

*Magne Mogstad og Li-Chun Zhang*

**På veien fra familie- til  
husholdningsregister:**

En metode for prediksjon av  
samboere uten barn

Notater

# Innhold

<b>1. Innledning .....</b>	<b>1</b>
<b>2. Viktige begreper, definisjoner og forutsetninger .....</b>	<b>3</b>
2.1 Definisjonen av husholdning, familie og bostedsadresse.....	3
2.2 Økonomisk velferd som begrep .....	4
2.3 Ekvivalentinntekt .....	4
2.4 Inntektsbegrepet .....	6
2.5 Datamateriale .....	7
<b>3. Analyse av inntektsulikhet ved bruk av familie og husholdning som økonomiske enheter .....</b>	<b>8</b>
3.1 Mål for inntektsulikhet.....	8
3.2 En kartlegging av avviket mellom opplysninger om familie fra DSF og opplysninger om husholdningen fra FoB2001. ....	9
3.3 Analyse av inntektsulikhet for totalpopulasjonen, i spesielle grupper i samfunnet og på kommunalt nivå med husholding og familie som økonomisk enhet. ....	13
<b>4. Modell for prediksjon av samboere uten barn fra registrerte enslige i DSF .....</b>	<b>18</b>
4.1 Teori for modellen.....	19
4.1.1 Populasjon og enheter.....	19
4.1.2 Desilgjennomsnittet av ekvivalentinntekten.....	19
4.1.3 Registerbasert metode .....	20
4.1.4 Samboerskap blant registrerte enslige .....	20
4.1.5 Antall matcher på hver adresse.....	21
4.1.6 Fordelingen til matching klassene .....	21
4.1.7 Stokastisk matching.....	22
4.2 Resultater ved bruk av modellen for prediksjon av samboere uten barn .....	23
4.3 Endring i den stokastiske matching prosessen .....	28
4.4 Resultater ved bruk av modellen for prediksjon av samboere uten barn etter endring i den stokastiske matching prosessen .....	29
4.4.1 Landsnivå .....	29
4.4.2 Kommunalt nivå .....	30
4.5 Analyse av inntektsulikhet ved å benytte det utvidete familiebegrepet som økonomisk enhet .	30
4.6 Utblikk.....	32
<b>5. Sammendrag.....</b>	<b>34</b>
<b>Referanser .....</b>	<b>36</b>
<b>Vedlegg: Tabellene 4.3, 4.4, 4.6 og 4.7 samt figur 4.1.....</b>	<b>38</b>
<b>De sist utgitte publikasjonene i serien Notater .....</b>	<b>50</b>

## Tabellregister

Tabell 2.1: Vektgrunnlaget for tre forskjellige ekvivalensskalaer der det kun fokuseres på størrelsen til husholdningen. E angir graden av stordriftsfordeler .....	6
Tabell 3.1a: Antall personer etter familiestørrelse (DSF) og husholdningsstørrelse (FoB2001). November 2001 .....	10
Tabell 3.1b: Husholdningstype fra FoB2001 og familietype fra DSF for individer hvor antall medlemmer i husholdningen ikke er lik antall medlemmer i familien. November 2001 .....	11
Tabell 3.2: Desilgruppens gjennomsnittlige ekvivalentinntekt for totalpopulasjonen basert på kvadratrotskalaen med familie og husholdning som økonomisk enhet. År 2000.....	14
Tabell 3.3: Gini-koeffisienten for fordelingen av inntekt etter skatt basert på kvadratrotskalaen med familie og husholdning som økonomisk enhet, for grupper av populasjonen dannet på bakgrunn av familietype fra DSF, alder og utdanningsnivå. År 2000.....	15
Tabell 3.4: Summarisk statistikk som viser avviket mellom Gini-koeffisientene for fordelingen av inntekt etter skatt på kommunenivå basert på kvadratrotskalaen med henholdsvis familie og husholdning som økonomisk enhet. År 2000 .....	16
Tabell 4.1: Fordelingen av samboere uten felles barn fra FoB2001 etter husholdningsstørrelse.....	23
Tabell 4.2a: Gruppeinndeling for kvinner etter utdanningsnivå og alder.....	24
Tabell 4.2b: Gruppeinndeling for menn etter utdanningsnivå og alder.....	24
Tabell 4.5: Fordelingen av de potensielle kvinnelige og mannlige samboerne fra DSF over 15 ulike grupper karakterisert etter alder og utdanning.....	27
Tabell 4.8: Gini-koeffisienten for fordeling av inntekt etter skatt basert på kvadratrotskalaen med familie, husholdning og et utvidet familiebegrep som økonomisk enhet. År 2000.....	31
Tabell 4.9: Desilgruppens gjennomsnittlige ekvivalentinntekt basert på kvadratrotskalaen med et utvidet familiebegrep som økonomisk enhet. År 2000.....	32
Tabell 4.10: Summarisk statistikk som viser nivået på Gini-koeffisientene for fordelingen av inntekt etter skatt på kommunenivå basert på kvadratrotskalaen med familie, husholdning og det utvidete familiebegrepet som økonomisk enhet. År 2000.....	32
Tabell 4.3: Fordelingen av de faktiske samboerne uten barn fra FoB2001 over samboergruppene .....	38
Tabell 4.4: Fordelingen av modellens predikerte samboere uten barn over samboergruppene .....	40
Tabell 4.6: Avvikene i de ulike samboergruppene fra den faktiske fordelingen av samboere over samboergruppene fra FoB2001 ved bruk av den endrede stokastiske matching prosessen .....	42
Tabell 4.7: Kommunetilhørigheten til de 299 samboerparene som modellen ved trekning 1 ikke allokerer en riktig samboergruppe til.....	45

## 1. Innledning \*

Når en analyserer inntektsulikhet og fattigdom i befolkningen, kan en ikke bare ta utgangspunkt i individenes personinntekt. En persons faktiske økonomiske velferd avhenger av husholdningens samlede inntekt, størrelse og sammensetning. Derfor bør en benytte (kost)husholdning som økonomisk enhet i analyse av økonomisk velferd. Det vil si at en må bruke husholdningens samlede disponible inntekt som grunnlag for å tallfeste husholdningsmedlemmenes økonomiske levestandard.

Sammensetningen til en husholdning kartlegges vanligvis ved å intervju personer eller bruke folketellinger som er gjennomført med tiårsintervaller, mens registerdata baseres på familier. En familie kan bestå av en enslig, ugift far/mor med barn, ektepar med og uten barn og samboerpar med minst ett felles barn. I Det sentrale folkeregisteret (DSF) identifiseres alle medlemmene i en familie gjennom sitt felles familienummer. En husholdning omfatter alle som bor og spiser sammen (kosthusholdning). Husholdningen kan altså inneholde mer enn én familie. En analyse av økonomisk velferd basert på DSF koblet med inntektsdata vil dermed ikke nødvendigvis gi et godt bilde av individenes velferd i samfunnet. Det er likevel ofte hensiktsmessig å benytte data fra DSF i slike analyser. Grunnen er at registerdata har en rekke viktige fordeler framfor intervjuundersøkelser og folketellinger:

- Registeropplysningene holdes fortløpende oppdatert. Det er dermed mulig å gjennomføre analyser for tidspunkter der det ikke foreligger folketellinger eller intervjuundersøkelser.
- Registeret inneholder opplysninger om hele befolkningen og gir dermed muligheter for å trekke utsagnskraftige konklusjoner om små grupper.
- Registeropplysningene er uten frafall, slik at eventuelle skjevheter som følge av lav respons på intervju blir borte. Dette gir normalt sikrere anslag på populasjonsstørrelser.
- Registerinformasjonen gir mulighet til å følge personer over tid og dermed undersøke mobilitet i inntekt og inntektsfordeling, også for grupper med relativt få personer. Det er også mulig å undersøke om inntektsulikhet er et forbigående fenomen eller et vedvarende problem, og hva som kjennetegner individer med lav inntekt over lengre tid.

I dette notatet viser vi hvordan vi kan løse det statistiske registerproblemet omtalt ovenfor ved å utvikle, anvende og evaluere en metode for produksjon av husholdningsutgaver av DSF. Et kritisk problem i denne sammenhengen er hvordan en kan predikere hvem som er samboere uten barn. Målet er å utvikle en statistisk modell for samboerskap uten barn basert på opplysninger om familie, husholdning, inntekt, utdanning, alder og bostedsadresse. Modellen er estimert på grunnlag av data fra Folke- og bolig tellingen 2001 (FoB2001) og DSF (2001) og blir brukt til å predikere hvem som er samboere uten barn blant registrerte enslige i DSF. Målet er å videreutvikle den metoden Åserud (2000) benyttet basert på data fra Inntekts- og

---

\* Takk til Audun Langørgen, Bård Lian og Rolf Aaberge for god hjelp og godt samarbeid i arbeidet med å utvikle modellen som er drøftet i dette notatet.

formuesundersøkelsen fra 1997 for å identifisere samboere. En svakhet med Åseruds metode er at han baserte seg på begrenset informasjon for et utvalg av befolkningen. Denne svakheten kan elimineres ved å bruke data fra FoB2001 fordi denne tellingen gir relevante data for hele befolkningen. Vi forventer derfor at metoden som utvikles i prosjektet vil gi et bedre grunnlag for å simulere en husholdningsutgave av DSF, og dermed også et bedre grunnlag for å studere økonomisk velferd i befolkningen. En slik metode kan benyttes i analyser av økonomisk velferd for år hvor vi ikke har data fra folketellinger og ønsker å benytte data for hele befolkningen fra DSF.

I dette notatet drøfter vi mulighetene for å bruke Det sentrale folkeregisteret (DSF) koblet med inntektsdata som grunnlag for studier av økonomisk velferd ved hjelp av en modell som produserer simulerte husholdningsversjoner av DSF. Formålet er å inkludere uregistrerte samlivsformer i DSF, og dermed benytte et tilnærmet husholdningsbegrep og ikke familie som økonomisk enhet. Et slikt simulert husholdningsregister vil kunne gi et bedre bilde av individenes faktiske samlivssituasjon enn registerets familiebegrep.

## 2. Viktige begreper, definisjoner og forutsetninger

Nedenfor vil vi se nærmere på begreper, definisjoner og forutsetninger som er sentrale for analyse av økonomisk velferd. Det er viktig å merke seg at våre analyser og resultater må forstås og fortolkes innenfor rammen av de definisjoner og antakelser som benyttes.

### 2.1 Definisjonen av husholdning, familie og bostedsadresse

I våre analyser i dette notatet vil det være avgjørende hvordan begreper som husholdning, familie og bostedsadresse defineres.<sup>1</sup> Vi vil bruke formell adresse for å matche samboere. Formell adresse defineres her som den bostedsadressen der den enkelte person var registrert bosatt i DSF per 3.11.2001. Det er i dette arbeidet nødvendig å fokusere på formell adresse fordi modellen senere skal kunne brukes til å lage husholdningsversjoner av DSF for år der vi ikke har opplysninger om faktisk adresse. I en analyse der det benyttes formell adresse og ikke faktisk adresse, vil det være verd å merke seg at en del borteboende studenter vil bli feilaktig behandlet som boende hos sine foreldre. Dermed vil de inngå i sine foreldres familie og husholdning. Dette vil ha betydning for både studentene og individene i deres familie fordi inntekten, størrelsen og sammensetningen til familiene og husholdningene avhenger av hvordan studenter behandles. Ifølge FoB2001 var det 82724 borteboende studenter som studerte ved et studiested i Norge på tellingstidspunktet.<sup>2</sup>

For vårt formål er det ønskelig å bruke kosthusholdning som økonomisk enhet fordi vi blant annet ønsker å si noe om stordriftsfordelene i konsumet. Kosthusholdning omfatter personer som bor i samme bolig og har felles kosthold. Dessverre gir ikke FoB2001 noen informasjon om kosthusholdninger. Derfor vil vi ta utgangspunkt i bohusholdning, definert som alle personer bosatt i samme bolig. Kvalitetsundersøkelsen for FoB2001 indikerer imidlertid at avviket mellom disse to husholdningstypene ikke vil være av avgjørende betydning.<sup>3</sup> Denne undersøkelsen består av 8446 intervjuobjekter, 8446 husholdninger og 16231 personer. Ved intervju fremkommer det her at det kun er 157 personer som deler bolig, men ikke kost med intervjuobjektet. Dette kan typisk være individer som bor i bokollektiv.

Det er verd å merke seg at familiebegrepet i DSF skiller seg vesentlig fra familiebegrepet i FoB2001. Familie defineres i DSF ved et familienummer knyttet til hver person. Familienummeret skal holdes løpende oppdatert ved meldinger om fødsler, vigslar, skilsmisser, dødsfall, flyttinger osv. Dette offisielle familienummeret kan enkelt utvides til å behandle samboere med felles barn som en familie. Da vil kvinnens offisielle familienummer bli familiens felles familienummer. Dermed vil samboere med minst ett felles barn få felles familienummer og bli behandlet som en familie i DSF; samboere med felles barn.<sup>4</sup> Derimot vil statistikk basert på registerdata ikke behandle samboere uten felles barn som en familie, men som ulike familier. En familie består i FoB2001 av personer som er bosatt i samme bolig, og som er knyttet til hverandre som ektefeller, registrerte partnere, samboere (med og uten felles barn), og/eller som foreldre og barn (uansett barnets alder). En familie kan høyst bestå av to påfølgende generasjoner. I FoB2001 identifiseres samboere uten felles barn ved hjelp av data fra boligskjemaet og vil bli definert som en familie så sant de tilhører samme bolig.

---

<sup>1</sup> Definisjonene av familie, husholdning og adresse er hentet fra "Om statistikken" på internettsadresse: <http://www.ssb.no/fob2001/om.html>.

<sup>2</sup> I Kirkeberg (2003) drøftes effektene av å utelate studenter og formuende i analyse av fattigdom.

<sup>3</sup> Kvalitetsundersøkelsen for FoB2001 er en del av SSBs Arbeidskraftundersøkelse, 4.kvartal 2001.

<sup>4</sup> Gruppen samboere med felles barn drøftes i Byberg et al. (2001) og Noack (2002).

Dette skiller seg fra DSF der samboere uten felles barn blir regnet som ulike familier. Vår modell skal anvendes på registerdata for å predikere samboere uten barn fra de registrerte enslige. Da er vi nødt til å bruke registerets familiebegrep.

Det er viktig å presisere at selv om familie og husholdning vil være de økonomiske enhetene i våre analyser, vil alle resultatene i dette arbeidet ha individ som analyseenhet.

## **2.2 Økonomisk velferd som begrep**

I økonomisk faglitteratur er begrepet økonomisk velferd ikke tillagt noen allment anerkjent betydning. I dette arbeidet skal vi bruke økonomisk velferd som en betegnelse for inntekten som er tilgjengelig for medlemmene i en husholdning eller i en familie. Vi retter altså ikke fokus på det faktiske konsumet de økonomiske ressursene gir medlemmene, men på mulighetene til konsum. I enkelte analyser av inntektsulikhet og fattigdom fokuseres det på individenes faktiske forbruk. Grunnen til at vi velger å legge inntekt og ikke forbruk til grunn for å definere økonomisk velferd, er at det er rimelig at manglende økonomiske ressurser, her i form av lav inntekt, må være årsaken til det lave forbruket for at en person skal defineres til å ha relativt lav økonomisk velferd. Et lavt forbruk alene gir oss ikke nødvendigvis noe informasjon om individers økonomiske velferd. En som velger å spare, men har høy inntekt, og dermed velger lavt forbruk kan naturligvis ikke oppfattes som et individ med lav økonomisk velferd. Ved å fokusere på inntekt for å definere økonomisk velferd vil individer som har lav lønnsinntekt, men stor formue og høyt faktisk forbruk, kunne være en problemgruppe. Men siden slike individer ofte vil ha aksjeutbytte, renteinntekter eller andre kapitalinntekter av deres formue og dermed en relativt høy inntekt etter skatt på tross av lav lønnsinntekt, vil de ikke bli definert som individer med lav økonomisk velferd. Dette på tross av at de har en lønnsinntekt som isolert ville gi materiell velferd under et minstenivå for samfunnets standarder. I dette arbeidet ser vi bort fra problemet med formuende individer med lav inntekt. En alternativ tilnærming ville vært å fjerne personer som har formue over en viss grense.

I praksis vil vi sjelden kjenne den interne fordelingen av inntekt i en familie eller en husholdning. Derfor vil dette arbeidet følge standard praksis som vil si at alle personene i en familie eller husholdning oppnår samme økonomiske velferd. Det er altså inntektene totalt i husholdningen/familien og husholdningens/familiens størrelse og sammensetning som er avgjørende for et individs økonomiske velferd.

## **2.3 Ekvivalentinntekt**

En analyse av økonomisk velferd i et samfunn krever sammenliknbarhet mellom personer. Ett av problemene vi står ovenfor ved velferdsanalyse er valg av målestokk for sammenlikning av den økonomiske velferden til personer i husholdninger og familier av forskjellig størrelse og sammensetning. Det er en vanlig oppfatning at kostnaden for å oppnå et visst velferdsnivå avhenger av størrelsen og kanskje også sammensetningen til husholdningen eller familien. Et samboerpar vil for eksempel kunne ha betydelige stordriftsfordeler i konsumet ved å dele utgifter til konsumgoder som bil, bolig, hytte, hvitevarer, mat osv. Vi antar at en økning i størrelsen på husholdningen eller familien fra for eksempel en til to personer ikke vil kreve en fordobling av inntekten for å holde velferdsnivået konstant.

For å sammenlikne husholdninger og familier av ulik størrelse og sammensetning ønsker vi å bruke en ekvivalensskala som omregner familie- eller husholdningsinntektene til sammenliknbare personinntekter kalt ekvivalentinntekter. I dette arbeidet vil vi anta at ekvivalensskalaene er uavhengige av hvilke inntekter husholdningen eller familien har. Denne antakelsen er forenkende og vanlig, men det kan være grunn til å stille spørsmål ved hvor rimelig en antakelse om uavhengighet er. Det finnes skalaer som ikke antar uavhengighet av inntektsnivå, men disse skalaene er mer kompliserte og ikke like vanlige i økonomisk litteratur.<sup>5</sup> I inntektsavhengige skalaer antas det ofte at stordriftsfordelene i konsumet øker med inntekten, med andre ord at økningen i husholdningsstørrelsen blir tillagt større vekt når den enslige er fattig enn når personen er velstående. Begrunnelsen for dette er at husholdninger med lave inntekter må prioritere nødvendighetsgoder som mat og klær, og ikke luksusgoder hvor stordriftsfordelene i konsumet er relativt større. Når studier begrenses til husholdninger med lave inntekter, vil bruk av en inntektsavhengig ekvivalensskala være mindre relevant.

Ekvivalentinntekten til et individ i en husholdning konstrueres ved å dividere husholdningsinntekten på en ekvivalensvekt. Ekvivalentinntektene til medlemmene i en husholdning indikerer forbrukspotensialet til hvert medlem av husholdningen, forutsatt at forbruket skal fordeles likt på husholdningsmedlemmene. For å finne ekvivalentinntekten kan vi vekte husholdningens totale inntekt etter husholdningens størrelse og sammensetning. Vekten skal representere en avveining mellom betydningen av felles konsum (stordriftsfordeler) og privat konsum. Ekvivalentinntekten til medlemmer av en familie defineres tilsvarende. Det er viktig å merke seg at valg av vekt vil være et verdivalg.

Det finnes en rekke forskjellige inntektsuavhengige ekvivalensskalaer som blir brukt til analyser av økonomisk velferd. I Norge og andre OECD-land er det vanlig å bruke OECD-skalaen som målestokk, selv om skalaen verken har et mer overbevisende teoretisk eller empirisk fundament enn mange alternative skalaer.<sup>6</sup> En annen mye brukt ekvivalensskala er den såkalte kvadratrotskalaen som i motsetning til OECD-skalaen ikke skiller mellom voksne og barn i husholdningen og familien. OECD-skalaen er blitt kritisert for at den vektlegger stordriftsfordelene for lite i flerpersonghusholdninger og flerpersongfamilier. Denne kritikken er særlig relevant når analysen er rettet mot fordelingen av inntekt over hele befolkningen.

