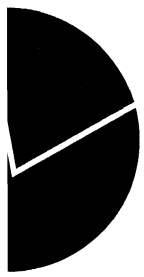


Ann Marit Kleive Holmøy

Dokumentasjonsrapport:
Beregning av vekter til Inntekts- og
formuesundersøkelsene 1993

Notater

STATISTISKE SENTRALBYRÅET



STATISTISKE SENTRALBYRÅET

INNHold

1. INNLEDNING	2
2. INNTEKTS- OG FORMUESUNDERSØKELSEN FOR PERSONER OG HUSHOLDNINGER (IF).....	2
3. INNTEKTS- OG FORMUESUNDERSØKELSEN FOR PERSONLIG NÆRINGSDRIVENDE (IFN).....	7
4. SAMVEIING AV IF- OG IFN-UTVALGENE	11
5. INNTEKTS- OG FORMUESUNDERSØKELSEN FOR SKATTEPLIKTIGE AKSJESELSKAP (IFS).....	16
6. KONKLUSJON	18
<i>Problemer</i>	18
<i>Forbedringer</i>	19
REFERANSER.....	21

1. Innledning

Beregning av vekter til inntekts- og formuesundersøkelsene er gjennomført separat for personer og husholdninger (IF) og for personlig næringsdrivende (IFN). I tillegg er det på grunnlag av filene til de to forannevnte undersøkelsene laget vekter til et samveid utvalg (IF/IFN). Denne tredelingen står i kontrast til tidligere år da det kun er blitt beregnet vekter til ett samveid utvalg. Erfaring og metoder kjent fra disse undersøkelsene er benyttet for å beregne vekter til inntekts- og formuesundersøkelsen for selskap (IFS). Vektberegningsteknikken som er benyttet for alle delutvalgene er kalibrering. For dokumentasjon av kalibrering vises det til Notat 94/23 og Notat 92/3.

Arbeidet med 1993-undersøkelsene kom i gang tidlig i forhold til tidligere år, men likevel ble tallene ferdige forholdsvis sent. På første arbeidsmøte 14. november 1994 ble det lagt opp til at utvalgene skulle være ferdig vektet innen første halvdel av 1995. Dette for å ha god tid til kontroll, fremskrivninger og tabellkjøringer. På grunn av for sen start, diverse komplikasjoner og manglende registrerte skjema, ble ikke filen med vektgrunnlaget levert før medio juni. Ferdig beregnede vekter ble levert i slutten av juni. Kommentarer og innspill kom på grunn av ferieavvikling sent tilbake, og dette førte, sammen med feil populasjon i IFN, til at de endelige vektene til IF og IFN ikke ble klare før i slutten av september. IF- og IFN-statistikken ble publisert i uke 39, og IFS-statistikken i uke 42. Vektene til IF/IFN ble klare på et tidligere tidspunkt, men de endelige vektene kunne ikke lages til før situasjonen i IF og IFN var avklart med tanke på kalibreringsvariable, ekstremobservasjoner o.l.

2. Inntekts- og formuesundersøkelsen for personer og husholdninger (IF)

Utvalget til undersøkelsen er satt sammen av utvalgene til leveårsundersøkelsen og paneldelen 1. år og 2. år i forbruksundersøkelsen. På grunnlag av intervju er kosthusholdningene til de uttrukne personene kartlagt, og husholdningsmedlemmene inkludert i utvalget. I de tilfeller husholdning p.g.a. frafall ikke kan kartlegges er familie satt lik husholdning. Personer og husholdninger fordeler seg på de ulike delutvalgene som vist i tabellen nedenfor.

Tabell 1. Fordeling av husholdninger og personer etter utvalg

Delutvalg	Husholdninger (IO)	Personer inkludert IO
Levekår	2.535	7.128
Forbruk 1. år	494	1.213
Forbruk 2. år	493	1.510
IF totalt	3.522	9.851

Av de 9851 personene ble 14 personer fra i alt 7 husholdninger betraktet som ekstremer. Disse ble gitt vekt=1 og tatt ut av de videre vektberegningene. En person er å betrakte som ekstrem hvis personens veide andel av den veide totalen blir for stor. For hver person beregnes

$$\hat{t}_{ij} = v_i * t_{ij(s)} \quad (1)$$

der v_i er person i sin vekt, $t_{ij(s)}$ er person i sin observasjon av variabel j i utvalget og \hat{t}_{ij} er person i sin andel til den veide totalen

$$\hat{t}_j = \sum_s v_i * t_{ij(s)} = \sum_s \hat{t}_{ij} \quad (2)$$

Hvis

$$a_{ij} = \frac{\hat{t}_{ij}}{\hat{t}_j} * 100\% \quad (3)$$

overskrider f.eks 20% er det rimelig å tro at person i har en ekstrem observasjon av variabel j i utvalget. Grensen for overskridelse kan variere for de ulike ligningsvariablene avhengig av antall observasjoner i utvalget og viktigheten av variabelen. Det er derfor viktig at det foretas en skjønnsmessig vurdering parallelt.

Utgangspunktet for beregning av vekter er trekkesannsynligheter definert som forholdet mellom antall uttrukne personer og antall trekkbare personer i utvalget. Basert på inverse trekkesannsynligheter har delutvalgene i IF en estimert aldersfordeling som vist i tabell 2.

Tabell 2. Estimert aldersfordeling i delutvalgene til IF basert på inverse trekkesannsynligheter

Aldersgruppe	Levekår	%-vist avvik	Forbruk 1. år	%-vist avvik	Forbruk 2. år	%-vist avvik	Populasjonen 31.12.1995
0 - 6 år	445.403	9.4	490.893	17.9	685.973	64.7	416.312
7 - 12 år	305.109	-2.8	409.078	30.3	487.325	55.0	313.914
13 - 15 år	158.067	-0.17	101.101	-36.1	196.324	23.9	158.335
16 - 19 år	227.012	0.32	124.126	-45.1	233.498	3.19	226.276
20 - 24 år	341.928	3.2	196.007	-40.8	299.017	-9.7	331.265
25 - 29 år	346.359	4.5	326.327	-4.6	382.077	11.7	341.955
30 - 39 år	581.685	-8.7	812.896	27.5	847.911	33.0	637.277
40 - 49 år	637.497	3.9	496.621	-19.0	671.220	9.4	613.468
50 - 59 år	409.759	0.72	416.675	2.42	352.106	-13.4	406.816
60 - 66 år	262.384	1.52	303.302	17.3	206.663	-42.3	258.446
67 + år	634.037	2.14	697.186	12.3	428.080	-31.0	620.751
Totalt	4.349.238	0.56	4.374.211	1.14	4.790.194	10.76	4.324.815

Tabellen viser at aldersfordelingen blir forholdsvis bra estimert i levekårsutvalget, mens det i forbruksutvalget 1. år og i forbruksutvalget 2. år er et betydelig avvik mellom estimerte tall og tall hentet fra registeret. Spesielt er avviket i yngste og eldste aldersgruppen stort. Tabellen avspeiler svakheten ved forbruksutvalgene, som til tross for et neglisjerbart frafall har en meget skjev aldersfordeling. For å rette opp den skjeve aldersfordelingen er de tre delutvalgene korrigert for alder.

