forsinkelsesavgift	7.771	4.614	9.526
Tilleggsskatt	6.990	10.423	8.508
Trygdeavgift lav sats	127.564	96.371	106.030
Trygdeavgift mellom sats	415.525	396.637	377.155
Trygdeavgift høy sats	140.488	140.037	143.108
Skattebegrensning § 78	42.881	20.805	28.590
Forsørgerfradrag	202.187	188.972	191.457
Nettoinntekt stat	601.226	552.055	540.523

*Dette gjelder formueskatt samlet til kommune og stat

Tabell 3.10. Antall personer i skatteklasse 1 og 2 blant personer i husholdninger og familier med næringsinntekt i 1990. Estimert ved aldersjusterte vektorer.

Skatteklasse	IFN	IF(næring)	Ligningsreg. (næring)
1	524.897	491.475	512.820
2	73.270	57.943	69.698

Tabell 3.11. Nettoinntekt og toppskattegrunnlag fordelt på skattegrunnlag og inntektintervaller. Aldersjusterte vektorer.

Klasse	Intervall	IFN		IF (næring)		Ligningsreg. (næring)	
		Ant.	Beløp	Ant.	Beløp	Ant.	Beløp
1							
Nettoinntekt							
	00 - 65	525	28.2	491	26.7	513	26.1
	65 - 130	342	16.9	336	17.0	324	16.1
	130 - 164	180	4.9	186	5.1	174	4.7
	164 - 207	113	3.7	119	3.8	109	3.0
	207 -	62	6.0	64	7.4	59	6.5
Toppkattgr.lag							
	00 - 207	534	60.7	502	58.9	513	56.8
	207 -	93	8.6	95	8.6	86	8.5
2							
Nettoinntekt							
	00 - 81	73	4.2	58	3.4	70	6.6
	81 - 162	39	2.4	32	2.0	35	2.1
	162 - 191	19	0.5	18	0.5	18	1.2
	191 - 216	14	0.3	14	0.3	13	0.3
	216 -	11	1.5	8	0.9	10	1.4
Toppkattgr.lag							
	00 - 249	76	9.2	64	7.9	70	7.9
	249 -	10	1.2	12	1.1	10	1.5

Tabellene 3.10 og 3.11 viser at både antall og beløp stemmer godt overens for IFN-utvalget, mens næringsdelen av IF ikke viser et like korrekt bilde av inntektsfordelingen i forhold til ligningsregisteret. Årsakene til dette kan være at populasjonen er fra 1990, mens tallene er fra 1991. I tillegg er det ikke sikkert at næringsdelen av IF er representativ for næringspopulasjonen.

3.4. Sammenkobling av IF og IFN

Fremgangsmåten for sammenkobling av IF- og IFN-utvalgene er beskrevet i kapittel 2.4. IF-utvalget er et utvalg som skal være representativt for hele populasjonen, mens IFN-utvalget skal være representativt for populasjonen som består av alle husholdninger som hadde positiv næringsinntekt i 1990. Det betyr at det er kun blant husholdninger med næringsinntekt i 1990 at det er aktuelt å koble de to utvalgene sammen. De nye vektene til det kombinerte utvalget kan beregnes ved å benytte (2.3) som grunnlag. En felles vektfaktor for det sammensatte utvalget er derfor gitt ved

$$(3.2) \quad w_L = w_{IF}$$

$$w_S = \frac{2007 * I_{IF} * w_{IF} + 15055 * I_{IFN} * w_{IFN}}{17062}$$

I kapittel 2.3 ble det argumentert for en aldersjustering av IF-utvalget. Det er tidligere vist at aldersfordelingen i næringsdelen til IF og i IFN er ulik, og estimert totalantall i de to utvalgene er også forskjellig. Dette betyr at det sammensatte utvalget ikke lenger vil ha en riktig aldersfordeling. Som illustrasjon er aldersfordelingen til det sammensatte utvalget vist i tabellen nedenfor.

Tabell 3.12. Fordeling av personer etter alder og kjønn etter sammenkobling av aldersjustert IF og IFN og populasjonen 01.01.92

Alder/kjønn	IF(aldersj.)+IFN	Populasjonen
00 - 06	394.272	398.008
07 - 12	311.950	311.434
13 - 15	166.349	161.165
16 - 19	253.501	250.124
20 - 24	339.760	336.925
25 - 29	330.115	328.371
30 - 34	325.626	317.121
35 - 39	312.989	310.003
40 - 44	316.316	310.976
45 - 49	272.026	270.157
50 - 54	198.942	212.215
55 - 59	183.514	182.471
60 - 66	279.343	273.615
67 -	585.157	570.000
Menn	2.092.560	2.092.142*
Kvinner	2.177.300	2.130.443*
Totalt	4.269.860	4.222.585

*Bergnet antall menn og kvinner over 67 år utenfor institusjon.

Til tross for at delutvalgene i IF er aldersjustert er det fortsatt en skjevhet i aldersfordelingen. Dette medfører at en korrigering for alder er nødvendig også for det sammensatte utvalget.

I de tre neste tabellene er tall for det sammensatte utvalget før og etter aldersjustering gitt sammen med ligningstallene.

Tabell 3.13. Oppblåste totalbeløp basert på sammensatt utvalg og aldersjustert sammensatt utvalg sammenlignet med ligningsregisterets totaltall.

Variabel	Sammen- satt utv.	Aldersjust.- sammens. utv.	Lignings- registeret
Pensjonsgiv. innt. lav sats	336.291	333.576	332.137
Pensjonsgiv. innt. jord/skog/fiske	9.806	9.844	10.295
Pensjonsgiv. innt. høy sats	19.042	19.173	19.437
Trygdegrunnlag	83.095	81.312	83.475
Grunnlag for beregning av toppskatt	446.502	442.307	443.448
Formueskatt kommune	2.538	2.501	2.606
Formueskatt stat	1.399	1.382	1.661
Inntektsskatt kommune	58.746	58.249	58.173
Inntektsskatt stat	9.026	8.979	9.037
Fellesskatt	16.326	16.178	16.308
Toppskatt stat	4.838	4.841	4.638
Skatt på betydelig aksjesalg	0.485	0.508	0.508
Forsinkelsesavgift	0.153	0.157	0.023
Tilleggs-skatt	0.255	0.244	0.273
Trygdeavgift lav sats	1.323	1.294	1.326
Trygdeavgift mellom sats	26.709	26.505	26.433
Trygdeavgift høy sats	2.176	2.194	2.328
Skattebegrensning § 78	1.421	1.377	1.477
Forsørgerfradrag	2.039	2.003	1.951
Nettoinntekt stat	365.465	361.857	365.323

Tabellen viser at de aldersjusterte tallene er lavere enn når kun IF er aldersjustert. Dette skyldes at vektene jevnt over justeres ned når aldersjusteringen skjer til slutt.

På helt tilsvarende vis kan antall bak de forskjellige summene i tabell 3.13 sammenlignes.

Tabell 3.14. Oppblåst antall basert på sammensatt utvalg, aldersjustert sammensatt utvalg sammenlignet med ligningsregisterets totaltall.