I dette notatet skal vi benytte kvadratrotskalaen. Når vi kun fokuserer på størrelsen til husholdningen og ikke på sammensetningen, kan vi generelt definere husholdningens ekvivalentinntekt slik:

$$y_j = \sum_{j \in h} x_j / u_h^E \text{ for } j \in h.$$

Her betegner  $h$  husholdningen som individ  $j$  tilhører,  $x_j$  er individets inntekt og  $u_h$  gir oss husholdningsstørrelsen. Dette innebærer at ekvivalentinntekten,  $y_j$ , er identisk for alle medlemmene av samme husholdning.  $E$  er en parameter som viser graden av stordriftsfordeler og kan ta verdier mellom 0 og 1. Ved å bruke  $E=1$  ser vi helt bort i fra stordriftsfordelene i konsumet og antar at alt konsum er privat konsum. Da må en husholdning med 4 personer ha

<sup>5</sup> Se Van Praag og Van der Sar (1988) og Aaberge og Melby (1998) for nærmere drøfting omkring inntektsuavhengige ekvivalensskalaer.

<sup>6</sup> Aaberge og Melby (1998).



4 ganger så høy inntekt som en husholding med 1 person for at individer skal ha lik ekvivalentinntekt. Hvis vi velger  $E=0$  vil vi tillegge felles konsum maksimal betydning. Begge disse ytterpunktene virker urimelige. En mye brukt ekvivalensskala er  $E=0,5$ , tidligere omtalt som kvadratrotskalaen. Tabellen nedenfor illustrerer betydningen av valg av vekt.

**Tabell 2.1: Vektgrunnlaget for tre forskjellige ekvivalensskalaer der det kun fokuseres på størrelsen til husholdningen. E angir graden av stordriftsfordeler**

Husholdningsstørrelse	$E=0,7$	$E=0,5$	$E=0,3$
2 personer	1,625	1,414	1,231
4 personer	2,639	2,000	1,516
6 personer	3,505	2,450	1,712

Når vi benytter oss av ekvivalensskalaer med små stordriftsfordeler i husholdingen, vil personer i store husholdninger få lavere ekvivalentinntekt enn personer i små husholdninger relativt til om vi bruker ekvivalensskalaer med store stordriftsfordeler. Dette vil kunne ha implikasjoner for analyse av økonomisk velferd i et samfunn.<sup>7</sup>

På tilsvarende måte blir familiens ekvivalentinntekt definert. Eneste forskjell er at nå blir familie og ikke husholding benyttet som økonomisk enhet.

## 2.4 Inntektsbegrepet

Til nå har vi brukt begrepet inntekt uten å ha gitt en presis definisjon. Tradisjonelt defineres inntekt i økonomisk litteratur som det maksimale forbruket en kan tillate seg uten at nettoformuen reduseres. Fordi det er mangelfulle data av nettoformue bruker vi i en grovere definisjon som er mye brukt i analyse av inntektsulikhet, *inntekt etter skatt*. Vi vil definere inntekt etter skatt tilsvarende som blant annet Aaberge et al. (1999). Det er verd å merke seg at selv om dette inntektsbegrepet kan sies å være en god indikator på husholdingenes økonomiske ressurser, så inkluderer det ikke alle viktige inntekter som vil ha betydning for aktørenes økonomiske velferd. Dette vil for eksempel gjelde verdien av offentlige tjenester, verdien av hjemmeproduksjon og "svarte" arbeidsinntekter.

Vår definisjon av inntekt etter skatt kan grovt skisseres slik:

$$\begin{aligned} \text{Inntekt etter skatt} &= \text{Lønnsinntekt} \\ &+ \text{Netto næringsinntekt før fondsavsetninger og avskrivninger} \\ &+ \text{Brutto kapitalinntekt} \\ &+ \text{Overføringer} \\ &- \text{Skatt} \end{aligned}$$

I definisjonen av inntekt etter skatt ovenfor benyttes brutto kapitalinntekt og det er ikke gjort noen fratrukk for renteutgifter i definisjonen. En alternativ innteksdefinisjon som er mye brukt i statistikken er *disponibel inntekt*, hvor kapitalinntekten bli regnet netto, og derfor fratrukket gjeldsrenter. Begrunnelsen for å ikke trekke fra gjeldsrentene er gjort ut fra at norsk inntektsstatistikken mangler et egnet mål på inntektsstrømmen av å ha bolig. Dagens praksis hvor inntekt av egen bolig kun verdsettes til 2,5 prosent av boligens ligningsverdi fratrukket et bunnfradrag, vil ifølge Epland (1998) føre til en klar undervurdering av inntekt av egen bolig fordi ligningsverdien i de fleste tilfeller vil ligge langt under markedsverdien. En mulig

<sup>7</sup> For drøfting omkring valg av ekvivalensskala se Atkinson et. al. (1995), Epland (1998) og Lund og Aaberge (1999).

tilnærming ville vært å estimere inntekt av egen bolig tilsvarende det beløpet en ville fått dersom boligen blir utleid til markedspris. Men denne tilnærmingen kan være problematisk å gjennomføre i praksis fordi dette krever relativt nøyaktige boligopplysninger for hvert enkelt individ.

I dette arbeidet vil familier og husholdninger med personer som har negativ privat inntekt etter skatt bli holdt utenfor analysen. Dette er en liten gruppe som ikke vil ha vesentlig betydning for analysen. Dessuten har vi også gjort noen justeringer for personer med negativ kapitalinntekt eller negativ næringsinntekt slik at disse inntektskomponentene i slike tilfeller blir satt lik null. Dette er i tråd med Langørgen og Aaberge (2002).

Valg av inntektsbegrep vil kunne være av avgjørende betydning for resultatene av analyser av økonomisk velferd. Derfor er det viktig å presisere at resultatene i dette notatet avhenger av vår definisjon av inntekt. En studie av Andersen et al. (2003) drøfter hvordan valg av inntektsbegrep påvirker resultater fra analyser av temporær og kronisk fattigdom. Videre i notatet vil vi omtale inntekt etter skatt som inntekt hvis ikke annet presiseres.

## **2.5 Datamateriale**

Det empiriske materialet som vi benytter i våre analyser i dette arbeidet, er hentet fra SSB sin FoB2001 gjennomført 3. november 2001. Dette suppleres med informasjon fra DSF fra 3. november 2001 og inntektsdata for år 2000. Vår basispopulasjon omfatter alle personer som ifølge DSF var bosatt i Norge 3. november 2001, og som da er inkludert i FoB2001. Dette er 4520947 personer. Lov om folkeregistrering fra 16. januar 1970 og forskrifter fra 1994 definerer hvem som er bosatt i Norge, og hvor de skal ha adresse. I våre analyser vil det være et avvik i tid mellom informasjon fra DSF 2001 og FoB2001 (begge fra 3. november 2001) og inntektsdata fra året 2000, fordi inntektsdata fra året 2001 foreløpig ikke har vært tilgjengelig.

I dette arbeidet trekker vi veksler på framgangsmåten i Åserud (2000). I tillegg tar vi i bruk relevant sekundærlitteratur for å sammenlikne og diskutere funnene i våre analyser.

Alle beregninger og resultater i dette arbeidet er gjennomført med programpakken SAS.

### 3. Analyse av inntektsulikhet ved bruk av familie og husholdning som økonomiske enheter

I kapittel 3 vil vi kartlegge avviket mellom opplysninger om familiesammensetningen fra DSF og opplysninger om husholdningssammensetningen fra FoB2001, og studere forskjellene i inntektsulikhet ved å bruke familie og husholdning som økonomisk enhet. Vi vil derfor først gi en omtale av metodene for måling av inntektsulikhet.

#### 3.1 Mål for inntektsulikhet

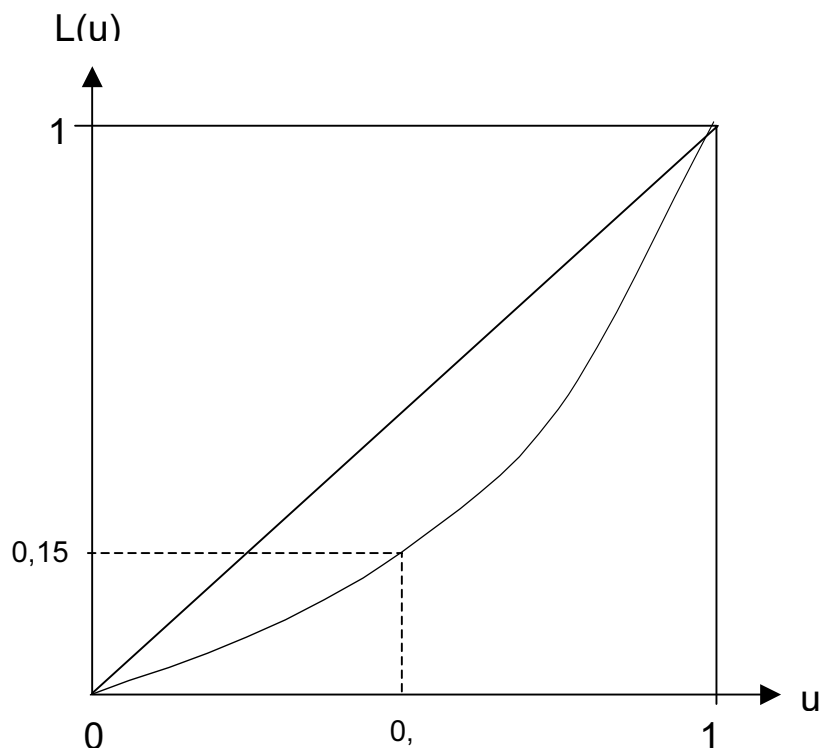
Vi vil i våre ulikhetsanalyser benytte vanlige metoder som Gini-koeffisient og desiler for å belyse fordelingen av inntekt i samfunnet. Utgangspunktet for Gini-koeffisienten er Lorenzkurven til den kumulative inntektsfordelingen. Vi lar  $X$  være en inntektsvariabel med den kumulative fordelingen  $F(\cdot)$  som er definert over  $[0, \infty)$  og med gjennomsnittet  $\mu$ . Her vil  $F^{-1}(0) \equiv 0$ . Lorenzkurven defineres da formelt som:

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

Lorenzkurven viser grafisk for hver  $u$  mellom 0 og 1, andelen av total inntekt som tilfaller den  $100u$  prosent av befolkningen med lavest inntekt, og er uavhengig av måleenhet.

Hvis alle enheter har samme inntekt, vil  $L(u)=u$  for alle  $u$ . Da vil Lorenzkurven være en rett linje med 45 graders vinkel, ofte omtalt som likhetskurven. Lorenzkurven ligger helt eller delvis under denne likhetskurven så sant noen inntekter er forskjellige. Hvis en person eier all inntekt vil Lorenzkurven følge x-aksen og  $L(u)=0$  før den spretter vertikalt opp til  $L(u)=1$  når  $u=1$ . Jo større inntektsulikheten i samfunnet er, jo lavere vil Lorenzkurven ligge.

**Figur 3.1: Eksempel på Lorenzkurve**



Gini-koeffisienten (G) defineres ved å måle avviket mellom Lorenzkurven og likhetskurven. Gini-koeffisienten er formelt definert ved:

$$G = 2 \int_0^1 (u - L(u)) du = 1 - 2 \int_0^1 L(u) du . ^8$$

Koeffisienten måles altså ved å multiplisere arealet mellom  $u$  og  $L(u)$  med 2. Parameteren  $G$  vil kunne variere mellom 0 og 1. Hvis avviket mellom Lorenzkurven og likhetskurven er 0, vil  $G=0$  og all inntekt er likt fordelt. Det andre ytterpunktet er  $G=1$  som innebærer at det er en person som mottar all inntekt, og de andre ikke mottar noe. Ulikheten vil altså være større jo større  $G$  er, med andre ord jo mer inntektsfordelingen avviker fra fullstendig likhet.

En annen metode vi vil benytte i vår analyse av inntektsfordelingen i samfunnet er desiler. Mens Gini-koeffisienten oppsummerer ulikheten i fordelingen av inntekt i en populasjon i et enkelt tall mellom 0 og 1, vil desilgjennomsnittene gi oss mer detaljert informasjon om inntektsulikheten. Desiltabeller fremkommer ved å dele inn populasjonen i 10 like store grupper. Desil 1 omfatter de ti prosent av populasjonen med lavest inntekt, mens desil 10 omfatter de ti prosent av populasjonen med høyest inntekt. Hvis inntektsfordelingen er helt jevn, vil alle de ti gruppene disponere 10 prosent hver av totalinntekten. I dette arbeidet vil vi fokusere på desilgjennomsnittet. Tolkningen av gjennomsnittsinntekten til desilgruppe 1 er gjennomsnittsinntekten til populasjonens 10 prosent fattigste. I en analyse av inntektsulikhet vil ofte inntektssammensetningen i en desilgruppene være interessant. For eksempel er det ikke slik at vi ved analyse av fattigdom bør fokusere ensidig på gjennomsnittsinntekten for de ti prosent fattigste. Ofte bør det suppleres med informasjon om inntektssammensetningen blant de fattige. Da får vi et bilde av den relative forskjellen i inntekt i desilgruppe 1. I dette arbeidet vil vi begrense oss til å kun se på de gjennomsnittlige ekvivalentinntekten til desilgruppene.

### **3.2 En kartlegging av avviket mellom opplysninger om familie fra DSF og opplysninger om husholdningen fra FoB2001.**

I dette delkapittelet vil vi studere forskjellene mellom familiesammensetningen i DSF og husholdningssammensetningen i FoB2001. En kartlegging av avviket mellom registeropplysningene og opplysningene fra folketellingen vil kunne gi en indikasjon på hvor egnet registerdata er som datagrunnlag for økonomiske analyser sammenliknet med data fra folketellinger. Problemstillingen som kan gjøre seg gjeldende, er at avviket mellom familie- og husholdningsbegrepet er for stort til at familie vil være egnet som økonomisk enhet i analyser av økonomisk velferd.

Basispopulasjon omfatter alle personer som ifølge DSF var bosatt i Norge 3. november 2001, og som da er inkludert i FoB2001. Fra basispopulasjonen fjernes individer fra husholdninger og familier med mer enn 17 medlemmer og individer fra felleleshusholdninger. En felleleshusholdning består av personer som er bosatt og får dekket sine behov for kost, pleie og omsorg på en institusjon. Til sammen er 35020 individer fjernet fra basispopulasjonen.

Tabell 3.1a viser forskjellen mellom registeropplysninger om familie fra DSF og opplysninger fra FoB2001 om husholdninger. Resultatene presentert i tabellen viser at det er et vesentlig

---

<sup>8</sup> Se for eksempel Bojer (1990) for en nærmere drøfting av ulikhetsmål.

avvik mellom familiestørrelsen og husholdningsstørrelsen. Av tabell 3.1a ser vi at 10,5 prosent av totalpopulasjonen (473054 individer) har annerledes familiestørrelse enn husholdningsstørrelse. Av disse har 443717 individer flere medlemmer i husholdningen enn familien, mens kun 29337 individer har større familiestørrelse enn husholdningsstørrelse. Videre viser tabellen at omtrent 39 prosent av totalt antall individer med avvik mellom husholdnings- og familiestørrelsen, har familiestørrelse lik 1 og husholdningsstørrelse lik 2. En stor del av disse er samboere uten felles barn.

**Tabell 3.1a: Antall personer etter familiestørrelse (DSF) og husholdningsstørrelse (FoB2001). November 2001**

Familiestørrelsen ifølge DSF	Husholdningsstørrelsen ifølge FoB2001						
	1	2	3	4	5	6 eller flere	Alle
1	733120	183535	34352	17601	8314	5046	981968
2	1013	879734	54240	15133	7274	5018	962412
3	1654	6296	711166	42640	11306	7622	780684
4	1948	696	7256	984400	25983	10065	1030348
5	1307	169	456	5599	530406	15588	553525
6 eller flere	521	64	40	231	2087	174047	176990
Alle	739563	1070494	807510	1065604	585370	217386	4485927

Tabell 3.1a viser at 4012873 individer har like mange medlemmer i husholdningen som i familien. Likevel vil familiesammensetningen og husholdningssammensetningen til disse individene kunne avvike. Medlemmene i et individs familie og husholdning behøver ikke være identiske selv om husholdningstørrelsen er lik familiestørrelsen.

For å studere samsvaret mellom opplysningene om husholdningssammensetningen fra FoB2001 og opplysningene om familiesammensetningen fra DSF, ser vi nærmere på familie- og husholdningssammensetningen til alle individer som lever i husholdninger og familier med familie- og husholdningsstørrelse mindre enn 6; til sammen 4264433 individer. Av disse avviker familiestørrelsen fra husholdningsstørrelsen hos 425799 individer. Til sammenlikning er det kun 2734 individer der familiestørrelsen er lik husholdningsstørrelsen og medlemmene i husholdningen ikke er identiske med medlemmene i familien. Dette viser at den største delen av avviket mellom opplysninger om husholdningssammensetningen fra folketellinger og opplysninger om familiesammensetningen fra registeret, fanges opp ved å studere avvik i antall medlemmer i husholdningen i forhold til familien.

Tabell 3.1b viser husholdningstypen og familietyper til individer hvor familiestørrelsen avviker fra husholdningsstørrelsen. Årsaken til at summen av individer med avvik mellom familiestørrelse og husholdningsstørrelse er større i tabell 3.1b enn tabell 3.1a, er at vi i tabell 3.1a ikke får frem avvik mellom familiestørrelse og husholdningstørrelse for individer som har familiestørrelse og husholdningsstørrelse større enn 6. Ved å studere familietyper og husholdningstypen til individer hvor familiestørrelsen avviker fra husholdningsstørrelsen, kan vi få et nærmere innblikk i sentrale kjennetegn til individer hvor familieopplysningene ikke samsvarer med husholdningsopplysningene.