I tillegg til aldersjustering er det ved beregning av vektene kalibrert med hensyn på totalbeløpene og totalantallene til ligningsvariablene

- netto formue stat
- personinntekt pensjoner
- personinntekt lønn
- personinntekt jord/skog/fiske
- personinntekt annen næring

Å kalibrere med hensyn på disse variablene vil si at vektene beregnes slik at totalbeløpene og totalantallene for disse variablene blir korrekt estimert. Nytt i årets undersøkelse er å kalibrere med hensyn på antall med positive observasjoner på de ulike ligningsvariablene. Undersøkelser i forbindelse med IF-92 viste at andel personer med observert verdi lik 0 på de ulike ligningsvariablene er høyere i

utvalget enn i registeret. Dette vil medføre en skjevhet i fordelingen mellom utvalg og register. Ved å gi de personene dette gjelder en dempet vekt, vil konsistensen i fordeling mellom utvalg og register forbedres betraktelig.

De tre delutvalgene ble hver for seg kalibrert som forklart, og deretter veid sammen. Ekstremutvalget ble inkludert i utvalget før de endelige totaltallene ble estimert.

Resultatene for IF-93 er vist i tabellene nedenfor.

Tabell 3. Estimerte totalbeløp for ligningsvariable i IF, sammenlignet med ligningsregisteret. Tall i millioner. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte vekter	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	422.638	421.982	0.15
Netto formue stat	414.877	414.877	0
Alminnelig inntekt	378.874	375.147	0.99
Toppskattegrunnlaget	482.703	481.964	0.153
Personinntekt pensjoner	90.618	90.618	0
Personinntekt lønn	358.850	358.850	0
Personinntekt jord/skog/fiske	7.776	7.776	0
Personinntekt annen næring	25.387	25.387	0
Skatt formue kommune	3.284	3.294	-0.31
Skatt inntekt kommune	64.905	64.144	1.18
Skatt formue stat	793	795	-0.25
Fellesskatt	21.635	21.371	1.23
Toppskatt	8.282	8.529	-2.89
Forsinkelsesavgift	18	28	-35.1
Tilleggsskatt	261	240	8.75
Skatt på etterbetalt pensjon	122	121	0.82
Trygdeavgift pensjon	2.733	2.713	0.74
Trygdeavgift lønn	27.620	27.602	0.065
Trygdeavgift næring opp til 12G	2.546	2.318	9.84
Trygdeavgift næring over 12G	109	146	-25.34
Trygdeavgift jord/skog/fiske	598	596	0.33
Forsørgerfradrag barn+ungdom	1.980	1.958	1.12
Finnmarksfradrag	408	449	-9.13
Fradrag BSU	147	134	9.70
Fradrag AMS	120	110	9.09
Skattebegrensning §78	2.276	2.390	-4.77
Anvendt godtgjørslfradrag	2.672	2.784	-4.02
Netto formue kommune som SA	447.946	449.587	-0.36
Alminnelig innt. kommune SA	382.278	379.406	0.75
Netto formue stat SA	447.946	449.570	-0.36

Av tabell 3 fremgår det at de estimerte totalbeløpene for viktige variable som formue, skatter og avgifter ligger nært opp til ligningsregisterverdiene. I tillegg blir også som en følge av vektbergningsmetoden, kalibreringsvariablene korrekt estimert. For øvrige ligningsvariable er avvikene større, noe som kan skyldes at antall observasjoner i utvalget av disse variablene er lavt. Hvorvidt det er over- eller

underestimering varierer for de ulike variablene, men der det er overestimering er denne forholdsvis lav, mens graden av underestimering er noe større.

Viktig bl.a. ved beregning av gjennomsnittstall, er antall personer med positive observasjoner som ligger til grunn for totalbeløpene. Tabell 4 viser hvordan de ulike antallene ble estimert.

Tabell 4. Estimert antall positive observasjoner for hver av ligningsvariablene i IF, sammenlignet med ligningsregisteret. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte vekter	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	1.108.393	1.108.536	-0.013
Netto formue stat	1.021.956	1.021.956	0
Alminnelig inntekt	2.963.885	2.954.171	0.33
Toppskattegrunnlaget	3.274.814	3.326.155	-1.54
Personinntekt pensjoner	1.099.643	1.099.643	0
Personinntekt lønn	2.420.563	2.420.563	0
Personinntekt jord/skog/fiske	98.475	98.475	0
Personinntekt annen næring	174.863	174.863	0
Skatt formue kommune	1.112.425	1.112.950	-0.05
Skatt inntekt kommune	2.963.885	2.984.270	-0.68
Skatt formue stat	1.021.956	1.022.003	-0.0046
Fellesskatt	2.963885	2.984.000	-0.67
Toppskatt	804.313	791.128	1.67
Forsinkelsesavgift	19.275	23.287	-17.2
Tilleggsskatt	26.893	22.726	18.0
Skatt på etterbetalt pensjon	11.714	9.488	23.1
Trygdeavgift pensjon	1.085.072	1.081.913	0.29
Trygdeavgift lønn	2.216.826	2.202.355	0.66
Trygdeavgift næring opp til 12G	170.567	170.158	0.24
Trygdeavgift næring over 12G	11.051	7.777	42.1
Trygdeavgift jord/skog/fiske	96.825	95.447	1.44
Forsørgerfradrag barn+ungdom	968.901	986.498	-1.78
Finnmarksfradrag	56.996	62.926	-9.43
Fradrag BSU	81.327	77.264	5.26
Fradrag AMS	125.905	123.277	2.12
Fradrag for etterbetalt pensjon	3.919	3.117	25.0
Skattebegrensning §78	497.965	525.544	-5.25
Anvendt godtgjørslfradrag	259.621	251.346	3.29
Netto formue kommune som SA	1.748.022	1.778.477	-1.71
Alminnelig innt. kommune SA	3.257.341	3.318.878	-1.85
Netto formue stat SA	1.748.022	1.778.416	-1.71

Tabell 4 viser tilsvarende som i tabell 3 at antall personer bak de ulike ligningsvariablene blir forholdsvis godt estimert. Avviket på trygdeavgiftene for næringsdrivende og noen av fradragene er spesielt merkbart. Dette er variable hvor det er få observasjoner i populasjonen og dermed også få observasjoner i utvalget. Hvorvidt det er over/underestimering varierer for de ulike ligningsvariablene, men det er en større grad av overestimering enn tilfellet var med de estimerte totalbeløpene.

I og med at IF er husholdningsstatistikk er fordeling av husholdninger etter størrelse interessant, spesielt ved sammenligning fra år til år. Tabell 5 viser det estimerte fordelingen basert på IF-utvalget og de ligningskalibrerte vektene.

Tabell 5. Estimert husholdningsfordeling IF

Antall personer i husholdningen	Antall husholdninger
1	745.518
2	500.865
3	272.121
4	254.980
5	112.834
6	21.555
7	3.283
8	2.248
9	571
11	150
Totalt	1.914.130

For å få et anslag på variasjonen i materialet er det beregnet 95 % konfidensintervall for noen viktige totaler. Disse intervallene er beregnet ved bruk av kryssvalidering og bootstrap. Metoden gir et bilde av utvalgsusikkerheten, og er beskrevet i Aasness & Co. Tabellen under viser estimerte totaler og intervallgrenser i tillegg til relativt konfidensintervall definert ved forholdet mellom konfidensgrensene og gjennomsnittsverdien. Konfidensintervallene kan tenkes på som et mål på hvor mye gjennomsnitt- og totaltall vil variere når i tenker oss at vi beregner tall ut fra ulike utvalg fra samme populasjon.

