

# RAPPORTER

## INNETKTSBEGREPER- INNETKTSFORDELING

AV  
TORSTEIN BYE OG TOR EIVIND HØYLAND

STATISTISK SENTRALBYRÅ  
CENTRAL BUREAU OF STATISTICS OF NORWAY

RAPPORTER FRA STATISTISK SENTRALBYRÅ 81/31

# INNTEKTSBEGREPER - INNTEKTSFORDELING

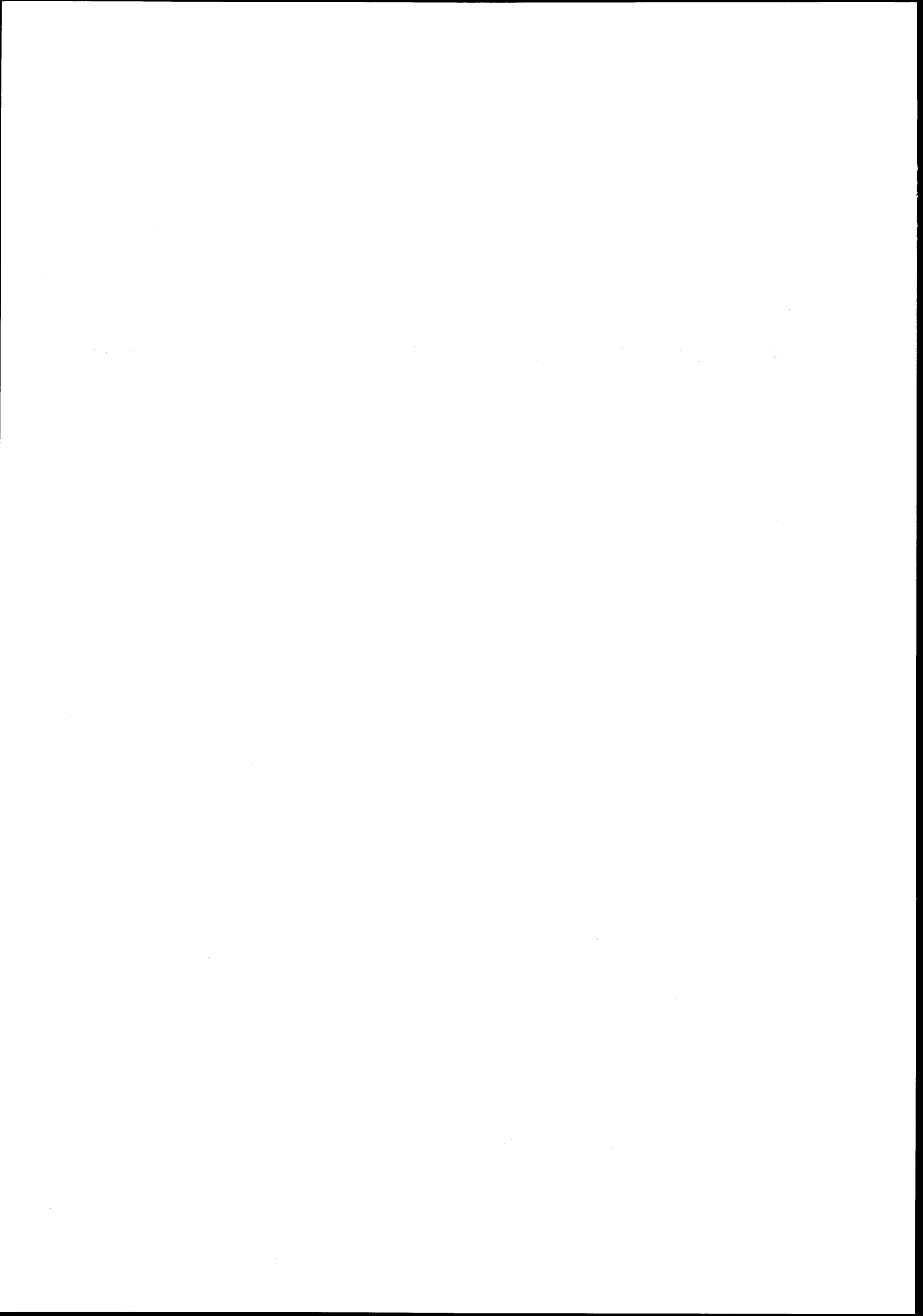
AV

TORSTEIN BYE OG TOR EIVIND HØYLAND

STATISTISK SENTRALBYRÅ  
OSLO — KONGSVINGER 1981

ISBN 82-537-1639-7

ISSN 0332-8422

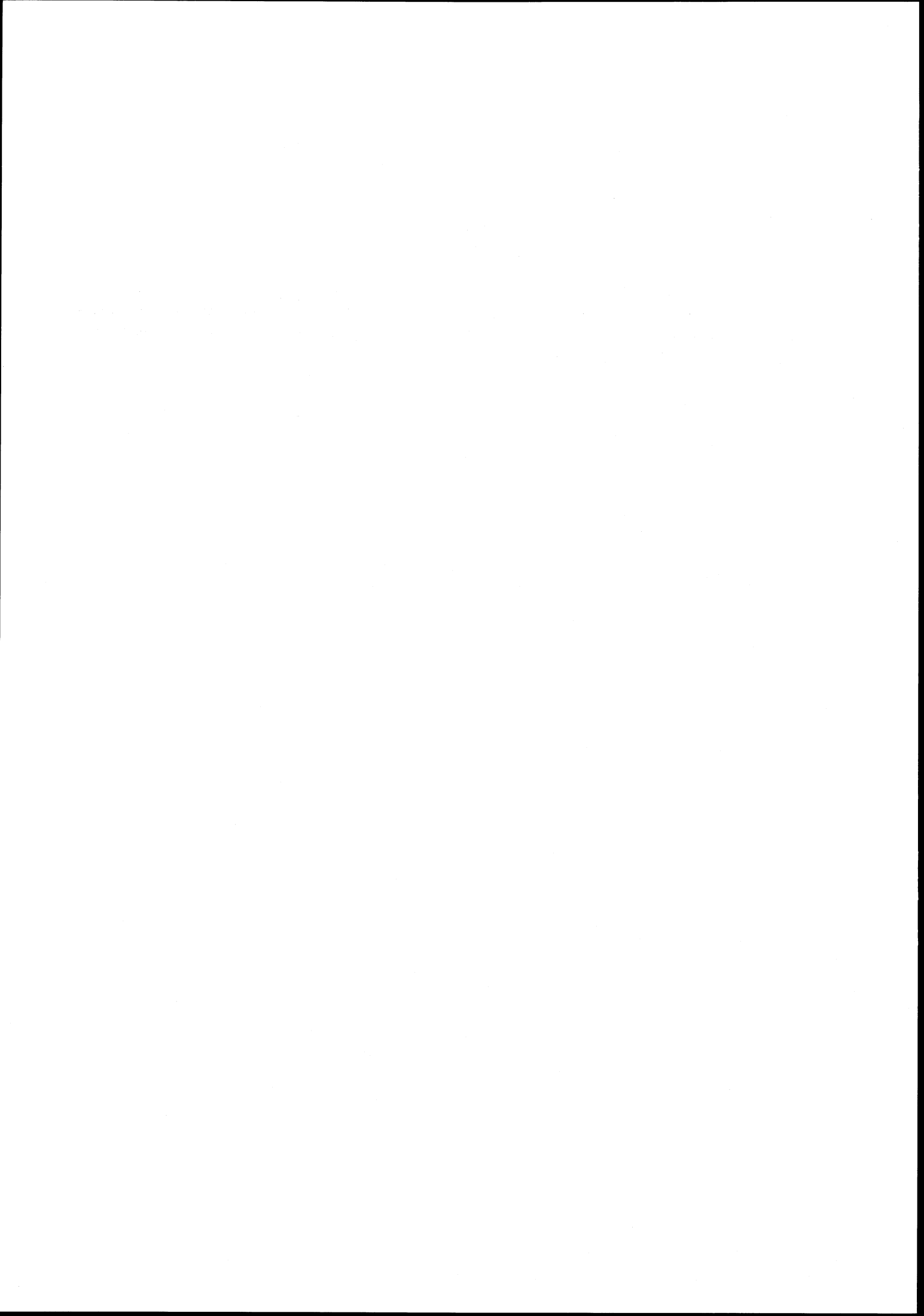


## FORORD

Denne rapporten gir et oversyn over de ulike inntektsbegreper som brukes i inntektsstatistikk fra Statistisk Sentralbyrå. Den beskriver også en del metoder til analyse av inntektsfordeling. Disse metodene blir anvendt til analyse av inntektsfordelingene i 1973, 1976 og 1979. Dataene til analysene er hentet fra inntekts- og formuesundersøkelsene.

Statistisk Sentralbyrå, Oslo, 4. desember 1981

Arne Øien



# INNHold

	Side
1. Innledning .....	7
2. Inntektsbegreper i Statistisk Sentralbyrå .....	7
2.1. Innledning .....	7
2.2. Lønnsstatistikk .....	7
2.3. Nasjonalregnskapets årslønnstall .....	8
2.4. Skattestatistikk .....	8
2.5. Inntektsstatistikk .....	8
2.6. Andre inntektsstatistikker .....	9
2.7. En sammenlikning av de ulike inntektsstatistikkene .....	9
2.8. En framstilling av forholdet mellom de ulike inntektsstatistikkene .....	11
3. Mål til beskrivelse av inntektsfordelingen .....	12
3.1. Innledning .....	12
3.2. Mål for beliggenhet av inntektsfordelingen .....	12
3.3. Mål for spredning .....	13
3.4. Sammenlikning av inntektsfordelinger .....	13
3.5. Variasjonskoeffisienten .....	14
3.6. Fraktiler .....	15
3.7. Lorenz-kurven .....	16
3.8. Gini-koeffisienten .....	17
4. Inntektsfordelingens avhengighet av inntektstakerheten og inntektsbegrepet belyst ved Lorenz-kurven og gini-koeffisienten .....	19
4.1. Innledning .....	19
4.2. Usikkerhet ved å benytte inntektsstatistikkens data .....	19
4.3. Problemstillinger .....	19
4.4. Konklusjoner .....	24
5. Lar inntektsfordelingen seg beskrive av en parametrisk kontinuerlig funksjon? .....	24
5.1. Innledning .....	24
5.2. Den empiriske fordelingen .....	25
5.3. Den lognormale fordelingen .....	27
5.4. Inntektsfordelingen beskrevet med en lognormal fordelingsfunksjon .....	27
5.5. Gammafordelingen .....	28
5.6. Inntektsfordelingen beskrevet med en gamma fordelingsfunksjon .....	29
5.7. En sammenlikning av funksjonene funnet i 4.4. og 4.6. og den empiriske inntektsfordelingen beskrevet i inntektsstatistikken .....	29
5.8. Gammafordelingens parametre .....	34
5.9. Oppsummering .....	34
 Vedlegg	
1. Litteraturliste .....	37
2. Tabeller .....	39
Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) .....	42



## 1. INNLEDNING

Formålet med denne rapporten er to-delt. I en første del ser vi på de statistikkene Statistisk Sentralbyrå utgir for å belyse lønns- og inntektsforholdene i Norge. I denne delen (kapittel 2) viser vi at de ulike inntektsstatistikkene har ulike formål og bruker forskjellige inntektsenheter og inntektsbegreper. En følge av dette er at de ulike statistikkene ikke uten videre kan sammenliknes. Skjematisk har vi imidlertid forsøkt å vise hvordan de likevel henger sammen (diagram side 11).

I andre del av rapporten (kapittel 3, 4 og 5) benytter vi én av inntektsstatistikkene til å analysere inntektsfordelingen. Her har vi foretatt et valg. De inntektsdata vi mener gir best grunnlag for å beskrive inntekt som levekårskomponent, er de tre-årslige inntektsundersøkelsene. I disse nyttes bl.a. et disponibelt inntektsbegrep med husholdning som enhet. Ut fra andre målesettinger kunne vi ha gjort andre valg av data.

For å sammenlikne graden av ulikhet i to eller flere inntektsfordelinger, må man ha en del mål som beskriver ulikhet. I kapittel 3 tar vi opp en del av de vanlige statistiske mål for dette formål, og drøfter kort deres anvendelsesmulighet i inntektsfordelingssammenheng. Variasjonskoeffisienten, Lorentz-kurven og Gini-koeffisienten er her sentrale mål.

En konkret drøftelse av inntektsfordelinger og ulikhet belyst ved Lorentz-kurven og Gini-koeffisienten er gjort i kapittel 4. Datagrunnlaget er Inntektsstatistikk 1973, 1976 og 1979. Vi tar for oss kun totaler og forsøker ikke å gå inn på fordelinger etter forskjellige sosioøkonomiske kjennetegn. Prinsippene skissert i kapitlet kan imidlertid også direkte nyttes for disse. Konklusjonene i kapitlet er at husholdningers inntektsfordeling er jevnere enn inntektstakeres inntektsfordeling. Videre er husholdningens disponible likningsinntektsfordeling jevnere enn husholdningens inntektsfordeling der nettoinntekt pluss særfradrag nyttes som inntektsbegrep. Uansett hvilket inntektsbegrep og hvilken inntektstakerenhet vi nytter finner vi ikke at det har skjedd en utjevning av fordelingene i perioden 1973 - 1979.

I kapittel 5 har vi forsøkt å beskrive inntektsfordelinger ved hjelp av parametriske fordelingsfunksjoner. Datagrunnlaget her er Inntektsstatistikk 1976. Vi benytter her både den lognormale fordelingen og gammafordelingen. Gammafordelingen benyttes til analyse av inntektsfordelingen. For å få til disse estimeringene med enkle metoder på et så stort datamateriale (ca. 15 000 enheter), har vi avkortet fordelingen slik at den 4-5 prosent øverste del av fordelingen (de med høyest inntekt) er holdt utenfor ved estimeringen. Dette skulle imidlertid ikke ha særlig stor innflytelse på resultatene. Konklusjonene i dette kapitlet samsvarer i store trekk med de i kapittel 4.

## 2. INNTEKTSBEGREPER I STATISTISK SENTRALBYRÅ

### 2.1. Innledning

I dette kapitlet tar vi for oss de enkelte typer inntektsstatistikk som utarbeides i Byrået og peker på omfang og begreper i disse. Vi vil forsøke å vise hvor de ulike statistikkene stemmer overens og hvor de adskiller seg fra hverandre.

### 2.2. Lønnstatistikk

Selve lønnsfastsettelsen i arbeidslivet skjer ikke ved en enkel tilpasning på arbeidsmarkedet, men ved forhandlinger mellom arbeidstakerorganisasjoner og arbeidsgiverorganisasjoner. I dag utarbeides det en rekke lønnstatistikker uavhengig av hverandre med oppgave å gi bakgrunn for de regelmessige lønnsforhandlinger som føres mellom disse organisasjonene. Disse statistikkene bygger på skjemaoppgaver fra bedrifter eller på administrative data. Enkelte statistikker bygger på utvalg og andre på tellinger. Enhetene i statistikkene er lønnstakere. Generelt er visse grupper av lønnstakere holdt utenfor, f.eks. eiere og eieres familie.

Når en skal måle belønningen for økonomisk innsats, må en trekke inn arbeidstiden. Her har man flere tidsmål å velge mellom; lønn pr. time, uke, måned eller år er det man vanligvis nytter. Innenfor hvert av disse fire tidsmålene er det rom for variasjoner og forskjellige definisjoner. Hva skal en f.eks. regne som et ukeverk; 30, 35 eller 40 timer? Tilsvarende spørsmål kan en stille om månedsverk og årsverk. Hvor går skillet mellom heltids- og deltidsansatte?



Hvilke lønnsposter som skal telle med er et annet spørsmål. Skal overtid og andre variable og/eller faste tillegg tas med? Felles for alle lønnsstatistikkene er at de bruker bruttoinntektsbegreper. Tallene er gitt før fradrag for skatt, pensjonsinnskott og før inntektstakeren har fått gjort sine lov-messige fradrag i likningsøyemed.

Når slike spørsmål er avklart er de definisjoner en har vedtatt avgjørende for om én statistikk vil bli sammenliknbar med en annen. Sammenliknbarhet er da også ofte et argument i avveiningen av hvilke begrep som skal benyttes i en statistikk.

### 2.3. Nasjonalregnskapets årslønnstall

Ved hjelp av nasjonalregnskapet kan en beregne årslønnstall, definert som bedriftenes lønnskostnader dividert på antall utførte årsverk. Tallene kan gis etter næring, og angir utviklingen av lønnskostnadene for alle norske bedrifter. De totale kostnadene i bedriftene, heriblant lønnskostnadene, benyttes ved beregningen av driftsresultatet i hver aktuell næring. Driftsresultatet danner bakgrunn for å beregne den funksjonelle inntektsfordelingen i Norge.

### 2.4. Skattestatistikk

Skattestatistikken er en årlig statistikk som bygger på Skattedirektoratets skattebånd. Statistikken gir tall for den inntekt og formue som ligger til grunn for skatteberegning og tall for utliknet skatt. Enheten er skatteyster, som skatteyster ved kommune-/statskattelikningen er regnet person, selskap eller annet skattesubjekt som er ilagt formues- eller inntektsskatt ved kommune-/statskattelikningen eller person som har opptjent inntekt ved sjømannskatteordningen.

Skattestatistikken skiller mellom personlig og ikke-personlig skatteyster. Personlig skatteyster kan omfatte flere personer. For eksempel er fellesliknet ektepar regnet som én skatteyster.

I skattestatistikken står tre inntektsbegreper sentralt. Det er nettoinntekt pluss særfradrag ved statskattelikningen, nettoinntekt pluss særfradrag ved kommuneskattelikningen og pensjonsgivende inntekt. Inntektsbegrepene er utvidet i forhold til lønnsstatistikken inntektsbegrep. Næringsinntekt og kapitalinntekter er kommet i tillegg og alle likningsmessige fradragposter til fradrag ved beregning av nettoinntekt pluss særfradrag ved stats- og kommuneskattelikningen, og næringsinntekten er kommet i tillegg ved beregning av pensjonsgivende inntekt.

Skattestatistikken kan teoretisk kobles til personregisteret. Dermed er det mulig å etablere husholdning som enhet i statistikken. Denne mulighet er imidlertid ennå ikke utnyttet.

