

Rapport
1/2001

**Analyse av
inntektsfordeling og
inntektsulikhet basert på
registerdata**

**En kartlegging av muligheter
og begrensninger**

Remy Åserud



*Stiftelsen Frischsenteret for samfunnsøkonomisk forskning
Ragnar Frisch Centre for Economic Research*

**Analyse av inntektsfordeling og inntektsulikhet basert
på registerdata**

En kartlegging av muligheter og begrensninger

Remy Åserud

Sammendrag: Denne rapporten tallfester og drøfter avvikene mellom registerdata og data hentet fra utvalgsundersøkelser med hovedvekt på forholdet mellom begrepene hushold og familie, og hvordan disse avvikene kan virke inn på den estimerte inntektsfordelingen. Vi får bekreftet at samboere uten felles barn er en av gruppene preget av lavt samsvar mellom hushold og familie, men så lenge befolkningen ses under ett betyr det lite for den estimerte inntektsfordelingen hvilke data som ligger til grunn. Når vi retter fokuset mot bestemte grupper i befolkningen tyder imidlertid resultatene på at bruk av familie som økonomisk gir opphav til for høye estimat på inntektsulikheten. For å redusere denne skjevheten introduseres en metode for å omgjøre det familie-baserte inntektsregisteret til et husholds-basert inntektsregister.

Nøkkelord: Inntektsfordeling, og -ulikhet, utvalgsundersøkelser, hushold/familie, logit-modell.

Kontakt: www.frisch.uio.no, tlf 22 95 88 10

Rapport fra prosjektet " Inntektsfordeling og inntektsmobilitet", (internt prosjektnummer 1502), finansiert av Arbeids- og administrasjonsdepartementet

* Takk til Frischsenteret for muligheten jeg fikk til å arbeide der som studentassistent for å skrive hovedoppgave, for tilrettelegging av datamaterialet og all støtte underveis. Jeg vil også takke min veileder, Rolf Aaberge, som med sin faglige styrke og mange inspirerende synspunkter har vært til uvurderlig hjelp under arbeidet.

Innhold

1. INNLEDNING	1
1.1 OVERSIKT.....	2
1.2 FORDELER VED ANALYSER BASERT PÅ REGISTERDATA	3
2. REGISTER OG HUSHOLD	4
2.1 OVERSIKT.....	4
2.2 AVVIK PÅ HUSHOLDSNIVÅ.....	5
2.3 AVVIK ETTER ALDER	6
2.4 OM ÅRSAKER TIL AVVIK.....	8
2.3 SAMBOERE UTEN BARN	9
2.4 OM MÅLING AV AVVIK MELLOM REGISTER OG INTERVJU	10
2.4.1 Tidsavvik mellom innhenting av registeropplysninger og gjennomføring av intervju.....	11
2.5 KONKLUSJON	12
3. FORDELINGSEFFEKTER.....	13
3.1 OM VEKTING OG UTVALG	14
3.2 GINI-KOEFFISIENTEN	15
3.3 VEKTER OG ESTIMERING AV FORDELING	16
3.4 HUSHOLD OG FORDELING	17
3.4.1 Fra egen inntekt til ekvivalentinntekt	17
3.4.2 Inntektseffekten.....	19
3.4.3 Størrelseseffekten.....	20
3.5 LAV INNTEKT	23
3.6 DESILFORDELINGER	26
3.7 SAMBOERE OG ENSLIGE	29
3.8 KONKLUSJON	33
3.8.1 Mulige løsninger.....	34
4. MODELL FOR Å FINNE SAMBOERE GJENNOM REGISTEROPPLYSNINGER.....	36
4.1 POPULASJON OG UTVALG	37
4.2 TEORI	38
4.3 TIDLIGERE STUDIER.....	40
5. ESTIMERING	42
5.1 OM DATAGRUNNLAGET.....	42
5.2 HVEM SKAL ”NØKKELEN” BRUKES PÅ?	43
5.3 MODELL 1	44
5.4 MODELL 2	48
5.5 DISKUSJON	52
REFERANSER:.....	55

1. Innledning

For å analysere fordelingen av inntekt og måle inntektsulikhet i befolkningen, er det nødvendig å ta utgangspunkt i husholdet som økonomisk enhet og benytte husholdets samlede disponible inntekt som grunnlag for å tallfeste enkeltmedlemmenes økonomiske levestandard. Husholdets sammensetning må vanligvis kartlegges ved å intervju personene som er trukket ut. Dette er fordi det bare er mulig å finne familieenheter i registerdataene (Det sentrale folkeregisteret), og familier er et snevrere begrep enn hushold. En familie kan bestå av en enslig, ugift far/mor med barn, ektepar med og uten barn. I Det sentrale folkeregisteret (DSF) identifiseres alle medlemmene i en familie gjennom et felles familienummer. Et hushold omfatter på sin side alle som bor og spiser sammen (kosthushold). Husholdet kan altså inneholde mer enn en familie.

I de fleste tilfeller vil ett hushold bestå av én familie slik at begrepene blir sammenfallende. For slike hushold er dermed registeropplysninger like gode som intervjuopplysninger. Det er hovedsakelig to grupper hushold vi *ikke* kan finne ved hjelp av registerdataene. Den ene gruppen er flerfamiliehushold, som er en forholdsvis liten gruppe. Ifølge tallene fra inntekts- og formuesundersøkelsen (IF) 1997 tilhører omlag 1,9 prosent av husholdene denne gruppen. Den andre gruppen er hushold som består av samboere uten felles barn. Disse vil være registrert som enslige dvs. som separate familier i registeret.

Problemet som da melder seg er om avviket mellom disse to begrepene er for stort til at familie kan brukes som grunnlag for studier av befolkningens økonomiske velferd. I tilfelle svaret er ja vil det være aktuelt å søke alternativ registerinformasjon og undersøke om det er mulig å konstruere hushold som ligger nærmere kosthusholdsdefinisjonen enn det familie gjør.

Formålet med denne oppgaven er å kartlegge mulighetene for å bruke Inntektsregisteret som grunnlag for studier av økonomisk velferd i befolkningen. For å gjennomføre denne evalueringen har vi brukt data fra SSBs inntekts- og formuesundersøkelse i 1997 siden denne undersøkelsen er basert på kosthushold (fra intervju) og dessuten inneholder alle inntektsopplysninger fra Inntekts- og formueundersøkelsen, som inntekt, alder, utdanning,

familienummer osv. Dataene omfatter opplysninger om nesten 40 000 personer i ca. 14 500 hushold.

1.1 Oversikt

Hovedformålet med dette prosjektet er å undersøke muligheten for å estimere inntektsfordelinger og inntektsulikhet på grunnlag av ”syntetiske” hushold, dvs. vi bruker registerdata og ikke intervju som grunnlag for å bestemme husholdssammensetningen blant personene i Inntektsregisteret eller et utvalg fra registeret. Oppgaven består av tre hoveddeler:

1. I kapittel 2 kartlegges samsvaret/avviket mellom intervju- og registeropplysninger. I denne delen vil vi analysere og tallfeste avviket mellom intervjuopplysninger om husholdssammensetning og registeropplysninger om familie. Metoden som brukes er stort sett frekvenstabeller av ulike bearbejdede og sorterte variable i datasettet. Analysene viser at samboere uten barn er en spesielt problematisk gruppe i så henseende, for eksempel bor over halvparten av alle som er registrert som enslige sammen med en eller flere andre. Vi fant også at borteboende studenter skaper avvik mellom registerfamilie og hushold. Dette skyldes reglene for folkeregistrering, der hovedregelen er at studenter ikke skal melde flytting til studiestedet. Generelt er avvik mellom registerfamilie og hushold mest vanlig blant personer i 20-årene.
2. Kapittel 3 undersøker i hvilken grad bruk av registeropplysninger gir opphav til skjevheter i estimering av inntektsfordelingen (inntektsulikhet) gjennom dannelsen av ekvivalentinntekt. Denne delen ser både på hele befolkningen og forskjellige undergrupper. Inntektsfordelingen blir hovedsakelig studert ved hjelp av Gini-koeffisienter, desilfordelinger og en lavinntektsgruppe. Analysene gir få sikre konklusjoner, men kan tyde på at bruk av familie som økonomisk enhet kan gi opphav til for høye estimanter på inntektsulikheten. Identifisering av samboere gir små utslag på fordelingene når vi ser på hele utvalget under ett. Men i bestemte alders- og husholdsgrupper gir en slik identifisering klart lavere observert ulikhet og færre i lavinntektsgruppen. For å estimere Gini-koeffisienter og desilfordelinger, som er en sentral del av dette kapitlet, brukes programmet Uroteng, skrevet av Tom Wennemoe og Rolf Aaberge ved Statistisk Sentralbyrå.
3. Kapittel 4 og 5 ser på mulighetene for å utvikle en modell som kan brukes til å predikere samboere uten barn i inntektsregisteret, vha. registerdata og adresseopplysninger. I

kapittel 4 presenteres den multinomiske logit-modellen og egenskaper ved data og bruksområde for modellen blir gjennomgått. Kapittel 5 inneholder to ulike logitmodeller estimert ved prosedyrene `proc logistic` og `proc catmod` i programpakken SAS. Modellene får stort sett signifikante parameterestimater med forventede fortegn. Den ”beste” modellen viser ved simuleringsforsøk god evne til å reprodusere fordelingen av samboere blant personer registrert som enslige.

1.2 Fordeler ved analyser basert på registerdata

Dersom vi ved hjelp av registerdata kan ”identifisere” personers faktiske hushold, eller noe i nærheten av dette, vil det åpne for flere interessante analyser. Analyser basert på registerdata har noen viktige fordeler framfor intervjuundersøkelser:

- De dekker et stort antall personer, og fjerner dermed problemer med for få observasjoner. Et stort antall observasjoner gir oss både sikrere estimater på ulike størrelser, og muligheter for å si noe om relativt små grupper.
- Registeropplysninger gir muligheter for å følge personer over tid, noe som gir anledning til å undersøke mobilitet i inntekt og inntektsfordeling. Vi kan også se på inntektsfordeling med lengre periode enn et år, og undersøke i hvilken grad inntektsfordelingen påvirkes av å forlenge regnskapsperioden til to, tre og fire år. I tillegg vil det være mulig å undersøke om lav inntekt er et forbigående eller kronisk problem, og videre studere hva som kjennetegner personer med lav inntekt over flere år.
- Registeropplysninger er uten frafall, slik at eventuelle skjevheter som følge av lav respons på intervju blir borte.

2. Register og hushold

2.1 Oversikt

Vi skal i det følgende se nærmere på samsvaret mellom registeropplysninger om familie og intervjuopplysninger om hushold. Dette er gjort for å kartlegge avviket mellom familie og hushold, og dermed finne ut hvor vi bør rette oppmerksomheten når vi senere skal se på inntektsfordelingen ved hjelp av registerdata. For å bestemme husholdssituasjonen til personene som trekkes ut for å være med i Inntekts- og Formuesundersøkelsene (IF) blir disse intervjuet. En god del (omlag 30 prosent) svarer av ulike grunner ikke på intervjuet, og for disse brukes da registeropplysninger om familie som hushold. Men for de som svarer på intervjuet, nettoutvalget, har vi både opplysninger fra register og intervju. Det er nettoutvalget til IF-1997 vi har brukt som grunnlag for analysen nedenfor. Nettoutvalget består av ca. 28.000 personer fordelt på 10.000 hushold.

Tabell 2.1: *Sammenlikning mellom antallet medlemmer i familien i henhold til register og antall medlemmer i husholdningen oppgitt ved intervju i IF-1997. Prosent.*

Antall personer i register-familien	Antall personer i husholdet ifølge intervju					
	1	2	3	4	5	6 el. flere
1	48	37	8	5	1	1
2	4	84	7	3	1	1
3	5	24	63	6	1	1
4	4	8	13	73	2	1
5	4	4	6	10	73	2
6 eller flere	5	5	3	4	13	69

I tabell 2.1 har vi foretatt en slik sammenlikning mellom register- og intervjuopplysninger. Tabellen viser fordelingen av husholdsstørrelse etter størrelsen på registerfamilien. Tabell 2.1 viser tydelig hvor store feilene blir når vi bruker registerfamilie som husholdsdefinisjon. F.eks ser vi at av de familiene som var registrert som bestående av 1 person, bestod husholdet (ved intervju) av 2 personer i hele 37 prosent av tilfellene, mens kun halvparten (48 prosent) av enperson familiene faktisk levde som enslige. Vi finner også betydelige avvik blant registerfamiliene bestående av 3 eller 4 medlemmer. Her skyldes en vesentlig del av avviket at det er færre medlemmer i husholdet enn i registerfamilien. Forklaringen på dette er først og

fremst at studenter som ikke er hjemmeboende blir registrert som del av familien til foreldrene i DSF. Det klart største problemet med registeropplysningene er at det blir for mange enslige og for få par.

2.2 Avvik på husholdsnivå

Tabell 2.2 gir et mer detaljert bilde av hvilke husholdstyper som i størst grad berøres av avvikene. Tabellen viser i hvilken grad personer er registrert i familier som er mindre enn, like store eller større enn husholdsstørrelsen de oppgir i intervju. Den uthevede kolonnen i midten viser andelen personer innen hver husholdstype som er registrert tilhørende en familie med like mange medlemmer som det er i husholdet personen lever i. Kolonnene på hver side viser andelen personer med avvik mellom registerfamiliestørrelse og husholdsstørrelse. Av tabellen kan vi lese både retningen på avviket (hushold>familie eller motsatt) og størrelsen

Tabell 2.2: Samsvaret mellom intervjudata (hushold) og registerdata (familie) for personer, etter ulike typer hushold i IF-1997. Prosent. (N=27954.)

Husholdstype	Hushold>familie			Hushold =familie	Familie>hushold			Alle
	3+pers.	2 pers.	1 pers.		1 pers.	2 pers.	3+pers.	
Enslige								
under 45 år	-	-	-	65	5	12	18	100
45-64 år	-	-	-	86	12	2	1	100
65 år og over	-	-	-	96	3	1	0	100
Par uten barn, eldste								
under 45 år	-	-	55	31	6	6	3	100
45-64 år	-	-	9	61	22	7	1	100
65 år og over	-	-	4	90	6	0	0	100
Par med barn, yngste								
0-6 år	2	2	3	91	2	0	0	100
7-17 år	3	2	4	78	2	0	0	100
18 år og over	1	1	4	76	16	2	0	100
Mor/far med barn, yngste								
0-17 år	0	2	2	83	11	2	0	100
18 år og over	0	0	0	88	10	1	0	100
Andre (flerfamiliehushold)	11	25	40	10	6	5	3	100
I alt	1	2	8	75	10	3	2	100

(1,2 eller 3 og flere personer). Det er i stor grad blant hushold av typen par uten barn vi finner det laveste samsvaret mellom intervju og register. For personer som er under 45 år og lever i parforhold uten barn, er samsvaret mellom størrelsen på registerfamilie og hushold bare på 31 prosent.

Det lave samsvaret skyldes vesentlig at så mye som 55 prosent ifølge registeropplysningene bor alene. Også for par uten barn mellom 45 og 64 år er det forholdsvis stort avvik, men for denne gruppen skyldes mesteparten av dette at registerfamilien er større enn husholdet. Dette kan i stor grad skyldes utflyttede barn, som på grunn av regelverket, fortsatt er registrert sammen med sine foreldre i folkeregisteret.

Samsvaret er meget dårlig for gruppen som sier de lever i flerfamiliehushold. Registeropplysningene forteller hvem som er i familie med hverandre, og kan derfor naturligvis ikke fange opp at flere familier deler hushold. Imidlertid er denne gruppen forholdsvis liten, bare omlag 3 prosent av personene i datamaterialet oppga at de bodde i et flerfamiliehushold.

Vi har nå sett at registeropplysninger gir et varierende grunnlag for å si noe om husholdsstørrelsen, særlig for unge enslige, og par under 65 år uten barn. For over en tredjedel av unge enslige vil bruk av registeropplysninger gi et større hushold enn de i virkeligheten lever i. Fordelingsmessig kan dette få to konsekvenser. For det første kan samlet husholdsinntekt bli høyere enn den skulle vært, fordi for mange personer blir koblet på. For det andre vil ekvivalentinntekten isolert sett bli lavere siden husholdsinntekten skal deles på flere. I utgangspunktet kan vi ikke si noe generelt om den totale effekten av disse to faktorene. For unge enslige kan det gi for høy ekvivalentinntekt, mens foreldrene og eventuelle fortsatt hjemmeboende søsken vil få for lav ekvivalentinntekt i forhold til deres faktiske økonomiske velferd. Skjevheten mellom familie- og husholdsstørrelse går andre veien for personer i gruppen par under 45 år uten barn. Disse er i hovedsak registrert som enslige, og derfor vil deres ekvivalentinntekt ikke reflektere de økonomiske fordelene ved å dele hushold; den blir for lav.

2.3 Avvik etter alder

Vi skal nå se litt nærmere på hvilken aldersgruppe som blir sterkest berørt av forskjeller mellom register og virkelig husholdsstatus. Tabell 2.3 nedenfor viser tydelig hvordan

samsvarsprosenten mellom husholdstatus og registerfamiliestatus varierer med alder. Tidlig i livet er det bra samsvar, mens det faller gjennom tenårene og ut i første halvdel av tjuårene. Dette kan som nevnt skyldes at mange flytter hjemmefra og deretter danner nye hushold som ikke fanges opp av registeret (DSF). En slik hypotese styrkes når vi ser på årsakene til avviket. For gruppen 15-19 år skyldes mye av avviket at det er færre i husholdet enn i registerfamilien, noe som tyder på utflyttede studenter. Det forholdsvis svake samsvaret vi observerer blant personer i tjuårene, skyldes derimot for en stor del at husholdet er større enn registerfamilien. Dette indikerer samboere uten barn, som altså ikke er registrert som samme familie i registerene. I trettiårsalderen er samsvaret forholdsvis bra igjen, mange av samboerne har giftet seg eller fått barn sammen og dermed øker samsvaret.