Tabell 6. Estimert utvalgsusikkerhet IF

Variabel	95 % KI for gjennomsnittet		95 % KI for totalen Tall i millioner		Relativt KI
	Gjennom- snittsverdi	Konfidens- grenser	Total- verdi	Konfidens- grenser	
Netto formue stat	95.932	1.281	414.877	5.542	0.01336
Pers.innt. pensjoner	20.952	91	90.618	395	0.00436
Pers.innt lønn	82.976	274	358.850	1.186	0.00331
Pers.innt. jord/skog/fiske	1.798	343	7.776	1.486	0.19120
Pers.innt. annen næring	5.869	360	25.387	1.558	0.06139
Toppskattegrunnlag	111.613	2.420	481.964	10.469	0.02169
Alminnelig inntekt	87.606	1.231	375.147	5.324	0.01405
Toppskatt	1.915	992	8.529	4.294	0.51843
Fellesskatt	5.002	1.744	21.371	7.546	0.34880
Tr.avg. pensjoner	632	116	2.713	5.042	0.18447
Tr.avg. lønn	6.386	172	27.602	747	0.02707
Tr.avg næring	588	29	2.318	1.267	0.04980
Tr.avg. næring over 12G	25	28	146	122	1.11419
Tr.avg. jord/skog/fiske	138	92	596	402	0.67219
Anvendt godtgjørlefradrag	618	12	2.784	55	0.02060

Tabell 6 viser at de fleste personinntektene, netto formue stat, alminnelig inntekt og trygdeavgiftene har forholdsvis liten spredning. Usikkerheten er noe større for personinntekt primærnæringer og tilhørende trygdeavgifter. Det kan forklares ved at det er forholdsvis få observasjoner for disse variablene, og dermed vil enkelte store verdier gi store utslag.

I forhold til IF-92 er IF-93 basert på et mindre utvalg, noe som kan gi større usikkerhet. På den annen side har utvalget gjennomgått grundige kontrollrutiner, slik at sjansen for feil i IF-utvalget ikke er så stor. Kalibreringsvariablene er de samme som i fjor, men i IF-93 korrigeres det som nevnt i tillegg til beløp, også for antall positive observasjoner av kalibreringsvariablene. Dette vil på grunnlag av analyser gjort på IF-92, gi et bedre resultat når man sammenligner inntekts- og formuesfordelingen i register og veid utvalg.

3. Inntekts- og formuesundersøkelsen for personlig næringsdrivende (IFN)

Tidligere år har som nevnt grunnlag for statistikk på personlig næringsdrivende vært hentet fra et felles utvalg for personer, husholdninger og personlig næringsdrivende. IFN-statistikk har dermed vært basert på opplysninger fra IFN-delen av utvalget og opplysninger om næringsdrivende i IF-delen. For IFN-delen av dette felles utvalget har det tidligere år (1991 og 1992) blitt dannet husholdninger på grunnlag av FOB90 og opplysninger fra ligningskontorene. Vektene har blitt beregnet på husholdningsnivå. I årets undersøkelse (IFN-93) er analyseenheten, redusert til å være næringsdrivende med eventuell ektefelle/registrert partner. I de tilfeller der samboere til næringsdrivende også driver næringsvirksomhet er disse inkludert i utvalget. IFN-statistikk baserer seg dermed på et eget utvalg bestående næringsdrivende fra ulike næringer, og analyseenheten er dermed personer og ektepar (i utvidet forstand).

Populasjonen til IFN ble dannet på grunnlag av opplysninger fra FOB90. Det er dannet en delpopulasjonen bestående av 1/12 av den opprinnelige populasjonen. Størstedelen av IFN-utvalget har vært med siden starten (IFN-91). En nærmere beskrivelse av utvalget er gitt i Notat 94/23. For en del av disse personene har næringen opphørt, og de er ikke lenger interessant i forhold til statistikk på næringsdrivende. De tilfellene hvor hverken IO eller ektefelle har noen næringsinntekt, er tatt ut av utvalget. Hvis IO's næring har opphørt, men ektefelle har næringsinntekt, er personene beholdt i utvalget på grunnlag av ektefelles næringsinntekt. De personene som er tatt ut av IFN-utvalget vil fortsatt være interessante i skatteforskningsøyemed, og vil dermed kunne inkluderes i det samveide utvalget som behandles senere. I tillegg til paneldelen av utvalget er det inkludert ca. 500 personer som ikke var næringsdrivende i 1990, men har registrert næringsvirksomhet i 1992. Frafallet p.g.a. opphør i næring var høyt også blant disse næringsdrivende.

Nytt av året er Selvangivelsesregisteret (SA) som gir mulighet for å kartlegge flere registeropplysninger om næringsdrivende enn det som har vært mulig tidligere. Tidligere undersøkelser har vist at det å ha en korrekt aldersfordeling i utvalget er av vesentlig betydning. Dette nye SA-registeret gjør det mulig å korrigere utvalget for alder, og i tillegg vil det også kunne finnes en del registerstørrelser fra ligningsregisteret som det kan kalibreres mot.

Hver person i utvalget får sin trekkesannsynlighet ut fra hovedinntektens art 1993 (strata). I de tilfeller der bare en av personene har næringsinntekt får eventuell ektefelle samme trekkesannsynlighet. I de tilfeller der begge har næringsinntekt blir trekkesannsynligheten til begge satt lik summen av de to trekkesannsynlighetene. Personene i det endelige utvalget fordeler seg etter næringsstrata med tilhørende trekkesannsynlighet som vist i tabellen nedenfor.

Tabell 7. Fordeling av intervjuobjekt etter næring

Næring	Antall IO	Trekkesannsynlighet
Primærnæring	295	295/(12*5095)
Håndverk	373	373/(12*1482)
Handel	366	366/(12*1479)
Industri	366	366/(12*530)
Frie yrker	573	573/(12*3290)
Lønn/pensjon > næringsinntekt	374	374/(12*9703)
Nye næringsdrivende	404	404/52000
Totalt antall IO	2751	

I alt 13 personer fra 7 «familier» er å betrakte som ekstremobservasjoner, og er tatt ut av utvalget og gitt vekt=1. Disse er tatt ut som ekstremer på grunnlag av meget høye observasjoner på registervariabel i forhold til andre, eller at kombinasjonene av næring og diverse skjema verdier er spesielle. Kriteriet er det samme som beskrevet for IF-utvalget.

Ut fra næringstilhørighet er den inverse trekkesannsynligheten funnet. Estimert aldersfordeling basert på denne vekten gir et bilde av hvordan aldersfordelingen er i utvalget i forhold til populasjonen.

Tabell 8. Estimert aldersfordeling IFN-utvalget basert på inverse trekkesannsynligheter

Aldersgruppe	IFN inverse trekkesanns.	Populasjonen 31.12.1995	%-vist avvik
16 - 19 år	1.158	2.353	-50.7
20 - 24 år	7.683	13.631	-43.6
25 - 29 år	27.576	34.845	-20.8
30 - 39 år	125.446	123.806	1.32
40 - 49 år	159.197	156.015	2.04
50 - 59 år	101.525	101.335	0.18
60 - 66 år	59.007	48.062	22.7
67 + år	57.205	47.170	21.3
Totalt	538.801	527.217	2.19

Tabell 8 viser at det spesielt for de yngste og eldste aldersgruppene er betydelige avvik mellom estimert og virkelig aldersfordeling. Det er en underestimering i yngste årsklasser og en overestimering i eldste årsklasser. For de mellom 30 og 60 år treffer det veide utvalget forholdsvis bra. Dette skyldes at det er i disse aldersgruppene hovedtyngden av næringsvirksomhet ligger, og at utvalget har flest observasjoner i disse aldersgruppene.