Variabel	Sammen- satt utv.	Aldersjust.- sammens. utv.	Lignings- registeret
Pensjonsgiv. innt. lav sats	2.380.266	2.353.408	2.379.559
Pensjonsgiv. innt. jord/skog/fiske	136.934	135.405	141.591
Pensjonsgiv. innt. høy sats	179.336	177.335	180.572
Trygdegrunnlag	1.085.404	1.063.967	1.140.339
Grunnlag for beregning av toppskatt	3.202.915	3.159.544	3.274.144
Formueskatt stat	761.365	749.949	745.000
Inntektskatt kommune	2.885.411	2.846.369	2.935.000
Inntektsskatt stat	1.247.147	1.233.013	1.259.000
Fellesskatt	2.923.846	2.884.271	2.985.790
Toppskatt stat	630.409	623.560	631.104
Skatt på betydelig aksjesalg	1.323	1.318	997
Forsinkelsesavgift	18.738	18.552	25.476
Tilleggsskatt	22.400	22.140	23.046
Trygdeavgift lav sats	1.025.844	1.005.698	1.067.248
Trygdeavgift mellom sats	2.182.944	2.159.742	2.169.987
Trygdeavgift høy sats	163.380	161.654	167.682
Skattebegrensning § 78	403.663	395.842	450.261
Forsørgerfradrag	1.015.998	1.003.494	977.058
Nettoinntekt stat	3.187.939	3.144.816	3.243.176

Tabell 3.14 viser ikke den samme tendensen som tabell 3.13. Det estimerte antallet synker når det sammensatte utvalget justeres for alder. Dette stemmer overens med at populasjonstallet justeres ned ved aldersjustering.

På tilsvarende måte som det ble gjort for IFN og næringsdelen av IF, er det foretatt en estimering av nettoinntekt og antall fordelt på skatteklasser. Resultatene er vist i de to neste tabellene.

Tabell 3.15. Antall skattytere fordelt på klasser

Skatteklasser	Sammen- satt utv.	Aldersjust.- sammens. utv.	Lignings- registeret
1	2.806.170	2.768.284	3.065.297
2	364.845	359.879	398.697

Tabell 3.16. Nettoinntekt og toppskattegrunnlag fordelt på skattegrunnlag og inntektsintervaller

Klasse	Intervall	Sammen- satt utv.		Aldersjust.- sms. utv.		Lignings- registeret	
		Ant.	Beløp	Ant.	Beløp	Ant.	Beløp
1							
Nettoinntekt							
	00 - 65	2.806	157.9	2.768	155.8	3.065	159.6
	65 - 130	1.926	95.2	1.902	94.1	1.924	94.6
	130 - 164	999	27.1	988	26.8	995	26.3
	164 - 207	588	17.7	582	17.5	570	17.1
	207 -	270	21.7	267	21.4	263	21.9
Toppskattgr.lag							
	00 - 207	2.833	346.0	2.795	341.8	3.065	346.2
	207 -	541	42.9	536	42.4	533	40.6
2							
Nettoinntekt							
	00 - 81	365	22.7	360	22.4	399	22.9
	81 - 162	210	12.6	207	12.5	219	12.9
	162 - 191	101	2.3	100	2.3	98	2.4
	191 - 216	66	1.3	65	1.3	67	1.4
	216 -	46	6.5	45	6.4	49	6.1
Toppskattgr.lag							
	00 - 249	369	50.1	364	49.4	399	49.3
	249 -	57	7.4	56	7.3	58	7.3

Mønsteret er det samme i alle tabeller i dette avsnittet. Ved å aldersjustere tilslutt i det sammensatte utvalget blir tallene lavere og ligger jevnt over litt lavere enn registerets tall. I tabell 3.15 fremgår det at i skatteklasse 1 er det en skjevhet i utvalget som gir et for lavt antall og beløp i inntektsintervallet 0 - 65 000 og tilsvarende for høyt antall og beløp i de høyere inntektsintervallene. I skatteklasse 2 er beløpene i det aldersjusterte utvalget likt verdiene i registeret, men fortsatt er antallet for lavt.

3.5. Kalibreringsvariable

Hittil er utvalget kun etterstratifisert etter alder på personnivå. For mindre justeringer av utvalget, kan en benytte etterstratifisering etter alder på husholdningsnivå kombinert med bruk av inntektsvariable. Metoden som brukes er beskrevet i Haldal(1992).

Opplegget er det samme som i foregående avsnitt. Først etterstratifiseres IF-utvalget etter alder på husholdningsnivå, og så kombineres dette utvalget med IFN-utvalget. Deretter etterstratifiseres dette sammensatte utvalget på nytt etter alder og følgende inntektsvariabler og formueskatt:

- Pensjonsgivende inntekt, lav sats (lønn)
- Pensjonsgivende inntekt jordbruk, skog og fiske
- Pensjonsgivende inntekt høy sats(annen næringsinntekt)
- Trygdegrunlaget
- Formueskatt stat*

* På grunn av manglende datagrunnlag er skatt brukt istedet for formuen.

I tabellene i dette avsnittet sammenlignes tallene fra ligningsregisteret med tallene fra det aldersjusterte sammensatte utvalget og fra det sammensatte utvalget som er aldersjustert og kalibrert mot de fem kalibreringsvariablene fra ligningsregisteret.

Tabell 3.17. Oppblåste totalbeløp for sammensatt aldersjustert utvalg, sammensatt aldersjustert og kalibrert utvalg sammenlignet med ligningsregisterets totaltall.

Variabel	Kun aldersj.	Aldersjust.- og kalibrert	Ligningsregisteret
Pensjonsgiv. innt. lav sats	333.576	332.137	332.137
Pensjonsgiv. innt. jord/skog/fiske	9.844	10.295	10.295
Pensjonsgiv. innt. høy sats	19.173	19.437	19.437
Trygdegrunnlag	81.312	83.475	83.475
Grunnlag for beregning av toppskatt	442.307	443.500	443.448
Formueskatt kommune	2.501	2.752	2.606
Formueskatt stat	1.382	1.661	1.661
Inntektskatt kommune	58.249	58.726	58.173
Inntektsskatt stat	8.979	9.281	9.037
Fellesskatt	16.178	16.319	16.308
Toppskatt stat	4.841	4.929	4.638
Skatt på betydelig aksjesalg	0.508*	0.897	0.508
Forsinkelsesavgift	0.157	0.159	0.023
Tilleggsskatt	0.244	0.254	0.273
Trygdeavgift lav sats	1.294	1.329	1.326
Trygdeavgift mellom sats	26.505	26.418	26.433
Trygdeavgift høy sats	2.194	2.209	2.328
Skattebegrensning § 78	1.377	1.363	1.477
Forsørgerfradrag	2.003	2.003	1.951
Nettoinntekt stat	361.857	364.493	365.323

*For variabelen skatt på betydelig aksjesalg er det en dramatisk økning av oppblåst beløp, det vil si vektene øker kraftig for disse personene.

På helt tilsvarende vis kan antall bak de forskjellige summene i tabell 3.17 beregnes og deretter sammenlignes med ligningsregisterets verdier.

Tabell 3.18. Oppblåst antall basert på sammensatt aldersjustert utvalg, sammensatt aldersjustert og kalibrert utvalg sammenlignet med ligningsregisterets totaltall.