### 2.5. Inntektsstatistikk

Inntektsundersøkelsene er utvalgsundersøkelser som gjøres hvert tredje år. Et av formålene med undersøkelsene er å skaffe oversikt over fordelingen av personer og husholdninger etter inntektens størrelse og andre kjennemerker, samt å gi en oversikt over sammensetningen av inntekter til ulike grupper.

Grunnlaget for inntektsstatistikken er hovedsaklig selvangivelsesopplysninger hentet fra skattelikningen for et utvalg av personlige inntektstakere. Personlig inntektstakere er alle personer som ble behandlet som selvstendig enhet ved skattelikningen det året undersøkelsen gjelder for. Inntektsstatistikken opererer i tillegg også med en husholdningsenhet.

Inntektsstatistikken har konsentrert seg om to inntektsbegreper. I likhet med skattestatistikken opereres her med nettoinntekt pluss særfradrag. Statistikken har imidlertid i tillegg oppgaver over de fleste inntekts- og fradragposter i likningsøyemed slik at man har kunnet lage seg et begrep som egner seg bedre i levekårssammenheng. Disponibel likningsinntekt bygger på nettoinntekt pluss særfradrag, men er korrigert for skattefrie overføringer fra det offentlige og overføringer til det offentlige. Oppgave over barnetrygd, bostøtte fra husbanken og stipendier fra Statens Lånekassee for Utdanning er innhentet i tillegg til likningsoppgavene. I prinsippet kan man lage seg mange inntektsbegreper med forskjellig innhold og beregne konsekvenser for det offentliges direkte skatteinntekter av å nytte disse begreper ved likningen.

## 2.6. Andre inntektsstatistikker

Med bakgrunn i skattestatistikkens materiale lages det en del inntektsstatistikker for ulike yrkesgrupper. Vi har inntektsstatistikk for leger og tannleger. Her kobles skattestatistikken mot helsepersonellregisteret. En annen statistikk som lages med basis i skattestatistikken er "Bøndenes inntekt og formue".

Det lages også inntektsstatistikk på oppdrag fra spesielle grupper. Nylig har en utarbeidd slike statistikker for privatpraktiserende tannleger og kunstnere. Det som kjennetegner disse statistikkene er at enheten og inntektsbegrepet tilpasses anvendelsesområdet for statistikken. Inntektsbegrepet er derfor ofte ulikt det som benyttes i Byråets egne inntektsstatistikker.

## 2.7. En sammenlikning av de ulike inntektsstatistikkene

Som vi ser, er det særlig to kjennemerker som skiller de ulike inntektsstatistikkene fra hverandre. Det er inntektsbegrepet og enheten i statistikkene.

Lønnsstatistikkene utarbeides for enkelte grupper av lønnstakere og gir bruttolønns tall for lønnstakerenheten. Statistikken er tilpasset behovet til partene i arbeidslivet. De egner seg derimot dårlig for levekårsbetraktninger.

Skattestatistikken gir summariske inntekts- og skattetall og individuelle tall for grupper av skatteyterenheter (fordelingstall). Skattestatistikken egner seg derfor til å belyse inntektsfordelingen for skatteyterenheter. Den er samtidig et utgangspunkt for analyse og prognosearbeid i det offentliges budsjettssammenheng. Skattestatistikken gir grunnlag for å beregne totaltall i inntektsstatistikken.

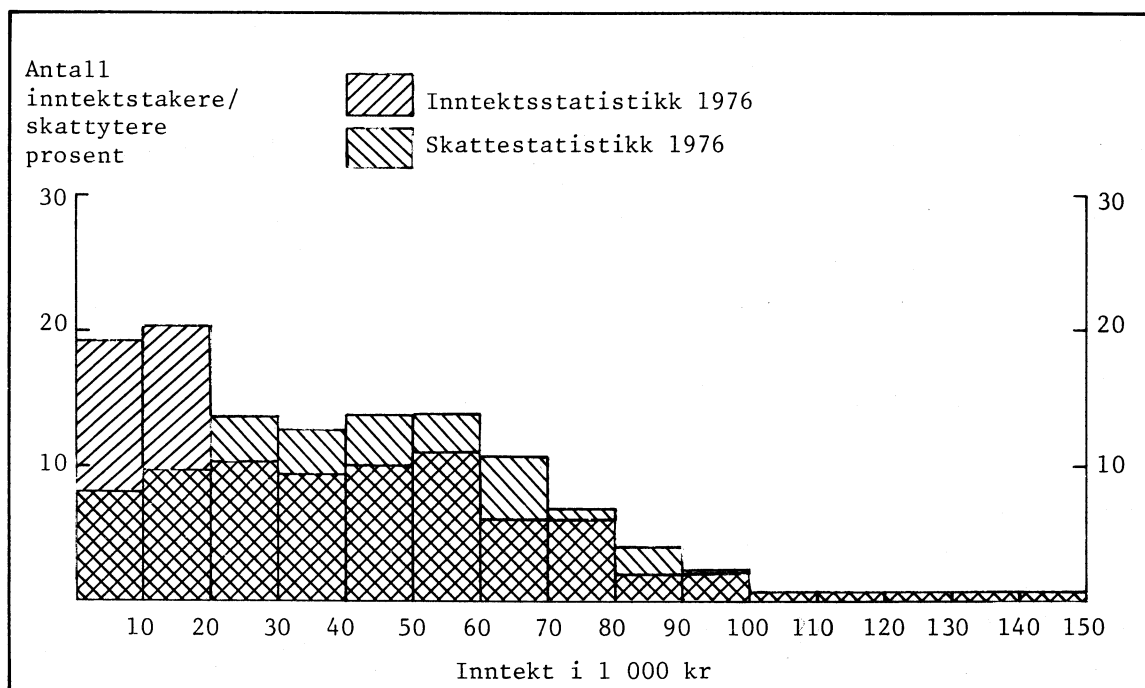
Inntektsstatistikken gir i tillegg til totaltall også sammensetningen av inntekten og fradragene. Dette gir et grunnlag for i skattemessig øyemed å beregne alternative inntektsbegreper og virkninger av dette på fordelingen og skatteinntektene. Videre har man i inntektsstatistikken muligheten til å vurdere inntektsfordelingen mellom husholdningene som er en interessant enhet i velferdssammenheng. Inntektsstatistikken opererer videre med et disponibelt inntektsbegrep som gjør det mulig å vurdere om de direkte skattene og stønadene som der er trukket inn, virker utjevne på fordelingen.

Vi skal se litt nærmere på de to siste statistikkene. Figur 2.1 viser hvor mange prosent av henholdsvis skatteytere og inntektstakere som har nettoinntekt pluss særfradrag på 0-10 000 kroner, 10-20 000 kroner osv. i henholdsvis skattestatistikken og inntektsstatistikken. En av forskjellene på de to statistikkene er definisjon av enheten. En stor del av de som har lav inntekt omfattes ikke av definisjonen skatteyter. Dette vil gjøre at fordelingen for inntektstakere er tyngre enn fordelingen for skatteytere ved lave inntekter. Sarme effekten ligger det i at skatteyterenheter kan omfatte to inntektstakere (fellesliknede ektefeller). Totaleffekten blir derfor at inntektsstatistikken viser større andel enheter med lave inntekter enn skattestatistikken.

I inntektsstatistikken går en videre i retning av å definere inntekt som en levekårskomponent. Enheten utvides til å omfatte hele husholdningen. Det har opplagt betydning for levekårene hvor mange inntekten skal brukes på. Det offentlige søker å ta hensyn til husholdningens sammensetning og hvert enkelt husholdningsmedlems inntekt i sin skatte- og stønadspolitikk. Inntektsstatistikken har oversikt over husholdningenes samlede disponibel likningsinntekt. Praktiske problemer har begrenset omfanget av de overføringer som inneholdes i dette begrep.

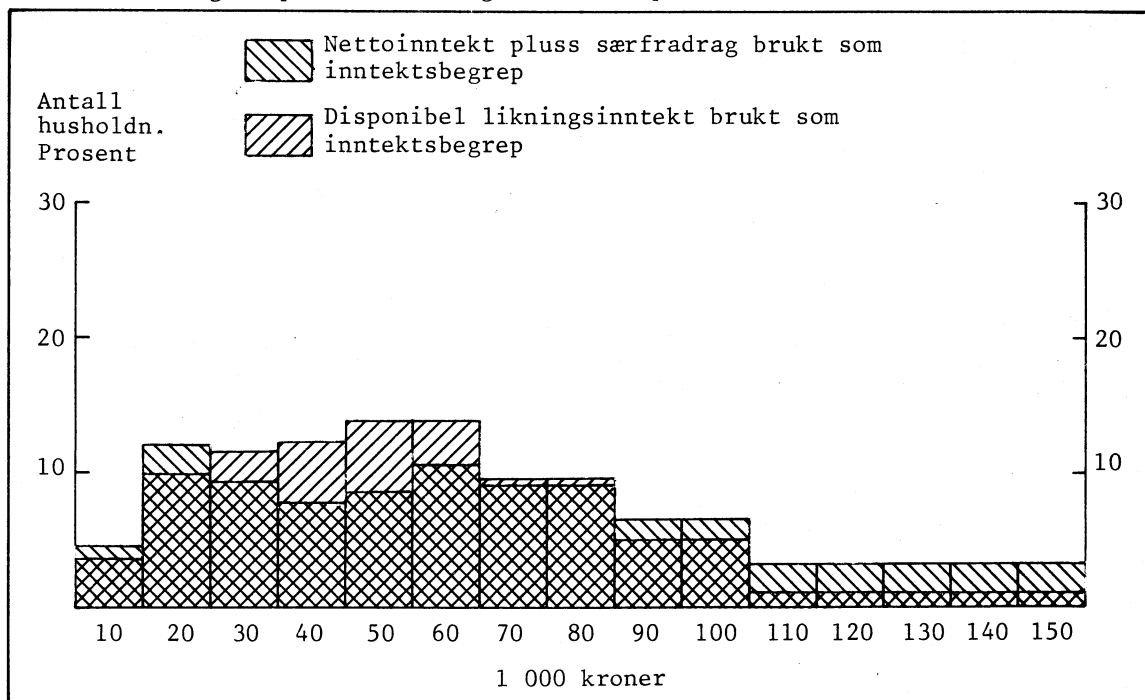
I figur 2.2 viser vi den relative inntektsfordelingen for husholdninger når vi benytter nettoinntekt pluss særfradrag og disponibel likningsinntekt. Det progressive skattesystem og stønadspolitikken gjør at fordelingen for husholdningens disponible likningsinntekt var forskjøvet i forhold til fordeling for husholdningers nettoinntekt pluss særfradrag, slik at vi fikk en "jevne" fordeling.

Figur 2.1 Inntektsfordelingen for inntektstakere og personlige skattytere i prosent. Nettoinntekt pluss særfradrag. 1976



K i l d e : Inntektsstatistikk 1976 og Skattestatistikk 1976

Figur 2.2 Inntektsfordeling for husholdningers nettoinntekt pluss særfradrag og disponibel likningsinntekt i prosent. 1976



K i l d e : Inntektsstatistikk 1976.

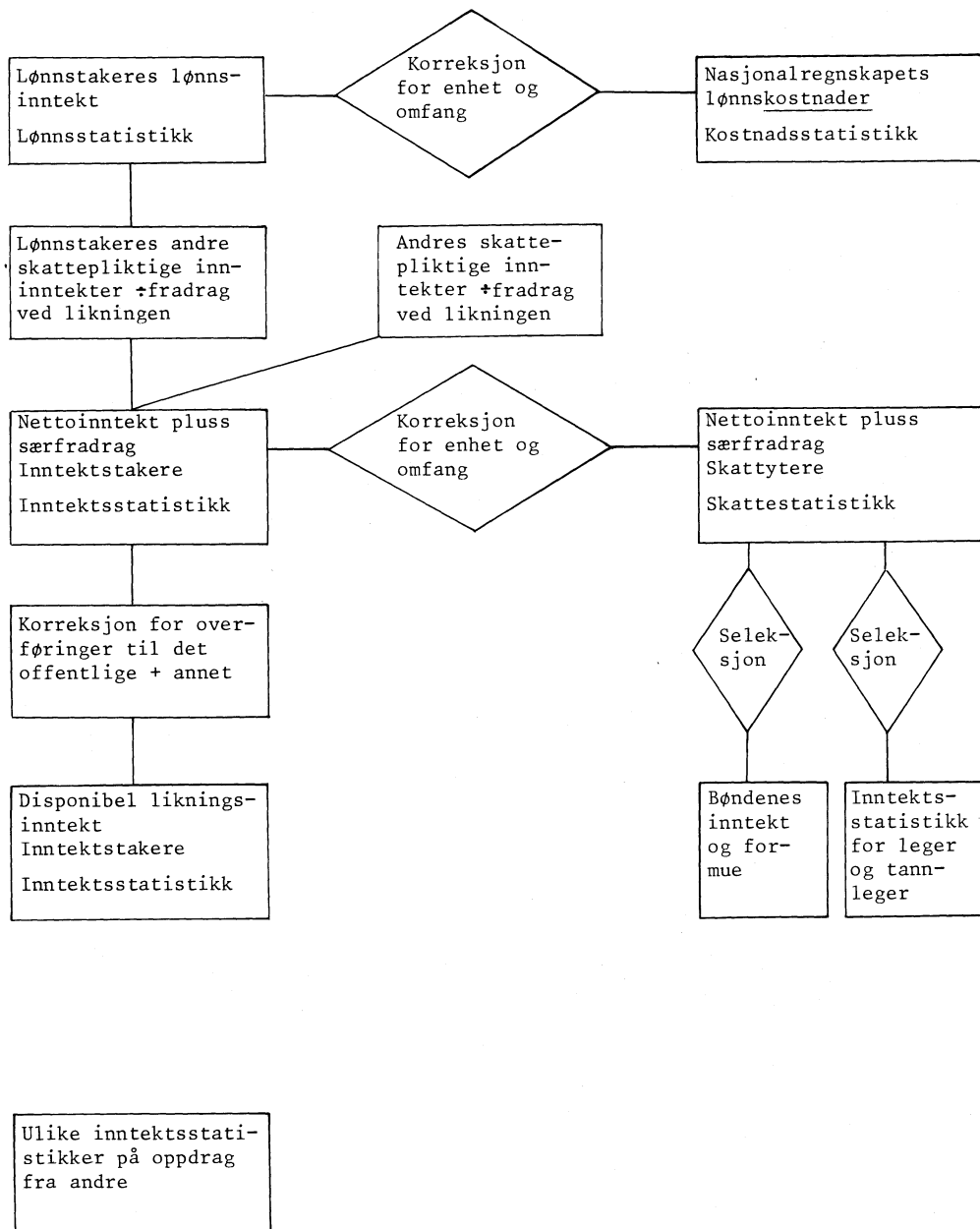
De statistikkene som er nevnt har alle sine spesielle formål. I inntektsfordelingsdiskusjoner kan alle statistikkene benyttes. Kommer man imidlertid over i mere velferdsbetraktninger er det naturlig å velge enten skattestatistikken eller inntektsstatistikken. Inntektsstatistikken har den fordel at den opererer med en husholdningsenhet og mulighet for alternative definisjoner av inntekt. Av denne årsak har det vært naturlig å velge nettopp inntektsstatistikken som grunnlag for de senere kapitler i denne rapporten der fordelingsproblemer tas opp.

## 2.8. En framstilling av forholdet mellom de ulike inntektsstatistikkene

I diagrammet nedenfor viser vi sammenhengen mellom inntektsstatistikkene.

Figur 2.8 En framstilling av forholdet mellom de ulike inntektsstatistikkene

I deagrammet nedenfor viser vi sammenhengen mellom inntektsstatistikkene.



### 3. MÅL TIL BESKRIVELSE AV INNTEKTSFORDELINGEN

#### 3.1. Innledning

Den mest alminnelige bruken av begrepet inntekt er som mål på velferd. I velferdssammenheng er det derfor interessant å se hvorledes inntekten fordeler seg i samfunnet, og også å registrere endringer i inntektsfordelingen over tid.

Ved en slik analyse må vi gjøre to ting:

- 1) Vi må definere inntekt
- 2) Vi må bestemme hvilken enhet i populasjonen inntekten skal knyttes til

I kapittel 1 har vi pekt på at vi har ulike valgmuligheter alt ettersom hvilke mål vi velger. Uansett hva vi velger vil vi anta at det til enhver inntektsenhet er knyttet én inntekt.

La oss anta at vi har en inntektsfordeling  $F$ . Vi lar  $a_1, a_2, \dots, a_n$  være et sett uavhengige observasjoner fra  $F$ . Da kan vi definere estimatoren:

$$(3.1.1) \hat{F}(x) = \frac{(\text{Antall } a_i \leq x)}{n}$$

$\hat{F}(x)$  er en konsistent estimator for den kumulative fordelingsfunksjonen  $F(x)$ . Vi betegner den tilhørende sannsynlighetstetthet ved  $f(x)$ .

Som vi tidligere har pekt på i kapittel 2 vil vi få ulike inntektsfordelinger alt etter hvilken inntektstakerenhet og hvilket inntektsbegrep vi bestemmer oss for å benytte. I utgangspunktet er derfor  $F(x)$  ukjent.

I inntektsstatistikken skaffer vi oss et utvalg av inntektsdata fra populasjonen. Vi vil i et senere kapittel undersøke om disse dataene indikerer at fordelingen  $F(x)$  er med i en kjent klasse av fordelinger.

#### 3.2. Mål for beliggenhet av inntektsfordelingen

Når beliggenheten av en inntektsfordeling skal karakteriseres er det særlig to størrelser som nyttes. Den ene er forventet gjennomsnittlig inntekt  $\mu$  der  $\mu$  er definert ved

$$(3.2.1) \mu = \int_0^{\infty} x f(x) dx$$

Den andre er medianinntekten, la oss kalle den  $a$ . Den er definert ved:

$$(3.2.2) F(a) = \frac{1}{2}$$

Populært vil vi si at halvparten av inntektstakerne har en inntekt mindre enn  $a$  og halvparten større. Gode estimatører for disse størrelsene vil være det aritmesiske gjennomsnittet og medianen blant observasjonene  $a_1, \dots, a_n$ . Vi definerer derfor

$$(3.2.3) \hat{\mu} = \frac{1}{n} \sum_{i=1}^n a_i \quad \text{og}$$

$$(3.2.4) \hat{a} = \hat{F}^{-1}\left(\frac{1}{2}\right)$$

Disse estimatene er konsistente for henholdsvis  $\mu$  og  $F^{-1}\left(\frac{1}{2}\right)$ .