Vekter til IFN er videre beregnet ved å kalibrere mot aldersfordelingen og totalbeløp og totalantall for ligningsvariablene

- netto formue stat
- personinntekt pensjoner
- personinntekt lønn

Tidligere år har det blitt kalibrert også mot total personinntekt jord/skog/fiske og total personinntekt annen næring, men når dette ble gjort for 93-utvalget, ble de estimerte tallene innen hver næring ikke gode nok. Dette skyldes at de næringsdrivende i annen næring representerer virksomheter i ulike størrelser, og en kalibrering kun på korrekt totalbeløp har ingen innvirkning på den estimerte fordelingen etter næring. Det ble derfor forsøkt med intervallvis kalibrering av personinntekt annen næring. Variabelen ble delt inn i 22 intervall, og vektene ble beregnet slik at antall i hvert delintervall ble korrekt estimert. Dette sikret et bedre estimert beløp innen hver næring. For å sikre best mulig gjennomsnittstall er det også justert for antall innen hver av yrkesgruppene primærnæring, håndverk, handel, industri, frie yrker og lønnsinntakere. Antall innen de ulike gruppene ble funnet ved å koble opplysninger om næring fra merverdiregisteret til næringsdrivende i selvangivelsesregisteret fra 1993.

Denne kalibreringen resulterte i estimerte totalbeløp som vist i tabell 9.

Tabell 9. Estimerte totalbeløp for ligningsvariable i IFN, sammenlignet med ligningsregisterets tall. Tall i millioner. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte vekker	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	130.598	130.570	0.021
Netto formue stat	129.949	129.949	0
Alminnelig inntekt	76.371	77.186	-1.06
Toppskattegrunnlaget	82.246	82.676	-0.52
Personinntekt pensjoner	7.707	7.707	0
Personinntekt lønn	45.463	45.463	0
Personinntekt jord/skog/fiske	7.868	7.742	1.63
Personinntekt annen næring	21.402	22.517	-4.95
Skatt formue kommune	1.123	1.127	-0.0035
Skatt inntekt kommune	13.823	13.980	-1.12
Skatt formue stat	295	298	-1.007
Fellesskatt	4.607	4.620	-0.28
Toppskatt	2.104	2.161	-2.63
Forsinkelsesavgift	14	13	7.70
Tilleggsskatt	99	132	-25.0
Skatt på etterbetalt pensjon	6	7	-14.28
Trygdeavgift pensjon	231	230	0.44
Trygdeavgift lønn	3.480	3.488	-0.23
Trygdeavgift næring opp til 12G	2.070	2.133	-2.95
Trygdeavgift næring over 12G	121	123	-1.63
Trygdeavgift jord/skog/fiske	605	593	2.02
Forsørgerfradrag barn+ungdom	386	217	77.0
Finnmarksfradrag	63	98	-35.7
Fradrag BSU	14	10	40.0
Fradrag AMS	31	26	19.0
Skattebegrensning §78	111	116	-4.3
Anvendt godtgjøringslefradrag	706	956	-26.1
Netto formue kommune som SA	133.067	132.697	0.28
Alminnelig innt. kommune SA	76.623	77.396	-0.99
Netto formue stat SA	133.067	132.696	0.28
Alm. næringsinntekt jord/skog/fiske	13.767	13.258	3.84
Alm. næringsinntekt annen næring	28.138	27.969	0.60

Tabell 9 viser at de estimerte totalene er gode i de fleste tilfeller bortsett fra for spesielle variable der det er få observasjoner i utvalget. De estimerte verdiene av viktige variable som alminnelig næringsinntekt jord/skog/fiske og annen næring avviker lite fra registerverdiene.

Tilsvarende tabell er laget for estimert antall bak de forskjellige ligningsvariablene.

Tabell 10. Estimert antall positive observasjoner for hver av ligningsvariablene i IFN, sammenlignet med ligningsregisterets tall. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte vekter	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	224.663	225.763	-0.49
Netto formue stat	217.212	217.212	0
Alminnelig inntekt	453.347	456.442	-0.68
Toppskattegrunnlaget	474.000	475.758	-0.37
Personinntekt pensjoner	95.420	95.420	0
Personinntekt lønn	322.095	322.095	0
Personinntekt jord/skog/fiske	102.815	97.937	4.98
Personinntekt annen næring	154.886	154.886	0
Skatt formue kommune	224.999	226.389	-0.61
Skatt inntekt kommune	453.347	456.448	-0.68
Skatt formue stat	217.212	217.213	0.0005
Fellesskatt	453.347	456.440	-0.68
Toppskatt	148.238	144.721	2.43
Forsinkelsesavgift	10.175	10.122	0.52
Tilleggsskatt	9.937	9.032	10.02
Skatt på etterbetalt pensjon	775	816	-5.02
Trygdeavgift pensjon	95.021	94.785	0.25
Trygdeavgift lønn	309.667	309.504	0.05
Trygdeavgift næring opp til 12G	151.028	153.373	-1.53
Trygdeavgift næring over 12G	6.867	6.770	1.43
Trygdeavgift jord/skog/fiske	98.892	94.904	4.2
Forsørgerfradrag barn+ungdom	199.686	212.560	-6.05
Finnmarksfradrag	8.062	12.127	-33.5
Fradrag BSU	7.342	5.781	27.2
Fradrag AMS	33.332	28.735	15.99
Skattebegrensning §78	22.466	24.393	-7.89
Anvendt godtgjørslfradrag	63.884	56.069	13.94
Netto formue kommune som SA	274.182	269.738	1.65
Alminnelig innt. kommune SA	467.825	469.754	-0.41
Netto formue stat SA	274.182	269.734	1.65

Tabell 10 viser at avviket mellom estimert og virkelig antall varierer for de ulike variablene. Estimeringsmetoden treffer bra på formue og skatt, men bommer på noen av fradragene og avgiftene.

Tilsvarende som for IF-undersøkelsen har vi for å få en oversikt over utvalgsusikkerheten, beregnet konfidensintervall for sentrale ligningsvariable.

Tabell 11. Estimert utvalgsusikkerhet IFN

Variabel	95 % KI for gjennomsnittet		95 % KI for totalen. Tall i millioner		Relativt KI
	Gjennomsnittsverdi	Konfidensgrenser	Totalverdi	Konfidensgrenser	
Netto formue stat	259.443	7.728	129.949	3.871	0.0298
Pers.innt. pensjoner	15.384	518	7.707	259	0.0337
Pers.innt lønn	90.764	19.834	45.463	9.935	0.2185
Pers.innt. jord/skog/fiske	15.704	2.915	7.868	1.460	0.1857
Pers.innt. annen næring	42.729	1.550	21.402	776	0.0363
Toppskattegrunnlag	164.195	4.518	82.246	2.263	0.0275
Alminnelig inntekt	152.467	1.726	76.371	864	0.0113
Toppskatt	4.200	106	2.104	53	0.0253
Fellesskatt	9.199	371	4.607	186	0.0404
Tr.avg. pensjoner	461	42	231	21	0.0918
Tr.avg. lønn	6.948	568	3.480	284	0.0818
Tr.avg næring	4.133	228	2.070	114	0.0554
Tr.avg. næring over 12G	243	5.180	121	25	0.2129
Tr.avg. jord/skog/fiske	1.208	118	605	59	0.0980
Alm.innt. annen næring	56.168	1.706	28.138	854	0.0304
Alm.innt. jord/skog/fiske	27.481	5.589	13.767	2.799	0.2034

Tabell 11 viser at det er forholdsvis jevn spredning i de ulike variablene omfattet av tabellen. En sammenligning av tabell 11 og tilsvarende tabell 6 for IF, viser at konfidensintervallene for IFN er videre, noe som indikerer større spredning i dette materialet.