Tabell 2.3: Samsvar mellom husholdsstørrelse og registerfamilie etter aldersgrupper. IF-1997. Prosent.

Alder	Hushold > registerfamilie			Hushold = familie	Hushold < registerfamilie		
	3+ pers.	2 pers.	1 pers.		1 pers.	2 pers.	3+ pers.
0-15	1	1	5	88	4	1	0
15-19	1	1	6	76	11	2	2
20-24	3	2	13	60	9	5	6
25-29	2	4	20	64	5	3	2
30-34	2	2	12	82	2	1	0
35-39	1	2	6	87	3	1	0
40-49	1	2	7	73	14	3	1
50-59	1	2	7	67	18	5	1
60-69	1	1	6	82	10	1	0
70-	1	2	5	88	4	0	0

I førti- og femtiårsalderen er samsvaret lavere igjen. Da skyldes forskjellen i stor grad at husholdet er mindre enn registerfamilien, og dette kan nok en gang forklares med utflyttede barn. I de eldste aldersgruppene er samsvaret mye bedre.

Det samme bølgemønsteret kommer tydelig frem i Figur 2.1. under, hvor vi betrakter alder som kontinuerlig variabel. Lavest samsvar mellom husholdsstørrelse og familiestørrelse observerer vi i to grupper; personer rundt 20 år og personer mellom 45 og 55 år. For denne siste gruppen skyldes avviket hovedsakelig at husholdet er mindre enn familien. Som antydnet ovenfor kan dette skyldes barn som har flyttet uten at dette blir registrert som endring i

Folkeregisteret. For den første gruppen skyldes avviket først at husholdet er større enn familien. Men rundt 23 års alderen er den viktigste årsaken at husholdet er større enn registerfamilien. Ved denne alderen begynner mange å bli ferdig med utdannelsen sin og de

Figur 2.1: Samsvar mellom hushold og registerfamilie for personer i IF-97 etter alder.

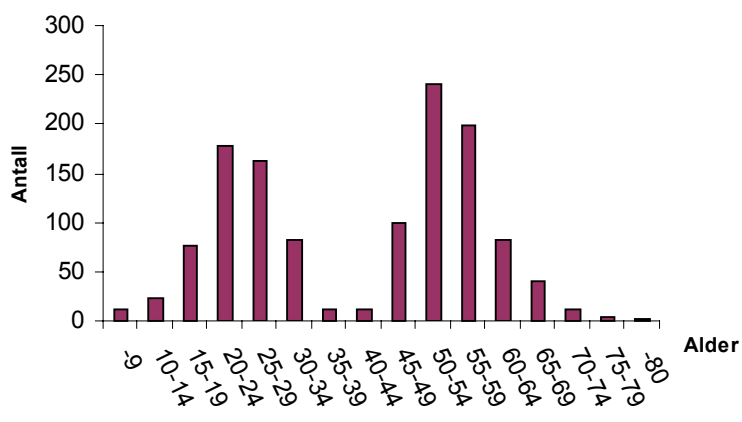


folkeregistrerer seg på sin nye adresse. Derfor er de ikke lenger registrert sammen med sine foreldre og eventuelt søsken, men for en stor del alene. Når man så finner seg en samboer slår dette ut i at husholdet i mange tilfeller blir større enn registerfamilien.

2.4 Om årsaker til avvik

Vi vil nå se nærmere på om de nevnte forklaringene vedrørende årsaker til avvik kan stemme. Vi tar utgangspunkt i personer som er 20-29 år gamle, er i nettoutvalget og har en registerfamiljestørrelse som avviker fra husholdsstørrelsen. I denne aldersgruppen har vi sett at

Figur 2.2: Personer som tilhører familien, men ikke husholdet til personer i 20-årene, etter alder.



avviket er forholdsvis stort. Hadde vi brukt registeropplysninger for å danne hushold for denne gruppen, ville vi feilaktig koblet på personene med samme familienummer. Ovenfor har vi gruppert disse personene, som vi kan kalle ”falske husholdsmedlemmer”, etter alder i figur 2.2. De ”falske husholdsmedlemmene” fordeler seg tydelig i to aldersgrupper. Den ene har tyngdepunkt rundt 50-årsalderen og består av foreldrene til personene på 20-29 år. Den andre gruppen har tyngdepunkt omlag ved 20-år og består mest sannsynlig av søsken. Disse tallene støtter hypotesen om at en stor del av avviket for personer i 20-årene, skyldes at selv om de har flyttet ut og bor alene, er de fortsatt registrert sammen med sin familie i folkeregisteret. Sannsynligvis dreier det seg som nevnt om studenter som etter reglene ikke skal melde fra til folkeregisteret dersom de flytter. For personer som er 20 år er omlag 50 prosent under utdanning, mens andelen synker til 25 prosent for 25-åringer (kilde Statistisk Årbok 1999). Det dreier seg derfor om en ganske stor gruppe personer. I Inntekts- og Forbruksundersøkelsene skal borteboende studenter regnes som bosatt på studiestedet dersom de har rett til å motta borteboerstipend, Hauge et.al. (2000). Opplysninger om hvem som mottar stipend fra lånekassen kan derfor hjelpe oss med å identifisere disse personene, slik at de ikke feilaktig havner i hushold med sin familie.

2.3 Samboere uten barn

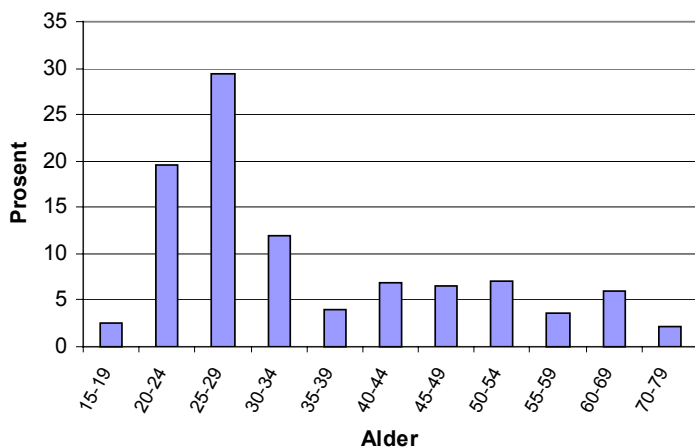
Vi vil nå se nærmere på gruppen samboere uten barn. Siden vi senere skal utarbeide en modell for å finne fordelingen av samboere uten barn, kan det være nyttig å fokusere spesielt på denne gruppen. Nettopp fordi samboere uten barn ikke får felles familienummer i Det sentrale folkeregisteret, er det lite statistikk over denne gruppen. I Ukens Statistikk nr.9, 1999 anslår SSB at det i 1998 var omlag 250.000 samboerpar i Norge. Bare 87.000 av disse parene har felles barn, noe som innebærer at ca. 2/3 av alle samboere, eller omlag 330.000 personer, er samboere uten barn. Dette utgjør omtrent 7.5 prosent av Norges befolkning.

Tabell 2.4: *Andel samboere etter alder i 1997. Prosent.*

Kjønn	Alder										
	16-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-69	70-79
Kvinner	7	35	36	23	16	9	9	6	6	2	1
Menn	1	17	37	31	20	13	11	9	6	4	2

Kilde: ukens statistikk nr.9 1998, SSB.

Figur 2.3: Samboere uten barn, fordeling etter alder. (N=1106)



Tabell 2.4 viser andelen personer som er samboere i ulike aldersgrupper. Det kommer tydelig frem at forekomsten av samboerskap er størst blant unge mennesker. Men denne tabellen gjelder alle samboere, også de med barn.

I utvalget til IF-1997 kan vi identifisere samboere uten barn. Datasettet inneholder opplysninger om 553 slike par. Som vi ser av figur 2.3, tilhører disse stort sett yngre aldersgrupper. Omlag halvparten av samboerne uten barn i datasettet er i 20-årene. 82 prosent av de 1106 personene i denne gruppen var registrert som enslige. De resterende 18 prosent var registrert i en familie bestående av 2 eller flere personer. Disse var fordelt på 4,5 prosent i familie med en annen, 6 prosent i familie med 2 andre, 4,5 prosent med 3 andre og 2,8 prosent i familie med 4 eller flere andre personer. Dette betyr at dersom vi skal finne samboere uten barn kun ved registerfamilie og adresse, kan vi ikke bare lete blant personer som er registrert som enslige. Antagelig vil det hjelpe dersom vi først identifiserer borteboende studenter (vha. opplysninger om stipend fra Lånekassen) slik at også disse er kandidater til å være samboere.

2.4 Om måling av avvik mellom register og intervju.

Vi har i det foregående målt avvik mellom register og hushold som forskjell i antall personer i registerfamilien og antall personer i husholdet. Dette målet på avvik fanger imidlertid ikke opp hele det reelle avviket. Selv om en person bor i et hushold som har like stort antall medlemmer som hans registerfamilie, kan husholdet bestå av *andre personer* enn hans registerfamilie. F.eks. kan en student være registrert sammen med en forelder, men leve i samboerskap med en annen. I dette tilfellet er registeropplysningene som de skal være,

personens hushold er like stort som registerfamilien, men husholdet består av andre personer enn registerfamilien. Et slikt avvik kan få like mye å si for en persons ekvivalentinntekt som avvik i antall personer. Problemet bør likevel være minimalt, fordi de langt fleste tilfellene blir fanget opp av registeret, det er ikke mange eksempler som det over.

2.4.1 Tidsavvik mellom innhenting av registeropplysninger og gjennomføring av intervju.

Det er viktig å påpeke at noen av avvikene mellom register- og intervjuopplysninger kan skyldes at disse opplysningene ikke er samlet inn på samme tidspunkt. Registeropplysningene er hentet fra årsskiftet, mens intervjuene foregår omlag 3 måneder senere, Sparby (1997). I løpet av denne tiden kan det skje endringer i familiene som dermed kan gi avvik. Folk kan f.eks. gifte seg, skille seg, få barn eller gå bort iløpet av disse månedene. Hvor stor andel av avviket som skyldes denne forskyvningen i tid er vanskelig å si ut fra dataene. Befolkningsstatistikk kan imidlertid gi en pekepinn på omfanget av slike familieendringer. Av 1000 ekteskap ble 19,4 oppløst ved død og 11,3 ved skilsmisse i 1997 (kilde: Statistisk Årbok 1999). Iløpet av tre måneder vil derfor omlag 5 og 3 av 1000 ekteskap bli oppløst pga. henholdsvis død og skilsmisse. (Antar at skilsmisser og dødsfall fordeler seg jevnt over året.) Hadde dette vært de eneste årsaker til avvik, skulle vi finne at husholdet i rundt 1 prosent av tilfellene var mindre enn registerfamilien. Som vi har sett, er omfanget av avviket i en helt annen størrelsesorden. Det er naturlig å anta at personer som dør mellom årsskiftet og intervju gir størst utslag blant eldre. Men nettopp blant eldre finner vi det høyeste samsvar mellom register og intervju. Dette underbygger antagelsen om at forsinkelsen mellom trekketidspunkt og intervju betyr lite for avvik i retningen hushold mindre enn registerfamilie.

Barnefødsler i tiden mellom trekke- og intervjutidspunkt vil gi flere personer i familien (og husholdet) enn det registeret viser. I datasettet har 3,3 prosent av husholdene barn på 1 år. På tre måneder kan vi derfor gå ut fra at i underkant av 1 prosent av husholdene får barn, og dermed en familie som er større enn registrene viser. Hovedtyngden av kvinnene som får barn er i alderen 25-34 år (Statistisk Årbok 1999). Ser vi på denne aldersgruppen isolert, finner vi at omlag 11 prosent av personene lever i en husholdning hvor det yngste barnet er 1 år. Dersom vi antar at dette er konstant, vil det si at omkring 3 prosent av personene i denne aldersgruppen får barn iløpet av månedene mellom trekketidspunkt og intervju. En del av avviket blant personer i alderen 25-34 kan dermed forklares med nyankomne barn, men det er

likevel langt igjen til det observerte avviket på mellom 26 og 16 prosent der husholdningen var større enn familien.

Dersom vi skulle bruke registerdata i fordelingsstudier direkte uten intervju, ville tidfestingsproblemene med fødsler, giftermål, skilsmisser og død bli borte. Det er likevel grunn til å kartlegge det avviket de skaper, fordi vi da får kunnskap om hvilket avvik vi kommer til å sitte igjen med, som ikke skyldes tidsavvik mellom trekke- og intervjutidspunkt.

2.5 Konklusjon

Som ventet er det samboere uten barn som byr på størst utfordringer når det gjelder å danne hushold på basis av registerdata. Dette kom tydelig frem allerede i tabell 2 der vi så at 55 prosent av personene under 45 år som lever i parforhold uten barn, var registrert som enslige. Vi har også sett at gruppen samboere uten barn utgjør et ikke ubetydelig antall personer, omlag 330.000. Et annet problem, er de mange unge enslige som er registrert sammen med sin familie, selv om de har flyttet ut. Dersom dette ikke dreier seg om registreringsfeil, noe som er lite sannsynlig, gjelder det borteboende studenter. Siden disse skal regnes som egne hushold, vil ikke folkeregisteret gi tilstrekkelig informasjon til å identifisere slike hushold. Sannsynligvis vil opplysninger om stipendtildeling fra Lånekassen være til god hjelp i disse tilfellene.

Fordelingsmessig vil det få konsekvenser at så mange samboere uten barn blir regnet som enslige ved bruk av familie som økonomisk enhet. Personer som lever i samboerskap uten felles barn vil dermed framstå med lavere økonomisk velferd enn det de i virkeligheten har. Bruk av familie som økonomisk enhet vil også påvirke borteboende studenter, som ofte er registrert sammen med sine foreldre. Ved bruk av familien som økonomisk enhet, vil de få en økonomisk velferd som reflekterer foreldrenes inntekt selv om de bor for seg selv. Dette vil sannsynligvis gi en for høy økonomisk velferd for denne gruppen iforhold til deres reelle situasjon.

3. Fordelingseffekter

Vi er interessert i personers husholdssammensetning fordi det først og fremst er inntektene i husholdet som er avgjørende for en persons økonomiske velferd. Videre er det en vanlig oppfatning at det er betydelige stordriftsfordeler i husholdet, slik at et toperson-hushold som har en samlet inntekt på f.eks. kr 500.000, vil ha det mer romslig økonomisk sett enn et hushold bestående av en enslig som har en inntekt på kr 250.000. For å sammenlikne hushold av ulik størrelse, bruker man en ekvivalensskala som omregner husholdsinntekter til sammenlignbare personinntekter. En mye brukt ekvivalensskala dividerer husholdets inntekt med kvadratroten til antall personer i husholdet, se Atkinson et. al. (1995), Epland (1998-a) og Lund og Aaberge (1999). Referansehusholdet i denne skalaen er altså den enslige. For at et par skal ha samme levestandard som en enslig som tjener kr 250.000 må paret ifølge denne skalaen tjene $kr\ 250.000 * 2^{1/2} = kr\ 353.553$ tilsammen; da vil hver av dem oppnå en økonomisk velferd tilsvarende en enslig med disponibel inntekt på kr 250.000.

Vi så i forrige kapittel at registeropplysninger om familie ikke var en fullgod erstatning for intervjuopplysninger om hushold. I dette avsnittet ønsker vi å finne svar på spørsmålet: I hvilken grad vil estimatene av inntektsfordeling og inntektsulikhet bli påvirket av at vi bruker registerfamilie istedet for hushold for å danne personers ekvivalentinntekt? En naturlig fremgangsmåte ville være å danne to sett ekvivalentinntekter for personene vi har intervjuopplysninger om, ett basert på husholdsopplysninger fra intervju og ett basert på registerdata. Her støter vi imidlertid på et problem. For personene som har svart på intervjuet har vi bare opplysninger om *husholdets* inntekter. De eneste opplysningene om personene i registerfamiliene, dersom denne avviker fra det reelle husholdet, er kjønn og fødselsår. Et eksempel: En person som er trukket ut for å være med i undersøkelsen er kanskje registrert sammen med sine foreldre og søsken, men under intervjuet kommer det fram at han bor alene. De eneste opplysningene vi da har om hans foreldre og søsken er fødselsår, kjønn og familienummer. Vi har ingen opplysninger om deres inntekt. Dermed kan vi ikke danne ekvivalentinntekter for personer i nettoutvalget med utgangspunkt i registeropplysninger. For frafallet er det motsatt; vi har her opplysninger om registerfamiliens inntekter, men selvsagt ikke for husholdets hvis disse to avviker fra hverandre. Ved sammenlikning av register og hushold, må vi altså samtidig sammenlikne personer i frafallgruppen med personer fra svargruppen. Disse gruppene kan imidlertid ikke betraktes som tilfeldige utvalg fra

bruttoutvalget. F.eks. har de som ikke svarer i gjennomsnitt lavere inntekt, lavere utdanning, er noe eldre, og har i gjennomsnitt omlag 2,43 personer i sin registerfamilie mot 2,72 personer for de som svarer.

En vei utenom dette problemet er å sammenlikne to grupper innenfor nettoutvalget. Den ene gruppen omfatter personer der hushold og registerfamilie er sammenfallende. Dermed kan vi se på denne gruppen som om registerdata er utgangspunkt for beregning av ekvivalentinntekt. Den andre gruppen er personer der husholdet er forskjellig fra registerfamilien, og for denne bruker vi husholdsopplysningene for å danne ekvivalentinntekter.