En tredje størrelse som karakteriserer en fordeling er fordelings mode (typetall). Hvis inntektsfordelingen er enpuklet ut fra inntektsbegrepet og inntektstakerenheten vi velger, vil moden fortelle oss hvilken inntekt som forekommer oftest (vi kan f.eks. måle i hele 1 000 kr), og hvis inntektsfordelingen er enpuklet, si noe om hvor tyngden i fordelingen ligger.

### 3.3. Mål for spredning

Når spredningen i en inntektsfordeling skal angis nyttes ofte varians som mål. Denne sier noe om gjennomsnittlig spredning fra gjennomsnittet. Variansen  $\sigma^2$  er definert ved:

$$(3.3.1) \sigma^2 = \int_0^{\infty} (x-\mu)^2 \cdot f(x) dx$$

Alternativt kan vi benytte standardavviket  $\sigma$ . Ut fra observasjonene våre kan vi estimere en konsistent forventningsrett estimator for  $\sigma^2$

$$(3.3.2) \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (a_i - \hat{\mu})^2$$

Et annet spredningsmål som ofte benyttes er interkvartilområdet ( $c_{0,25}$ ,  $c_{0,75}$ ) der  $c_{0,25}$  og  $c_{0,75}$  er definert slik at

$$(3.3.3) F(c_{0,25}) = 0,25$$

$$F(c_{0,75}) = 0,75$$

Det vil si at halvparten av inntektstakerne har en inntekt mellom  $c_{0,25}$  og  $c_{0,75}$ . En fjerdepart av inntektstakerne har en inntekt under  $c_{0,25}$  og en fjerdepart har en inntekt over  $c_{0,75}$ .

Størrelsene  $\mu$ ,  $a$ ,  $\sigma^2$  og interkvartilområdet ( $c_{0,25}$ ,  $c_{0,75}$ ) er avhengig av hvilken måleskala som benyttes. En endring av måleskalaen vil derfor påvirke størrelsene av dem.

### 3.4. Sammenlikning av inntektsfordelinger

For å se om en inntektsfordeling er jevnere enn en annen kunne vi sammenlikne noen av størrelsene funnet i kapittel 3.3 og se på utviklingen av dem i ulike år. På grunn av inntektsutviklingen har imidlertid måleskalaen økt fra år til år. Vi kan derfor ikke se om det kun er fordelingsendringer som er bakgrunn for endringene i størrelsene  $\mu$ ,  $a$  eller  $\sigma^2$ .

I publikasjonen "Inntektsforholdene i Norge" NOU 1974:34 har en prøvd å stille opp betingelser for en endring i inntektsfordelingen ved å se på endringer i  $\bar{a}$  og  $\sigma_a^2$  simultant. Den sier at en tilstrekkelig betingelse for å påstå at en inntektsfordeling  $F_1(x)$  er jevnere enn annen  $F_2(x)$  er

$$\mu_1 > \mu_2 \text{ og samtidig } \sigma_1^2 < \sigma_2^2$$

Dette vil vi heretter kalle NOU's kriterium.

Ut fra inntektsstatistikken kan vi utlede følgende tabell. Vi benytter inntektstakerenheten husholdning.

Tabell 3.1. Estimert gjennomsnittsinntekt og standardavvik\* i ulike år

k	$\hat{\mu}_k$	$\hat{\sigma}_k$	$\hat{\mu}_k^{\text{korr}}$	$\hat{\sigma}_k^{\text{korr}}$
1973 .....	36 900	22 900	36 900	22 900
1976 .....	60 300	36 300	45 200	27 200
1979 .....	74 900	48 700	45 400	29 600

\*) Ved beregning av  $\hat{\mu}$  og  $\hat{\sigma}$  er noen få inntektstakere med høy inntekt holdt utenfor. Se forøvrig kap. 5.

Tabell 3.1 viser at hvis vi benytter oss av inntektsbegrepet nettoinntekt pluss særfradrag og bruker inntektstakerenheten husholdning så har  $\hat{\mu}$  økt fra 1973 til 1976 og fra 1976 til 1979. Det samme har  $\sigma^2$ . Derfor synes det rimelig å hevde at NOU's kriterium ikke er oppfylt.

Inflasjonen er blant årsakene til en inntektsøkning i 1970-årene. Dette kan vi se på som at måleskalaen for inntekt har endret seg. La oss derfor studere om NOU's kriterium er oppfylt, om vi istedenfor nominelle inntektstall benytter oss av realinntekt. Fremdeles nytter vi inntektsbegrepet nettoinntekt pluss særfradrag, men vi deflaterer inntektstallene fra 1976 og 1979 med konsumprisindeksen (1973 som basisår). Dette er en vanlig metode. De indeksskorrigerte verdiene for  $\hat{\mu}_k$  og  $\hat{\sigma}_k$  vil vi benevne  $\hat{\mu}_k^{\text{korr}}$  og  $\hat{\sigma}_k^{\text{korr}}$ .

La  $\gamma_i$  være konsumprisindeksen i år. Omregnet til 1973=100 har vi da

$$\begin{aligned} \gamma_{73} &= 100 \\ (3.4.1) \quad \gamma_{76} &= 133,3 \\ \gamma_{79} &= 164,8 \end{aligned}$$

Estimatorene  $\hat{\mu}_k^{\text{korr}}$  og  $\hat{\sigma}_k^{\text{korr}}$  er beregnet i tabell 2.1. Det synes også her rimelig å hevde at NOU's kriterium ikke er oppfylt.

Vi har også sett på fordelingen av disponibel likningsinntekt på samme vis som vi gjorde for nettoinntekt pluss særfradrag. Både da vi brukte nominell disponibel likningsinntekt og "real" disponibel likningsinntekt (se tabell 3.2) virket det rimelig å hevde at NOU's kriterium ikke var oppfylt når vi ser på inntektsutviklingen fra 1973 til 1979.

Tabell 3.2. Estimert gjennomsnittlig disponibel likningsinntekt og standardavvik\* i ulike år

k	$\hat{\mu}_k$	$\hat{\sigma}_k$	$\hat{\mu}_k^{\text{korr}}$	$\hat{\mu}_k^{\text{korr}}$
1973 .....	32 100	17 900	32 100	17 900
1976 .....	52 200	28 300	39 200	21 200
1979 .....	63 700	36 800	38 700	22 300

\*) Se note side 15.

NOU stilte opp en tilstrekkelig betingelse for å kunne påstå at en inntektsfordeling er jevnere enn en annen. For å finne nødvendige betingelser må vi lete etter andre mål.

### 3.5. Variasjonskoeffisienten

Ved sammenlikninger av inntektsfordelinger er en ofte interessert i mål som er invariante overfor de endringene i fordelingen som skyldes endringer av skala. Variasjonskoeffisienten har denne egenskapen. Den er definert ved

$$(3.5.1) \quad V = \frac{\sigma}{\mu}$$

Ut fra inntektsstatistikken kan vi beregne en estimator for variasjonskoeffisienten ved å benytte de ulike estimatorene for gjennomsnitt og standardavvik som tidligere er funnet.

Estimatoren blir

$$(3.5.2) \quad \hat{V} = \frac{\hat{\sigma}}{\hat{\mu}}$$

Vi har tatt med estimatorene for variasjonskoeffisientene for inntektsfordelingen i årene 1973, 1976 og 1979 for de inntektsbegrep som er benyttet i inntektsstatistikken i tabell 3.3.

Tabell 3.3. Variasjonskoeffisienten i ulike år

År inntektsbegrep K	1973	1976	1979
Nettoinntekt osv. ....	0,619	0,602	0,650
Disponibel likningsinntekt	0,559	0,542	0,578

Vi ser at variasjonskoeffisienten har steget både for nettoinntekt og disponibel likningsinntekt fra 1973 til 1979. Det vil si at spredningen relativt har økt mer enn gjennomsnittet. Dette kan tas som en indikator for at inntektsfordelingen er blitt mindre jevn i løpet av 70-åra.

Vi vil ikke her ta standpunkt til om endringen er signifikant, men bare konstatere at vi har fått en økning i varianskoeffisienten.

### 3.6. Fraktiler

Median og kvartiler deler en fordeling i like store deler og gir oss dermed et visst grunnlag for å vurdere spredningen i materialet. Imidlertid sier dette ikke noe om spredningen innen delene. Vi kan nå forfølge prinsippet med å dele fordeling og dermed få f.eks. desiler (10 deler) eller percentiler (100 deler). I prinsippet kan vi dele opp i et vilkårlig antall deler. Desiler er definert slik:

$$(3.6.1) F(c_1) = \frac{1}{10}$$

$$F(c_2) = \frac{2}{10}$$

$$\vdots$$

$$F(c_9) = \frac{9}{10}$$

og for percentiler:

$$(3.6.2) F(c_1) = \frac{1}{100}$$

$$F(c_2) = \frac{2}{100}$$

$$\vdots$$

$$F(c_{99}) = \frac{99}{100}$$

og generelt

$$(3.6.3) F(c_i) = \frac{i}{n} \quad i = 1, \dots, n-1$$

der  $n$  er antall deler vi ønsker å dele fordeling opp i.

Fellesbetegnelsen på median, kvartil, desil, percentil etc. er fraktiler. De mest benyttede fraktiler i inntektsundersøkelsen er median og desiler.

I undersøkelser med mange inntektstakere viser det seg ofte at medianinntekten ligger noe under gjennomsnittsinntekten. I de seneste inntektsundersøkelsene er estimert medianinntekt for husholdningene 10-15 prosent lavere enn estimert forventet inntekt. For inntekttakere er estimert medianinntekt 15-20 prosent lavere enn estimert forventet inntekt.

Denne forskjellen på medianinntekt og gjennomsnittsinntekt innebærer at inntektsfordelingen ikke er symmetrisk omkring gjennomsnittet, men at den er skjev til høyre. Denne forskjellen er vist i tabell 3.4 og tabell 3.5



Tabell 3.4. Inntekt - inntektstakere (realinntekt i parentes, basisår 1973\*)

	1973	1976	1979
Gjennomsnitt	25 900	36 400 (27 300)	46 600 (28 300)
Median .....	21 500	29 700 (22 300)	39 300 (23 800)

\*) Indeks se side 16.

Tabell 3.5. Inntekt husholdninger (realinntekt i parentes, basisår 1973\*)

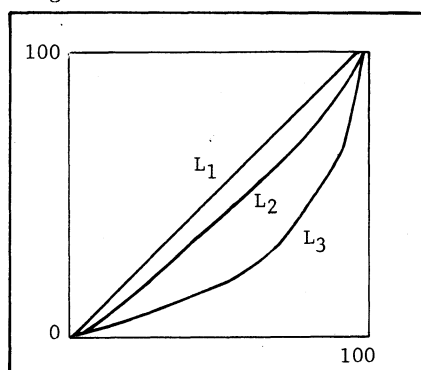
	1973	1976	1979
Gjennomsnitt	40 300	66 400 (49 800)	80 900 (49 100)
Median .....	35 500	60 000 (45 000)	70 900 (43 000)

\*) Indeks se side 16.

### 3.7. Lorentz-kurver

Når vi har rangert inntektstakerne etter inntektens størrelse har vi en mulighet for å regne ut hvor mange prosent av den totale inntekt en vilkårlig prosent av inntektstakerne med minst inntekt har. I prinsippet kan vi dele opp de rangerte inntektstakerne i et stort endelig antall grupper. En kontinuerlig kurve som viser sammenhengen mellom prosent av de inntektstakerne som har lavest inntekt og deres prosentandel av den totale inntekten, kalles Lorentz-kurve. Vi vil senere få vise at vi ikke trenger å dele opp i mer enn 10 grupper (desiler) for å få en god tilnærming til en kontinuerlig kurve.

Figur 3.1 Lorentzkurve



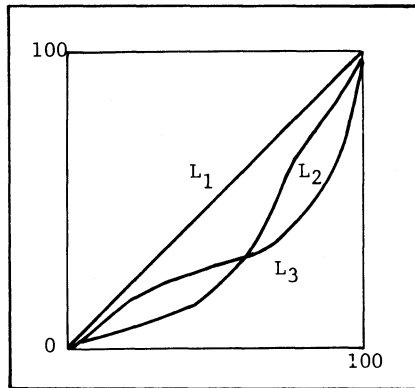
I figur 3.1 har vi tegnet inn to Lorentzkurver. Langs den vannrette aksene måler vi inntektstakernes andel av totalinntekt og langs den loddrette måles andel av antall inntektstakere. Diagonalen  $l_1$  er en kurve som viser at alle inntektstakerne har samme inntekt. Dette vil vi kalle en jevn fordeling.  $l_2$  er en kurve som sier at noen har lav inntekt og noen har høy inntekt. Dette vil vi si er en ujevn fordeling. Desto større avstand det er fra  $l_2$  til  $l_1$  desto mere ujevn vil vi si at fordelingen er.  $l_3$  er således en mer ujevn fordeling enn  $l_2$ .

Ved avgjørelsen om den ene eller andre kurven påviser den jevneste fordelingen må usikkerheten ved undersøkelsen bringes inn. Dette impliserer ofte skjønn.

### 3.8. Ginikoeffisienten

Anta at vi har to inntektsfordelinger representert ved hver sin Lorentzkurve. Hvis kurvene ikke skjærer hverandre, kan vi avgjøre hvilken som viser den jevneste fordelingen (jfr. figur 3.1). Hvis kurvene skjærer hverandre kan vi ikke finne direkte av figuren (figur 3.2) hvilken fordeling som er jevnest. En måte å avgjøre dette på da er å beregne Ginikoeffisienten.

Figur 3.2 Lorentzkurver som skjærer hverandre

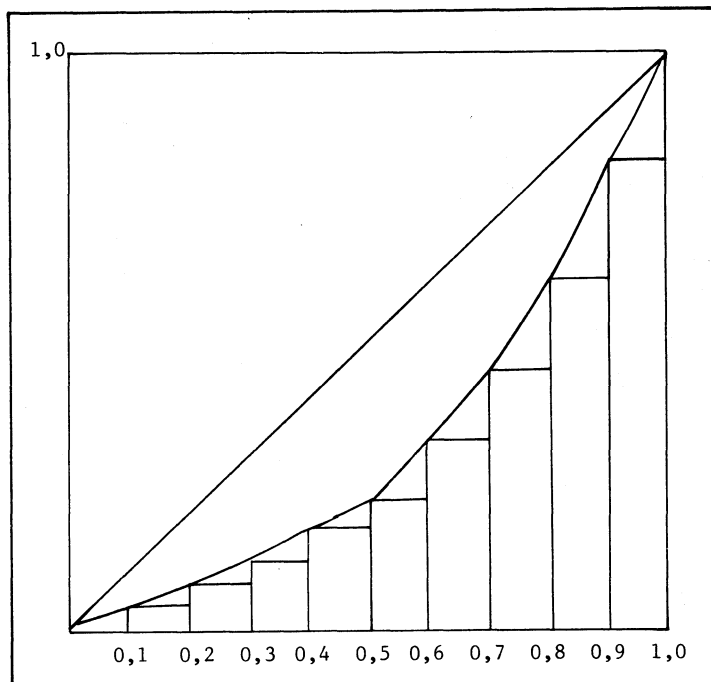


Ginikoeffisienten er definert som flateinnholdet av segmentet mellom Lorentzkurven og diagonalen dividert på flateinnholdet av trekanten under diagonalen.

Vi ser at den fordelingen som har den minste Ginikoeffisienten er den jevneste fordelingen etter våre definisjoner.

Ved beregning av Gini-koeffisienten burde vi kjent den teoretiske fordelingen for å kunne integrere arealet under kurvene. Denne kjenner vi imidlertid ikke. Vi skal nå vise at ved å benytte desiltabeller (bare 10 grupper) kan vi få en god tilnærming til den kontinuerlige kurven ved flateinnholdsregninger.

Figur 3.3 Lorentzkurve estimert ved bruk av desiler



Valg av enhet er her uten betydning. Vi regner i andeler. For å beregne arealet mellom kurven og diagonalen beregner vi først arealet under kurven for deretter å trekke dette fra arealet av hele trekanten under diagonalen.

Arealet under kurven vil etter at vi har laget 9 rektangler og 10 trekanter på grunnlag av desiltabellene, kunne beregnes.

Høyden i trekantene er henholdsvis  $x_1, x_2, \dots, x_{10}$

Høyden i rektanglene blir da:  $1 - x_{10}, 1 - (x_9 + x_{10}), \dots, 1 - \sum_{i=2}^{10} x_i$

Alle grunnlinjer i trekantene og rektanglene er lik 0,1.

Totalt flateinnhold av trekantene:

$$(3.8.1) A_1 = \frac{0,1 \cdot \sum_{i=1}^{10} x_i}{2} = 0,05$$

siden

$$\sum_{i=1}^{10} x_i = 1$$

Totalt flateinnhold av rektanglene:

$$(3.8.2) A_2 = 0,1 \sum_{i=1}^{10} (10-i)x_i = 1 - 0,1 \sum_{i=1}^{10} ix_i$$

En estimator for ginikoeffisienten kan derfor finnes ved.

$$(3.8.3) \hat{G} = \frac{0,5 - (A_1 + A_2)}{0,5} = 1 - 2(A_1 + A_2)$$

Den unøyaktigheten vi får i  $A_1$  og dermed i  $\hat{G}$  fordi vi benytter oss av desiler istedenfor den kontinuerlige estimerte fordelingsfunksjonen  $\hat{F}$  er neglisjerbar. Vi har

$$(3.8.4) \frac{A_1}{A_2} = \frac{0,1 \sum_{i=1}^{10} x_i}{2(1 - 0,1 \sum_{i=1}^{10} ix_i)}$$

Hvis nå:  $x_1 = x_2 = x_3 = \dots = x_{10} = 0,1$

Så har vi

$$(3.8.5) \frac{A_1}{A_2} = \frac{1}{20 - 0,2 \sum_{i=1}^{10} i} = \frac{1}{9} \Rightarrow A_1 = \frac{1}{9} A_2$$

Feilen vi gjør er en liten del av  $A_1$  og  $A_1$  er en liten del av  $A_2$ . Totalt er altså feilen neglisjerbar.