**Tabell 3.1b: Husholdningstype fra FoB2001 og familietype fra DSF for individer hvor antall medlemmer i husholdningen ikke er lik antall medlemmer i familien. November 2001**

Antall Prosent Rad Prosent Kolonne prosent	Familietype									
	Husholdningstype	Enslig	Ektepar uten hjemmeboende barn	Ektepar med hjemmeboende barn	Enslig mor med barn	Enslig far med barn	Samboerpar med felles barn	Registrert partnerskap uten barn	Registrert partnerskap med barn	Andre med flere
	70734	0	5760	23041	5159	140	0	0	64	104898
	14,73	0	1,2	4,8	1,07	0,03	0	0	0,01	21,85
	67,43	0	5,49	21,97	4,92	0,13	0	0	0,06	
Ikke i noe parforhold	28,43	0	6,44	26,48	22,98	1,55	0	0	96,97	
	1171	23214	46154	78	30	0	0	0	2	70649
	0,24	4,83	9,61	0,02	0,01	0	0	0	0	14,71
	1,66	32,86	65,33	0,11	0,04	0	0	0	0	
Ektepar	0,47	100	51,6	0,09	0,13	0	0	0	3,03	
	0	0	0	0	0	0	110	8	0	118
	0	0	0	0	0	0	0,02	0	0	0,02
	0	0	0	0	0	0	93,22	6,78	0	
Registrerte partnere	0	0	0	0	0	0	100	61,54	0	
	3480	0	1262	3631	2196	4813	0	0	0	15382
	0,72	0	0,26	0,76	0,46	1	0	0	0	3,2
	22,62	0	8,2	23,61	14,28	31,29	0	0	0	
Samboere med felles barn	1,4	0	1,41	4,17	9,78	53,15	0	0	0	
	168116	0	2470	22396	5079	49	0	0	0	198110
	35,01	0	0,51	4,66	1,06	0,01	0	0	0	41,26
	84,86	0	1,25	11,3	2,56	0,02	0	0	0	
Samboere uten felles barn, motsatt kjønn	67,58	0	2,76	25,74	22,62	0,54	0	0	0	
	4235	0	22	291	39	1	0	0	0	4588
	0,88	0	0	0,06	0,01	0	0	0	0	0,96
	92,31	0	0,48	6,34	0,85	0,02	0	0	0	
Samboere uten felles barn, samme kjønn	1,7	0	0,02	0,33	0,17	0,01	0	0	0	
	1	0	33646	124	35	0	0	5	0	33811
	0	0	7,01	0,03	0,01	0	0	0	0	7,04
	0	0	99,51	0,37	0,1	0	0	0,01	0	
Barn bosatt med ektepar/registrerte partnere	0	0	37,61	0,14	0,16	0	0	38,46	0	
	34	0	0	37282	9860	4048	0	0	0	51224
	0,01	0	0	7,76	2,05	0,84	0	0	0	10,67
	0,07	0	0	72,78	19,25	7,9	0	0	0	
Barn bosatt med samboere	0,01	0	0	42,85	43,92	44,7	0	0	0	
	994	0	137	167	51	4	0	0	0	1353
	0,21	0	0,03	0,03	0,01	0	0	0	0	0,28
	73,47	0	10,13	12,34	3,77	0,3	0	0	0	
Samboere med felles barn utenfor husholdningen	0,4	0	0,15	0,19	0,23	0,04	0	0	0	
Sum	248765	23214	89451	87010	22449	9055	110	13	66	480133
	51,81	4,83	18,63	18,12	4,68	1,89	0,02	0	0,01	100

Tabell 3.1b viser at nesten 96 prosent av alle individene hvor familiestørrelsen avviker fra husholdningsstørrelsen har en av disse husholdningstypene: Ikke i noe parforhold, ektepar, samboere av motsatt kjønn uten felles barn og barn bosatt med samboere/ektepar/registrerte partnere. Nedenfor vil vi både se nærmere på hovedtrekkene som karakteriserer individene

med avvik mellom husholdningsstørrelse og familiestørrelse i disse husholdningsgruppene og peke på flere mulige forklaringer til dette avviket:

1. Samboere av motsatt kjønn uten felles barn.
  - a. Samboere uten barn hvor:
    - i. Begge partene er registrert enslige i DSF. Partene vil da tilhøre hver sin familie, men en felles husholdning. Til sammen er dette 141000 individer og omtrent 30 prosent av antall individer i tabell 3.1b.
    - ii. Minst en av partene ikke er registrert enslig. Dette kan for eksempel skyldes at samboerparet lever sammen med mor og far til den ene parten. Da vil de to parene tilhøre samme husholdning i FoB2001, men ulike familier i DSF.
  - b. Samboere med særkullsbarn<sup>9</sup> hvor :
    - i. Begge har særkullsbarn og partene vil tilhøre hver sin familie, men en felles husholdning. Familietypene vil være *enslig mor med barn* kombinert med *enslig far med barn*.
    - ii. Kun en part har særkullsbarn og partene vil tilhøre hver sin familie, men en felles husholdning. Familietypene vil være *enslig mor/far med barn* kombinert med *enslig*.
2. Ikke i noe parforhold.
  - a. Kan for det første skyldes individer som lever i samliv bestående av flere generasjoner. En familie i DSF kan i høyden bestå av to påfølgende generasjoner. Dette vil ofte være ektepar med barn eller enslig mor eller far med barn som bor med sin enslige mor/far som har husholdningstype *ikke i noe parforhold*. Individer i et slikt samliv vil kunne tilhøre en felles husholdning, men flere familier.
  - b. Kan for det andre skyldes bokollektiv av individer som ikke lever i ett parforhold. Individene som lever i et slikt samliv er i DSF registrert som enslige eller enslige foreldre med barn, mens de ifølge FoB2001 er medlemmer av samme husholdning.
  - c. Kan for det tredje skyldes individer som har vært gift/samboere eller fått barn, men som på tellingstidspunktet er enslig mor/far med eller uten barn og bor hos sine foreldre. På grunn av at individer som har giftet seg og /eller fått barn aldri igjen blir regnet til foreldrenes familie i DSF, vil individer i et slikt samliv om de flytter tilbake til sine foreldre tilhøre samme husholdning som sine foreldre i FoB2001, men flere familier i DSF.
3. Ektepar.
  - a. En del individer har *ektepar* som husholdningstype og *ektepar med/uten hjemmeboende* barn som familietype. Forskjell i familiestørrelse og husholdningsstørrelse for disse individene kan for det første skyldes at ekteparet lever sammen med minst ett annet par (vennepar, foreldre til den ene parten, søsken med partner, osv.). For det andre kan det forklares ut fra tilsvarende begrunnelser som er gitt i punkt 2a og 2c.

---

<sup>9</sup> Samboerskap med særkullsbarn innebærer at et par er samboere uten felles barn, men der minst en av partene har barn med en annen person. Da vil for eksempel kvinnen som har barn med en annen mann, være registrert som enslig mor med barn, mens samboeren kan være registrert som enslig. Deres husholdningstørrelse vil ikke være to, men tre eller flere alt etter hvor mange særkullsbarn husholdningen har.

4. Barn bosatt med ektepar/samboerpar.
  - a. Skyldes at barn som lever i et samliv i et av de tre punktene omtalt ovenfor vil ha ulik husholdningssammensetning og familiesammensetning. Det er verd å merke seg at i DSF vil barn tilhøre sin foreldres familie uavhengig av alder, mens i FoB2001 må personen være under 18 for å være en del av sine foreldres familie.

På bakgrunn av tabell 3,1b kan vi ikke fastslå hvor stor del av avviket mellom opplysninger om familiesammensetningen fra DSF og opplysninger om husholdningssammensetningen fra FoB2001 som skyldes flergenerasjoner, særkullsbarn, bokollektiv, osv. En slik tallfesting krever en nærmere studie av familiesammensetningen og husholdningssammensetningen til individene hvor husholdningsstørrelsen avviker fra familiestørrelsen.

I tillegg til årsakene diskutert i de fire punktene ovenfor kan en del av avviket mellom opplysninger om familiesammensetningen fra DSF og opplysninger om husholdningssammensetningen fra FoB2001 skyldes feil i DSF, feil i FoB2001 eller at DSF ikke har like oppdaterte opplysninger om individenes samlivsform som FoB2001.

### **3.3 Analyse av inntektsulikhet for totalpopulasjonen, i spesielle grupper i samfunnet og på kommunalt nivå med husholding og familie som økonomisk enhet.**

Fra delkapittel 3.2 fremgikk det at det var et vesentlig avvik mellom opplysninger om familiesammensetningen fra DSF og opplysninger om husholdningssammensetningen i FoB2001. Videre i kapitlet undersøker vi hvordan valg av datagrunnlag påvirker inntektsulikheten for totalpopulasjonen, i spesielle grupper i samfunnet og blant de enkelte kommunene. For å analysere inntektsulikhet vil vi benytte programmet Uroteng konstruert av Rolf Aaberge og Tom Wennemo fra Statistisk sentralbyrå (SSB). Ved bruk av dette programmet kan vi estimere Gini-koeffisienter og desilgjennomsnitt. Programmet skal brukes på data fra FoB2001 og DSF koblet med inntektsdata for år 2000. Da vil vi få mulighet til å sammenlikne analyse av inntektsulikhet ved å bruke familie og husholdning som økonomisk enhet. Ved analyse basert på data fra FoB2001 vil vi beregne individenes ekvivalentinntekt ved å bruke husholdningens samlede inntekt og husholdningsstørrelsen. Når vi benytter data fra DSF, finner vi individenes ekvivalentinntekt ved å bruke familiens samlede inntekt og familiestørrelsen.

Fra basispopulasjonen som omfatter alle personer registrert boende i Norge 3. november 2001, tar vi bort individer hvor det mangler sentral informasjon og individer fra familier og husholdninger med medlemmer som har negativ privat inntekt etter skatt. I tillegg vil vi se bort fra husholdninger og familier med mer enn 17 medlemmer, som typisk kan være en institusjon. Til sammen fjerner vi omtrent 125000 personer fra basispopulasjonen. Vår populasjon består da av 4394931 individer.

Gini-koeffisienten for fordelingen av inntekt etter skatt for totalpopulasjonen når vi benytter kvadratrotskalaen, er 0,271 med familie som enhet og 0,265 med husholdning som enhet. Gini-koeffisienten er altså høyere ved familie som enhet enn ved husholdning som enhet. Riktignok er ikke forskjellen stor, men vi kan fastslå at bruk av familie som enhet vil gi en noe høyere inntektsulikhet. Dette er i samsvar med resultatene til Åserud (2000) basert på data fra en utvalgsundersøkelse.



Tabell 3.2 gir en mer detaljert beskrivelse av fordelingen av inntekt og forskjellen i inntektsulikhet ved å bruke familie og husholdning som økonomisk enhet for totalpopulasjonen. Resultatene presentert i tabellen bekrefter inntrykket av at inntektsulikheten er noe større med familie som økonomisk enhet enn med husholdning som økonomisk enhet. Kolonne 4 og 7 viser desilgruppens relative gjennomsnittsinntekt; gjennomsnittsinntekten til desilgruppen målt i forhold til populasjonens gjennomsnittsinntekt. En sammenlikning av desilgruppens relative gjennomsnittsinntekt ved å bruke familie som enhet med desilgruppens relative gjennomsnittsinntekt ved å benytte husholdning som enhet, viser at de relative gjennomsnittsinntektene er lavere i desilgruppe 1 og 2 og høyere i desilgruppe 9 og 10 med familie som enhet enn med husholdning som enhet. I de andre desilgruppene er de relative gjennomsnittsinntektene omtrent like med husholdning som enhet som med familie som enhet.

**Tabell 3.2: Desilgruppens gjennomsnittlige ekvivalentinntekt for totalpopulasjonen basert på kvadratrotskalaen med familie og husholdning som økonomisk enhet. År 2000**

Økonomisk enhet	Husholdning			Familie		
	Desilgruppe	Gj.snitt (kr)	Desilgr. gj.snitt/ populasjonens gj.snitt	Desilgruppe	Gj.snitt (kr)	Desilgr. gj.snitt/ populasjonens gj.snitt
	0.00-0.10	89800	0,39	0.00-0.10	83200	0,37
	0.10-0.20	132600	0,58	0.10-0.20	126800	0,57
	0.20-0.30	157000	0,68	0.20-0.30	150600	0,68
	0.30-0.40	177600	0,77	0.30-0.40	171000	0,77
	0.40-0.50	196200	0,85	0.40-0.50	189600	0,85
	0.50-0.60	214800	0,93	0.50-0.60	208000	0,93
	0.60-0.70	235600	1,02	0.60-0.70	228200	1,02
	0.70-0.80	261500	1,14	0.70-0.80	253700	1,14
	0.80-0.90	300000	1,30	0.80-0.90	292100	1,31
	0.90-1.00	535000	2,33	0.90-1.00	525300	2,36

En forklaringen på at inntektsulikheten er større når vi benytter familie som enhet istedenfor husholdning som enhet, er at familiebegrepet gir flere enslige enn husholdningsbegrepet. Mange av de som er enslige i DSF, tilhører egentlig en husholdning med flere medlemmer. Hvis inntektene er helt jevnt fordelt i samfunnet, vil dette føre til større ulikhet fordi de enslige ikke har stordriftsfordeler i konsumet og dermed får lavere ekvivalentinntekt. Er det slik at de som er enslige ved bruk av familie som enhet, og ikke enslige ved bruk av husholdning som enhet, i tillegg er relativt fattige, vil ulikheten i samfunnet bli ytterligere forsterket ved bruk av familie som enhet.<sup>10</sup>

Vi så at det er relativt liten forskjell i inntektsulikhet for totalpopulasjonen ved å benytte familie og husholdning som enhet. Men selv om forskjellen i inntektsulikhet er relativt liten for totalpopulasjonen, vil det kunne være vesentlig større forskjell ved analyse av spesielle grupper i samfunnet eller av enkelte kommuner. Gini-koeffisientene ved slike analyser måler inntektsulikheten blant individene i de enkelte gruppene.

<sup>10</sup> Tabell C.4. Aaberge og Langørgen (2003) viser individene registrert bosatte i Norge sin inntektsfordeling i desiler etter alder kombinert med familietype (enslige/aleneførelde eller par). Her benyttes et inntektsbegrep korrigeret for kommunale tjenesters bidrag til økonomisk velferd. Denne tabellen viser at en stor del av individene i laveste inntektsdesil er enslige mellom 16 og 66 år, mens en stor del av individene i det rikeste desilet tilhører et par og er mellom 16 og 66 år.

Tabell 3.3 illustrerer at det vil kunne være både større og mindre forskjeller i inntektsulikhet avhengig av valg av økonomisk enhet ved analyse av spesielle grupper i samfunnet, i forhold til analysen for totalpopulasjonen. For de fleste gruppene i tabell 3.3 er avviket i Gini-koeffisient med husholdning og familie som enhet omtrent like stort som for totalpopulasjonen. Det kan være verd å merke seg differansen i Gini-koeffisient i aldersgruppen 16-29 år med henholdsvis familie og husholdning som økonomisk enhet. I denne aldersgruppen er Gini-koeffisienten 0,14 større med familie som enhet enn med husholdning som enhet. Hva kan dette skyldes? Dette er en aldersgruppe hvor en betydelig andel har relativt lave private inntekter etter skatt, for eksempel studenter med lav lønnsinntekt eller individer som lever i et samliv der en av partene er hjemmeværende med barn og uten lønnsinntekt eller deltidsarbeidende. En analyse med husholdning som enhet vil dermed kunne redusere inntektsulikheten i gruppen i forhold til om en benytter familie som enhet, fordi vi vil kunne få en utjevning av inntekten ved å legge sammen inntektene til samboerpar med ulik privat inntekt.

**Tabell 3.3: Gini-koeffisienten for fordelingen av inntekt etter skatt basert på kvadratrotskalaen med familie og husholdning som økonomisk enhet, for grupper av populasjonen dannet på bakgrunn av familietype fra DSF, alder og utdanningsnivå. År 2000**

	Økonomisk enhet:	
	Familie	Husholdning
	Gini-koeffisient	
<u>Familietype fra DSF:</u>		
Enslig	0,299	0,302
Ektepar uten barn	0,275	0,274
Ektepar med barn	0,240	0,241
Enslig mor med barn	0,212	0,232
Enslig far med barn	0,268	0,260
Samboerpar med minst ett felles barn	0,193	0,193
Partnerskap uten barn	0,220	0,219
Partnerskap med barn	0,159	0,160
<u>Aldersgruppe:</u>		
0-15	0,243	0,236
16-29	0,282	0,268
30-44	0,244	0,241
45-66	0,273	0,267
67-79	0,244	0,244
80 år og eldre	0,274	0,268
<u>Utdanningsgruppe:</u>		
Kun grunnskole, ingen eller uoppgitt utdanning	0,256	0,248
Videregående skole	0,260	0,255
Universitets- eller høyskoleutdanning	0,282	0,276

Ved å studere avvikene i inntektsulikhet i tabell 3.3 nærmere, ser vi at Gini-koeffisientene til de ulike gruppene i de fleste tilfeller er større med familie som enhet enn med husholdning som enhet, akkurat som resultatet var for totalpopulasjonen. Men det er et par unntak, for eksempel er Gini-koeffisienten for de registrerte enslige større med husholdning som enhet enn med familie som enhet. Vi undervurderer altså inntektsulikheten blant de registrerte enslige ved å benytte familie og ikke husholdning som enhet. En kunne jo tro at inntektsulikheten ville reduseres ved å bruke husholdning som enhet, fordi vi vil få en

utjevning av inntekten ved å legge sammen inntektene til samboerpar med ulik privat inntekt. Økningen i inntektsulikhet ved å benytte husholdning og ikke familie som enhet skyldes trolig at individene som ifølge DSF er enslige, men som ifølge FoB2001 ikke er enslige (for eksempel samboere uten barn), jevnt over har relativt høyere privat inntekt etter skatt enn individene som ifølge både FoB2001 og DSF er enslige. Dermed vil individene som i utgangspunktet har relativt høy privat inntekt etter skatt, få enda høyere ekvivalentinntekt ved å benytte husholdning som enhet i forhold til om familie var den økonomiske enheten. Dette kan forklare at inntektsulikheten øker i gruppen registrert enslige om husholdning benyttes som den økonomiske enheten, i forhold til om analysen er basert på familie som enhet. Faktisk har i overkant av 60 % av de som ifølge DSF er enslige, men ikke er enslige i FoB2001, større privat inntekt etter skatt enn medianinntekten etter skatt til individene som ifølge både FoB2001 og DSF er enslige.

Den summariske statistikken presentert i tabell 3.4 er ment å belyse forskjellen i inntektsulikhet ved å analysere inntektsfordelingen på kommunenivå ved å benytte husholdning og familie som enhet, målt ved avviket i Gini-koeffisient. Det er verd å merke seg at standardavviket i tabellen ikke må tolkes som et mål på utvalgsusikkerheten, men som et mål på variasjonen på Gini-koeffisienten mellom kommunene.

**Tabell 3.4: Summarisk statistikk som viser avviket mellom Gini-koeffisientene for fordelingen av inntekt etter skatt på kommunenivå basert på kvadratrotskalaen med henholdsvis familie og husholdning som økonomisk enhet. År 2000**

Gini-koeffisient med familie som enhet - Gini-koeffisient med husholdning som enhet			
Gj.snitt	St. avvik	Min	Max
0,005	0,004	-0,039	0,014

Av tabell 3.4 ser vi at forskjellen i inntektsulikhet avhengig av valg av økonomisk enhet kan være større på kommunalt nivå enn for totalpopulasjonen. Avviket for totalpopulasjonen var kun 0,006, mens avviket på kommunenivå varierte mellom -0,039 og 0,014. Dette innebærer at analyse på kommunenivå basert på informasjon fra DSF vil kunne være vesentlig forskjellig fra en analyse basert på data fra FoB2001, selv om forskjellene ikke er store for totalpopulasjonen. For de årene der vi ikke har data fra folketellinger, vil det derfor kunne være et større problem å basere analysene av økonomisk velferd på data fra DSF ved regional analyse enn ved analyse av totalpopulasjonen.

Resultatene fra dette kapitlet viser at det kan være et problem å benytte familie som økonomisk enhet. I delkapittel 3.2 viser vi at dette i stor grad skyldes at samboere uten felles barn blir behandlet som enslige (eller enslige med barn) i DSF, mens de i virkeligheten er samboere og drar nytte av stordriftsfordeler i konsumet. En sentral gruppe av samboerne uten felles barn er samboere uten barn i det hele tatt. Et sentralt formål i dette notatet er å konstruere en modell som predikerer samboerskap uten barn. Dette vil være et viktig skritt mot å lage et komplett simulert husholdningsregister (SHR) der vi bruker husholdning som økonomisk enhet.

Et viktig spørsmål vil i denne sammenheng være hva som blir konsekvensene av å betrakte samboere uten barn som enslige. Ved å betrakte samboere uten barn som enslige, vil vi ikke få et godt mål på deres økonomiske velferd. Vi vil undervurdere den økonomiske velferden til et samboerpar der partene tjener like mye, sammenliknet med en situasjon der partene ble

betraktet som enslige. Hvis den ene parten tjener tilstrekkelig mer enn den andre parten, vil det at vi bruker familie som enhet gjøre at vi overvurdere velferden til den parten som tjener mest, mens vi undervurderer velferden til den parten som tjener minst. Kan vi gjøre en a priori betraktning om effekten på inntektsulikheten i et samfunn av å betrakte samboere uten barn som enslige? Åpenbart ikke. Den første grunnen er som nevnt ovenfor at individer med tilstrekkelig større inntekt enn sin samboer vil få økt sin ekvivalentinntekt ved å bli behandlet som enslig. Den andre grunnen er at samboerparet kan være i en høyinntektsgruppe så vel som i en lavinntektsgruppe, og ulikhet er jo et relativt mål. Dette innebærer at hvis vi har et samboerpar der begge har relativt lav (og identisk) inntekt, vil vi overvurdere inntektsulikheten i samfunnet ved å bruke familie som enhet. Derimot vil effekten av å bruke familie som enhet for et samboerpar der begge har relativt høy (og identisk) inntekt, være at vi undervurderer inntektsulikheten. Som vi ser er det åpent for motstridende effekter på inntektsulikheten av å betrakte samboere uten barn som enslige.