Det kan være vanskelig å sammenligne kvalitetene ved årets IFN-undersøkelse med tidligere års undersøkelser. Dette fordi årets undersøkelse som nevnt avviker fra de andre ved at den baserer seg på et eget utvalg bestående kun av næringsdrivende med ektefelle. I tillegg er det benyttet andre metoder for vektberegning i og med at det tas hensyn til fordeling av personinntekt annen næring. Det nye SA-registeret har også gitt større muligheter for kontroll av IFN-utvalget, spesielt gjelder dette konsistens i aldersfordeling og i fordeling av næringsdrivende etter hovedinntektens art. Alt dette medfører at kvaliteten på vektene og dermed også på undersøkelsen, vil bli forbedret i forhold til tidligere år.

4. Samveining av IF- og IFN-utvalgene

Utvalget som skal benyttes til skatteforskning og fordelingsanalyser er satt sammen av IF-utvalget og IFN-utvalget påkoblet øvrige familiemedlemmer. Vektene beregnes slik at alle i samme husholdning (IF) og familie (IFN) får samme vekt. Det vil si at det er en inkonsistens mellom enhetene vektene beregnes på grunnlag av. Alternativet vil være å benytte familie som enheten på hele utvalget, men det ble vurdert som et dårligere alternativ, og i tillegg ville ikke allerede innhentet informasjon om husholdningene bli benyttet. Trekkessannsynligheter er for IF-delen beregnet på grunnlag av trekkesannsynlighetene til de ulike delutvalgene, og for IFN-delen av utvalget basert på hvilken næring personen/familien tilhører. De som ble tatt ut av IFN-utvalget på grunn av opphørt næring, tas inn igjen i det samveide utvalget som representanter for populasjonen ikke-næringsdrivende. Disse personene får beregnet en trekkesannsynlighet ut fra hvilket strata de tilhørte når de ble trukket ut til IFN-utvalget i 1991. Disse personene kan tenkes på som avganger fra populasjonen næringsdrivende og ekvivalent som tilganger til populasjonen ikke-næringsdrivende. Denne trekkesannsynligheten sammenlignes med trekkesannsynligheten til de i IF-utvalget, og fordi trekkesannsynligheten til IFN-utvalget var større enn trekkesannsynligheten til IF-utvalget, justeres IFN-sannsynligheten ned med en faktor på 10.

I alt 39 personer fra 15 husholdninger/familier er på grunnlag av formelene (1) - (3) tatt ut av utvalget og behandlet som ekstremobservasjoner. Disse husholdningene er med unntak av én husholdning, ekstremobservasjoner i IF eller IFN. Ekstremene får vekt=1, og inkluderes i utvalget etter at det er funnet kalibrerte vekter.

Ved en samveining av de to delutvalgene veies vektene sammen i de tilfeller der det er næringsinntekt i familien/husholdningen. I de rene lønsmottakerhusholdningene beholdes vekten fra IF-utvalget. Den estimerte aldersfordelingen basert på inverse trekkesannsynligheter for det sammenkoblede utvalget er vist i tabellen nedenfor.

Tabell 12. Estimert aldersfordeling IF/IFN-utvalget basert på inverse trekkesannsynligheter

Aldersgruppe	IF/IFN	Populasjonen 31.12.1995	%-vist avvik
0 - 6 år	490.160	416.312	17.7
7 - 12 år	347.153	313.914	10.6
13 - 15 år	167.572	158.335	5.8
16 - 19 år	234.885	226.276	3.8
20 - 24 år	340.820	331.265	2.8
25 - 29 år	375.776	341.955	9.9
30 - 39 år	699.564	637.277	9.7
40 - 49 år	665.093	613.468	8.4
50 - 59 år	425.042	406.816	4.5
60 - 66 år	285.724	258.446	10.5
67 + år	646.172	620.751	4.1
Totalt	4.677.965	4.324.815	8.2

Av tabell 12 fremgår det at aldersfordelingen i det sammensatte utvalget også er skjev. Jevnt over blir estimert antall innen hver aldersgruppe estimert for høyt, og en justering for alder er nødvendig.

For å skape størst mulig grad av konsistens mellom de ulike undersøkelsene, baserer valg av kalibreringsvariable for IF/IFN-utvalget seg på hva som ble valgt for IF-utvalget. Estimerte totalbeløp er dermed funnet ved kalibrering med hensyn på aldersfordelingen og totalbeløpene til følgende ligningsvariable

- netto formue stat
- personinntekt pensjoner
- personinntekt lønn
- personinntekt jord/skog/fiske
- personinntekt annen næring

I tillegg er det kalibrert med hensyn på antall positive observasjoner for alle disse ligningsvariablene. Resultatene for totalbeløp og totalantall er vist i tabellen nedenfor.

Tabell 13. Estimerte totalbeløp for ligningsvariable i IF/IFN, sammenlignet med ligningsregisterets tall. Tall i millioner. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte veker	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	422.340	421.985	0.084
Netto formue stat	414.877	414.877	0
Alminnelig inntekt	377.157	375.147	0.54
Toppskattegrunnlaget	482.381	481.964	0.086
Personinntekt pensjoner	90.618	90.618	0
Personinntekt lønn	358.850	358.850	0
Personinntekt jord/skog/fiske	7.776	7.776	0
Personinntekt annen næring	25.387	25.387	0
Skatt formue kommune	3.289	3.294	-0.15
Skatt inntekt kommune	64.638	64.144	0.77
Skatt formue stat	794	795	-0.12
Fellesskatt	21.547	21.371	0.82
Toppskatt	8.743	8.529	2.51
Forsinkelsesavgift	22	28	-21.4
Tilleggsskatt	248	240	3.33
Skatt på etterbetalt pensjon	122	131	-6.87
Trygdeavgift pensjon	2.732	2.713	0.70
Trygdeavgift lønn	27.657	27.602	0.199
Trygdeavgift næring opp til 12G	2.413	2.318	4.09
Trygdeavgift næring over 12G	174	146	19.2
Trygdeavgift jord/skog/fiske	598	596	0.33
Forsørgerfradrag barn+ungdom	1.998	1.958	2.04
Finnmarksfradrag	382	449	-14.9
Fradrag BSU	154	134	14.9
Fradrag AMS	118	110	7.27
Skattebegrensning §78	2.281	2.390	-4.56
Anvendt godtgjøringsfradrag	2.391	2.784	-14.11
Netto formue kommune som SA	447.633	449.587	-0.44
Alminnelig innt. kommune SA	380.771	379.406	0.359
Netto formue stat SA	447.633	449.570	-0.43
Alm.næringsinnt. jord/skog/fiske	14.093	13.320	5.8
Alm.næringsinnt. annen næring	30.880	28.033	10.15

Estimeringsmetoden gir forholdsvis bra resultat for formue og trygdeavgifter bortsett fra for noen av fradragene. Dette skyldes antakelig få observasjoner i utvalg og register. Vestentlige variable som alminnelig næringsinntekt primærnæring og annen næring avviker mer fra registerverdiene enn hva tilfellet var for IFN-utvalget vist i tabell 9. Tilsvarende tabell for antall personer med positive verdier for de ulike ligningsvariablene er vist nedenfor.