#### 4. INNTEKTSFORDELINGENS AVHENGIGHET AV INNTEKTSTAKERENHETEN OG INNTEKTSBEGREPET, BELYST VED LORENZ-KURVER OG GINI-KOEFFISIENTER

##### 4.1. Innledning

I dette kapittel ser vi på en del spørsmål angående enhet og begrepsdefinisjoner. Vi skal se at fordelingene er forskjellige alt etter hvilke begrep vi benytter. Til å belyse disse sider ved inntektsfordelingene, vil vi benytte Lorenz-kurver og Gini-koeffisienter beregnet på grunnlag av data fra Inntektsstatistikkene 1973, 1976 og 1979. Til slutt i kapitlet er Gini-koeffisientene og tallgrunnlaget for kurvene (akkumulerte deciler) plassert i tabell 4.1. Ginikoeffisientene blir kommentert sammen med kurvene. Lorentzkurvene vil gi en indikasjon på hvor "jevnheten" i to kurver er forskjellig, mens Gini-koeffisientene tar standpunkt til hele fordelingen samlet.

##### 4.2. Usikkerhet ved å benytte inntektsstatistikens data

Ved de fleste statistiske undersøkelser vil det være en eller annen form for usikkerhet ved tallene. Vi skal her kort nevne noen usikkerhetsfaktorer ved inntektsundersøkelsenes tallmateriale.

- i) Inntektsundersøkelsene er utvalgsundersøkelser og vil inneholde de vanlige usikkerhetsfaktorer ved slike undersøkelser når de benyttes til å si noe om fordelingen i hele universet (her alle inntektstakerne i Norge). Totaltallene vil imidlertid være gode, da de er funnet ved oppblåsing mot Skattestatistikens telling.
- ii) Inntektsopptjeningen slik den måles, inneholder ofte (for lønnstakere) akkord, overtid etc. Dette vil skape svingninger i inntektene selv om det ikke er strukturelle endringer. Dette gjør at en ikke uten videre kan betrakte inntektsforskjeller som uttrykk for lønnsinntektsforskjeller der bare fastlønn inngår.
- iii) I Inntektsundersøkelsene har en ikke direkte måleproblemer i forbindelse med innsamlingen, da den benytter administrative data. Vi har heller ikke frafallsproblemer, da svarprosenten er 100. Likevel må vi imidlertid kunne si at vi har tildels store måleproblemer. Inntektsundersøkelsene pretenderer å si noe om inntektsfordelingen i Norge. Vi skal imidlertid huske på at vi opererer med en legal økonomi og en svart økonomi. Den sistnevnte beregnes ikke og vil bli holdt utenfor her. Skattevesenet har også opplagt store måleproblemer. I vår undersøkelse må vi altså nøye oss med å snakke om fordelingen av likningsmessig oppgitt inntekt.

Ovenstående tilsier at vi må ha et visst utslag i våre målevariable før vi kan påstå at én fordeling er jevnere enn en annen. Ideelt burde vi ha avgjort på et teoretisk grunnlag hvor store utslag vi må ha i våre målevariable for å kunne trekke "sikre" konklusjoner. Dette har vi ikke forsøkt å gjøre her. Problemet er nærmere omtalt i kapittel 5, der vi forsøker å beskrive inntektsfordelingen med kjente fordelingsfunksjoner.

##### 4.3. Problemstillinger

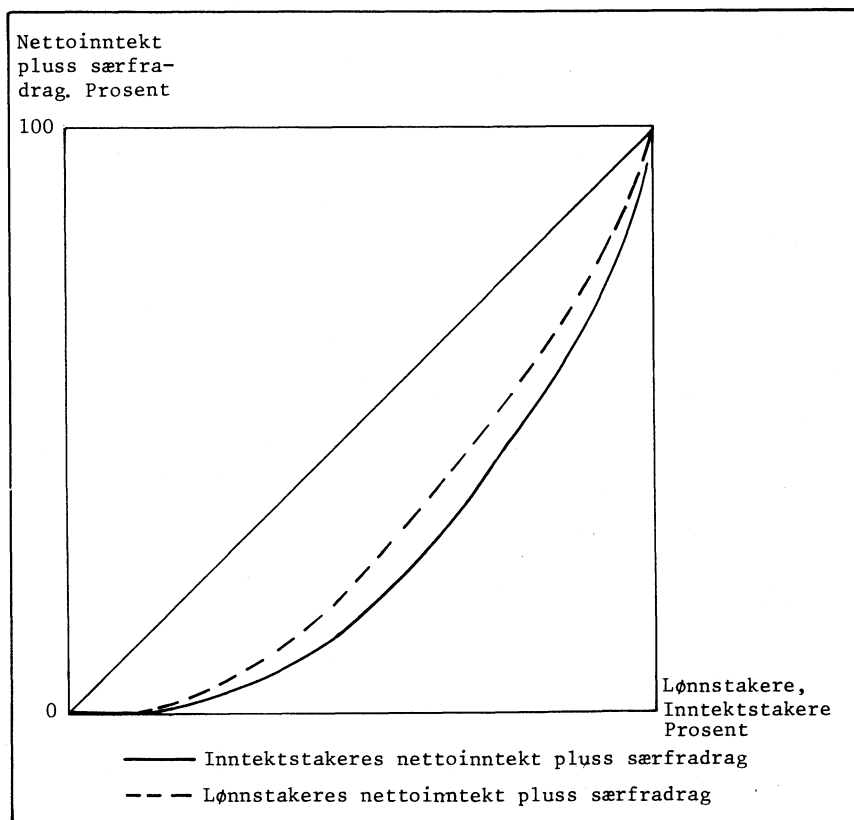
De fleste interesseorganisasjoner i arbeidslivet har i 1970-årene vært opptatt av lønnsutjevning. Våre Lorenz-kurver er ikke særlig godt egnet til å gi svar på om dette virkelig er skjedd (jfr. enhetsdefinisjonene og inntektsdefinisjonen i inntektsstatistikkene). Dette skyldes innholdet i inntektsbegrepet og mangel på homogenitet i inntektstakerenheten m.h.p. inntektsbegrepet (dvs. inntektstakerenheten omfatter lønnstakere, personlig næringsdrivende, pensjonister, trygdede, skoleelever etc. med forskjellige typer inntekt-lønnsinntekt, næringsinntekt, trygder etc.). Vi har lagt inn en kurve for lønnstakeres inntekt for å vise at denne ikke kan representeres av fordelingen for inntektstakernes inntekt.

Apriori antakelser om at høyinntektshusholdninger eller lavinntektshusholdninger har flest inntektstakere pr. husholdning, og dermed spekulasjoner om at dette virker til større skjevhet, henholdsvis mindre skjevhet i inntektsfordelingen, kan man forsøke å belyse ved analyser av Lorentzkurvene. Jfr. forøvrig påstanden om at både mannen og kona i husholdningen må arbeide hvis man skal klare seg. Inntektsfordelingen for inntektstakere og inntektsfordelingen for husholdninger benyttes da ved sammenlikningen.

Det offentlige skatte- og stønadspolitikk har som hovedformål å virke utjevne spesielt på husholdningenes inntekter. Inntektsstatistikens begrep disponibel likningsinntekt tar hensyn til den

direkte skattepolitikken og enkelte stønader. Ved å sammenlikne fordelingen over husholdningenes nettoinntekt pluss særfradrag og fordelingen over husholdningenes disponible likningsinntekt, vil man få en indikasjon på om denne politikken er "effektiv". "Effektiv" er her satt i gåseøyne da man for å vurdere dette skikkelig burde trekke inn et mål for hva man ønsker å oppnå og deretter måle resultatet i forhold til virkemiddelbruken og målet.

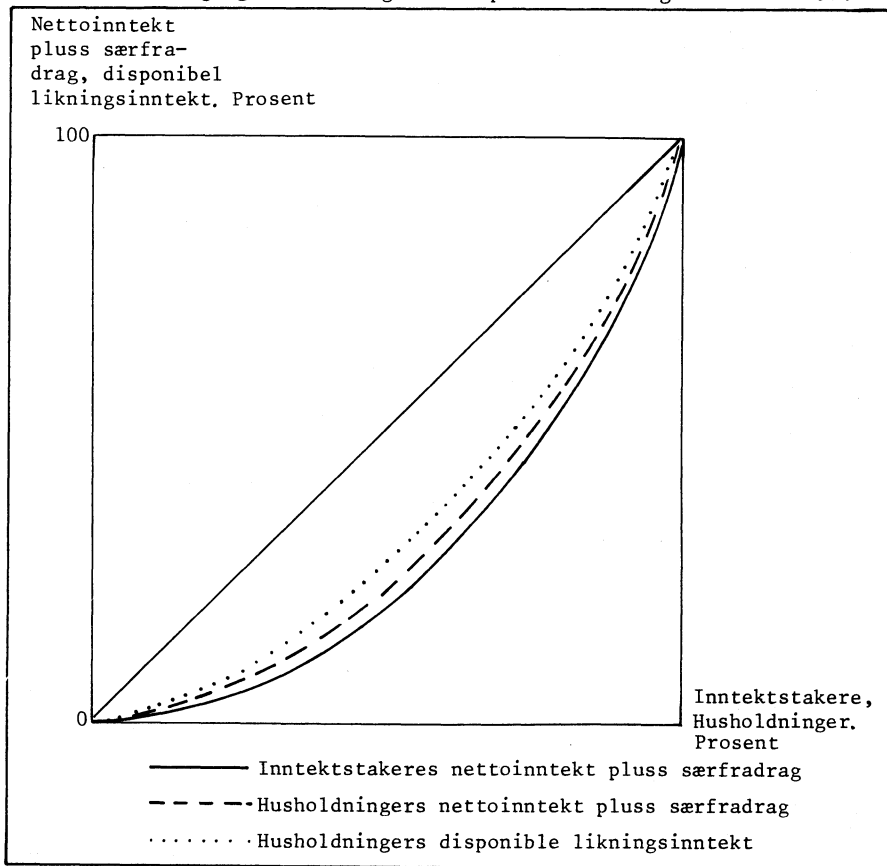
Figur 4.1 Lorenz- kurven og Gini-koeffisienten benyttet på inntektsstatistikdata for 1976



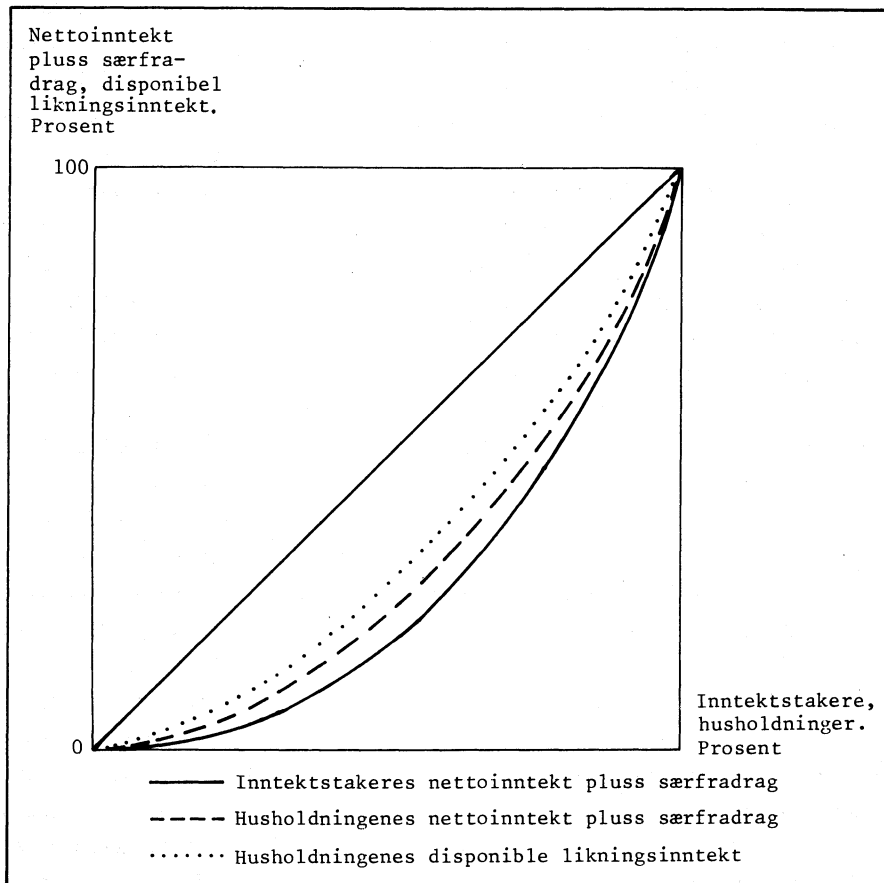
Figur 4.1 (jfr. tabell 4.1) viser Lorentzkurvene for fordelingene av lønnstakeres og inntektstakeres nettoinntekt pluss særfradrag 1976. Vi ser av figuren at fordelingene er forskjellige og at lønnstakeres inntektsfordeling er jevnere enn inntektstakernes. Gini-koeffisientene er på henholdsvis 0,363 og 0,443. Med vårt store utvalg (16 000 enheter) kan det vises at variansen på koeffisientene er meget liten. At Ginikoeffisienten er såvidt forskjellig, vil vi ta som et tegn på at lønnstakeres inntektsfordeling ikke uten videre kan representeres av inntektstakernes inntektsfordeling. Mange forhold kan være årsak til differansen mellom fordelingene. Inntektstakerenhetene omfatter pensjonister og trygdede samt skoleelever og vernepliktige foruten lønnstakere og selvstendige næringsdrivende. De førstnevnte er typisk lavinntektstakere og vil trekke i retning av en mindre jevn fordeling. De personlig næringsdrivende er på den annen side stort sett høyinntektstakere. Dette bidrar til å gjøre inntektsfordelingen mere ujevn i den andre retningen.

Allerede her illustreres betydningen av valg av enheter når en skal diskutere inntektsfordelinger. Vi vil nå forlate lønnstakerenheten og konsentrere oss om inntektstakerenheten og husholdningsenheten.

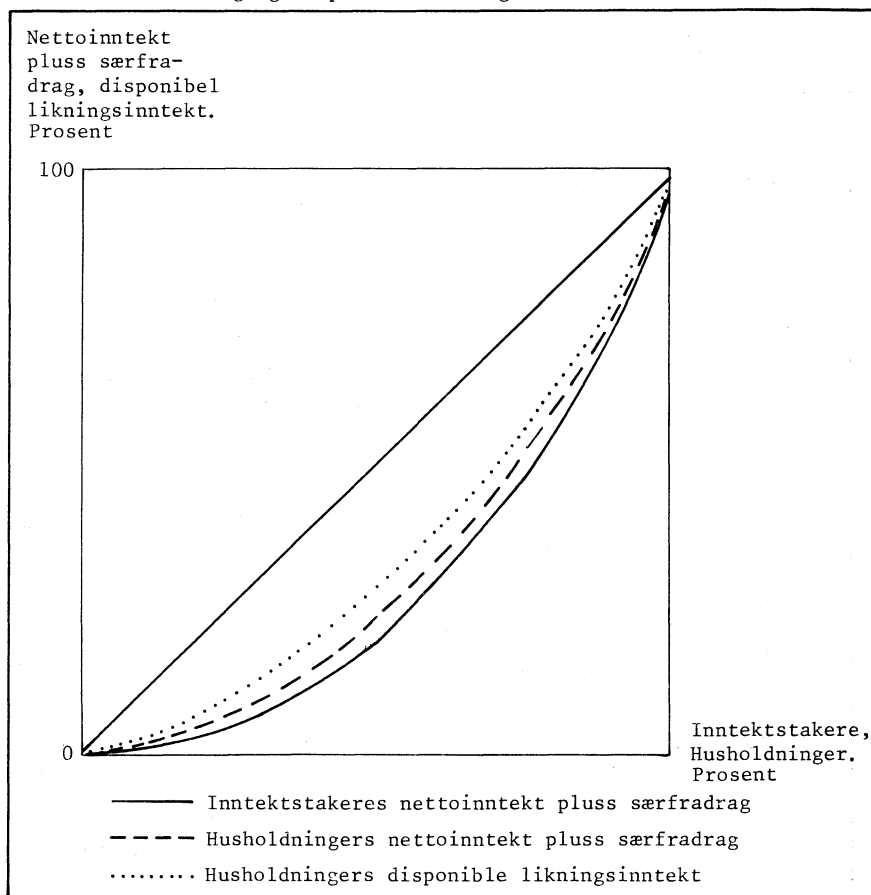
Figur 4.2 Lorenz-kurver over inntektstakeres nettoinntekt pluss særfradrag, husholdningenes nettoinntekt pluss særfradrag og husholdningenes disponible likningsinntekt. 1973



Figur 4.3 Lorenz-kurver over Inntektstakeres nettoinntekt pluss særfradrag og husholdningers nettoinntekt pluss særfradrag og disponible likningsinntekt. 1976



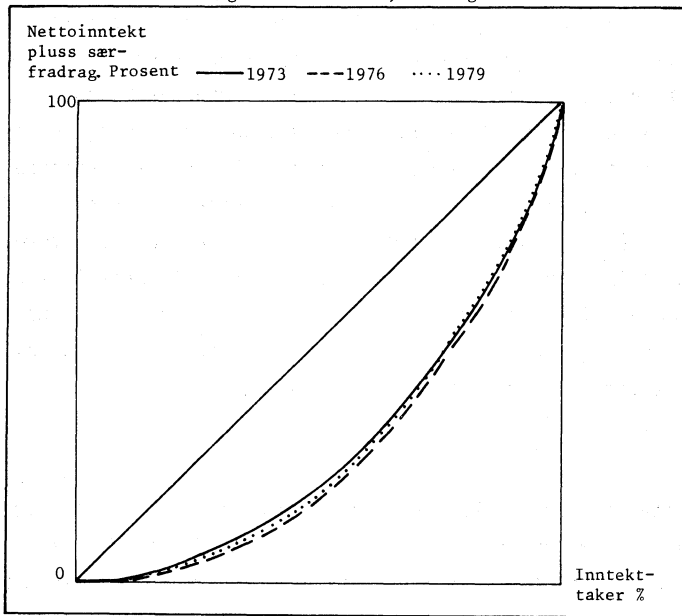
Figur 4.4 Lorentz-kurver over inntektstakeres nettoinntekt pluss særfradrag og husholdningers nettoinntekt pluss særfradrag og disponibel likningsinntekt. 1979