Et viktig unntak fra denne generelle beskrivelsen, er personer som er registrert som enslige. Deres registerfamilie er jo ikke annet enn personen selv, så for disse kan vi danne ekvivalentinntekt på basis av registeropplysninger og sammenlikne med deres ekvivalentinntekt på basis av husholdsopplysninger. Vi fastslo i forrige kapittel at det største problemet ved å bruke registerfamilie som økonomisk enhet, er samboere uten barn, som oftest er registrert som enslige. Vi kan derfor finne ut hvilken innvirkning det har på inntektsfordelingen å identifisere samboere blant personer registrert som enslige i DSF.

Ekvivalentinntekten er i det følgende regnet ut ved hjelp av følgende formel:

$$\text{Ekvivalentinntekt} = \frac{\text{Husholdsinntekt}}{(\text{Antall personer i husholdet})^E}$$

der $E = 0,5$ dersom ikke annet er oppgitt.

3.1 Om vekting og utvalg

Vektene i datasettet er beregnet ut i fra to formål. For det første skal de sørge for at utvalget reflekterer populasjonen mht. alder og befolkningsstatistikk. Men de blir også kalibrert slik at utvalgets inntektsvariable summerer seg til kjente totalverdier. Variablene dette gjelder er lønnsinntekt, personinntekt fra selvstendig sysselsetting i primærnærings, andre næringer og i liberale næringer, pensjonsgivende inntekt og netto formue som gir skatt til staten (Epland 1998-b). Avviket mellom utvalget og fasitsvar fra likningsregisteret som kalibreringen rettet opp, var i størrelsesorden fra -15 prosent på pensjonsinntekt til 33 prosent på personinntekt fra liberale næringer, Holdt et.al (1999).

I de følgende beregninger vil vi istedet benytte vektorer som kun retter opp at personer som tilhører ulike hushold/familiestørrelser har ulik sannsynlighet for å komme med i utvalget, dvs. det er bare trekkesannsynlighetene som er avgjørende for størrelsen på vektene. Ved et rent tilfeldig utvalg av hushold ville alle personer ha samme sannsynlighet for å bli trukket ut og dermed fått tildelt samme vekt. Men siden vi trekker et utvalg personer og husholdet/familie siden kobles på, vil en person som tilhører et stort hushold ha større sjanse til å bli med i utvalget enn en som lever i et mindre hushold. Vi sier at disse vektene baserer seg på inverse trekkesannsynligheter. Grunnen til at vi heller bruker slike vektorer er at i fordelingsstudier er det ikke totalnivået på inntekter som interesserer, men derimot den relative fordelingen av inntekter blant befolkningen. Dessuten vil kravet om konsistens med observerte totaler kunne lede til skjevheter i den relative fordelingen av inntekter.

I utvalget til IF inngår fire delutvalg: Inntekts- og formuesundersøkelsens ordinære panel, levekår panel, og levekår tverrsnitt og et tilleggsutvalg av næringsdrivende. Vi vil bruke alle utvalgene utenom tilleggsutvalget av næringsdrivende, fordi dette utvalget ikke er representativt uten bruk av de originale vektene.

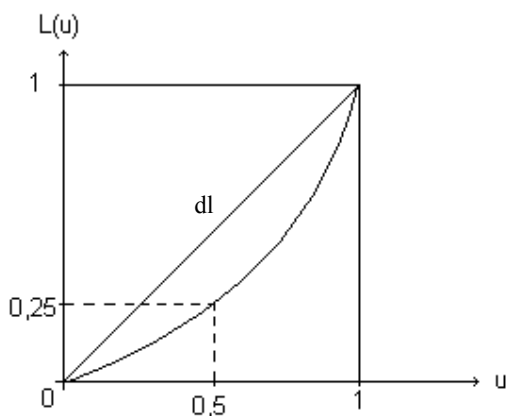
3.2 Gini-koeffisienten

Gini-koeffisienten er en vanlig måte å oppsumere graden av ulikhet i en inntektsfordeling. Den er et tall mellom 0 og 1 der 0 betyr at alle personer har nøyaktig like stor inntekt og 1 betyr at en person har all inntekten i økonomien og alle andre ikke har noenting. Formelt tar man utgangspunkt i Lorenzkurven til den kumulative inntektsfordelingen. Vi lar X være en inntektsvariabel med kumulativ fordeling $F(\cdot)$ og gjennomsnitt μ . F er definert over $[0, \infty)$, F^{-1} er den inverse til F og $F^{-1}(0) = 0$. Lorenzkurven $L(\cdot)$ til F er da definert som følger:

$$L(u) = \frac{1}{\mu} \int_0^u F^{-1}(t) dt, \quad 0 \leq u \leq 1.$$

Lorenzkurven sier hvor stor andel av den totale inntekten i samfunnet som tilkommer de fattigste $u \cdot 100\%$ av befolkningen. Figur 3.1 under gir et eksempel på en Lorenzkurve. I ette tilfellet er kurven tegnet slik at den halvparten av befolkningen som har lavest inntekt tilsammen får 25 prosent av all inntekt i samfunnet. Jo lavere og lengere mot høyre kurven ligger, jo større er inntektsforskjellene i økonomien. Den diagonale linjen, dl , på figuren kan betraktes som en Lorenz-kurve for en inntektsfordeling der alle har nøyaktig lik inntekt. Gini-

Figur 3.1: Eksempel på Lorenzkurve



koeffisienten tar utgangspunkt i dette ekstreme tilfellet der alle har lik inntekt og måler avviket fra denne situasjonen til den observerte. Formelt er Gini-koeffisienten definert slik, Bojer (1990):

$$G = 2 \int_0^1 (u-L(u))du = 1-2 \int_0^1 L(u)du$$

Med andre ord er Gini-koeffisienten lik arealet mellom u og L(u) multiplisert med 2. Fordelen med denne koeffisienten er at den oppsummerer mye informasjon om fordelingen i en økonomi i et enkelt tall mellom 0 og 1. Gini-koeffisienten måler bare relative inntektsforskjeller, og dermed blir det f.eks. mulig å sammenlikne inntektsfordelingen i land med forskjellige valutaer uten å gå veien om kjøpekraftspariteten.

3.3 Vekter og estimering av fordeling

I tabell 3.1 har vi beregnet Gini-koeffisienter med standardavvik på to ulike måter. Som vi ser øker Gini-koeffisienten i alle de tre delutvalgene en del ved bruk av vekter basert på inversetrekkesanssynligheter. I Inntekts- og formuesundersøkelsen for 1996 ble Gini-koeffisienten beregnet til 0,278, og to år før til 0,279 (NOS C 581). For IF-1997 er det foreløpig ikke publisert tall for inntektsulikhet. Vi ser at Gini-koeffisientene beregnet på basis av trekkesanssynlighetene er omlag 10 prosentpoeng større enn de vi finner på grunnlag av de

Tabell 3.1: Ginikoeffisienter for personers ekvivalentinntekt etter ulike delutvalg og vekter.

Utvalg	Vekter	Gini-koeffisient	Standardavvik
Hele utvalget	Originale	0,259	0,004
	Nye	0,271	0,004
Utvalget uten selvstendig næringsdrivende	Originale	0,258	0,004
	Nye	0,266	0,005
Kun tilleggsutvalg av næringsdrivende	Originale	0,276	0,007
	Nye	0,286	0,006

vektene som Statistisk Sentralbyrå benytter. Dette viser at kravet om konsistens med kjente totalverdier for inntekter og skatter skaper skjevheter i estimeringen av Gini-koeffisienten som leder til for trekkesannsynlighetene er omlag 10 prosentpoeng større enn de vi finner på grunnlag av de vektene som SSB benytter. Dette viser at kravet om konsistens med kjente totalverdier for inntekter og skatter kan skape skjevheter i estimeringen av Gini-koeffisienten som igjen kan lede til for lave anslag på inntektsulikheten. Imidlertid kan det forholdsvis store avviket også skyldes at SSBs vekter retter opp eventuell skjev aldersfordeling i utvalget som følge av frafall. I det følgende vil vi benytte vektene basert på inverse trekkesannsynligheter dersom ikke annet er oppgitt.

3.4 Hushold og fordeling

Et viktig formål med dette kapitlet er å finne ut om intervjuopplysninger om hushold lar seg erstatte av registeropplysning om familie med tanke på å estimere mål på ulikhet i inntektsfordelingen. Som nevnt er husholdet viktig i slike studier fordi vi antar at en persons økonomiske velferd i større grad er avhengig av husholdets inntekt og størrelse enn personens egen inntekt i seg selv. Dersom vi hadde brukt personlig inntekt som utgangspunkt for inntektsanalyser, og ikke ekvivalentinntekt, hadde det ikke vært nødvendig å bestemme personers hushold, og man kunne uten problemer bruke inntektsregisteret direkte. Men resultatene av slike analyser ville blitt ubrukelige med hensyn til økonomisk velferd. Det er nettopp derfor vi analyserer ekvivalentinntekt istedenfor personlig inntekt, og dermed må husholdsstatus klarlegges.

3.4.1 Fra egen inntekt til ekvivalentinntekt

I tabell 3.2 har vi regnet ut Gini-koeffisienten med tilhørende standardavvik for fire ulike utvalg av hushold. Hvert utvalg representerer en måte å danne hushold på. Den første gruppen består av personene i nettoutvalget hvis hushold var identisk med registerfamilie. Den andre gruppen er også fra nettoutvalget, men består av personer fra husholdene som på en eller annen måte ikke stemte med en registerfamilie. I den tredje gruppen, har vi personene fra frafallet. Som vi vet blir husholdene til disse personene dannet kun ved hjelp av registeropplysninger om familie. Den siste gruppen er summen av alle personer som enten tilhører nettoutvalget og ikke har avvik mellom registerfamilie og hushold, eller frafallet. Det vil si at alle i denne gruppen har hushold identisk med registerfamilie i dataene. Årsaken til denne inndelingen, er at vi ønsker å se om måten husholdet blir bestemt på gir utslag på estimatene til Gini-koeffisienten.

I et forsøk på å kontrollere for at personene i utgangspunktet ikke er tilfeldig fordelt på utvalgene, har vi også beregnet Gini-koeffisienter på basis av personenes egne inntekt etter skatt. Tanken er at forskjellen mellom koeffisientene basert på henholdsvis egen inntekt og ekvivalentinntekt sier oss noe om hvordan husholdsdannelsen virker inn på ulikhetsmålet.

For alle de fire utvalgene i tabell 3.2 ser vi at Gini-koeffisienten er betraktelig lavere når grunnlaget er ekvivalentinntekt enn når grunnlaget er egen inntekt etter skatt. Men vi ser også at det er betydelige forskjeller mellom utvalgene. For begge inntektstypene er ulikheten større i frafallsgruppen enn i de andre gruppene. Vi ser også at det er i denne gruppen forskjellen mellom koeffisientene basert på egen inntekt og ekvivalentinntekt er lavest, nemlig 0,056. Den største differansen mellom Gini-koeffisientene finner vi blant personer hvis familie og hushold var identisk. Felles for disse to gruppene er at registerfamilie kunne vært/har blitt benyttet som økonomisk enhet.

For gruppen bestående av personer med hushold som økonomisk enhet dvs. de fra nettoutvalget med familie \neq hushold, er differansen 0,058, omlag den samme som for frafallet. Dette gir oss en indikasjon på at bruk av registerfamilie som økonomisk enhet ikke påvirker estimatet på Gini-koeffisienten på noen systematisk måte i forhold til om vi bruker husholdet.

Tabell 3.2: *Gini-koeffisienter med standardavvik for inntekt etter skatt og ekvivalentinntekt ($E=0,5$) for ulike utvalg av personer over 17 år. (IF-1997 uten tilleggsutvalg for næringsdrivende.)*

Personer	Personlig inntekt etter skatt		Ekvivalentinntekt		Dif-feranse	Prosent av hele utvalget
	Gini	Std.-avvik	Gini	Std.-avvik		
i nettoutvalget som hadde familie = hushold	0,338	0,009	0,254	0,008	0,084	55,8
i nettoutvalget som hadde familie \neq hushold	0,328	0,008	0,270	0,006	0,058	19,6
i frafallet	0,356	0,015	0,300	0,012	0,056	26,7
i frafallet + de i nettoutvalget som hadde hushold = familie	0,346	0,008	0,270	0,006	0,076	80,4

Det er imidlertid mulig å analysere overgangen fra egen inntekt til ekvivalentinntekt nærmere. Vi kan tenke oss at beregning av ekvivalentinntekten til en person er en prosess som starter med personens egen inntekt og ender med hans ekvivalentinntekt. I løpet av denne prosessen får husholdet personen lever i to effekter: *Hvilke* personer man deler hushold med virker inn, og hvor *mange* personer man deler hushold med virker inn. Dersom man deler hushold med personer med høy inntekt virker det positivt inn på ekvivalentinntekten. Denne effekten kan vi kalle inntektseffekten. Antallet personer i husholdet virker slik at jo flere personer i husholdet, jo lavere ekvivalentinntekt for personene som tilhører det. Denne effekten kan vi kalle størrelseseffekten. Vi vil i dette avsnittet undersøke hvordan husholdsdannelsen påvirker disse to effektene mht. estimerer på Gini-koeffisienten.

3.4.2 Inntektseffekten

I tabell 3.3 under har vi forsøkt å isolere denne "inntektseffekten" av husholdet/registerfamilien, ved å se på forskjellen mellom Gini-koeffisienter beregnet på grunnlag av egen inntekt etter skatt og på grunnlag av den samlede husholdsinntekten etter skatt. I begge tilfeller er det personer som er analyseenheten, når vi ser på samlet husholdsinntekt har vi gitt hver person inntekten til hele husholdet/registerfamilien han lever i. Dette tilsvarer en ekvivalentinntekt der husholdets størrelse ikke betyr noe, kun den samlede inntekten. Man kan si at husholdet er preget av fullstendige stordriftsfordeler.

Tabell 3.3: *Gini-koeffisienter med standardavvik for personlig inntekt etter skatt og samlet husholdsinntekt etter skatt ($E=0,5$) for ulike utvalg av hushold. (IF-1997 uten tilleggsutvalg for næringsdrivende.)*

Personer	Personlig inntekt etter skatt		Husholds-inntekt etter skatt		Differanse	N
	Gini	Std.-avvik	Gini	Std.-avvik		
i nettoutvalget som hadde familie = hushold	0,338	0,009	0,305	0,006	0.033	12825
i nettoutvalget som hadde familie \neq hushold	0,328	0,008	0,303	0,005	0.025	4667
i frafallet	0,356	0,015	0,374	0,011	-0.018	6356
i frafallet + de i nettoutvalget som hadde hushold = familie	0,346	0,008	0,334	0,006	0.012	19181

Igjen er det stor forskjell mellom underutvalgene, og igjen er det frafallet som skiller seg ut. Denne gruppen er den eneste med større ulikhet når samlet husholdsinntekt er grunnlaget i forhold til når grunnlaget er personlig inntekt. Årsaken kan være at enslige er overrepresentert i frafallet. Enslige vil naturligvis sitte med samme beløp enten man ser på personlig inntekt eller den samlede inntekten i husholdet. De som ikke er enslige får derimot ofte høyere husholdsinntekt enn egen inntekt. Siden enslige er overrepresentert, kan det føre til at ulikheten blir større i dette utvalget enn i de andre så lenge husholdsinntekt er grunnlaget.

For de to gruppene som tilhører nettoutvalget, er ulikheten med grunnlag i personlige inntekter større enn ulikheten med grunnlag i husholdsinntekter. Effekten er altså den motsatte av det vi fant i frafallsgruppen. I disse to gruppene er det relativt sett flere hushold med to inntekter og færre enslige. Dette bidrar til at det er relativt færre med lave husholdsinntekter. Vi kan anta at relativt få enslige, sammen med det faktum at en del kvinner fortsatt er deltidsansatte/hjemmeværende og dermed har lav/ingen egen inntekt, gjør at ulikheten blir lavere med husholdsinntekt som grunnlag.

3.4.3 Størrelseseffekten

I tabell 3.4 under ser vi nærmere på det vi kalte størrelseseffekten på ekvivalentinntekten. For alle utvalgene har vi igjen beregnet to Gini-koeffisienter. Den første tar utgangspunkt i den samlede inntekten (etter skatt) i husholdet personen lever i. Den andre tar utgangspunkt i

Tabell 3.4: *Gini-koeffisienter med standardavvik for samlet husholdsinntekt etter skatt og ekvivalentinntekt (etter skatt og med $E=0,5$) for ulike utvalg av personer. (IF-1997 uten tilleggsutvalg for næringsdrivende.)*

Personer	Husholds- inntekt etter skatt		Ekvivalentinntekt		Dif- feranse	N
	Gini	Std.-avvik	Gini	Std.-avvik		
i nettoutvalget som hadde familie = hushold	0,305	0,006	0,254	0,008	0,051	12825
i nettoutvalget som hadde familie \neq hushold	0,303	0,005	0,270	0,006	0,033	4667
i frafallet	0,374	0,011	0,300	0,012	0,074	6356
i frafallet + de i nettoutvalget som hadde hushold = familie	0,334	0,006	0,270	0,006	0,064	19181

personenes ekvivalentinntekt. Ved å sammenlikne disse to inntektene, ønsker vi å finne ut om det er forskjeller mellom utvalgene og dermed forskjeller mellom hvordan husholdet blir bestemt.

