

# RAPPORTER

## OM MÅLING AV ULIKSKAP

AV  
ROLF AABERGE

STATISTISK SENTRALBYRÅ  
CENTRAL BUREAU OF STATISTICS OF NORWAY

RAPPORTER FRA STATISTISK SENTRALBYRÅ 82/9

# OM MÅLING AV ULIKSKAP

AV  
ROLF AABERGE

STATISTISK SENTRALBYRÅ  
OSLO — KONGSVINGER 1982  
ISBN 82-537-1721-0  
ISSN 0332-8422



## FORORD

Eit av hovudføremåla med levekårsstatistikken er å gi informasjon om ulikskap i levekåra. For at vi skal kunne presentere informasjon om ulikskap på ein meiningsfull måte, må vi operasjonalisere dette omgrepet. I rapporten drøftar vi kjende metodar for måling av ulikskap i ei ressursfordeling (f.eks. ei inntektsfordeling). Vi foreslår dessutan eit nytt prinsipp for konstruksjon av slike metodar. Eit av kapitla i rapporten gir ein illustrasjon av den praktiske bruken og nytten til eit sett metodar som er utvikla etter dette prinsippet.

Resultata og metodane i denne rapporten dannar grunnlaget for ein pågåande analyse av ulikskap i fordelinga til ei rekkje levekårsvariable (inntekt, butettheit, talet på sjukedagar, talet på veker arbeidsledig eller permittert, osv.). Resultata frå dette arbeidet vil bli publiserte i serien Rapportar for 1982.

Statistisk Sentralbyrå, Oslo, 19. januar 1982

Arne Øien



## INNHALD <sup>\*)</sup>

	Side
1. Innleiing .....	7
2. Ulikskapskurver .....	11
3. Ulikskapsmål .....	19
4. Gjennomsnittet som mål for ulikskap .....	22
5. Estimering av ulikskapskurver og ulikskapsmål .....	25
5.1 Estimering av ulikskapskurver .....	25
5.2 Estimering av ulikskapsmål .....	27
5.3 Estimering av fraktildifferansefunksjonen .....	29
6. Eksempel på ulikskapsanalyse av inntektsfordelingar for åra 1970 og 1976 .....	31
6.1 Personinntekt .....	31
6.2 Hushaldsinntekt .....	48
 Appendiks 1. Asymptotisk teori for empiriske ulikskapskurver og ulikskapsmål .....	 55
a. Asymptotiske resultat for empiriske ulikskapskurver	55
b. Asymptotiske resultat for estimatorane til ulik- skapsmåla A, B, C, D og G .....	67
c. Asymptotiske resultat for den empiriske fraktil- differansefunksjonen .....	74
 Appendiks 2. Formell drøfting av eigenskapar til ulikskapsmåla A, B, C, D og G. ....	 76
 Litteratur .....	 78
 Utkome i serien Rapportar frå Statistisk Sentralbyrå (RAPP) ...	 79

\*) Takk til Arne S. Andersen, John Dagsvik og Tor Haldorsen for nyttige kommentarar til manuskriptet.



## 1. INNLEIING

For å drøfte ulikskap i fordelinga av ein ressurs, er det i samfunnsvitskapane vanleg å ta utgangspunkt i likskapsfordelinga. Vi har likskap i fordelinga av ein ressurs viss og berre viss alle individa i populasjonen tek imot ein like stor del av ressursen. Sidan det eksisterer uendeleg mange forskjellige fordelingar som viser avvik frå likskapsfordelinga, blir følgjelig ulikskap eit mangfoldig og innfløkt omgrep. Dette gjer det problematisk å måle graden av ulikskap i fordelinga av ein ressurs. Dei mange og tildels svært forskjellige forslaga på mål for ulikskap frå den samfunnsvitskaplege litteraturen gir nettopp uttrykk for at ulikskap er eit fleirdimensjonalt omgrep. Hovedtyngda av desse måla er utvikla i samband med den merksemda økonomane har vigd fordelinga av inntekt. Også i denne rapporten har vi avgrensa oss til å drøfte måling av ulikskap i inntektsfordelingar. Vi vil derfor understreke at dei generelle resultatane som blir presenterte i rapporten også kan brukast til ulikskapsanalyse av andre ressursfordelingar.

Føremålet med arbeidet har vore å utvikle nye og drøfte eksisterande teknikkar for måling og analyse av ulikskap i fordelinga av ein ressurs. Vi har funne det nyttig å skilje mellom følgjande to problem når vi skal vurdere ulikskap av eit sett samanliknbare ressursfordelingar:

- (i) Gi ei detaljert samanlikning av fordelingane. Dette vil vere spesielt interessant når vi skal studere eventuelle endringar i mønsteret av ulikskap over tid.
- (ii) Ranging av fordelingane, dvs. ordne fordelingane etter graden av "ulikskap". For å gjere dette, må vi fastsetje rangeringskriterium. For eksempel kan vi ønskje eit kriterium som reflekterer bestemte trekk av ulikskap ved fordelinga. Det er hovudsakleg dette problemet dei tradisjonelle måla for ulikskap er retta mot.

Sidan den kumulative fordelingsfunksjonen ikkje har interessant økonomisk ulikskapstolking, er det naturleg å søke etter transformasjonar av fordelingsfunksjonen som har tiltrekkande tolking for å belyse problem (i). Slike transformasjonar vil vi kalle ulikskapskurver. I den økonomiske faglitteraturen er Lorenz-kurva (nærmare omtalt i kapittel 2) eit kjent eksempel på ei ulikskapskurve. Ein sentral eigenskap ved Lorenz-kurva er at den er invariant ovafor skalaendringar i den underliggjande fordelinga. I faglitteraturen er det vanleg å føreta den avgrensinga av omgrepet ulikskap som følgjer av skalainvariansprinsippet, dvs. om alle inntektene i populasjonen får same prosentvise tillegg (reduksjon) så fører ikkje dette til endringar i ulikskapen



innafor populasjonen. Ei anna viktig avgrensing av omgrepet ulikskap følgjer av overføringsprinsippet. Dette prinsippet seier at ulikskapen i populasjonen blir redusert viss vi overfører inntekt frå ein rikare til ein fattigare person og denne overføringa ikkje er så stor at mottakaren blir rikare enn givaren. Lorenz-kurva tilfredsstillar overføringsprinsippet. I klassen av transformasjonar som tilfredsstillar skalainvarians- og overføringsprinsippet er det ein-eintydig samheng mellom Lorenz-kurva og den kumulative fordelingsfunksjonen. Dermed blir Lorenz-kurva ein transformasjon av den kumulative fordelingsfunksjonen som tek vare på all tilgjengeleg informasjon om ulikskap.

I kapittel 2 vil vi presentere ei rekkje andre ulikskapskurver som alle har den omtalte skalainvarians-eigenskapen og dessutan tilfredsstillar overføringsprinsippet. Eit felles trekk ved desse ulikskapskurvene er at dei gir grunnlag for å oppnå ei meir mikroorientert tolking av ulikskapen i populasjonen enn det Lorenz-kurva gir.

Når vi skal rangere fleire fordelingar etter graden av ulikskap, vil "plasseringa" av Lorenz-kurvane i forhold til likskapskurva (den Lorenz-kurva vi får når alle individa har same inntekt) avgjere resultatet. For at dette skal ha meining, må vi avbilde Lorenz-kurva på den reelle tallinja. Ei slik avbilding blir kalla ulikskapsmål. I praksis må vi derfor velje ein funksjonal (dvs. ein funksjon av Lorenz-kurva) som avbilder Lorenz-kurva på den reelle tallinja på ein slik måte at skalainvarians- og overføringsprinsippet er oppfylt. Viss Lorenz-kurvane ikkje skjer kvarandre, vil alle funksjonalar i denne klassen gi same rangering av fordelingane. I tilfelle Lorenz-kurvane skjer kvarandre vil val av funksjonal vere avgjerande for kva slags rangering vi ender opp med. Den funksjonalen som avbilder informasjonen i Lorenz-kurva ved arealet mellom den aktuelle Lorenz-kurva og likskapskurva (heretter kalla areal-funksjonalen) peiker seg ut på grunn av tiltrekkande geometrisk tolking. Dette ulikskapsmålet (multiplisert med 2) er kjent under namnet Gini-koeffisienten og tilfredsstillar overføringsprinsippet og kravet om skalainvarians. Den enkle geometriske tolkinga av Gini-koeffisienten gjer det vidare lett å studere meir detaljert kva som har ført til den aktuelle rangeringa. Dette kan vi gjere ved å framstille dei tilsvarande Lorenz-kurvane grafisk.

For å byggje opp eit sett konkurrentar til Gini-koeffisienten, kan vi avbilde Lorenz-kurva ved andre funksjonalar som tilfredsstillar overføringsprinsippet og kravet om skalavarians. Men dermed taper vi den tiltrekkande geometriske tolkinga som areal-funksjonalen gir. Eit naturleg alternativ vil derfor vere å halde fast på areal-funksjonalen og avbilde ein-eintydige

transformasjonar av Lorenz-kurva. Vi må da kreve at transformasjonane skal ha sjølvstendige ulikskapstolkingar. Dette siste prinsippet ligg til grunn for dei forslaga på mål for ulikskap som blir presenterte i denne rapporten.

Ein sams eigenskap for ulikskapsmåla er at dei tek verdien 0 viss og berre viss alle individa i populasjonen har lik inntekt. Holdninga til kravet om at alle individa i ei gruppe eller ein populasjon må ha same inntekt for å oppnå fullstendig rettferd, varierer sjølvsagt med politisk oppfatning. Men uansett kva slags standpunkt ein tek til dette spørsmålet, vil likskap (slik som definert ovafor) fungere som eit informativt referansepunkt.

I tidlegare økonomisk faglitteratur fekk problemet med å velje mellom dei ulike ulikskapsmåla ein sentral plass; det blei da lagt vekt på om måla var reknemessig enkle å handtere, om dei hadde tilfredsstillande variasjonsområde og om dei hadde evne til å formidle informasjon frå heile fordelinga. Slike diskusjonar gav ikkje den nødvendige avklaringa av problemet med å velje eit ulikskapsmål framfor eit anna. Dalton (1920) gav eit interessant bidrag ved å peike på at ein eller annan velferdsside er knytt til eit kvart mål for ulikskap. Dalton (1920) meinte derfor det ville vere naturleg å konstruere mål for ulikskap på grunnlag av ein eller annan velferdsfunksjon. Denne tilnærminga har vore inspirasjonen til dei resultata som ligg føre i Atkinson (1969). I den artikkelen blir det føreslått å rangere inntektsfordelingane etter storleiken på forventa velferd

$$(1.1) \quad \int_0^a U(y) dF(y) \quad , \quad 0 \leq y \leq a$$

der  $F(\cdot)$  er ein kumulativ fordelingsfunksjon og  $U(\cdot)$  er ein velferdsfunksjon. For å oppnå rangering av eit sett fordelingar, må vi gjere føresetnader om forma på velferdsfunksjonen  $U(\cdot)$ ; vi må velje ein spesifisert velferdsfunksjon. I praksis vil dette seie at vi må postulere ein funksjon som reflekterer eit sett gitte velferdskrav. Problemet er at det ville bli vanskeleg å finne fram til eit sett velferdskrav som politikarar, fagfolk, osv. ville akseptere som rimelege. Det er derfor berre når  $U(F^{-1}(t))$  (for kvar  $t$ ) har sjølvstendig økonomisk tolking at kriteriet (1.1) kan betraktast som ikkje-kontroversielt. Da er vi sikra presis innsikt i kva slags grunnlag (1.1) rangerer fordelingane etter ved å studere plottet av punktane  $(t, U(F^{-1}(t)))$ .

Atkinson (1969) vel å studere ulikskapsmåla (1.1) som er avgrensa ved klassen av veksande og konkave velferdsfunksjonar  $U(\cdot)$  og påviser interessante samanhengar mellom dette settet av mål og Lorenz-kurva (Lorenz-kurva er nærmare omtalt i kapittel 2). Atkinson viser blant anna at eitkvart

mål (1.1) der  $U(\cdot)$  er ein veksande og konkav velferdsfunksjon gir same rangering av to (eller fleire) fordelingar viss og berre viss dei tilsvarende Lorenz-kurvane ikkje skjer kvarandre. Resultata seier at vi i såfall kan skilje mellom (rangere) to eller fleire fordelingar ved (1.1) utan ytterlegare spesifisering av forma på den veksande og konkave velferdsfunksjonen  $U(\cdot)$ . Følgjeleg har vi også at Gini-koeffisienten vil gi same rangering som (1.1) med  $U(\cdot)$  veksande og konkav, når Lorenz-kurvane ikkje skjer kvarandre. Vanskane oppstår viss Lorenz-kurvane skjer kvarandre. Da kan vi alltid finne ein veksande og konkav velferdsfunksjon  $U(\cdot)$  som er slik at den fordelinga som har høgst Gini-koeffisient blir rangert gunstigast ved målet (1.1).

Implisitte føresetnader om forma på velferdsfunksjonen ligg til grunn for eitkvart mål for ulikskap. Atkinson peikar på at ei rekkje av føresetnadene som ligg til grunn for dei klassiske måla er av tvilsam karakter, og føreslår derfor at vi skal angripe problemet direkte ved å spesifisere velferdsfunksjonen. Av same grunn har Sen (1972) funne det nyttig å dele dei eksisterande ulikskapsmåla opp i to klassar; dei som er baserte på ein eksplisitt velferdsfunksjon og dei som ikkje er det. Som vi har peika på tidlegare vil det alltid vere ulike oppfatningar om kva slags form velferdsfunksjonen skal ha, men den direkte framgangsmåten tjener til å understreke at eitkvart mål for ulikskap involverer vurdering av velferd.

Ved å krevje at ulikskapsmåla skal vere avbendingar av ulikskapskurver med tiltrekkande økonomisk tolking der avbendinga har ei enkel geometrisk tolking, unngår vi dei nemde innvendingane mot respektive klassiske ulikskapsmål og ulikskapsmål utvikla etter velferdsprinsippet. Vi vil da vere i stand til å gjennomskue kva slags inntektsulikskapar i populasjonen som blir oppsummerte av målet. Det einaste ulikskapsmålet frå litteraturen som fullt ut tilfredsstillar dette kravet er Gini-koeffisienten. I denne rapporten presenterer vi eit nytt sett ulikskapsmål, som saman med Gini-koeffisienten er utvikla etter eitt og same prinsipp. Desse ulikskapsmåla vil skilje seg frå kvarandre ved at dei informerer om og legg vekt på forskjellige sider ved ulikskap i fordelinga. I praksis kan vi derfor velje ut det (eller dei) av desse måla som er mest informativt i forhold til det konkrete problemet vi ønskjer å studere.

I kapitla 2 og 3 drøftar vi ulikskapskurver og ulikskapsmål som informerer om ulikskap i fordelinga av ein ressurs innafor ein gitt populasjon; vi oppnår ein beskrivelse av korleis den samla ressursen (f.eks. inntekt) fordelar seg på individa i ein populasjon. Ved å samanlikne fleire populasjonar etter denne ramma, er det graden av ulikskap innafor dei enkelte populasjonane som avgjer den rangeringa vi endar opp med.

Varierande tilgang på ressursar (samla inntekt) skaper i tillegg ulikskap mellom populasjonane. Samanlikning av gjennomsnittsinntektene i populasjonane gir oss da summarisk informasjon om ulikskap mellom desse populasjonane. I kapittel 4 vurderer vi gjennomsnittet (forventninga) som mål for ulikskap mellom populasjonar.

I kapittel 5 viser vi korleis dei teoretiske ulikskapskurvene og ulikskapsmåla kan estimerast, medan appendiks 1 inneheld asymptotiske resultat for estimatorane som blei drøfta i kapittel 5. I kapittel 6 analyserer vi ulikskapane i eit sett inntektsfordelingar for åra 1970 og 1976. I dette kapitlet går vi nærmare inn på korleis resultatata frå kapitla 2, 3, 4 og 5 kan brukast i praksis.

## 2. ULIKSKAPSKURVER

For å studere ulikskap i fordelinga av ein ressurs, er det naturleg å ta utgangspunkt i den kumulative fordelingsfunksjonen. Problemet er at det er vanskeleg å tolke avvik frå likskapsfordelinga ved ein direkte studie av denne funksjonen. For å bøte på dette vil vi søke etter transformasjonar av den kumulative fordelingsfunksjonen som har tiltrekkande ulikskapstolkning. Frå den økonomiske faglitteraturen er Lorenz-kurva (Lorenz (1905)) eit velkjent eksempel på ein slik transformasjon. La  $X$  vere ein ressursvariabel med kumulativ fordelingsfunksjon  $F$  og forventning  $EX = \mu$ . Lorenz-kurva  $L$  for  $F$  er definert ved

$$(2.1) \quad L(u) = \frac{1}{\mu} \int_0^u F^{-1}(t) dt, \quad 0 \leq u \leq 1.$$

Lorenz-kurva er opprinneleg utvikla for å informere om ulikskap i inntektsfordelingar. For å gi drøftinga av  $L(\cdot)$  og konkurrerende ulikskapskurver eit konkret innhald, skal vi heretter la  $X$  vere ein inntektsvariabel. La  $[0, \infty)$  vere definisjonsområdet til  $F$ .

Definisjonen (2.1) gir ikkje noko umiddelbart forslag til korleis Lorenz-kurva kan tolkast. I den økonomiske faglitteraturen har ein sprunge over dette problemet ved å basere tolkinga av  $L(u)$  på den naturlege estimatoren for  $L(u)$ . Ved å innføre funksjonen  $H(\cdot)$  som er definert ved

$$(2.2) \quad H(u) = E(X \mid X \leq F^{-1}(u)) = \frac{1}{u} \int_0^u F^{-1}(t) dt,$$

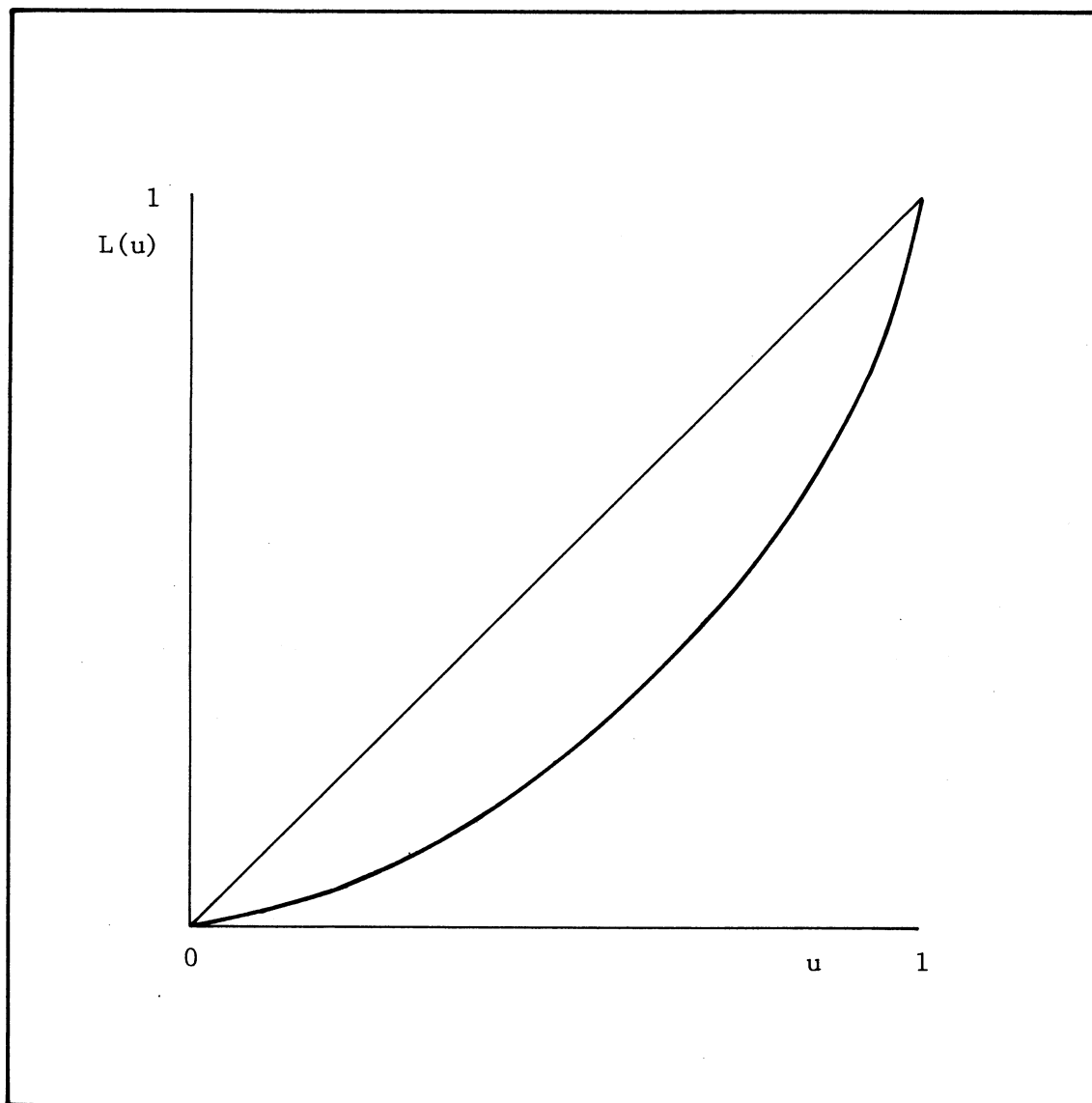
skal vi vise at det er dekning for den tradisjonelle tolkinga av Lorenz-kurva. Legg merke til at  $\mu = H(1)$ . Frå (2.1) og (2.2) får vi at  $L(u)$  kan skrivast på forma

$$(2.3) \quad L(u) = \frac{uH(u)}{H(1)} = \frac{E(NuX \mid X \leq F^{-1}(u))}{E(NX)}$$

der  $N$  er talet på individ i populasjonen. (2.3) viser at  $L(u)$  uttrykkjer forholdet mellom forventta sum av inntektene til den nedre  $100u$ -prosenten i populasjonen (inntekter som er mindre eller lik  $u$ -fraktilen til  $F$ ) og forventta totalinntekt i populasjon.