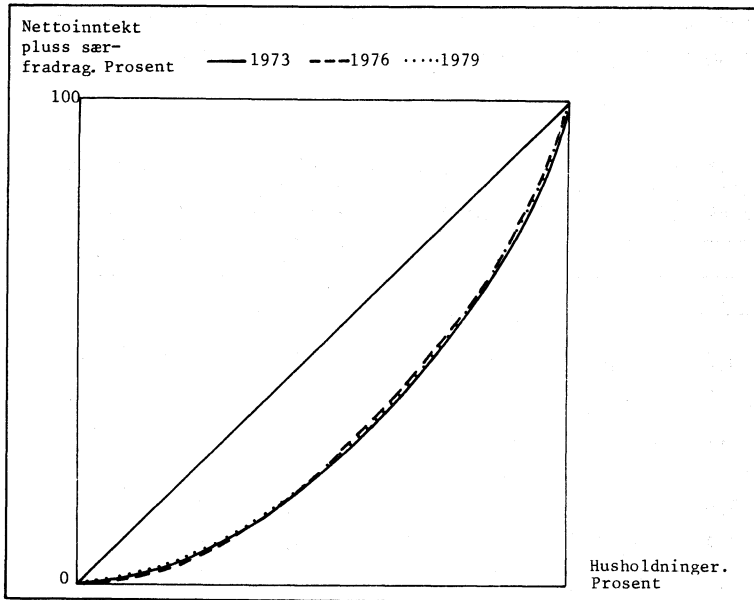
Figur 4.2 - 4.4 (jfr. tabell 4.1) viser fordelingen for inntektstakeres nettoinntekt pluss særfradrag, husholdningenes nettoinntekt pluss særfradrag og husholdningenes disponible likningsinntekter for årene henholdsvis 1973, 1976 og 1979. Vi ser at fordelingen for alle år blir jevnere når vi går over fra å betrakte inntektstakernes til å betrakte husholdningenes fordeling m.h.p. nettoinntekt pluss særfradrag. Gini-koeffisientene er her henholdsvis 0,415 og 0,373 (-73), 0,443 og 0,365 (-76) og, 0,421 og 0,369 (-79). Kombinasjonen av Gini-koeffisientutslagene størrelse og det at tre påfølgende undersøkelser viser samme tendens, tyder på at husholdningenes inntektsfordeling er jevnere enn inntektstakernes inntektsfordeling m.h.p. nevnte inntektsbegrep. En forklaringsfaktor her kan være at i lavinntektsfamilier må flere i husholdningen ut i arbeid simpelthen for å kunne oppeholde et visst forbruk. Tradisjonelt har også høyinntektsfamiliens barn hatt en større mulighet for å sende sine barn på skoler, mens lavinntektsfamiliens barn ofte rekrutteres til arbeidslivet i ung alder. Et moment som skulle veie mot her, er at utdanningsnivå blant familiemedlemmer i høyinntektsfamilier kanskje skulle føre til større yrkesaktivitet enn i lavinntektsfamilier. Inntektsstatistikken viser også at høyinntektshusholdninger har flere antall inntektstakere i gjennomsnitt enn lavinntektshusholdninger.

Videre viser figurene 4.2 - 4.4 at en overgang fra nettoinntekt pluss særfradrag til disponibel likningsinntekt m.h.p. fordeling blant husholdningene, vil føre til en jevnere fordeling. Gini-koeffisientene er her henholdsvis 0,373 og 0,314 (-73), 0,365 og 0,304 (-76) og, 0,369 og 0,300 (-79). Størrelsesorden på utslagene samt parvise like tendenser over observasjonsårene gjør at vi også her finner det rimelig å hevde at fordelingen er blitt jevnere ved ovennevnte overgang. Forklaringsfaktoren her er selvfølgelig skattesystemets progressive karakter m.h.p. nettoinntekt pluss særfradrag og de støtteordninger som er inneholdt i disponibel likningsinntektsbegrepet. Vi kan altså slutte at skatte- og stønadspolitikken er "effektiv" m.h.p. målsettingen om ujevning. Hvor effektiv den er er det imidlertid ikke mulig å si noe om uten å sette utslagene her i forhold til størrelsesorden på virkemidlene. Dette skal vi ikke gjøre her. Et interessant aspekt er det imidlertid at hvis en studerer Gini-koeffisientene, vil en ikke få helt klare konklusjoner m.h.p. hva som betyr mest i inntektsomfordelingen beskrevet ved figur 4.2 - 4.4: overgang til husholdningsenhet fra inntektstakerenhet eller overgang fra nettoinntekt pluss særfradrag til disponibel likningsinntektsbegrep for husholdningene.

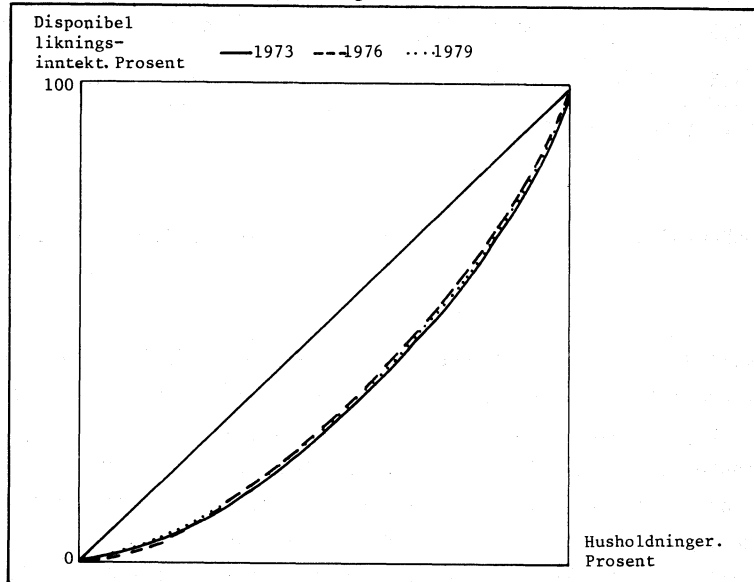
Figur 4.5 Lorentz-kurver over inntektstakeres nettoinntekt plus særfradrag for årene 1973, 1976 og 1979



Figur 4.6 Lorentz-kurver over husholdningens nettoinntekt plus særfradrag for årene 1973, 1976 og 1979



Figur 4.7 Lorentz-kurver over husholdningens disponible likningsinntekt for årene 1973, 1976 og 1979





Av figur 4.5 til 4.7 ser vi at det ikke er noen klar utjevning i inntektsfordelingene m.h.p. noen av begrepene/enhetene fra 1973 - 1979. Gini-koeffisientene er her henholdsvis 0,415, 0,443, 0,421 for inntektstakeres nettoinntekt pluss særfradrag, 0,373, 0,365, 0,369 for husholdningenes nettoinntekt pluss særfradrag og 0,314, 0,304, 0,300 for husholdningenes disponible likningsinntekt.

Dette er kanskje litt overraskende når man tenker på hvor framtreddende inntektsutjevningssjonesjonen har vært i løpet av disse årene. Inntektsutjevningssjonesjonen har imidlertid ofte gått på lønnsutjevning. Om dette har skjedd gir ikke disse kurvene svar på. Jf. forøvrig diskusjon under figur 4.1. Husholdningenes nettoinntekter er det vanskelig å styre mot inntektsutjevning. Denne fordelingen er direkte avhengig av antall inntektstakere i husholdningen. Husholdningenes disponible likningsinntekt er det imidlertid større mulighet til å påvirke via skatte- og stønadspolitikken. Det ser ikke ut som dette har skjedd i særlig stor grad gjennom 70-årene.

#### 4.4. Konklusjoner

Lønnstakeres inntektsfordeling kan ikke representeres av inntektstakeres inntektsfordeling, hovedsakelig pga. mangel på homogenitet i inntektstakerenheten. Det foregår en inntektsutjevning ved overgang fra inntektstakerenheten til husholdningsenheten og fra nettoinntekt pluss særfradrag til disponibel likningsinntekt. Det første pga. at antall inntektstakere i husholdningene varierer det andre hovedsakelig på grunn av skattepolitikken. Vi kan ikke påstå at inntektsfordelingen er blitt jevnere gjennom 70-årene for noen av begrepene inntektstakers nettoinntekt pluss særfradrag, husholdningers nettoinntekt pluss særfradrag eller husholdningers disponible likningsinntekt.

Tabell 4.1. Lønnstakeres lønnsinntekt (LI), inntektstakeres nettoinntekt pluss særfradrag (II), husholdningenes nettoinntekt pluss særfradrag (HI) og disponibel likningsinntekt (HD) etter akkumulerte deciler og fordelingenes Gini-koeffisienter (G)

Akkumulerte deciler Prosent	1973			1976				1979		
	II	HI	HD	LI	II	HI	HD	II	HI	HD
10 .....	0,6	1,6	2,4	0,3	0,3	1,4	2,2	0,4	1,7	2,7
20 .....	3,3	4,7	6,5	2,5	2,1	4,3	6,3	2,7	4,6	6,9
30 .....	7,3	9,1	12,0	7,2	5,6	8,9	12,1	6,6	9,0	12,7
40 .....	12,7	15,4	19,1	14,5	10,3	15,6	19,4	11,7	15,3	19,8
50 .....	19,9	23,5	27,6	23,8	17,1	23,9	28,1	18,9	23,4	28,2
60 .....	29,4	33,0	37,4	34,8	26,8	33,7	38,2	28,7	33,1	38,1
70 .....	41,3	44,2	48,7	47,2	39,4	45,2	49,7	41,1	44,6	49,7
80 .....	55,4	57,7	61,8	61,1	54,5	58,9	63,1	55,8	59,4	63,1
90 .....	72,5	74,5	77,5	77,1	72,6	75,6	78,8	73,3	75,2	78,7
100 .....	100,0	100,0	100,0	100,2	100,2	99,9	100,1	100,1	100,0	100,1
G .....	0,415	0,373	0,314	0,363	0,443	0,365	0,304	0,421	0,369	0,300

Kilde: Inntektsstatistikk 1973, 1976, 1979

## 5. LAR INNTEKTSFORDELINGEN SEG BESKRIVE AV EN PARAMETRISK KONTINUERLIG FUNKSJON

### 5.1. Innledning

I dette kapitlet diskuterer vi noen av de teoretiske fordelingsfunksjonene mange har hevdet kan beskrive inntektsfordelingen. I NOU 44/77 blir det hevdet at det er "vanlig" å anta at en inntektsfordeling best kan beskrives ved en lognormal fordelingsfunksjon.

To amerikanske forskere, A.B.L. Salem og T.D. Mount, har i *Econometrica* Vol. 42 nr. 6 1974 sett på fordelingen av familieinntekten i USA og påstår at inntektsfordelingen bedre kan beskrives ved en topa-parameter gamma fordelingsfunksjon. På tross av at inntektsfordelingen i USA og Norge ikke uten videre kan sammenliknes, vil vi undersøke om inntektsfordelingen i Norge kan beskrives av en slik funksjon.

Vi vil se på hvorledes en fordelingsfunksjon fra klassen av lognormale fordelingsfunksjoner og en fra klassen av gammafunksjoner beskriver inntektsfordelingen i Norge. I begge tilfellene estimeres parametrene i fordelingene ved hjelp av momentmetoden. Det betyr at vi "setter" forventning og varians i fordelingene lik henholdsvis middeltall og empirisk varians i de observerte inntektsfordelingene.

Etter at parametrene på denne måten er estimert, plotter vi de to fordelingene og vurderer hvilken vi synes passer best. Begge disse fordelingsfunksjonene har lett tolkbare parametre og vi kan derfor benytte endringene i dem fra år til år til å si noe om inntektsutviklingen. Dette er en av de store fordelene ved å beskrive inntektsfordelingen med en slik parametrisk fordelingsfunksjon.

De tre siste inntektsundersøkelsenes oversikt over husholdningenes inntekt kan sammenliknes år for år. I alle tre undersøkelsene har vi noen få husholdninger som har en inntekt høyt over gjennomsnittet. Disse husholdningene har vi valgt å se bort fra. Dette har ført til at når vi ser på husholdningens nettoinntekt pluss særfradrag, bare får med de ca. 96 prosent av husholdningene som har lavest inntekt. Når det gjelder disponibel likningsinntekt har vi tatt med ca. 99 prosent av husholdningene som har lavest disponibel likningsinntekt.

Denne avgrensingen har ført til at bare husholdninger med mindre enn 100 000 kroner i inntekt er med i inntektsfordelingene for 1973. Bare husholdninger med mindre enn 150 000 kr er med i inntektsfordelingene for 1976 og bare de husholdningene med mindre enn 200 000 kr i inntekt er med for 1979.

Bakgrunnen for at vi har valgt å gjøre denne avgrensningen er at disse vanskeliggjør tilpasningen av den parametriske fordelingsfunksjonen. Tabeller som viser husholdningene gruppert etter husholdningenes inntekt er gitt i vedlegg 2.

## 5.2. Den empiriske fordelingen

Vi har beregnet en del empiriske størrelser for de tre inntektsundersøkelsene 1973, 1976 og 1979. I det etterfølgende er:

$\bar{x}_a^I$  - gjennomsnittlig nettoinntekt pluss særfradrag (I) i år a pr. husholdning.

$\bar{x}_a^D$  - gjennomsnittlig disponibel likningsinntekt (D) i år a pr. husholdning.

$S_a^I$  - empirisk standardavvik i fordelingen for nettoinntekt pluss særfradrag.

$S_a^D$  - empirisk standardavvik i fordelingen for disponibel likningsinntekt.

Inntektsstatistikken er basert på en utvalgsundersøkelse av ca. 15 000 inntektstakere, dvs. ca. 8 000 husholdninger. Inntektsstatistikken opererer med oppblåste tall.\* Husholdningene opererer i tillegg med ulik vekt i statistikken. Beregningen av den empiriske variansen  $S^2$  har derfor skjedd på følgende måte. Vi vet at

$$(5.2.1) (S_a^k)^2 = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ija}^k - \bar{x}_a^k)^2$$

Her betegner  $X_{ija}^k$  - inntekt i år a for husholdning j i inntektsgruppe i når vi benytter inntektsbegrep k  
 i - inntektsgruppe i  
 $m_i$  - antall i inntektsgruppe i oppblåst  
 n - antall inntektsgrupper  
 N - totalt antall husholdninger (oppblåst tall)  
 k = I (Nettoinntekt pluss særfradrag), D (disponibel likningsinntekt)

\* Se inntektsstatistikk 1973 s. 8

Av (5.2.1) får vi

$$(5.2.2) (S_a^k)^2 = \sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ija}^k - \bar{x}_{i \cdot a}^k)^2 + \sum_{i=1}^n \frac{M_i}{N} (\bar{x}_{i \cdot a}^k - \bar{x}_a^k)^2$$

Her betegner  $\bar{x}_{i \cdot a}$  gjennomsnittsinntekten i inntektsgruppe  $i$ . Den empiriske variansen er blitt en sum av variansen innen hver inntektsgruppe  $V_1$  og variansen mellom inntektsgruppene  $V_2$ .

$V_2$  lar seg lett beregne fra tabellene i vedlegg 2. For å gi en tilnærming for  $V_2$  kan vi benytte følgende metode hentet fra målefeilsteorien.

I intervall  $i$  har vi  $m_i$  husholdninger med  $m_i$  ulike inntekter. disse fordelt med en avstand  $\frac{1}{m_i}$  i intervallet  $(C_0, C_1)$  slik at husholdning nr.  $i$  har inntekt  $C_0 + i \cdot \frac{1}{m_i} \cdot C_0$  er nedre inntektsgrense og  $C_1$  er øvre inntektsgrense i intervallet.

La  $V_1^{(i)}$  være variansen innen inntektsgruppe  $i$ .

Vi får da:

$$(5.2.3) V_1^{(i)} \approx \frac{m_i \cdot (C_1 - C_0)^2}{12} \quad \text{når } m_i \text{ er stor}$$

Da blir:

$$(5.2.4) V_1 = \sum_i^n V_1^{(i)}$$

Vi har i tillegg beregnet at selv om inntektene var fordelt slik at de gav størst mulig varians innen hver inntektsgruppe, ville  $V_1$  ikke overstige 10 prosent av den totale variansen. Feilen vi gjør ved å benytte oss av (5.2.3) er derfor relativt liten. Vi vil benytte denne senere ved estimering av parametrene i de ulike fordelingsfunksjonene.

Resultatene vi får av de empiriske beregningene m.h.p. nettoinntekt pluss særfradrag, er gitt i tabell 5.1. Populasjonen inneholder som nevnt bare de 96 prosent av husholdningene med lavest inntekt.

Tabell 5.1. Husholdningenes gjennomsnittlige nettoinntekt pluss særfradrag og empirisk standardavvik i fordelingen i ulike år. Alle tall gitt i kroner

	$\bar{x}_a^I$	$s_a^I$
1973 .....	36 900	22 900
1976 .....	60 300	36 300
1979 .....	74 900	48 700

Resultatene vi får av de tilsvarende empiriske beregningene m.h.p. disponibel likningsinntekt, er gitt i tabell 5.2. Populasjonen inneholder bare de 99 prosent av husholdningene med lavest inntekt.

Tabell 5.2. Husholdningenes gjennomsnittlige disponible likningsinntekt og standardavvik i fordelingen i ulike år. Alle tall gitt i kroner

a	$X_a^D$	$S_a^D$
1973 .....	32 100	17 900
1976 .....	52 200	28 300
1979 .....	63 700	36 800

Vi ser av disse tabellene at husholdningenes gjennomsnittlige inntekt har økt med 63,4 prosent fra 1973 til 1976 og med 24,2 prosent fra 1976 til 1979. Den gjennomsnittlige disponible likningsinntekten økte med 62,6 prosent fra 1973 til 1976 og med 22,0 prosent fra 1976 til 1979. Vi ser også at gjennomsnittlig disponibel likningsinntekt var ca. 86 prosent av nettoinntekt pluss særfradrag og at dette forholdet har vært tilnærmet konstant i de tre inntektsårene vi har sett på.

Ut fra disse empiriske størrelsene vi har funnet, har vi søkt å finne en teoretisk parametrisert fordelingsfunksjon som kunne beskrive inntektsfordelingene. Som nevnt i innledningen vil vi se på en funksjon fra klassen av lognormale fordelingsfunksjoner og en fra klassen av gamma fordelingsfunksjoner og sammenlikne disse.