Variabel	Kun aldersj.	Aldersjust.- og kalibrert	Ligningsregisteret
Pensjonsgiv. innt. lav sats	2.356.958	2.353.211	2.379.559
Pensjonsg. innt. jord/skog/fiske	137.381	140.787	141.591
Pensjonsgiv. innt. høy sats	180.428	181.356	180.572
Trygdegrunnlag	1.060.091	1.066.553	1.140.339
Grunnlag for beregn av toppskatt	3.162.096	3.161.831	3.274.144
Formueskatt stat	750.020	760.611	745.000
Inntektskatt kommune	2.849.737	2.849.721	2.935.000
Inntektsskatt stat	1.237.084	1.242.734	1.259.000
Fellesskatt	2.887.129	2.887.345	2.985.790
Toppskatt stat	626.330	627.074	631.104
Skatt på betydelig aksjesalg	1.269	1.459	997
Forsinkelsesavgift	19.244	19.284	25.476
Tilleggsskatt	22.053	22.141	23.046
Trygdeavgift lav sats	1.001.865	1.008.220	1.067.248
Trygdeavgift mellom sats	2.164.152	2.160.788	2.169.987
Trygdeavgift høy sats	164.464	165.001	167.682
Skattebegrensning § 78	391.297	387.028	450.261
Forsørgerfradrag	997.526	997.219	977.058
Nettoinntekt stat	3.147.702	3.147.744	3.243.176

På tilsvarende måte som tidligere kan estimert antall skatteyttere og estimert totalbeløp fordelt på skatteklassene finnes.

Tabell 3.19. Antall skattytere fordelt på klasser

Skatteklasser	Kun aldersj.	Aldersjust.- og kalibrert	Ligningsregisteret
1	2.772.967	2.772.651	3.065.297
2	358.384	358.827	398.697

Tabell 3.20. Nettoinntekt og toppskattgrunnlag fordelt på skattegrunnlag og inntektsintervaller

Klasse	Intervall	Kun aldersj.		Aldersjust.- og kalibrert		Ligningsregisteret	
		Ant.	Beløp	Ant.	Beløp	Ant.	Beløp
1							
Nettoinntekt							
	00 - 65	2.773	156.0	2.768	156.1	3.065	159.6
	65 - 130	1.907	94.4	1.902	94.7	1.924	94.6
	130 - 164	993	26.9	988	27.0	995	26.3
	164 - 207	584	17.6	582	17.8	570	17.1
	207 -	269	21.6	267	22.6	263	21.9
Toppskattgrunnlag							
	00 - 207	2.799	343.0	2.795	342.7	3.065	346.2
	207 -	539	42.9	536	43.2	533	40.6
2							
Nettoinntekt							
	00 - 81	358	22.4	360	22.4	399	22.9
	81 - 162	208	12.5	207	12.5	219	12.9
	162 - 191	100	2.3	100	2.3	98	2.4
	191 - 216	66	1.3	65	1.4	67	1.4
	216 -	46	6.5	45	7.4*	49	6.1
Toppskattgrunnlag							
	00 - 249	362	49.4	364	49.5	399	49.3
	249 -	56	7.4	56	8.1*	58	7.3

*For nettoinntekter over 216 000 og toppskattgrunnlag over 249 000 i klasse 2 er det et betydelig avvik som oppstår på grunn av kalibreringen.

3.6. Diskusjon

Tabellene i de foregående avsnittene viser at mye av forskjellen blir rettet opp ved å etterstratifisere utvalget med hensyn på alder. Etterstratifisering med hensyn på noen ligningsvariable ga mindre utslag, men resultatene ble bedre. Det er også gjort forsøk på å benytte intervallfordelingen til nettoinntekt og toppskattgrunnlaget som kalibreringsvariable, men det oppsto da sammenbrudd i vektene i form av negative verdier. Årsaken til at dette fører galt avsted kan være en for sterk vridning av utvalget for å få betingelsene oppfylt.

Det fremgår av tabell 3.20 at det laveste inntektsintervallet estimeres for lavt både med hensyn på antall og beløp, mens i de høyere intervallene ligger de estimerte verdiene jevnt over for høyt. Dette medfører at for alle personer med inntekter over laveste intervall skal vekten både opp og ned for å få betingelsene oppfylt. Det kan være at en kom lenger ved en svakere betingelse, nemlig at samlet beløp til skattyterne i hvert intervall skulle være riktig istedet for at samlet beløp totalt i hvert intervall uavhengig av skattyternes nettoinntekt skal stemme. Da ville hver person kun få en betingelse istedet for inntil fem.

Tabell 3.20 viser også at ved kalibrering estimeres nettoinntektene for lavt i de fleste intervaller. Dette kan skyldes at kalibreringen med hensyn på ligningsvariablene skaper en skjevhet i utvalget i forhold til populasjonen.

4. IF/IFN 1992

4.1 Utvalg

Inntekts- og formuesundersøkelsene i 1992 er i likhet med tilsvarende undersøkelse for 1991, satt sammen av utvalg trukket etter forskjellige utvalgsplaner. Undersøkelsen kan deles i to deler, en for husholdninger og en for selvstendig næringsdrivende. Utvalget til den ordinære inntekts- og formuesundersøkelsen (IF92) er det fortsatt naturlig å dele videre inn i tre deler; leveårsutvalget og paneldelene av forbruksutvalget som var med første gang i h.h.v. 1991 og 1992. For hvert av utvalgene blir husholdningsmedlemmene til IO inkludert i utvalget.

Til inntekts- og formuesundersøkelsen for næringsdrivende (IFN92) er det trukket et utvalg blant de som var registrert som selvstendig næringsdrivende i 1990. Fremgangsmåten ved trekking av dette utvalget er beskrevet i kapittel 2.1. Utvalget er nesten identisk med utvalget til IF91/IFN91.

De ulike utvalgene er koblet til ligningsregisteret for 1992 for å få verdier for endel inntekts- og formuesvariable. Antallet innen de ulike utvalgene fordeler seg som vist i tabellen nedenfor.

Tabell 4.1. Antall husholdninger og personer i de ulike utvalgene.

Utvalg	Husholdninger	Personer
IF, i alt	3.508	9.361
Levekår	2.539	6.919
Forbruk 92	492	1.228
Forbruk 91	477	1.214
IFN	4.555	14.752
Totalt	8.063	24.113

IFN-utvalget er basert på Folketellingen i 1990 (FOB90). Det medfører at de som er nyetablerte som selvstendig næringsdrivende etter 1990 ikke har hatt mulighet til å bli trukket ut i utvalget. På grunn av dette kan det bli en skjev fordeling av antall næringsdrivende i forhold til det virkelige antallet i 1992.