Også i denne sammenhengen er det frafallet som tar oppmerksomheten, med den største differansen mellom Gini-koeffisientene basert på de to inntektsvariablene. Sannsynligvis er det sammensetningen av denne gruppen, med overrepresentasjon av enslige, som gir seg uttrykk på denne måten. Vi så i forrige tabell (3.3) at ulikheten ble stor blant frafallet når vi så på husholdningsinntekter. Når vi så går over til ekvivalentinntekt igjen, faller ulikheten mest i denne gruppen. De mange enslige vil ha lik ekvivalentinntekt som husholdsinntekt, mens blant hushold bestående av flere personer er ekvivalentinntekten naturligvis lavere enn husholdsinntekten. Resultatet er at vi får en utjevning av inntektene, og lavere Gini-koeffisient. Poenget med de to siste tabellene var altså å se om vi kunne finne ut noe mer om familie og hushold som økonomisk enhet enn det som var mulig ved kun å sammenlikne personlig inntekt og ekvivalentinntekt (som i tabell 3.2). Som vi så, skiller personer som tilhører frafallet seg mer ut enn man får inntrykk av ved kun å se på tallene i tabell 3.2. I forbindelse med vår problemstilling om muligheter for bruk av registerfamilie som økonomisk enhet kan vi slutte to ting:

For det første er det størst ulikhet i inntektsfordelingen målt ved Gini-koeffisienten blant personer som ikke svarer. Hvorvidt dette skyldes at fordelingen virkelig er preget av større ulikhet i denne gruppen (f.eks. dersom det er de ”rike” og de ”fattige” som ikke svarer) eller det er begrensningene i husholdsopplysninger (som gir for mange enslige), er usikkert. Antageligvis er begge effekter tilstede.

For det andre ser det ikke ut til at bruk av registerfamilie istedenfor hushold som økonomisk enhet gir systematiske utslag på estimer på inntektsfordelingen i den ene eller andre retningen. De to gruppene med registerfamilie som økonomisk enhet skiller seg ikke fra gruppen med hushold som økonomisk enhet, hverken når det gjelder nivået på Gini-koeffisientene eller differansene.

Et viktig poeng, som så langt ikke har vært diskutert, er hvilke effekter det vil gi på estimatene av Gini-koeffisienten i hele befolkningen dersom vi kun hadde brukt registerfamilie som økonomisk enhet. Vi kan ikke si noe sikkert om dette fordi vi ikke vet

hvordan bruk av registerfamilie ville slå ut for gruppen som svarte og hadde hushold \neq registerfamilie.

Men noe kan vi lese ut av tabell 3.2. Ulikheten er større i gruppen med avvik mellom hushold og registerfamilie enn i gruppen med hushold identisk med registerfamilie, så lenge ekvivalentinntekt er grunnlaget. Dersom ulikheten måles ut fra personlig inntekt blir resultatet motsatt; større ulikhet for gruppen med familie=hushold og lavere for gruppen med familie \neq hushold. Utjevningseffekten av å bruke ekvivalentinntekt ser altså ut til å være større dersom familie brukes som økonomisk enhet enn dersom husholdet brukes. Dette kan tyde på at bruk av registerfamilie som økonomisk enhet ville gi lavere estimater på ulikheten dersom vi tar for oss hele befolkningen.

Vi kan også dra nytte av det vi vet om sammensetningen av gruppen med familie \neq hushold. Fra forrige kapittel vet vi bl.a at vi har en overrepresentasjon av personer i 20-årene og av personer mellom 40 og 60 år i denne gruppen. Det er rimelig å anta at personer i 20-årene generelt har noe lavere inntekt enn gjennomsnittet, og at de mellom 40 og 60 år har inntekt noe over snittet. Dette gir større ulikhet innad i denne gruppen enn det vi ville fått uten denne overrepresentasjonen. Hvorvidt denne ulikheten innad i gruppen smitter over på helheten kan vi ikke vite.

Selv om vi ikke kan si noe sikkert, tyder altså tallene i tabell 3.2, og det vi vet om hvordan denne gruppen er sammensatt, på at registerfamilie som økonomisk enhet kan gi noe for lave estimater på ulikheten.

3.5 Lav inntekt

Vi har nå sett litt på graden av ulikhet i fordelingene, men en Gini-koeffisient sier ikke alt om inntektsfordelingen i et utvalg/populasjon. Skal inntektsfordelingen studeres, er det viktig å se om de ulike utvalgene også gir ulike nivåer på andelen med lav inntekt. Lav inntekt defineres ofte i denne sammenhengen som inntekt under 50 prosent av medianinntekten. I tabell 3.5 har vi regnet ut andelen med lav inntekt etter denne definisjonen i de ulike utvalgene, og på basis av de to inntektsvariablene inntekt etter skatt og inntekt etter skatt pr. forbruksenhet. Som med Gini-koeffisienten, har det en utjevne effekt å se på ekvivalentinntekt fremfor personlig inntekt. Andelen med lav inntekt varierer mellom 14,7 og 20,7 prosent når vi tar utgangspunkt i personenes inntekt etter skatt. Ser vi på inntekt etter skatt pr. forbruksenhet, blir andelen lavere, men vi ser fortsatt tydelige forskjeller mellom utvalgene, spesielt mellom de to delene av nettoutvalget. Blant personer med familie = hushold er det 6,6 prosent med lav inntekt, mot så mye som 14,3 prosent blant de som hadde avvik mellom intervju- og registeropplysninger. Dette til tross for at begge gruppene hadde forholdsvis like tall for andel med lav inntekt når inntekt etter skatt lå til grunn. Det kan derfor virke som om det er dannelsen av hushold, med tilhørende ekvivalentinntekt som gir forskjellig utslag.

Tallene i tabell 3.5 illustrerer også hvor viktig det er å beregne en ekvivalentinntekt på bakgrunn av den samlede inntekten i husholdet en person lever i. Dersom vi måler andelen

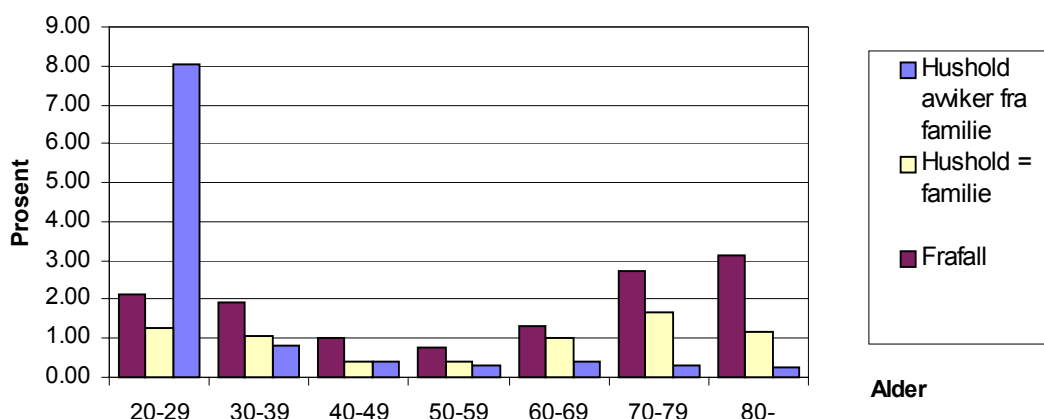
Tabell 3.5: *Andel personer over 17 år med inntekt (inntekt etter skatt og inntekt pr. forbruksenhet etter skatt) lavere enn 50 prosent av medianinntekten etter utvalg.*

Personer	Andel personer med inntekt lavere enn 50 % av medianen		50 % av medianinntekt	
	Inntekt etter skatt	Inntekt etter skatt pr. forbr. Enhet.	Innt. etter skatt	Innt. etter skatt pr. forbr. enhet
i nettoutvalget som hadde familie = hushold	17,0	6,6	72 976	89 465
i nettoutvalget som hadde familie ≠ hushold	20,7	14,3	71 267	98 880
i frafallet	14,5	12,5	62 033	89 310
i hele utvalget	17,1	9,9	69 531	91 385

personer med lav inntekt på bakgrunn av egen inntekt etter skatt, gir dette omlag 17 prosent med lav inntekt. En av årsakene er at hjemmeværende og deltidsarbeidende har ingen eller lav egen inntekt, og dermed ofte havner under 50 prosent av medianinntekten. Dette behøver ikke bety at de har en lav økonomisk velferd, ofte kompenseres deres egne lave inntekt av at andre i husholdet har en normal inntekt. Og det er nettopp det vi kontrollerer for når vi beregner ekvivalentinntekten.

For å få et skarpere bilde av hvem de med lav inntekt er, ser vi i Figur 3.2 nærmere på alderssammensetningen blant de med lav inntekt. Søylene angir hvor mange prosent av de med lav inntekt som tilhører de forskjellige aldersgrupper og utvalg. Det fremkommer tydelig at personer fra to av gruppene er forholdsvis likt fordelt aldersmessig. Både blant personer med lav ekvivalentinntekt fra frafallet og blant de fra nettoutvalget som hadde familie = hushold, var det en viss overvekt av eldre. Da er det helt motsatt i gruppen av personer med lav inntekt blant de som svarte og hadde *avvik* mellom intervju og register. Den største andelen av disse er mellom 20 og 29 år. Dette tyder på at med husholdet basert på registeropplysninger om familie, fanges ikke unge med lav ekvivalentinntekt opp. Sannsynligvis er dette de borteboende studentene vi finner igjen her. Grafen viser tydelig at aldersfordelingen blant personer med lav ekvivalentinntekt ville vært en helt annen dersom registerdata hadde blitt brukt direkte for å danne hushold. Først og fremst hadde vi ikke oppdaget at unge mennesker er relativt sterkt representert i lavinntektsgruppen. Dernest kunne konklusjonen blitt at fattigdom stort sett var et eldreproblem. Figuren illustrerer hvor viktig

Figur 3.2: *Personer med ekvivalentinntekt under 50 prosent av medianen, etter alder og utvalg.*



det er å identifisere borteboende studenter, dersom vi bruker registeropplysninger om familie som økonomisk enhet.

I tabell 3.5 ser vi igjen på de tre delutvalgene, men har nå delt personene inn i tre aldersgrupper. I den nederste raden er alle delutvalgene slått sammen. Det ser her ut til å være en tendens til at ulikheten blir mindre jo eldre personene blir.

For den yngste gruppen, 20-34 år, er ulikheten størst blant personene som ikke svarte, og lavest blant de respondentene som ikke hadde avvik mellom registerfamilie og hushold. Som vi har sett er avvik mellom registerfamilie og virkelig hushold spesielt vanlig i denne aldersgruppen. Høy Gini-koeffisient blant personer fra avvik-gruppen er ser ut til å trekke opp koeffisienten for alle unge sett under ett. Vi får her en indikasjon på at det, særlig for yngre, er viktig å få med de som har avvik mellom hushold og registerfamilie. Standardavvikene er imidlertid høye fordi det blir forholdsvis få observasjoner i hver gruppe. Dette gjør det vanskelig å si noe sikkert.

Når man sammenlikner Gini-koeffisienter fra forskjellige grupper, gir det oss et mål på ulikhet innad i hver gruppe. Men det sier oss ikke noe om forholdet i inntektsnivå mellom gruppene. Dermed vet vi heller ikke hvilken effekt ulikheten i en gruppe har på ulikheten i hele befolkningen. Et eksempel vil klargjøre dette poenget. Vi kan tenke oss at vi har en befolkning som kan deles i to grupper. Hver gruppe for seg har en Gini-koeffisient på f.eks 0,25. Men dette trenger ikke bety at Gini-koeffisienten blir 0,25 når vi slår sammen de to gruppene og ser på hele befolkningen. Dersom inntektene i den ene gruppen har et lavt nivå sammenliknet med inntektene i den andre gruppen, kan Gini-koeffisienten for hele befolkningen sett under ett bli høyere. Generelt kan vi altså ikke si noe om hvordan hver gruppe bidrar til den totale ulikheten i befolkningen ved å sammenlikne Gini-koeffisienter for de enkelte undergruppene.

Tabell 3.5: *Gini-koeffisienter for personers ekvivalentinntekt ($E=0,5$) etter aldersgrupper. Tilhørende standardavvik i parentes.*

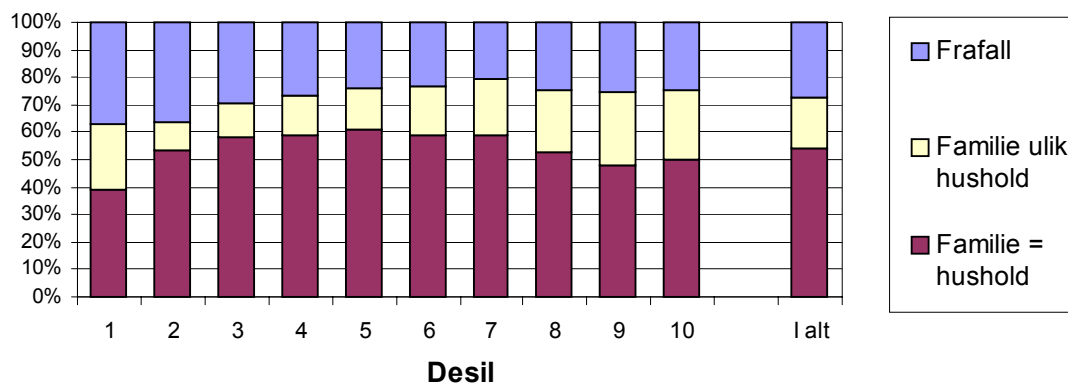
Utvalg	20-34 år		35-66 år		67 år og over	
Hushold = familie	0,255	(0,022)	0,237	(0,008)	0,230	(0,020)
Hushold \neq familie	0,299	(0,009)	0,219	(0,007)	0,260	(0,022)
Frafall	0,300	(0,026)	0,287	(0,018)	0,211	(0,014)
Alle	0,278	(0,013)	0,248	(0,006)	0,228	(0,013)

3.6 Desilfordelinger

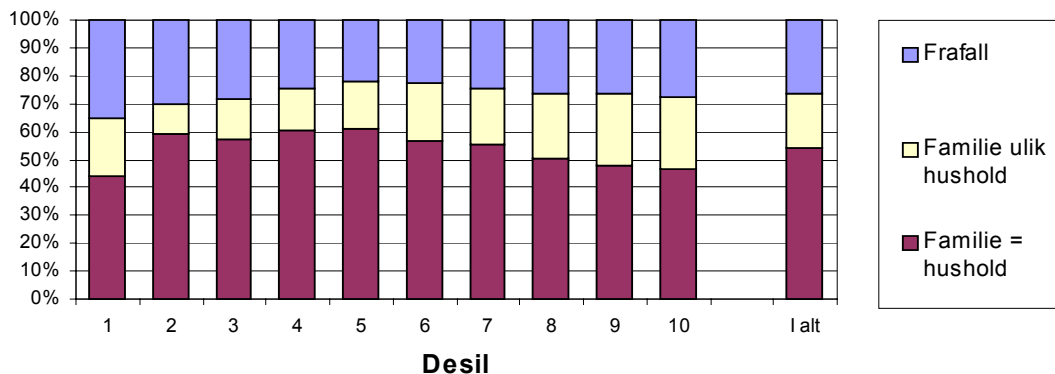
Med figur 3.3 har vi forsøkt å rette på dette. Der er alle husholdene i utvalget delt opp etter desiler. Desilene fremkommer ved å dele inn husholdene i ti like store grupper der den første inneholder den tidelen som har lavest inntekt, den andre inneholder tidelen med nest lavest inntekt osv. Desil nr 10 består således av den rikeste tidelen av husholdene i utvalget. Videre er hver desil delt i tre avhengig av hvilket av delutvalgene husholdene i denne desilen tilhører. F.eks. kommer omlag 40 prosent av husholdene i første desil fra respondentene som hadde familie = hushold, 37 prosent av husholdene kommer fra frafallet, mens de resterende 23 prosentene var fra hushold som svarte og hadde familie \neq hushold. Den siste søylen i diagrammet er fordelingen av hushold i hele utvalget. Frafallet er altså betydelig overrepresentert i den laveste desilen.

Sammensetningen av hushold varierer noe fra desil til desil. Hushold som ikke svarte på intervjuet, dvs. frafallsgruppen, er sterkere representert i de nederste og øverste desilene enn på midten. Dette kan tyde på at frafallsgruppen bidrar til større ulikhet når hele utvalget er sett under ett. Hushold som var blant respondentene og hadde avvik mellom registerfamilie og hushold, er også noe sterkere representert i de første og siste desilene enn i de midterste. Vi kan anta at også denne gruppen er med på å trekke ulikheten i hele utvalget opp. Den siste gruppen er også den største, nemlig de som svarte og hadde hushold = familie. Disse var sterkest representert i de midterste desilene og bidro til å jevne ut inntektsfordelingen i hele utvalget.

Figur 3.3: Fordeling av hushold etter desiler av ekvivalentinntekt pr. forbruksenhet.



Figur 3.4: Fordeling av personer over 17 år etter desiler av ekvivalentinntekt.



Dette forholdet mellom gruppene som figur 3.3 har avdekket, stemmer godt overens med informasjonen i Gini-koeffisientene i de tre delutvalgene. De viste at både frafallet og hushold som hadde avvik mellom familie og hushold, hadde høyere Gini-koeffisient enn gruppen av hushold som ikke hadde slikt avvik. Og ulikheten i hele utvalget sett under ett, er noe høyere enn for gruppen som ikke hadde avvik, men ikke så høy som frafall- og avvikgruppene. Ved hjelp av figur 3.3 har vi fått klarlagt at selve fordelingene av inntekt i stor grad er overlappende. Dette ser vi av at alle desilene inneholder representanter for alle gruppene hushold. Dersom inntektene i f.eks frafallet hadde vært på et jevnt over lavere nivå enn i de andre gruppene, ville de første desilene vært dominert av denne gruppen. Slik er det ikke, og Gini-koeffisientene fra tabell 3.3 er dermed mulige å sammenlikne fordi inntektene i hver av gruppene stort sett fordeler seg over samme intervall.