## 4. Modell for prediksjon av samboere uten barn fra registrerte enslige i DSF

Kapittel 3 viste at det er et vesentlig avvik mellom opplysninger om familie fra DSF og opplysninger om husholdning fra FoB2001, og at dette avviket vil kunne ha betydning for resultatene i en fordelingsanalyse. I de siste 30 årene har det i Norge så vel som i en rekke andre industrialiserte land blitt mer vanlig å leve i samboerskap. En vesentlig del av avviket mellom DSF og FoB2001 skyldes nettopp samboerskap uten felles barn.

Vi skal i dette kapitlet konstruere en modell som skal brukes til å predikere samboere uten barn. Fra FoB2001 kjenner vi husholdningsstørrelsen, det totale samboerantallet og hvem som er samboere med hvem. Vi kobler på individenes inntekt og utdanning til DSF som inneholder informasjon om registeradresse, alder, bokommune, familietype og familiestørrelse. Inntektsdata og informasjon om utdanningsnivå er fra året 2000 da slike data enda ikke er tilgjengelig for året 2001. Øvrige registerdata er fra november 2001 da SSB gjennomførte FoB2001.

Vi understreker at denne modellen ikke har som formål å forklare samboerskap. Det må ikke tolkes slik at modellen gir oss opplysninger om hva som karakteriserer individers valg av samlivsform, men den kan si noe om hvilke kjennetegn som er viktige for å karakterisere personene som lever i samboerskap uten barn. Formålet med modellen er å predikere samboerskap for å lage en simulert husholdningsversjon av DSF som skal brukes som grunnlag for studier av økonomisk velferd. Om vi matcher en registrert enslig kvinne som lever i samboerskap med en mann som faktisk er enslig, trenger det ikke være et problem så lenge mannens inntekt er på et tilsvarende nivå som den virkelige samboerens inntekt.

Målet for denne modellen er å videreutvikle den metoden Åserud (2000) konstruerte for prediksjon av samboerskap uten felles barn.<sup>11</sup> Vi ønsker å utvikle en modell som predikerer hvem av de registrert enslige som er samboere uten barn ved hjelp av registerinformasjon som utdanning og alder. Vi er ikke bare interessert i å si noe om hvem av de registrerte enslige som er enslige, og hvem som er samboere, men også eventuelt samboere med hvem. Åserud brukte et utvalg fra Inntekts- og formuesundersøkelsen fra 1997 for å identifisere samboere uten felles barn, og grupperte individene i grupper ut fra utdanning og alder. På bakgrunn av informasjon fra dette utvalget konstruerte Åserud en modell som ved å utnytte informasjon om forholdet mellom personenes utdanning og alder klarer å gjenskape fordelingen vi observerer i utvalget. Dette gjelder både i forhold til antall enslige og samboere, og når det gjelder hvem som er samboer med hvem. Da vi testet denne modellen på hele populasjonen i DSF, kom det frem noen svakheter i modellens beskrivelse av samboere uten barn i registeret. Vi sammenliknet de simulerte resultatene med FoB2001, der husholdning brukes som økonomisk enhet og samboere uten barn er identifisert.

---

<sup>11</sup> Åserud (2000) benyttet en multinomisk logit-modell for å predikere samboerskap. Så vidt vi kjenner til er en slik metode ikke benyttet tidligere for å predikere samboerskap. Imidlertid har Texmon (1999), Berrington og Diamond (2000) og Ermisch og Francesconi (2000) benyttet liknende metoder for å si noe om hva som karakteriserer samboerskap mer generelt.

I registerdataene, i motsetning til i utvalgsundersøkelsen, hadde vi en ekstra betingelse på problemet vårt i form av en adressebeskranking. I registeret vil kun de enslige som bor på en adresse med minst en annen enslig av motsatt kjønn være potensielle samboere. Denne typen informasjon kan ikke benyttes når dataene stammer fra en utvalgsundersøkelse. Ved bruk av Åseruds modell på registerdata overvurderer vi antall enslige og undervurderer samboerantallet, og vi vil heller ikke få en god prediksjon av hvem som er samboere med hvem. Heller ikke ved ulike endringer av Åseruds modell klarte vi på en tilfredstillende måte å gjenskape samboernes faktiske fordeling av hvem som er samboere med hvem.<sup>12</sup> Det viste seg at modellen for prediksjon av samboerskap konstruert fra en populasjon uten adressebeskranking ikke uten videre kan brukes til å predikere samboerskap i registerdata når vi tar hensyn til adresse. Dette skyldes at Åserud hadde begrenset informasjon fra et utvalg uten adresseopplysninger. Ved å benytte resultater fra FoB2001 som gir relevant data for hele befolkningen, ønsker vi å konstruere en modell som er mer egnet til å predikere hvem som er samboere uten barn fra registrerte enslige i DSF.

## 4.1 Teori for modellen

### 4.1.1 Populasjon og enheter

Vi definerer  $U = \{1, \dots, j, \dots, N\}$  som en populasjon av  $N$  personer.  $U_h = \{1, \dots, j, \dots, N_h\}$  betegner den delpopulasjon av  $N_h$  personer som i FoB2001 er i en husholding med  $h$  personer, gitt at  $h < 18$ . Her vil  $U = \cup_{h=1}^{17} U_h$  og  $N = \sum_h N_h$ . Vi lar  $H_h$  betegne antall husholdinger i  $U_h$ . Da vil  $u_h = N_h/H_h$  for  $j \in U_h$ , hvor  $u_h = h$  for  $h < 18$ . Dette vil si at  $u_h$  er størrelsen på husholdningen til individ  $j$  hvis individet  $j \in U_h$  og husholdningsstørrelsen er mindre enn 18.

Vi lar  $\Omega_g = \{1, \dots, j, \dots, M_g\}$  betegne den delpopulasjon av  $M_g$  personer som ifølge registeret tilhører en familie av størrelsen  $g$ , gitt at  $g < 18$ . Det følger at  $U = \cup_{g=1}^{17} M_g$  og  $N = \sum_g M_g$ . Vi lar  $F_g$  være antall familier i  $\Omega_g$ . Da vil  $w_g = M_g/F_g$  for  $j \in \Omega_g$ , hvor  $w_g = g$  for  $g < 18$ . Dette vil si at  $w_g$  er størrelsen på familien til individ  $j$  hvis individ  $j \in \Omega_g$  og familiestørrelsen er mindre enn 18.

### 4.1.2 Desilgjennomsnittet av ekvivalentinntekten

Vi lar  $x_j$  stå for den private inntekten til  $j \in U_h$ . Husholdningens ekvivalentinntekt er da definert som:

$$y_j = \sum_{j \in c} x_j / \sqrt{u_h} \text{ for } j \in U_h.$$

Her betegner  $c$  husholdningen som  $j$  tilhører. Dette innebærer at  $y_j$  er identisk for alle medlemmene av samme husholdning og  $\sum_{j \in c} y_j = u_h y_j = \sum_{j \in c} x_j \sqrt{u_h}$ .

Vi lar  $U_{(k)} = \{1, \dots, j, \dots, N/10\}$  være det  $k$ te desilet til  $U$  i stigende rekkefølge, slik at hvis  $i \in U_{(k)}$ ,  $j \in U_{(l)}$  og  $k < l$  vil  $y_i < y_j$ . Vi lar  $I_j(h, k) = 1$  hvis  $j \in U_h$  og  $j \in U_{(k)}$ , med andre ord hvis

<sup>12</sup> Ved endringer av Åseruds modell benyttes Sundvoll, A. og H. M. Teigum (1998), Dale, T. (1997) og Øyangen, I. (1999).

individ  $j$  i FoB2001 har husholdingsstørrelse  $h$  og tilhører desil  $k$ . Ellers er  $I_j(h,k)=0$ . Det  $k$ te desilgjennomsnittet av ekvivalentinntekten til husholdningen er definert ved:

$$\bar{y}_k = (10/N) \sum_{j \in U^{(k)}} y_j = (10/N) \sum_{j \in U^{(k)}} x_j \sqrt{n_j} = (10/N) \sum_{h=1}^{17} u_h^{1/2} X_{hk}, \text{ hvor } X_{hk} = \sum_{j \in U} I_j(h,k) x_j.$$

Familiens ekvivalentinntekt etter skatt er definert ved:

$$z_j = \sum_{j \in b} x_j / \sqrt{w_g}, \text{ for } j \in \Omega_g.$$

Her betegner  $b$  familien som  $j$  tilhører. Dette innebærer at  $z_j$  er identisk for alle medlemmene av samme familie og  $\sum_{j \in b} z_j = w_g z_j = \sum_{j \in b} x_j \sqrt{w_g}$ .

Vi lar  $\Omega_{(l)} = \{1, \dots, j, \dots, N/10\}$  være det  $l$ te desilet til  $U$  i stigende rekkefølge. Dette innebærer at hvis  $i \in \Omega_{(k)}$ ,  $j \in \Omega_{(l)}$  og  $k < l$  vil  $z_i < z_j$ . Vi lar  $I_j(g,l)=1$  hvis  $j \in \Omega_{(l)}$  og  $j \in \Omega_{(g)}$ , altså hvis individet i registeret tilhører en familie med familiestørrelse  $g$  og desil  $l$ . Ellers er  $I_j(g,l)=0$ . Verdiene til  $I_j(g,l)$  er kjent for hele populasjonen fra DSF koblet med inntektsdata.

#### 4.1.3 Registerbasert metode

Anta at vi har konstruert en fullstendig simulert husholdningsversjon (SHR) av DSF. Vi lar  $\hat{U}_h = \{1, \dots, j, \dots, \hat{N}_h\}$  betegne en populasjon av  $\hat{N}_h$  personer fra konstruerte hushold med  $h$  personer, gitt at  $h < 18$ . Da vil  $\hat{N} = \sum_h \hat{N}_h$ . For eksempel vil  $\hat{U}_3 = \{1, \dots, j, \dots, \hat{N}_3\}$  betegne populasjonen av  $\hat{N}_3$  personer som har konstruert husholdningsstørrelse lik 3. Vi lar  $\hat{H}_h$  betegne antall husholdninger i  $\hat{U}_h$  og  $\hat{u}_h = \hat{N}_h / \hat{H}_h$  for  $j \in \hat{U}_h$ , der  $\hat{u}_h = h$  for  $h < 18$ .

Den konstruerte husholdningens ekvivalentinntekt er da definert som:

$$\hat{y}_j = \sum_{j \in \hat{c}} x_j / \sqrt{\hat{u}_h} \text{ for } j \in \hat{U}_h, \text{ hvor } \hat{c} \text{ er den konstruerte husholdningen til individ } j.$$

Her vil  $\hat{U}_{(k)}$  stå for det  $k$ te desilet til  $U$  i stigende rekkefølge for  $\hat{y}_j$ . Vi lar  $\hat{I}_j(h,k)=1$  hvis  $j \in \hat{U}_h$  og  $j \in \hat{U}_{(k)}$ , og  $\hat{I}_j(h,k)=0$  ellers. Et estimat på det  $k$ te desil gjennomsnittet  $\bar{y}_k$  er nå gitt ved  $\hat{\bar{y}}_k$  som er gjennomsnittet til  $\{\hat{y}_j; j \in \hat{U}_{(k)}\}$  og er definert tilsvarende som  $\bar{y}_k$  tidligere.

#### 4.1.4 Samboerskap blant registrerte enslige

Vårt mål for modellen er å simulere en husholdningsversjon av DSF som vi kaller SHR. Et slikt register vil ikke lenger ha registerets familiebegrep som økonomisk enhet, men et familiebegrep som nærmer seg husholdningsbegrepet til FoB2001. Vi tar utgangspunkt i DSF med familie som økonomisk enhet. Her vil  $\hat{I}_j(h)=1$  hvis individets husholdningssammensetning er identisk med individets familiesammensetning, og  $\hat{I}_j(h)=0$  ellers. Den estimerte gjennomsnittlige ekvivalentinntekten for det  $l$ te desil er gitt ved:

$$\hat{\bar{y}}_l = \bar{z}_l = (10/N) \sum_{g=1}^{17} w_g^{1/2} X_{gl}, \text{ hvor } X_{gl} = \sum_{j \in U} I_j(g,l) x_j.$$

Mange personer som er registrerte enslige, er i virkeligheten samboere uten felles barn og har husholdningsstørrelse lik 2. Vi lar derfor SHR være identisk med DSF og bruker familie som

økonomisk enhet i tilfeller der  $j \in \Omega_g$  og  $g > 1$ . For hver  $j \in \Omega_1$  setter vi enten  $\hat{I}_j(h) = 1$  for  $h=1$  og  $j$  forblir enslig også i SHR. Eller så matcher vi  $j$  med en annen registrert enslig  $i \in \Omega_1$ , og setter  $\hat{I}_j(h) = \hat{I}_i(h) = 1$  for  $h=2$ , og individ  $i$  og  $j$  er samboere uten barn i SHR. Vi ser altså bort i fra samboerskap med særkullsbarn. Vår modell har som formål å predikere samboerskap blant registrerte enslige av motsatt kjønn i DSF.

#### 4.1.5 Antall matcher på hver adresse

Det er  $a = \{1, \dots, A\}$  adresser både i DSF og FoB2001. Vi lar  $\Omega_{1a}$  betegne et sett av registrerte enslige på adressen  $a$  og  $M_{1a}$  betegne antall personer i  $\Omega_{1a}$ . Vi lar  $M_{1a,0}$  være antall kvinner i  $\Omega_{1a}$  og  $M_{1a,1}$  være antall menn i  $\Omega_{1a}$ . Matchpotensialet på en adresse vil være det maksimale antall mulige matcher på hver adresse. På adressen  $a$  er matchpotensialet gitt ved  $\alpha_a = \min(M_{1a,0}, M_{1a,1})$ . Hvis det for eksempel er tre menn og en kvinne som er registrert enslige ved en adresse, vil det maksimale antall mulige matcher av samboerpar (der begge er registrert enslige og av motsatt kjønn) være en. Matchpotensialet er da en.

Vår metode forutsetter at det totale matchpotensialet er større enn eller lik antall faktiske samboerpar uten barn som er registrert enslige. Denne betingelsen vil nødvendigvis være oppfylt da de faktiske samboerne uten barn også vil være potensielle samboere uten barn.

Vi allokere antall matcher proporsjonalt med matchpotensialet til adressen og får antall samboere vi skal matche på adressen  $a$  gitt ved  $\eta_a = (\alpha_a / \alpha) \eta$ . Vi lar  $\eta$  betegne summen av antall samboere uten barn fra FoB2001 som er registrert som enslige i DSF og  $\alpha$  betegne det totale matchpotensialet i populasjonen. Med andre ord antall matcher vi skal trekke ut på en adresse, er lik matchpotensialet på adressen multiplisert med totalt antall samboere uten barn fra FoB2001 som er registrert som enslige i DSF, dividert på det totale matchpotensialet i populasjonen. Vår allokering av antall matcher kan raffineres ved å for eksempel skille mellom type adresse, alder på de registrerte enslige på adressen etc. Dette vil kunne forbedre modellen. Vi kan tenke oss en adresse med mange registrerte enslige av begge kjønn der alle personene er eldre. Her vil modellen vår tilsi at mange samboere skal matches, selv om vi vet at eldre ikke er typiske samboere uten barn. Modellen slik den er formalisert ovenfor, tar ikke innover seg spesielle kjennetegn ved de ulike adressene ved allokering av antall matcher på hver adresse.

Et problem vil være at  $\eta_a$  ofte ikke vil være et heltall. Vi løser dette ved å bruke en metode kalt "bracketing".  $\gamma_a$  defineres som det heltall som tilfredstiller at  $\gamma_a \geq \eta_a > \gamma_a - 1$ . For eksempel: hvis  $\eta_a = 1,2$  vil  $\gamma_a$  være 2. Vi lar  $\tilde{n}_a$  være antall matcher som faktisk skal gjennomføres på adressen  $a$ . Vi velger da  $\tilde{n}_a = \gamma_a - 1$  med sannsynlighet  $\gamma_a - \eta_a$  og  $\tilde{n}_a = \gamma_a$  med sannsynlighet  $1 + \eta_a - \gamma_a$ . Forventingen til  $\tilde{n}_a$  er  $\eta_a$  ved bracketing. Ved uavhengig bracketing på hver adresse vil store talls lov implisere at  $\sum_a \tilde{n}_a / A$  konvergerer mot  $\eta / A$ . Dette sikrer oss at antall matcher som skal allokere, blir tilnærmet likt antall faktisk samboere.

#### 4.1.6 Fordelingen til matching klassene

Vi lar  $(i, j)$  betegne at individ  $i$  og  $j$  er et samboerpar, hvor  $i, j \in \Omega_{1a}$  på en adresse  $a$ , med andre ord individene  $i$  og  $j$  er registrerte enslige av ulikt kjønn på adressen  $a$ . Vi klassifiserer hvert par  $(i, j)$  i matching klasser som vi betegner ved  $d_{ij} = d$  for  $d = 1, \dots, D$ . Matching klassene vil her bli definert på bakgrunn av informasjon om individenes alder, utdanning og kjønn. Inndelingen av parene i matching klasser innebærer at vi kategoriserer parene i ulike



samboergrupper ut fra eksogene kjennetegn. Videre i dette notatet vil begrepet matching klasse være ekvivalent med samboergruppe.

Vi lar  $\delta = \{(i, j); i, j \in \Omega_1 \text{ og } i, j \in U_2\}$  være de registrerte enslige som ifølge FoB2001 lever i samboerskap og  $H_\delta$  være antall par i  $\delta$ . Dersom  $d_{ij} = d$  for  $(i, j \in \delta)$ , skal  $I_{ij}(d)=1$ , og ellers skal  $I_{ij}(d)=0$ .

På bakgrunn av FoB2001 finner vi  $f_d = H_\delta^{-1} \sum_{(i,j) \in \delta} I_{ij}(d)$ .  $f_d$  er sjansen for å tilhøre klasse  $d$  for et samboerpar uten barn, og vi finner det ved å summere de faktisk samboerparene uten barn som er registrerte enslige i klasse  $d$  dividert med totalt antall par som er registrert enslige, men lever i samboerskap uavhengig av klasse. Vi lar  $\mathbf{f} = (f_1, \dots, f_D)$  betegne sannsynlighetsvektoren til samboergruppene. For eksempel vil et tilfeldig par  $(i, j)$  ha sannsynlighet for å være i gruppe 1 lik  $f_1 = \sum I_{ij}(1) / H_\delta$ .

#### 4.1.7 Stokastisk matching

Vi vil nedenfor gjøre rede for hvordan vi skal benytte stokastisk matching for å oppnå et enkelt SHR-register. Multiple SHR-registre kan vi konstruere ved å gjennomføre uavhengige repetisjoner av den samme prosedyren.

Vi lar  $D_a = \{d_{ij}; i \in M_{1a,0} \text{ og } j \in M_{1a,1}\}$  inneholde alle mulige matching klasser representert på adressen  $a$ . Da vil  $f_{d;a} = (f_d / \sum_{d \in D_a} f_d)$  for  $d \in D_a$ , være den standardiserte fordelingen til matching klassen på adressen  $a$ . Vi standardiserer sannsynlighetene etter hvilke klasser som er representert på adressen.

Vi lar  $\hat{\delta}_a$  inneholde de  $\tilde{n}_a$  faktiske matchene av samboere på adresse  $a$ . Vi generer  $\hat{\delta}_a$  ved følgende repetisjoner. Initialt setter vi  $M_{1a}^{(1)} = M_{1a}, D_a^{(1)} = D_a$  og  $f_{d;a}^{(1)} = f_{d;a}$ . For  $k=1, \dots, \tilde{n}_a$ :

1. trekker vi  $d$  fra  $D_a^{(k)}$  etter sannsynlighetsfordelingen  $f_{d;a}^{(k)}$ , med andre ord vi trekker klasse etter den standardiserte sannsynlighetsfordelingen.
2. legger vi  $(i_k, j_k)$  til  $\hat{\delta}_a$ , trukket tilfeldig fra  $\{(i, j); i \in M_{1a,0}^{(k)}, j \in M_{1a,1}^{(k)} \text{ og } d_{ij}=d\}$ . Dette vil si at vi trekker ut ett par fra klassen vi har trukket ut i trinn 1.
3. setter vi  $M_{1a}^{(k+1)} = M_{1a}^{(k)} - \{(i_k, j_k)\}$  og får  $D_a^{(k+1)}$  og  $f_{d;a}^{(k+1)}$  fra  $D_a^{(k)}$ . Med andre ord etter at et par  $(i_k, j_k)$  er trukket, standardiseres sannsynlighetene  $f_{d;a}$  på nytt.