Ein viktig eigenskap ved Lorenz-kurva er at den er invariant ovafor skalatransformasjonar av  $X$ , dvs. om alle inntektene i populasjonen får same prosentvise tillegg (reduksjon) så endrar ikkje Lorenz-kurva form. Når  $F$  er likskapsfordelinga (alle individa i populasjonen får same inntekt) blir  $L(u) = u$  (diagonalen i fig. 1). Den andre ekstremfordelinga oppstår når den totale inntekta går til ein person. Da blir  $L(u) = 0$  for  $u < 1$  og  $L(1) = 1$ . For andre fordelingar blir  $L(\cdot)$  ein konveks funksjon. Lorenz-kurva informerer derfor om mindre grad av ulikskap i fordelinga av inntekt til nærare kurva ligg diagonalen i fig. 1.

Figur 1. Illustrasjon av ei Lorenz-kurve



Horisontalaksen ( $u$ ) representerer populasjonspartar som er ordna etter inntektsstorleikane, medan vertikalaksen ( $L(u)$ ) meddeler kor stor del av dei totale inntektene som tilhøyrer den  $100u$ -prosenten av inntektstakarane med lågaste inntekter.

Ved å byggje på funksjonane  $E(X | X \leq F^{-1}(\cdot))$  og  $E(X | X \geq F^{-1}(\cdot))$  har vi utvikla eit sett konkurrerande ulikskapskurver til Lorenz-kurva. Desse ulikskapskurvene gir ei meir mikroorientert innsikt i ulikskapen innafor populasjonen enn Lorenz-kurva. La ulikskapskurvene  $M(\cdot)$  og  $N(\cdot)$  vere definert ved respektive

$$(2.4) \quad M(u) = \frac{E(X | X \leq F^{-1}(u))}{EX}, \quad 0 \leq u \leq 1$$

og

$$(2.5) \quad N(u) = \frac{E(X | X \leq F^{-1}(u))}{E(X | X \geq F^{-1}(1-u))}, \quad 0 \leq u \leq 1.$$

Som vi skal sjå har både M-kurva og N-kurva ei rekkje tiltrekkande eigenskapar. Sidan

$$E(X | X \geq F^{-1}(u)) = \frac{1}{1-u} \int_u^1 F^{-1}(t) dt = \frac{1}{1-u} (\mu - \int_0^u F^{-1}(t) dt)$$

får vi frå (2.2) og (2.3) at

$$(2.6) \quad E(X | X \geq F^{-1}(u)) = \frac{\mu}{1-u} (1-L(u)).$$

Frå (2.2) og (2.3) har vi vidare at

$$E(X | X \leq F^{-1}(u)) = \frac{\mu}{u} L(u).$$

M-kurva og N-kurva kan derfor uttrykkjast ved Lorenz-kurva på følgjande måtar,

$$(2.7) \quad M(u) = \frac{L(u)}{u}$$

og

$$(2.8) \quad N(u) = \frac{L(u)}{1-L(1-u)}$$

Sidan  $\lim_{u \rightarrow 0} M(u) = \lim_{u \rightarrow 0} N(u) = 0$

er både  $M(u)$  og  $N(u)$  veldefinert i 0.

Når  $F$  er likskapsfordelinga blir  $M(u) = N(u) = 1$  for  $u > 0$  og  $M(0) = N(0) = 0$ .

Den andre ekstremfordelinga (all inntekt går til ein person) gir

$M(u) = N(u) = 0$  for  $u < 1$  og  $M(1) = N(1) = 1$ . For andre kumulative for-

delingar  $F$  tek både M-kurva og N-kurva ulike funksjonsformer, men alltid

slik at  $0 \leq M(u) \leq 1$ ,  $M(0) = 0$ ,  $M(1) = 1$  og  $0 \leq N(u) \leq 1$ ,  $N(0) = 0$ ,  $N(1) = 1$ ,

dvs. universa av M-kurver og N-kurver er begge avgrensa av eit kvadrat med sider lik 1. Til nærare ei M-kurve og ei N-kurve ligg den øvre horisontalsida i kvadratet, til mindre grad av ulikskap målt ved respektive  $M(\cdot)$  og  $N(\cdot)$  har vi i den underliggjande fordelinga  $F$ .  $M(u)$  informerer om forholdet mellom gjennomsnittsinntekta til den nedre  $100u$ -prosenten og gjennomsnittsinntekta for heile populasjonen (totalgjennomsnittet) og  $N(u)$  om forholdet mellom gjennomsnittsinntektene til den nedre og den øvre  $100u$ -prosenten. Sidan ekstremkurvene både for  $M(\cdot)$  og  $N(\cdot)$  blir 0 (når all inntekta går til ein person) og 1 (når inntekta er likt fordelt) og  $M(\cdot)$  og  $N(\cdot)$  dessutan har ei presis økonomisk tyding, blir det i praksis enkelt å tolke dei grafiske framstillingane av desse funksjonane. I klassen av funksjonar  $M(\cdot)$  representerer den rette linja  $u$  (diagonalen i kvadratet med hjørnene  $(0,0)$ ,  $(0,1)$ ,  $(1,0)$  og  $(1,1)$ ) den uniforme fordelinga (sjå tabell 1). Sidan det er lett å lese avvik frå ei rett linje, gir dette opphav til interessant ekstra informasjon. I kapittel 6 viser vi korleis dette kan utnyttast i samband med ein studie av inntektsdata frå 1970 og 1976. Kommentaraane til dei grafiske framstillingane av M-kurva og N-kurva for ulike populasjonar gir dessuten innsikt i informasjonsevna til begge desse ulikskapskurvene. Som vi skal sjå i kapittel 6 vil den grafiske framstillinga av to M-kurver og to N-kurver skilje betre mellom ulikskapane i dei to underliggjande fordelingane enn den grafiske framstillinga av dei to tilsvarande Lorenz-kurvene.

Vi skal no presentere to andre konkurrentar til Lorenz-kurva. La ulikskapskurvene  $P_1(\cdot)$  og  $P_2(\cdot)$  vere definert ved respektive

$$(2.9) \quad P_1(u) = \frac{E(X \mid X \leq F^{-1}(u))}{E(X \mid X \geq F^{-1}(u))}, \quad 0 \leq u \leq 1$$

og

$$(2.10) \quad P_2(u) = \frac{E(X \mid X \geq F^{-1}(u))}{EX}, \quad 0 \leq u \leq 1$$

Ved å kombinere (2.2), (2.3) og (2.6) får vi at  $P_1$ -kurva og  $P_2$ -kurva kan uttrykkjast ved Lorenz-kurva på følgjande måtar,



$$(2.11) \quad P_1(u) = \left(\frac{1-u}{1-L(u)}\right) \left(\frac{L(u)}{u}\right)$$

og

$$(2.12) \quad P_2(u) = \frac{1-L(u)}{1-u}$$

$$\text{der } \lim_{u \rightarrow 0} P_1(u) = 0, \quad \lim_{u \rightarrow 1} P_1(u) = \frac{\mu}{F^{-1}(1)}$$

$$\text{og } \lim_{u \rightarrow 1} P_2(u) = \frac{F^{-1}(1)}{\mu}$$

$P_1(u)$  informerer om forholdet mellom gjennomsnittsinntektene til den nedre  $100u$ -prosenten og den øvre  $100(1-u)$ -prosenten av inntektsfordelinga. Dersom vi for eksempel har at  $P_1(1/4) = 1/5$ , så tjener den nedre fjerdedelen av populasjonen gjennomsnittleg femteparten av det den øvre  $3/4$ -delen tjener.  $P_1(u)$  har variasjonsområde  $[0, \mu/F^{-1}(1)]$ , medan  $P_2(u)$  har variasjonsområde  $[1, F^{-1}(1)/\mu]$ .  $P_2(u)$  informerer om forholdet mellom gjennomsnittsinntekta til den øvre  $100(1-u)$ -prosenten og totalgjennomsnittet. Både  $P_1$ -kurva og  $P_2$ -kurva blir lik 1 når  $F$  er likskapsfordelinga. For andre fordelingsfunksjonar  $F$  tek både  $P_1$ -kurva og  $P_2$ -kurva funksjonsformer forskjellig fra 1, men alltid slik at

$$0 \leq P_1(u) \leq 1 \quad 0 \leq u \leq 1$$

og

$$1 \leq P_2(u) \leq \frac{1}{1-u} \quad 0 \leq u < 1.$$

Til nærare ei  $P_1$ -kurve og ei  $P_2$ -kurve ligg den horisontale linja 1, til mindre grad av ulikskap målt ved respektive  $P_1(\cdot)$  og  $P_2(\cdot)$  har vi i den underliggjande fordelinga  $F$ . Legg merke til at universet av  $P_2$ -kurver er avgrensa av funksjonen  $1/(1-u)$  og linja 1 medan universet av  $P_1$ -kurver er avgrensa av kvadratet med hjørnene  $(0,0)$ ,  $(0,1)$ ,  $(1,0)$  og  $(1,1)$ .

Det følger frå (2.7), (2.8), (2.9) og (2.10) at det er ein-eintydig samanheng mellom respektive M-kurva, N-kurva,  $P_1$ -kurva og  $P_2$ -kurva og Lorenz-kurva. Sidan Lorenz-kurva er invariant ovafor skalatransformasjonar av  $X$ , følgjer det derfor at også  $M(\cdot)$ ,  $N(\cdot)$ ,  $P_1(\cdot)$  og  $P_2(\cdot)$  oppfyller skalainvarians-prinsippet.

Ulikskapskurvene  $L$ ,  $M$ ,  $N$ ,  $P_1$  og  $P_2$  tilfredsstiller alle overførings-prinsippet, dvs. om vi føretek ei inntektsoverføring  $b$  frå ein person med inntekt  $x_1$  til ein person med lægre inntekt  $x_2$  og denne overføringa er slik at  $x_2 \leq x_1 - b$ , da får vi redusert ulikskapen i fordelinga. Dette prinsippet blei introdusert av Dalton (1920).

I tabell 1 gir vi funksjonsformene til Lorenz-kurva, M-kurva, N-kurva,  $P_1$ -kurva og  $P_2$ -kurva for fire kjende fordelingsklasser.

Tabell 1. Ulikskapskurvene  $L(\cdot)$ ,  $M(\cdot)$ ,  $N(\cdot)$ ,  $P_1(\cdot)$  og  $P_2(\cdot)$  generert av nokre vanlege kumulative fordelingsfunksjonar

Ulikskapskurver	Kumulativ fordelingsfunksjon	Lik $F(x) = \begin{cases} 1, & x > a \\ 0, & x < a \end{cases}$	Ekspensiell $F(x) = 1 - e^{-\lambda x}, x > 0$	Uniform $F(x) = \frac{x}{\theta}, 0 < x < \theta$	Pareto $F(x) = 1 - \left(\frac{\alpha}{x}\right)^\alpha$ $x > \alpha, \alpha > 1$
$L(u)$		$u$	$u + (1-u) \log(1-u)$	$u^2$	$1 - (1-u) \frac{(\alpha-1)}{\alpha}$
$M(u)$		$1$	$1 + \frac{(1-u)}{u} \log(1-u)$	$u$	$\frac{1 - (1-u) \frac{(\alpha-1)}{\alpha}}{u}$
$N(u)$		$1$	$\frac{u + (1-u) \log(1-u)}{u(1-\log u)}$	$\frac{u}{2-u}$	$\frac{(1-\alpha)}{u^\alpha} - \left(\frac{1-u}{u}\right) \frac{(\alpha-1)}{\alpha}$
$P_1(u)$		$1$	$\frac{u + (1-u) \log(1-u)}{u(1-\log(1-u))}$	$\frac{u}{1+u}$	$\frac{\frac{1}{(1-u)^\alpha} - (1-u)}{u}$
$P_2(u)$		$1$	$1 - \log(1-u)$	$1+u$	$(1-u)^{-\frac{1}{\alpha}}$

Vi vil nemne 4 andre ulikskapskurver (flater) utan å gi ei nærare drøfting av deira eigenskapar.

$$(2.13) \quad R(u) = \frac{1}{EX} [E(X \mid X \geq F^{-1}(u)) - E(X \mid X \leq F^{-1}(u))], \quad 0 \leq u \leq 1.$$

$R(u)$  er forskjellen mellom gjennomsnittsinntektene til dei som tjener respektive meir og mindre enn  $u$ -fraktilen (den nedre  $100u$ -prosenten) relativt til totalgjennomsnittet.

$$(2.14) \quad S(u) = \frac{E(X \mid F^{-1}(u) \leq X \leq F^{-1}(1-u))}{EX}, \quad 0 \leq u \leq \frac{1}{2}.$$

$S(u)$  er forholdet mellom gjennomsnittsinntekta til dei som tjener meir enn den nedre  $100u$ -prosenten samt mindre enn den øvre  $100u$ -prosenten og totalgjennomsnittet.

$$(2.15) \quad T(u,v) = \frac{E(X \mid F^{-1}(u) \leq X \leq F^{-1}(v))}{EX}, \quad 0 \leq u \leq v \leq 1.$$

$T(u,v)$  er forholdet mellom gjennomsnittsinntekta til dei som tjener meir enn  $u$ -fraktilen samt mindre enn  $v$ -fraktilen og totalgjennomsnittet.

$$(2.16) \quad W(u,v) = \frac{E(X \mid X \leq F^{-1}(u))}{E(X \mid X \geq F^{-1}(v))}, \quad 0 \leq u, v \leq 1.$$

$W(u,v)$  uttrykker forholdet mellom gjennomsnittsinntektene til dei som tjener mindre enn den nedre  $100u$ -prosenten og dei som tjener meir enn den øvre  $100(1-v)$ -prosenten i populasjonen.

Ved å nytte definisjonen av betinga forventning og (2.1), (2.2), (2.3) og (2.6), finn vi følgjande samanheng mellom desse ulikskapskurvene (flatene) og Lorenz-kurva,

$$(2.17) \quad R(u) = \frac{u-L(u)}{u(1-u)},$$

$$(2.18) \quad S(u) = \frac{1}{1-2u} (L(1-u) - L(u)),$$

$$(2.19) \quad T(u,v) = \frac{1}{v-u} (L(v) - L(u))$$

og

$$(2.20) \quad W(u,v) = \left(\frac{L(u)}{u}\right) \left(\frac{1-v}{1-L(v)}\right).$$

Legg merke til at både M-kurva, N-kurva og  $P_1$ -kurva er med i klassen (2.20).

Vi har at

$$L(u) = u W(u,0) ,$$

$$M(u) = W(u,0) ,$$

$$N(u) = W(u,1-u) ,$$

$$P_1(u) = W(u,u) ,$$

og

$$P_2(u) = \frac{1}{1-u} (1-u W(u,0)).$$

### 3. ULIKSKAPSMÅL

Den vanlege angrepsmåten ved samanlikning av inntektsfordelingar er å rangere fordelingane etter verdiane på eit bestemt ulikskapsmål. Vi har definert og drøfta dette omgrepet i innleiinga. Tre av dei mest brukte ulikskapsmåla er Gini-koeffisienten, variasjonskoeffisienten (det relative standardavviket) og det relative forventningsavviket. Sidan slike mål er summariske er det viktig å kjenne deira tolking og eigenskapar for å bruke dei på ein fornuftig måte.

Gini-koeffisienten  $G$  er definert som arealet mellom diagonalen og Lorenz-kurva (sjå fig. 1) multiplisert med 2, dvs.  $G$  er definert ved

$$(3.1) \quad G = 2 \int_0^1 (u-L(u)) du$$

og er derfor ei oppsummering med enkel geometrisk tolking av den ulikskapsinformasjonen som Lorenz-kurva formidler.  $G$  har variasjonsområdet  $[0,1]$  og viser ein mindre grad av ulikskap til mindre verdi den tek.

Mange av dei klassiske ulikskapsmåla er konstruerte på grunnlag av statistiske mål for spreiding. For å kunne gjennomskue kva slags ulikskapar i populasjonen eit slikt mål oppsummerer, vil vi undersøke om det er mogleg å tolke det ved arealvurderingar av ei ulikskapskurve. Vi skal vise at det relative forventningsavviket og variasjonskoeffisienten kan tolkast på denne måten, men at dei to ulikskapskurvene som vi kan knytte til desse to måla ikkje har spesielt tiltrekkande økonomisk tolking samanlikna med ulikskapskurvene som blir gitt i kapittel 2. Det relative forventningsavviket  $\Delta$  er definert ved

$$(3.2) \quad \Delta = \frac{E |X - \mu|}{\mu}$$

for  $F$  definert på  $[0, \infty)$ .

(3.2) gir

$$\Delta = \frac{1}{\mu} \int_0^{\infty} |x - \mu| dF(x) = \int_0^1 \frac{|F^{-1}(u) - \mu|}{\mu} du .$$

Dette viser at  $\Delta$  oppsummerer informasjonen frå ulikskapskurva

$$Q_1(u) = \frac{|F^{-1}(u) - \mu|}{\mu} , \quad 0 \leq u \leq 1 .$$

$Q_1(u)$  gir absoluttavviket mellom  $u$ -fraktilen og forventninga relativt til forventninga.  $\Delta$  oppsummerer derfor alle desse relative avvika ved arealet mellom  $u$ -aksen og ulikskapskurva  $Q_1(\cdot)$ .

Variasjonskoeffisienten  $V$  er definert ved

$$(3.3) \quad V = \frac{(E(X - \mu)^2)^{\frac{1}{2}}}{\mu} .$$

Sidan det er ein-eintydig samanheng mellom  $V$  og  $V^2$  for  $V \geq 0$  kan vi velje å studere  $V^2$ . (3.3) gir

$$V^2 = \frac{1}{\mu^2} \int_0^{\infty} (x - \mu)^2 dF(x) = \int_0^1 \left( \frac{F^{-1}(u) - \mu}{\mu} \right)^2 du$$

som viser at  $V^2$  oppsummerer informasjonen frå ulikskapskurva

$$Q_2(u) = \left( \frac{F^{-1}(u) - \mu}{\mu} \right)^2 , \quad 0 \leq u \leq 1 .$$

$Q_2(u)$  gir det relative kvadratavviket mellom  $u$ -fraktilen og forventninga.  $V^2$  oppsummerer derfor informasjonen frå alle dei relative kvadratavvika ved arealet mellom  $u$ -aksen og ulikskapskurva  $Q_2(\cdot)$ .

Både  $Q_1$ -kurva og  $Q_2$ -kurva viser ein mindre grad av ulikskap til nærare dei ligg  $u$ -aksen. Ei viktig innvending mot både  $Q_1$ -kurva og  $Q_2$ -kurva, er at dei ikkje tilfredsstillar overføringsprinsippet. Etter ei inntekts-overføring frå ein rikare til ein fattigare person vil den nye  $Q_1$ -kurva skjere den gamle. Det tilsvarande forholdet har vi for  $Q_2$ -kurva, medan Lorenz-kurva og dei andre ulikskapskurvene som blei drøfta i kapittel 2 vil vise ei mindre grad av ulikskap etter ei slik overføring. Når det gjeld dei avleidde måla  $\Delta$  og  $V^2$ , er det lett å vise at  $V^2$  tilfredsstillar og at  $\Delta$  ikke tilfredsstillar overføringsprinsippet. Effekten på  $V^2$  av ei overføring  $b$  frå ein person med inntekt  $b_1$  til ein person med inntekt  $b_2$  der  $b$  er slik at  $b_2 \leq b_1 - b$ , er proporsjonal med  $[(b_1 - b_2) - b]b$ , dvs. ei ubetydeleg

inntektsoverføring frå ein svært rik til ein svært fattig person kan føre til ein betydeleg reduksjon i verdien på  $V^2$ .

I kapittel 2 peika vi på at ulikskapskurvene  $L(\cdot)$ ,  $M(\cdot)$ ,  $N(\cdot)$ ,  $P_1(\cdot)$  og  $P_2(\cdot)$  tilfredsstilte både skalainvarians- og overføringsprinsippet. På grunn av samanhengen mellom Lorenz-kurva ( $L$ ) og Gini-koeffisienten (sjå (3.1)) følgjer det opplagt at også Gini-koeffisienten tilfredsstiller både skalainvarians- og overføringsprinsippet. Vi skal no presentere 4 ulikskapsmål som er arealavbildingar av ulikskapskurvene  $M(\cdot)$ ,  $N(\cdot)$ ,  $P_1(\cdot)$  og  $P_2(\cdot)$  og som tilfredsstiller skalainvarians- og overføringsprinsippet. La  $A$ ,  $B$ ,  $C$  og  $D$  vere definert ved respektive

$$(3.4) \quad A = \int_0^1 (1-M(u))du ,$$

$$(3.5) \quad B = \int_0^1 (1-N(u))du ,$$

$$(3.6) \quad C = \int_0^1 (1-P_1(u))du$$

og

$$(3.7) \quad D = \int_0^1 (P_2(u)-1)du.$$

Vi ser at  $A$ ,  $B$  og  $C$  er definert som areala mellom den horisontale linja 1 (likskapskurva) og dei respektive ulikskapskurvene  $M(\cdot)$ ,  $N(\cdot)$  og  $P_1(\cdot)$ , medan  $D$  er definert som arealet mellom  $u$ -aksen og  $P_2$ -kurva. Alle desse fire måla viser ein mindre grad av ulikskap til mindre verdier dei tek. Både  $A$ ,  $B$  og  $C$  har variasjonsområdet  $[0,1]$ , medan  $D$  har variasjonsområdet  $[0,\infty)$ . I tabell 2 gir vi ei oversikt over kva slags verdier  $A$ ,  $B$ ,  $C$ ,  $D$  og  $G$  tek for ekstremfordelingane og for den uniforme fordelinga. La I representere likskapsfordelinga og III den andre ekstremfordelinga (all inntekta går til ein person). II er den uniforme fordelinga  $(0,a)$ .