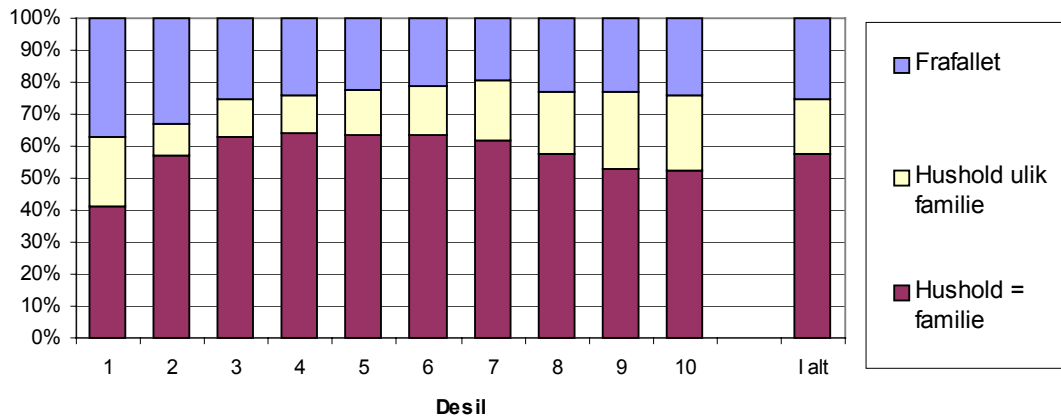
Figur 3.4 er helt lik figur 3.3 bortsett fra at vi nå har personer som enhet og ikke hushold. Det er personers økonomiske velferd vi ønsker å konsentrere oppmerksomheten om. Mønsteret er imidlertid svært likt det som kom frem for hushold, og årsakene sannsynligvis de samme.

Med hensyn til figurene 3.3 og 3.4 må det bemerkes at disse ikke beskriver den virkelige desilfordelingen, siden personene eller husholdene ikke er vektet på noen måte. Likevel kan de være nyttige som en beskrivelse av kvaliteten i datagrunnlaget.

For å rette på problemet med vektingen, har vi laget figur 3.5. Her er programmet Uroteng¹ blitt brukt for å finne desilgrensene. Fordelen med å bruke dette programmet, er at det tar

¹ Program (av Rolf Aaberge og Tom Wennemo, SSB) som regner ut Gini-koeffisienter, og to andre ulikhetsmål, Lorentzkurver og kvintiler samt dekomponerer disse etter inntektskilder.

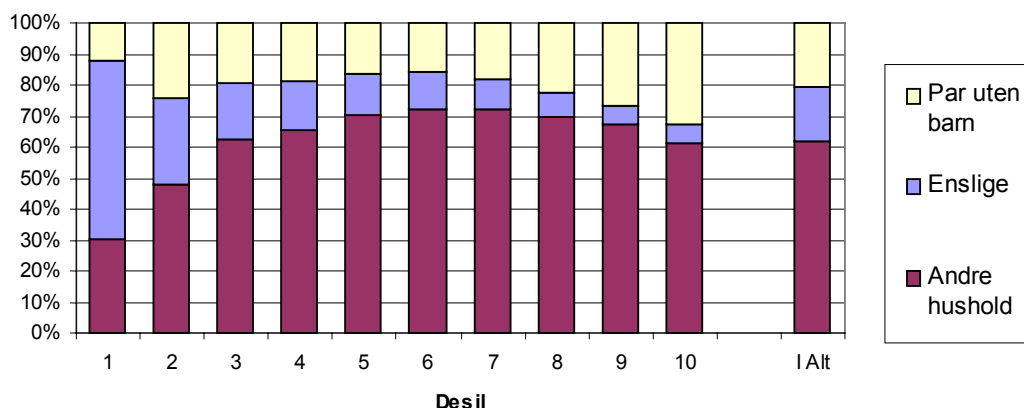
Figur 3.5: Fordeling av personer etter delutvalg og desiler av ekvivalentinntekt etter skatt. Alle personer. Vektet for å motvirke ulik trekkesannsynlighet.



hensyn til vektene. Som vi ser, er det samme tendens som i de to foregående figurene til at frafallet og respondentene med hushold ulik familie er underrepresentert i de midterste desilene og til en viss grad overrepresentert i de ytterste. Dette er en bekreftelse på at de to gruppene trekker opp ulikheten i totalen, og at gruppen som har hushold = familie alene gir et skjevt bilde av den reelle inntektsfordelingen i samfunnet.

Vi vet at det største problemet ved å bruke registerfamilie som økonomisk enhet, er at samboere uten barn blir betraktet som enslige. Dermed får de redusert sin økonomiske velferd sammenliknet med den de ville fått dersom de hadde blitt regnet som samboere. I figur 3.6 nedenfor har vi laget en desilfordeling tilsvarende de vi allerede har sett på. Men denne gangen ser vi på fordelingen av husholdstyper innad i hver desil. Vi har konsentrert oss om gruppene enslige, par uten barn og andre hushold. Dette er gjort for å fokusere på forskjellene

Figur 3.6: Fordeling av personer etter husholdstype og desiler av ekvivalentinntekt.



mellom enslige og par uten barn. Det kommer tydelig frem at enslige er forholdsvis sterkt overrepresentert i de første desilene. I desilen bestående av personene med de laveste ekvivalentinntektene, er omlag 57 prosent enslige. Dette tallet synker ettersom vi går oppover i fordelingen til omlag 6 prosent i de to ”rikeste” desilene.

Når det gjelder par uten barn, er denne gruppen hushold mye jevnere fordelt over desilene. Dette indikerer at for mange enslige, som vi får ved å bruke familien som økonomisk enhet, vil trekke opp ulikheten i fordeling av ekvivalentinntekt. Dersom vi klarer å danne samboerpar vil sannsynligvis ulikheten bli mindre, siden par uten barn er fordelt noe jevnere over desilene. Uansett viser figur 3.6 at det er viktig å finne samboere uten barn, kanskje først og fremst fordi tallet på enslige da blir mer korrekt.

3.7 Samboere og enslige.

I dette avsnittet vil vi undersøke i hvilken grad Gini-koeffisienten blir påvirket av om vi klarer å identifiserer samboere uten barn. I tabell 3.6 har vi beregnet Gini-koeffisienter for tre ulike utvalg; hele utvalget, kun respondentene og personer som er mellom 20 og 35 år. Den siste gruppen er med fordi det var blant personer i denne alderen at avviket mellom registerfamilie og hushold var mest vanlig. For de som i intervju oppga at de var samboere, men i folkeregisteret var oppført som enslige (eller tilhørende en annen familie), danner vi to sett ekvivalentinntekter. Et der de betraktes som enslige, og et der de betraktes som samboere. I datasettet er det 1028 personer som faller i kategorien samboere uten barn.

Som vi ser i tabell 3.6, har det forholdsvis liten effekt på Gini-koeffisienten for hele utvalget om samboerne blir identifisert og får ekvivalentinntekt som samboere, eller om de regnes for enslige. Koeffisienten er 0,002 større når de får sin ekvivalentinntekt beregnet som enslige og dette er ikke en signifikant forskjell. Blant respondentene er effekten 0,003 for høy Gini-koeffisient når samboere betraktes som enslige. Det gir et riktigere inntrykk å måle effekten i denne gruppen enn i hele utvalget. Det er bare blant respondentene vi kan finne samboere uten barn. Hele utvalget inneholder jo også personer som ikke svarte på intervjuet, og samboere uten barn blant disse vil være representert med ekvivalentinntekt som enslig uansett. Når vi så ser på effekten i hele utvalget av at samboere *blant respondentene* regnes som enslige, blir det for få samboere i forhold til det totale antallet personer, og tilsvarende

Tabell 3.6: Gini-koeffisienter i fordelingen av ekvivalentinntekt ($E=0,5$), for personer.

Grunnlag for ekvivalentinntekt	Utvalg					
	Hele utvalget		Kun respondentene		Personer mellom 20 og 35 som svarte	
Hushold (samboere som par)	0,265	(0,005)	0,255	(0,005)	0,267	(0,014)
Familie (Samboere som enslige)	0,267	(0,005)	0,258	(0,005)	0,275	(0,015)

reduisert effekt. Måler vi derimot effekten blant respondentene stemmer forholdet mellom samboere uten barn og totalt antall personer bedre, og vi måler den fulle effekten.

Blant personer mellom 20 og 35 år som svarte, var effekten av å beregne samboere som enslige betraktelig større, nemlig 0,008. Dette gir nok en indikasjon på at problemet med samboerne er størst innenfor denne aldersgruppen. Her var også standardavviket en god del større, noe som delvis skyldes færre observasjoner. Vi finner derfor ingen signifikante effekter i denne gruppen heller.

Hvorvidt samboere ble regnet som enslige eller som par hadde altså forholdsvis liten effekt på Gini-koeffisienten. Vi vil nå se på andelen personer med lav inntekt, og hvilken effekt identifisering av samboere har på denne.

Det er ikke gitt at andelen fattige vil stige dersom samboere uten barn blir regnet som enslige. Riktignok vil personer som lever i samboerskap der begge har inntekt stort sett få lavere ekvivalentinntekt dersom de blir regnet som enslige. Unntaket er personer som har vesentlig større inntekt enn sin samboer, disse vil få økt ekvivalentinntekt av å bli regnet som enslig. Selv om vi antar at de fleste får redusert sin ekvivalentinntekt, blir det ikke automatisk flere fattige. Medianinntekten i samfunnet vil også falle, og med den grensen for lav inntekt. Isolert sett betyr det færre personer med ekvivalentinntekt under 50 prosent av medianen.

I tabell 3.7 tar vi for oss de samme tre utvalgene som vi gjorde i tabell 3.6, og regner ut andelen fattige både dersom samboere blir regnet som et par i beregningen av ekvivalentinntekt og dersom de blir regnet som enslige. Forskjellen er ikke særlig stor når vi ser på hele utvalget, men som nevnt er det bedre å se på forskjellen blant respondentene. Og i

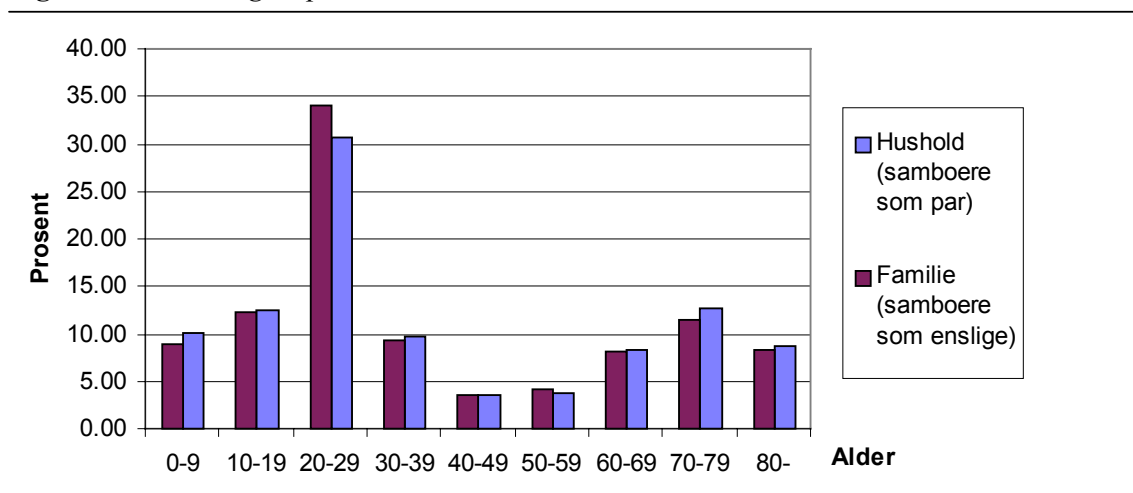
Tabell 3.7: *Andel personer med lav ekvivalentinntekt (under 50 % av medianen), etter ulike utvalg og grunnlag for ekvivalentinntekt.*

Grunnlag for ekvivalentinntekt	Utvalg		
	Hele utvalget	Kun respondentene	Personer mellom 20 og 35 som svarte
Hushold (samboere som par)	8,36	7,02	11,23
Familie (Samboere som enslige)	8,69	7,59	13,14

denne gruppen er det en noe tydeligere forskjell. Dersom samboere ikke identifiseres og dermed blir regnet som enslige, blir den estimerte andelen med lav inntekt omlag 0,6 prosentpoeng høyere enn den skulle vært. I aldersgruppen 20-35 år øker andelen med lav inntekt med nesten 2 prosentpoeng fra 11,23 til 13,14 dersom samboerne blir regnet som enslige. Det vil altså være en viss fordel å identifisere samboere uten barn dersom vi skal få gode estimater på andelen personer med lav inntekt. Men det er først når vi velger ut aldersgruppen der samboerskap er mest vanlig at forskjellene blir betydelige.

I figur 3.7 ser vi nærmere på aldersfordelingen av personer med lav inntekt. Dersom vi bruker familie som økonomisk enhet (dvs. lar samboere være enslige) i beregning av ekvivalentinntekt, får dette ikke store konsekvenser i forhold til om vi hadde brukt hushold (dvs. regnet samboere som par). Den største forskjellen finner vi i aldersgruppen 20-29 år, som blir noe sterkere representert blant personer med lav inntekt ved bruk av familie som økonomisk enhet. Andelen av de fattige som tilhører denne aldersgruppen, øker fra 30,7 til

Figur 3.7: *Fordeling av personer med lav inntekt etter alder. Prosent.*



34,2 prosent. Siden personer i 20-årene øker sin andel, må nødvendigvis en eller flere andre aldersgrupper bli redusert. Reduksjonen er jevnt fordelt over nesten alle de andre aldersgruppene, noe som er med på å understreke hvor isolert effekten av å bruke familie som økonomisk enhet er.

Hvordan går det med fordelingen av personer med lav inntekt i forhold til husholdstilknytning? Som det fremgår av tabell 3.8, gir familie som økonomisk enhet forholdsvis store utslag på denne variabelen. Særlig er det gruppen par uten barn, der samboerne naturligvis hører hjemme, som blir påvirket. Med hushold som økonomisk enhet tilhører 11,8 prosent av personene med lav inntekt denne gruppen. Men dersom vi bruker familie som økonomisk enhet, blir dette tallet nesten halvert og bare 6,7 prosent tilhører gruppen par uten barn. Familie som økonomisk enhet gir imidlertid en betydelig økning av enslige med lav inntekt. De andre gruppene blir i liten grad påvirket av hvordan samboerne blir behandlet. I motsetning til de andre dimensjonene vi har sett på i dette avsnittet, finner vi altså at valg av økonomisk enhet gir vesentlige forskjeller når det gjelder husholdstilknytning blant personer med lav inntekt.

Til sist i dette avsnittet vil vi undersøke om familie som økonomisk enhet gir utslag på fordelingen av personer med lav inntekt med hensyn til kjønn. Tabell 3.9 viser at vi får en liten økning i andelen kvinner blant de med lav inntekt. Forskjellen er imidlertid ikke stor, men dette skyldes sannsynligvis at det er få samboere uten barn i forhold til det totale antall personer med lav inntekt.

Dersom vi ser nærmere på de personene som faller under grensen for lav inntekt når familie er økonomisk enhet men ikke når vi bruker hushold, finner vi imidlertid store kjønnsforskjeller.

Tabell 3.8: *Fordeling av personer med lav ekvivalentinntekt etter hushold. Prosent.*

Husholdstype	Økonomisk enhet i ekvivalentinntekt	
	Hushold (samboere som par)	Familie (samboere som enslige)
Enslig	57,2	65,3
Par uten barn	11,8	6,7
Par med barn	14,5	13,1
Andre	16,6	15,0
Tilsammen	100,1	100,1

Tabell 3.9: Fordeling av personer med lav ekvivalentinntekt etter kjønn. Prosent.

Kjønn	Økonomisk enhet i ekvivalentinntekt	
	Hushold (samboere som par)	Familie (samboere som enslige)
Kvinner	60,8	61,6
Menn	39,2	38,4
Tilsammen	100,0	100,0

Omlag 70 prosent av disse personene er nemlig kvinner. Dette kan skyldes at kvinner generelt har et lavere lønnsnivå enn menn, og at det dermed er de som ”tjener” mest på at husholdet blir brukt som økonomisk enhet.

3.8 Konklusjon

I dette kapitlet har vi forsøkt å finne eventuelle utslag på ulike fordelingsmål av å erstatte husholdet som økonomisk enhet med familien. Dette er gjort for å få en oversikt over hvilke muligheter og begrensninger som ligger i bruk av registerdata som grunnlag for fordelingsstudier. Vi har sett på tre forskjellige fordelingsmål; Gini-koeffisienten, en fattigdomsgrense ved 50 prosent av medianinntekten og fordeling etter desiler.

Analyser ved hjelp av Gini-koeffisienter antydnet at bruk av familie som økonomisk enhet både kan trekke i retning av høyere og lavere observert ulikhet i samfunnet. For mange enslige i forhold til par kan gi en noe større observert ulikhet, mens for få studenthushold vil trekke i retning av lavere observert ulikhet.

Familieopplysninger som grunnlag for ekvivalentinntekt gir også utslag i gruppen av personer med lav inntekt. Siden borteboende studenter ikke identifiseres, domineres lavinntektsgruppen av eldre personer, og vi får ikke med at også unge i virkeligheten er forholdsvis sterkt representert blant de med lav inntekt.

Desilfordelingene viste først og fremst to generelle poenger i forbindelse med registerdata og fordeling. I forhold til personer som har registerfamilie=hushold, vil personene som har avvik mellom registerfamilie og hushold bidra til å øke den observerte ulikheten i økonomien. Dette

kan tyde på at å bruke familieopplysninger direkte vil gi for lav ulikhet. Det andre poenget er at for mange enslige særlig vil slå ut i at disse blir overrepresentert i de lavere desilene.

Vi forsøkte også å lage en oversikt over eventuelle konsekvenser av at samboere uten barn blir identifisert og dermed ikke blir behandlet som enslige ved beregning av ekvivalentinntekt. Generelt viste resultatene at en slik identifisering får liten betydning så lenge vi fokuserer på hele befolkningen sett under ett. Dersom vi ser på mindre deler av befolkningen, f.eks. aldersgrupper eller typer hushold, blir konsekvensene derimot merkbare. Hovedresultatene er at både ulikheten og antallet med lav inntekt øker blant unge mennesker når samboere blir betraktet som enslige. Videre blir enslige betydelig sterkere representert blant personer med lav inntekt enn de skulle vært.