Tidligere IF-undersøkelser har vist at det er av avgjørende betydning at aldersfordelingen i utvalget er riktig. Da det ikke kan tas direkte hensyn til aldersfordelingen når utvalget trekkes, løses problemet istedet ved å etterstratifisere vektene etter antall innen ulike aldersgrupper. I forhold til undersøkelsen i 1991 er det valgt å benytte færre og lengre intervall for de over 30 år. Dette kan begrunnes med at det er liten forskjell i yrkesaktiviteten til en person i aldersgruppen 30-34 år og en person i gruppen 35-39 år, o.s.v. Aldersinndelingen for de under 30 er den samme som tidligere. For 1991 ble det korrigert for antall pensjonister som var på institusjon. Dette medførte bl.a. at trygdene ble dårlig estimert. For å unngå denne feilestimeringen er det for 1992 ikke gjort noen korrigerings. Selv om de som er på institusjon ikke har mulighet til å komme med i noen av delutvalgene, vil det skape større skjevhet i totalene å utelate dem. Dette kan imidlertid føre til en økning i antall trygdede fra 1991 til 1992.

Til tross for at Forbruk 91 skulle være trukket som et familieutvalg viste det seg ved å anta at også dette utvalget var et personutvalg at det ble oppnådd totaler som lå klart nærmere verdiene fra befolkningsregisteret. Alle tre delutvalgene er derfor behandlet som personutvalg. Benyttes inverse trekk sannsynligheter som vekter for personene i utvalget gir det da oppblåsningsfaktorer som er omvendt proporsjonale med antall personer over 15 år i husholdningen. Delutvalgene i IF92 får dermed følgende aldersfordeling.

Tabell 4.2. Aldersfordeling i utvalgene basert på trekksannsynlighetene.

Alder	Levekår	Forbruk		Befolkningen 01.01.93
		Panel 91	Panel 92	
00-06	417.434	497.430	438.863	407.968
07-12	319.160	362.298	329.686	311.335
13-15	165.764	135.065	200.299	157.448
16-19	227.436	130.378	150.553	235.745
20-24	347.759	227.968	175.784	336.171
25-29	298.661	371.457	370.302	336.613
30-39	596.736	772.942	716.531	633.125
40-49	642.386	510.762	518.744	602.840
50-59	394.063	384.624	385.769	392.598
60-66	290.561	346.067	301.943	265.551
67 +	628.858	681.701	852.614	619.773
Totalt	4.328.818	4.420.687	4.441.488	4.299.167

Av tabell 4.2 fremgår det at alderstotalene blir rimelig godt estimert, der resultatene fra levekårsutvalget ligger nærmest de virkelige verdiene. Studeres derimot de estimerte totalene i de enkelte aldersklassene nærmere, viser det store avvik fra verdiene fra befolkningsregisteret. Spesielt er avviket stort når data fra de to forbruksutvalgene benyttes. Dette viser at aldersfordelingen til dels er ganske skjev i utvalget og en etterstratifisering etter alder vil kunne slå gunstig ut på estimeringen av ligningstotalene.

4.2. Aldersjustering av IF-utvalget

For å rette opp skjevheten i aldersfordelingen i utvalgene er ligningstotalene forsøkt estimert ved å benytte aldersjusterte vektorer. Korrigeringen skjer direkte på husholdningsnivå. Siden det er naturlig å se hele IF under ett, veies vektene sammen for å finne felles vektorer for IF-utvalget.

$$(4.1) \quad w_{IF} = \frac{6919 * I_L * w_L + 1228 * I_{92} * w_{92} + 1214 * I_{91} * w_{91}}{9361}$$

Her er w_L , w_{92} , w_{91} vektorer beregnet enten som inverse trekksannsynligheter eller ved kalibrering etter alder. I_L , I_{92} , I_{91} er indikatorfunksjoner som er lik 1 hvis personen er med i det aktuelle utvalget og 0 ellers. Benyttes (4.1) som oppblåsningsfaktor, resulterer det i følgende estimerte totaler for hele IF-utvalget.

Tabell 4.3. Oppblåste totalbeløp fra IF-utvalget basert på inversetrekkssannsynligheter, aldersjusterte vektorer sammenlignet med ligningsregisterets totaltall.

Variabel	Inverse	Aldersjust.	Ligningsreg.
Netto formue kommune	406.969	383.417	389.824
Netto formue stat	399.481	376.384	382.504
Alminnelig inntekt	356.769	353.262	345.505
Alminnelig inntekt stat	272	284	659
Toppskattegrunnlag	474.127	471.709	464.010
Pers.innt. pensjoner (lav)	91.712	85.891	85.849
Pensjon fra delingsforetak	0	0	4
Pers.innt. lønn	350.836	354.315	346.279
Lønn fra delingsforetak	181	164	594
Pers.innt. jord/skog/fiske	7.505	7.484	8.440
Pers.innt. annen næring	24.012	23.843	22.835
Formueskatt kommune	3.122	2.936	2.991
Inntektskatt Kommune	60.640	60.086	58.423
Formueskatt stat	743	697	709
Inntektskatt stat	167	174	119
Fellesskatt	20.213	20.028	19.474
Toppskatt	7.954	7.997	7.626
Skatt på norsk aksjeutb.	368	364	530
Skatt på utenl. aksjeutb.	138	130	4
Forsinkelsesavgift	16	16	19
Tilleggsskatt	444	440	238
Trygdeavgift pensjoner (lav)	2.754	2.579	2.574
Trygdeavgift lønn (mellom)	27.125	27.388	26.744
Tr.avg. lønn fra delingsf.	14	12	46
Trygdeavgift næring (høy)	2.480	2.462	2.311
Trygdeavgift jord/skog/fiske	571	571	647
Forsørgerfradrag	2.023	1.981	1.954
Fradrag for BSU	85	89	98
Fradrag for AMS	44	43	40
Fradrag etter 65 % regelen	23	21	139
Skattebegrensing etter § 78	2.390	2.248	2.355
Nettoformue kommune s/a	434.521	410.516	417.729
Alminnelig inntekt kommune s/a	360.361	357.032	349.840
Netto formue stat s/a	434.521	410.516	417.669
Alminnelig inntekt stat s/a	272	284	733

Av tabell 4.3 kan en se at oppblåste beløp estimert ved bruk av inverse trekkssannsynligheter stort sett ligger høyere enn verdiene fra ligningsregisteret. Tallene som er funnet ved hjelp av aldersjusterte vektorer viser også noe avvik, men det er ingen entydig tendens til avvik i den ene eller andre retningen. Men i forhold til resultatene ved bruk av inverse trekkssannsynligheter er det prosentvise avviket fra registerverdiene mindre når aldersjusterte vektorer benyttes. Dette underbygger ideen om at aldersfordelingen i utvalget har betydning for estimering av totalene.

På samme vis kan estimert antall personer bak ligningstotalene sammenlignes. Resultatet er gitt i tabellen nedenfor.

Tabell 4.4. Oppblåste antall fra IF-utvalget basert på inverse trekksannsynligheter og aldersjusterte vektorer sammenlignet med ligningsregisterets totaltall.