Denne prosessen gjentas helt til alle par vi skal trekke på adressen er trukket. Deretter neste adresse.

Et eksempel kan være til hjelp for å forstå metoden. På adresse  $a$  er det 3 registrert enslige menn (A, B og C) og 3 registrert enslige kvinner (M, L og K) som er 17 år eller mer. Det er altså 9 potensielle samboerpar som alle tilhører forskjellige samboergrupper og maksimalt kan det predikeres 3 samboerpar på adressen, altså matchpotensialet på 3. Anta at det med vår modell skal trekkes to samboerpar på adresse  $a$ . Den standardiserte sannsynligheten for at samboerparet  $\{A, L\}$  som tilhører gruppe 4, skal trekkes ut som det første samboerparet, er sannsynligheten for at et tilfeldig samboerpar skal tilhøre gruppe 4 dividert med summen av sannsynlighetene til de potensielle samboerparene på adressen. Vi får altså en relativ sannsynlighet for at par  $\{A, L\}$  skal trekkes ut som det første par. Vi antar at  $\{A, L\}$  trekkes ut. Da har vi fått trukket ut ett av de to samboerparene som modellen sier skal trekkes ut på denne adressen. Vi må nå fjerne alle potensielle samboerpar der enten A eller L (eller begge) inngår. Det vil si samboerparene:  $\{A, M\}$ ,  $\{A, L\}$ ,  $\{A, K\}$ ,  $\{B, L\}$  og  $\{C, L\}$ . Deretter standardiserer vi sannsynlighetene igjen og trekker et samboerpar til ut fra den nye sannsynlighetsfordelingen. Da har vi

predikert like mange samboerpar som vi ønsket å predikere, og vi kan gjør samme prosedyre på neste adresse. Hvis vi på denne adresse ikke skal trekke noen samboerpar, forblir våre potensielle samboere enslige, og vi fortsetter på neste adresse.

#### 4.2 Resultater ved bruk av modellen for prediksjon av samboere uten barn

For å identifisere potensielle samboere blant de registrerte enslige er vi nødt til kun å ta utgangspunkt i informasjon fra DSF og ikke bruke vår husholdningsinformasjon fra FoB2001. Vi plukker ut alle menn og kvinner som er registrert som enslige med alder høyere eller lik 17 år. Disse vil i utgangspunktet være potensielle samboere uten barn. Vår modell skal predikere samboerskap av ulikt kjønn uten barn der begge er registrert som enslige. Dette krever at de potensielle samboerne ikke bare selv er registrert enslige, men også lever på en adresse med minst en annen registrert enslig av motsatt kjønn. Matchpotensialet på adressene er da til sammen 202568 mulige par, det vil si 405136 individer.

På bakgrunn av matchpotensialet skal vi allokere antall matcher på hver adresse. Vi må da identifisere personene som er samboere uten barn der begge er registrert som enslige. Av 4394931 individer totalt er det 195696 individer som ifølge FoB2001 er samboere uten felles barn som er registrerte enslige, og hvor vi har tilstrekkelig informasjon om deres samboer, det vil si  $\eta = 195696$ . Tabell 4.1 viser fordelingen over husholdningsstørrelsen til disse personene. Som vi ser, er den klart største gruppen nettopp samboere uten barn med husholdningsstørrelse lik to.

**Tabell 4.1: Fordelingen av samboere uten felles barn fra FoB2001 etter husholdningsstørrelse**

Husholdningsstørrelse	Antall	Prosent
2	140100	71,591
3	30994	15,838
4	17047	8,711
5	5410	2,765
6	1532	0,783
7	380	0,194
8	147	0,075
9	42	0,021
10	22	0,011
11	8	0,004
12	4	0,002
13	0	0
14	4	0,002
15	6	0,003
Alle	195696	100

Av de 140100 som har husholdningsstørrelse lik 2 er det 137382 som lever i et samboerskap med en annen person som er registrert enslig. Avviket kan skyldes at registeropplysningene ikke er tilstrekkelig oppdatert, og for eksempel at personen nylig er separert, og nå lever i et samboerskap med en ny partner. Antall samboere som skal allokeres ved hjelp av vår modell,

er altså 137382 individer. Dette vil si 68691 samboerpar med ulikt kjønn der begge er registrert enslig og er 17 år eller mer.

Det neste vi gjør, er å klassifisere både de faktiske samboerparene og de potensielle samboerparene ut fra deres alder og utdanning. Det er rimelig å anta at et viktig kjennetegn ved samboerskap er forholdet mellom samboernes alder og utdanning. Dette er i tråd med resultatene fra Åserud (2000). Forsøk i hans arbeid viste at alder og utdanningsnivå var de variablene, som var tilgjengelige i datagrunnlaget, som var best egnet til å identifisere samboerskap. Tanken er at en ung kvinne med høy utdanning vil ha større sannsynlighet for å være samboer med en ung mann med høy utdanning enn med en eldre mann med lav utdanning. Ut fra vår gruppering av de faktiske samboerparene vil vi få sannsynligheter for at et tilfeldig samboerpar skal tilhøre en viss gruppe. At et individ er i en klasse, vil si at han tilhører en gruppe definert ut fra bestemte kriterier. Vi har 15 ulike klasser for menn og kvinner. Tabell 4,2a viser gruppeinndelingen for kvinner etter alder og utdanningsnivå.

**Tabell 4.2a: Gruppeinndeling for kvinner etter utdanningsnivå og alder**

Kvinnens utdanningsnivå	Kvinnens alder				
	17-24	25-33	34-45	45-60	61 og eldre
Kun grunnskole, ingen eller uoppgitt utdanning	1	4	7	10	13
Videregående skole	2	5	8	11	14
Universitets- eller høyskolenivå	3	6	9	12	15

Tabell 4,2b som viser gruppeinndelingen for kvinner etter alder og utdanningsnivå. En sammenlikning av tabell 4,2a med tabell 4,2b viser at mannens aldersdeling er definert forskjellig fra kvinnens. Dette skyldes at menn ofte er noe eldre før de går inn i et samliv.

**Tabell 4.2b: Gruppeinndeling for menn etter utdanningsnivå og alder**

Kvinnens utdanningsnivå	Mannens alder				
	17-26	27-35	36-45	45-60	61 og eldre
Kun grunnskole, ingen eller uoppgitt utdanning	1	4	7	10	13
Videregående skole	2	5	8	11	14
Universitets- eller høyskolenivå	3	6	9	12	15

Vi har altså 15 grupper for kvinner og 15 grupper for menn etter alder og utdanning. Dette gir 225 ulike kombinasjoner av samboergrupper der partene har motsatt kjønn. Den enkelte samboergruppe blir bestemt slik:

$$\text{Samboergruppe} = 15 (\text{Kvinnegruppe} - 1) + \text{Mannsgruppe}$$

Et samboerpar hvor for eksempel kvinnen tilhører gruppe 11 og mannen tilhører gruppe 3 vil tilhøre samboergruppe 153.

Noen eksempler vi kanskje gjøre metoden klarere. Samboergruppe 1 består av en mann fra gruppe 1 og en kvinne fra gruppe 1. Kvinnens gruppenummer øker med 1 for hver femtende samboergruppe. Dette innebærer at samboergruppe 1 til 15 består av en kvinne fra gruppe 1, samboergruppe 16 til 30 består av kvinne fra gruppe 2, mens samboergruppe 31 til 45 består av kvinne fra gruppe 3, osv. Slik fortsetter prosessen til og med samboergruppe 225 hvor kvinnens gruppenummer er 15. Mannens gruppenummer øker med 1 for hver av de første 15 samboergruppene. For den sekstende samboergruppen vil mannens gruppenummer bli satt lik 1 igjen.

Deretter gjentar prosessen seg for de neste 15 samboergruppene og for den 31 samboergruppen blir mannens gruppe satt lik 1 igjen. Slik fortsetter prosessen til og med samboergruppe 225.

Ikke alle samboergruppene er representert blant de faktiske samboerparene, og sannsynligheten for at et potensielt samboerpar skal tilhøre deres klasse er dermed lik null. For at et samboerpar skal være potensielle samboere, må sannsynligheten for at de skal være et samboerpar være større enn null. Derfor fjernes de potensielle samboerne som er i samboergrupper som ikke er representert blant de faktiske samboerparene. Et eksempel på en slik klasse er gruppe 15. Dette er ikke uventet, da det intuitivt kan virke usannsynlig at en ung kvinne med lav utdanning er samboer med en eldre mann med høy utdanning.

Tabell 4.3 presentert i vedlegget, som viser fordelingen av de faktiske samboerne uten barn fra FoB2001 over samboergruppene, er med på å forsterke inntrykket vårt av at individers valg av samliv ofte innebærer valg av partner med relativ lik alder og utdanning. Tabellen viser også at grupper som for eksempel 17, 65 og 81 er grupper der det er relativt mange samboere uten barn der begge er registrert enslige i DSF. Dette er i tråd med våre forventninger. Det virker rimelig at det er relativt mange samboerpar i samboergruppene 17, 65 og 81 da disse samboergruppene innebærer at mannen og kvinnen begge tilhører henholdsvis gruppe 2, 5 og 6.

Det er 200385 potensielle samboere når vi ser bort i fra de potensielle samboerpar som er i samboergrupper der det ikke er faktiske samboere ifølge FoB2001. På bakgrunn av matchpotensialet på hver adresse allokterer vi 68622 samboerpar på adresser med potensielle samboerpar. Ved å sammenlikne antall faktiske samboere med antall samboere som skal allokteres, ser vi at det er 69 flere faktiske samboere. Store talls lov skulle sikre oss at det matchede antallet samboer konvergerer mot det faktiske antallet samboere. Avviket er marginalt og skyldes at den uavhengige bracketingen ikke er gjort på et stort nok antall adresser.

Vi benytter den stokastiske matching prosessen til å trekke samboerpar helt til alle par vi skal predikere på de ulike adressene er trukket ut. Ved å koble disse samboerparene med data fra DSF får vi et enkelt simulert husholdningsregister. Tabell 4.4 i vedlegget viser resultatene av prediksjonen av samboere uten barn fra de registrerte enslige i DSF.

Ved å sammenlikne resultatene presentert i tabellene 4.3 og 4.4, begge i vedlegget, får vi et inntrykk av hvor egnet modellen er til å predikere samboere uten barn der begge er registrerte enslige, og er av motsatt kjønn. En grafisk fremstilling er gitt i figur 4.1 i vedlegget. Figur 4.1 er et histogram over to sider som er ment å illustrere modellens fordeling av hvem som er samboere med hvem er. Vi ønsker å belyse forskjellen mellom den faktiske fordelingen av samboere over de 225 gruppene og den predikerte fordelingen. Det faktiske antall samboere i de enkelte gruppene er gitt av den mørke stolpen, mens det predikerte antall samboere er gitt av den lyse stolpen.

Modellens struktur sikrer oss at det aggregerte antall predikerte samboere er tilnærmet lik det faktisk antall samboere fra FoB2001. Spørsmålet er om modellen på en tilfredstillende måte klarer å gjenskape fordelingen av hvem som er samboere med hvem. Vi ser av tabell 4.3, tabell 4.4 og figur 4.1 at modellens prediksjoner er gode i den forstand at det grovt sett blir trukket et høyt antall predikerte samboere i de gruppene som ifølge FoB2001 har et høyt antall faktisk samboere, og omvendt. Prediksjonen av samboere er beheftet med usikkerhet, og det er ikke slik at vi forventer at modellen skal gi oss det faktiske antall samboere i hver av

de 225 gruppene. Det som er viktig er at modellen gir oss et godt helhetsbilde av fordelingen av samboerskap uten barn blant registrerte enslige. Modellen viser en svakhet ved prediksjon av samboerskap i intervallet fra gruppe 184-210, som inneholder eldre kvinner med lav og middels utdanning. Her predikerer vi gjennomgående alt for mange samboere. Hva skyldes dette?

Vi vil undersøke dette ved å se nærmere på fordelingen over grupper blant de potensielle samboerne. Det totale matchpotensialet ( $\alpha$ ) er 200385. Dette vil være det maksimale antall personer som kan leve i et samboerskap uten barn under betingelsene av at både de og deres partner er over 17 år, registrert enslig, boende på samme adresse og av motsatt kjønn. Vi får det ved å summere minimum(kvinner, menn) på hver adresse. Til sammen er det 511266 individer som er potensielle samboere på de aktuelle adressene, hvorav 248455 er menn og 262811 er kvinner. Disse er delt i grupper etter kjønn, alder og utdanning som angitt i tabell 4.2. Fordelingen av de potensielle samboerne over gruppene vil kunne være med på å forklare observerte forskjeller mellom tabell 4.3 og tabell 4.4.

Ved å allokere antall matcher på adresser på bakgrunn av matchpotensialet alene får vi et problem. Vi ser i tabell 4.5 at antallet potensielle samboerne som er kvinner med lav og middels utdanning og er i den eldste aldersgruppen, er betydelig. Det er rimelig at mange av de registrerte enslige vil være eldre kvinner med lav og middels utdanning. Kvinner lever ofte lenger enn menn og det vil dermed være relativt mange enslige kvinner i den eldste aldersgruppen. At det er vesentlig færre eldre kvinner med høy utdanning som er potensielle samboere, er heller ikke overraskende. Høyere utdanning var ifølge tall fra FoB2001 uvanlig blant kvinner i denne aldersgruppen.<sup>13</sup>

Hvis det er relativt mange potensielle enslige i noen grupper, vil de også være hyppige representert som potensielle samboere på de ulike adressene. En adresse med stort matchpotensialet og individer kun fra gruppe 13 og 14, får allokert like mange matcher som en adresse med like stort matchpotensial, men med individer fra grupper som har større sannsynlighet for å være samboere. Selv om potensielle samboere med individer fra gruppe 13 og 14 initialt har lav sannsynlighet, vil samboere fra disse gruppene ofte blir trukket ut. Dette skyldes at det er sannsynlighetsfordelingen på den enkelte adressen og ikke sannsynlighetene til hver samboergruppe som er bestemmende for matchingen på adressen. En adresse med mange potensielle samboerpar vil få allokert en god del matcher uavhengig av om alle de potensielle samboerparene består av individer fra grupper som ikke ofte er faktiske samboere. Da vil vi også nødvendigvis få predikert en rekke samboere fra for eksempel grupper i intervallet 184-210 som består av kvinner fra gruppe 13 og 14. Det er nettopp dette vi observerte da vi sammenliknet tabell 4.3 og tabell 4.4. Dette skyldes at modellens allokering av matcher på adresser ikke tar innover seg de potensielle samboernes fordeling over de ulike gruppene. Ved å raffinere modellens allokering av matcher vil vi kunne forbedre modellens prediksjoner ytterligere.

---

<sup>13</sup> Informasjon hentet fra "Utdanningsnivå i befolkningen" på internetadresse: [www.ssb.no/utniv/](http://www.ssb.no/utniv/).

**Tabell 4.5: Fordelingen av de potensielle kvinnelige og mannlige samboerne fra DSF over 15 ulike grupper karakterisert etter alder og utdanning**

<b>Kvinner</b>		
Gruppe	Antall	Prosent
1	3110	1,18
2	22928	8,72
3	3431	1,31
4	4008	1,53
5	24487	9,32
6	31331	11,92
7	4638	1,76
8	15873	6,04
9	11991	4,56
10	13958	5,31
11	25230	9,60
12	10489	3,99
13	52009	19,79
14	32415	12,33
15	6913	2,63
Alle	262811	100
<b>Menn</b>		
Gruppe	Antall	Prosent
1	4643	1,87
2	26949	10,85
3	5904	2,38
4	8368	3,37
5	37649	15,15
6	29485	11,87
7	8058	3,24
8	23560	9,48
9	12315	4,96
10	14158	5,70
11	26144	10,52
12	12409	4,99
13	17909	7,21
14	15430	6,21
15	5474	2,20
Alle	248455	100

### 4.3 Endring i den stokastiske matching prosessen

Et alternativ til å forbedre modellens allokering av matcher på adresser er å endre den stokastiske matching prosessen som ga overrepresentering av samboere i grupper med mange registrerte enslige. Dette vil nok være betydelig enklere og trolig gi oss en bedre prediksjon av samboere uten barn enn en raffinering av allokeringemetoden i modellen.

Vi lar fortsatt  $(i, j)$  betegne at to individer,  $i$  og  $j$  er et samboerpar, hvor  $i, j \in \Omega_{1a}$  på en adresse  $a$ . Dette vil si at individene  $i$  og  $j$  er registrerte enslige av ulikt kjønn på adressen  $a$ . Vi klassifiserer hvert par  $(i, j)$  i matching klasser som vi betegner ved  $d_{ij} = d$  for  $d = 1, \dots, D$ . Matching klassene vil helt tilsvarende som tidligere bli definert ved å utnytte informasjon om individenes alder, utdanning og kjønn. Vi lar  $\delta = \{(i, j); i, j \in \Omega_1 \text{ og } i, j \in U_2\}$  være de registrerte enslige som ifølge FoB2001 lever i samboerskap, og  $H_\delta$  være antall par i  $\delta$ . Dersom  $d_{ij} = d$  for  $(i, j \in \delta)$ , skal  $I_{ij}(d) = 1$ , og ellers skal  $I_{ij}(d) = 0$ .  $\sum_{(i,j) \in \delta} I_{ij}(d)$  vil være å forstå som summen av antall par registrerte enslige, men faktisk samboende uten barn i klasse  $d$ .

I denne andre versjonen av modellen er det fortsatt matchpotensialet ( $\alpha$ ) lik 200385 og 68691 faktiske samboerpar. Vi vet matching klassene både til de potensielle og til de faktiske samboerne. Tidligere allokerte vi antall matcher til hver adresse på bakgrunn av adressens matchpotensialet. Deretter trakk vi samboerpar uten barn på hver adresse ut fra sannsynlighetsfordelingen på den enkelte adressen. Antall samboere på en adresse ble da bestemt ut fra matchpotensialet på adressen, mens gruppefordelingen til de predikerte samboerne er bestemt på bakgrunn av sannsynlighetsfordelingen til samboergruppene representert på adressen og ikke samboergruppens initiale sannsynlighetsfordeling. Til forskjell fra denne metoden vil vi nedenfor utvikle en metode hvor vi fokuserer på den enkelte samboergruppe og ikke den enkelte adresse når vi predikerer samboerskap uten barn. Ved å benytte denne andre metoden trekker vi samboere uten barn i den enkelte gruppe tilfeldig fra de potensielle samboerne i denne gruppen, uavhengig av matchpotensialet på de potensielle samboernes adresse. Dermed unngår vi en overrepresentering av samboere i grupper med mange registrerte enslige. Metoden vi vil bruke nå består av to trekninger. Trekning 1 er slik:

Vi lar  $N_d = (f_d \times \eta)$  betegne antall samboere vi ønsker å predikere i gruppe  $d$ , hvor  $N_d$  vil være lik antall faktiske samboere i gruppe  $d$ . Initialt setter vi  $\Omega_{1a}^{(d,k)} = \Omega_{1a}$  for adresse  $a = 1, \dots, A$ . For klassene  $d = 1, \dots, D$  og det predikerte samboerantallet  $k = 1, \dots, N_d$ :

- Lar vi  $\Omega_{1a,1}^{(d,k)}$  betegne de mennene som er potensielle samboere tilgjengelige på adresse  $a$ , mens  $\Omega_{1a,0}^{(d,k)}$  betegner de kvinnene som er potensielle samboere tilgjengelig på adresse  $a$ . Vi lar  $\beta_{d,k;a} = \{d_{ij} = d; i \in \Omega_{1a,0}^{(d,k)} \text{ og } j \in \Omega_{1a,1}^{(d,k)}\}$  betegne alle potensielle par av matching klasse  $d$  på adresse  $a$ .
- Trekker vi tilfeldig et par  $(i, j)$  på adresse  $a$  fra alle de  $\sum_a \beta_{d,k;a}$  parene tilgjengelige.
- Legger vi  $(i, j)$  til  $\hat{\delta}$  og fjerner  $i$  fra  $\Omega_{1a,0}^{(d,k)}$  og  $j$  fra  $\Omega_{1a,1}^{(d,k)}$ .