### 5.3. Den lognormale fordeling

At en variabel  $X$  er fordelt lognormal  $(\mu, \sigma)$ , vil si at  $\ln X$  er normalfordelt med forventning  $\mu$  og varians  $\sigma^2$ . Fordelingen lar seg tabulere ut fra det kjennskap vi har til normalfordelingen.

Vi har:

$$(5.3.1) X = e^Y \quad \text{der } Y \text{ er } N(\mu, \sigma^2).$$

La  $M_Y(t)$  være den momentgenererende funksjonen for  $Y$ .

Da er

$$(5.3.2) M_Y(t) = E(e^{tY}) = E(X^t)$$

Den momentgenererende funksjonen til en normalfordelt variabel kjenner vi. Momentprinsippet gir oss derfor følgende estimator for  $\mu$  og  $\sigma^2$ :

$$(5.3.3) \hat{\mu} = \ln \bar{X} - \frac{1}{2} \ln \left( \frac{S^2}{\bar{X}^2} + 1 \right)$$

$$\hat{\sigma}^2 = \ln \left( \frac{S^2}{\bar{X}^2} + 1 \right)$$

(5.3.3) gir oss momentestimatoren til parametrene som bestemmer forventning og varians i den lognormale fordelingsfunksjonen. Vi ser av (5.3.3) at estimatoren bare avhenger av gjennomsnittet  $\bar{X}$  og den empiriske variansen  $S^2$ . Estimatorene er derfor lette å beregne. Men en estimator må vurderes ut fra sine samplingsegenskaper. Det kan vises at momentestimatorer under relativt generelle betingelser, er konsistent og tilnærmet normale når vi har et stort antall observasjoner.

### 5.4. Inntektsfordelingen beskrevet med en lognormal fordelingsfunksjonen

Vi vil anta vi har  $N$  husholdninger i Norge. Hver av disse har en nettoinntekt  $X_i$ ,  $i = (1, 2, \dots, n)$ . Vi antar nå at  $X_i$  er lognormalt fordelt slik at  $\ln X_i$  er  $N(\mu^I, (\sigma^I)^2)$ .  $I$  står for

nettoinntekt pluss særfradrag. D står for disponibel likningsinntekt. Vi kan ut fra opplysningene i inntekts- og formueundersøkelsen 1976 estimere  $\mu^I$  og  $(\sigma^I)^2$  ut fra kjennskapet til  $\bar{X}$  og  $S^2$  i inntektsstatistikken. Vi lar  $X_i^I$  være netto inntekt pluss særfradrag for husholdning i.

Estimater for  $\mu^I$  og  $(\sigma^I)^2$  blir:

$$(5.4.1)* \quad \hat{\mu}_{76}^I = \ln \bar{X}_{76}^I - \frac{1}{2} \ln \left[ \frac{(S_{76}^I)^2}{(X_{76}^I)^2} + 1 \right] = 6,247$$

$$(\hat{\sigma}_{76}^I)^2 = \ln \left[ \frac{(S_{76}^I)^2}{(X_{76}^I)^2} + 1 \right] = 0,309$$

Disse er beregnet etter at en liten høyinntektsgruppe er holdt utenfor.

I tillegg vet vi at det til hver husholdning er tilordnet en disponibel likningsinntekt  $Y$ . La  $Y$  være lognormal  $(\mu^D, (\sigma^D)^2)$ . D står for disponibel likningsinntekt. Inntektsstatistikken gir oss igjen estimater for  $\mu^D$  og  $(\sigma^D)^2$ .

$$(5.4.2)* \quad \hat{\mu}_{76}^D = 6,258$$

$$(\hat{\sigma}_{76}^D)^2 = 0,258$$

Ut fra disse estimatene kan vi tabellere  $F^I(x)$  og  $F^D(y)$ .  $F^I(x) \sim \text{lognormal}(6,247, 0,309)$  og  $F^D(y) \sim \text{lognormal}(6,258, 0,258)$ .

Vi har delt opp  $F^I(x)$  og  $F^D(y)$  i ulike inntektsgrupper i tabell 3 og tabell 4 og vedlegg 2. Grupperingen er gjort for å kunne sammenlikne de teoretiske fordelingene med den observerte inntektsfordelingen i inntektsstatistikken som er gitt i inntektsgrupper.

### 5.5. Gammafordelingen

At en variabel  $X$  er gammafordelt  $(\alpha, \lambda)$  vil si at den har en sannsynlighetstetthet  $g(x)$  lik:

$$(5.5.1) \quad g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{der } x > 0, \alpha > 0, \lambda > 0 \text{ og}$$

$\Gamma(\alpha)$  er gammafunksjonen definert ved

$$(5.5.2) \quad \Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$$

Momentprinsippet gir oss følgende estimatører for  $\alpha$  og  $\lambda$ .

$$(5.5.3) \quad \hat{\alpha} = \frac{\bar{X}^2}{S^2}$$

$$\hat{\lambda} = \frac{\bar{X}}{S^2}$$

På tilsvarende måte som i kapittel 5.3 ser vi at estimatorene kun er avhengig av gjennomsnittet og empirisk varians. Når vi har et stort antall observasjoner, vet vi som sagt at momentestimatørene under relativt generelle forhold er konsistente og tilnærmet normale.

\*) Tallene for  $\bar{X}_{76}^I$ ,  $\bar{X}_{76}^D$ ,  $S_{76}^I$  og  $S_{76}^D$  er hentet fra kapittel 5.2.

### 5.6. Inntektsfordelingen beskrevet med en gamme fordelingsfunksjon

Vi vil som i kapittel 5.4 anta at det finnes  $N$  husholdninger i Norge. Hver av disse har en nettoinntekt  $X_i^I$   $i = (1, 2, \dots, N)$ . Vi antar at  $X_i^I$  er gammafordelt  $(\alpha^I, \lambda^I)$ .

I Inntekts- og formuesundersøkelsen 1976 har vi en observasjon av inntekten. Vi kan ut fra opplysninger derfra estimere  $\alpha^I$  og  $\lambda^I$  ut fra kjennskapet til  $\bar{X}^I$  og  $(S^I)^2$  i inntektsstatistikken.

Estimater for  $\alpha^I$  og  $\lambda^I$  blir:

$$(5.61) \quad \hat{\alpha}_{76}^I = \frac{(\bar{X}_{76}^I)^2}{(S_{76}^I)^2} = 2,76$$

$$\hat{\lambda}_{76}^I = \frac{\bar{X}_{76}^I}{(S_{76}^I)^2} = 4,58 \cdot 10^{-3}$$

Vi kan endre inntektsbegrepet og si at til enhver husholdning er tilordnet en disponibel likningsinntekt  $Y$ . Vi lar  $Y$  være gammafordelt  $(\alpha^D, \lambda^D)$ . Inntektsstatistikken gir oss da følgende estimator for  $\alpha^D$  og  $\lambda^D$ :

$$(5.6.2)* \quad \hat{\alpha}_{76}^D = 3,40$$

$$\hat{\lambda}_{76}^D = 6,51 \cdot 10^{-3}$$

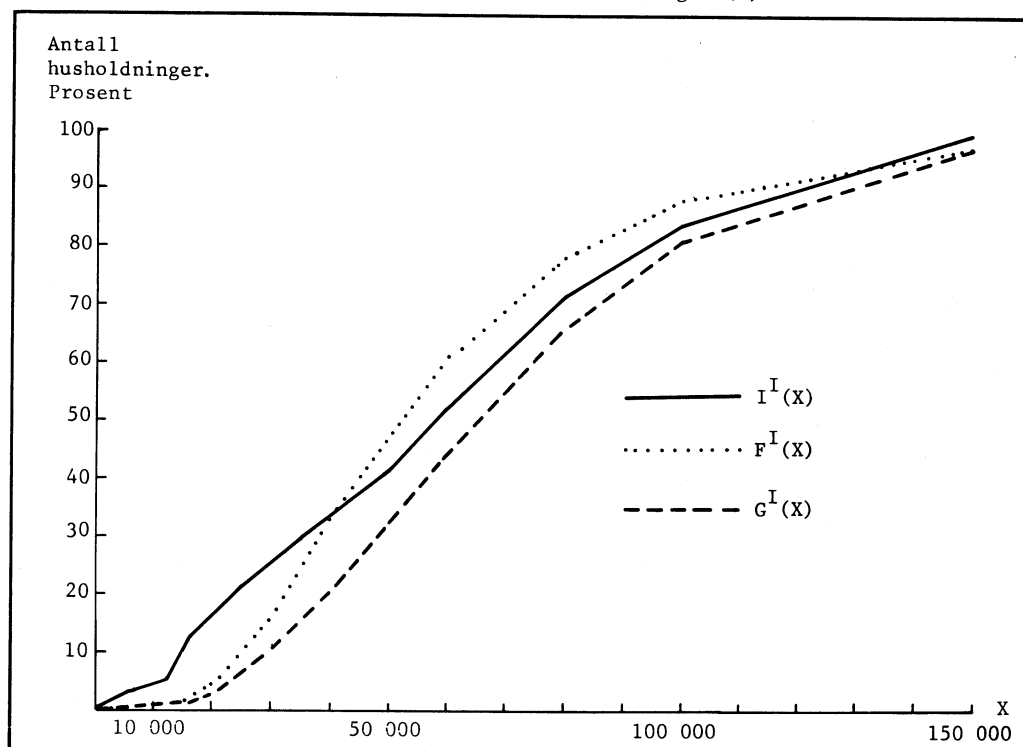
Ut fra disse estimatene kan vi tabulere  $G^I(x)$  og  $G^D(y)$ .

På samme måte som i kapittel 5.2 har vi delt inn  $G^I(x)$  og  $G^D(y)$  i ulike inntektsgrupper. Dette er gjort i tabell 5 og tabell 6 i vedlegg 2.

### 5.7. En sammenlikning av funksjonene funnet i 5.4 og 5.6 og den empiriske inntektsfordelingen beskrevet i inntektsstatistikken

Ut fra tabellene i vedlegg 1 har vi tegnet inn den empiriske inntektsfordelingen fra inntektsstatistikken og de to teoretiske fordelingsfunksjonene som skulle beskrive den. Dette er gjort for de to inntektsbegrepene nettoinntekt pluss særfradrag og disponibel likningsinntekt. Vi kunne få et teoretisk mål for hvor godt fordelingsfunksjonene beskriver inntektsfordelingen ved å legge et konfidensbånd om den observerte inntektsfordelingen. På grunn av det store antall observasjoner i inntektsstatistikken er det omfattende å beregne et slikt konfidensbånd. En løsning kunne vært å benytte et utvalg av observasjonene, og deretter estimere konfidensbåndet. Siden vi ikke har hatt ressurser til å gjøre dette, har vi måttet nøye oss med de grafiske framstillingene, og resonnerer på intuitivt grunnlag.

\*) Se note side 28

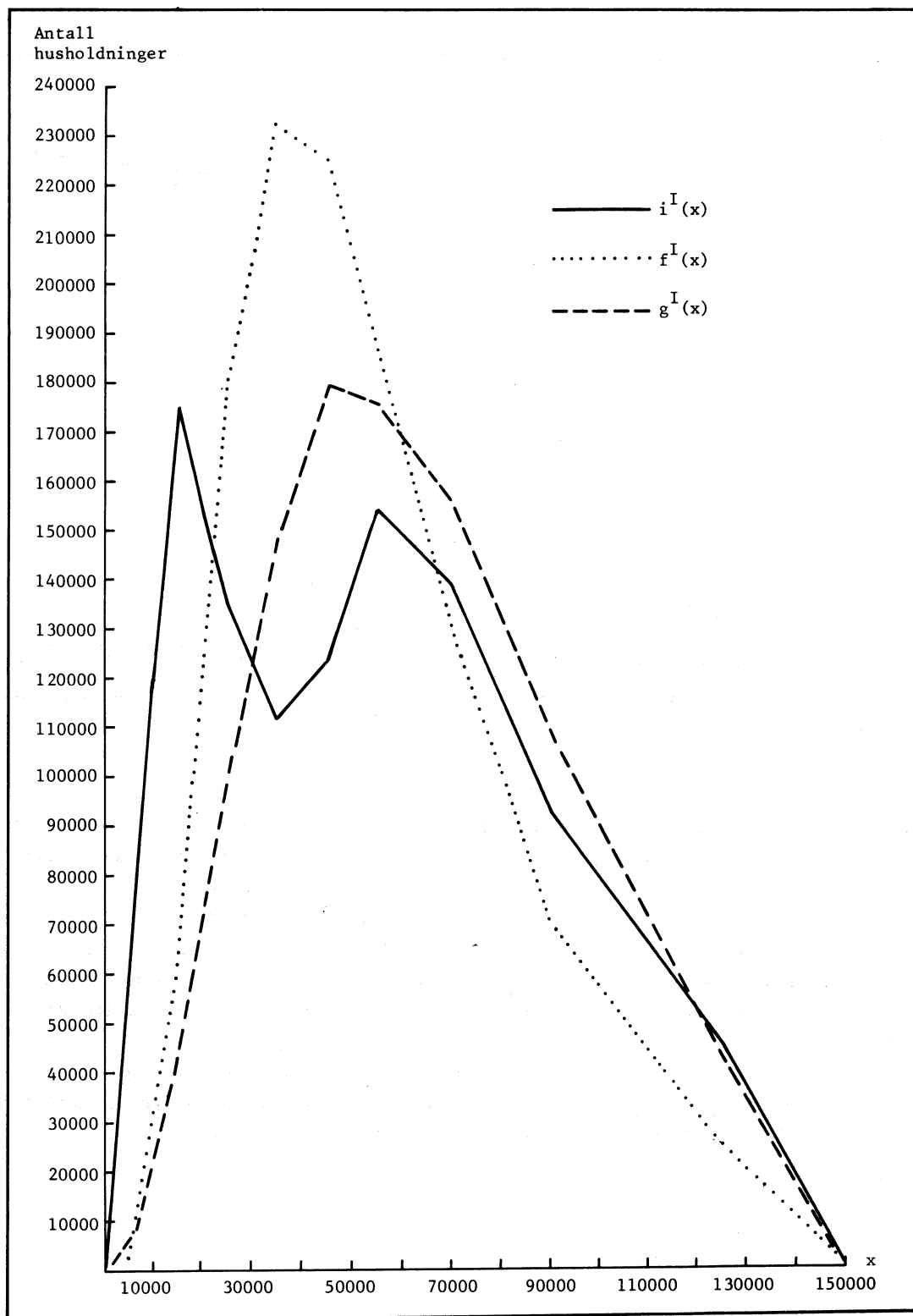
Figur 5.1 En grafisk framstilling av  $I^I(x)$ ,  $F^I(x)$  og  $G^I(x)$ 

I figur 5.1 har vi tegnet de kumulative fordelingsfunksjonene  $I^I(x)$ ,  $F^I(x)$  og  $G^I(x)$  der  $x$  er nettoinntekt pluss særfradrag. Tallene finnes i tabell 1, tabell 3 og tabell 5 i vedlegg 2.

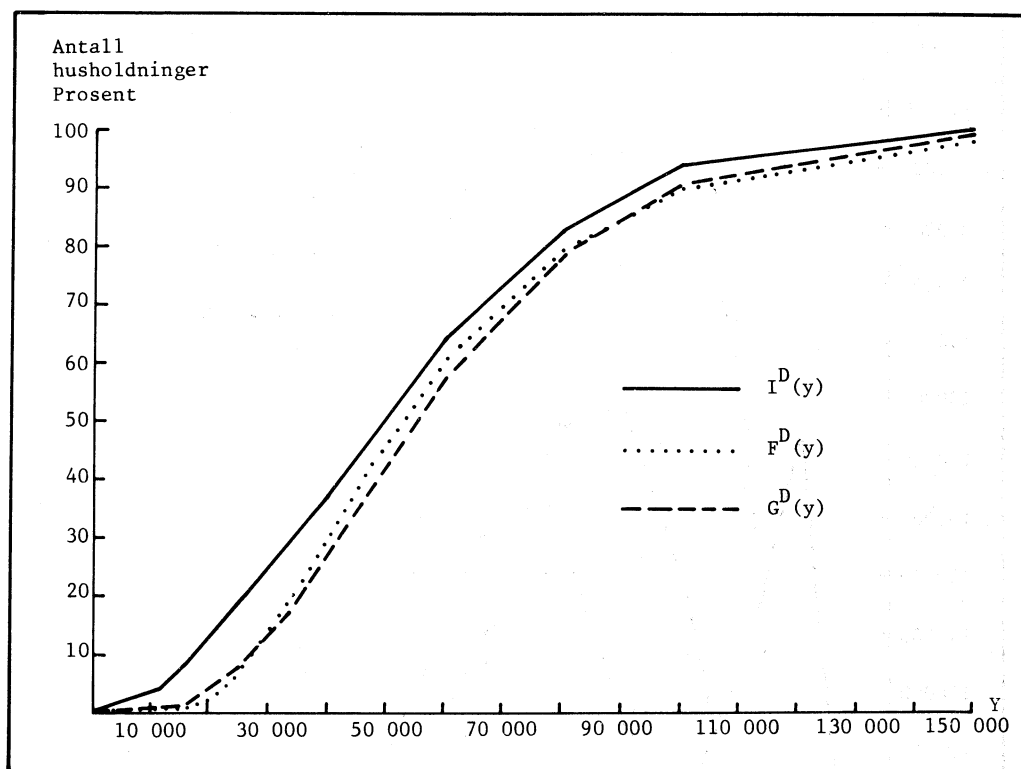
Figur 5.2 viser tettheten  $i^I(x)$ ,  $f^I(x)$  og  $g^I(x)$  til fordelingsfunksjonene  $I^I(x)$ ,  $F^I(x)$  og  $G^I(x)$ .

For å tegne funksjonen  $I^I(x)$  har vi benyttet os av inntektsstatistikkens grupperte data. For å kunne sammenlikne  $F^I(x)$  med  $I^I(x)$  har vi beregnet det antall som skulle hatt inntekt i de ulike intervallene hvis inntektsfordelingen hadde hatt  $F^I(x)$  som fordelingsfunksjon. Denne har vi tabulert i tabell 3. Vi har også gjennomført tilsvarende for  $G^I(x)$ . Dette er tabulert i tabell 5. Vi har så slått sammen en del av inntektsgruppene slik at vi har fått en bredde på 10 000 kr i hvert inntektsintervall. For høye inntekter har vi vært nødt til å bruke 20 000 kr som intervallbredde og i det siste intervallet, 100 000 - 150 000, har vi vært nødt til å benytte 50 000 kr som bredde i intervallet. Vi har dermed fått 8 inntektsintervaller. Hver av disse har vi latt representere med en representantverdi  $x_i$  som er midtpunktet i intervallet. I figur 5.2 har vi skissert "tettheten" i inntektsfordelingen ved å trekke en rett linje mellom  $i^I(x_i)$  og  $i^I(x_j)$ . Det tilsvarende har vi gjort for å skissere  $f^I(x)$  og  $g^I(x)$ .