Tabell 14. Estimert antall positive observasjoner for hver av ligningsvariablene i IF/IFN sammenlignet med ligningsregisterets tall. Uthevete tall er kalibreringsvariable

Variabel	Alders- lignings- og antallskalibrerte vekter	Ligningsregisteret 31.12.1993	%-vist avvik
Netto formue kommune	1.105.283	1.108.536	-0.29
Netto formue stat	1.021.956	1.021.956	0
Alminnelig inntekt	2.941.693	2.984.171	-1.42
Toppskattegrunnlaget	3.262.872	3.326.155	-1.9
Personinntekt pensjoner	1.099.642	1.099.643	0
Personinntekt lønn	2.420.563	2.420.563	0
Personinntekt jord/skog/fiske	98.475	98.475	0
Personinntekt annen næring	174.863	174.863	0
Skatt formue kommune	1.108.580	1.112.950	-0.39
Skatt inntekt kommune	2.941.693	2.984.270	-1.43
Skatt formue stat	1.021.956	1.022.003	-0.0046
Fellesskatt	2.941.954	2.984.000	-1.41
Toppskatt	806.472	791.128	1.94
Forsinkelsesavgift	21.394	23.287	-8.13
Tilleggsskatt	28.709	22.726	26.3
Skatt på etterbetalt pensjon	10.716	9.488	12.9
Trygdeavgift pensjon	1.083.692	1.081.913	0.16
Trygdeavgift lønn	2.220.603	2.202.355	0.82
Trygdeavgift næring opp til 12G	172.118	170.158	1.15
Trygdeavgift næring over 12G	10.770	7.777	38.5
Trygdeavgift jord/skog/fiske	95.456	95.447	0.0094
Forsørgerfradrag barn+ungdom	987.639	986.498	0.12
Finnmarksfradrag	55.068	62.926	-12.5
Fradrag BSU	84.811	77.264	9.76
Fradrag AMS	125.029	123.277	1.42
Fradrag for etterbetalt pensjon	3.243	3.117	4.04
Fradrag nedsettelse 65 %	4.296	5.978	-28.1
Skattebegrensning §78	492.443	525.544	-6.3
Anvendt godtgjørslfradrag	260.455	251.346	3.62
Netto formue kommune som SA	1.747.465	1.778.477	-1.74
Alminnelig innt. kommune SA	3.247.415	3.318.878	-2.15
Netto formue stat SA	1.747.465	1.778.416	-1.74

For det samveide utvalget bommer estimeringsmetoden på skattene og enkelte av fradragene. Viktige variable som formue og trygdeavgifter estimerer metoden bra. Graden av over/underestimering er den samme.

På grunn av statistikkens formål er vektene i det samveide utvalget beregnet slik at alle i samme husholdning har lik vekt. Det er derfor interessant å få en oversikt over fordelingen av husholdninger etter størrelse. Husk at for IFN-delen er familie satt lik husholdning. Tabellen nedenfor viser en slik fordeling.

Tabell 15. Fordeling av husholdninger etter størrelse IF/IFN93 sammenlignet med IF/IFN92 og IF93

Husholdnings- størrelse	Antall husholdninger IF/IFN93	Antall husholdninger IF/IFN92	Antall husholdninger IF93
1	773.299	773.266	745.518
2	494.119	456.912	500.865
3	264.056	261.056	272.121
4	265.092	253.990	254.980
5	108.344	109.766	112.834
6	21.177	22.562	21.555
7	3.136	4.737	3.283
8	1.719	1.842	2.248
9	343	243	571
10	76	199	0
11	170	0	150
Totalt	1.932.036	1.884.573	1.914.130

Sammenlignes estimert husholdningsfordeling for IF93 med estimert husholdningsfordeling for IF/IFN92 og 93 i tabell 13, fremgår det at IF/IFN-utvalgene estimerer flere enpersonhusholdninger enn IF gjør. Dette kan forklares ved at i IFN-delen av det samveide utvalget er det en overvekt av enpersonhusholdninger bl.a. fordi samboere ikke blir registrert. De to IF/IFN-utvalgene er forholdsvis like bortsett fra for to-personshusholdninger. Totalt sett resulterer IF/IFN-utvalgene i flere husholdninger enn IF-utvalget, noe som også kan skyldes at familie defineres som husholdning.

Tabell 16. Estimert utvalgsusikkerhet IF/IFN

	95 % KI for gjennomsnittet		95 % KI for totalen Tall i millioner		Relativt KI
	Gjennomsnittsverdi	Konfidensgrenser	Totalverdi	Konfidensgrenser	
Netto formue stat	95.915	1838	414.877	7.950	0.01917
Pers.innt. pensjoner	20.953	127	90.618	550	0.00608
Pers.innt lønn	82.973	4457	358.850	19.278	0.05372
Pers.innt. jord/skog/fiske	1.798	147	7.776	638	0.08211
Pers.innt. annen næring	5.868	191	25.387	829	0.03267
Toppskattegrunnlag	111.536	804	482.381	3.480	0.00722
Alminnelig inntekt	87.205	710	377.157	3.070	0.00814
Toppskatt	2.021	363	8.743	1.572	0.17987
Fellesskatt	4.982	1007	21.547	4.357	0.20222
Tr.avg. pensjoner	631	34	2.732	147	0.05423
Tr.avg. lønn	6.395	154	27.657	669	0.02420
Tr.avg næring	557	195	2.413	845	0.35058
Tr.avg. næring over 12G	40	23	174	102	0.58697
Tr.avg. jord/skog/fiske	138	14	598	63	0.10549
Alm.innt. annen næring	7.138	57	30.880	250	0.00812
Alm.innt. jord/skog/fiske	3.258	14	14.093	63	0.00448

Tabell 16 viser at variasjonen er mindre i IF/IFN-utvalget enn i IF- og IFN-utvalget vist i tabell 6 og tabell 11. Spesielt gjelder denne forbedringen variable som alminnelig inntekt jord/skog/fiske og annen næring. Estimaten i IF/IFN har mindre usikkerhet enn estimaten i den rene næringsundersøkelsen IFN,

noe som mest sannsynlig forårsakes av at utvalget er større. Det blir derfor et vurderingsspørsmål hvorvidt man velger et utvalg som gir mindre usikkerhet fremfor et utvalg som estimerer bedre, men har større variasjon.

5. Inntekts- og formuesundersøkelsen for skattepliktige aksjeselskap (IFS)

Hovedformålet med IFS er å skaffe til veie data for skatteanalyser, som er egnet både til å vise hvordan dagens bestemmelser virker, og som et grunnlag for å beregne hvordan nye regler kan slå ut. Gjennom Etterskuddsregisteret (E-reg) hentes det informasjon som kan benyttes til å lage statistikk over aksjeselskapers tilpasning til skatteregler, oversikt over utlignet skatt for ulike selskapstyper, analyser av hvordan endringer i skatteregler vil slå ut for aksjeselskaper, og analyser av hvordan virkningen av skattleggingen av aksjeselskaper endrer seg over tid. Trekkenhet og analyseenhet er foretak hentet fra E-reg. Utvalget er trukket på grunnlag av E-reg, men for å danne et best mulig grunnlag for stratifisering hentes så mye informasjon som mulig ut fra andre registre. Utvalgsplanen utformes slik at alle typer selskaper målt etter størrelse og næring er representert. I utgangspunktet består utvalget av en totaltellingsdel og en ordinær utvalgsdel. For mer utfyllende opplysninger henvises til utvalgsplanen for IFS 93 (SRB 15.juni 1995 og SRB 17. oktober 1995).

På grunn av de registrene som benyttes, og på grunn av de krav som stilles til utvalgsplanene, oppstår det enkelte problemer. Utvalget er trukket på grunnlag av E-reg 92, mens populasjonen det skal sies noe om er representert i E-reg 93. Dette avviker i tid mellom trekketidspunkt og populasjonstidspunkt kompenseres ved at det ble trukket et tilleggsutvalg fra Brønnøysundregisteret som skal representere de nyetablerte i populasjonene. Disse tilgangene er både totaltellingsforetak og utvalgsforetak. I tillegg er det en del foretak som av ulike årsaker går over fra utvalgsforetak til totaltellingsforetak. Totaltellingsforetakene skulle ideelt sett hatt vekt lik 1, men fordi utvalget ikke dekker alle foretak, har vi valgt å la foretakene få en vekt større enn 1. Utvalgsdelen av tilleggsutvalget inkluderes i den opprinnelige utvalgsdelen.