Ved å bruke denne metoden forsøker vi å sikre at det predikeres det ønskede antall samboere i både hver gruppe og totalt. Det kan bli et problem at vi ved trekking av samboere i de siste klassene ikke har nok potensielle samboere. Dette vil eventuelt være fordi vi har "brukt opp" for mange av gruppenes initiale potensielle samboere i tidligere trekninger. Vi vil derfor begynne å predikere samboere i den klassen med lavest antall potensielle samboere. Hvis vi likevel får situasjoner der vi ønsker å trekke et samboerpar i en gruppe der det ikke er flere

potensielle samboere, så gjennomfører vi en ny trekning; trekning 2. I trekning 2 vil vi trekke et tilfeldig samboerpar fra de potensielle samboerparene i kommunen som ikke er trukket ut i trekning 1. Da vil vi predikere samboerpar med feil gruppetilhørighet, men det totale antall predikerte samboere i kommunen blir riktig.

Et eksempel vil kanskje gjøre metoden klarere. Vi starter den stokastiske matchingen med klasse X som er den klassen med færrest potensielle samboere i kommune Y. Her er det 2 faktiske samboerpar uten barn ifølge FoB2001 og seks potensielle samboerpar ifølge FoB2001. Vi trekker tilfeldig ett samboerpar fra denne klassen. Det viser seg at dette paret tilhører adresse a. På adresse a fjerner vi alle par der minst ett av individene tilhører det uttrukne paret. Deretter trekker vi et nytt tilfeldig samboerpar fra de resterende potensielle samboerne. Denne prosessen bruker vi til å trekke de 2 samboerparene vi ønsket å predikere. Slik gjør vi for alle grupper i alle kommuner.

Et alternativ ville vært å brukt denne metoden på landsnivå. Hvis vi unngår å trekke samboerpar med trekning 2, vil vi sikre oss at antall predikerte samboere i hver klasse, og dermed også totalt, stemmer overens med tallene fra FoB2001. Men vi ville ikke få predikert verken et riktig antall samboere totalt eller i hver gruppe for hver enkelt kommune. Da denne modellen senere bør være anvendbar til regional analyse, er det viktig at metoden også er god til å predikere samboerskap på kommunenivå. Derfor vil vi gjennomføre den nye matching prosessen drøftet ovenfor på kommunalt nivå, og etterpå kan vi aggregere resultatene til landsnivå. Da vil vi potensielt kunne få riktig antall samboere både totalt og i de enkelte gruppene i hver kommune, så vel som på landsnivå.

#### **4.4 Resultater ved bruk av modellen for prediksjon av samboere uten barn etter endring i den stokastiske matching prosessen**

Vi innførte ekstra restriksjoner i den stokastiske matching prosessen fordi vi ønsket at modellen skulle predikere samboerskap bedre både på landsnivå og på kommunalt nivå. Nedenfor vil vi først se på landsnivå, før vi fokuserer på kommunenivå. Dette kan virke unødvendig da det er slik at hvis prediksjonene stemmer på kommunenivå, vil de nødvendigvis også stemme på landsnivå. Men ved å se på landsnivå får vi en grei oversikt over hvilke grupper det eventuelt er feil i modellens prediksjoner. I delkapittelet der fokus er på kommunenivå, ønsker vi å vise i hvilke kommuner modellen eventuelt predikerer feil.

##### **4.4.1 Landsnivå**

Tabell 4.6 som er presentert i vedlegget viser resultatene fra modellens prediksjoner av samboere på landsnivå ved å aggregere resultatene for den enkelte kommune.

Vi ser i kolonne 2 i tabell 4.6 at modellens prediksjoner etter trekning 1 er radikalt forbedret. For kun 299 par oppstår problemet med at vi ønsker å trekke et samboerpar i en gruppe der det ikke er flere potensielle samboere. Dette innebærer at for kun 299 av 68691 samboerpar vil modellens predikerte samboergruppe ikke samsvare med den faktiske samboergruppen etter trekning 1. Dette vil si at modellen ved trekning 1 klarer å gjenskape 99,6 prosent av de faktiske samboernes gruppefordeling på landsnivå. Den fjerde kolonnen gir oss avviket fra den faktiske gruppefordelingen til samboerne gitt i tabell 4.3, etter trekning 2. Ved trekning 2 vil vi, akkurat som før vi endret matching prosessen, gjennomgående predikere for mange samboere i grupper der det er mange potensielle enslige. Et eksempel på en slik gruppe er gruppe 193 bestående av et par der begge er i gruppe 13 (dvs. eldre mann med lav utdanning og eldre kvinne med lav utdanning), der det ved trekning 2 predikeres 46 for mange samboerpar. Det er ikke overraskende at vi prediker for mange samboere i gruppe 193, da det



finnes mange eldre med lav utdanning som er registrert enslige og dermed potensielle samboere uten barn.

Ved å benytte den endrede stokastiske matching prosessen klarte modellen for det første å predikere et korrekt antall samboere uten barn på landsnivå. Derneft klarte denne modellen å gjenskape den faktiske fordelingen av samboere uten barn over de ulike samboergruppene nesten perfekt. Forskjellen mellom den faktiske gruppefordelingen og den predikerte gruppefordelingen er nå marginal og vil være av liten betydning for en analyse av inntektsulikhet.

#### **4.4.2 Kommunalt nivå**

Før vi endret matching prosessen, predikerte vi samboerskap på landsnivå. En slik metode sikret ikke at vi fikk et korrekt antall samboere i hver kommune. Den nye matching prosessen har derimot en slik struktur at vi kan predikere samboerskap på kommunalt nivå, og etterpå aggregere resultatene til landsnivå. I delkapittel 4.4.1 så vi at denne modellen sikret oss et riktig antall samboere på landsnivå etter trekning 2. Et riktig antall samboere predikeres også i den enkelte kommune.

Tabell 4.7 som er presentert i vedlegget viser kommunetilhørigheten til de 299 samboerparene som modellen ikke allokerer en riktig samboergruppe til ved trekning 1. Tabellen viser at de 299 samboerparene tilhører typisk mindre kommuner med få potensielle samboere. For eksempel vil ingen av de 299 samboerparene være bosatt i noen av de fire største bykommunene (Oslo, Bergen, Trondheim og Stavanger). Få potensielle samboere vil gi større sjanse for at vi ved trekking av samboere ikke har nok potensielle samboere. Dette fordi vi har "brukt opp" for mange av gruppenes initiale potensielle samboere i tidligere trekninger.

#### **4.5 Analyse av inntektsulikhet ved å benytte det utvidete familiebegrepet som økonomisk enhet**

I kapittel 3 studerte vi forskjellen i ulikhet i inntektsfordelingen i samfunnet ved å bruke familie og husholdning som økonomisk enhet. Nå vil vi gjennomføre en tilsvarende analyse der vi tar utgangspunkt i registerets familiebegrep, men inkluderer våre predikerte samboerpar. Ved å koble disse samboerparene med DSF får vi et simulert husholdningsregister. Dette SHR-registeret benytter et utvidet familiebegrep som økonomisk enhet, der samboere uten barn ikke lenger er definert som enslige. Ved å predikere samboerskap uten barn tar vi dermed et skritt mot et fullstendig simulert husholdningsregister der husholdning, og ikke familie, vil være den økonomiske enheten.

Ved å gjennomføre en inntektsanalyse av gruppene faktiske samboere uten barn og predikerte samboere uten barn, som begge består av 137382 individer, ser vi at forskjellen i inntektsulikhet er relativt liten mellom disse to gruppene. Faktisk er avviket mellom Gini-koeffisientene kun på 0,002. En slik marginal forskjell i Gini-koeffisient vil være en indikasjon på at vi ved simuleringen av samboere klarer å få representert individer med liknende inntekt som de faktiske samboerne. Det er likevel verd å merke seg at siden de predikerte samboerne til dels er forskjellige fra de faktiske samboerne vil Gini-koeffisientene ikke være direkte sammenliknbare, og resultatene fra en sammenlikning må derfor tolkes med forsiktighet. En kan for eksempel tenke seg en situasjon der modellens predikerte samboere systematisk har en høyere inntekt enn de faktiske samboerne uten barn. I en slik situasjon vil Gini-koeffisienten for gruppen faktiske samboere kunne være lik Gini-koeffisienten for

gruppen predikerte samboere, på tross av at simuleringen av samboere ikke klarer å få representert individer med liknende inntekt som de faktiske samboerne.

Vi kan konstruere multiple SHR-registre ved å gjennomføre uavhengige repetisjoner av den samme prosedyren. Vi slår resultatene fra de multiple husholdningsregistrene sammen for å skaffe en Monte Carlo approksimering til en analyse som gir oss forventningsstørrelser av Gini-koeffisienten og desilgjennomsnittene. Forventningen approksimeres ved å beregne gjennomsnittet for de multiple SHR-registrene. En slik metode kalles multippel imputering og er omtalt i Rubin (1997). Han anbefaler å benytte mellom 5 og 20 repetisjoner av prosedyren, avhengig av forskjellene i SHR-registrenes aktuelle størrelser. Variasjonen i Gini-koeffisientene og desilgjennomsnittene mellom SHR-registrene viser seg å være marginal og uten vesentlig betydning for analyse av inntektsulikhet. Forskjellen mellom den største verdien til Gini-koeffisienten og den minste verdien til Gini-koeffisienten blant de 5 SHR-registrene er 0,00009. Dermed vil det i hvert fall ikke være behov for mer enn 5 SHR-registre for å få en god approksimering av forventningsstørrelsene.

I hvert av disse 5 SHR-registrene er det gjennomført to trekninger for å predikere samboere uten barn, trekning 1 og trekning 2. Hver av disse trekningene innebærer usikkerhet rundt hvilke potensielle samboere som blir trukket. Et individ som blir trukket til å være samboer i ett SHR-register, er ikke nødvendigvis samboer i et av de andre registrene. I praksis vil de predikerte samboernes ekvivalentinntekt variere mellom de enkelte SHR-registrene. Vi tar gjennomsnittet av Gini-koeffisienten og desilgjennomsnittene og får dermed en approksimasjon på forventningen til størrelsene.

Tabell 4.8 viser at Gini-koeffisienten er lavere ved å bruke det utvidete familiebegrepet enn ved å benytte registerets familiebegrep. Dessuten ser vi at Gini-koeffisienten med det utvidete familiebegrepet som enhet, er nærmere Gini-koeffisienten vi får med husholdning som økonomisk enhet, enn Gini-koeffisienten vi får ved å benytte familie som enhet. Det virker rimelig at inntektsulikheten i samfunnet reduseres ved at vi tar innover oss at noen av de registrerte enslige lever i samboerskap uten felles barn og dermed drar fordel av stordriftsfordeler. Men som diskutert i kapittel 3 er dette ikke noe vi kan slå fast a priori. Våre resultater er i tråd med Åserud (2000) sine resultater fra analyser basert på utvalgsdata. Han fikk også som resultat at inntektsulikheten gikk ned når samboere uten barn behandles som en felles husholdning.

**Tabell 4.8: Gini-koeffisienten for fordeling av inntekt etter skatt basert på kvadratrotskalaen med familie, husholdning og et utvidet familiebegrep som økonomisk enhet. År 2000**

Økonomisk enhet:	Gini-koeffisient
Husholdning	0.265
Familie	0.271
Utvidet familiebegrep	0,269

**Tabell 4.9: Desilgruppenes gjennomsnittlige ekvivalentinntekt basert på kvadratrotskalaen med et utvidet familiebegrep som økonomisk enhet. År 2000**

Økonomisk enhet	Utvidet familiebegrep		
	Desilgruppe	Gj.snitt (kr)	Desilgr. gj.snitt/ populasjonens gj.snitt
	0.00-0.10	85400	0,38
	0.10-0.20	128700	0,57
	0.20-0.30	152600	0,68
	0.30-0.40	173000	0,77
	0.40-0.50	191600	0,85
	0.50-0.60	210100	0,93
	0.60-0.70	230700	1,03
	0.70-0.80	256500	1,14
	0.80-0.90	294800	1,31
	0.90-1.00	528000	2,35

Av tabell 4.8 ser vi at vi får et bedre bilde på inntektsulikheten for totalpopulasjonen ved å benytte det utvidete familiebegrepet som enhet i forhold til å bruke familie som enhet. Den summariske statistikken presentert i tabell 4,10 er ment å belyse forskjellen i inntektsulikhet når inntektsfordelingen analyseres på kommunenivå ved å benytte husholdning, familie og det utvidete familiebegrepet som enhet, målt ved avviket i Gini-koeffisient. Det er verd å merke seg at standardavviket i tabellen ikke må tolkes som et mål på utvalgsusikkerheten, men som et mål på variasjonen på Gini-koeffisienten mellom kommunene.

**Tabell 4.10: Summarisk statistikk som viser nivået på Gini-koeffisientene for fordelingen av inntekt etter skatt på kommunenivå basert på kvadratrotskalaen med familie, husholdning og det utvidete familiebegrepet som økonomisk enhet. År 2000**

Økonomisk enhet	Gini-koeffisient			
	Gj.snitt	St. avvik	Min	Max
Familie	0,234	0,036	0,176	0,526
Utvidet familiebegrep	0,232	0,036	0,171	0,524
Husholdning	0,229	0,036	0,180	0,517

Tabell 4,10 viser at det gjennomsnittlige avviket i Gini mellom analysene med henholdsvis det utvidete familiebegrepet og husholdning som enhet er mindre enn det gjennomsnittlige avviket i Gini mellom analysene med henholdsvis familie og husholdning som enhet. En fullstendig analyse av inntektsfordelingen på kommunenivå viser at i 338 av kommunene er Gini-koeffisienten nærmere Gini-koeffisienten med husholdning som enhet når vi benytter det utvidete familiebegrepet som enhet i forhold til om vi benytter familie som enhet. For år der vi ikke har folketellinger og ønsker å benytte registerdata i analyser av fordelingen av inntekt kan det derfor være viktig å basere regionale analyser på SHR og ikke DSF.

#### 4.6 Utblikk

Mot slutten av 1990-tallet fikk inntektsulikhet og fattigdomsproblematikken økt oppmerksomhet. Fattigdom var spesielt i fokus i forkant av Stortingsvalget i 2001. I den siste tiden har det på nytt blitt satt fokus på fattigdom og inntektsulikhet både i økonomisk

litteratur, i den politiske diskusjon og i media. En av årsakene kan være at regjeringen sist høst la frem en egen stortingsmelding hvor fattigdomsproblemet drøftes relativt inngående (Sosialdepartementet, 2002). Dette arbeidet gir ingen beskrivelse av regionale forskjeller i fattigdom og inntektsulikhet, men i kjølvannet av stortingsmeldingen har det blitt gjennomført arbeid som studerer den geografiske variasjonen i fattigdom og inntektsulikhet, blant annet Kirkeberg (2003) og Mogstad (2003).

Til forskjell fra Kirkeberg som benyttet opplysninger fra en utvalgsundersøkelse med et begrenset antall observasjoner, brukte vi data fra FoB2001 og DSF med opplysninger om hele befolkningen i analysen av inntektsulikhet.<sup>14</sup> Ved å benytte registeropplysninger eller folketellingsopplysninger er det mulig å bryte ned resultatene til å beskrive og drøfte forskjeller i inntektsulikhet på kommunenivå. En viktig anvendelse av modellen utviklet i dette arbeidet er at vi skal kunne bruke den til å predikere samboerskap for år hvor vi kun har registerinformasjon og ikke kan bruke informasjon om husholdningenes faktiske sammensetning fra folketellinger. I dette notatet brukte vi modellen til å konstruere husholdningsversjoner av DSF for året 2000. Tilsvarende planlegges det å benytte modellen til å konstruere simulerte husholdningsregistre for tidligere årganger av DSF. Dette vil kreve en interpolasjon av de faktiske samboertallene fra folketellingene i 2001 og 1990. Dette muliggjør å benytte registerdata for å studere utviklingen av inntektsulikhet og fattigdom over tid. En slik studie kan brukes til å belyse både kronisk fattigdom og mobilitet i inntekt og inntektsulikhet på kommunalt nivå.

Det simulerte husholdningsregisteret som er utviklet i dette arbeidet, vil være et første skritt mot et system av simulerte husholdningsregistre med husholdning som økonomisk enhet. Resultatene fra dette arbeidet viste at det fortsatt er forskjeller i inntektsulikhet når husholdning brukes som enhet i forhold til om vi benytter det utvidete familiebegrepet som enhet. Dette skyldes at samboere uten barn kun er en del av forklaringen på avviket mellom DSF og FoB2001. Andre grupper vil for eksempel være samboere med særkullsbarn, samboere der en av partene ikke er registrert enslig, bokollektiver, flergenerasjonsfamilier osv. I tillegg kan en del av avviket mellom opplysninger om familie fra FoB2001 og opplysninger om husholdning fra DSF, skyldes feil i DSF, feil i FoB2001 eller at DSF ikke har like oppdaterte opplysninger om individenes samlivsform som FoB2001 har. Modellen vi utviklet i dette arbeidet som på en tilfredstillende måte predikerte samboerpar uten barn, bør kunne utvides til også å predikere andre typer husholdninger som består av flere familier. Et eksempel er samboere med særkullsbarn hvor minst en av partene ikke vil være registrert enslig, men enslig mor/far med barn. Da er riktignok husholdningsstørrelsen usikker. Men dette kan løses ved først å predikere samboere med ett særkullsbarn, deretter to særkullsbarn osv. Ved å benytte versjoner av metoden introdusert i dette notatet kan det konstrueres simulerte husholdningsregistre som inkluderer flere uregistrerte samlivsformer enn samboerskap uten barn, og dermed bruke et tilnærmet husholdningsbegrep og ikke det utvidete familiebegrepet som økonomisk enhet.

---

<sup>14</sup> Mogstad (2003) gjennomfører analyse av fattigdom på kommunenivå basert på data fra SHR-registre med et utvidet familiebegrep som enhet og FoB2001 med husholdning som enhet.

## 5. Sammendrag

For å analysere inntektsulikhet og fattigdom i befolkningen kreves det ikke bare opplysninger om individets personinntekt, men også informasjon om husholdningens inntekt, størrelse og sammensetning. Husholdningsdata er gjerne basert på folketellinger som gjennomføres med tiårsintervaller eller intervjuundersøkelser fra et utvalg av befolkningen. En relevant alternativ datakilde er Det sentrale folkeregisteret (DSF) koblet med inntektsdata. DSF holdes fortløpende oppdatert og inneholder opplysninger om hele befolkningen, men er basert på familie som økonomisk enhet.<sup>15</sup> Familie er et snevrere begrep enn (kost)husholdning, som omfatter alle som bor og spiser sammen. Derfor vil en analyse av økonomisk velferd basert på registerdata ikke nødvendigvis gi et godt bilde av inntektsulikhet eller fattigdom i et samfunn.