Et annet og litt overraskende funn i dette kapitlet, var hvor stor forskjell det ble på de estimerte Gini-koeffisientene når vi byttet ut datasettets originalvekter med vektorer som kun reflekterer trekkesannsynlighet. Trekkesannsynlighetsvektene ga noe større estimert ulikhet i utvalget. En mulig årsak kan være kalibreringen av vektene som bl.a. medfører at svært høye inntekter er blitt gitt lave originale vektorer for å få inntektene i utvalget til å stemme med kjente totaler.

3.8.1 Mulige løsninger

Hva kan gjøres for å bøte på disse problemene? Når det gjelder borteboende studenter, er disse forholdsvis lette å finne ved hjelp av opplysninger fra Lånekassen om hvem som får borteboerstipend. Men her bør det nevnes at mange studenter får økonomisk hjelp fra foreldrene slik at deres økonomiske velferd til en viss grad avhenger av foreldrenes. Derfor er det et valg om man vil betrakte studenter som egne hushold eller som del av foreldrenes hushold. Studenter er uansett ganske problematiske i fordelingssammenhenger i og med at mange lever på lån og stipend fra Lånekassen. Det er bare stipenddelen som kommer med i klassifiseringsvariabelen 'inntekt etter skatt', slik at denne variabelen i liten grad reflekterer den reelle økonomiske situasjonen. I inntektsstatistikker kommer derfor studenter ofte uforholdsmessig dårlig ut, og kan lett forårsake skjevheter. En måte å løse dette problemet på kan være å la studielån og avdrag på studielån inngå i inntekten (med hhv. positivt og negativt fortegn. Denne løsningen brukes i Sverige, Epland (1998-c). En annen utvei er å presentere inntektsstatistikk for hushold/personer i to omganger; en inkludert og en uten studenter. Se f.eks. Epland (1998-a.)

For samboere uten barn finnes det ingen registre som kan rette opp feilen vi gjør ved å benytte familie som økonomisk enhet. I neste kapittel ser vi på mulighetene for å utvikle en modell som kan predikere hvem som er samboere og hvem som er enslige ut fra registeropplysninger.

4. Modell for å finne samboere gjennom registeropplysninger

Et viktig formål med dette prosjektet er å undersøke om det er mulig å identifisere samboere uten felles barn gjennom felles adresse (og leilighetsnummer hvis tilgjengelig) hentet fra Det sentrale folkeregisteret, og kjennetegn ved individet som alder, utdanning, eller andre opplysninger som er gitt i datagrunnlaget for IF-97. Poenget er altså å finne en ”nøkkel” som forteller oss hvem som er samboere med hvem når vi har registeropplysningene om personene som er bosatt på en gitt adresse. Dersom man bor i blokk, har man samme adresse som mange andre, slik at felles adresse ikke er nok til å finne ut hvem som deler hushold.

I kontrast til virkeligheten, vil så å si alle som har felles adresse i utvalget til IF-97, tilhøre samme hushold. Årsaken til dette er måten personene har kommet med i utvalget på. Først er en person trukket ut tilfeldig. Deretter er denne personen, som kalles kontaktperson, intervjuet om sitt hushold, og dersom det er flere personer i dette husholdet vil hver av disse også få sin egen personfil i datamaterialet. Dersom kontaktpersonen lever alene, vil man gå videre for å intervju neste kontaktperson. For å få med to personer som har identisk adresse men ikke deler hushold, må to kontaktpersoner ha identisk adresse. Sannsynligheten for at dette skjer er relativt liten. Vi vil derfor, i utvalget, nesten ikke finne noen personer som har felles adresse men ikke lever i samme hushold. Adresseopplysninger vil da være nok til at vi kan finne samboere i utvalget. Men slik er det ikke i virkeligheten. Det vil derfor, på et senere tidspunkt, bli aktuelt å bruke et fullstendig adresseregister for å få testet hvor bra modellen vår blir.

Vi må antagelig i første omgang konsentrere oss om å finne affiniteter (”nytte”) ved hjelp av en multinomisk logitmodell. Denne modellen brukes ofte når den avhengige variabelen er diskret, som f.eks. samboer eller ikke samboer. Disse affinitetene sier noe om hvor stor tilbøyelighet en person har til hvert enkelt utfall. Det som blir estimert, er ulike parametre som vil gi et estimat på hvor stor sannsynlighet en person har for de ulike utfallene (samboer, enslig) som funksjon av en vektor av uavhengige, observerbare variable. For å ”forklare” variasjonen i den avhengige variabelen vil vi benytte personers alder og utdanningsnivå, men også forsøke med andre variable som det finnes opplysninger om i datagrunnlaget, f.eks. inntekt.

4.1 Populasjon og utvalg

Vi plukker ut alle menn (eller kvinner) fra datafilene, som er blant respondentene og har svart at de er enslige eller lever som samboer uten felles barn. Som tidligere forklart er det først og fremst denne gruppen som byr på problemer mht. bruk av registerfamilie som økonomisk enhet (bortsett fra borteboende studenter). Deretter definerer vi alternative grupper som individene i denne populasjonen kan befinne seg i. Hovedgruppene er enslig og samboer. Men det kan være en fordel å dele opp gruppen samboer i flere undergrupper kjennetegnet av egenskaper ved den kvinnen mannen bor sammen med. Disse egenskapene kan f.eks. være alder og utdannelsesnivå. Ved å dele kvinnene inn i tre aldersgrupper og tre utdannelsesgrupper får vi ni muligheter som vist i tabell 4.1.

Tabell 4.1. *Eksempel på gruppeinndeling og -nummerering av kvinnens alder og utdannelsesnivå.*

Kvinnens utdannelsesnivå	Kvinnens alder		
	18-30 år	30-45 år	45 og over
Kun grunnskole	1	2	3
Videregående skole	4	5	6
Universitets- eller høyskolenivå	7	8	9

Grunnen til at vi deler inn i grupper på denne måten, er at dette vil gi bedre resultater når vi skal finne ut hvem som er samboer med hvem i tilfeller med mange kombinasjonsmuligheter. Vi kan tenke oss at en mann som er registrert som enslig bor på samme adresse som et antall kvinner som også er registrert som enslige, og dermed er kandidater til å være samboer med den uttrukne mannen. Da hjelper det lite å vite om mannen har høy affinitet for å være samboer så lenge vi ikke vet hvem av kvinnene vi skal velge som samboer med ham. Men takket være gruppeinndelingen har vi nå også estimater for mannens affinitet mht. hvilken gruppe samboeren skal tilhøre. Dermed har vi større sjanse for å velge riktig kvinne, eller ihvertfall en som tilhører riktig gruppe, og dermed blir feilen mindre enn den kunne blitt dersom vi hadde benyttet en tilfeldig trukket kvinne som samboer til mannen.

Det er ikke sikkert at så mange som ni grupper er det beste. Har vi for mange grupper blir antall observasjoner av hver gruppe relativt få, og estimatene kan bli svært usikre. Velger vi for få grupper, kan resultatet bli at vi ikke treffer godt nok på egenskapene til kvinnen i det

”syntetiske” samboerskapet. Vi må heller ikke glemme at det ikke spiller så stor rolle om vi finner akkurat de personene som er samboere i virkeligheten. Det som er avgjørende, er at vi får et best mulig grunnlag for å estimere fordelingen av inntekt og inntektsulikhet basert på registerdata.

4.2 Teori

I vårt system kan altså mannen ”velge” en av 10 grupper; enten å være enslig, eller en av gruppene 1-9 i tabell 4.1. Vi kaller dette mulighetsområdet for B , og lar B være hele universet av mulige ”valg”. Det er en viss sansynlighet, $P_j(B)$, for at et individ skal ”velge” gruppe j , $j=1,2,\dots,10$. Det må da være slik at

$$\sum_{j=1}^{10} P_j(B) = 1 \quad (1)$$

siden valg av en gruppe utelukker andre grupper. Hvis det eksisterer et sett av stokastiske variable U_j , $j \in B$ som er slik at

$$P_j(B) = P\left(U_j = \max_{k \in B} U_k\right) \quad (2)$$

utgjør valgsansynlighetene en stokastisk nyttemodell (random utility model), Dagsvik (2000). Vi kan anta at disse nyttenivåene har en struktur som angitt i likning (3)

$$U_j = v_j + \varepsilon_j \quad (3)$$

der v_j er et deterministisk ledd og ε_j er en tilfeldig variabel (støy), som er ekstremverdifordelt av type III. Vi kan spesifisere v_j som en funksjon av observerbare karakteristika og ukjente parametre

$$v_j = \beta_j x = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 = \beta_{j0} + \beta_{j1}alder + \beta_{j2}utdannelse \quad (4)$$

En mye brukt funksjonsspesifikasjon som tilfredsstill (1) og som følger fra (2) og (3), er den multinomiske logit-modellen (på matrisform):

$$P_j(B) = H_j(X; \beta) \equiv \frac{e^{X\beta_j}}{\sum_{k=1}^{10} e^{X\beta_k}} \quad (5)$$

Og det er også via denne funksjonsspesifikasjonen vi vil få estimert β 'ene, nærmere bestemt gjennom en maximum likelihood estimering. Siden $j = 1, 2, 3, \dots, 10$, ser vi at det blir 30 ukjente β 'er, men bare 27 kan bestemmes. Vi kan derfor sette β_{10} , β_{11} og β_{12} lik 0.

Vi lar $Y_{ij} = 1$ dersom individ i faller i kategori j og 0 ellers. $H_j(X_i; \beta)$ er uttrykk for den multinomiske logit sannsynligheten som gitt i (5) der X_i er vektoren av uavhengige variable (alder og utdanningsnivå) for individ i . Likning (6) er et uttrykk for den totale sannsynlighet i observasjonene, der N er antall observasjoner.

$$\prod_{i=1}^N \prod_{j=1}^{10} H_j(X_i; \beta)^{Y_{ij}} = \prod_{i=1}^N \prod_{j=1}^{10} \left(\frac{e^{X_i \beta_j}}{\sum_{k=1}^{10} e^{X_i \beta_k}} \right)^{Y_{ij}} \quad (6)$$

Vi tar logaritmen til (6) og får loglikelihood funksjonen (7):

$$l = \sum_{i=1}^N \sum_{j=1}^{10} Y_{ij} \log H_j(X_i; \beta) \quad (7)$$

De ukjente parametrene blir nå estimert ved å maksimere (7) med hensyn på parametrene. Hele denne operasjonen er noe dataproedyren i programpakken SAS vil ta seg av, men jeg har her, med svært god hjelp fra Dagsvik (2000), forsøkt å gi en beskrivelse av det teoretiske fundamentet for metoden. Selv om dette prosjektet ikke har som formål å analysere partnervalg, vil "nyttemodellen" (3) gjøre det enklere å simulere samboerskap blant personer som er registrert som enslige i DSP i forhold til å bruke den multinomiske modellen (5).

Når så våre ukjente parametre er estimert kan vi finne en gitt manns estimerte affinitet for gruppetilhørighet dersom vi kjenner hans alder og utdanningsnivå ved å sette inn for v_j og

trekke en ekstremverdifordelt ϵ_j i (3) for hver j . Naturligvis må vi også gjøre hele denne operasjonen en gang til, bare at menn og kvinner nå bytter plass, slik at vi kan finne affinitetene for gruppetilhørighet også for kvinnene som trekkes ut for å være med i undersøkelsen.

Men selv om vi vet hvilken gruppe mannen har størst affinitet for, er det flere problemer. For det første kan det hende at det ikke befinner seg noen kvinne på den aktuelle adressen som tilhører gruppen modellen har "valgt". Dette kan løses ved at vi finner en kvinne i den gruppen som har fått høyest affinitet blant de gruppene som er mulige på den aktuelle adressen. Et annet problem oppstår når det på mannens adresse er flere kvinner i "vinnergruppen". Sannsynligvis gjør vi ikke noen stor feil dersom vi bare velger en av kvinnene ved å trekke tilfeldig. En annen og muligens noe bedre metode, er å finne ut for hvilken gruppe de aktuelle kvinnene har høyest affinitet og dermed se hvilken som passer best til den uttrukne mannen. Denne fremgangsmåten bør kanskje benyttes hver gang, uansett om vi får et entydig valg i første omgang eller ikke.

4.3 Tidligere studier

Det ser ikke ut til å ha vært foretatt andre studier som er direkte sammelignbare, dvs studier som gjør bruk av en multinomisk logit-modell for å estimere sannsynligheter for samboerskap. Imidlertid har metoden tidligere blitt brukt for å si noe om samboere mer generelt.

En britisk studie, Berrington og Diamond (2000), tok i bruk en multinomisk logit-modell for å finne faktorer som påvirker valg av første samlivsform og alder ved inntreden i samboerskap eller ekteskap. Studien fokuserte bl.a. på sosiale problemer, boligsituasjon, religiøsitet og foreldres sosiale bakgrunn og samlivshistorie, men hadde også med individenes alder og utdannelsesnivå. De fant at lengre utdanning forsinket starten på første samliv en del, og en tendens til at ekteskap i større grad enn samboerskap ble utsatt.

En annen britisk studie, Ermisch og Francesconi (2000), bruker metoden til å kaste lys over årsaker til at samboerskap blir avbrutt, enten av ekteskap eller oppløsning. Analysen finner få signifikante faktorer, men forfatterne mener at utsikter til høyere inntekter, særlig for mannen i forholdet, øker sannsynligheten for at samboerskapet varer eller ender i ekteskap.

I Norge har Inger Texmon (1999) gjennomført en studie som likner på Berringtons. En del av problemstillingen var å finne variable som kan være med på å forklare valg av første samlivsform og samlivsform ved første barns fødsel. Særlig alder, bostedslandsdel og religiøs aktiv/ikke aktiv ga signifikante parametre. Dess yngre, dess større sannsynlighet for at samboerskap ble valgt som første samlivsform. Personer bosatt på Østlandet eller i Trøndelag og Nord-Norge velger oftere samboerskap enn personer bosatt på Sør- og Vestlandet. Religiøst aktive mennesker har større sannsynlighet for å gifte seg enn for å bli samboere. Utdanningsnivå fikk ikke signifikante variable, men det kan se ut til at menns giftermåls-tilbøyelighet øker litt ved høyere utdanning.

5. Estimering

Vi skal i dette kapitlet estimere en modell/nøkkel som kan brukes til å simulere husholdet til personer som er registrert som enslige i DSP. Vi vet jo at en god del av disse er samboere. Poenget er at modellen skal hjelpe oss til å finne ut hvem som er samboer med hvem og hvem som er enslig. Det er viktig å understreke at vi ikke ønsker å utarbeide en omfattende teori som forklarer samboerskap. Modelloppbygningen vil derfor ikke bli påvirket av teoretiske føringer når det f.eks. gjelder metodevalg eller valg av uavhengige variable. Vi ønsker kun at nøkkelen blir en praktisk og brukbar tilnærming til intervjuopplysninger om husholdet i forbindelse med inntekts- og fordelingsundersøkelser. I slike undersøkelser er det ikke avgjørende at vi identifiserer nøyaktig hvilke personer som bor sammen. Anta f.eks. at en mann, registrert som enslig, men i virkeligheten er samboer, er trukket ut til å være med i en fordelingsstudie. Det viktigste er at vi finner en samboer med inntekt på samme nivå som den virkelige samboeren. Vi vil altså få akkurat det samme resultatet (nivået på ekvivalentinntekt) dersom vi finner feil person men riktig inntekt.

5.1 Om datagrunnlaget

I dette avsnittet ser vi nærmere på hvilke personer vi skal bruke som datagrunnlag i estimeringen. Dette grunnlaget bør i størst mulig grad tilsvare de personene som en eventuell ”nøkkel” skal brukes på.

Problempersonene er i første rekke de som er registrert som enslige. Disse kan leve som enslige eller som samboere uten felles barn, og må naturligvis være med i datagrunnlaget. Det er disse vi skal bruke ”nøkkelen” på. For denne gruppen er opplysningene i DSP som de skal være, og vi kan anta at adressen i registeret er korrekt. Dermed har vi en kobling (adressen) til mulige andre personer som er registrert som enslige og derfor også er kandidater til å være samboer eller enslig. På denne måten får vi oversikt over alle mulige konstellasjoner av enslige og samboerpar på en adresse.

Men som vi har sett tidligere, er det en hel del personer, som oftest studenter, som er registrert med sine foreldre og eventuelle øvrige familie, men som i virkeligheten bor alene eller som samboer. Det burde ikke være noe problem å identifisere disse personene ved hjelp av opplysninger om borteboerstipend fra Lånekassen. En slik identifisering gjør at også disse

personene havner i gruppen av kandidater til å være enslige eller leve i samboerskap. Men for denne gruppen støter vi på et nytt problem; den folkeregistrerte adressen stemmer ikke med den faktiske. Studenter skal i henhold til regelverket ikke melde flytting til sitt bosted under studietiden. Dette fører til at folkeregisteret ikke gir noen mulighet for å koble en borteboende student med andre mulige samboerkandidater. For at borteboende studenter skal kunne være med, trenger vi opplysninger om semesteradresse fra Lånekassen. I tilfelle slike opplysninger ikke foreligger, vil det ikke bli mulig å estimere husholdet for denne gruppen ved hjelp av modellen vi skal utvikle. Denne modellen krever at vi har identifisert alle registrerte enslig på en gitt adresse, slik at vi får oversikt over alle potensielle samboerpar.

Modellen i denne oppgaven fokuserer kun på personer som er registrert som enslige, og vi ser altså bort fra gruppen av borteboende studenter. Som datagrunnlag for å estimere parametrene i modellen bruker vi derfor alle personer i utvalget til IF-97 som svarer at de lever i par, men som ikke er registrert på samme familie, samt alle enslige. I datasettet er det 474 par, 742 enslige menn og 860 enslige kvinner.