For å kunne ta best mulig hensyn til representativiteten og dekke de ulike tilfellene på en best mulig måte, er populasjonen delt inn i flere delpopulasjoner ut fra hvordan utvalgene fra disse er trukket. Tabell 17 nedenfor viser de ulike delutvalgene og delpopulasjonene, og hvordan vektene er beregnet.

Tabell 17. Oversikt over delutvalg og delpopulasjoner

Utvalg	Kommentar	Populasjon
3.313	Utvalgsdel inkludert tilganger utvalg. Vekt beregnet ved kalibrering etter sum positiv alminnelig inntekt i hver næring og antall med positiv, 0, og negativ alminnelig inntekt.	99.090
56	Ble trukket som utvalgsforetak, men har siden trekketidspunktet blitt totaltellingsforetak. Vekt beregnet manuelt ut fra fordeling på positiv alminnelig inntekt.	314
50	Tilganger totaltelling. Trukket for å representere de nye i populasjonen. Vekt beregnet manuelt ut fra fordeling på positiv alminnelig inntekt.	179
25	Ekstremer kommet til i løpet av vektberegningene + de med negativ vekt. Vekt=1.	25
2.684	Totaltellingsforetak, (ekstremer, total 92, total 93) Vekt=1.	2.684
6.128	Sum	102.292

Denne fremgangsmåten resulterte i vektorer som ga korrekte estimater av positiv alminnelig inntekt innen hver næring, i tillegg til gode estimater av skattene. Andre variable er det vanskelig å si noe om fordi registervariablene er ufullstendige. Vektene sikrer også en rimelig fordeling av antall foretak innen hver næring. Vektene til IFS-93 er først og fremst beregnet med tanke på å gi nivå-tall for 1993, og ikke for å kunne sammenligne IFS-tall over tid.

Innen hver næring fordeler vektene seg som vist i tabell 18.

Tabell 18. Fordeling av vekt etter næring

Næring	Antall obs.	Sum vekt	Gj.snitt	Max. vekt	Antall pop
0124	1.259	27.252	21.64	138.03	27.241
3	784	10.024	12.78	45.44	9.982
5	465	8.306	17.86	48.62	8.297
6	2.135	33.332	15.61	52.49	33.321
7	260	4.132	15.89	105.70	4.145
8	930	14.099	15.16	75.99	14.150
9	295	5.144	17.43	56.76	5.156
Totalt	6.128	102.292	16.69	138.03	102.292

Tabell 15 viser at gjennomsnittlig vekt er forholdsvis lik for de ulike næringene. Unntaket er gruppen hvor næring 0, 1, 2 og 4 er slått sammen. Av ulike årsaker har noen foretak i denne gruppen fått en forholdsvis høy vekt. Estimert antall innen hver næring er noenlunde i overensstemmelse med antall i populasjonen. Disse vektberegningene resulterer i følgende estimerte totaltall totalt og innen hver næring.

Tabell 19. Estimerte totaler fra IFS-utvalget sammenlignet med tall fra Etterskuddsregisteret. Beløp i millioner

Variabel	Estimert total alle delutvalg (6128 obs.)	Fasittall fra Etterskuddsregisteret (102292 obs.)	%-vist avvik
Antall sysselsatte	720.511	639.240	12.7
Positiv alminnelig inntekt	49.640	50.586	-1.87
Negativ alminnelig inntekt	-11.539	-11.871	2.8
Omsetting	1.330.368	910.234	46.0
Aksjekapital	167.349	148.160	12.9
Kommuneskatt	5.454	5.567	-2.03
Fellesskatt	8.403	8.562	-1.86
Sum skatt	14.058	19.343	-27.3
Driftsinntekter	1.247	951	31.2
Totalkapital	1.105	1.190	-7.14
Positiv alm. inntekt næring 0,1,2,4	9.374	9.888	-5.19
Positiv alm. inntekt næring 3	10.344	10.440	-0.92
Positiv alm. inntekt næring 5	1.665	1.674	-0.54
Positiv alm. inntekt næring 6	13.139	13.015	0.95
Positiv alm. inntekt næring 7	1.838	1.879	-2.16
Positiv alm. inntekt næring 8	12.355	12.748	-3.08
Positiv alm. inntekt næring 9	923	938	-1.6

Tabellen viser at positiv og negativ alminnelig inntekt totalt og innen hver næring estimeres meget godt. Det samme gjør også totalkapital og skattene unntatt sum skatt. Antall syssesatte estimeres for høyt. Dette kan skyldes at de foretakene som er med i utvalget har gjennomgående for lave beløp på positiv alminnelig inntekt i forhold til antall sysselsatte. For de øvrige variablene underestimeres totalene sammenlignet med totaltallene fra Etterskudsregisteret.

Tabell 20. Estimert utvalgsusikkerhet IFS

Variabel	95 % KI for gjennomsnittet		95 % KI for totalen Tall i millioner		Relativt KI
	Gjennomsnittsverdi	Konfidensgrenser	Totalverdi	Konfidensgrenser	
Pos.alm.innt. næring 1,2,4	91.534	37.810	9.374	3.868	0.41307
Pos.alm.innt. næring 3	101.229	37.747	10.344	3.861	0.37288
Pos.alm.innt. næring 5	16.287	8.094	1.665	828	0.49696
Pos.alm.innt. næring 6	128.365	13.043	13.139	1.334	0.10161
Pos.alm.innt. næring 7	17.963	10.807	1.838	1.105	0.60163
Pos.alm.innt. næring 8	120.835	27.465	12.355	2.809	0.22729
Pos.alm.innt. næring 9	9.037	2.944	923	301	0.32582
Aksjekapital	1.634.367	8.225	167.349	841	0.01
Kommuneskatt	53.314	5.384	5.454	550	0.10
Sum skatt	137.417	14.771	14.058	1.511	0.11
Totalkapital	11.728	1.714	1.105	175	0.15

Det er visse avvik fra tidligere IFS-undersøkelser både når det gjelder utvalget og vektberegningene. Variablene det etterstratifiseres etter er de samme som i fjor, men ideen med å dele utvalg og populasjon i mindre enheter er ikke blitt forsøkt tidligere. I tillegg har det ikke tidligere vært kalibrert med hensyn på totalbeløp innen hver næring.

6. Konklusjon

Problemer

Arbeidet med IFS har stort sett gått uten alvorlige problemer. Det at det var inkonsistens mellom trekke- og populasjonstidspunkt har skapt en del ekstraarbeid, men fordi det gjennom hele prosessen har vært en dialog mellom seksjon 720 og seksjon 420 har de fleste problemer blitt løst i fellesskap. Som nevnt er det visse forandringer i forhold til i fjor, men forhåpentligvis vil årets vektresultater i et analysegrunnlag som er representativt og konsistent for populasjonen.

Når det gjelder de andre inntekts- og formuesundersøkelsene har arbeidet p.g.a diverse problemer som har oppstått, blitt ganske mye forsinket både i forhold til i fjor og i forhold til planlagte tidsfrister.