Tabell 2. Verdiane til ulikskapsmåla  $A$ ,  $B$ ,  $C$ ,  $D$  og  $G$  for den uniforme fordelinga II og dei to ekstremfordelingane I og III.

Ulikskaps- mål \ Fordelings- funksjon	G	A	B	C	D
I	0	0	0	0	0
II	.333	.500	.614	.693	.500
III	1	1	1	1	$\infty$

For å bruke ulikskapsmåla på ein fornuftig måte, har vi tidlegare understreka kor viktig det er å kjenne deira tolking og eigenskapar. Ulikskapsmåla A, B, C, D og G har det til felles at dei er utvikla etter eitt og same prinsipp (sjå innleiinga). På ein lettfatteleg måte kan kvart av desse ulikskapsmåla drøftast i forhold til ei ulikskapskurve med sjølvstendig tolking. Forskjellige eigenskapar ved ulikskapskurvene vil komme til uttrykk i eigenskapane til dei respektive ulikskapsmåla.

I appendiks 2 viser vi at G vil leggje størst vekt på inntektsoverføringar i den sentrale delen (toppen) av fordelinga, medan A, B og C vil gi størst vekt til overføringar i den nedre delen av fordelinga. Vi føreset at inntektsoverføringane skjer etter overføringsprinsippet og at inntektsfordelinga er eintoppa. Vidare har vi at A legg større vekt enn både B og C på overføringar i den nedre delen av fordelinga.

D vil leggje størst vekt på inntektsoverføringar i den øvre delen av fordelinga.

#### 4. GJENNOMSNIETTET SOM MÅL FOR ULIKSKAP

I kapitla 2 og 3 har vi utvikla og drøfta skalainvariante ulikskapskurver og ulikskapsmål som informerer om ulikskap i fordelinga av ein ressurs innafor ein gitt populasjon. Vi oppnår dermed ein beskrivelse av korleis den samla ressursen (f.eks. inntekt) som tilhøyrrer eller blir teke i mot av ein populasjon, fordeler seg på individa i denne populasjonen. Ved å samanlikne fleire populasjonar etter denne ramma, er det graden av ulikskap innafor dei enkelte populasjonane som avgjer rangeringa vi endar opp med. Inntektsnivået (forventa inntekt) for dei enkelte populasjonane får inga betydning for rangeringsresultatet.

I dei situasjonane der det er naturleg å betrakte eit sett populasjonar som konkurrentar om ein total ressurs (f.eks. samla inntekt), vil det vere interessant å finne ut korleis inntektene blir fordelt både mellom og innafor (del)populasjonane. Dei omtalte kurvene og måla i kapitla 2 og 3 kan vi bruke for å skaffe oss informasjon om ulikskap innafor kvar enkelt av dei aktuelle (del)populasjonane. Ved å samanlikne gjennomsnittsinntektene for (del)populasjonane, oppnår vi summarisk informasjon om ulikskap mellom (del)populasjonane. Sidan same forskjellen i gjennomsnittsinntekt mellom to populasjonar kan oppstå for uendeleg mange sett av fordelingsfunksjonar, er det problematisk å bruke gjennomsnittet (forventninga) som mål for ulikskap mellom populasjonar. For å illustrere dette skal vi spesielt studere inntektene

for populasjonane gitt ved personar under 25 år i respektive 1970 (X) og 1976 (Y). Gjennomsnittsinntektene for denne gruppa var kr 10 500 i 1970, og kr 18 700 i 1976, dvs. i gjennomsnitt steig inntektene med kr 8 200. Ved å betrakte funksjonen  $\Lambda(\cdot)$  (definert ved (4.1)), oppnår vi bakgrunnsinformasjon for endringa i gjennomsnittsinntekt. Vi vil kalle denne funksjonen for fraktildifferansefunksjonen.

$$(4.1) \quad \Lambda(t) = G^{-1}(t) - F^{-1}(t), \quad 0 \leq t \leq 1$$

der  $G(y)$ ,  $0 \leq y \leq a_2$  og  $F(x)$ ,  $0 \leq x \leq a_1$  er dei kumulative inntektsfordelingane for respektive 1976 og 1970.

La  $\mu_1 = EX$  og  $\mu_2 = EY$ . Sidan

$$\mu_2 - \mu_1 = \int_0^{a_2} y \, dG(y) - \int_0^{a_1} x \, dF(x) = \int_0^1 G^{-1}(t) dt - \int_0^1 F^{-1}(t) dt$$

gir (4.1)

$$(4.2) \quad \mu_2 - \mu_1 = \int_0^1 \Lambda(t) dt,$$

som viser at forskjellen i gjennomsnitt (forventning) kan tolkast ved ei arealoppsummering av den ulikskapsinformasjonen som fraktildifferansefunksjonen formidler.  $\Lambda(t)$  uttrykkjer forskjellen mellom 100 t-prosent fraktilane for dei to fordelingane G og F. Når vi analyserer inntektsfordelingar, blir fraktildifferansane  $\Lambda(t)$ ,  $10 t = 1, 2, \dots, 10$  ofte kalla desilinntektsdifferansar. Ved hjelp av tabell 23 frå Inntektsstatistikk 1970 og tabell 31 frå Inntektsstatistikk 1976 (NOS-publikasjonar) finn vi følgjande estimat for desilinnntektsdifferansane:

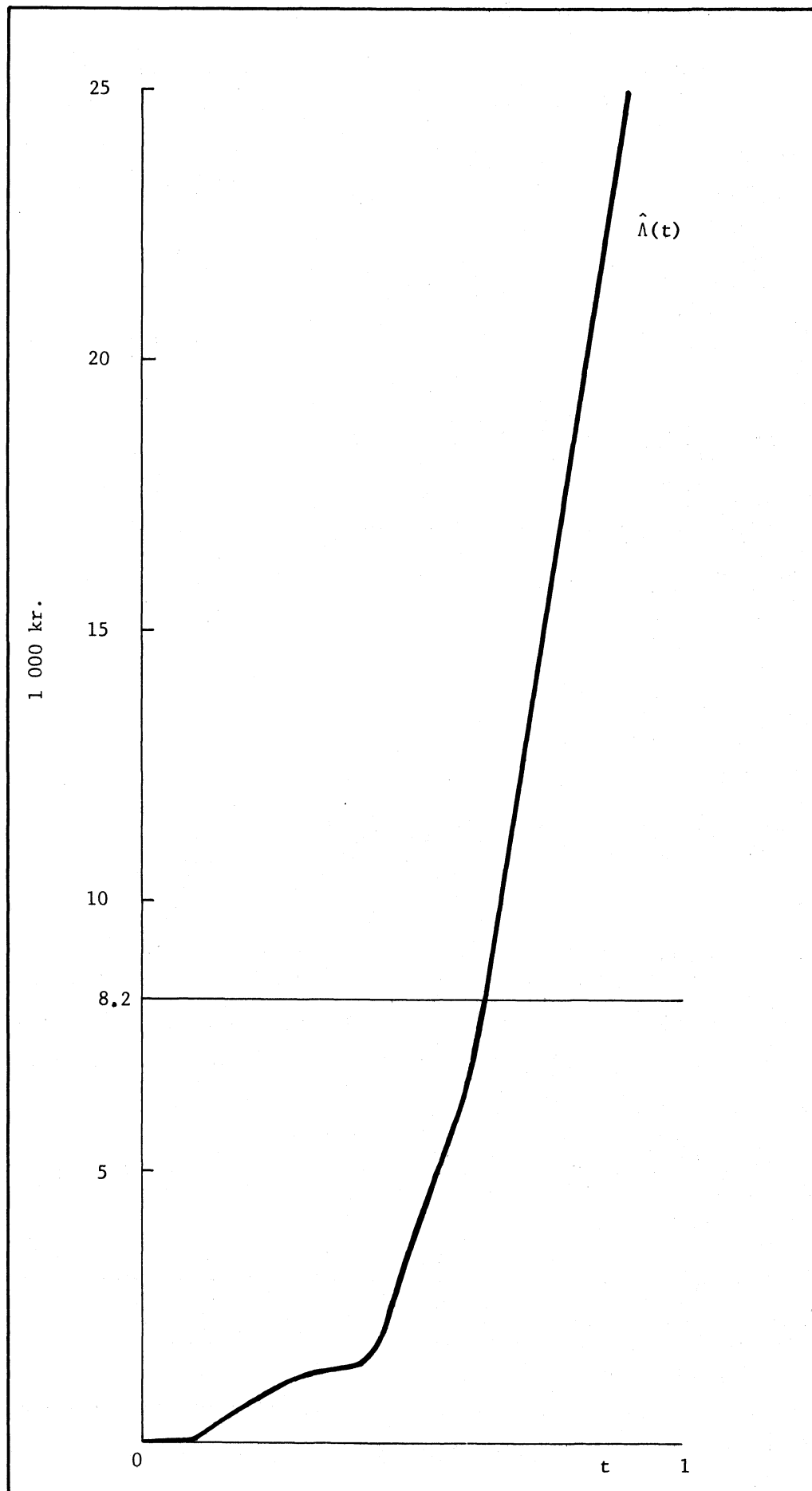
Tabell 3. Estimat for desilinntektsdifferansar mellom 1976 og 1970 for inntektsgruppa personar under 25 år (1000 kr).

t	.10	.20	.30	.40	.50	.60	.70	.80	.90
$\hat{\Lambda}(t)$	0	.70	1.19	1.30	3.73	6.27	12.60	19.57	24.98

På grunnlag av punkttestimata i tabell 3 har vi framstilt fraktildifferansefunksjonen grafisk i figur 2. t-aksen gir uttrykk for populasjonsandelar.



Figur 2. Fraktildifferansefunksjonen mellom åra 1976 og 1970 for personar under 25 år. (1 000 kr)



Dersom alle inntektene hadde auka like mykje ville plottet av  $\hat{\Lambda}(t)$  ha falle saman med den horisontale linja 8.2. I staden viser det aktuelle plottet at  $\hat{\Lambda}(t)$  veks kraftig med veksande  $t$ . Dei små inntektene har hatt ein ubetydeleg auke frå 1970 til 1976, medan dei store inntektene har auka kraftig. Det er stigningen i dei store inntektene som fører til at gjennomsnittsinntekta nærmast blir dobla frå 1970 til 1976. Forskjellen i gjennomsnittsinntekt mellom åra 1970 og 1976 blir det same som arealet under funksjonen  $\hat{\Lambda}(\cdot)$ .

Vi har no demonstrert informasjonsverdien til gjennomsnittet og fraktildifferansefunksjonen ved å studere inntektsutviklinga for personar under 25 år. Vel så interessant vil det vere å nytte denne reidskapen saman med metodane frå kapitla 2 og 3 for analyse av inntektsdata til ulike populasjonar på same tidspunkt. Dette blir gjort i kapittel 6.

## 5. ESTIMERING AV ULIKSKAPSKURVER OG ULIKSKAPSMÅL

I dette kapittel skal vi vise korleis ulikskapskurvene i kapittel 2, måla i kapittel 3 og fraktildifferansefunksjonen i kapittel 4 kan estimerast både på grunnlag av kontinuerlege og grupperte data.

La  $X_1, X_2, \dots, X_n$  vere eit sett uavhengige stokastiske variable med kumulativ fordelingsfunksjon  $F$  og la  $F_n$  vere den empiriske fordelingsfunksjonen, dvs.  $F_n(x) = (\# X_i \leq x)/n$ .

### 5.1. Estimering av ulikskapskurver

Da det er naturleg å bruke den empiriske fordelingsfunksjonen  $F_n$  til å estimere  $F$ , oppnår vi eit konsistent estimat for  $H(u)$  ved å erstatte  $F$  med  $F_n$  i uttrykket (2.2) for  $H(u)$ . Dette gir

$$(5.1) \quad \hat{H}(u) = \frac{1}{u} \int_0^u F_n^{-1}(t) dt$$

$$\text{der} \quad F_n^{-1}(t) = \inf\{x : F_n(x) \geq t\},$$

som vi kan skrive på forma

$$(5.2) \quad \hat{H}(u) = \frac{1}{nu} \left( \sum_{j=1}^i X_{(j)} + (nu-i)X_{(i+1)} \right), \quad u \in \left[ \frac{i}{n}, \frac{i+1}{n} \right)$$

$$\text{der} \quad X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

Frå (5.2) får vi

$$(5.3) \quad \hat{H}\left(\frac{i}{n}\right) = \frac{1}{i} \sum_{j=1}^i X_{(j)}, \quad i=1, 2, \dots, n.$$

Legg merke til at

$$\hat{H}(1) = \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j.$$

Ved å erstatte  $H(u)$  med  $\hat{H}(u)$  i uttrykket (2.3) for  $L(u)$ , får vi følgende konsistente estimat for  $L(u)$ ,

$$(5.4) \quad \hat{L}(u) = \frac{u\hat{H}(u)}{\hat{H}(1)}$$

sidan både  $M(u)$ ,  $N(u)$ ,  $P_1(u)$  og  $P_2(u)$  kan uttrykkjast ved  $L(u)$  (sjå kapittel 2), oppnår vi konsistente estimat for desse ulikskapskurvene ved å erstatte  $L(u)$  med  $\hat{L}(u)$  i uttrykka for dei. Dette gir

$$(5.5) \quad \hat{M}(u) = \frac{\hat{L}(u)}{u},$$

$$(5.6) \quad \hat{N}(u) = \frac{\hat{L}(u)}{1-\hat{L}(1-u)},$$

$$(5.7) \quad \hat{P}_1(u) = \left(\frac{1-u}{u}\right) \left(\frac{\hat{L}(u)}{1-\hat{L}(u)}\right)$$

og

$$(5.8) \quad \hat{P}_2(u) = \frac{1-\hat{L}(u)}{1-u}.$$

(5.3) og (5.4) gir spesielt for  $u = i/n$ ,

$$(5.9) \quad \hat{L}\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^i X_{(j)}}{\sum_{j=1}^n X_j}, \quad i=1, 2, \dots, n$$

$$(5.10) \quad \hat{M}\left(\frac{i}{n}\right) = \frac{\bar{X}(i)}{\bar{X}}, \quad i=1, 2, \dots, n$$

der

$$\bar{X}(i) = \frac{1}{i} \sum_{j=1}^i X_{(j)}$$

$$(5.11) \quad \hat{N}\left(\frac{i}{n}\right) = \frac{\bar{X}(i)}{\bar{X}'(i)}, \quad i=1, 2, \dots, n$$

der

$$\bar{X}'(i) = \frac{1}{i} \sum_{j=1}^i X_{(n+1-j)}$$

$$(5.12) \quad \hat{P}_1\left(\frac{i}{n}\right) = \frac{\bar{X}(i)}{\bar{X}'_{(n+1-i)}}, \quad i=1, 2, \dots, n$$

og

$$(5.13) \quad \hat{P}_2\left(\frac{i}{n}\right) = \frac{\bar{X}'_{(n+1-i)}}{\bar{X}}, \quad i=1, 2, \dots, n$$

$\hat{M}(u)$ ,  $\hat{N}(u)$ ,  $\hat{P}_1(u)$  og  $\hat{P}_2(u)$  kan alle reknast ut på grunnlag av  $\hat{L}(u)$ . Som vi skal sjå i kapittel 6, fører dette til enkle rekneoperasjonar når data er tilgjengeleg på tilsvarende måte som ved desiltabellane i NOS-publikasjonane Inntektsstatistikk 1970 og 1976.

## 5.2 Estimering av ulikskapsmål

I dette avsnittet skal vi gi konsistente estimatorar for ulikskapsmåla A, B, C, D og G.

Ved å erstatte  $L(u)$  med  $\hat{L}(u)$  i uttrykket (3.1) for G, får vi følgende konsistente estimator for Gini-koeffisienten G,

$$(5.14) \quad \hat{G} = 2 \int_0^1 (u - \hat{L}(u)) du = \frac{\sum_{i=1}^n (X_{(i)} - \frac{2}{n+1} \sum_{j=1}^i X_{(j)})}{\sum_{j=1}^n X_j}$$

I litteraturen er det vanleg å operere med

$$(5.15) \quad \tilde{G} = \frac{\sum_{i=1}^{n-1} i(n-1) (X_{(i+1)} - X_{(i)})}{(n-1) \sum_{j=1}^n X_j} = \frac{n+1}{n-1} \hat{G}$$

som estimator for Gini-koeffisienten. Sidan  $\tilde{G}$  berre skiljer seg frå  $\hat{G}$  ved faktoren  $(n+1)/(n-1)$ , blir valet mellom desse to estimatorane ei smaksak.

På same måte som for  $G$ , får vi konsistente estimatorar for  $A$ ,  $B$ ,  $C$  og  $D$  ved å erstatte dei teoretiske ulikskapskurvene med dei tilsvarande empiriske i uttrykka for kvart av desse måla,

$$(5.16) \quad \hat{A} = \int_0^1 (1-\hat{M}(u))du = 1 - \frac{\frac{1}{n} \sum_{i=1}^n \bar{X}(i)}{\bar{X}},$$

$$(5.17) \quad \hat{B} = \int_0^1 (1-\hat{N}(u))du = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\bar{X}(i)}{\bar{X}'(i)}$$

$$(5.18) \quad \hat{C} = \int_0^1 (1-\hat{P}_1(u))du = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\bar{X}(i)}{\bar{X}'_{(n+1-i)}}$$

og

$$(5.19) \quad \hat{D} = \int_0^1 (\hat{P}_2(u)-1) du = \frac{\frac{1}{n} \sum_{i=1}^n \bar{X}'_{(n+1-i)}}{\bar{X}} - 1.$$

**Merknad:** Alle desse estimatorane kan reknast direkte ut på grunnlag av kjennskapen til dei empiriske ulikskapskurvene. La  $\Gamma(\cdot)$  vere ei ulikskapskurve og la  $\hat{\Gamma}(u_i)$ ,  $i=0, 1, \dots, k+1$  vere  $k+2$  estimerte punkt for  $\Gamma(\cdot)$  der  $u_0=0$  og  $u_{k+1}=1$ . Da kan vi estimere parameteren

$$(5.20) \quad \gamma = \int_0^1 \Gamma(u)du$$

ved

$$\begin{aligned}
 (5.21) \quad \tilde{\gamma} &= \sum_{i=0}^k [\hat{\Gamma}(u_i)(u_{i+1} - u_i) \\
 &+ \frac{1}{2}(\hat{\Gamma}(u_{i+1}) - \hat{\Gamma}(u_i))(u_{i+1} - u_i)] \\
 &= \frac{1}{2} \sum_{i=0}^k (\hat{\Gamma}(u_{i+1}) + \hat{\Gamma}(u_i))(u_{i+1} - u_i)
 \end{aligned}$$

$\tilde{\gamma}$  er ein konsistent estimator for  $\gamma$ , men den vil overestimere  $\gamma$  når  $\Gamma(u)$  er ein konveks funksjon og  $k$  er liten, dvs. når vi har delt materialet inn i få grupper. Gastwirth (1972) har vist at feilen ved å estimere Gini-koeffesienten på denne måten er neglisjerbar når  $k > 20$ .

La  $u_i = i/10$ ,  $i=0, 1, \dots, 10$  og anta  $\hat{\Gamma}(0) = 0$  og  $\hat{\Gamma}(1) = 1$ .

Da blir

$$(5.22) \quad \tilde{\gamma} = \frac{1}{20} (2 \sum_{i=1}^9 \hat{\Gamma}(\frac{i}{10}) + 1).$$

I Byrået er det vanleg å publisere både person- og hushaldsinntekt etter desilar. Tabellar med slik informasjon gjer det mogleg å estimere måla A, B og G direkte ved formelen (5.22). For å kunne estimere måla C og D ved (5.21) treng vi i tillegg eit estimat for  $(F^{-1}(1))$ , dvs. vi treng eit anslag for den største inntekta i populasjonen (den øverste desilinntekta). I NOS-publikasjonane om inntektsstatistikk blir det berre publisert tal for dei 9 minste desilinntektene. Av denne grunn har vi i kapittel 6 avgrensa oss til å bruke ulikskapsmåla G, A og B i ulikskapsanalysen av inntektsfordelingar for 1970 og 1976.

### 5.3 Estimering av fraktildifferansefunksjonen

La  $Y_1, Y_2, \dots, Y_m$  vere eit sett uavhengige stokastiske variable med kumulativ fordelingsfunksjon  $G$ . Utvala  $(X_1, X_2, \dots, X_n)$  og  $(Y_1, Y_2, \dots, Y_m)$  er stokastisk uavhengige. La  $G_m$  vere den empiriske fordelingsfunksjonen  $G_m(y) = (\# Y_i \leq y)/m$ .

Det er naturleg å nytte

$$(5.23) \quad \hat{\Lambda}(t) = G_m^{-1}(t) - F_n^{-1}(t)$$

der

$$G_m^{-1}(t) = \inf \{y: G_m(y) \geq t\},$$

til å estimere  $\Lambda(t)$ .

Sidan  $F_n$  og  $G_m$  er konsistente estimat for  $F$  og  $G$ , blir  $\hat{\Lambda}(t)$  eit konsistent estimat for  $\Lambda(t)$ . Vi vil kalle  $\hat{\Lambda}$  den empiriske fraktildifferansefunksjonen.