Variabel	Inverse	Aldersjust.	Ligningsreg.
Netto formue kommune	1.103.368	1.050.332	1.079.043
Netto formue stat	1.014.691	966.909	991.137
Alminnelig inntekt	2.939.864	2.907.533	2.946.578
Alminnelig inntekt stat	1.926	2.025	7.853
Toppskattegrunnlag	3.255.590	3.241.632	3.296.261
Pers.innt. pensjoner (lav)	1.117.326	1.052.103	1.085.137
Pers.innt. lønn	2.369.917	2.404.682	2.397.606
Lønn fra delingsforetak	954	866	2.943
Pers.innt. jord/skog/fiske	101.142	100.602	105.492
Pers.innt. annen næring	165.843	165.832	160.990
Formueskatt kommune	1.108.319	1.055.481	1.083.590
Inntektskatt Kommune	2.939.864	2.907.533	2.946.661
Formueskatt stat	1.014.691	966.909	991.226
Inntektskatt stat	1.926	2.025	8.052
Fellesskatt	2.939.864	2.907.533	2.946.544
Toppskatt	792.745	791.461	771.995
Skatt på norsk aksjeutb.	200.311	192.855	189.994
Skatt på utenl. aksjeutb.	1.381	1.248	3.282
Forsinkelsesavgift	17.127	18.248	19.961
Tilleggsskatt	27.123	26.734	20.614
Trygdeavgift pensjoner (lav)	1.111.469	1.046.614	1.066.372
Trygdeavgift lønn (mellom)	2.185.846	2.207.794	2.183.946
Tr.avg. lønn fra delingsf.	954	866	2.920
Trygdeavgift næring (høy)	163.042	162.973	156.836
Trygdeavgift jord/skog/fiske	97.078	96.575	102.351
Forsørgerfradrag	1.000.648	986.181	985.461
Frdrag for BSU	48.793	51.984	56.330
Frdrag for AMS	61.837	61.463	60.652
Frdrag etter 65 % regelen	1.244	968	5.014
Skattebegrensing etter § 78	515.360	485.304	524.338
Nettoformue kommune s/a	1.719.943	1.669.614	1.738.030
Alminnelig inntekt kommune s/a	3.228.988	3.213.288	3.281.604
Netto formue stat s/a	1.719.943	1.669.614	1.737.905
Alminnelig inntekt stat s/a	1.926	2.025	10.394

I motsetning til resultatene i tabell 4.3 viser tabell 4.4 at de to estimeringsmetodene er positivt korrelerte. Noen av antallene ligger lavere enn registerverdiene og noen ligger høyere, men mønsteret er det samme for begge estimeringsmetodene. Det prosentvise avviket er imidlertid størst når en bruker inverse trekksannsynligheter som oppblåsningsfaktorer.

Studerer de tre delutvalgene hver for seg ser en den samme tendensen som ved estimering av alder. Data fra levkårsutvalget gir gode estimater, mens data fra de to forbruksutvalgene resulterer i verdier som ligger langt fra registerverdiene. Mye av dette kan forklares ved utvalgsstørrelsen, men det kan også tyde på at forbruksutvalgene er lite representative for den inntektsmottakende delen av befolkningen.

4.3. Kontroll av IFN-utvalget og næringsdelen av IF-utvalget

Som for undersøkelsen i 1991 er det interessant å sammenligne IFN-utvalget og næringsdelen av IF-utvalget. Strata A, B og C fra IFN-utvalget behandles sammen med den delen av IF-utvalget som hadde næringsinntekt i 1992 i tillegg til vanlig lønnsinntekt.

Når resten av husholdningen til IO skal inkluderes er næring ikke alltid kjent. De klassifiseres da etter metoden som er beskrevet i kapittel 2.2. Trekksannsynligheten til de næringsdrivende avhenger av

hvilket stratum de tilhører, mens trekksannsynligheten til hele husholdningen kan defineres som summen av trekksannsynlighetene til de selvstendige i husholdningen, inkludert IO.

IF-utvalget ble etterstratifisert etter alder slik at totalene innen hver aldersklasse skulle stemme. Dette ga som vist i tabellene 4.3 og 4.4 en bedre estimering av både totalbeløp og totalantall. Det er rimelig å tro at det også finnes en skjevhet i aldersfordelingen i IFN-utvalget. Etterstratifisering etter alder er imidlertid ikke like enkelt for IFN-utvalget siden antall innen hvert næringsstratum ikke er kjent fra noe register.

Næringsdelen av IF-utvalget er de som eventuelt i tillegg til vanlig lønnsinntekt har næringsinntekt. I IF-utvalget i 1992 var det 1945 personer i denne kategorien. Hvis en sammenligner estimert aldersfordeling for næringsdelen av IF og IFN med aldersfordelingen i populasjonen av næringshusholdninger i 1990, vil det være avvik mellom utvalgsverdiene og registeret. Dette kan sees ut fra følgende tabell.

Tabell 4.5. Fordeling av personer etter alder i næringshusholdningene i 1990, basert på næringsutvalgene, og i populasjonen av næringsfamilier i 1990

Alder	IFN	IF (næring)	Populasjonen næringsfamilie 1990
00-06	63.600	66.787	60.655
07-12	64.720	64.243	65.213
13-15	36.388	34.095	35.670
16-19	56.787	54.842	56.062
20-24	66.313	55.447	58.362
25-29	50.760	40.191	43.820
30-39	113.013	105.045	110.921
40-49	135.016	130.905	130.007
50-59	91.129	88.995	85.403
60-66	53.820	50.186	48.923
67 +	69.974	51.250	52.418
Totalt	801.524	741.993	747.454

Tabell 4.5 viser at for de yngre aldersklasser er de estimerte tallene omtrent like for de to utvalgene, mens for IFN-utvalget er tallene høyere i aldersgruppene over 20 år. Noe av avviket kan forklares med at næringsdrivende som har etablert seg etter 1990 ikke er med i registeret. Tabell 4.6 viser også at en etterstratifisering etter alder kan være hensiktsmessig for IFN-utvalget, men det eksisterer ingen annen populasjon enn den fra FOB90, og avviket fra denne populasjonen i forhold til de virkelige verdiene vil øke etterhvert.

4.4. Sammenkobling av IF og IFN

Etter aldersjustering av IF-utvalgene kobles IF- og IFN-utvalgene sammen og felles vektor beregnes. En slik sammenkobling er kun aktuell blant de husholdningene som hadde næringsinntekt i 1990. I IF-utvalget var det 1945 personer som hørte til husholdninger der det var registrert næringsinntekt i 1990. Det gir da i tillegg til de 14752 fra IFN-utvalget totalt 16697 personer i næringshusholdningene. En felles vekt for IF og IFN kan dermed skrives

$$(4.2) \quad w_s = \frac{1945 * w_{IF} * I_{IF} + 14752 * w_{IFN} * I_{IFN}}{16697}$$

Her er w_{IF} , w_{IFN} vektor beregnet enten som inverse trekksannsynligheter eller aldersjusterte. I_{IF} , I_{IFN} er indikatorvariable som definert tidligere. For de husholdningene som kun hadde lønnsinntekt blir vekten for det sammensatte utvalget den samme. Det sammensatte utvalget gir en estimert aldersfordeling som vist i tabellen nedenfor. Vektene til IF-delen av det sammensatte utvalget er de

aldersjusterte vektene funnet tidligere, mens til IFN-delen av utvalget er de inverse trekksannsynlighetene brukt som oppblåsingsfaktorer.

Tabell 4.6. Fordeling av personer i ulike aldersgrupper etter sammenkobling av aldersjustert IF og IFN og populasjonen 01.01.93.