I dette notatet kartla vi avviket mellom opplysninger om familie fra DSF og opplysninger om husholdning fra Folke- og bolig tellingen 2001 (FoB2001). Deretter så vi nærmere på forskjellene i inntektsulikhet i år 2000 ved å bruke familie og husholdning som økonomisk enhet. I analysene av inntektsulikhet i dette arbeidet ble inntekt etter skatt benyttet som inntektsbegrep, individ som analyseenhet og kvadratrot skalaen ble benyttet til å omregne familie- eller husholdningsinntektene til sammenliknbare personinntekter kalt ekvivalentinntekter. Resultatene viser at omtrent 30 prosent av avviket mellom DSF og FoB2001 skyldtes individer som er registrerte som enslige i DSF, men som ifølge FoB2001 lever i samboerskap uten barn med en annen registrert enslig. Analysene av inntektsulikhet med familie og husholdning som enhet viste at inntektsulikheten for totalpopulasjonen er større med familie som enhet i forhold til om vi brukte husholdning som enhet. Metodene som ble benyttet for måling av inntektsulikhet var Gini-koeffisient og desiler. Selv om forskjellen i inntektsulikhet avhengig av valg av enhet ikke er stor for totalpopulasjonen er forskjellen større ved analyse på kommunalt nivå og innenfor spesielle grupper i samfunnet. Av denne grunn fant vi det nødvendig å utvikle en metode for produksjon av husholdningsutgaver av DSF.

Modellen vi utviklet og drøftet er en statistisk modell som benytter registerinformasjon om de registrerte ensliges alder, utdanning, kjønn, kommune og bostedsadresse for å predikere samboerskap i denne gruppen. Ved å utnytte informasjon fra FoB2001 og DSF om forholdet mellom de faktiske samboerne uten barn sin utdanning, alder, kommune og bostedsadresse estimeres modellen. Den estimerte modellen klarte for det første å predikere et korrekt antall samboere på landsnivå. Dernest klarte denne modellen nesten perfekt å gjenskape den faktiske fordelingen av samboere uten barn over de 225 ulike samboergruppene karakterisert ved alder og utdanning. Forskjellen mellom den faktiske gruppefordelingen og den predikerte gruppefordelingen av samboere var kun 299 av 68691 samboerpar. Videre klarte modellen også å predikere et riktig antall samboere uten barn i den enkelte kommune.

Ved å behandle de predikerte samboerne uten barn som samboere og ikke som enslige, får vi et simulert husholdningsregister, SHR, med et utvidet familiebegrep som økonomisk enhet. Vi konstruerte multiple SHR-registre ved å gjennomføre uavhengige repetisjoner av modellens prosedyre. Ved å benytte en Monte Carlo approksimering på de multiple SHRene,

---

<sup>15</sup> For analyser av økonomisk velferd på regionalt nivå kan det være nødvendig å benytte registerdata eller data fra folketellinger. Empiri fra analyser av økonomisk velferd på regionalt nivå basert på intervjuundersøkelser må tolkes med varsomhet fordi antall observasjoner i utvalget setter begrensninger for i hvilken grad statistikken kan brytes ned på mindre geografiske områder.

beregnet vi forventningsstørrelser av Gini-koeffisienten og desilgjennomsnittene når vi benyttet det utvidete familiebegrepet som enhet. Resultatene av denne inntektsanalysen viser at inntektsulikheten med det utvidete familiebegrepet som enhet, er nærmere inntektsulikheten med husholdning som enhet sammenliknet med analysen hvor familie var brukt som enhet. Årsaken til at resultatene fra inntektsanalysen basert på SHR ikke samsvarer perfekt med resultatene fra inntektsanalysen basert på FoB2001, er at samboere uten barn kun er en del av forklaringen på avviket mellom opplysninger om familiesammensetningen fra DSF og opplysninger om husholdningssammensetningen fra FoB2001. Andre grupper av betydning er for eksempel samboere med særkullsbarn, flergenerasjonsfamilier og bokollektiver. På tross av at det fortsatt er et avvik mellom data fra FoB2001 og SHR, illustrerer resultatene i dette notatet at vi får et bedre bilde av inntektsulikheten ved å benytte data fra SHR i forhold til å benytte data fra DSF. SHR vil derfor gi et bedre datagrunnlag enn DSF for analyse av økonomisk velferd i befolkningen for år der vi ikke har data fra folketellinger og ønsker å benytte registerinformasjon istedenfor intervjuundersøkelser.

## Referanser

Aaberge, R. og I. Melby (1998): The Sensitivity of Income Inequality to Choice of Equivalence Scales, *Review of Income and Wealth*, Vol.44 No.4 1998, International Association for Research in Income and Wealth, New York, USA.

Aaberge, R., A. Andersen og T. Wennemo (1999): Temporær og kronisk fattigdom i Norge. 1979-1996, Notater 99/32, Statistisk sentralbyrå.

Aaberge, R. og A. Langørgen (2003): Measuring the Benefits from Public Services. The Effects of Local Government Spending on the Distribution of Income in Norway, Discussion Paper 339/03, Statistisk Sentralbyrå.

Andersen, A., J. Epland, T. Wennemo og R. Aaberge (2003): Økonomiske konjunkturer og fattigdom: En studie basert på norske inntektsdata, 1979-2000, *Tidsskrift for Velferdsforskning*, 2, 2003.

Atkinson, A.B., L. Rainwater og T.M. Smeeding (1995): *Income distribution in OECD countries*, Social Policy Studies 18, OECD.

Berrington, A. og Diamond I. (2000): Marriage or cohabitation: a competing risks analysis of first partnership formation among the 1958 British cohort, *Journal of the Royal Statistical Society*, 163, The Royal Statistical Society, London.

Bojer, H. (1990): *Inntekt og ulikhet*, Rapport 6/1990, Senter for anvendt forskning. Samfunnsøkonomisk institutt, Universitet i Oslo.

Byberg, I. H., A.H. Foss og T. Noack (2001): *Gjete kongens harer - rapport med å få samboere mer innpasset i statistikken*, Rapporter 2001/40, Statistisk sentralbyrå.

Dale, T. (1997): Samordnet levekårsundersøkelsen 1997 - panelundersøkelsen. Dokumentasjonsrapport, Notater 97/58, Statistisk Sentralbyrå.

Epland, J. (1998): *Endringer i fordelingen av husholdningsinntekt:1986-1996*, Rapporter 98/17, Statistisk sentralbyrå.

Ermisch, J. og M. Franchesconi (2000): Cohabitation in Great Britain: not for long but there to stay, *Journal of the Royal Statistical Society*, 163, The Royal Statistical Society, London.

Kirkeberg M. I. (2003): Fattigdom og inntektsfordeling: Oslo - flest fattige og størst ulikhet, *Samfunnsspeilet* 08/2003, Statistisk sentralbyrå.

Langørgen, A. og R. Aaberge (2002): "Fordelingsvirkninger av kommunal tjenesteproduksjon", *Norsk Økonomisk Tidsskrift* 1/02, Statistisk sentralbyrå.

Lund, K., og R. Aaberge (1999): "Effekten av valg av ekvivalensskala på tallfesting av omfang, fordeling og utvikling av fattigdom i Norge 1982-1995", Notater 99/32, Statistisk sentralbyrå.

Mogstad, M. (2003): Analyse av fattigdom og inntektsulikhet basert på registerdata, Hovedoppgave, Samfunnsøkonomisk institutt, Universitet i Oslo.

Noack, T. (2002): Samboere med felles barn. En gruppe med mye gjennomtrekk, *Økonomiske analyser* 3/2002, Statistisk sentralbyrå.

Rubin, D. (1997): *Multiple Imputation for Nonresponse in Surveys*, New York, Wiley.

Sosialdepartementet (2002): *Stortingsmelding nr. 6 (2002-2003)*, Tiltaksplan mot fattigdom, Sosialdepartementet.

Statistisk sentralbyrå (2002): "Om statistikken". - URL:  
<http://www.ssb/fob2001/om.html>

Statistisk sentralbyrå (2002): "Utdanningsnivå i befolkningen". - URL:  
<http://www.ssb.no/utniv/>

Sundvoll, A. og H.M. Teigum (1998): Samordna levekårsundersøkelse 1997-tversnittundersøkelsen. Dokumentasjonsrapport, Notater 98/34, Statistisk sentralbyrå.

Texmon, I. (1999): Samliv i Norge mot slutten av 1900-tallet. En beskrivelse av endringer og mangfold, Samboere og Samfunnet, NOU 1999:25. Barne- og familiedepartementet, Oslo.

Van Praag, B.M.S. and N.L. Van der Sar (1988): Household cost functions and equivalence scales, *Journal of Human Resources*, 23, 1988.

Øyangen, I. (1999): Inntekts og formuesundersøkelsen 1998, Dokumentasjonsrapport, Notater 99/68, Statistisk sentralbyrå.

Åserud, R. (2000): Analyse av inntektsfordeling og inntektsulikhet basert på registerdata, Hovedoppgave, Samfunnsøkonomisk institutt, Universitet i Oslo.



## Vedlegg: Tabellene 4.3, 4.4, 4.6 og 4.7 samt figur 4.1

**Tabell 4.3: Fordelingen av de faktiske samboerne uten barn fra FoB2001 over samboergruppene**

Gruppe:	Antall	Prosent	Gruppe:	Antall	Prosent	Gruppe:	Antall	Prosent
1	78	0,114	76	73	0,106	151	0	0
2	281	0,409	77	978	1,424	152	2	0,003
3	27	0,039	78	885	1,288	153	0	0
4	42	0,061	79	402	0,585	154	13	0,019
5	145	0,211	80	3521	5,126	155	61	0,089
6	50	0,073	81	6741	9,814	156	10	0,015
7	7	0,01	82	83	0,121	157	161	0,234
8	16	0,023	83	580	0,844	158	591	0,86
9	3	0,004	84	736	1,071	159	153	0,223
10	0	0	85	11	0,016	160	1403	2,042
11	1	0,001	86	54	0,079	161	3436	5,002
12	0	0	87	62	0,09	162	1115	1,623
13	0	0	88	0	0	163	481	0,7
14	0	0	89	2	0,003	164	648	0,943
15	0	0	90	5	0,007	165	231	0,336
16	369	0,537	91	3	0,004	166	0	0
17	5530	8,051	92	5	0,007	167	0	0
18	635	0,924	93	1	0,001	168	0	0
19	273	0,397	94	27	0,039	169	5	0,007
20	2460	3,581	95	107	0,156	170	10	0,015
21	720	1,048	96	33	0,048	171	7	0,01
22	21	0,031	97	110	0,16	172	28	0,041
23	137	0,199	98	323	0,47	173	149	0,217
24	34	0,049	99	75	0,109	174	139	0,202
25	3	0,004	100	169	0,246	175	190	0,277
26	12	0,017	101	285	0,415	176	769	1,12
27	2	0,003	102	39	0,057	177	1067	1,553
28	0	0	103	12	0,017	178	75	0,109
29	0	0	104	6	0,009	179	160	0,233
30	0	0	105	1	0,001	180	206	0,3
31	27	0,039	106	2	0,003	181	0	0
32	630	0,917	107	25	0,036	182	1	0,001
33	330	0,48	108	2	0,003	183	0	0
34	34	0,049	109	119	0,173	184	1	0,001
35	374	0,544	110	621	0,904	185	5	0,007
36	348	0,507	111	249	0,362	186	1	0,001
37	3	0,004	112	410	0,597	187	2	0,003
38	15	0,022	113	1731	2,52	188	7	0,01
39	10	0,015	114	528	0,769	189	1	0,001
40	0	0	115	371	0,54	190	195	0,284
41	1	0,001	116	1072	1,561	191	221	0,322
42	1	0,001	117	300	0,437	192	45	0,066
43	0	0	118	31	0,045	193	1106	1,61
44	0	0	119	40	0,058	194	801	1,166
45	0	0	120	9	0,013	195	98	0,143
46	39	0,057	121	2	0,003	196	1	0,001
47	90	0,131	122	12	0,017	197	4	0,006
48	32	0,047	123	2	0,003	198	1	0,001

49	137	0,199	124	41	0,06	199	0	0	
50	383	0,558	125	259	0,377	200	1	0,001	
51	234	0,341	126	472	0,687	201	0	0	
52	43	0,063	127	120	0,175	202	2	0,003	
53	103	0,15	128	658	0,958	203	5	0,007	
54	36	0,052	129	779	1,134	204	2	0,003	
55	6	0,009	130	82	0,119	205	104	0,151	
56	16	0,023	131	345	0,502	206	202	0,294	
57	9	0,013	132	441	0,642	207	83	0,121	
58	2	0,003	133	6	0,009	208	449	0,654	
59	0	0	134	9	0,013	209	688	1,002	
60	1	0,001	135	13	0,019	210	260	0,379	
61	122	0,178	136	0	0	211	0	0	
62	1346	1,959	137	0	0	212	0	0	
63	300	0,437	138	0	0	213	1	0,001	
64	562	0,818	139	6	0,009	214	0	0	
65	4895	7,126	140	29	0,042	215	0	0	
66	2040	2,97	141	4	0,006	216	0	0	
67	227	0,33	142	117	0,17	217	0	0	
68	945	1,376	143	304	0,443	218	2	0,003	
69	268	0,39	144	37	0,054	219	0	0	
70	40	0,058	145	1019	1,483	220	12	0,017	
71	80	0,116	146	1627	2,369	221	28	0,041	
72	21	0,031	147	253	0,368	222	43	0,063	
73	0	0	148	382	0,556	223	49	0,071	
74	4	0,006	149	350	0,51	224	127	0,185	
75	1	0,001	150	78	0,114	225	146	0,213	
							Kumulativ:	frekvens	prosent
								68691	100

**Tabell 4.4: Fordelingen av modellens predikerte samboere uten barn over samboergruppene**

Gruppe	Antall	Prosent	Gruppe:	Antall	Prosent	Gruppe	Antall	Prosent
1	80	0,117	76	36	0,052	151	0	0,000
2	199	0,290	77	867	1,263	152	88	0,128
3	12	0,017	78	453	0,660	153	0	0,000
4	38	0,055	79	228	0,332	154	49	0,071
5	122	0,178	80	3046	4,439	155	267	0,389
6	26	0,038	81	6819	9,937	156	94	0,137
7	11	0,016	82	66	0,096	157	159	0,232
8	31	0,045	83	535	0,780	158	796	1,160
9	10	0,015	84	593	0,864	159	202	0,294
10	0	0,000	85	21	0,031	160	1150	1,676
11	12	0,017	86	84	0,122	161	3041	4,432
12	0	0,000	87	93	0,136	162	845	1,231
13	0	0,000	88	0	0,000	163	525	0,765
14	0	0,000	89	33	0,048	164	615	0,896
15	0	0,000	90	21	0,031	165	178	0,259
16	225	0,328	91	4	0,006	166	0	0,000
17	4246	6,188	92	24	0,035	167	0	0,000
18	273	0,398	93	1	0,001	168	0	0,000
19	178	0,259	94	41	0,060	169	15	0,022
20	2147	3,129	95	103	0,150	170	64	0,093
21	563	0,820	96	34	0,050	171	48	0,070
22	40	0,058	97	97	0,141	172	30	0,044
23	211	0,307	98	247	0,360	173	172	0,251
24	55	0,080	99	44	0,064	174	155	0,226
25	22	0,032	100	134	0,195	175	117	0,170
26	64	0,093	101	186	0,271	176	571	0,832
27	29	0,042	102	16	0,023	177	727	1,059
28	0	0,000	103	19	0,028	178	63	0,092
29	0	0,000	104	13	0,019	179	130	0,189
30	0	0,000	105	2	0,003	180	129	0,188
31	6	0,009	106	7	0,010	181	0	0,000
32	276	0,402	107	66	0,096	182	242	0,353
33	87	0,127	108	5	0,007	183	0	0,000
34	9	0,013	109	107	0,156	184	52	0,076
35	180	0,262	110	685	0,998	185	399	0,581
36	133	0,194	111	252	0,367	186	141	0,205
37	2	0,003	112	270	0,393	187	105	0,153
38	15	0,022	113	1420	2,069	188	373	0,544
39	12	0,017	114	372	0,542	189	85	0,124
40	0	0,000	115	270	0,393	190	640	0,933
41	4	0,006	116	902	1,314	191	1009	1,470
42	4	0,006	117	205	0,299	192	221	0,322
43	0	0,000	118	62	0,090	193	3654	5,325
44	0	0,000	119	70	0,102	194	2086	3,040
45	0	0,000	120	10	0,015	195	176	0,256
46	25	0,036	121	4	0,006	196	29	0,042
47	60	0,087	122	49	0,071	197	214	0,312
48	8	0,012	123	12	0,017	198	28	0,041
49	84	0,122	124	27	0,039	199	0	0,000
50	239	0,348	125	242	0,353	200	181	0,264
51	114	0,166	126	459	0,669	201	0	0,000

52	30	0,044	127	64	0,093	202	54	0,079
53	67	0,098	128	474	0,691	203	214	0,312
54	17	0,025	129	544	0,793	204	108	0,157
55	15	0,022	130	48	0,070	205	225	0,328
56	22	0,032	131	276	0,402	206	602	0,877
57	9	0,013	132	289	0,421	207	262	0,382
58	10	0,015	133	31	0,045	208	996	1,451
59	0	0,000	134	21	0,031	209	1432	2,087
60	1	0,001	135	20	0,029	210	415	0,605
61	76	0,111	136	0	0,000	211	0	0,000
62	1148	1,673	137	0	0,000	212	0	0,000
63	119	0,173	138	0	0,000	213	7	0,010
64	351	0,511	139	33	0,048	214	0	0,000
65	4227	6,160	140	132	0,192	215	0	0,000
66	1596	2,326	141	30	0,044	216	0	0,000
67	156	0,227	142	126	0,184	217	0	0,000
68	858	1,250	143	355	0,517	218	39	0,057
69	205	0,299	144	47	0,068	219	0	0,000
70	53	0,077	145	881	1,284	220	26	0,038
71	127	0,185	146	1231	1,794	221	59	0,086
72	44	0,064	147	186	0,271	222	62	0,090
73	0	0,000	148	378	0,551	223	67	0,098
74	27	0,039	149	292	0,426	224	153	0,223
75	16	0,023	150	46	0,067	225	150	0,219
<b>Kumulativ:</b>							frekvens	prosent
							68622	100,00

**Tabell 4.6: Avvikene i de ulike samboergruppene fra den faktiske fordelingen av samboere over samboergruppene fra FoB2001 ved bruk av den endrede stokastiske matching prosessen**

Gruppe	Antall samboerpar gruppen mangler før trekning 2	Antall samboerpar gruppen er allokert etter trekning 2	Differanse i absolutt verdi
1	1	1	0
5	2	0	2
6	0	1	1
7	0	1	1
11	0	2	2
16	2	1	1
17	8	5	3
18	4	0	4
19	0	1	1
20	11	2	9
21	5	2	3
22	0	1	1
23	1	1	0
24	1	1	0
26	0	2	2
32	3	2	1
33	3	0	3
34	1	0	1
35	1	0	1
36	3	0	3
39	0	1	1
41	0	1	1
46	0	1	1
47	0	1	1
50	1	0	1
51	3	0	3
52	1	0	1
53	2	0	2
57	0	1	1
58	0	2	2
62	6	0	6
63	2	0	2
64	6	3	3
65	10	4	6
66	4	3	1
67	3	0	3
68	4	4	0
69	3	0	3
73	0	1	1
74	0	1	1
76	1	0	1
77	1	0	1
78	4	1	3
79	3	1	2
80	13	3	10
81	34	0	34
82	1	0	1

83	2	0	2
84	1	1	0
85	1	0	1
86	1	0	1
89	0	1	1
91	0	1	1
95	1	0	1
96	0	1	1
97	2	0	2
98	0	1	1
99	1	0	1
100	2	0	2
101	4	0	4
103	0	2	2
104	1	0	1
109	1	0	1
110	1	3	2
111	0	1	1
112	7	3	4
113	12	1	11
114	1	0	1
115	2	0	2
116	13	3	10
117	1	0	1
119	1	1	0
120	1	1	0
125	2	0	2
126	2	1	1
127	2	1	1
128	2	0	2
129	2	0	2
131	4	1	3
132	4	0	4
137	0	2	2
139	1	0	1
140	0	3	3
142	1	2	1
143	2	2	0
144	1	0	1
145	4	4	0
146	4	1	3
147	1	2	1
148	3	4	1
149	0	2	2
152	0	3	3
154	0	2	2
155	1	5	4
156	0	2	2
157	2	1	1
158	4	4	0
159	1	0	1
160	6	3	3
161	12	2	10
162	2	0	2
163	5	3	2
164	3	2	1
165	2	0	2