For  $t \in [\frac{i}{n}, \frac{i+1}{n}]$  blir  $F_n^{-1}(t) = \inf \{x : F_n(x) \geq t\} = X_{(i)}$  og

$G_m^{-1}(t) = \inf \{y : G_m(y) \geq t\} = Y_{(\langle mt \rangle)}$  der  $\langle v \rangle =$  minste heile tal  $\geq v$  og

$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(m)}$  er settet av dei ordna  $Y$ -observasjonane.

Dermed får vi

$$(5.24) \quad \hat{\Lambda}(t) = Y_{(\langle mt \rangle)} - X_{(i)}, \quad t \in [\frac{i}{n}, \frac{i+1}{n}] .$$

Når utvala er like store blir

$$\hat{\Lambda}(t) = Y_{(i)} - X_{(i)}, \quad t = [\frac{i}{n}, \frac{i+1}{n}], \quad m=n.$$

Når funksjonen  $\hat{\Lambda}(t)$  blir plotta for alle  $t$  er det betre å nytte gjennomsnitta av venstre inversen og høgre inversen i staden for venstre inversen. Høgre inversen  $T^{-I}$  av ein funksjon  $T$  er definert ved

$$T^{-I}(v) = \sup \{u : T(u) \leq v\} .$$

I så fall vil vi plotte funksjonen

$$(5.25) \quad \tilde{\Lambda}(t) = \frac{1}{2} [G_m^{-1}(t) + G_m^{-I}(t) - F_n^{-1}(t) - F_n^{-I}(t)]$$

som for  $t \in [\frac{i}{n}, \frac{i+1}{n}]$  reduserer seg til uttrykket

$$(5,25) \quad \tilde{\Lambda}(t) = \frac{1}{2} [Y_{(\langle mt \rangle)} + Y_{([\langle mt \rangle + 1])} - X_{(i)} - X_{(i+1)}]$$

der  $[v] =$  største heile tal  $\leq v$ .

Når utvala er like store blir

$$\tilde{\Lambda}(t) = \frac{1}{2} [Y_{(i)} + Y_{(i+1)} - X_{(i)} - X_{(i+1)}], \quad t \in [\frac{i}{n}, \frac{i+1}{n}], \quad m=n.$$

**Merknad:** Når data er tilgjengeleg på tilsvarande måte som ved desilinntekts-tabellane i NOS-publikasjonane Inntektsstatistikk 1970 og 1976, blir utrekninga av  $\hat{\Lambda}(t)$  for  $t = .10, .20, \dots, .90$  enkel. Desse funksjonsverdiane blir lik differansane mellom parvise desilinntekter for dei to aktuelle populasjonane (sjå tabell 3).

## 6. EKSEMPEL PÅ ULIKSKAPSANALYSE AV INNTEKTSFORDELINGAR FOR ÅRA 1970 OG 1976

Fordelinga av inntekt blir sett på som eit vesentleg uttrykk for velferdsvariasjonar blant folk. I Norge har ein f.eks. nytta progressiv skattleggjing som eit virkemiddel for å redusere den velferdsulikskap som kjem til uttrykk i fordelinga av nettoinntekt (inntekt før skatt). Effekten av dette kan vi måle ved å analysere fordelingane av nettoinntekt og disponibel inntekt (inntekt etter skatt) på den måten vi foreslår i denne rapporten.

Vesentleg på grunn av relativ høg inflasjon og skattefrie rentefrådrag har det etter kvart blitt svært gunstigt å ta opp lån til ulike formål. Etter som låneevna blant anna er avhengig av bruttoinntekta, vil ein studie i fordelinga av bruttoinntekt gi eit vel så interessant bilde på velferdsulikskap som det fordelinga av nettoinntekt gir. Mangel på tilgjengeleg data om bruttoinntekt hindrar oss i å få belyst dette utifrå aktuelle bruttoinntektsfordelingar.

På grunnlag av dei inntektsopplysningane som er tilgjengelege i NOS-publikasjonane Inntektsstatistikk 1970 og 1976, skal vi ved hjelp av ulikskapskurvene  $L(\cdot)$ ,  $M(\cdot)$  og  $N(\cdot)$  og dei tilhøyrande ulikskapsmåla studere fordelinga av personinntekt før skatt (nettoinntekt) innafor fem aldersgrupper og fordelinga av hushaldsinntekt før og etter skatt (nettoinntekt og disponibel inntekt) for alle hushald og for 3 grupper som er bestemt ved talet på inntektstakarar i hushaldet. Fordelingar av disponibel inntekt inngår ikkje i den offisielle statistikken, men er tilgjengeleg som maskintabellar i Statistisk Sentralbyrå.

For dei lesarane som ikkje har sett seg grundig inn i kapitla 1, 2 og 3, finn vi det nødvendig å understreke at dei omtalte metodane frå desse kapitla er instrument for å gjere funn om ulikskap innafor populasjonar. I dette kapitlet viser vi korleis dei kan brukast utan å gjere forsøk på å forklare eller gi årsaker til funna.

I samband med dataanalysar, er det viktig å gi kvalitetsvurderingar av estimerte parametrar og kurver. I appendiks 1 presenterer vi det nødvendige verktøyet som skal til for å føreta slike kvalitetsvurderingar. Men mangelen på dataprogram gjer at vi må nøye oss med å publisere estimerte ulikskapskurver og ulikskapsmål.

### 6.1. Personinntekt

Frå Inntektsstatistikk 1970 skal vi gjere bruk av tabell 19. Den gir fordelinga av personinntekt før skatt (nettoinntekt) etter alder og desilar. Vi gir att denne tabellen i tabell 4 nedafor.



Tabell 4. Personinntekt i grupper for alder etter desilar (1970).  
Andeler.

Desil nr.	Alder	Under 25 år	25-39 år	40-59 år	60-69 år	70 år og over
	1		.001	.008	.001	.014
2		.012	.030	.031	.035	.053
3		.026	.053	.049	.043	.056
4		.044	.077	.068	.056	.061
5		.069	.093	.086	.075	.069
6		.097	.106	.100	.095	.081
7		.125	.119	.114	.116	.093
8		.158	.134	.130	.137	.112
9		.192	.157	.156	.165	.149
10		.276	.223	.256	.264	.294
I alt		1.000	1.000	1.000	1.000	1.000

La  $Z_i$  vere den delen av den totale inntekta som tilhøyrrer desil nr.  $i$ .  
Frå dei tabellerte  $Z_i$ -verdiane (tabell 4), kan vi rekne ut Lorenz-kurve-  
verdiar av følgjande enkle samanheng

$$(6.1) \quad \hat{L}(i/10) = \sum_{j=1}^i Z_j, \quad i = 1, 2, \dots, 9.$$

Ved å gjere bruk av dei funksjonelle samanhengane mellom  $\hat{M}(u)$ ,  $\hat{N}(u)$  og  $\hat{L}(u)$   
(kap. 5.1), finn vi vidare  $\hat{M}(i/10)$  og  $\hat{N}(i/10)$  for  $i = 1, 2, \dots, 9$ . Dei  
tabellerte verdiane for dei empiriske ulikskapskurvene  $\hat{L}$ ,  $\hat{M}$  og  $\hat{N}$  er rekna  
ut på grunnlag av tabell 4 ovafor og tabell 36 frå Inntektsstatistikk 1976.

Tabell 5. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva  
for personar under 25 år. Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
		L(u)	1970	.001	.013	.039	.084	.152	.249	.374
	1976	.001	.010	.029	.059	.109	.188	.305	.476	.698
M(u)	1970	.01	.07	.13	.21	.30	.42	.53	.66	.80
	1976	.01	.05	.10	.15	.22	.31	.44	.60	.78
N(u)	1970	.004	.03	.06	.11	.18	.27	.39	.54	.73
	1976	.003	.02	.04	.07	.12	.20	.31	.48	.70

Tabell 6. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for personar i alderen 25-39 år. Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	1970	.008	.037	.091	.168	.261	.367	.486	.620	.777
	1976	.001	.017	.058	.130	.224	.336	.462	.605	.771
M(u)	1970	.08	.19	.30	.42	.52	.61	.69	.78	.86
	1976	.01	.09	.19	.33	.45	.56	.66	.76	.86
N(u)	1970	.04	.10	.18	.27	.35	.44	.53	.65	.78
	1976	.004	.04	.11	.20	.29	.39	.49	.62	.77

Tabell 7. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for personar i alderen 40-59 år. Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	1970	.010	.041	.091	.158	.244	.344	.459	.589	.745
	1976	.004	.029	.074	.141	.230	.337	.458	.593	.753
M(u)	1970	.10	.21	.30	.40	.49	.57	.65	.74	.83
	1976	.04	.15	.25	.35	.46	.56	.65	.74	.84
N(u)	1970	.04	.10	.17	.24	.32	.41	.51	.61	.75
	1976	.02	.07	.14	.21	.30	.39	.49	.61	.76

Tabell 8. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for personar i alderen 60-69 år (1970) og 60-66 år (1976). Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	1970	.014	.049	.092	.148	.223	.318	.434	.571	.736
	1976	.010	.041	.082	.138	.214	.314	.436	.576	.740
M(u)	1970	.14	.25	.31	.37	.45	.53	.62	.71	.82
	1976	.10	.21	.27	.35	.43	.52	.62	.72	.82
N(u)	1970	.05	.11	.16	.22	.29	.37	.48	.60	.75
	1976	.04	.10	.15	.20	.27	.36	.47	.60	.75

Tabell 9. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for personar over 69 år (1970) og over 66 år (1976). Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	1970	.032	.084	.141	.201	.270	.351	.444	.557	.706
	1976	.018	.065	.122	.180	.243	.317	.407	.520	.678
M(u)	1970	.32	.42	.47	.50	.54	.59	.63	.70	.78
	1976	.18	.33	.41	.45	.49	.53	.58	.65	.75
N(u)	1970	.11	.19	.25	.31	.37	.44	.52	.61	.73
	1976	.06	.14	.21	.26	.32	.39	.46	.56	.69

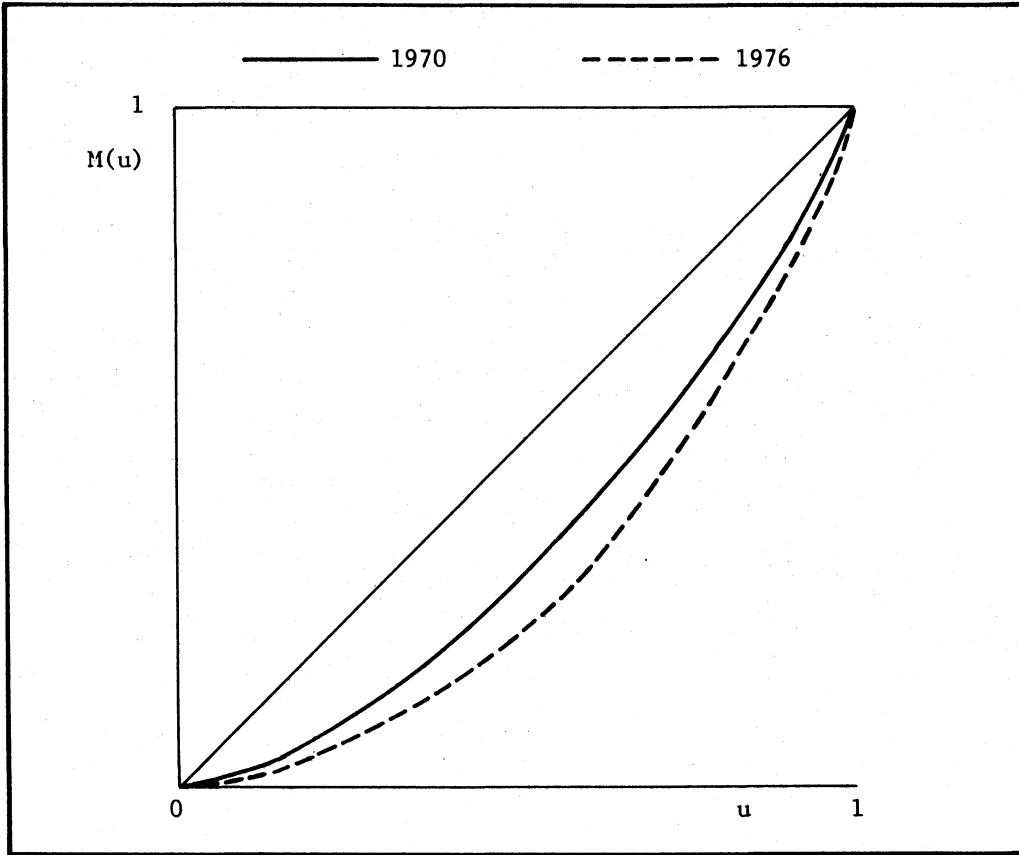
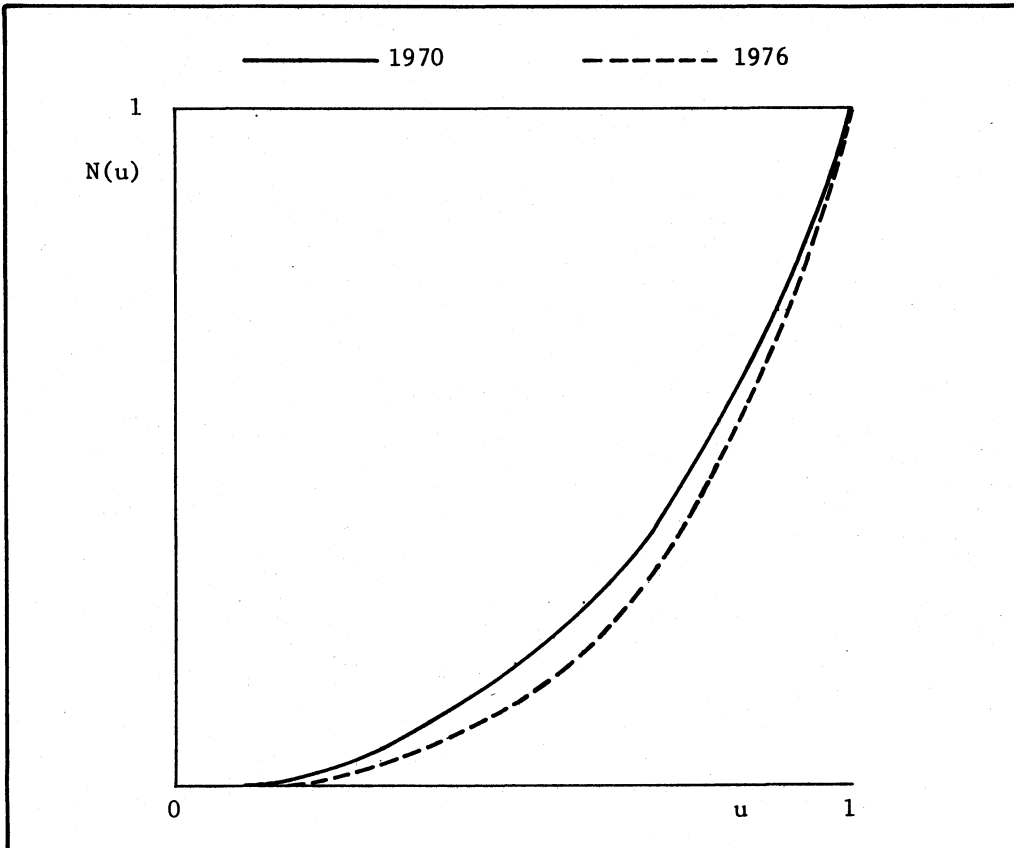
Tabell 10. Gjennomsnittsinntekter for fem aldersgrupper i 1970 og 1976

Aldersgruppe år	Under 25 år	25-39 år	40-59 år	60-69 år (1970) 60-66 år (1976)	Over 69 år (1970) Over 66 år (1976)
1970	10 500	24 000	25 600	19 100	12 700
1976	19 500	43 500	48 500	41 300	23 300

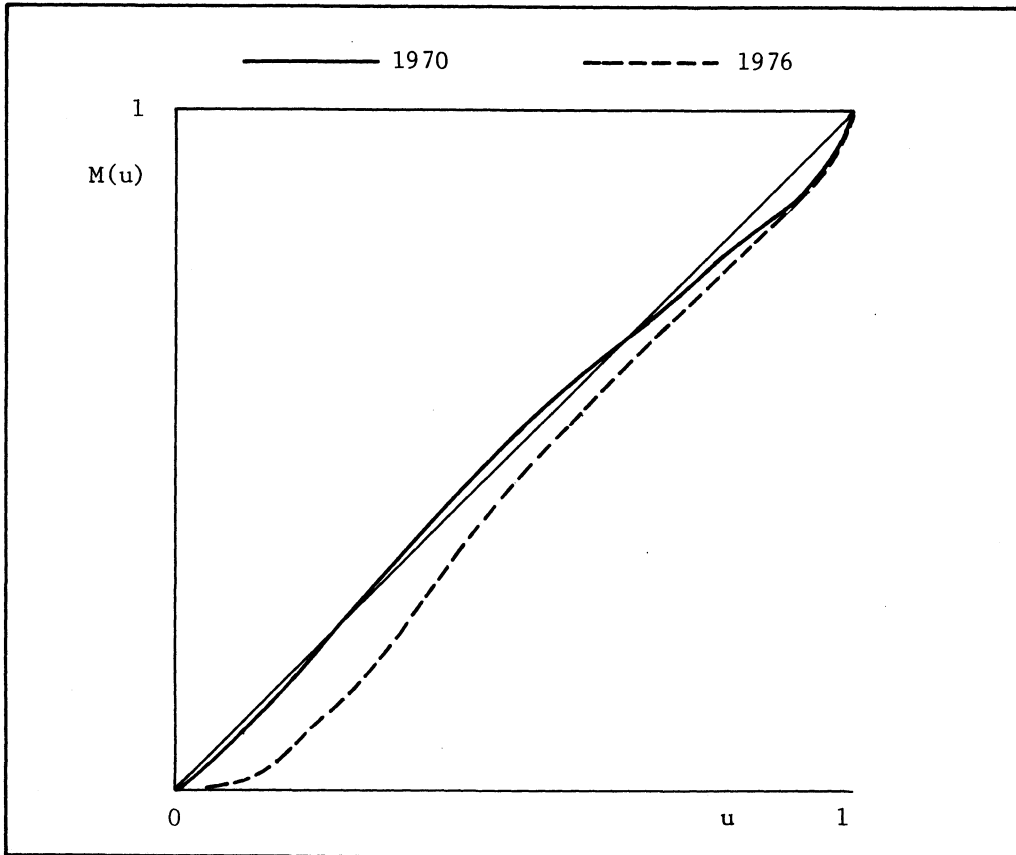
Kvart enkelt av tala i tabellane ovafor er estimat for parametrar med presis økonomisk tyding. For ein fast  $u$  kan vi sjå på  $L(u)$ ,  $M(u)$  og  $N(u)$  som parametrar. I samband med drøftinga av eigenskapar for ulikskapskurvene i kapittel 2, omtalte vi også den økonomiske tolkninga av desse parametrane. Vi vil derfor nøye oss med å kommentere nokre få av tala i tabellane ovafor.

Frå tabell 5 får vi at  $M(.50)=.30$  i 1970. Det vil seie at den nedre halvparten av inntektstakarane under 25 år hadde ei gjennomsnittleg nettoinntekt som var i underkant av ein tredjedel av totalgjennomsnittet. I 1976 hadde den tilsvarande gruppa av inntektstakarar ei gjennomsnittleg nettoinntekt som var i underkant av ein fjerdedel av totalgjennomsnittet,  $M(.50)=.22$ . Ved å betrakte  $N(u)$  får vi eit bilde på forholdet mellom dei minste og dei største nettoinntektene. Vi ser at  $N(.10)$  er lik .004 og .003 for respektive 1970 og 1976. Det vil seie at den nedre 10-prosenten av inntektstakarane under 25 år i gjennomsnitt har 4 promille av gjennomsnittleg nettoinntekt til den øvre 10-prosenten i 1970 og 3 promille i 1976. Hvis vi ser på det tilsvarande forholdet for aldersgruppa 25-39 år, finn vi frå tabell 6 4 prosent i 1970 og 4 promille i 1976.