Alder	IF(aldersj)+IFN	Populasjonen
00-06	405.152	407.968
07-12	311.756	311.335
13-15	159.473	157.448
16-19	237.463	235.745
20-24	345.770	336.171
25-29	345.950	336.613
30-39	640.164	633.125
40-49	606.471	602.840
50-59	394.482	392.598
60-66	268.761	265.551
67 +	636.315	619.773
Totalt	4.351.763	4.299.167

Tabell 4.6 viser at til tross for at IF-utvalget er justert etter alder, er det fortsatt avvik mellom den estimerte aldersfordelingen og registerfordelingen. Av den grunn bør også det sammensatte utvalget etterstratifiseres etter alder.

4.5. Kalibreringsvariable

På samme vis som for 1991 ble det benyttet variable fra ligningsregisteret for en ytterligere justering av vektene. Følgende variable ble benyttet til denne etterstratifiseringen:

- Nettoformue stat som gir skatt.
- Personinntekt pensjoner (lav sats).
- Personinntekt lønn.
- Personinntekt jord, skog, fiske.
- Personinntekt annen næring.

Kalibreres det mot disse variablene i tillegg til antallet innen hver aldersgruppe fås estimerte totaler som vist i tabellen nedenfor. Til sammenligning er også resultatene for det sammensatte utvalget (IF (aldersjustert) +IFN) tatt med.

Tabell 4.7. Oppblåste totalbeløp for sammensatt aldersjustert utvalg, sammensatt aldersjustert og kalibrert utvalg, sammenlignet med ligningsregisterets totaltall.

Variabel	Kun aldersjust.	Aldersjust. og kalibrert	Ligningsregisteret
Netto formue kommune	369.766	389.161	389.824
Netto formue stat	362.852	382.504	382.504
Alminnelig inntekt	354.597	348.273	345.505
Alminnelig inntekt stat	288	275	659
Toppskattegrunnlag	476.018	463.783	464.010
Pers.innt. pensjoner (lav)	87.038	85.849	85.849
Pers.innt. lønn	356.645	346.279	346.279
Lønn fra delingsforetak	297	291	594
Pers.innt. jord/skog/fiske	8.136	8.440	8.440
Pers.innt. annen næring	23.879	22.835	22.835
Formueskatt kommune	2.791	2.984	2.991
Inntektskatt Kommune	60.156	59.064	58.423
Formueskatt stat	653	709	709
Inntektskatt stat	17	16	119
Fellesskatt	20.053	19.688	19.474
Toppskatt	8.178	7.711	7.626
Skatt på norsk aksjeutb.	282	325	530
Skatt på utenl. aksjeutb.	0.6	0.6	4
Forsinkelsesavgift	19	18	19
Tilleggsskatt	336	319	238
Trygdeavgift pensjoner (lav)	2.614	2.578	2.574
Trygdeavgift lønn (mellom)	27.560	26.753	26.744
Tr.avg. lønn fra delingsf.	23	22	46
Trygdeavgift næring (høy)	2.447	2.358	2.311
Trygdeavgift jord/skog/fiske	624	648	647
Forsørgerfradrag	1.995	1.991	1.954
Fradrag for BSU	94	90	98
Fradrag for AMS	41	41	40
Fradrag etter 65 % regelen	19	25	139
Skattebegrensing etter § 78	2.335	2.279	2.355
Nettoformue kommune s/a	397.277	415.919	417.729
Alm. innt. kommune s/a	358.440	352.093	349.840
Netto formue stat s/a	397.277	415.919	417.669
Alminnelig inntekt stat s/a	288	275	733

Tabell 4.7 viser at når totalene tvinges til å stemme for noen av ligningsvariablene, vil også estimatene for de øvrige ligningstotalene ligge nærmere registerverdiene. I forhold til kun å korrigere for alder i IF-utvalget er forbedringen stor. Tilsvarende estimering kan gjøres på antall personer bak de registrerte verdiene for hver av ligningsvariablene. Resultatet er gitt i tabellen nedenfor.

Tabell 4.8. Oppblåst antall basert på sammensatt aldersjustert utvalg, sammensatt aldersjustert og kalibrert utvalg, sammenlignet med ligningsregisterets totaltall.

Variabel	Kun aldersjust.	Aldersjust. og kalibrert	Ligningsregisteret
Netto formue kommune	1.061.052	1.060.581	1.079.043
Netto formue stat	979.920	982.365	991.137
Alminnelig inntekt	2.948.543	2.897.272	2.946.578
Alminnelig inntekt stat	2.056	1.969	7.853
Toppskattegrunnlag	3.292.842	3.238.309	3.296.261
Pers.innt. pensjoner (lav)	1.072.867	1.058.104	1.085.137
Pers.innt. lønn	2.431.148	2.387.379	2.397.606
Lønn fra delingsforetak	1.455	1.442	2.943
Pers.innt. jord/skog/fiske	104.709	108.097	105.492
Pers.innt. annen næring	168.690	167.911	160.990
Formueskatt kommune	1.065.768	1.065.074	1.083.590
Inntektskatt Kommune	2.948.652	2.897.380	2.946.661
Formueskatt stat	979.920	982.365	991.226
Inntektskatt stat	2.056	1.969	8.052
Fellesskatt	2.948.819	2.897.540	2.946.544
Toppskatt	797.257	772.679	771.995
Skatt på norsk aksjeutb.	198.669	197.338	189.994
Skatt på utenl. aksjeutb.	1.727	1.958	3.282
Forsinkelsesavgift	19.602	19.256	19.961
Tilleggsskatt	24.831	24.522	20.614
Trygdeavgift pensjoner (lav)	1.065.366	1.050.591	1.066.372
Trygdeavgift lønn (mellom)	2.224.697	2.181.688	2.183.946
Tr.avg. lønn fra delingsf.	1.455	1.442	2.920
Trygdeavgift næring (høy)	164.813	163.931	156.836
Trygdeavgift jord/skog/fiske	101.411	104.681	102.351
Forsørgerfradrag	996.184	992.009	985.461
Frdrag for BSU	54.368	51.932	56.330
Frdrag for AMS	62.968	62.218	60.652
Frdrag etter 65 % regelen	2.089	2.347	5.014
Skattebegrensing etter § 78	502.768	491.059	524.338
Nettoformue kommune s/a	1.694.186	1.677.756	1.738.030
Alm. innt. kommune s/a	3.264.441	3.210.618	3.281.604
Netto formue stat s/a	1.694.186	1.677.756	1.737.905
Alminnelig inntekt stat s/a	2.056	1.969	10.394

I motsetning til resultatene for totalbeløpene i tabell 4.7 viser tabell 4.8 at for totalantallene oppnås det ikke en like stor forbedring ved å kalibrere etter de valgte ligningstotaler. Det er heller ingen entydig tendens til at estimatene funnet ved alderskalibrerte vektorer er de beste estimatene.