5.2 Hvem skal ”nøkkelen” brukes på?

Med utgangspunkt i register- og adresseopplysninger ønsker vi å finne ut hvem, blant de registrerte enslige, som virkelig lever alene og hvem som lever i samboerskap. I mange tilfeller byr ikke dette på noe problem. Generelt henviser en adresse til en bygning, og i denne bygningen kan det være en eller flere boliger, men hver bolig kan bare ha en adresse. Dersom en person er alene om å ha en viss adresse, kan vi derfor slutte at denne personen lever som enslig. Også dersom kun to av motsatt kjønn deler en adresse, vil det være overveiende sannsynlig at disse er samboere, dersom ikke aldersforskjellen mellom dem er for stor. I tilfeller som disse, er adresse nok til å finne husholdet til personene. Men hvor mange gjelder dette?

Bygninger som tilhører kategoriene 'frittliggende enebolig' og 'våningshus på gård' inneholder oftest en enkelt bolig. For slike boliger vil adressen være identisk med en bolig. Ifølge SSB² lever 35 prosent av enslige mellom 16 og 44 år i frittliggende enebolig eller i våningshus på gård. For disse vil med andre ord adresseopplysning alene være nok for å bestemme husholdet. I gruppen par mellom 16 og 44 år uten barn, bor 51 prosent i slike

² Levekårsundersøkelsen 1997: Boforhold, fritid og vold: Tabell 10A, Statistisk sentralbyrå.

boliger, og også for mange av disse husholdene vil adresseopplysninger alene være tilstrekkelig. Andelen blant enslige og par uten barn som bor i enebolig eller våningshus er enda høyere i aldersgruppen 45-66 år, der 52 prosent av de enslige og 72 prosent av par uten barn lever i slike boliger. Men som vi har sett er det stort sett unge mennesker som lever i samboerskap, og det vil derfor være riktigere å bruke tallene for aldersgruppen 16-44 år som et anslag på hvor mange registrerte enslige vi klarer å plassere i hushold kun på grunnlag av adresseopplysninger.

I tilfellene der en adresse henviser til en bygning med flere boliger, som blokker, bygårder og horisontalt delte småhus, vil ikke adresse alene kunne si oss noe om husholdet. Adressen forteller oss bare hvem som deler bygning og ikke hvem som deler hver av boligene i bygningen. Blant personene som er bosatt på en viss adresse, plukker vi ut de som er registrert som enslige. Det er disse vi ved hjelp av modellen omtalt i forrige kapittel vil forsøke å predikere husholdssammensetningen til.

5.3 Modell 1

Modell (1) under estimerer affiniteter for å tilhøre de 10 ulike gruppene satt opp i tabell 4.1. Gruppe 1-9 betyr at person j er samboer og gruppenummeret forteller hvilken kategori samboeren tilhører. Disse kategoriene består av forskjellige nivåer på samboers alder og utdanning, se tabell 5.1. Personer som er enslige, tilhører gruppe 10. Aldersinndelingen er noe forskjellig for kvinner og menn for å sikre en jevnere fordeling av antallet personer i de ulike gruppene, og fordi menn ofte er noe eldre før de går inn i et samliv.

Tabell 5.1: *Gruppenndeling av samboere etter alder og utdanning. Gruppenummer brukt i modell 1 og 2.*

		Høyeste fullførte utdanning		
		Menner	Kvinner	
				Lav (0-10 år)
				Middels (Videreg. skole)
				Høyere (Univ. / Høyskole)
Alder		Menner	Kvinner	
	17-26	17-24		1
	27-35	25-33		2
	36-	34-		3
				4
				5
				6
				7
				8
				9

Modell 1 er en noe forenklet form av opplegget i kapittel 4.2. Forenklingen består i at β_1 og β_2 er uavhengige av samboergruppen. Grunnen til å først forsøke denne forenklede modellen er at vi har for få observasjoner i hver gruppe til å estimere brukbare stigningsparametre for hver samboergruppe.

Modell 1:

$$U_{ij} = v_{ij} + e_{ij} = \alpha_i + \beta_1 \text{alder}_j + \beta_2 \text{utdannelse}_j + e_{ij} \quad (8)$$

Uavhengige variable er alder og utdanningsnivå. Forsøk viste at disse variablene ga de beste resultatene av de som var tilgjengelige i datagrunnlaget. Alder er personenes alder i antall år og utdanning har verdien 1 til 8³. Igjen er i samboergruppe (eventuelt enslig, $i=10$) og j er person. Parametrene er estimert ved prosedyren PROC LOGISTIC i SAS. Prosedyren ga konvergerende parameterestimer. For hvert kjønn gir modellen estimer på 9 konstantledd i tillegg til α_1 og α_2 . Siden vi ikke har parameter for $i=10$ settes $\alpha_{j10} = 0$. Som nevnt i forrige kapittel er e_{ij} et tilfeldig restledd trukket fra en ekstremverdifordeling (type III). Teknisk vil dette si at $e_{ij} = -\log(-\log r)$ der r er et tilfeldig tall mellom 0 og 1 fra en uniform fordeling og generert av dataprogrammet. e_{ij} vil derimot følge en ekstremverdifordeling av type III, som kjennetegnes ved et tyngdepunkt i underkant av verdien 1 og med en tung hale til høyre.

I tabell 5.2 er estimeringsresultatene brukt i modell 1. Som det kommer frem av tabellen, er det ganske stor forskjell mellom kjønnene. For kvinnene gir modellen bare signifikante parameterestimer, mens det blant mennene bare er sju signifikante parametre. Det vi imidlertid er mer opptatt av, er hvor godt egnet modellen er til å gjenskape den virkelige fordelingen av samboere og enslige, siden det er denne som igjen er avgjørende for fordelingen av ekvivalentinntekt.

Dersom vi bruker de estimerte sannsynlighetene fra modell 1 direkte, uten å gå veien om nyttefunksjonen i likning (3) i forrige kapittel, får alle personene størst sannsynlighet for å tilhøre gruppe 10. Det betyr at ifølge de estimerte sannsynlighetene er alle enslige. Som vi vet gjelder dette bare for 64 prosent av kvinnene og 61 prosent av mennene i dataene. Modellen fungerer med andre ord ikke særlig godt når det gjelder å predikere personers samboergruppe ut fra de estimerte sannsynlighetene.

³ Tallene 1 til 8 representerer nivået på høyeste fullførte utdanning. 1 er for eksempel førskole og 8 er utdanning på forskernivå. "Standard for utdanningsgruppering i offentlig norsk statistikk", *Standarder for norsk statistikk 7*.

Tabell 5.2: Estimerte parametre i modell 1.

Parameter	Menn			Kvinner		
	Koeffisient	St.avvik	P-verdi	Koeffisient	St.avvik	P-verdi
Alder	-0,041	0,004	<0,001	-0,073	0,004	<0,0001
Utdanning	-0,004	0,043	0,933	-0,126	0,048	0,008
Konstantledd, i=1	-2,458	0,329	<0,001	-0,804	0,343	0,019
Konstantledd, i=2	-0,884	0,277	0,001	0,706	0,304	0,020
Konstantledd, i=3	-0,474	0,273	0,082	1,186	0,302	<0,001
Konstantledd, i=4	-0,339	0,272	0,212	1,451	0,301	<0,001
Konstantledd, i=5	0,015	0,270	0,956	1,896	0,302	<0,001
Konstantledd, i=6	0,567	0,269	0,035	2,314	0,304	<0,001
Konstantledd, i=7	0,982	0,270	<0,001	2,710	0,307	<0,001
Konstantledd, i=8	1,093	0,270	<0,001	2,890	0,308	<0,001
Konstantledd, i=9	1,197	0,270	<0,001	3,061	0,310	<0,001
Antall signifikante konstantl. ($\alpha=0,05$)		6 av 9			9 av 9	
-2 log likelihood		3434,3			3338,3	

P-verdiene er beregnet ut fra en forutsetning om at parameterestimatenes er chi-kvadrat-fordelt.

Når relasjonen (8) blir brukt til å simulere fordelingen av samboere og enslige i utvalget, og vi dermed bruker parameterestimatenes i tillegg til et restledd trukket fra ekstremverdifordelingen, blir resultatene noe mer tilfredsstillende. Andelen personer som blir predikert til å være enslige blir om lag halvparten av kvinnene og 38 prosent av mennene. Dette stemmer litt bedre enn resultatet av sannsynlighetene, men det blir noe for få enslige i forhold til samboere.

Simuleringen foregår slik at hver person blir predikert å tilhøre den gruppen som i følge (8) gir høyest nivå på U. Det vil si

$$\text{predikert gruppe for person } j = \text{maks}(U_{ji}, i = 1, 2, \dots, 10).$$

For eksempel vil en kvinnes affinitet for å tilhøre gruppe 2 (dvs. ha en samboer i gruppe2) når hun er 27 år og har utdanning på nivå 6 bli:

$$U_{j2} = 0,706 - 0,073*27 - 0,126*6 + e_{ji}$$

På tilsvarende måte blir affinitetene simulert for alle personer og alle grupper.

Når det stokastiske restleddet kommer inn i bildet, kan det gi store utslag på gruppene med få personer. For å kontrollere for disse tilfeldige utslagene, er simuleringen foretatt 30 ganger på hele utvalget, med nytt tilfeldig restledd for hver av personene i hver simulering. Til slutt er det beregnet et gjennomsnitt av predikert antall personer i hver av gruppene i de 30 omgangene. Tallene for disse simuleringsgjennomsnittene og den virkelige observerte fordelingen er i tabell 5.3.

Vi så at modellen ikke treffer spesielt godt på andelen samboere i forhold til enslige. Tabell 5.3 viser også den predikerte fordelingen av personer på de 9 samboergruppene. En ønsket egenskap, er at modellen gjenspeiler de observerte andelene personer i hver av de 9 samboergruppene. Hver gruppe representerer ulike kombinasjoner og alder og utdanning, og både alder og utdanning påvirker personers inntekt og dermed fordelingen av ekvivalentinntekt. Skal modellen til slutt resultere i en inntektsfordeling som likner den virkelige, er det en klar fordel at alle undergruppene 1-9 er representert i riktig forhold til

Tabell 5.3: *Virkelig og predikert antall personer i hver samboergruppe. Predikert antall er gjennomsnittet av 30 simuleringer med modell 1.*

Gruppe	Menn		Kvinner	
	Virkelig	Predikert	Virkelig	Predikert
1	24	5	27	4
2	80	22	77	18
3	42	35	47	28
4	17	41	32	35
5	52	57	63	53
6	102	97	68	84
7	97	147	77	127
8	30	163	41	151
9	29	185	42	176
10	742	463	860	661

hverandre. Den predikerte fordelingen i tabellen nedenfor gir imidlertid en klar overrepresentasjon av personer i de siste av de ni gruppene og en tilsvarende underrepresentasjon av de første gruppene. En slik fordeling innebærer at det blant de modellpredikerte samboerne blant annet vil være for mange eldre med høy utdanning (gruppe 9) og for få yngre med lav utdanning (gruppe 1) i forhold til den observerte, reelle fordelingen. Dette gir i neste trinn en skjev fordeling av ekvivalentinntekt.

I forhold til målet om å bruke modellen for å danne en fordeling av ekvivalentinntekt har modellen derfor to svakheter. For det første gir den for stor andel samboere, som igjen vil bety at mange får for høy ekvivalentinntekt. Gjennomsnittet vil også bli overestimert. For det andre predikerer modellen for mange samboere i gruppen eldre med høy utdanning og for få i gruppen yngre med lav utdanning. Dette vil ytterligere øke ekvivalentinntekten til de predikerte samboerne, siden eldre med høy utdanning stort sett har høyere inntekt enn yngre med lav utdanning. Resultatet blir altså at de som kommer gunstig ut mht. ekvivalentinntekt, (samboerne) i for stor grad har høy inntekt, og de enslige, som i utgangspunktet har lav ekvivalentinntekt, i for stor grad har lav inntekt. Modell 1 vil derfor ikke gi noe godt bilde verken av fordelingen av samboere eller av fordelingen av ekvivalentinntekt.

5.4 Modell 2

Modell 2 tilsvarer fremgangsmåten skissert i kapittel 4.2. I forhold til modell 1 er den mer fleksibel siden også koeffisientene til alder og utdanning tillates å varierer fra gruppe til gruppe. I modell 1 ligger det innebygget en antagelse om at stigningsparametrene i regresjonen er like for alle grupper. Dette impliserer at regresjonslinjene er parallelle for hver gruppe, den eneste forskjellene er konstantleddet som bestemmer høydenivået på linjene. Modell 2 tillater at linjene "spriker" dersom dette passer bedre med observasjonene. Modellen betrakter den uavhengige variabelen som en kategori- eller nominellverdi og ikke som en ordinal verdi. Vi kan altså si at modell 2 er mer tilpasningsdyktig i forhold til dataene sammenliknet med modell 1. Den største fordel er likevel at simuleringer med denne modellen alltid vil gi riktig andel personer i hver av de 10 gruppene. Dette gir et bedre og mer solid utgangspunkt i forhold til fordelingen av ekvivalentinntekt.

Det er rimelig å anta at et viktig aspekt ved inngåelse av samboerskap (og andre samliv) er *forholdet* mellom samboernes alder, utdanning og andre variable, og ikke bare nivået på egen

alder og utdanning. Korrelasjonskoeffisientene mellom samboeres alder og utdanningsnivå er hhv 0,95 og 0,49 blant samboerne uten barn i dataene. Dette drar vi nytte av når vi lar β 'ene variere fra gruppe til gruppe. Et eksempel vil klargjøre:

En kvinne som selv er ung og har høy utdanning vil ha forholdsvis høy sannsynlighet for å være samboer med en mann som også er ung og har høy utdanning (gruppe 3), og lav sannsynlighet for å være samboer med en mann som er eldre og har lav utdanning (gruppe 7). Vi vil derfor forvente at estimeringen reflekterer dette ved at sannsynligheten for å ha en samboer fra gruppe 3 faller med egen alder og øker med eget utdanningsnivå. Altså negativ alderskoeffisient og positiv utdanningskoeffisient. Sannsynligheten for å være samboer med en mann fra gruppe 7 vil i motsetning øke med egen alder og falle med eget utdanningsnivå og dermed gi negativ utdanningskoeffisient og positiv alderskoeffisient.

I modell 1 fikk vi forholdsvis svake estimater på koeffisientene. Antagelig var dette fordi modellen tvang alder og utdanning til å ha samme effekt uansett gruppe. Kun konstantleddet fikk variere fra gruppe til gruppe, men dette utnytter naturligvis ikke informasjonen om forholdet mellom egen og samboers alders- og utdanningsnivå. Denne økte utnyttelsen av informasjonen i dataene vil forhåpentligvis gi seg utslag i mer treffsikre estimater.

Modell 2:

$$U_{ij} = v_{ij} + e_{ij} = \alpha_i + \beta_{1i}ALD_j + \beta_{2i}UTD_j + e_{ij} \quad (9)$$

Der i er samboergruppe 1,2,...,10 og j er person. Parametrene i (9) er estimert ved PROC CATMOD i programpakken SAS. ALD er aldersgruppe definert ved intervallene i tabell 5.1, og UTD er utdanningsnivå, lav/middels/høy utdanning, definert i samme tabell. Forskjellen fra modell 1 er, ved siden av at altså β 'ene får variere fra gruppe til gruppe, at de uavhengige variablene er representert gjennom kategorivariable.

Desverre har den økte fleksibiliteten i modell 2 en pris. For å få tilstrekkelig variasjon i observasjonene, og dermed gode estimater, er det nødvendig med mange observasjoner for hver alders- og utdanningskombinasjon. Det er derfor personenes egne, observerte alder er erstattet med tre aldersgrupper, de samme aldersgruppene som inndelingen av samboerne i

tabell 5.1 er basert på. Det samme gjelder utdanningsnivået som i modell 2 er representert ved lav/middels/høy utdanning, se samme tabell.

Vi går nå over til å se på estimeringsresultatene av modell 2 som er i tabell 5.4. Som det kommer fram der gir modellen mange signifikante parametre, i tillegg til at både aldersgruppe og utdanningsnivå som samlede Wald chi-kvadrat observatorer er klart signifikante. Sannsynlighetsmaksimeringsrutinene konvergerer for både kvinner og menn.

Parameterestimer med standardavvik og p-verdier er i tabell 5.5 under. Som vi ser blir antagelsene vi gjorde i eksemplet over om fortegnene på noen av parameterestimatene bekreftet. For både kvinner og menn får vi negative alderskoeffisienter og positive utdannelseskoeffisienter for den estimerte sannsynligheten for å ha en samboer i gruppe 3 (lav alder, høy utdanning). Og som ventet er fortegnene motsatt når det gjelder sannsynligheten for å ha en samboer i gruppe 7 (høy alder, lav utdanning).