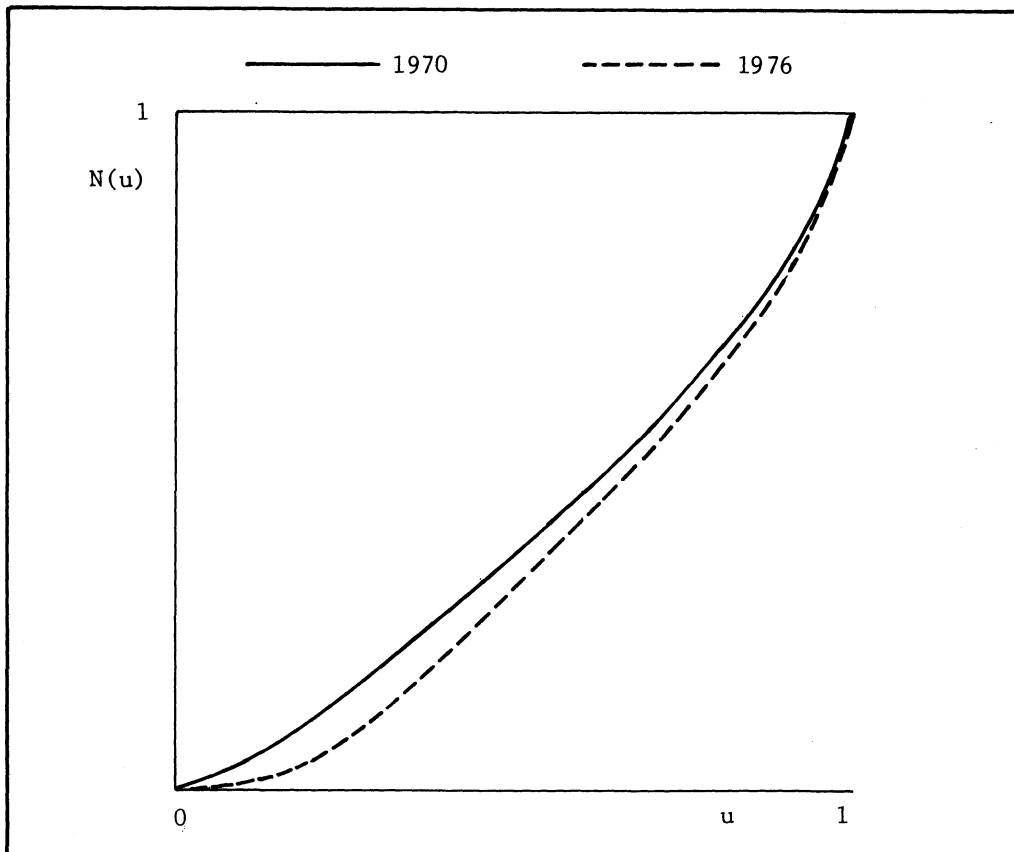
Ved å framstille dei empiriske ulikskapskurvene  $\hat{M}(\cdot)$  og  $\hat{N}(\cdot)$  grafisk, vil vi få eit meir oversiktleg bilde av den innsikta tabellane 5, 6, 7, 8 og 9 gir. Det ideelle ville ha vore å plotte funksjonane  $\hat{M}(\cdot)$  og  $\hat{N}(\cdot)$  på grunnlag av råmaterialet, men mangel på dataprogram gjer at vi må nøye oss med å basere den grafiske framstillinga på dei estimerte punktane frå tabellane ovafor.

Figur 3. Ulikskapskurver  $M(u)$  for personar under 25 år. NettoinntektFigur 4. Ulikskapskurver  $N(u)$  for personar under 25 år. Nettoinntekt

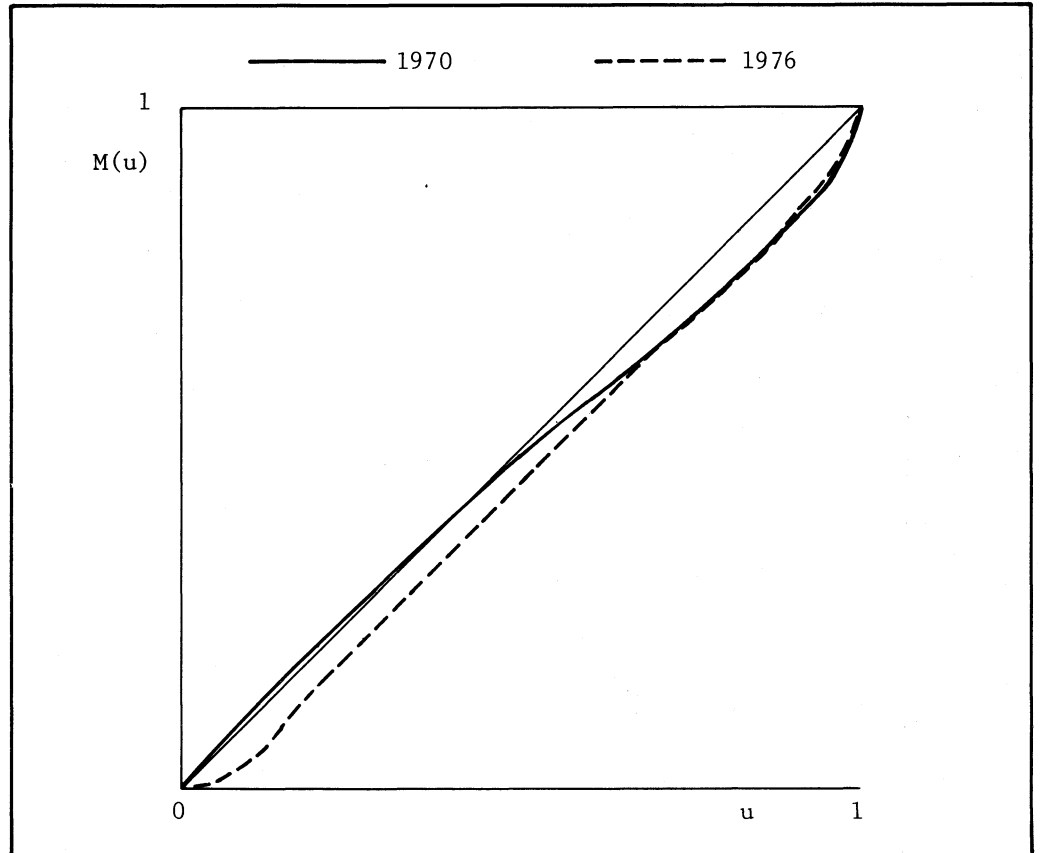
Figur 5. Ulikskapskurver  $M(u)$  for personar i alderen 25-39 år.  
Nettoinntekt



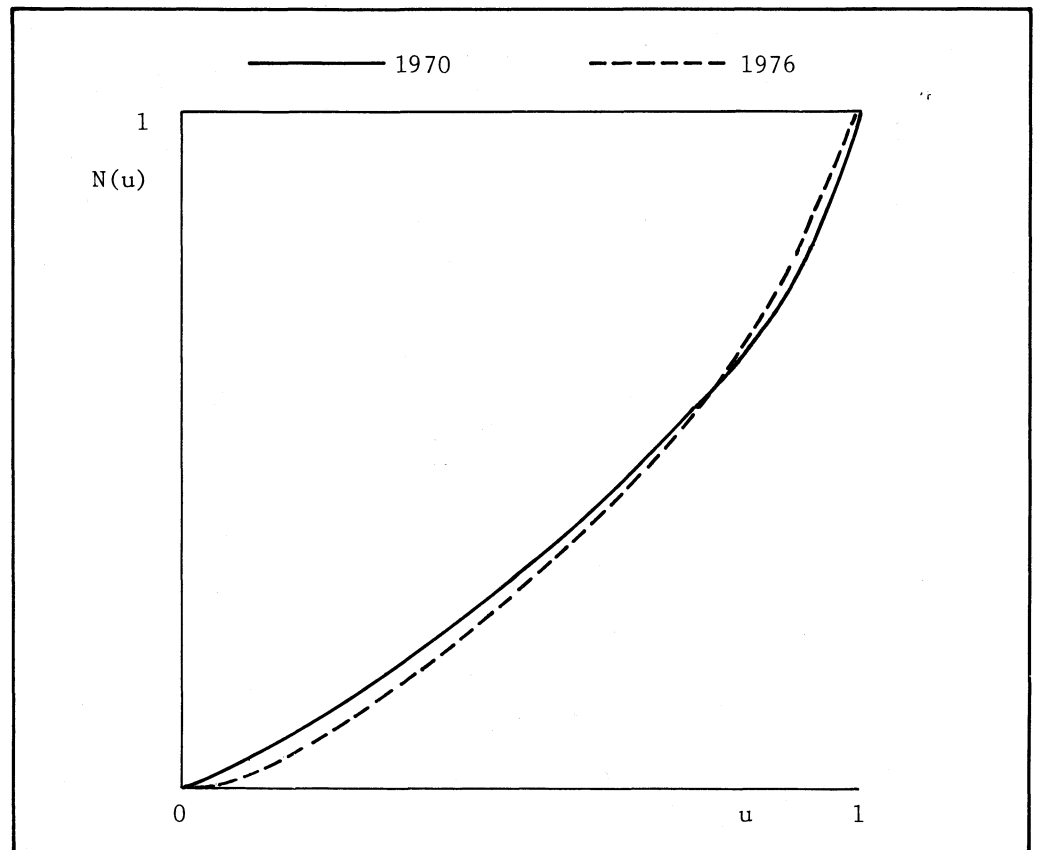
Figur 6. Ulikskapskurver  $N(u)$  for personar i alderen 25-39 år.  
Nettoinntekt



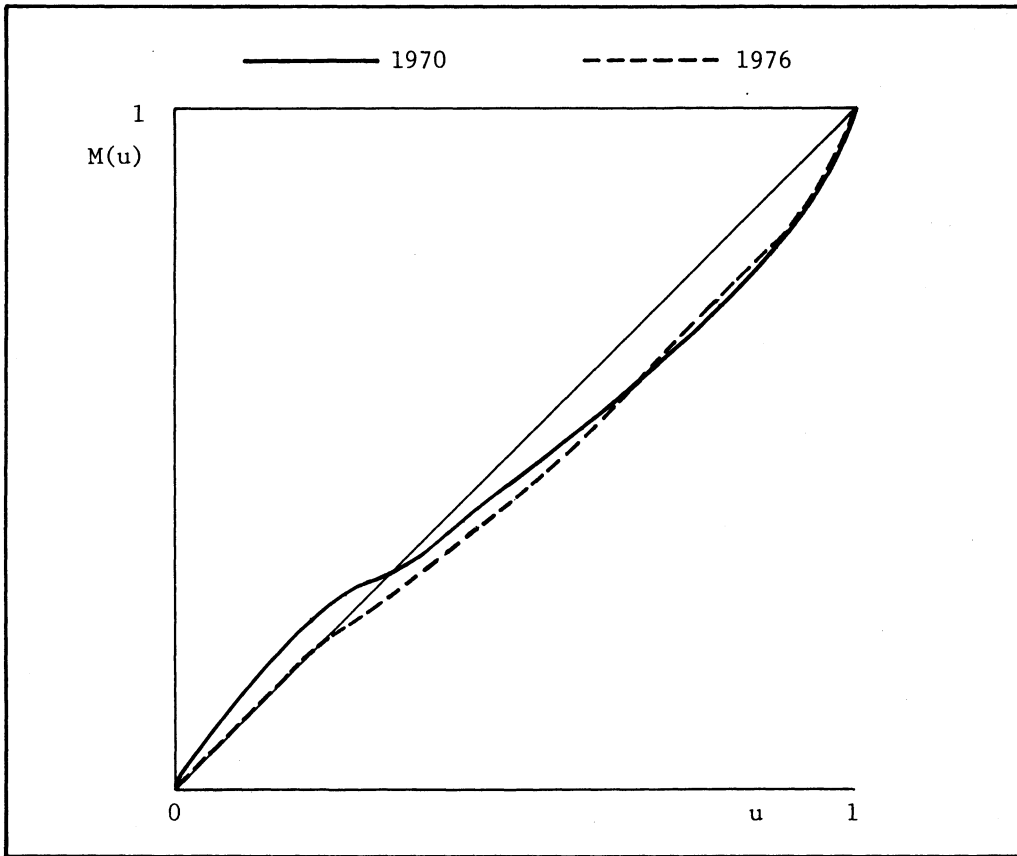
Figur 7. Ulikskapskurver  $M(u)$  for personar i alderen 40-59 år.  
Nettoinntekt



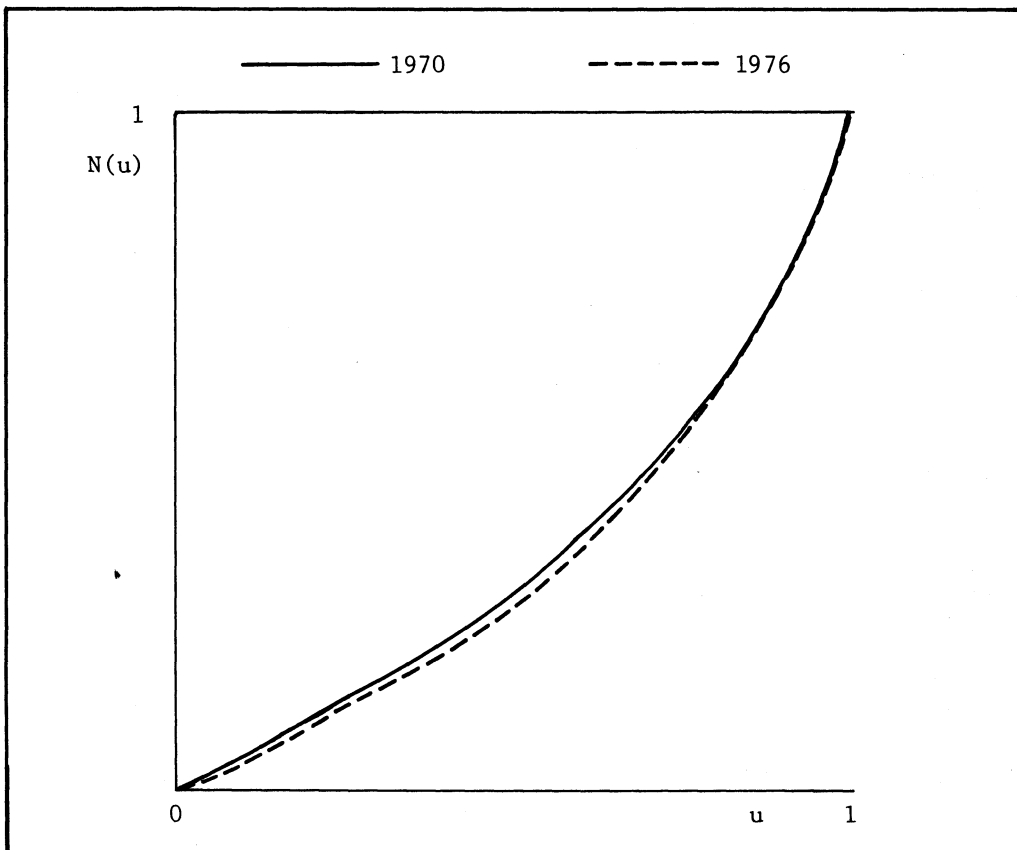
Figur 8. Ulikskapskurver  $N(u)$  for personar i alderen 40-59 år.  
Nettoinntekt



Figur 9. Ulikskapskurver  $M(u)$  for personer i alderen 60-69 år (1970) og 60-66 år (1976). Nettoinntekt

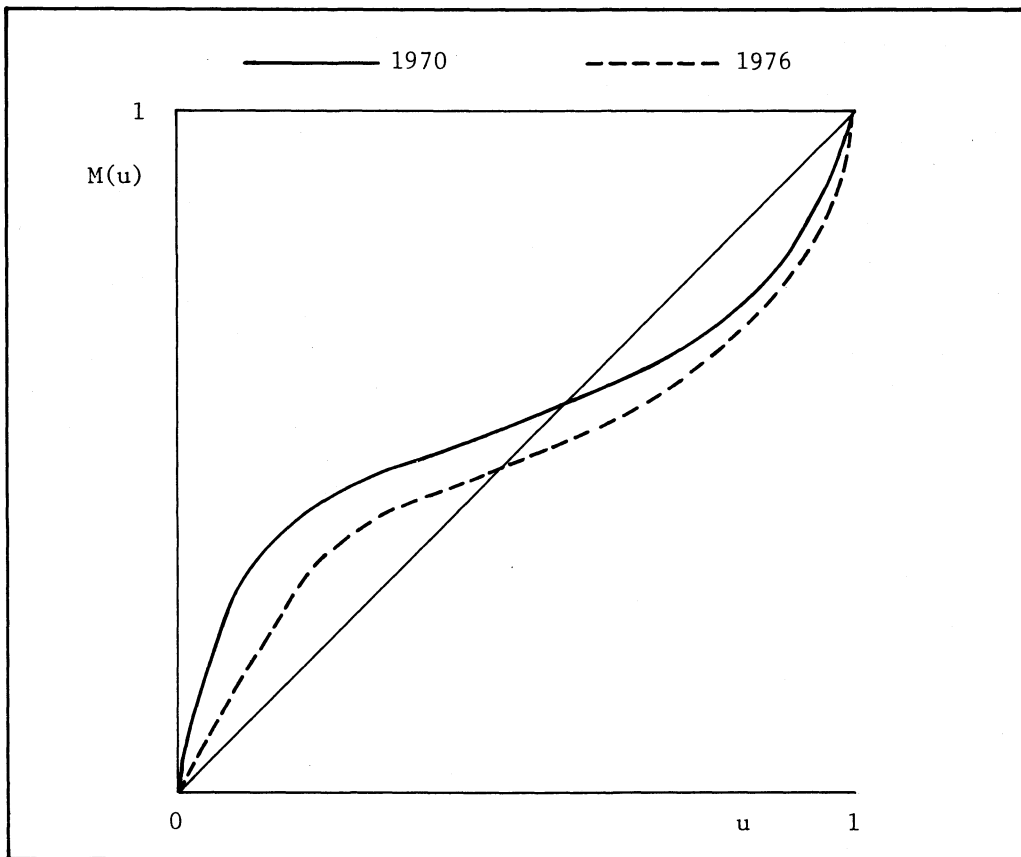


Figur 10. Ulikskapskurver  $N(u)$  for personer i alderen 60-69 år (1970) og 60-66 år (1976). Nettoinntekt

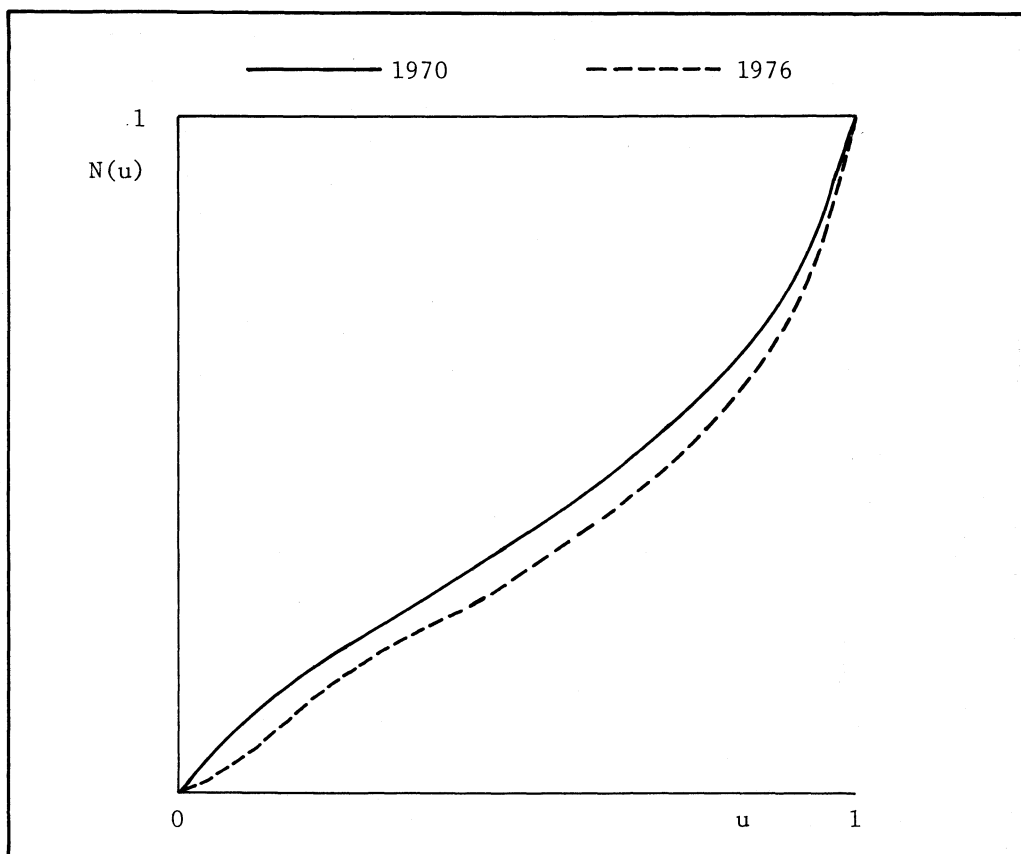




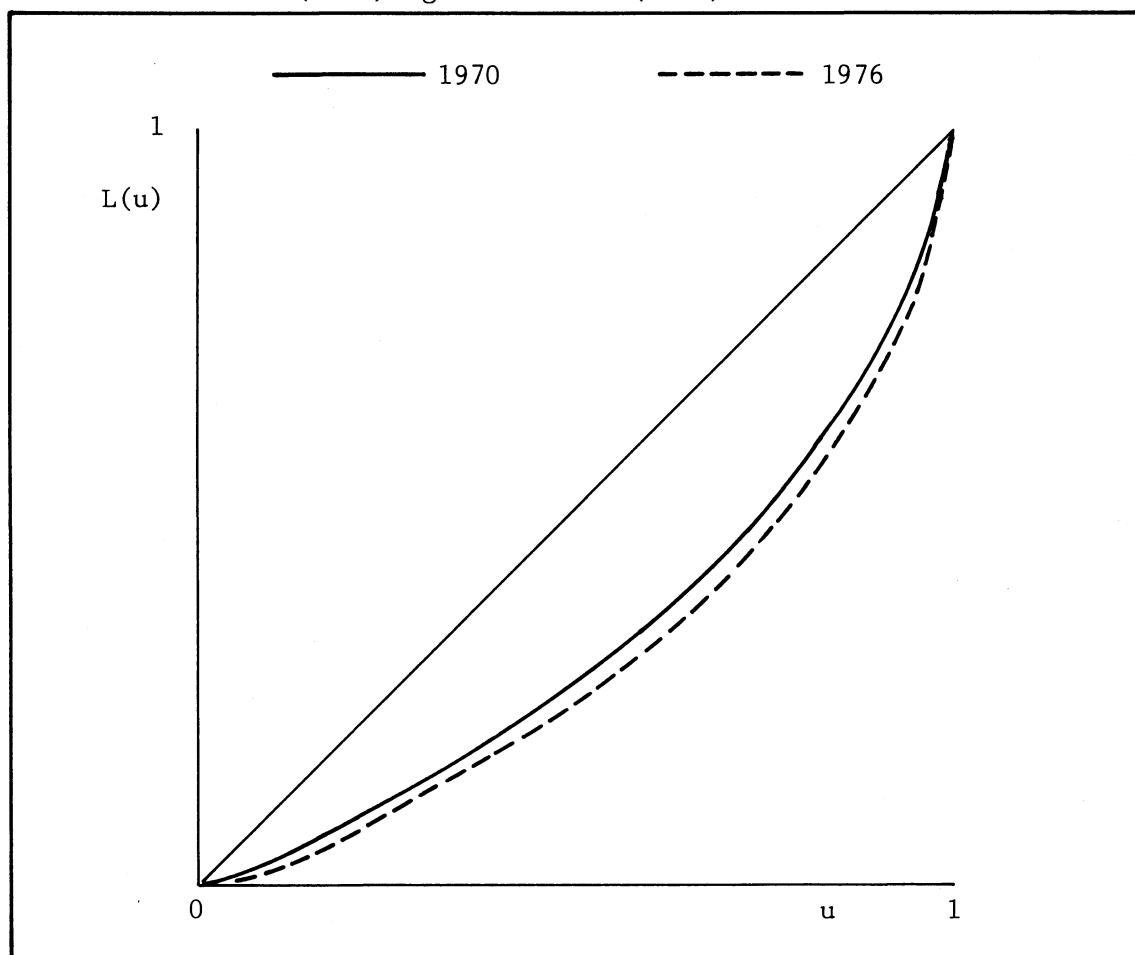
Figur 11. Ulikskapskurver  $M(u)$  for personar over 69 år (1970) og over 66 år (1976). Nettoinntekt



Figur 12. Ulikskapskurver  $N(u)$  for personar over 69 år (1970) og over 66 år (1976). Nettoinntekt



Figur 13. Ulikskapskurver  $L(u)$  (Lorenz-kurver) for personar over 69 år (1970) og over 66 år (1976). Nettoinntekt



Før vi går igang med å kommentere figurane, skal vi kort repetere nokre av eigenskapane til ulikskapskurvene  $M(\cdot)$ ,  $N(\cdot)$  og  $L(\cdot)$ .  $M(\cdot)$  og  $N(\cdot)$  vil uttrykkje ein større grad av likskap til nærare desse kurvene ligg den horisontale linja 1. For Lorenz-kurva  $L(\cdot)$  blir diagonalen (sjå figur 13) ein tilsvarande referanse. Sidan  $M(u)=u$  for den uniforme fordelinga, blir diagonalen ein nyttig tilleggsreferanse for denne ulikskapskurva. Vi kan snakke om fordelingar som gir ein større eller mindre grad av ulikskap enn den uniforme fordelinga gir.