4.6. Fordelingen i utvalget

Aldersjustering og kalibrering etter enkelte ligningsvariable virker positivt inn på estimeringen av totalbeløpene. Med fornuftige valg av kalibreringsvariable oppnås den ønskede effekten av metoden, nemlig at totalene for de andre ligningsvariablene blir estimert bedre. Et spørsmål kan være om det skapes nye skjevheter i utvalget ved å kreve at totalsummene for noen variable skal være eksakte. En måte å få et innblikk i fordelingen i utvalget på, er å studere prosentilene ved hjelp av **Mean Tukey Difference Plot**. Dette er det aktuelt å gjøre for ligningsvariablene

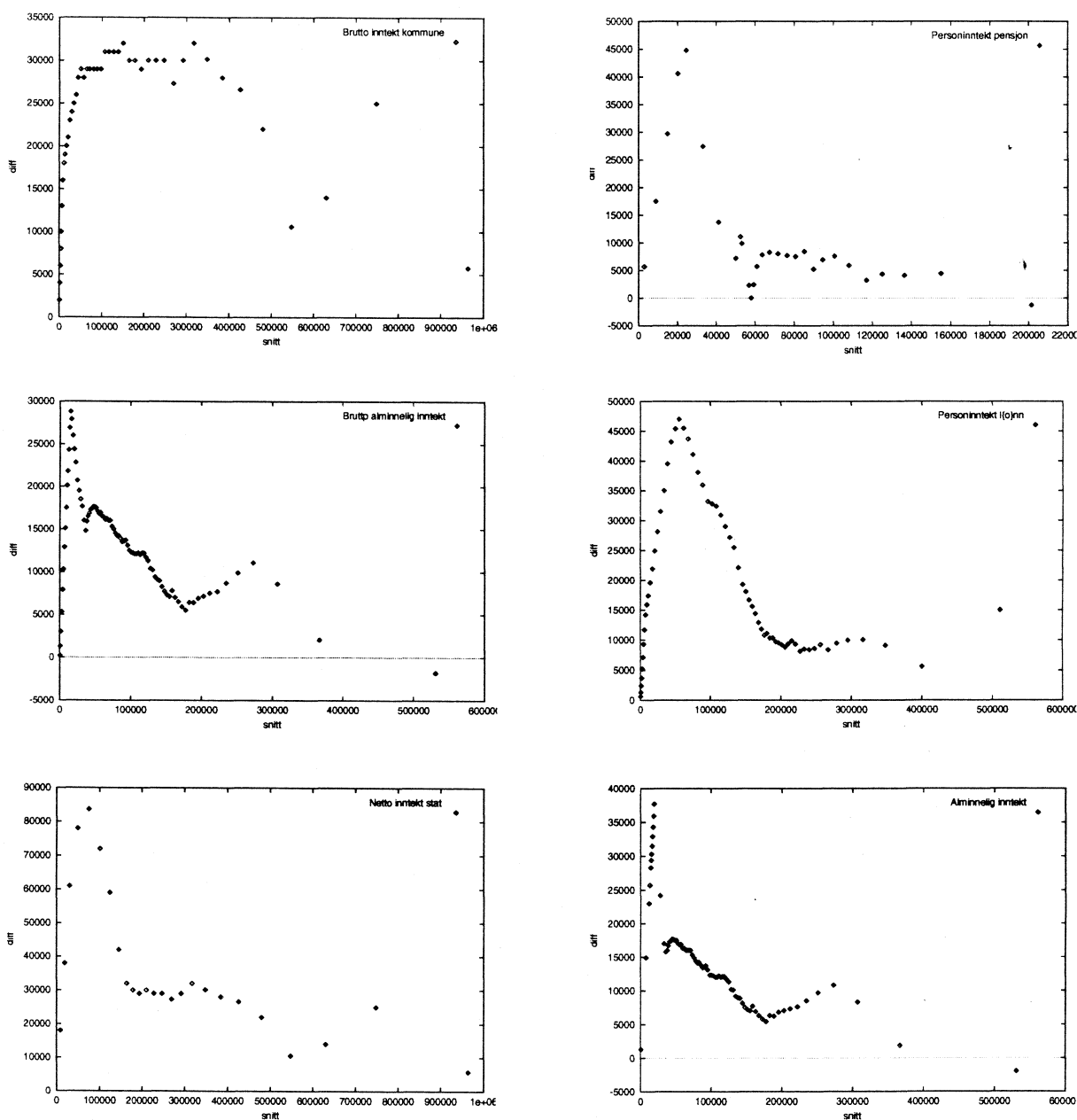
- Netto formue stat som gir skatt.
- Alminnelig inntekt stat som gir skatt.
- Personinntekt pensjoner.
- Personinntekt lønn.
- Netto formue kommune som på s/a.
- Alminnelig inntekt kommune som på s/a.

For hver av disse ligningsvariablene defineres r_i som i 'te prosentil i ligningsregisteret og u_i som i 'te prosentil i utvalget med hensyn på de kalibrerte vektene. Mean Tukey Difference kan da skrives som

$$(4.3) \quad t_i = \frac{r_i - u_i}{\frac{1}{2}(r_i + u_i)}$$

Plottes teller (4.3) mot nevner i samme uttrykk for $i=0.5, \dots, 99.5$ gir det et bilde av forholdet mellom prosentilverdiene i ligningsregisteret og i utvalget. Hvis verdiene primært er positive indikerer det at registerverdiene er høyere enn verdiene i utvalget og omvent.

Figur 4.1. Plott av prosentilene i utvalg og register for utvalgte ligningsvariable



Figur 4.1 viser tydelig at omtrent alle punktene har positive koordinater, og det betyr at prosentilene til registerverdiene ligger høyere enn tilsvarende prosentil i utvalget. Det er dermed rimelig å tro at det er en skjevhet mellom fordelingene til variablene i utvalget og i registeret. Prosentilene i utvalget er funnet på grunnlag av de kalibrerte vektene. Spørsmålet blir da hvor mye av denne skjevheten som er i det opprinnelige utvalget og hvor mye som skapes av kalibreringen. Det bør i det videre arbeidet undersøkes hvor stor effekt en slik skjevhet får inn på estimeringen av ligningstotalene, og hvilke metoder som eventuelt kan benyttes for å etterstratifisere slik at fordelingene blir mer sammenfallende.

For kalibreringsvariablene er som kjent vektene beregnet slik at de oppblåste totaltallene stemmer med ligningsregisterets verdier. Dette medfører at de positive og negative punktene i plottene i figur 4.1 skal oppveie hverandre. Det ser ikke ut til at dette er tilfelle for plottene av kalibreringsvariablene vist i figur 4.1. Årsaken til dette må være at det er endel informasjon i halen til de største ordnede observasjonene som ikke kommer med når prosentilene beregnes. Det vil derfor bli arbeidet videre med å studere effekter i halene til fordelingene.

4.7. Konklusjoner

Som for tilsvarende undersøkelse i 1991, gir en korrigerende for alder det største utslaget både på ligningsverdiene og ligningsantallene. Ved å tvinge totalene til noen av ligningsvariablene til å stemme rettes de mindre skjevhetene opp. Denne forbedringen ble best for ligningsbeløpene, mens effekten ikke ble like stor for antallet bak de ulike ligningsvariablene. Dette viser at hvis aldersfordelingen er riktig i utvalget er mulighetene for å få korrekt estimerte totaler størst. Mindre endringer tar kalibreringen seg av.