Av figur 5.1 og 5.2 ser vi at både den lognormale fordelingen og gammafordelingen underestimerer antall husholdninger med lav inntekt. Den lognormale fordelingen overestimerer antallet husholdninger med middels inntekt og underestimerer antall husholdninger med høy inntekt. Den lognormale fordelingsfunksjonen gir oss en jevnere inntektsfordeling enn inntektsstatistikkens data gir. Gammafordelingen overestimerer antall husholdninger med middels og høy inntekt, men ikke i så stor grad som den lognormale fordelingsfunksjonen. Det synes som en gamma fordelingsfunksjonen bedre tar vare på skjevheten i inntektsfordelingen.

Figur 5.2 En grafisk framstilling av  $i^I(x)$ ,  $f^I(x)$ ,  $g^I(x)$ 

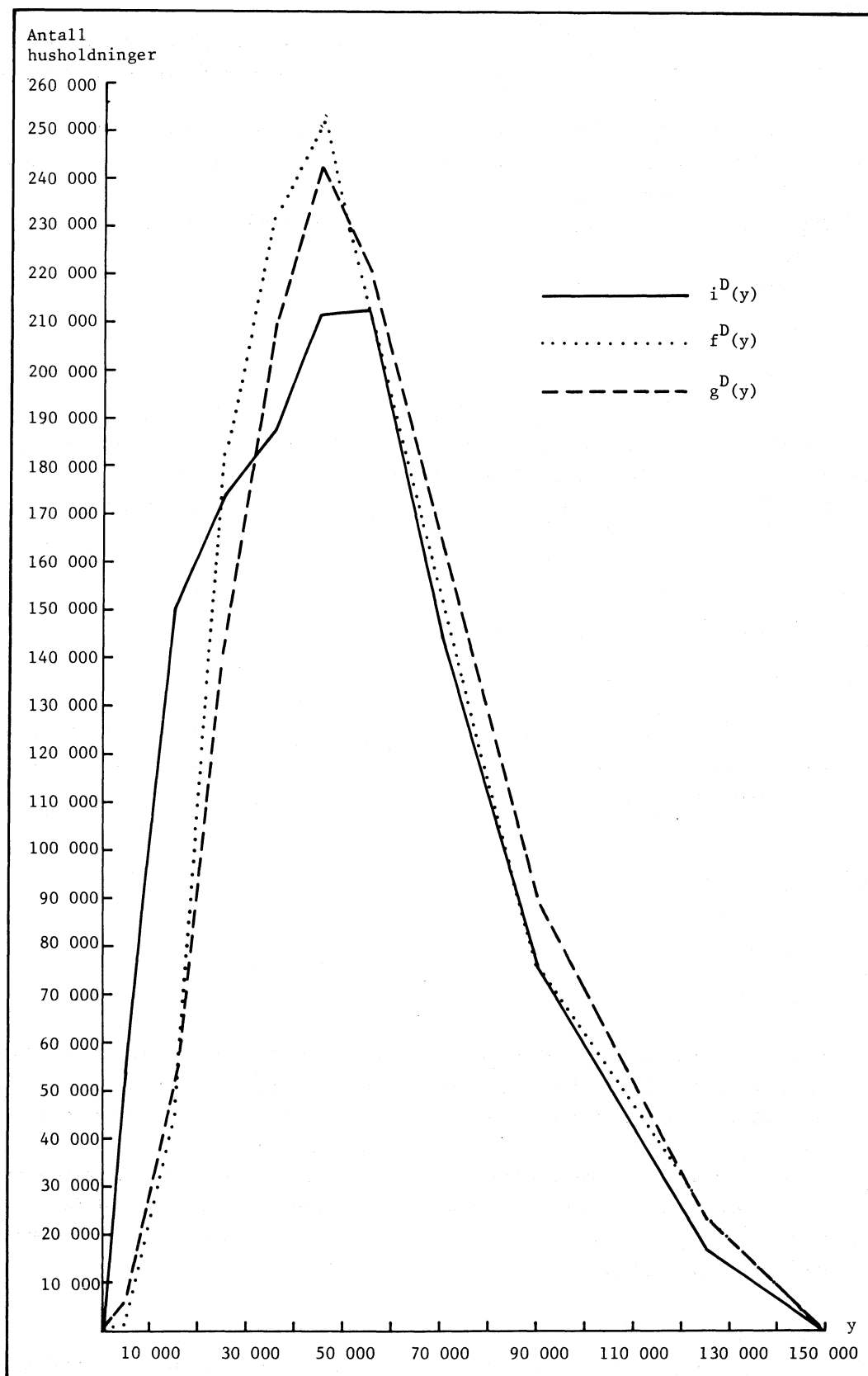


Figur 5.3 En grafisk framstilling av  $I^D(y)$ ,  $F^D(y)$ ,  $G^D(y)$ 

I figur 5.3 og 5.4 har vi tegnet tilsvarende figur som i 5.1 og 5.2 men byttet ut begrepet nettoinntekt pluss særfradrag med inntektsbegrepet disponibel likningsinntekt. Vi ser av figur 5.3 og 5.4 at  $F^D(y)$  og  $G^D(y)$  beskriver funksjonen  $I^D(y)$  bedre enn de tilsvarende funksjonen gjorde i figur 5.1 og 5.2.

Fortsatt ser vi at både den lognormale fordelingen og gammafordelingen underestimerer antall husholdninger med lav inntekt og overestimerer husholdninger med middelsinntekt. Når vi benytter inntektsbegrepet disponibel likningsinntekt ser vi at vi ikke uten videre kan avgjøre hvilken av de to teoretiske fordelingsfunksjonene vi bør velge. Vi gjør ingen stor feil ved å foretrekke den ene framfor den andre.

For drøftinger videre har vi valgt å beskrive inntektsfordelingen ved hjelp av en gammafordeling. En fordel med gammafordelingen er at den har lettere tolkbare parametre enn den lognormale fordelingen, og de kan lettere fortelle om hvorledes skjevheten i inntektsfordelingen er.

Figur 5.4 En grafisk framstilling av  $i^D(y)$ ,  $f^D(y)$ ,  $g^D(y)$ 

### 5.8. Gammafordelingens parametre

Vi har i kapittel 5.7 funnet at vi ønsker å benytte oss av en gamma fordeling når vi skal beskrive inntektsfordelinger. Selv om beskrivelsen av inntektsfordelingen ikke er perfekt kan endringen i parametrene si noe om de ulike inntektsfordelingene.

I en to parametrers gammafordeling (se 5.5.1) er parameteren  $\lambda$  en ren skalaparameter. Hvis vi innfører  $Z = c \cdot X$  ( $c$  er en konstant) og  $X$  er gamma  $(\alpha, \lambda)$  så er  $Z \sim \text{gamma}(\alpha, \frac{\lambda}{c})$ .  $\alpha$  er derimot en såkalt fasongparameter (shape parameter).

Ved å studere forskjellige gammafordelinger hvor  $\lambda$  holdes fast og  $\alpha$  varierer finner vi at en økning i  $\alpha$  medfører at vi får en jevnere fordeling.

Fra kapittel 5.5 har vi at

$$(5.8.1) \quad \hat{\alpha}^I = \frac{(\bar{X}^I)^2}{(S^I)^2}$$

Fra kapittel 2 kjenner vi igjen variasjonskoeffisienten invertert og kvadrert. Som estimator for  $\alpha$ -parameteren i gamma fordelingen får koeffisienten en mye videre tolkning.

Vi har beregnet  $\hat{\alpha}^I$  for årene 1973, 1976 og 1979 ved hjelp av inntektsstatistikken og har funnet

$$(5.8.2) \quad \hat{\alpha}_{1973}^I = 2,60$$

$$\hat{\alpha}_{1976}^I = 2,76$$

$$\hat{\alpha}_{1979}^I = 2,37$$

Vi ser at  $\hat{\alpha}$  økte fra 1973 til 1976, men avtok så fra 1976 til 1979. Dette er en indikasjon på at inntektsfordelingen jevnet seg ut fra 1973 til 1976, men deretter igjen ble mer ujevn fram til 1979.

Tilsvarende kan vi for inntektsbegrepet disponibel likningsinntekt beregne

$$(5.8.3) \quad \hat{\alpha}_{1973}^D = 3.22$$

$$\hat{\alpha}_{1976}^D = 3.40$$

$$\hat{\alpha}_{1979}^D = 3,00$$

Vi ser at inntektsfordelingen er jevnere når vi benytter inntektsbegrepet disponibel likningsinntekt enn nettoinntekt pluss særfradrag. Dette virker rimelig pga. statens skatte og stønadspolitik.

Resultatene stemmer godt overens med det vi så for nettoinntekt pluss særfradrag. Vi har en indikasjon på at inntektsfordelingen i 1979 var mindre jevn enn i 1973.

Vi ser altså at ved å beskrive inntektsfordelingen med en parametrisk fordelingsfunksjon har vi relativt enkelt kunnet si noe om utviklingen av hele inntektsfordelingen ut fra utviklingen av estimatene for parametrene. Dette er en av årsakene til at en ofte prøver å beskrive et empirisk materiale med en parametrisk fordelingsfunksjon.

### 5.9. Oppsummering

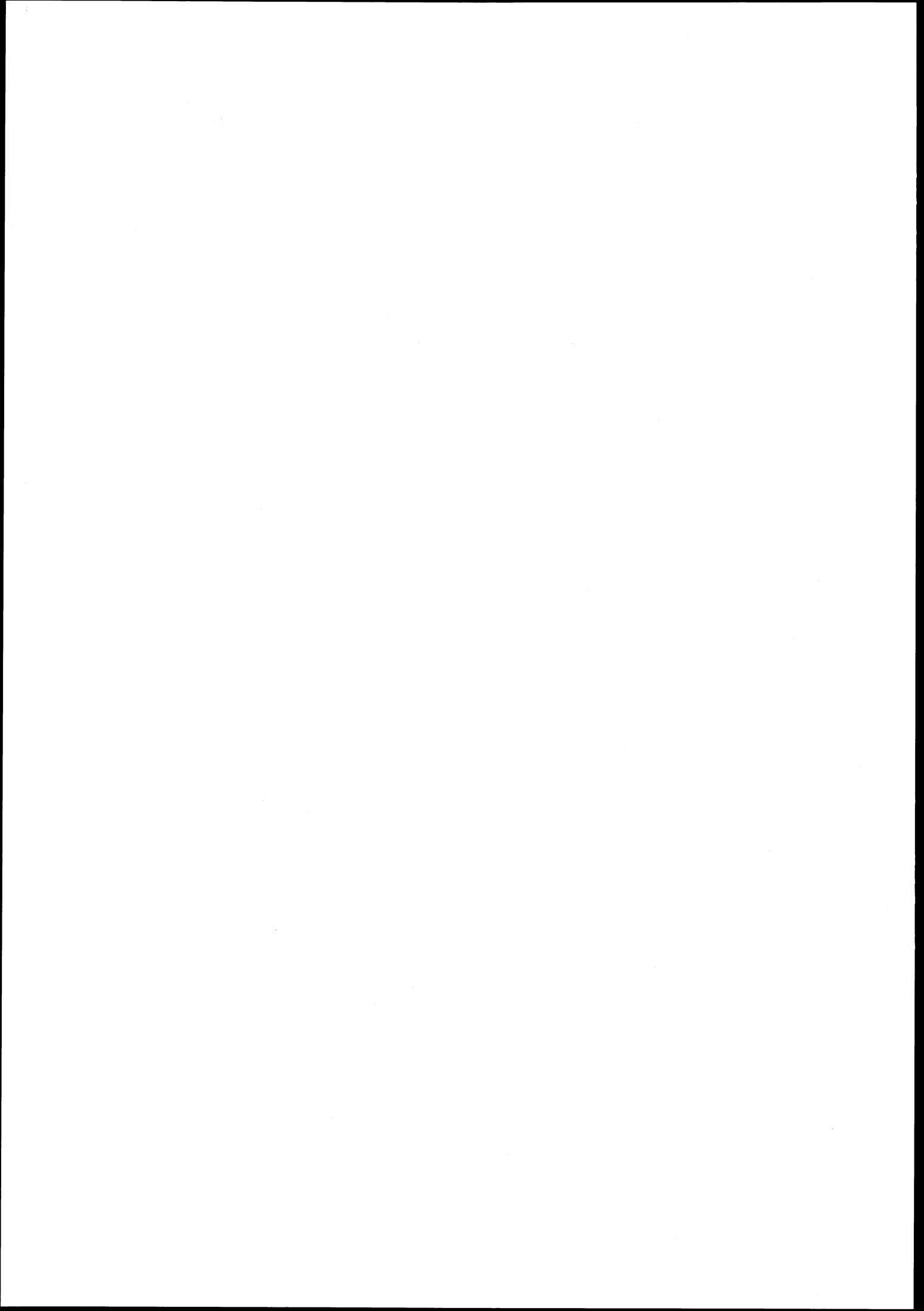
Ved å beskrive den empiriske inntektsfordelingen  $I(x)$  med en parametrisk fordelingsfunksjon  $F(x)$ , har vi med relativt enkle metoder kunnet si noe om ulike inntektsfordelinger. For å avgjøre om tilpasningen er tilfredsstillende kreves bruk av statistiske metoder. I dette kapitlet har vi imidlertid kun benyttet oss av intuitive resonnementer. Hovedmålsettingen vår har vært å peke på metoder som kan benyttes til å studere forskjellige inntektsfordelinger. De samme metodene vil selvsagt kunne nyttes selv om inntektsbegrepet og inntektstakerenheten er valgt på en annen måte.

Vi har sett på to typer fordelingsfunksjoner. Det kan selvfølgelig finnes andre.

Selv om vi "fjerner" husholdningene med særlig høy inntekt, så er det de husholdningene med lav som skaper de største problemene for tilpasning av fordelingsfunksjonene. Hadde vi også "fjernet" husholdninger med lave inntekter kunne vi kanskje funnet en fordelingsfunksjon som bedre beskrev inntektsfordelingen for de husholdningene som hadde mellom 20 000 og 150 000 kr inntekt.

Vi har valgt å beskrive inntektsfordelingen i Norge med en gamma fordelingsfunksjon. Estimatorene til parametrene i fordelingen er funnet ved momentmetoden. Det finnes muligens bedre estimeringsmetoder enn denne. Det ville kreves økte ressurser om de skulle finnes og brukes. Siden antallet observasjoner i inntektsstatistikken er mange er det tvilsomt om gevinsten i nøyaktighet ville forsvart det økte ressursforbruket.

Enkelte hevder på bakgrunn av vanskelighetene med å beskrive inntektsfordelingen godt med enkelte prametricke fordelingsfunksjoner at en bør tilpasse forskjellige fordelingsfunksjoner i de ulike inntektsintervallene. Vi er ikke enig i dette. Selv om vi ved en slik framgangsmåte kan oppnå en bedre tilpasning mister vi samtidig fordelene ved at endringen i estimatene for parametrene er relativt enkle å tolke.



## LITTERATUR

NOS Lønnsstatistikk 1976

NOS Skattestatistikk 1976

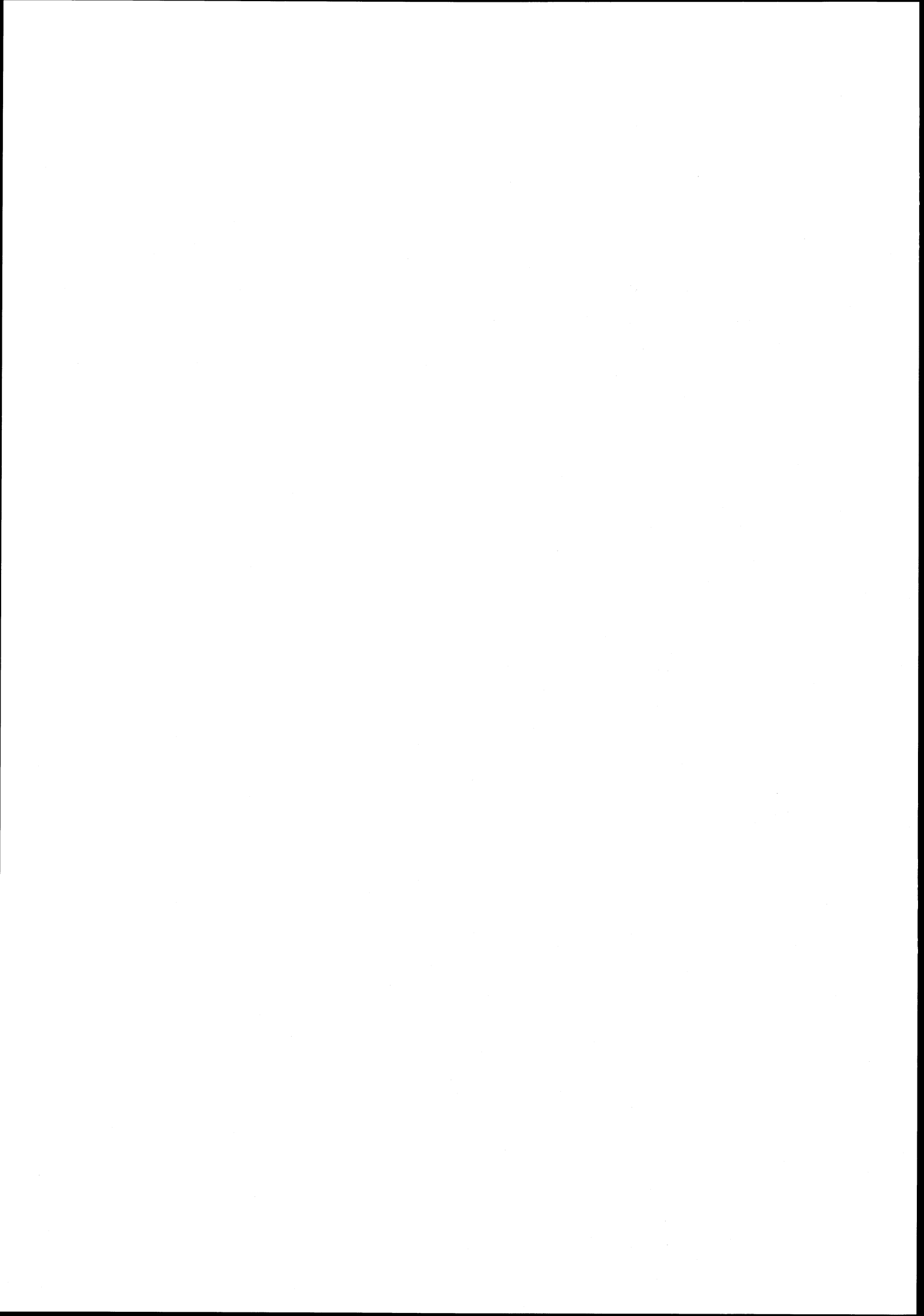
NOS Inntektsstatistikk 1976

NOS Statistikk over lavinntektsgrupper

Høyland, A. Sannsynlighetsregning 1 og 2

NOU: 1974:34 Inntektsfordeling i Norge

Salem, A.B.L. Mount T.D., Econometrica Vol. 42 nr. 6 1974



## TABELLER

V e d l e g g 2

Tabell 1. Husholdninger med nettoinntekt pluss særfradrag mindre enn 150 000 kr. 1976