Vi ser av figurane ovafor at inntektsfordelingane er blitt meir ulike i 1976 enn i 1970 for alle dei aktuelle aldersgruppene. Tendensen til større grad av ulikskap er spesielt markert for dei yngste og dei eldste aldersgruppene. Vi legg elles merke til at figurane gir eit ymt om at ulikskapen i fordelinga av nettoinntekt er størst innafor dei yngste aldersgruppene både i 1970 og 1976. Vi skal seinare gi ei meir presis handsaming av ulikskapsrangering for desse fem fordelingane både i 1970 og 1976. Vi skal da nytte tre ulike kriteria (ulikskapsmål) for rangering.

Diagonalen i figurane 3, 5, 7, 9 og 11 representerer den uniforme fordelinga, dvs. når alle dei aktuelle inntektsstorleikane er like ofte representert i populasjonen. Av figur 3 ser vi at inntektsfordelinga for den yngste aldersgruppa både i 1970 og 1976 har ein markert større grad av ulikskap enn det vi ville få om inntektene var uniformt fordelt. For aldersgruppene 25-39 år og 40-59 år (fig. 5 og 7) ser vi at inntektene er nærmast uniformt fordelt i 1970, medan begge 1976-fordelingane gir ein større grad av ulikskap enn den uniforme fordelinga. For dei to eldste aldersgruppene viser mønsteret i figurane 9 og 11 for 1970, at dei øvre og dei nedre delene av inntektstakarane hadde større gjennomsnittsinntekt enn det dei ville fått om inntektene var uniformt fordelt. Dette forholdet har endra seg i 1976. Da får dei øvre delene av populasjonane enda høgre gjennomsnittsinntekter relativt til totalgjennomsnittet, medan dei relative gjennomsnittsinntektene til dei nedre delene av populasjonen er blitt mindre.

I figur 13 har vi gitt ei grafisk framstilling av to Lorenz-kurver på grunnlag av nettoinntektsfordelingar for personar over 69 år i 1970 og for personar over 66 år i 1977. Vi har tidlegare peika på at den grafiske framstillinga av Lorenz-kurva skil dårlegare mellom to ulike inntektsfordelingar enn det ulikskapskurvene  $M(\cdot)$  og  $N(\cdot)$  gjer. Figurane 11, 12 og 13 gir ein god illustrasjon av dette poenget.

Ved å nytte den funksjonelle samanhengen mellom  $L(u)$  og respektive  $P_1(u)$  og  $P_2(u)$  skal vi rekne ut funksjonsverdiar for  $P_1$ -kurva og  $P_2$ -kurva på grunnlag av tabell 9.

Tabell 11. Estimerte funksjonsverdiar av  $P_2$ -kurva og  $P_1$ -kurva for personar over 69 år (1970) og over 66 år (1976). Nettoinntekt

	u år	.10	.20	.30	.40	.50	.60	.70	.80	.90
$P_2(u)$	1970	1.08	1.15	1.23	1.33	1.46	1.62	1.85	2.22	2.94
	1976	1.09	1.17	1.25	1.37	1.51	1.71	1.98	2.40	3.22
$P_1(u)$	1970	.30	.37	.38	.38	.37	.36	.34	.31	.27
	1976	.16	.28	.32	.33	.32	.31	.29	.27	.23

Frå tabell 11 ser vi f.eks. at gjennomsnittsnettoinntekta til den øvre 10-prosenten er nærmast 3 gonger totalgjennomsnittet ( $P_2(.90)$ ) i 1970 og 3.2 gonger totalgjennomsnittet i 1976. Vidare finn vi at den nedre 10-prosenten har ei gjennomsnittleg nettoinntekt som er i underkant av ein tredjedel (30%) av gjennomsnittsnettoinntekta til den øvre 90-prosenten ( $P_1(.10)$ ) i 1970, medan den nedre 10-prosenten har ei gjennomsnittleg nettoinntekt som ligg i underkant av ein sekstedel (16%) av gjennomsnittsnettoinntekta til den øvre 90-prosenten i 1976. Den dramatiske endringa frå 1970 til 1976 vil først og fremst skuldast endringar i populasjonen. Delen av yrkesaktive over 66 år i 1976 er større enn delen av yrkesaktive over 69 år i 1970.

Ei naturleg oppsummering av den drøftinga vi har gjort ovafor, vil vere å rangere dei fem inntektsfordelingane etter graden av ulikskap. Som rangeringskriteria skal vi nytte ulikskapsmåla A, B og G; dei oppsummerer ulikskap formidla ved respektive M-kurva, N-kurva og Lorenz-kurva. A, B og G legg vekt på forskjellige sider for ulikskap ved ei og same fordeling, og vil derfor naturlegvis kunne ende opp med forskjellige rangering for eit og same sett av fordelingar. I kapittel 3 har vi vurdert og dessutan samanlikna eigenskapane for A, B og G.

Vi kan estimere A, B og G direkte på grunnlag av tala i tabellane 5, 6, 7, 8 og 9 ved å nytte prosedyren (5.22) (sjå Merknad, kapittel 5.2). Vi får da følgjande tabell, der 1 representerer den yngste aldersgruppa, 2 den nest yngste aldersgruppa, osv.

Tabell 12. Estimerte ulikskapsmål for 10 nettoinntektsfordelingar

Ulikskaps- mål		Aldersgruppe nr. i	1	2	3	4	5
		$\hat{A}_i$	1970	.637	.505	.521	.530
	1976	.684	.559	.546	.546	.513	
$\hat{B}_i$	1970	.719	.616	.635	.647	.597	
	1976	.756	.659	.651	.656	.641	
$\hat{G}_i$	1970	.466	.336	.364	.384	.342	
	1976	.524	.378	.376	.388	.390	

I 1970 gir derfor A, B og G følgjande rangering

$$\begin{aligned}
 &A_5 < A_2 < A_3 < A_4 < A_1 \\
 (6.2) \quad &B_5 < B_2 < B_3 < B_4 < B_1 \\
 &G_2 < G_5 < G_3 < G_4 < G_1
 \end{aligned}$$

A og B gir same rangering, medan G i motsetnad til A og B rangerer aldersgruppe 2 gunstigare enn aldersgruppe 5. Forklaringa på dette er at A og B legg større vekt enn G på ulikskapsaspektet i den nedre delen av fordelinga. Vi ser frå tabell 12 at nokre av forskjellane som ligg til grunn for (6.2) er små og derfor kan skuldast tilfeldige variasjonar. For å dra sluttningar av typen (6.2) bør ein nytte simultane sluttningreglar, men manglande programutrustning gjer dette umogleg i vår situasjon. Det store talet på observasjonar (ca. 8 000 i 1970 og ca. 12 000 i 1976) gir forholdsvis presise estimat og dermed reduksjon i sjansen for at (6.2) er feil.

For 1976 gir tabell 12 opphav til følgjande rangeringar etter respektive A, B og G

$$\begin{aligned}
 & A_5 < A_3 = A_4 < A_2 < A_1 \\
 (6.3) \quad & B_5 < B_3 < B_4 < B_2 < B_1 \\
 & G_2 \leq G_3 < G_4 \leq G_5 < G_1
 \end{aligned}$$

Uansett kva slags ulikskapsmål vi ønskjer å rangere fordelingane etter, ser vi frå (6.2) og (6.3) at rangeringa av dei fem nettoinntektsfordelingane har endra seg frå 1970 til 1976.

Vi har no gitt ei vurdering av korleis nettoinntektene i 1970 og 1976 er fordelt på individa innafor fem ulike populasjonar. Resultatet viste at nettoinntektene var minst ulikt fordelt (målt ved A eller B) innafor den eldste aldersgruppa både i 1970 og 1976.

Ved å samanlikne gjennomsnittsinntektene i tabell 10, oppnår vi summarisk informasjon om ulikskap mellom populasjonane. Tabellen viser at den yngste og den eldste aldersgruppa har dei minste gjennomsnittsinntektene både i 1970 og 1976. Forskjellen i gjennomsnittsinntekt mellom den nest yngste og den eldste aldersgruppa i 1970 er kr 11 300. På same måte som i kapittel 4 vil vi oppnå bakgrunnsinformasjon om forskjellen i gjennomsnittsinntekt ved å studere fraktildifferansefunksjonen  $\Lambda(\cdot)$ .

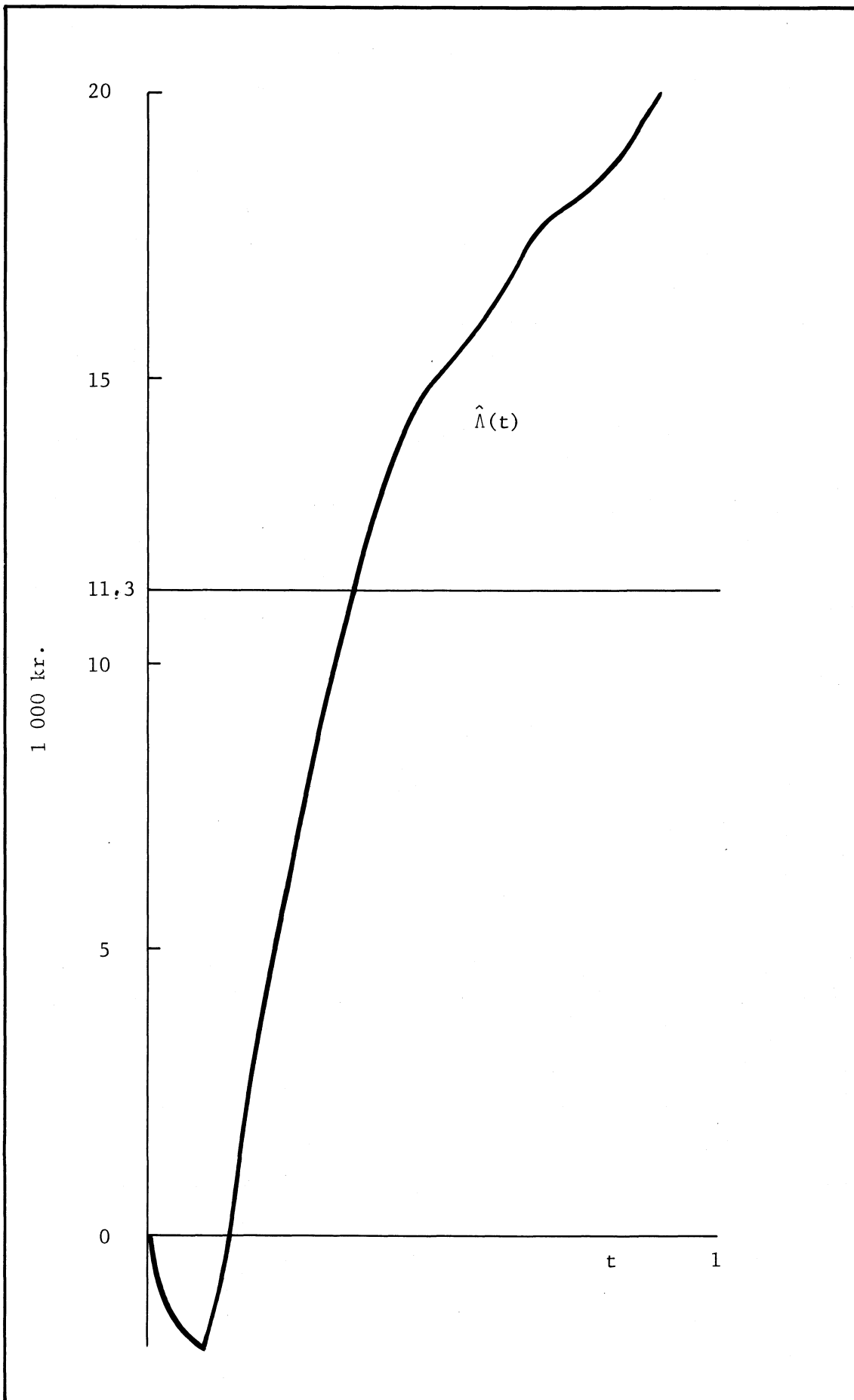
Tabell 13. Estimert for desilinntektsdifferansar mellom aldersgruppene 25-39 år og over 69 år i 1970 (1 000 kr)

t	.10	.20	.30	.40	.50	.60	.70	.80	.90
$\hat{\Lambda}(t)$	-2.1	3.4	8.8	12.7	15.0	16.2	17.8	18.5	20.0

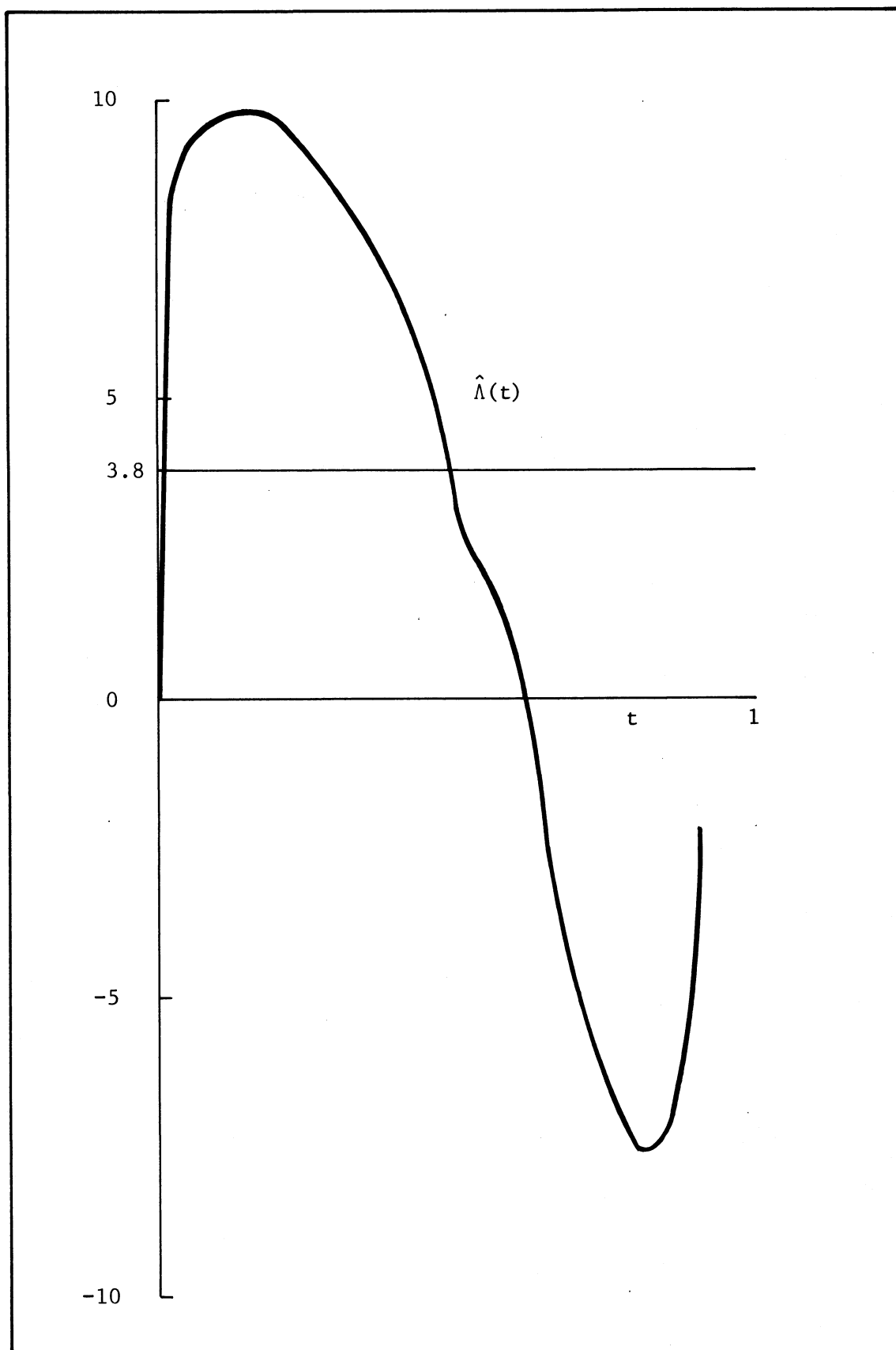
På grunnlag av punkttestimata i tabell 13 har vi framstilt fraktildifferansefunksjonen grafisk i figur 14. Plottet av  $\hat{\Lambda}(\cdot)$  viser at dei store og mellomstore nettoinntektene er betydeleg større for aldersgruppa 25-39 år samanlikna med aldersgruppa over 69 år. Når det gjeld dei aller minste inntektene, er desse størst for aldersgruppa over 69 år. Det er forskjellane i dei store nettoinntektene som fører til at forskjellen i gjennomsnittsinntekt mellom dei to gruppene blir så pass stor som kr 11 300.

Forskjellen i gjennomsnittsinntekt mellom den eldste og den yngste aldersgruppa i 1976 er kr 3 800. I tabell 14 gir vi dei fraktildifferanseestimata som byggjer på desilinntektene i tabell 31 frå Inntektsstatistikk 1976.

Figur 14. Fraktildifferansefunksjonen mellom aldersgruppene 25-39 år og over 69 år i 1970. (1 000 kr)



Figur 15. Fraktildifferansefunksjonen mellom aldersgruppene over 66 år og under 25 år i 1976





Tabell 14. Estimert for desilinntektsdifferansar mellom aldersgruppene over 66 år og under 25 år i 1976 (1 000 kr)

t	.10	.20	.30	.40	.50	.60	.70	.80	.90
$\hat{\Lambda}(t)$	9.7	9.7	8.5	6.8	3.1	.7	-5.0	-7.7	-2.2

Figur 15. viser eit heilt anna mønster enn figur 14. Dei minste og mellomstore nettoinntektene for kvar av dei to gruppene er betydeleg større for den eldste aldersgruppa, medan dei store nettoinntektene har ein tendens til å vere større i den yngste aldersgruppa. Det er den store forskjellen på dei minste og mellomstore nettoinntektene som gjer at gjennomsnittsnettoinntekta blir størst i den eldste av dei to aldersgruppene.