5. Diskusjon

5.1. Noen kritiske merknader til utvalgene i IF91 og IF92

For alle tre delutvalgene som ligger til grunn for IF-undersøkelsene i 1991 og 1992 ble det påvist en skjevhet i aldersfordelingen. Denne skjevheten ble rettet opp ved å etterstratifisere etter alder, men med den forutsetning at alle i husholdningen skulle få lik vekt. Denne korrigeringen gjorde at ligningstotalene ble bedre estimert enn ved kun å benytte inverse trekksannsynligheter som vekter. Det er rimelig å anta at også IFN-utvalget har en skjev aldersfordeling. Fordi det ikke finnes noen registerpopulasjon for næringsdrivende, bortsett fra FOB90, er det ikke mulig å korrigere for alder på samme måte som for IF-utvalgene.

Aldersjustering basert på husholdningen

Tabellene i kapittel 3 viser at det er liten forskjell mellom å aldersjustere på person- eller husholdnings-nivå. Når det aldersjusteres på husholdningsnivå kan en persons vekt øke (minke) selv om det er for mange (få) i personens aldersgruppe avhengig av de andre i husholdningens alder. Dette kan en imidlertid forsvares ved at det er for få (mange) husholdninger med denne alderssammensetningen.

Ligningsregisteret - hva er populasjonen?

De kjente totaltallene som er benyttet i estimeringen er kjent fra ligningsregisteret. Utvalget inneholder kun befolkningen utenom institusjoner, mens ligningsregisteret inneholder skatteyttere på institusjoner. Dette medfører at tallene ikke er helt sammenlignbare. Det er rimelig å tro at fordelingen til institusjonsbefolkningen er annerledes for mange ligningsvariable slik at skjevheter vil uansett bli et problem.

5.2. Arbeidsoppgaver fremover - forslag til forbedringer

Fordeling i utvalget

Som vist i figur 4.1 er det en skjevhet i utvalget i forhold til registeret. Videre ser det ut som det spesielt i halene kan skje ubehagelige ting. Det er derfor viktig å studere mer i detalj hvordan kalibreringen påvirker fordelingen i halene.

Flere kategoriske kalibreringsvariable?

Det er rimelig å stille spørsmål om det er riktig eller tilstrekkelig å kun korrigere for alder i utvalget. I tillegg kan også variable som kjønn og familiesammensettingen benyttes som kategoriske kalibreringsvariable. I FOB90 var det et hovedpoeng å bassere husholdnings-statistikken på en "kalibrering" mot registervariabelen familiestørrelse. I tillegg kan det være et alternativ å korrigere IF-utvalget for korrekte skatteklasser.

I stedet for å kalibrere etter ligningstotaler kan det være en ide å ta utgangspunkt i totalantall og totalbeløp innen ulike inntektsklasser. For IF vil det være naturlig å gruppere etter skatteklasse og inntektsintervall. Disse tallene er kjent fra ligningsregisteret slik at en korrigerende av vektene vil kunne skje direkte ved å benytte tilsvarende kalibreringsteknikk som for ligningstotalene.

For IFN vil det i tillegg til inntektsintervall være aktuelt å gruppere etter type næring. Imidlertid oppstår det et problem her da det ikke finnes noen registerpopulasjon etter næring. En løsning vil kunne være å utnytte informasjonen som er kjent fra FOB90 men da vil ikke de endringene som skjer etter 1990 komme med i denne populasjonen. Siden endringene i forhold til FOB90 er kjent for de som er med i utvalget vil endringene i populasjonen kunne estimeres ut fra disse opplysningene.

Et annet problem som kan oppstå, primært for IFN-utvalget, er at ved en så fin inndeling vil det være klasser/grupper som har ingen eller veldig få observasjoner. Det kan da være interessant å trekke inn metoder fra *Small Area Estimation* (se Thomsen og Spjøtvoll 1988). Ideen her er å utnytte verdier/informasjon fra større grupper som startverdier for en iterativ rutine for å finne totalverdier.

Aldersjustering av IFN

Den skjevheten i aldersfordelingen som er i IF-utvalget, er det naturlig å forvente at også finnes i IFN-utvalget. For å rette opp denne skjevheten kan det være en ide å bruke aldersfordelingen i næringsdelen av IF som fasitvariable, og etterstratifisere IFN-utvalget slik at det får samme aldersfordeling som dette utvalget. Imidlertid er det stor forskjell i antall observasjoner i de to utvalgene. (I 1992 var det 1945 i IF-utvalget som var med i næringshusholdninger, mens det i IFN-utvalget var 14752 observasjoner.) Metodisk sett er det derfor vanskelig å forsvare en stratifisering av et stort utvalg etter et mye mindre. Et annet alternativ kan være å gjøre bruk av resultatene fra FOB90, men de vil ikke fange opp de nyetablerte næringshusholdningene.

Det bør også arbeides videre med å forbedre registerpopulasjonen, da spesielt næringspopulasjonen.

Endringer 91-92

Som kjent ble skattereformen innført i 1992 og dermed er det ulike regler som ligger til grunn ved beregning av inntekts- og formuesvariable. For å kunne se effekten av skattereformen burde det kunne gjøres en analyse på endringene inntekts- og formuesnivå. Det må utredes hvilke felles vektorer som skal brukes.

EDB-siden

For å lette arbeidet i fremtidige inntekts- og formuesundersøkelser, bør det lages et meny (vindu) basert program, slik at det eneste arbeidet er å legge inn observasjonene og noen antallsvariable før både de aldersjusterte og de kalibrerte vektene beregnes. Dette problemet kan løses i SAS ved fsviw, men det er ingen ideell løsning. Muligens er en kombinasjon av SAS og Windows det beste. Fordi endel eksisterende programvare som benyttes er laget i SAS, er det videre arbeidet bundet til denne programvaren.

Denne problemstillingen har også generell interesse for bruk av kalibrering i andre undersøkelser.

Litteraturstudier og samarbeid.

Som en del av det videre arbeidet er det viktig å knytte kontakter opp mot andre miljøer som arbeider med de samme problemstillingene. Bl.a. er det aktive miljø innen kalibreringsteknikker både i Sverige og i Frankrike. I tillegg har Statistics Canada gjort forsøk med kalibrering. Det er viktig å sammenligne metodene og undersøke de erfaringer de forskjellige miljøene har med de ulike kalibreringsteknikkene.

6. Referanser

Heldal, Johan (1992): A Method for Calibration of Weights in Sample Surveys. Interne notater 92/3.

Heldal, Johan og Jon Ivar Røstadsand (1994): Hva kan inntektsundersøkelsen for husholdningene fortelle oss? Om bruk av utvalgsundersøkelser i nasjonalregnskapet. Økonomiske analyser 2/94.

NOU 93:17 (1993): Levekår i Norge.

Sparby, Grete (1994): Inntekts- og formuesundersøkelsen 1990. Dokumentasjon. Notater 94/8.

Spjøtvoll, Emil og Ib Thomsen (1988): Application of some Empirical Bayes Methods to Small Area Statistics. Reprint series no. 32, 1988.

Statistisk sentralbyrå

Oslo
Postboks 8131 Dep.
0033 Oslo

Tlf.: 22 86 45 00
Fax: 22 86 49 73

Kongsvinger
Postboks 1260
2201 Kongsvinger

Tlf.: 62 88 50 00
Fax. 62 88 50 30



Statistisk sentralbyrå
Statistics Norway