Den største årsaken til at vektene ble klare såpass sent var at arbeidet med vektning ikke kom i gang før midt i juni, på et tidspunkt hvor vektene etter planen allerede skulle vært ferdig. Denne forsinkelsen skyldes først og fremst stor arbeidsmengde ved seksjon 420. I tillegg var det i vektefilen en del feil som det tok tid å rette opp. Dette gjaldt spesielt IFN-utvalget der det var med ca 500 personer som ikke hadde noen næringsinntekt, eller som ikke hadde noen ektefelle med næringsinntekt. Disse personene ble tatt ut av IFN-utvalget, men ved samveilingen ble de inkludert i IF/IFN-utvalget med en konstruert trekkesannsynlighet. Disse personene burde vært tatt ut av vektefilen på et mye tidligere tidspunkt. De første vektsettene til uttalelse ble levert i månedsskiftet juni/juli, men på grunn av bl.a. ferieavvikling kom det ingen reaksjoner før midt i august. Endelige vektsett ble levert midt i september.

Vektberegningene til årets IF-undersøkelser for husholdninger og personlig næringsdrivende er for første gang todelt. I fjor ble det beregnet kun ett sett vektorer til et sammensatt IF/IFN-utvalg som dannet grunnlag for IF- og IFN-statistikk. Det samme utvalget ble benyttet som grunnlag for skatteøkonomiske analyser. I år ble IF-statistikk laget på grunnlag av et eget IF-utvalg med husholdning som enhet, IFN-statistikk ble laget på grunnlag av et eget IFN-utvalg med person/ektefelle som enhet, og grunnlaget for skatteøkonomiske analyser baserte seg på et samveid utvalg av IF og IFN (inkludert familiemedlemmer). Disse tre delutvalgene har hvert sitt tilhørende vektsett, og skal hver for seg være representative for de respektive populasjonene samtidig som det skal være en viss konsistens mellom de veide utvalgene. Det har derfor tatt en del tid å finne den løsningen som ville ivareta de fleste interesser og som ga mest mulig tilfredsstillende resultater. Dette betyr at arbeidsmengden ved årets undersøkelser har vært betraktelig større enn tidligere. I tillegg har det vært en del problemer med populasjonstallene for IFN. Det måtte etableres nye filer for å kunne gi populasjonstall for IFN. I utgangspunktet var det ved en feil med noen aktive eiere i populasjonen som bidro til at populasjonstallene ble alt for høye, og det var vanskelig å få tilfredsstillende kravene til utvalget. Da denne feilen ble rettet opp ble konsistensen mellom IFN-utvalg og IFN-populasjon betraktelig forbedret.

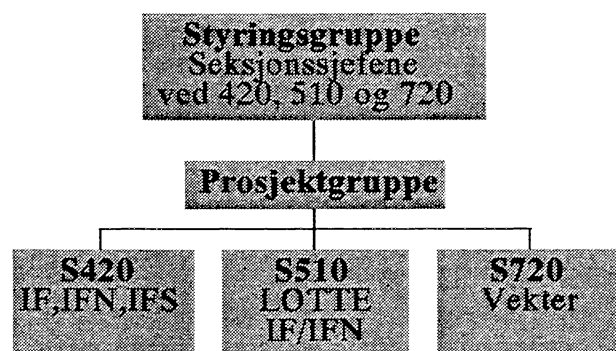
Usikkerhetsberegningene viser at IFN-utvalget har større utvalgsusikkerhet enn det samveide IF/IFN-utvalget, samtidig som de estimerte totaltallene for næringsvariable blir forholdsvis like i de to utvalgene. Det bør foretas tilsvarende tabellkjøringer på næringsdelen av IF/IFN som på IFN. Dette for å få et sammenligningsgrunnlag for å vurdere hvorvidt IFN-statistikk skal produseres ut fra et eget utvalg eller hvorvidt resultatet blir tilnærmet det samme ved å benytte det samveide utvalget.

Det endelige resultatet tyder på å være bedre enn tidligere år. Fordi utvalget takler flere kalibreringsvariable, er antakelig utvalget mer homogent enn før. Det å korrigere for antall med positiv observasjon på enkelte av ligningsvariablene, vil resultere i økt konsistens mellom utvalg og register. Spesielt i IFN er denne konsistensen forsøkt ivaretatt i og med at vi har justert for antall i delintervaller av en av de viktigste ligningsvariablene; personinntekt annen næring. Det bør for senere IF-undersøkelser diskuteres hvorvidt det bør tas med flere kategoriske kalibreringsvariable, og eventuelt en finere inndeling av disse. En del av arbeidsoppgavene fremover vil være å dokumentere forbedringen i form av analyser på fordelingen av inntekt og formue i populasjonen i forhold til i de veide utvalgene, og sammenligne med tilsvarende analyser for 1992.

Totalt sett kan man si at i arbeidet med IF-undersøkelsene bør hele prosessen fra datainnsamling til ferdig produkt sees på som en helhet. Det er viktig at datagrunnlaget gjennomgår en kvalitetskontroll før vektene beregnes. Dette for å unngå misforståelser og unødig arbeid. Denne kvalitetskontrollen bør bestå av å kontrollere for ulogiske sammenhenger mellom ligningsvariable, ekstreme observasjoner, spesielle familie / husholdningssammensetninger og andre feil i datamaterialet. I tillegg er det nødvendig med ekstremkontroll etter at vektene er beregnet.

Forbedringer

I et forsøk på å forbedre samarbeidet og kommunikasjonen og dermed også det endelige statistikkproduktet, er det foreslått å opprette en prosjektgruppe og en styringsgruppe organisert som vist nedenfor.



Prosjektgruppens viktigste arbeidsoppgave vil være å koordinere arbeidet med alle IF-undersøkelsene og på den måten sørge for at sluttproduktet; IF-statistikken og datagrunnlaget for LOTTE blir best mulig. En målsetning må være at forståelsen for de enkelte seksjoners arbeid og behov økes. Gruppen bør møtes minst en gang hver 14.dag (kan diskuteres) og ha mandat til i samarbeid med styringsgruppen, å ta beslutninger vedrørende utvalgsproblematikk, tidsplaner, kvalitetskontroll, vektberregning og andre løpende avgjørelser. Det er meningen av prosjektgruppen skal arbeide såpass tett at problemer og misforståelser med skjema, utvalg og populasjonstall unngås. Medlemmene i prosjektgruppen er ansvarlig for rapportering til sin egen seksjon/gruppe, og å sørge for at de oppgavene den enkelte seksjon er pålagt blir gjennomført. Arbeidet med IF-undersøkelsene bør følge de retningslinjene som er vedtatt i de respektive seksjoners virksomhetsplaner.

Styringsgruppen bør så snart som mulig sette sammen prosjektgruppen slik at arbeidet med IF-undersøkelsene for 1994 kan komme igang.

Referanser

Johan Heldal, *Notat 92/3, A Method for Calibration of Weights in Sample Surveys.*

Ann Marit Kleive og Leiv Solheim, *Notat 94/23, Kalibrering av vekter.*

Anne Sørbråten, *Notat datert 15. juni 1995, Utvalgsplan IFS.*

Anne Sørbråten, *Notat datert 17. oktober 1995, Beregning av vekter IFS93.*

Jørgen Aasness, Erik Fjærli, Hanne A. Gravningsmyhr, Ann Marit Kleive Holmøy and Bård Lian
Working Papers Amalgamated Series no.9533, Departement of Applied economics, University of
Cambridge, *The Norwegian Microsimulation Model LOTTE:
Applications to personal and Corporate Taxes and Social Security Benefits.*

Statistisk sentralbyrå

Oslo
Postboks 8131 Dep.
0033 Oslo

Telefon: 22 86 45 00
Telefaks: 22 86 49 73

Kongsvinger
Postboks 1260
2201 Kongsvinger

Telefon: 62 88 50 00
Telefaks: 62 88 50 30

ISSN 0806-3745



Statistisk sentralbyrå
Statistics Norway