173	1	1	0
175	2	1	1
176	6	3	3
177	2	1	1
178	1	0	1
179	1	1	0
180	1	0	1
182	0	1	1
184	0	4	4
185	0	5	5
186	0	3	3
187	0	10	10
188	0	5	5
189	0	1	1
190	0	12	12
191	0	8	8
192	0	1	1
193	0	46	46
194	0	20	20
195	0	4	4
197	0	3	3
199	0	1	1
200	0	1	1
201	0	3	3
202	0	2	2
203	0	3	3
204	0	3	3
205	2	3	1
206	2	3	1
207	1	1	0
208	1	17	16
209	0	4	4
210	1	3	2
215	0	1	1
223	1	1	0
<b>Totalt</b>	<b>299</b>	<b>299</b>	<b>416</b>

**Tabell 4.7: Kommunetilhørigheten til de 299 samboerparene som modellen ved trekning 1 ikke allokerte en riktig samboergruppe til**

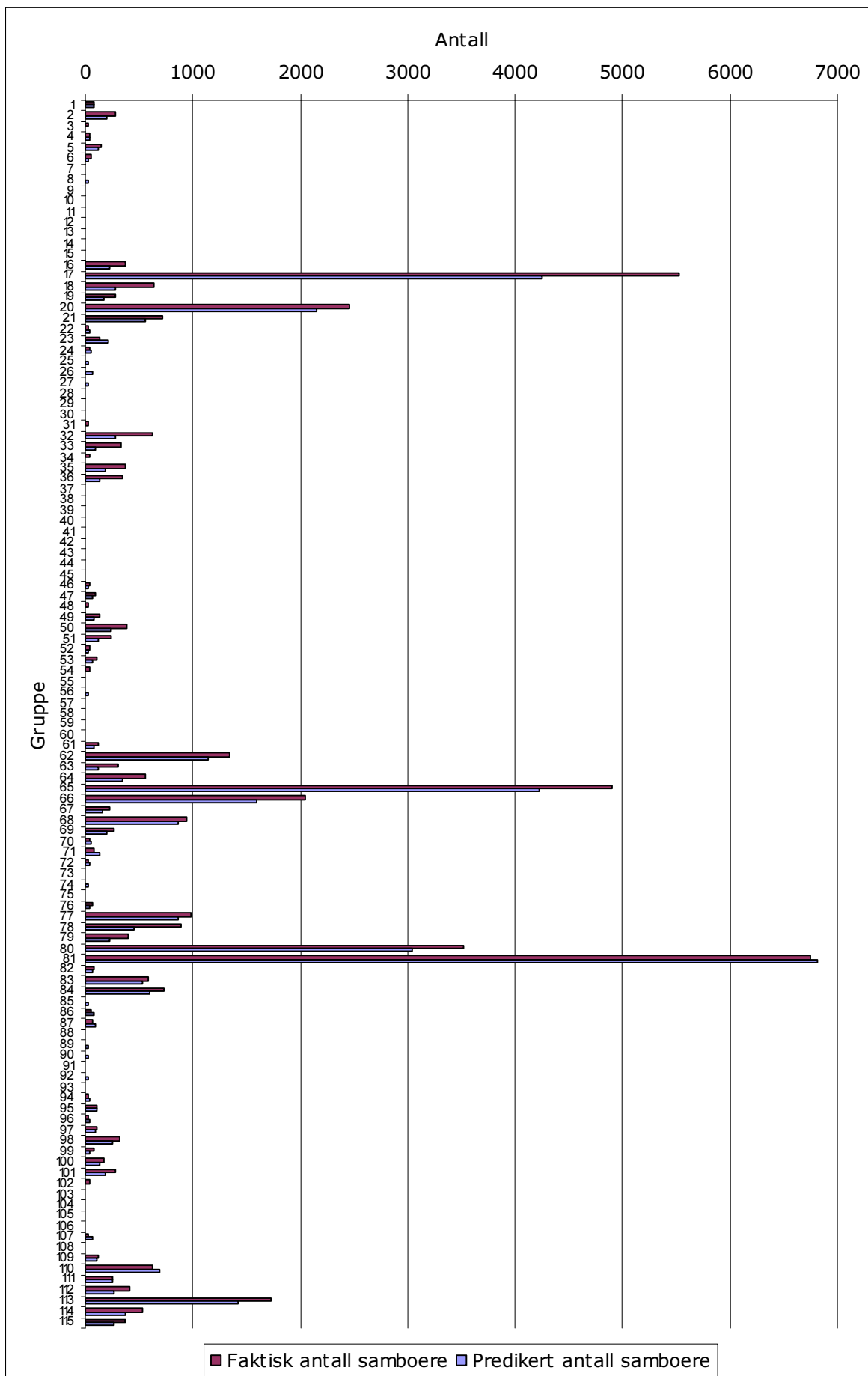
Kom. navn	Gruppe	Avvik fra den faktiske fordelingen etter trekning 1	Kom. navn	Gruppe	Avvik fra den faktiske fordelingen etter trekning 1
Halden	193	1	Fjell	68	1
Sarpsborg	191	1	Fjell	190	1
Aremark	190	1	Askøy	142	1
Marker	193	2	Osterøy	193	1
Marker	194	1	Meland	23	1
Rømskog	193	1	Meland	26	1
Trøgstad	65	1	Øygarden	193	1
Trøgstad	164	1	Austrheim	17	1
Trøgstad	187	1	Masfjorden	208	1
Skiptvet	188	1	Flora	68	1
Rakkestad	143	1	Flora	137	1
Råde	190	1	Flora	161	1
Våler	193	1	Høyanger	193	1
Hobøl	11	1	Balestrand	201	1
Hobøl	41	1	Aurland	11	1
Vestby	215	1	Aurland	193	1
Ski	154	1	Lærdal	152	1
Ski	155	1	Luster	193	1
Frogn	142	1	Askvoll	205	1
Oppegård	116	1	Fjaler	6	1
Oppegård	185	1	Jølster	209	1
Oppegård	194	1	Selje	158	1
Oppegård	201	1	Gloppen	209	1
Rælingen	208	1	Kristiansund	203	1
Skedsmo	187	1	Kristiansund	208	1
Skedsmo	200	1	Vanylven	209	1
Nittedal	58	1	Sande	210	1
Nittedal	116	1	Herøy	186	1
Nittedal	193	1	Herøy	208	1
Nittedal	197	1	Ulstein	20	1
Gjerdrum	127	1	Ulstein	66	1
Nes	176	1	Hareid	21	1
Eidsvoll	64	1	Ørsta	190	1
Nannestad	66	1	Stranda	188	1
Hurdal	80	1	Stranda	193	1
Hamar	193	1	Stranda	194	1
Ringsaker	184	1	Sykkylven	176	1
Løten	140	1	Skodje	189	1
Nord-Odal	113	1	Giske	184	1
Nord-Odal	193	1	Giske	208	1
Sør-Odal	74	1	Rauma	148	1
Eidskog	16	1	Rauma	185	2
Åsnes	158	1	Nesset	21	1
Åsnes	194	1	Nesset	208	1
Åsnes	202	1	Midsund	112	1
Trysil	32	1	Midsund	199	1
Os	103	1	Sandøy	190	1
Lillehammer	111	1	Sandøy	193	1
Gjøvik	131	1	Fræna	145	1
Lesja	194	1	Frei	17	1

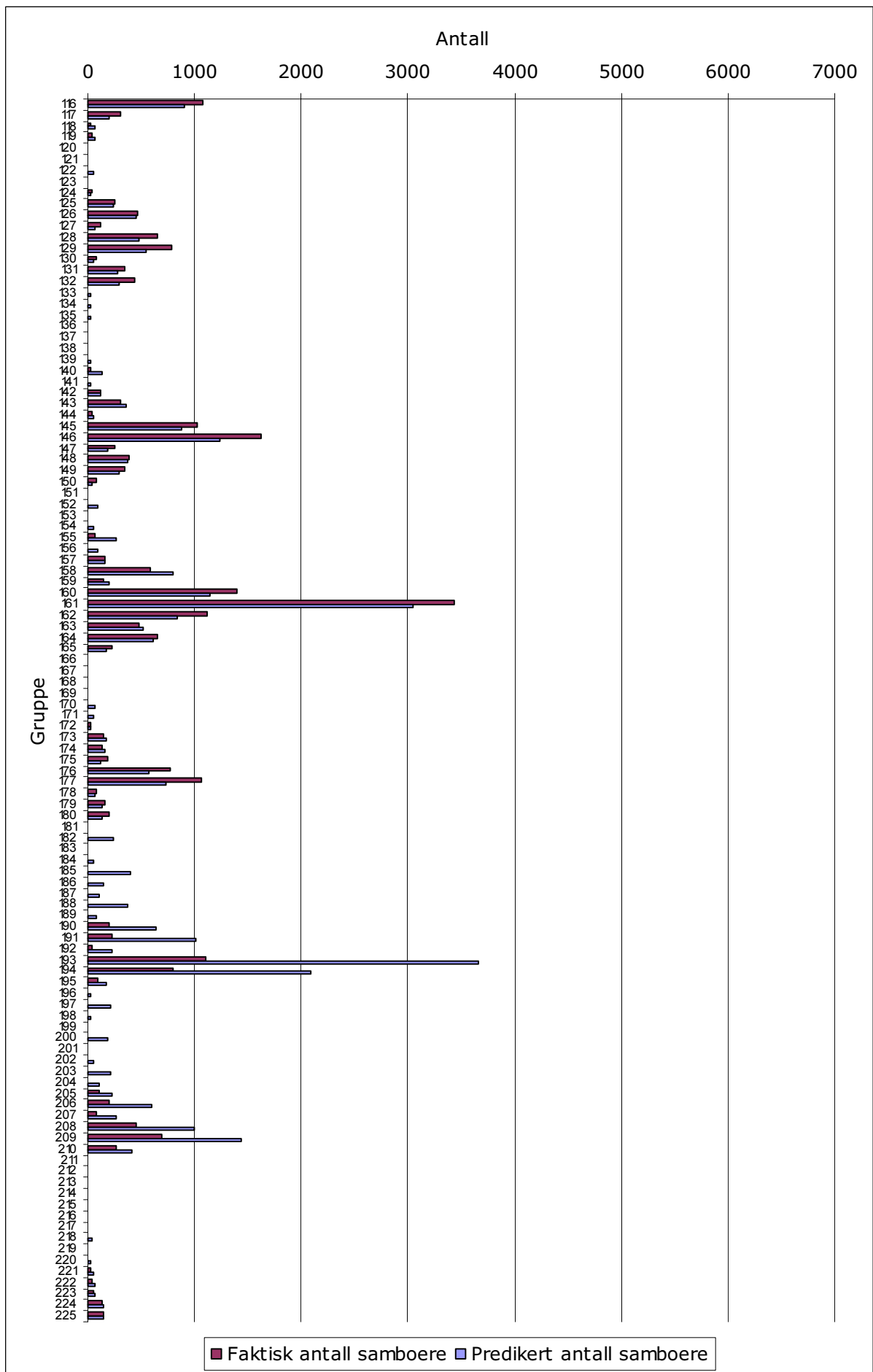


Lom	187	1	Frei	73	1
Sel	208	1	Aure	188	1
Sør-Fron	89	1	Aure	194	1
Gausdal	187	1	Aure	208	1
Vestre Toten	68	1	Hemne	187	1
Jevnaker	190	1	Hemne	190	1
Gran	187	1	Hitra	187	1
Søndre Land	80	1	Frøya	185	1
Sør-Aurdal	208	1	Frøya	208	1
Etnedal	157	1	Agdenes	47	1
Etnedal	193	1	Roan	193	1
Etnedal	210	1	Meldal	119	1
Nord-Aurdal	7	1	Meldal	148	1
Nord-Aurdal	64	1	Meldal	188	1
Øystre Slidre	84	1	Rørøros	191	1
Vang	205	1	Holtålen	163	1
Hole	145	1	Holtålen	194	1
Flå	203	1	Skaun	149	1
Gol	223	1	Steinkjer	158	1
Hemsedal	20	1	Meråker	208	1
Hemsedal	64	1	Stjørdal	197	1
Hemsedal	202	1	Levanger	194	1
Ål	112	1	Verdal	68	1
Ål	173	1	Verdal	194	1
Hol	66	1	Verdal	204	1
Hol	79	1	Mosvik	193	1
Sigdal	177	1	Namdalseid	193	1
Sigdal	193	1	Inderøy	1	1
Krødsherad	195	1	Grong	22	1
Øvre Eiker	152	1	Grong	116	1
Øvre Eiker	210	1	Grong	160	1
Lier	58	1	Fosnes	193	1
Lier	155	1	Flatanger	182	1
Lier	193	1	Vikna	193	1
Røyken	145	1	Nærøy	208	1
Røyken	193	1	Leka	80	1
Borre	190	1	Bodø	126	1
Svelvik	147	1	Bodø	155	1
Våle	193	1	Sømna	184	1
Stokke	187	1	Sømna	187	1
Nøtterøy	160	1	Vega	147	1
Porsgrunn	208	1	Vega	186	1
Kragerø	163	1	Alstahaug	156	1
Drangedal	65	1	Alstahaug	190	1
Bø	39	1	Alstahaug	204	1
Bø	146	1	Leirfjord	179	1
Sauherad	193	1	Vefsn	17	1
Seljord	188	1	Vefsn	193	1
Seljord	194	1	Lurøy	193	1
Nissedal	155	1	Rødøy	112	1
Nissedal	190	1	Saltdal	194	1
Fyresdal	175	1	Skjerstad	164	1
Tokke	152	1	Røst	193	2
Tokke	187	1	Flakstad	110	1
Tokke	193	1	Flakstad	186	1
Vinje	185	1	Vestvågøy	154	1
Lillesand	110	1	Bø	17	1

Lillesand	161	1	Bø	91	1
Lillesand	191	1	Bø	191	1
Bykle	149	1	Sortland	193	1
Bykle	158	1	Andøy	194	1
Mandal	26	1	Moskenes	193	1
Mandal	194	1	Kvæfjord	46	1
Farsund	176	1	Gratangen	137	1
Flekkefjord	194	1	Bardu	110	1
Vennesla	140	1	Målselv	32	1
Søgne	65	1	Målselv	78	1
Åseral	193	1	Målselv	195	1
Sirdal	120	1	Dyrøy	160	1
Sirdal	208	1	Torsken	193	1
Sokndal	208	1	Torsken	197	1
Klepp	207	1	Lenvik	17	1
Time	155	1	Lenvik	143	1
Gjesdal	24	1	Lenvik	163	1
Gjesdal	190	1	Lenvik	193	1
Randaberg	184	1	Balsfjord	205	1
Randaberg	195	1	Karlsøy	193	2
Sauda	194	1	Karlsøy	194	1
Kvitsøy	193	1	Lyngen	193	1
Tysvær	208	1	Storfjord	194	1
Karmøy	193	1	Nordreisa	98	1
Etne	191	1	Vardø	156	1
Etne	194	1	Vardø	204	1
Etne	208	1	Vadsø	19	1
Stord	194	1	Vadsø	148	1
Stord	201	1	Hammerfest	193	1
Tysnes	195	1	Hasvik	140	1
Kvinnherad	145	1	Kvalsund	193	1
Ulvik	190	1	Nordkapp	191	1
Granvin	209	1	Karasjohka-Karasjok	206	1
Voss	65	1	Gamvik	96	1
Voss	148	1	Gamvik	191	1
Kvam	194	1	Berlevåg	192	1
Fusa	193	1	Deatnu - Tana	103	1
Os	203	1	Deatnu - Tana	191	1
Os	206	2	Båtsfjord	57	1
Sund	193	1	Sør-Varanger	193	1

**Figur 4. 1: Faktiske og predikerte fordelinger av samboere over samboergruppene.**





## De sist utgitte publikasjonene i serien Notater

- |         |  |         |  |
|---------|--|---------|--|
| 2003/28 | A. K. Mevik: Usikkerhet i konjunkturbarometeret. 50s.  | 2003/43 | Å. Cappelen, T. Eika, P.R Johansen og J.A. Jørgensen: Makroøkonomiske konsekvenser av lavere aktivitet i oljevirksomheten framover. 30s.   |
| 2003/29 | A. Finstad og K. Rypdal: Bruk av helse- og miljøfaglige produkter i hisholdningene - et forprosjekt. 33s.  | 2003/44 | L. Østby: Innvandring fra nye EU- land; fortid, nåtid og mulig framtid. 44s.   |
| 2003/30 | T. Jørgensen: Dokumentasjon av prosjektet "Overgang utdanning-arbeid". Årgangene 1999-2000. 54s.   | 2003/45 | T. Dale, H. Høie og A-K.Johnsen: Evaluering av "Naturressurser og miljø" 30s.  |
| 2003/31 | Å. Cappelen og L.S. Stambøl: Virkninger av å fjerne regionale forskjeller i arbeidsgiveravgiften og noen mulige mottiltak. 35s.  | 2003/46 | L. Solheim: Foreløpige landstall i KOSTRA. Prinsipper, metoder, produksjon og eksemper. 76s  |
| 2003/32 | A. Rognan: Forprosjekt om studenters levkår. 31s.  | 2003/47 | A. Hurlen Foss: kvaliteten i boligdelen av Folke- og bolig tellingen. 32s.   |
| 2003/33 | S. Vatne Pettersen: Bosettningsmønster og segregasjon i storbyregionene. Ikke-vestlige innvandrere og grupper med høy og lav utdanning. Utredninger til Storbymeldingen, del 1. 71s. | 2003/48 | E. Siig Meen og O. Rognstad: Jordbrukstelling 1999- dokumentasjon. 105s.   |
| 2003/34 | A. Barstad og M.I. Kirkeberg: Levkår og ulikhet i storby. Utredninger til Storbymeldingen, del 2. 95s.   | 2003/49 | L.Rogstad: Statistiske temakart og PX-Map. 32s.  |
| 2003/35 | E.H. Nymoen, L. Østby og A. Barstad: Flyttinger og pendling i storbyregionene. Utredninger til Storbymeldingen del 3. 75s.   | 2003/50 | E. Holmøy: Velferdsregnskap - et mulig teoretisk rammeverk.35s.  |
| 2003/36 | A. Andersen , T. Løwe og E. Rønning: boforhold i storby. Utredninger til Storbymeldingen, del 4. 82s.  | 2003/51 | C. Wiecek: Undersøkelse om fremtidsplaner, familie og samliv. Dokumentasjonsrapport. 59s.  |
| 2003/37 | D. Sve: Seksualitet og helse. Dokumentasjon av datafangsten. 19s.  | 2003/52 | KOSTRA: Arbeidsgrupperapporter 2003. 153s.   |
| 2003/39 | C. Nordseth og T. Sandnes: FD - Trygd. Dokumentasjonsrapport. 1992-2001. 110s.   | 2003/53 | A. Haglund: Rapport fra arbeidsgruppa om forslag til arbeidsdeling mellom Brønnøysundregistrene (BR) og Statistisk sentralbyrå (SSB). 40s. |
| 2003/40 | A. Langørgen og R. Åserud: Faktorer bak kommunale variasjoner i utgifter til sosialhjelp i 2000. 20s.  | 2003/54 | E. Eng Eibak: Forventningsindikator - konsumprisene. Mai - november 2003. 19s.   |
| 2003/41 | T.M. Normann: Omnibusundersøkelsen februar/mars 2003. Dokumentasjonsrapport. 35s   | 2003/55 | G. Daugstad: Levkår for ungdom i større byer. 80s.   |
| 2003/42 | D.E. Somervoll: TROLL kan temmes. Kort innføring i Trollprogrammering. 13s.  | 2003/56 | A. Vedø og D. Rafat: Sammenligning av utvalgsplaner i AKU. 17s.  |
|         |  | 2003/57 | L. Belsby: Frafall og vekter i Tidsbruksundersøkelsen 2000-2001. 20s.  |
|         |  | 2003/58 | L.Belsby: Vekter i Forbruksundersøkelsen. 28s.   |