## 6.2. Hushaldsinntekt

Vi skal no studere fordelinga av hushaldsinntektene før og etter skatt (nettoinntekt og disponibel inntekt) i 1976 for alle hushald og for 3 grupper som er bestemt ved talet på inntektstakarar i hushaldet. Tabell 94 i Inntektsstatistikk 1976 gir oss nettoinntekt for hushald etter desilar og talet på inntektstakarar i hushaldet, medan disponibel inntekt berre er tilgjengeleg i form av maskintabellar i Statistisk Sentralbyrå. For å rekne ut punkttestimat for ulikskapskurvene  $L(\cdot)$ ,  $M(\cdot)$  og  $N(\cdot)$  går vi fram på same måte som vist i kapittel 6.1. Vi oppnår da følgjande tabellar og figurar,

Tabell 15. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for alle hushald (1976)

	u	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	INNTEKT									
	NETTO	.014	.043	.089	.156	.239	.337	.452	.589	.756
	DISPONIB.	.022	.063	.121	.194	.281	.382	.497	.630	.787
M(u)	NETTO	.14	.22	.30	.39	.48	.56	.65	.74	.84
	DISPONIB.	.22	.32	.40	.49	.56	.64	.71	.79	.87
N(u)	NETTO	.06	.10	.16	.24	.31	.40	.50	.62	.77
	DISPONIB.	.10	.17	.24	.31	.39	.47	.57	.67	.80

Tabell 16. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for hushald med 1 inntektstakar (1976)

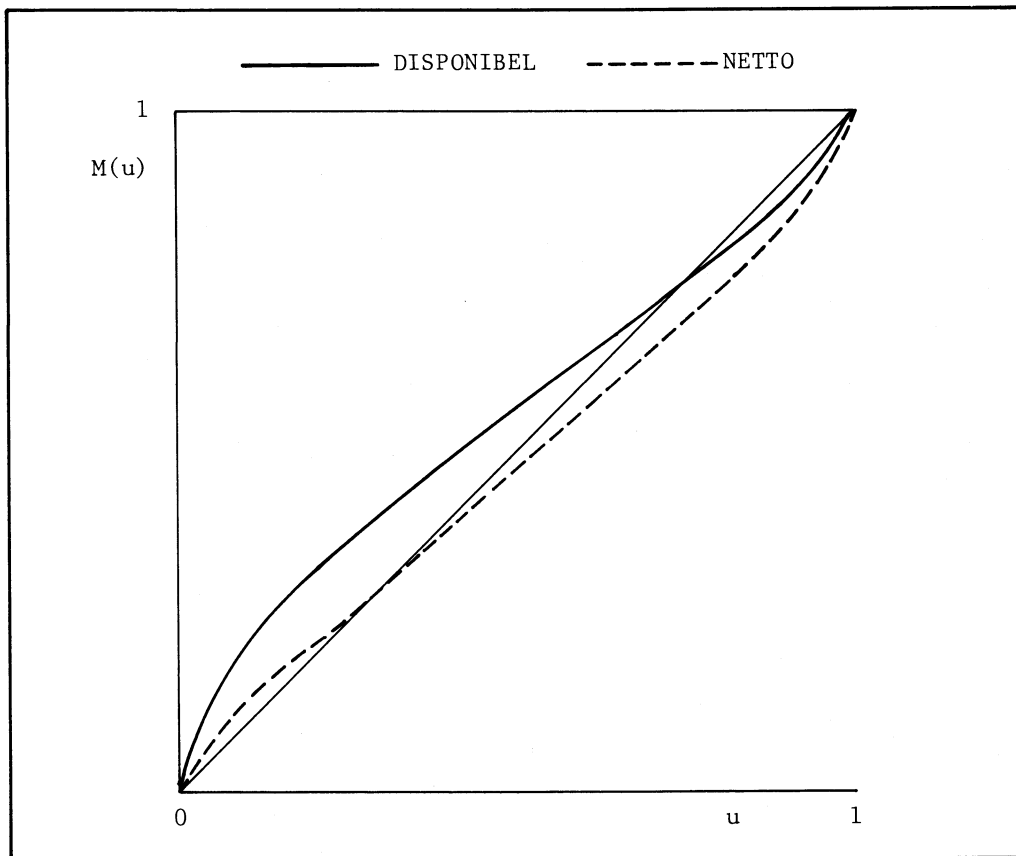
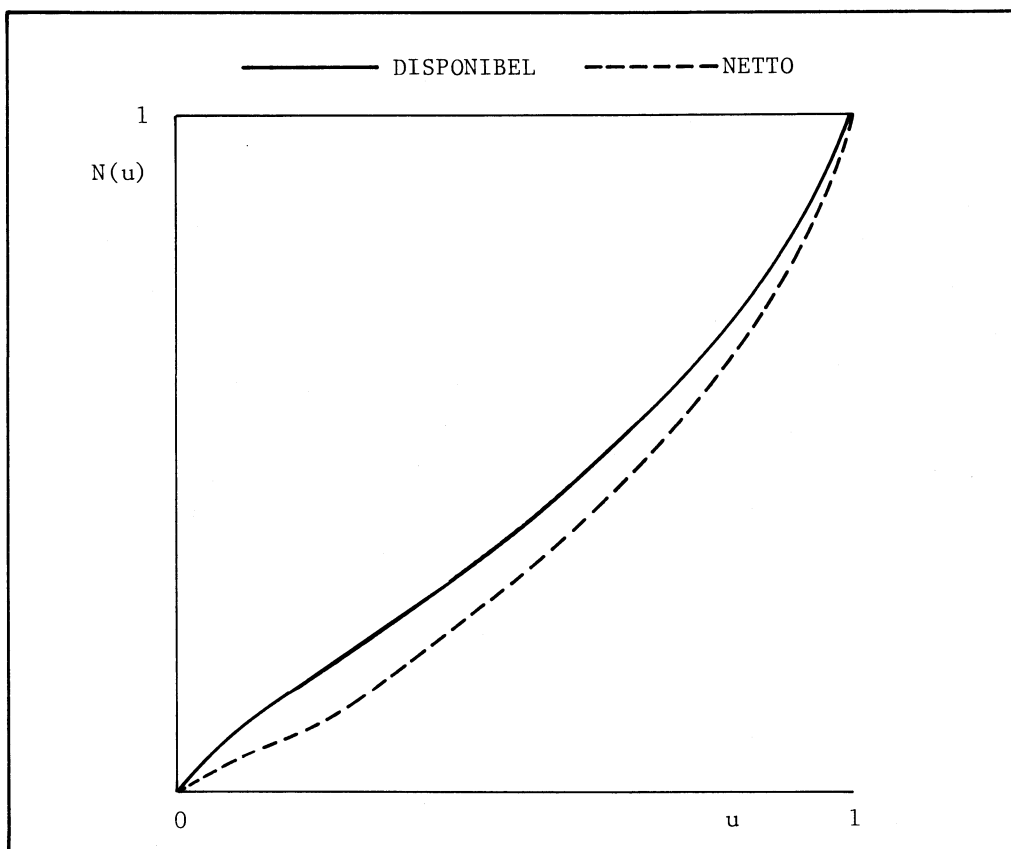
	INNTEKT <sup>u</sup>	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	NETTO	.012	.045	.086	.140	.215	.313	.434	.576	.745
	DISPONIB.	.020	.067	.124	.196	.283	.386	.505	.640	.795
M(u)	NETTO	.12	.23	.29	.35	.43	.52	.62	.72	.83
	DISPONIB.	.20	.34	.41	.49	.57	.64	.72	.80	.88
N(u)	NETTO	.05	.11	.15	.20	.27	.36	.47	.60	.75
	DISPONIB.	.10	.19	.25	.32	.39	.48	.58	.69	.81

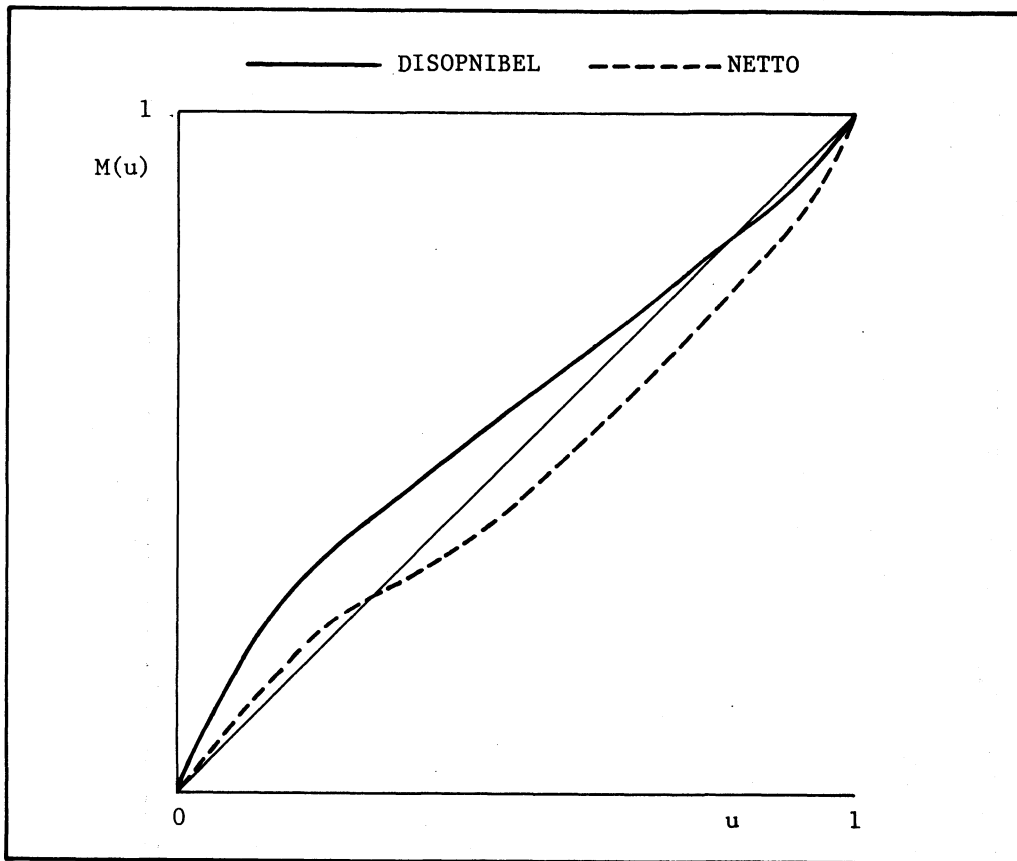
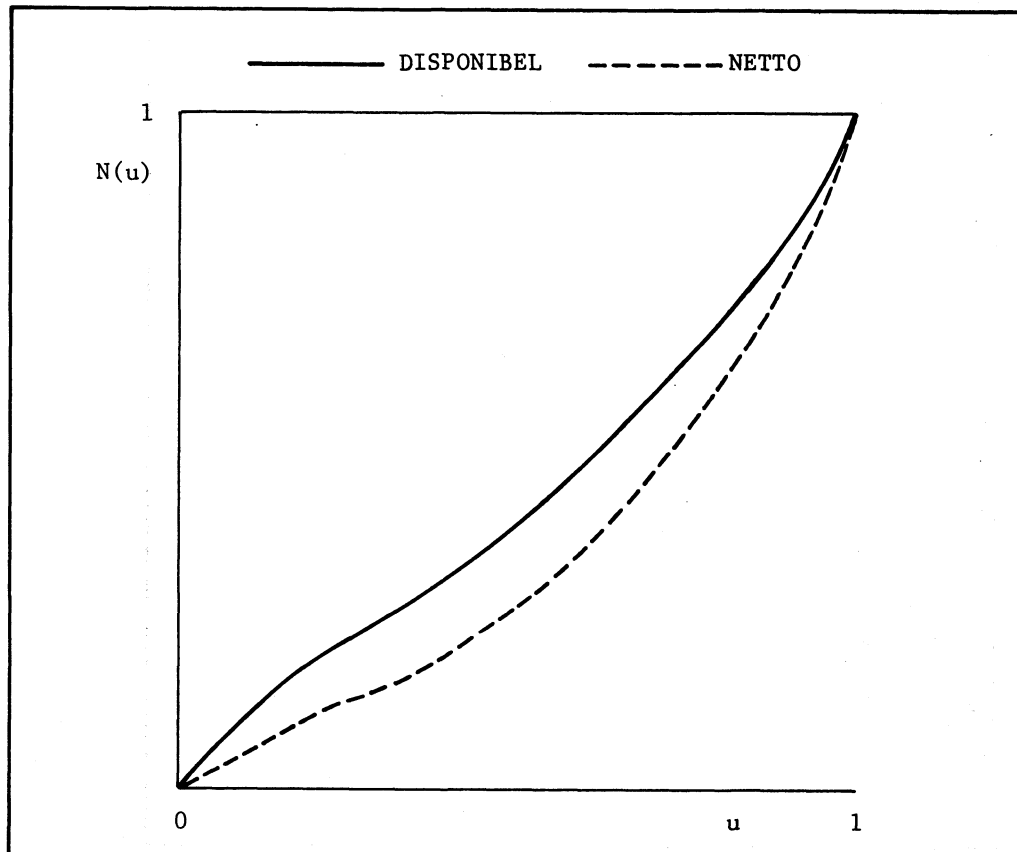
Tabell 17. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for hushald med 2 inntektstakarar (1976)

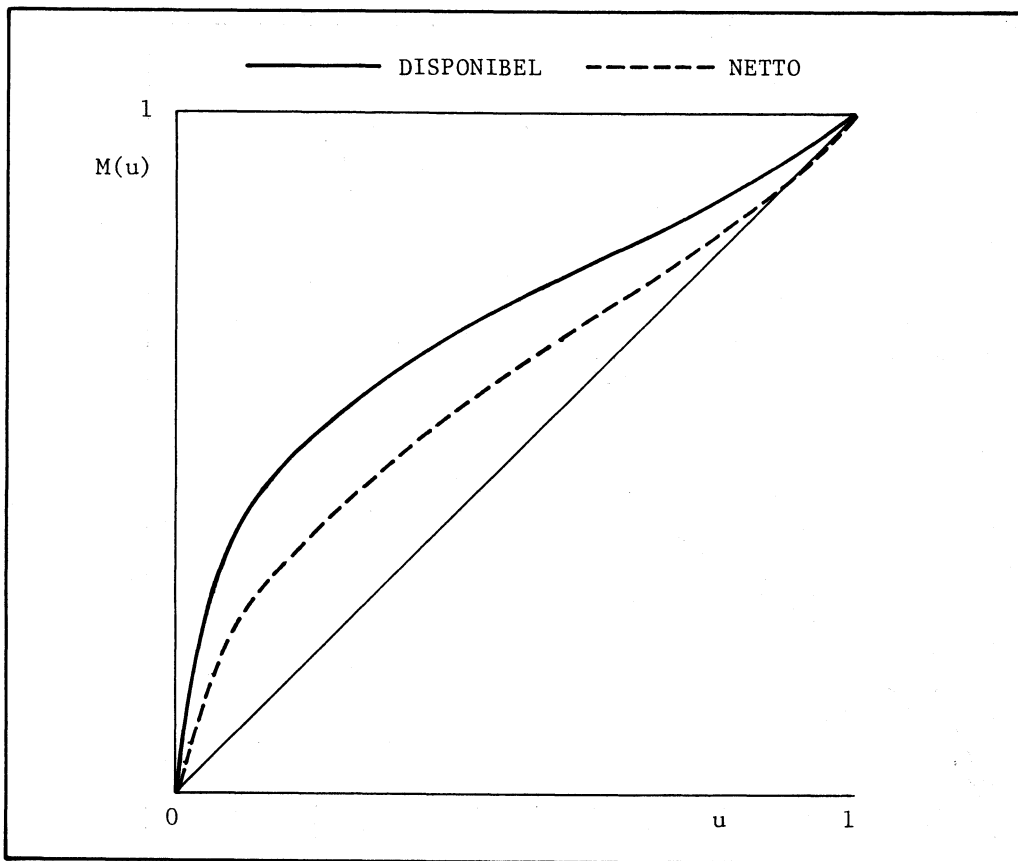
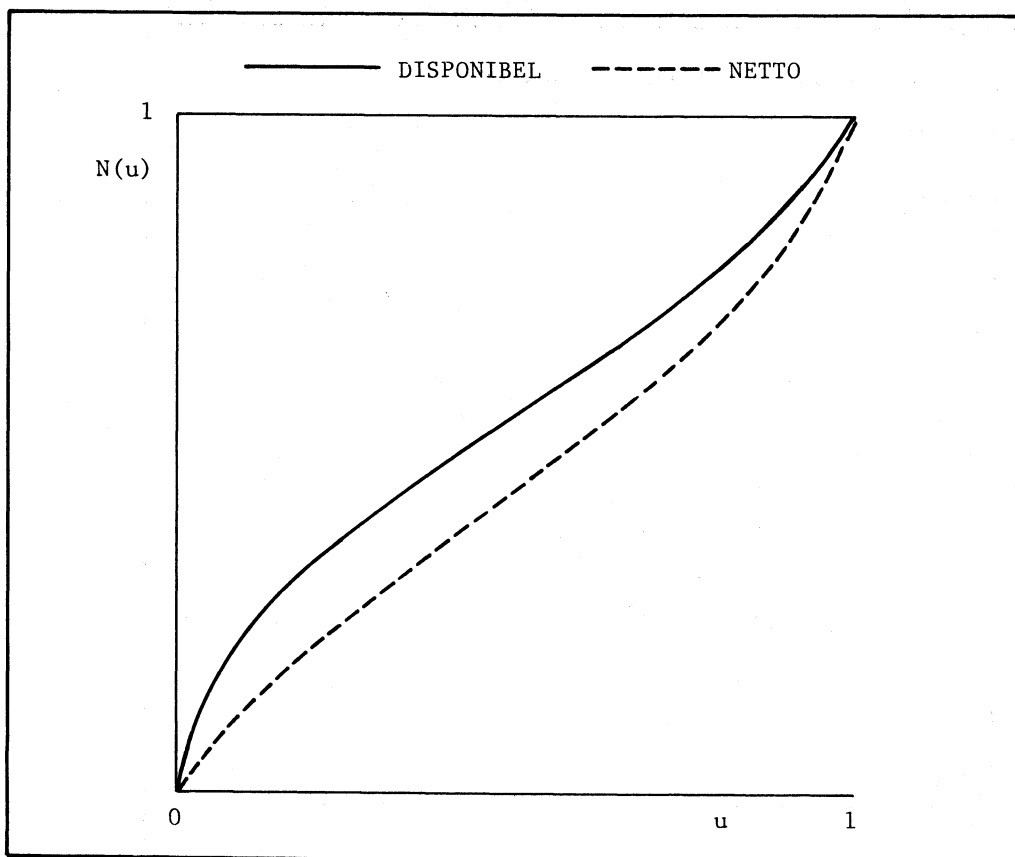
	INNTEKT <sup>u</sup>	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	NETTO	.026	.074	.141	.221	.311	.412	.525	.652	.798
	DISPONIB.	.040	.102	.179	.266	.361	.464	.576	.698	.833
M(u)	NETTO	.26	.37	.47	.55	.62	.69	.75	.82	.89
	DISPONIB.	.40	.51	.60	.67	.72	.77	.82	.87	.93
N(u)	NETTO	.13	.21	.30	.38	.45	.53	.61	.70	.82
	DISPONIB.	.24	.34	.42	.50	.56	.63	.70	.78	.87

Tabell 18. Estimerte funksjonsverdiar av Lorenz-kurva, M-kurva og N-kurva for hushald med 3 inntektstakarar (1976)

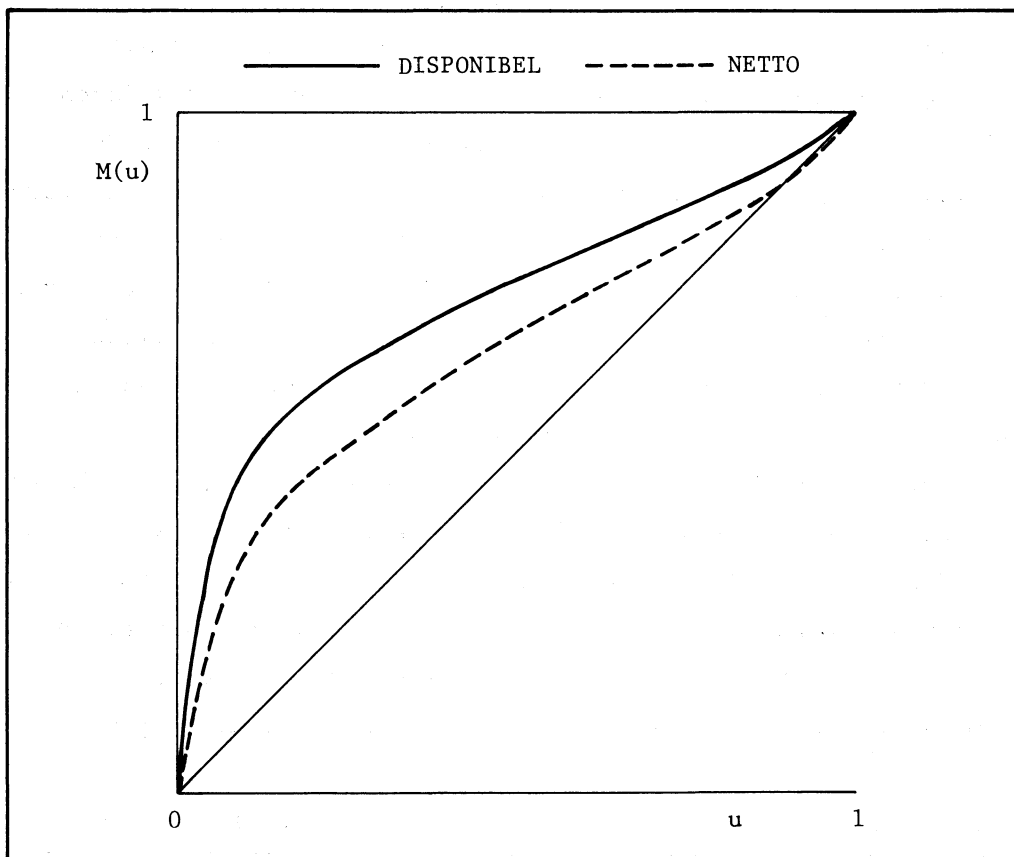
	INNTEKT <sup>u</sup>	.10	.20	.30	.40	.50	.60	.70	.80	.90
L(u)	NETTO	.036	.094	.165	.247	.338	.438	.548	.670	.810
	DISPONIB.	.048	.117	.196	.283	.377	.477	.585	.703	.835
M(u)	NETTO	.36	.47	.55	.62	.68	.73	.78	.84	.90
	DISPONIB.	.48	.59	.65	.71	.75	.80	.83	.88	.93
N(u)	NETTO	.19	.28	.37	.44	.51	.58	.66	.74	.84
	DISPONIB.	.29	.39	.47	.54	.61	.67	.73	.80	.88