Husholdningens inntekt. Kr	Tallet på husholdninger	Prosent	Inntekt	Gjennomsnitts- inntekt
			Mill.kr	Kr
HUSHOLDNINGER I ALT .....	1 461 836	100,0	88 220,1	60 300
Under 4 000 .....	38 190	2,6	51,1	1 300
4 000 - 5 999 .....	8 092	0,5	40,7	5 000
6 000 - 7 999 .....	10 629	0,7	74,0	7 000
8 000 - 9 999 .....	8 548	0,6	77,9	9 000
10 000 - 11 999 .....	8 090	0,5	89,2	11 000
12 000 - 15 999 .....	111 584	7,6	1 556,1	14 000
16 000 - 19 999 .....	56 372	3,9	1 018,6	18 100
20 000 - 24 999 .....	78 092	5,3	1 759,0	22 500
25 000 - 29 999 .....	57 766	4,0	1 589,3	27 500
30 000 - 34 999 .....	56 380	3,9	1 854,0	32 900
35 000 - 39 999 .....	55 461	3,8	2 092,0	37 700
40 000 - 49 999 .....	123 399	8,4	5 607,8	45 400
50 000 - 59 999 .....	154 589	10,6	8 542,7	55 300
60 000 - 79 999 .....	276 913	18,9	19 352,2	69 900
80 000 - 99 999 .....	186 632	12,8	16 769,0	89 900
100 000 - 149 999 .....	231 099	15,8	27 736,5	120 000

Tabell 2. Husholdninger med disponibel likningsinntekt mindre enn 150 000 kr. 1976

Husholdningens inntekt. Kr	Tallet på husholdninger	Prosent	Inntekt	Gjennomsnitts- inntekt
			Mill.kr	Kr
HUSHOLDNINGER I ALT .....	1 523 690	100,0	79 512,8	52 200
Under 4 000 .....	30 439	2,0	65,0	2 100
4 000 - 5 999 .....	5 089	0,3	25,1	4 900
6 000 - 7 999 .....	9 479	0,6	67,1	7 100
8 000 - 9 999 .....	10 176	0,7	90,3	8 900
10 000 - 12 999 .....	9 257	0,6	103,1	11 100
12 000 - 15 999 .....	58 004	3,8	881,1	15 200
16 000 - 19 999 .....	82 964	5,4	1 484,6	17 900
20 000 - 24 999 .....	79 071	5,2	1 797,8	22 700
25 000 - 29 999 .....	95 015	6,2	2 627,0	27 600
30 000 - 34 999 .....	89 949	5,9	2 938,1	32 700
35 000 - 39 999 .....	96 899	6,4	3 661,0	37 800
40 000 - 49 999 .....	211 867	13,9	9 612,9	45 400
50 000 - 59 999 .....	212 320	13,9	11 728,9	55 200
60 000 - 79 999 .....	290 166	19,0	20 300,9	70 000
80 000 - 99 999 .....	153 551	10,1	13 685,5	89 100
100 000 - 149 999 .....	89 444	5,9	10 444,4	116 800



Tabell 3. Inntektsfordelingen beskrevet med en lognormal fordelingsfunksjon  $F^I(x)^*$ . Nettoinntekt pluss særfradrag

Inntekts- intervall. Kr	Beregnet antall husholdninger	Prosent
0 - 3 999 ....	..	0,0
4 000 - 5 999 ....	..	0,0
6 000 - 7 999 ....	1 462	0,1
8 000 - 9 999 ....	1 462	0,1
10 000 - 11 999 ....	2 924	0,2
12 000 - 15 999 ....	20 466	1,4
16 000 - 19 999 ....	38 008	2,6
20 000 - 24 999 ....	77 477	5,3
25 000 - 29 999 ....	97 943	6,7
30 000 - 34 999 ....	114 023	7,8
35 000 - 39 999 ....	118 409	8,1
40 000 - 49 999 ....	225 123	15,4
50 000 - 59 999 ....	188 577	12,9
60 000 - 79 999 ....	261 669	17,9
80 000 - 99 999 ....	143 260	9,8
100 000 - 149 999 ....	128 642	8,8

\*)  $F^I(x) \sim \text{lognorm}(6,247, 0,309)$ .

Tabell 4. Inntektsfordelingen beskrevet med en lognormal fordelingsfunksjon  $F^D(y)^*$ . Disponibel likningsinntekt

Inntekts- intervall. Kr	Beregnet antall husholdninger	Prosent
0 - 3 999 ....	..	0,0
4 000 - 5 999 ....	..	0,0
6 000 - 7 999 ....	..	0,0
8 000 - 9 999 ....	1 524	0,1
10 000 - 11 999 ....	1 524	0,1
12 000 - 15 999 ....	12 100	0,8
16 000 - 19 999 ....	28 950	1,9
20 000 - 24 999 ....	68 566	4,5
25 000 - 29 999 ....	112 753	7,4
30 000 - 34 999 ....	102 087	6,7
35 000 - 39 999 ....	129 514	8,5
40 000 - 49 999 ....	251 409	16,5
50 000 - 59 999 ....	210 269	13,8
60 000 - 79 999 ....	300 167	19,7
80 000 - 99 999 ....	152 369	10,0
100 000 - 149 999 ....	123 419	8,1

\*)  $F^D(y) \sim \text{lognorm}(6,258, 0,258)$

Tabell 5. Inntektsfordelingen beskrevet med en toparameter gamma fordelingsfunksjon  $G^I(x)^*$ . Nettoinntekt pluss særfradrag

Inntektsintervall. Kr	Beregnet antall husholdninger	Prosent
0 - 3 999 ....	..	0,0
4 000 - 5 999 ....	1 462	0,1
6 000 - 7 999 ....	1 462	0,1
8 000 - 9 999 ....	2 924	0,2
10 000 - 11 999 ....	2 924	0,2
12 000 - 15 999 ....	17 542	1,2
16 000 - 19 999 ....	23 389	1,6
20 000 - 24 999 ....	46 777	3,2
25 000 - 29 999 ....	54 088	3,7
30 000 - 34 999 ....	67 244	4,6
35 000 - 39 999 ....	80 401	5,5
40 000 - 49 999 ....	179 806	12,3
50 000 - 59 999 ....	176 882	12,1
60 000 - 79 999 ....	312 833	21,4
80 000 - 99 999 ....	219 275	15,0
100 000 - 149 999 ....	226 585	15,5

\*)  $G^I(x) \sim \text{gamma}(2,76, 4,58 \cdot 10^{-3})$ .

Tabell 6. Inntektsfordelingen beskrevet med en toparameter gamma fordelingsfunksjon  $G^D(y)^*$ . Disponibel likningsinntekt

Inntektsintervall. Kr	Beregnet antall husholdninger	Prosent
0 - 3 999 ....	..	0,0
4 000 - 5 999 ....	..	0,0
6 000 - 7 999 ....	1 524	0,1
8 000 - 9 999 ....	4 571	0,3
10 000 - 11 999 ....	4 571	0,3
12 000 - 15 999 ....	16 761	1,1
16 000 - 19 999 ....	28 950	1,9
20 000 - 24 999 ....	60 948	4,0
25 000 - 29 999 ....	83 803	5,5
30 000 - 34 999 ....	100 564	6,6
35 000 - 39 999 ....	106 658	7,0
40 000 - 49 999 ....	242 267	15,9
50 000 - 59 999 ....	220 935	14,5
60 000 - 79 999 ....	332 164	21,8
80 000 - 99 999 ....	182 843	12,0
100 000 - 149 999 ....	124 943	8,2

\*)  $G^D(y) \sim \text{gamma}(3,40, 6,51 \cdot 10^{-3})$ .

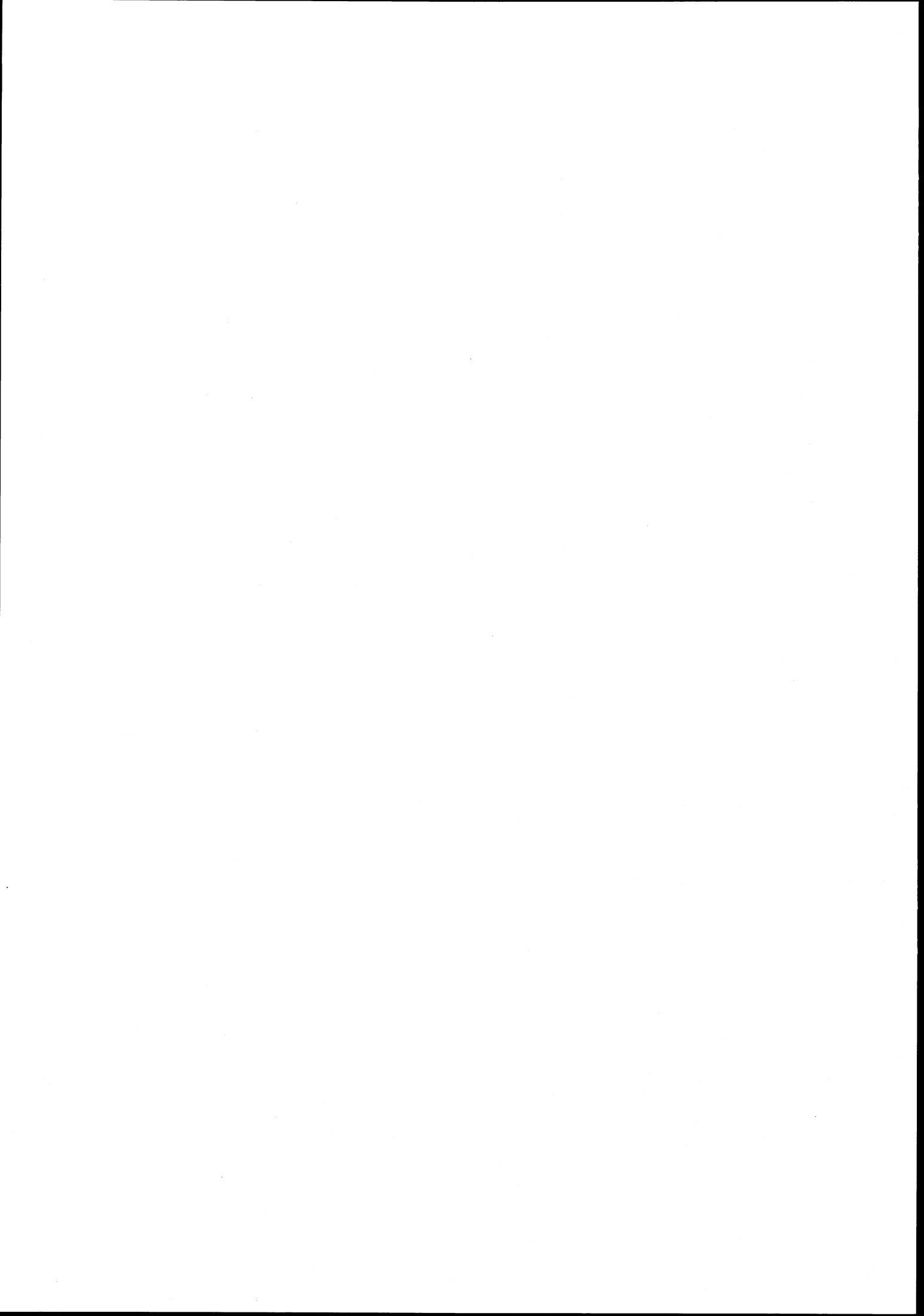
Trykt 1981

- Nr. 81/1 Erling J. Fløttum: National Accounts of Norway System and Methods of Estimation Sidetall 101 Pris kr 15,00 ISBN 82-537-1555-2
- 81/2 Referansearkiv for naturressurs- og forurensningsdata 2. utgave Sidetall 424 Pris kr 20,00 ISBN 82-537-1233-2
- 81/3 Byggekostnadsindeks for boliger Sidetall 127 Pris kr 15,00 ISBN 82-537-1232-4
- 81/4 Intervjuernes erfaringer fra arbeidskraftundersøkelsene Rapport fra 99 intervjuere Field Work Experiences with the Labour Force Sample Survey Reports from 99 Interviewers Sidetall 40 Pris kr 10,00 ISBN 82-537-1234-0
- 81/5 Strukturundersøkelse for bygg og anlegg Vann- og kloakkanlegg Sidetall 62 Pris kr 15,00 ISBN 82-537-1235-9
- 81/6 Erling Siring og Ib Thomsen: Metoder for estimering av tall for fylker ved hjelp av utvalgsundersøkelser Sidetall 42 Pris kr 10,00 ISBN 82-537-1509-9
- 81/7 Arne Ljones og Hans Viggo Sæbø: Temperaturkorrigering av energiforbruket Sidetall 43 Pris kr 10,00 ISBN 82-537-1507-2
- 81/8 En analyse av faktorinnsatsen i Norges utenrikshandel med utviklingsland og industriland Sidetall 55 Pris kr 15,00 ISBN 82-537-1506-4
- 81/9 Petter Longva: A System of Natural Resource Accounts Eit rekneskapssystem for naturressursar Sidetall 26 Pris kr 10,00 ISBN 82-537-1540-4
- 81/10 Stein Erland Brun: Tilgangen på arbeidskraft i fylkene for årene 1971 - 1979 Sidetall 72 Pris kr 15,00 ISBN 82-537-1514-5
- 81/11 Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1979 Sidetall 264 Pris kr 20,00 ISBN 82-537-1515-3
- 81/12 Helge Brunborg, Jan Mønnesland og Randi Selmer: Framskrivning av folkemengden etter ekteskapeleg status Sidetall 75 Pris kr 11,00 ISBN 82-537-1541-2
- 81/13 Adne Cappelen: Importinnhold i sluttleveringer Sidetall 20 Pris kr 10,00 ISBN 82-537-1545-5
- 81/14 MODIS IV Dokumentasjonsnotat nr. 16 Endringer i utgave 78-1 og 79-1 Sidetall 100 Pris kr 15,00 ISBN 82-537-1549-8
- 81/15 Skatter og overføringer til private Historisk oversikt over satser mv. Årene 1969 - 1981 Sidetall 74 Pris kr 15,00 ISBN 82-537-1554-4
- 81/16 Helgeturer 1978/79 Sidetall 23 Pris kr 10,00 ISBN 82-537-1560-9
- 81/17 Roy Østensen: Eie og bruk av personbil Foreløpige tall for 1979 og 1. kvartal 1980 Sidetall 42 Pris kr 10,00 ISBN 82-537-1566-8
- 81/18 Svein Homstvedt, Øyvind Lone og Tore Nesheim: Jordbruksareal ifølge jordregister og utvalgstillinger. Metodiske forskjeller belyst med materiale fra Trøgstad kommune Sidetall 62 Pris kr 15,00 ISBN 82-537-1598-6
- 81/19 Arne Faye: Holdninger til norsk utviklingshjelp 1980 Sidetall 62 Pris kr 15,00 ISBN 82-537-1562-5
- 81/20 Knut Fredrik Strøm: Konkurser i industri og varehandel Utvikling, hyppighet og omfang Sidetall 31 Pris kr 10,00 ISBN 82-537-1569-2
- 81/21 Frank Foyn: Miljøverninvesteringer i industrien. Problemer ved kartlegging av data Sidetall 34 Pris kr 10,00 ISBN 82-537-1591-9
- Nr. 81/22 Petter R. Koren: Etterspørsel etter energi i norsk industri Sidetall 27 Pris kr 10,00 ISBN 82-537-1592-7
- 81/23 Harald Bergland og Adne Cappelen: Produktivitet og sysselsetting i industrien Sidetall 75 Pris kr 15,00 ISBN 82-537-1600-1
- 81/24 Levekårsundersøkelsen 1980 Dokumentasjon Del I Sidetall 67 Pris kr 15,00 ISBN 82-537-1612-5
- 81/25 Tor Haldorsen: Norske ferieformer Sidetall 112 Pris kr 15,00 ISBN 82-537-1611-7
- 81/26 Aktuelle skattetall 1981 Current Tax Data Sidetall 46 Pris kr 10,00 ISBN 82-537-1610-9

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

Trykt 1981

- Nr. 81/27 Tiril Vogt: Planregnskap Ressursregnskap for fysisk planlegging Sidetall 70  
Pris kr 15,00 ISBN 82-537-1614-1
- 81/28 Figurer i publikasjoner Sidetall 115 Pris kr 15,00 ISBN 82-537-1628-1
  - 81/29 Tor Skoglund: Utprøving av modellen REGION mot fylkesfordelte nasjonalregnskapsdata for perioden 1973 - 1976 Sidetall 42 Pris kr 10,00 ISBN 82-537-1626-5
  - 81/31 Torstein Bye og Tor Eivind Høyland: Inntektsbegreper - Inntektsfordeling Sidetall 43 Pris kr 10,00 ISBN 82-537-1639-7
  - 81/32 Totalregnskap for fiske- og fangstnæringen 1977 - 1980 Sidetall 39 Pris kr 10,00 ISBN 82-537-1634-6





**Pris kr 10,00**

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

**ISBN 82-537-1639-7  
ISSN 0332-8422**