Tabell 5.4: *Analyse av sannsynlighetsmaksimering, modell 2.*

Kilde	Menn			Kvinner		
	DF	Chi-kvadrat	P-verdi	DF	Chi-kvadrat	P-verdi
Konstantledd	9	139,40	<0,001	9	122,63	<0,001
Aldersgruppe	9	195,52	<0,001	9	237,73	<0,001
Utdanningsnivå	9	84,01	<0,001	9	76,87	<0,001
Antall signifikante parametere ($\alpha=0,05$)	17 av 27			18 av 27		

Tabell 5.5: *Parameterestimer fra modell 2.*

Grupper	Gruppe	Menn			Kvinner		
		Estimat	St.avvik	P-verdi	Estimat	St.avvik	P-verdi
Konstantledd	1	-0,819	0,868	0,345	1,261	0,803	0,117
	2	1,743	0,523	0,001	1,477	0,543	0,007
	3	-0,103	0,720	0,887	-1,296	0,712	0,069
	4	-0,315	0,988	0,750	-0,894	0,763	0,241
	5	-0,788	0,593	0,184	-0,861	0,573	0,133
	6	-3,725	0,554	<0,001	-4,326	0,755	<0,001
	7	-11,455	2,146	<0,001	-7,366	1,582	<0,001
	8	-7,251	1,582	<0,001	-10,733	3,074	0,001
	9	-12,946	2,251	<0,001	-9,186	1,405	<0,001
Aldersgruppe	1	-1,594	0,344	<0,001	-2,402	0,386	<0,001
	2	-2,124	0,242	<0,001	-2,472	0,255	<0,001
	3	-2,168	0,341	<0,001	-1,834	0,257	<0,001
	4	-0,828	0,311	0,008	-1,016	0,229	<0,001
	5	-0,694	0,185	<0,001	-1,153	0,176	<0,001
	6	-0,606	0,148	<0,001	-0,869	0,187	<0,001
	7	3,481	0,708	<0,001	1,852	0,505	<0,001
	8	1,662	0,501	0,001	2,702	1,004	0,007
	9	2,733	0,717	0,001	1,665	0,435	0,001
Utdanningsnivå	1	0,117	0,309	0,704	-0,182	0,292	0,533
	2	-0,209	0,188	0,265	0,254	0,192	0,187
	3	0,393	0,250	0,115	0,871	0,247	<0,001
	4	-1,016	0,400	0,011	-0,065	0,241	0,788
	5	-0,227	0,204	0,264	0,364	0,179	0,042
	6	1,295	0,182	<0,001	1,575	0,241	<0,001
	7	-0,175	0,154	0,255	-0,165	0,179	0,357
	8	-0,194	0,258	0,452	-0,011	0,224	0,963
	9	1,081	0,262	<0,001	0,854	0,187	<0,001

Modell 2 sikrer at fordelingen av personer på de 10 gruppene blir som den skal. Tilhører for eksempel 3,5 prosent av de registrerte enslige kvinnene gruppe 3, vil simulering ved hjelp av modell 2 også alltid gi 3,5 prosent i gruppe 3 dersom antallet simuleringer (dvs. trekninger) blir stort nok. Tabell 5.6 viser virkelig og predikert antall personer i hver gruppe. Antallet predikerte i hver gruppe framkommer også her som et gjennomsnitt etter 30 runder med simuleringer på personene i datasettet. Det er tydelig at fordelingene er svært like allerede etter 30 trekninger. Årsaken til dette gode samsvaret ligger i modellspesifikasjonen, der de uavhengige variablene er aggregert opp på samme nivå som grupperingen av den uavhengige variabelen.

Tabell 5.6: *Virkelig og predikert antall personer i hver samboergruppe. Predikert antall er gjennomsnittet av 30 trekninger.*

Gruppe	Menn		Kvinner	
	Virkelig	Predikert	Virkelig	Predikert
1	24	25	27	28
2	80	81	77	79
3	42	42	47	46
4	17	17	32	31
5	52	52	63	63
6	102	101	68	69
7	97	96	77	77
8	30	29	41	40
9	29	31	42	41
10	742	742	860	860

5.5 Diskusjon

I dette kapitlet har vi sett på mulighetene for å finne samboerpar uten felles barn ut fra opplysninger om adresse, alder og utdanningsnivå. Av de to ulike modellene som ble forsøkt virker det som om den siste, modell 2, er den mest robuste. I modell 2 tillates stigningskoeffisientene å variere for ulike verdier av de uavhengige variablene, mens de i modell 1 ble holdt konstante. Et mulig ankepunkt ved modell 2, er hvordan alder som uavhengig variabel i denne modellen ble gruppert, slik at hver observasjon av alder var

representert av en aldersgruppe i stedet for virkelig alder i antall år, slik tilfellet var i modell 1. Som følge av denne grupperingen mister vi noe informasjon i overgangen fra en (tilnærmet) kontinuerlig variabel til en ordinal kategorivariabel. Dette var nødvendig for å få nok observasjoner i hver kategori til at modell 2 kunne gjennomføres. Noe av det samme gjelder utdanningsvariabelen, som i modell 1 hadde verdier fra 1 til 8, men som i modell 2 ble gruppert til lav, middels og høyere utdanning, altså kun tre nivåer. Hovedårsaken til at modell 2 er å foretrekke, ligger i at denne bruker informasjonen som ligger i *forholdet* mellom to personers aldersgruppe og utdanningsnivå, og at denne modellen gjensker den observerte fordelingen av personer både i forhold til de ulike samboergruppene og i forhold til kategoriene samboer - enslig.

Generelt må man også være forsiktig med å trekke konklusjoner om en modell ut i fra hvor godt den passer til dataene den selv bygger på. Andre data vil naturligvis alltid være noe forskjellige fra dataene som i denne oppgaven ble brukt til å estimere parametrene i modellen. Samsvaret mellom modellprediksjoner og observerte realiteter vil derfor kunne bli et annet dersom modellene ble forsøkt på andre data, dvs. andre personer enn de registrerte enslige i utvalget til IF-97. Modell 2 vil imidlertid alltid gi riktig fordeling av personer i de ulike gruppene uansett om vi simulerer med et annet datasett som grunnlag. Dette vil gjelde så lenge de uavhengige variablene det simuleres på grunnlag av er aggregert til de samme alders- og utdanningsgrupper vi benyttet til å estimere parametrene i modellen.

Hvor godt modellen fungerer i virkeligheten som nøkkel til å identifisere samboere uten barn kan vi imidlertid ikke si noe om. Det vil være nødvendig å teste ut modellen på et utvalg der vi har med adresseopplysninger. I datasettet brukt i våre estimeringer er det ikke adresseopplysninger, men det ville heller ikke hjulpet om personenes adresser ble koblet på. Årsaken ligger i hvordan utvalget er satt sammen og er nærmere redegjort for i innledningen til kapittel 4. For å finne ut om modellen fungerer, må den derfor testes på et utvalg bestående av alle registrerte enslige i et område.

Målet med dette kapitlet var å forsøke å lage en modell for å estimere fordelingen av samboere uten barn blant personer registrert i DSF som enslige. Modellen er ikke på noen måte ment brukt for å beregne sannsynligheter for enkeltindividers samlivsform eller egenskaper ved en eventuell partner. Det må også vektlegges at identifisering av fordelingen av samboere bare er et middel for å komme fram til en inntektsfordeling basert på

ekvivalentinntekt. For at det skal være mulig å danne ekvivalentinntektene til en gruppe personer må husholdet til hver enkelt være kjent. Men for å få et bilde av en *fordeling* av ekvivalentinntekter over en populasjon, er det ikke nødvendig å kjenne sammensetningen av hvert enkelt hushold. Så lenge vi er opptatt av egenskapene ved *fordelingen* av ekvivalentinntekt, som for eksempel Gini-koeffisienten og andelen personer under 50 prosent av medianinntekten, betyr ikke kjennskap til enkelthusholdene noe. Derfor betyr det lite om modellprediksjonene treffer virkeligheten på individnivå. Det viktige er om hele fordelingen av hushold, og dermed ekvivalentinntekter, treffer den virkelige fordelingen.

Dette kapitlet har uansett vist at det er mulig å estimere en modell som stort sett har signifikante parametre med fortegn som forventet og som meget godt gjensker fordelingen av samboere blant personer registrert som enslige. Videre studier vil være nødvendig for å teste modellen på virkelige observasjoner og eventuelt raffinere modellen ytterligere.

Referanser:

- Atkinson, A.B., Rainwater, L. og Smeeding, T.M. (1995): "Income distribution in OECD countries: evidence from the Luxembourg Income Study." *Social policy studies* nr.18. OECD, Paris.
- Berrington, A. og Diamond, I. (2000): "Marriage or coabitation: a competing risks analysis of first-partnership formation among the 1958 British birth cohort." *Journal of the Royal Statistical Society*, 163. The Royal Statistical Society, London.
- Bojer, H. (1990): "Inntekt og ulikhet", *Serien for studenter nr 21*, Sosialøkonomisk intstitutt, Universitetet i Oslo.
- Dagsvik, John K. (2000): "Probabilistic Models for Qualitative Choice Behavior – An introduction" *Documents* 2000/1, Statistisk Sentralbyrå.
- Epland, J. (1998-a): "Endringer i fordelingen av husholdningsinntekt:1986-1996." *Rapporter* 98/17, Statistisk Sentralbyrå.
- Epland, J. (1998-b): "Income distribution data for Norway: Robustness assessment report". Statistisk Sentralbyrå.
- Epland, J. (1998-c): "Inntekt etter skatt – revisjon av inntektsregnskapet i inntekts- og formuesundersøkelsen for husholdninger." *Notater* 98/43, Statistisk Sentralbyrå.
- Epland, J. (1997): Inntekts- og formuesundersøkinga – kombinasjon av intervjudata og registerdata i produksjonen av husholdsstatistikk, i Gulbransen, Moen og Keilman (red): *Norske husholdninger i forandring*, Norges forskningsråd.
- Ermisch, J. og Francesconi, M. (2000): "Cohabitation in Great Britain: not for long but here to stay." *Journal of the Royal Statistical Society*, 163. The Royal Statistical Society, London.
- Hauge, I., Hendriks, C., Hokstad, Ø. og Hustoft, A.G. (2000): "Standard for begreper og kjennemerker knyttet til familie- og husholdningsstatistikken." *Rapporter* 2000/5, Statistisk Sentralbyrå.
- Holdt, P. M., Kristoffersen, S. og Pedersen, V. (1999): "Beregning av vektorer til inntekts- og formuesundersøkelsene 1997" *Notater* 1999/78, Statistisk Sentralbyrå.
- Lund, K. og Aaberge, R. (1999): "Ekvivalensskala og tallfesting av omfang, fordeling og utvikling av fattigdom i Norge 1982-1995." i Birkeland, E. (Red.): "Forskjeller i levekår – hefte 1: Inntekt." *Notater* 99/32. Statistisk Sentralbyrå.
- Sparby, G. (1997): "Inntekts- og formuesundersøkelsen 1992. Dokumentasjon." *Notater* 97/12. Statistisk Sentralbyrå.
- Texmon, I. (1999): "Samliv i Norge mot slutten av 1900-tallet. En beskrivelse av endringer og mangfold." i *Samboerne og Samfunnet*, NOU 1999:25. Barne- og familiedepartementet, Oslo.

Publikasjoner fra Frischsenteret

Alle publikasjoner er tilgjengelig i Pdf-format på : www.frisch.uio.no

Rapporter

1/1999	Arbeidsledighet, arbeidsmarkedspolitikk og jobbsøking i Norge	Knut Røed, Hege Torp, Tom Erik Aabø
2/1999	Egenskaper ved tildelingsformer for nasjonale klimagasskvoter	Rolf Golombek, Michael Hoel, Snorre Kverndokk, Ove Wolfgang
3/1999	Regionale virkninger av økte elektrisitetspriser til kraftkrevende industri	Nils-Henrik M. von der Fehr, Trond Hjørungdal
4/1999	Bedriftsnedleggelse og klimakvoter i norsk industri	Rolf Golombek, Arvid Raknerud
5/1999	Utdanning og livsinntekt i Norge	Oddbjørn Raaum, Tom Erik Aabø, Thomas Karterud
1/2000	Hvem er de ledige? En økonometrisk analyse av arbeidsledighetens sammensetning i Norge på 1990-tallet	Morten Nordberg
2/2000	Effektivitet i pleie- og omsorgssektoren	Dag F. Edvardsen, Finn R. Førsund, Eline Aas
3/2000	Norge i liberalisert europeisk energimarked	Finn Roar Aune, Rolf Golombek, Knut Einar Rosendahl, Sverre A.C. Kittelsen
4/2000	Hvem vil og hvem får delta? Analyser av rekruttering og utvelgelse av deltakere til arbeidsmarkedstiltak i Norge på 1990-tallet	Knut Røed, Hege Torp, Irene Tuveng, Tao Zhang
5/2000	Deregulering av det vest-europeiske gassmarkedet - korttidseffekter	Rolf Golombek, Sverre A.C. Kittelsen, Ove Wolfgang
6/2000	Oversikt over litteratur om svart arbeid og skatteunndragelser	Erling Eide
7/2000	Arbeidstilbud i vedvarende gode tider	Christian Brinch
8/2000	Miljøreguleringer av norsk treforedlingsindustri	Rolf Golombek, Arent Greve, Ken Harris
1/2001	Analyse av inntektsfordeling og inntektsulikhet basert på registerdata. En kartlegging av muligheter og begrensninger	Remy Åserud

Arbeidsnotater

1/1999	Kan markedskreftene temmes i lønnsdannelsen?	Colin Forthun
2/1999	Inntektseffekter av utdanning i Norge – en litteraturoversikt	Oddbjørn Raaum
1/2000	Empirical Specification of the Model in "Early Retirement and Economic Incentives"	Erik Hernæs, Steinar Strøm
2/2000	Forholdene på arbeidsmarkedet, økonomiske incentiver og risikoen for å bli yrkeshemmet	Christian L. Wold Eide
3/2000	Koordinering av inntektsoppgjørene i Norge og Sverige 1961-1999	Bergljot Bjørnson Barkbu
4/2000	Insentivvirkninger av skatte- og pensjonsregler	Fredrik Haugen
5/2000	Dynamisk arbeidstilbud	Merethe Nordling

Memoranda

Serien publiseres av Sosialøkonomisk institutt, Universitetet i Oslo, i samarbeid med Frischsenteret. Listen under omfatter kun memoranda tilknyttet prosjekter på Frischsenteret. En komplett oversikt over memoranda finnes på www.sv.uio.no/sosoek/memo/.

3/1999	The Economics of Screening Programs	Steinar Strøm
7/1999	What hides behind the rate of unemployment? Micro evidence from Norway	Knut Røed, Tao Zhang
9/1999	Monte Carlo Simulations of DEA Efficiency Measures and Hypothesis Tests	Sverre A.C. Kittelsen
11/1999	Efficiency and Productivity of Norwegian Colleges	Finn R. Førsund, Kjell Ove Kalhagen
13/1999	Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies	Tor Jakob Klette, Jarle Møen, Zvi Griliches
14/1999	Unemployment Duration in a Non-Stationary Macroeconomic Environment	Knut Røed, Tao Zhang
16/1999	The effect of schooling on earnings: The role of family background studied by a large sample of Norwegian twins	Oddbjørn Raaum, Tom Erik Aabø

17/1999	Early Retirement and Economic Incentives	Erik Hernæs, Marte Sollie, Steinar Strøm
18/1999	Fewer in Number but Harder to Employ: Incidence and Duration of Unemployment in an Economic Upswing	Erik Hernæs
19/1999	Progressiv Taxes and the Labour Market	Knut Røed, Steinar Strøm
22/1999	Inequality, Social Insurance and Redistribution	Karl Ove Moene, Michael Wallerstein
24/1999	Do Voluntary Agreements Lead to Cost Efficiency	Rolf Golombek, Espen R. Moen
25/1999	Rent Grabbing and Russia's Economic Collapse	Sheetal K. Chand and Karl Ove Moene
28/1999	The role of foreign ownership in domestic environmental regulation under asymmetric information	Jon Vislie
29/1999	Labor unions versus individualized bargaining with heterogeneous labor	Jon Strand
32/1999	Efficiency in the Provision of Municipal Nursing – and Home-Care Services: The Norwegian Experience	Espen Erlandsen, Finn R. Førsund
33/1999	Effects of Progressive Taxes under Decentralized Bargaining and Heterogeneous Labor	Jon Strand
34/1999	Reflections on Abatement Modelling	Ove Wolfgang
35/1999	Crime Induced Poverty Traps	Halvor Mehlum, Karl Ove Moene, Ragnar Torvik
36/1999	Statistical Discrimination and the Returns to Human Capital and Credentials	Christian Brinch
38/1999	Relative Unemployment Rates and Skill-Biased Technological Change	Knut Røed
2/2000	Married Men and Early Retirement Under the AFP Scheme	Ole J. Røgeberg
4/2000	Family Labor Supply when the Husband is Eligible for Early Retirement: Some Empirical Evidences	Jia Zhiyang
5/2000	Earnings Assimilation of Immigrants in Norway - A Reappraisal	Pål Longva, Oddbjørn Raum
13/2000	Family Labour Supply when the Husband is Eligible for Early Retirement	Erik Hernæs, Steinar Strøm

15/2000	Labour Market Transitions and Economic Incentives	Knut Røed, Tao Zhang
19/2000	Have the Relative Employment Prospects for the Low-Skilled Deteriorated After All?	Knut Røed, Morten Nordberg
23/2000	A Note on the Weibull Distribution and Time Aggregation Bias	Knut Røed, Tao Zhang
34/2000	CO2 mitigation costs and ancillary benefits in the Nordic countries, the UK and Ireland: A survey	Snorre Kverndokk, Knut Einar Rosendahl
36/2000	A turning point in the development of Norwegian economics - the establishment of the University Institute of Economics in 1932	Olav Bjerkholt
40/2000	Health Insurance: Treatment vs. Compensation	Geir B. Asheim, Anne Wenche Emblem, Tore Nilssen
41/2000	Private health care as a supplement to a public health system with waiting time for treatment	Michael Hoel, Erik Magnus Sæther



Frischsenteret

Stiftelsen Frischsenteret for samfunnsøkonomisk forskning er en uavhengig stiftelse opprettet av Universitetet i Oslo. Frischsenteret utfører samfunnsøkonomisk forskning i samarbeid med Sosialøkonomisk institutt ved Universitetet i Oslo. Forskningsprosjektene er i hovedsak finansiert av Norges forskningsråd, departementer og internasjonale organisasjoner. De fleste prosjektene utføres i samarbeid mellom Frischsenteret og forskere ved andre norske og utenlandske forskningsinstitusjoner.

Frischsenteret
Gaustadalléen 21
0349 Oslo
Tlf: 22958810
Fax: 22958825
frisch@frisch.uio.no
www.frisch.uio.no