Figur 16. Ulikskapskurver  $M(u)$  for alle hushald (1976)Figur 17. Ulikskapskurver  $N(u)$  for alle hushald (1976)

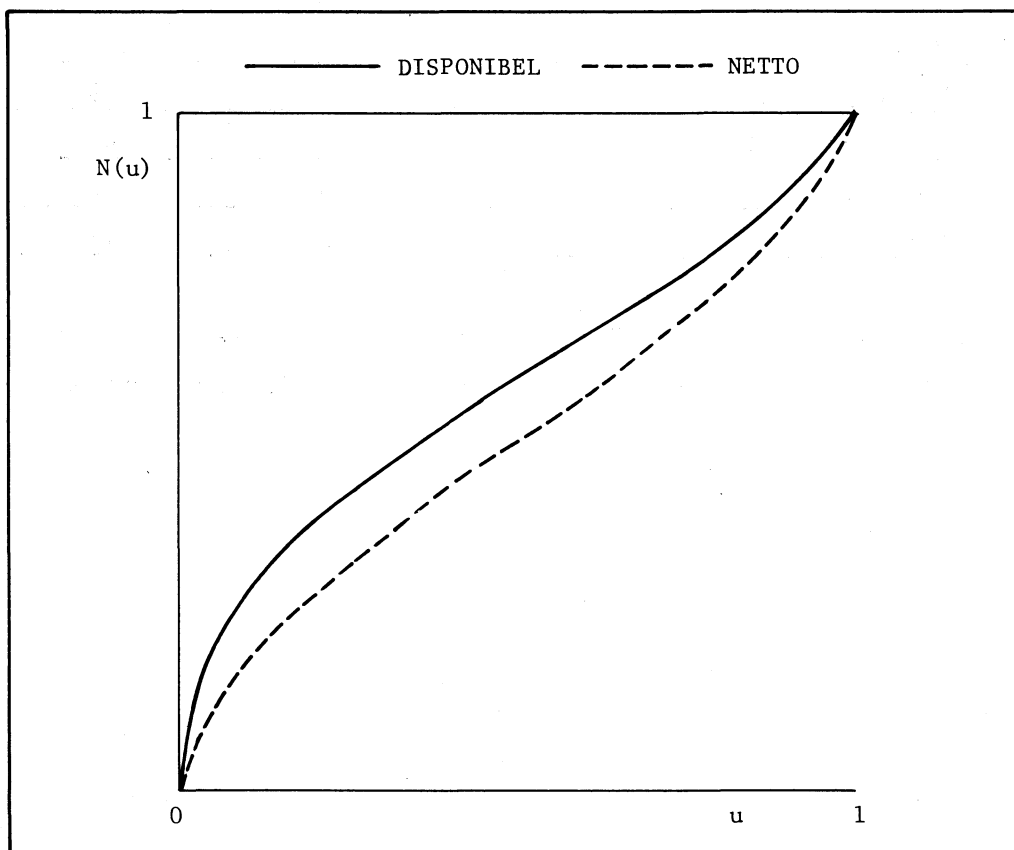
Figur 18. Ulikskapskurver  $M(u)$  for hushald med 1 inntektstakar (1976)Figur 19. Ulikskapskurver  $N(u)$  for hushald med 1 inntektstakar (1976)

Figur 20. Ulikskapskurver  $M(u)$  for hushald med 2 inntektstakarar (1976)Figur 21. Ulikskapskurver  $N(u)$  for hushald med 2 inntektstakarar (1976)

Figur 22. Ulikskapskurver  $M(u)$  for hushald med 3 eller fleire inntektstakarar (1976)



Figur 23. Ulikskapskurver  $N(u)$  for hushald med 3 eller fleire inntektstakarar (1976)



Tala i tabellane 15-18 har ei presis økonomisk tyding som er omtalt i kapittel 2. Figurane 16-23 gir eit oversiktleg bilde av informasjonen i tabellane 15-18. Dei heiltrukne kurvene ligg utan unntak nærmast den horisontale linja 1. Dette viser at disponibel inntekt er meir likt fordelt enn nettoinntekt innafor alle hushald og i tillegg innafor kvar av dei 3 hushaldsgruppene. Ved å samanlikne dei heiltrukne og dei stipla kurvene for kvar av figurane ovafor, får vi eit inntrykk av kva slags likskaps-effekt (målt ved respektive M-kurva og N-kurva) den progressive skattleggjinga hadde på desse 4 nettoinntektsfordelingane i 1976.

Frå figurane 18-23 ser vi at inntektene (målt ved både M og N) er meir likt fordelt innafor hushald med 3 eller fleire inntektstakarar enn innafor hushald med 1 og med 2 inntektstakarar. Vidare er det større grad av likskap i inntektsfordelingane for hushald med 2 inntektstakarar enn for hushald med 1 inntektstakar. Denne rangeringa får vi stadfesta i tabell 19 som gir dei aktuelle estimata for ulikskapsmåla A, B og G.

Tabell 19. Estimerte ulikskapsmål for 6 inntektsfordelingar i 1976

Ulikskapsmål		Talet på inntektstak. i hushaldet		
		1	2	3 eller fleire
$\hat{A}$	Nettoinntekt	.54	.41	.36
	Disp. inntekt	.44	.32	.29
$\hat{B}$	Nettoinntekt	.65	.54	.49
	Disp. inntekt	.57	.45	.41
$\hat{G}$	Nettoinntekt	.38	.26	.24
	Disp. inntekt	.30	.20	.18

Tabell 19 er ei oppsummering av informasjonen frå ulikskapskurvene  $M(\cdot)$ ,  $N(\cdot)$  og  $L(\cdot)$  gitt ved respektive A, B og G og gir grunnlag for kompakte samanlikningar mellom 3 hushaldsgrupper.

## APPENDIKS 1

## ASYMPTOTISK TEORI FOR EMPIRISKE ULIKSKAPSKURVER OG ULIKSKAPSMÅL

På grunnlag av dei empiriske ulikskapskurvene (kapittel 5.1) og dei tilhøyrande teoretiske ulikskapskurvene (kapittel 2), kan vi konstruere forskjellige empiriske prosessar. Ved å studere dei asymptotiske eigenskapane til desse prosessane, har vi funne asymptotiske fordelingar til vega utgåver av dei empiriske ulikskapskurvene (for faste punkt på kurva). Desse resultatane kan vi nytte til å etablere testar (desisjonsreglar) for hypotesar om faste punkt på ulikskapskurvene. Dei tilhøyrande asymptotiske variansane vil gi informasjon om kor presist ei ulikskapskurve blir estimert i eit gitt punkt. Alle desse resultatane blir presenterte og drøfta i kapittel A. I kapittel B gir vi asymptotiske variansar og fordelingar for vega utgåver av estimatorane til dei ulikskapsmåla som blir drøfta i kapittel 3. I kapittel C vurderer vi asymptotiske eigenskapar til den empiriske fraktildifferansefunksjonen.

a. Asymptotiske resultat for empiriske ulikskapskurver

Alle dei forskjellige empiriske ulikskapskurvene som blei definert i kapittel 5 kan uttrykkjast som funksjonar av blant anna  $\hat{H}(u)$  som er definert ved (5.1). Følgjeleg vil det vere av interesse å studere statistiske eigenskapar for  $\hat{H}(u)$ .

La  $D[0,1]$  vere rommet av funksjonar på  $[0,1]$  som er høgrekontinuerlege og har grense frå venstre. På dette rommet skal vi bruke Skorokhod-topologien (sjå kapittel 3, Billingsley (1968)).  $\sqrt{n}(\hat{H}(u)-H(u))$  og dei andre funksjonane som blir drøfta i dette kapitlet, er alle med i rommet  $D[0,1]$ .

La  $W_0(t)$  vere ei Brownsk bru på  $[0,1]$ , dvs. ein Gaussisk prosess med forventning 0 og kovariansfunksjon  $s(1-t)$ ,  $0 \leq s \leq t \leq 1$ .

Setning 1. Føreset at definisjonsområdet til den kumulative fordelingsfunksjonen  $F$  er eit endeleg intervall. Føreset vidare at  $F(x)$  har ein kontinuerleg derivert  $f(x)$  som er slik at  $0 < f(x) < \infty$  på definisjonsområdet til  $F$ . Da vil

$$\sqrt{n}(\hat{H}(u)-H(u)) \xrightarrow{D} \frac{1}{u} \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt.$$



Prov: Frå definisjonane av  $H$  og  $\hat{H}$  får vi at

$$\sqrt{n}(\hat{H}(u) - H(u)) = \frac{1}{u} \int_0^u \sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) dt$$

La  $U_1, U_2, \dots, U_n$  vere uavhengige stokastiske variable der kvar  $U$  er uniformt fordelt over  $[0, 1]$ . La  $\Gamma_n$  vere den empiriske fordelingsfunksjonen for  $U_1, U_2, \dots, U_n$  og la  $\Gamma_n^{-1}(t) = \inf \{u : \Gamma_n(u) \geq t\}$ . Da har vi at

$$\Gamma_n(F(x)) = \frac{1}{n} (\#U_i \leq F(x)) = \frac{1}{n} (\#F^{-1}(U_i) \leq x) \stackrel{D}{=} F_n(x)$$

sidan  $\Pr(F^{-1}(U_i) \leq x) = \Pr(U_i \leq F(x)) = F(x)$  og

$F^{-1}(\Gamma_n^{-1}(t))$  er uavhengig av  $F_n^{-1}(t)$ . Dermed blir  $F^{-1}(\Gamma_n^{-1}(t)) \stackrel{D}{=} F_n^{-1}(t)$ .

Vi får da at

$$\begin{aligned} \sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) &\stackrel{D}{=} \sqrt{n}(F^{-1}(\Gamma_n^{-1}(t)) - F^{-1}(t)) \\ &= \frac{F(\Gamma_n^{-1}(t)) - F^{-1}(t)}{\Gamma_n^{-1}(t) - t} \sqrt{n}(\Gamma_n^{-1}(t) - t) \end{aligned}$$

Det følgjer frå provet av Teorem 4.1, Doksum (1974) at

$(F(\Gamma_n^{-1}(t)) - F^{-1}(t)) / (\Gamma_n^{-1}(t) - t)$  konvergerer uniformert mot

$1/f(F^{-1}(t))$  og frå Shorack (1972) at  $\sqrt{n}(\Gamma_n^{-1}(t) - t) \stackrel{D}{\rightarrow} W_0(t)$ .

Dermed har vi at

$$\sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) \stackrel{D}{\rightarrow} \frac{W_0(t)}{f(F^{-1}(t))}$$

Sidan  $\int_0^u (W_0(t)/f(F^{-1}(t))) dt$  er kontinuerleg som funksjon av  $W_0(t)/f(F^{-1}(t))$

i Skorokhod-topologien (følgjer frå argumentasjonen i kapittel 4.4, Durbin (1973)) får vi resultatet i Setning 1 ved å bruke Teorem 5.1, Billingsley (1968).

Det følger no direkte frå Setning 1 at  $\hat{H}(u)$  er eit konsistent estimat for  $H(u)$ . Frå provet av Setning 1 oppnår vi elles følgjande resultat,

Korollar 1. Under føresetnadene i Setning 1 blir

$$\text{var } F_n^{-1}(t) \approx \frac{t(1-t)}{nf^2(F^{-1}(t))} .$$

Merknad 1: Korollar 1 gir at  $F_n^{-1}(1/2)$  er ein konsistent estimator for medianen med tilnærma varians  $[4nf^2(F^{-1}(1/2))]^{-1}$ .

Setning 2. La  $W_0(t)$  vere ei Brownsk bru og  $F$  ein kumulativ fordelingsfunksjon. Under føresetnadene i Setning 1 blir

$$\Pr\left(\int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt \leq k\right) = \Pr\left(\sum_{j=1}^{\infty} q_j(u) Z_j \leq k\right)$$

der  $q_j(u)$  er gitt ved

$$(1) \quad q_j(u) = \frac{\sqrt{2}}{j\pi} \int_0^u \frac{\sin(j\pi t)}{f(F^{-1}(t))} dt$$

og  $Z_1, Z_2, \dots$  er uavhengige og standard normalfordelte stokastiske variable.

Prov. La

$$V_N(t) = \frac{\sqrt{2}}{f(F^{-1}(t))} \sum_{j=1}^N \frac{\sin(j\pi t)}{j\pi} Z_j$$

Ved å nytte følgjande resultat (sjå f.eks. Hajek & Sidak (1967))

$$(2) \quad 2 \sum_{j=1}^{\infty} \frac{\sin(j\pi t) \sin(j\pi s)}{(j\pi)^2} = t(1-s), \quad 0 \leq t \leq s \leq 1,$$

ser vi at prosessen  $V_N(\cdot)$  er Gaussisk med forventning 0 og kovariansfunksjon

$$\text{cov}(V_N(t), V_N(s)) =$$

$$\frac{2}{f(F^{-1}(t)) f(F^{-1}(s))} \sum_{j=1}^N \frac{\sin(j\pi t) \sin(j\pi s)}{(j\pi)^2} \rightarrow \text{cov}(V(t), V(s))$$

$$\text{der } V(t) = \frac{W_0(t)}{f(F^{-1}(t))} .$$

For å prove at  $V_N(\cdot)$  går i fordeling mot  $V(\cdot)$ , følger det no frå Teorem 3.2 og Teorem 3.1.b, Hajek & Sidak (1967) at det er tilstrekkeleg å vise at

$$E[V_N(t) - V_N(s)]^4 \leq M(t-s)^2$$

der  $M$  er uavhengig av  $N$ .

For normalfordelte variable  $X$  med forventning 0 har vi at

$$EX^4 = 3[EX^2]^2 .$$

Dette gir

$$\begin{aligned} E[V_N(t) - V_N(s)]^4 &= 3[\text{var}(V_N(t) - V_N(s))]^2 = \\ &3 \{2 \text{ var} [ \sum_{j=1}^N \frac{1}{j\pi} (\frac{\sin(j\pi t)}{f(F^{-1}(t))} - \frac{\sin(j\pi s)}{f(F^{-1}(s))}) Z_j ]\}^2 = \\ &3 \{2 \sum_{j=1}^N [ \frac{1}{j\pi} (\frac{\sin(j\pi t)}{f(F^{-1}(t))} - \frac{\sin(j\pi s)}{f(F^{-1}(s))}) ]^2\} \leq \\ &12 \{ \frac{1}{f^2(F^{-1}(t))} \sum_{j=1}^{\infty} (\frac{\sin(j\pi t)}{j\pi})^2 + \frac{1}{f^2(F^{-1}(s))} \sum_{j=1}^{\infty} (\frac{\sin(j\pi s)}{j\pi})^2 - \\ &\frac{2}{f(F^{-1}(t)) f(F^{-1}(s))} \sum_{j=1}^{\infty} \frac{\sin(j\pi t) \sin(j\pi s)}{(j\pi)^2} \} = \\ &3 \{ \frac{t(1-t)}{f^2(F^{-1}(t))} + \frac{s(1-s)}{f^2(F^{-1}(s))} - 2 \frac{\text{cov}(W_0(t), W_0(s))}{f(F^{-1}(t)) f(F^{-1}(s))} \}^2 \end{aligned}$$

Sidan  $0 < f(x) < \infty$  på definisjonsområdet til  $F$ , finst det ein konstant  $M$  slik at  $f(F^{-1}(t)) \geq M \frac{1}{4}$  for alle  $t \in [0,1]$ . Altså får vi at

$$E[V_N(t) - V_N(s)]^4 \leq 3M \{t(1-t) + s(1-s) -$$

$$2 \operatorname{cov}(W_0(t), W_0(s))\}^2 \leq 3M(t-s)^2(1 - |t-s|)^2 \leq 3M(t-s)^2.$$

Vi har dermed vist at  $V_N(\cdot)$  konvergerer i fordeling mot  $V(\cdot)$ . Følgjeleg gir Teorem 5.1, Billingsley (1968) (grunngevinga for å nytte dette teoremet er gitt i provet av Setning 1) at

$$\int_0^u V_N(t) dt = \sqrt{2} \sum_{j=1}^N \frac{Z_j}{j\pi} \int_0^u \frac{\sin(j\pi t)}{f(F^{-1}(t))} dt \xrightarrow{D} \int_0^u V(t) dt,$$

og vi kan konstatere at

$$\int_0^u (W_0(t)/f(F^{-1}(t))) dt \quad \text{og} \quad \sum_{j=1}^{\infty} q_j(u) Z_j$$

har same fordeling.

La funksjonane  $\tau^2(u)$  og  $\lambda(u,v)$  vere definert ved respektive

$$(3) \quad \tau^2(u) = 2 \int_0^{F^{-1}(u)} \int_0^y F(x)(1-F(y)) dx dy, \quad 0 \leq u \leq 1$$

og

$$(4) \quad \lambda(u,v) = \int_{F^{-1}(u)}^{F^{-1}(v)} \int_0^{F^{-1}(u)} F(x)(1-F(y)) dx dy, \quad 0 \leq u \leq v \leq 1.$$

Korollar 2. Under føresetnadene i Setning 1 blir

$$\operatorname{cov}(\hat{H}(u), \hat{H}(v)) \approx \frac{1}{nuv} (\tau^2(u) + \lambda(u,v))$$

for  $0 < u \leq v \leq 1$  når  $n$  er stor.

Prov. Setning 1 og Setning 2 gir for store  $n$  og for  $0 < u \leq v \leq 1$  at

$$\text{cov}(\hat{H}(u), \hat{H}(v)) \approx \frac{1}{nuv} \text{cov}\left(\sum_{i=1}^{\infty} q_i(u)Z_i, \sum_{j=1}^{\infty} q_j(v)Z_j\right) =$$

$$\frac{1}{nuv} \sum_{j=1}^{\infty} q_j(u)q_j(v) =$$

$$\frac{1}{nuv} \sum_{i=1}^{\infty} \int_0^v \int_0^u \frac{2}{f(F^{-1}(t))f(F^{-1}(s))} \frac{\sin(i\pi s)\sin(i\pi t)}{(i\pi)^2} dt ds =$$

$$\frac{1}{nuv} \int_0^v \int_0^u \frac{2}{f(F^{-1}(t))f(F^{-1}(s))} \sum_{i=1}^{\infty} \frac{\sin(i\pi s)\sin(i\pi t)}{(i\pi)^2} dt ds =$$

$$\frac{1}{nuv} \left[ 2 \int_0^u \int_0^s \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds + \int_u^v \int_0^u \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds \right] =$$

$$\frac{1}{nuv} \left[ 2 \int_0^{F^{-1}(u)} \int_0^y F(x)(1-F(y)) dx dy + \right.$$

$$\left. \int_{F^{-1}(u)}^{F^{-1}(v)} \int_0^y F(x)(1-F(y)) dx dy \right] = \frac{1}{nuv} [\tau(u) + \lambda(u,v)].$$

Under utrekninga av tilnærminga for kovariansen mellom  $\hat{H}(u)$  og  $\hat{H}(v)$  har vi gjort bruk av Fubinis teorem (sjå Royden (1963)) og (2).

Merknad 2: Korollar 2 gir at  $\sqrt{n} \hat{H}(u)$  har asymptotisk varians  $(\tau(u)/u)^2$  for  $0 < u \leq 1$ .

Merknad 3: Sidan  $\hat{H}(1) = \bar{X}$ , gir Korollar 2 at

$$\frac{\tau(1)}{n} = \frac{2}{n} \int_0^{\infty} [(1-F(y)) \int_0^y F(x) dx] dy$$

er eit tilnærma uttrykk for variansen til gjennomsnittet av  $n$  uavhengige observasjonar frå  $F$ .

Setning 1 saman med Setning 2 og Korollar 2 gir følgjande nyttige resultat.

Korollar 3. Føreset at  $\tau^2(u) < \infty$  for  $0 < u \leq 1$ . Under føresetnadene i Setning 1 blir da

$$\frac{\sqrt{n} u(\hat{H}(u) - H(u))}{\tau(u)}$$

asymptotisk  $N(0,1)$  for fast  $u$ ,  $0 < u \leq 1$ .

La  $\mu = EX$ . Vi skal no bruke resultatene i Setning 1 og Setning 2 til å vise at den empiriske prosessen  $\sqrt{n} (\hat{L}(u) - L(u))$  konvergerer mot ein Gaussisk prosess.

Setning 3. Under føresetnadene i Setning 1 blir

$$\lim \Pr(\sqrt{n} (\hat{L}(u) - L(u)) \leq k) = \Pr\left(\sum_{j=1}^{\infty} h_j(u) Z_j \leq k\right)$$

der

$$h_j(u) = \frac{1}{\mu} (q_j(u) - q_j(1) L(u))$$

og  $Z_1, Z_2, \dots$  er uavhengige og standard normalfordelte variable.

$$\text{Prov. Sidan } \hat{L}(u) = \frac{u\hat{H}(u)}{\bar{X}},$$

$$L(u) = \frac{uH(u)}{\mu}, \quad \bar{X} = \hat{H}(1) \text{ og } \mu = H(1) \text{ får vi}$$

$$\sqrt{n} (\hat{L}(u) - L(u)) = \frac{1}{\bar{X}} (Y_n(u) - L(u)Y_n(1))$$

der

$$Y_n(u) = \sqrt{n} u(\hat{H}(u) - H(u)).$$

Setning 1 gir at

$$Y_n(u) \xrightarrow{D} Y(u)$$

og dermed at

$$Y_n(u) - L(u)Y_n(1) \xrightarrow{D} Y(u) - L(u)Y(1)$$

der

$$Y(u) = \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt.$$

Sidan  $\bar{X} \xrightarrow{P} \mu$  får vi derfor ved å bruke Cramer-Slutsks setning at

$\sqrt{n} (\hat{L}(u) - L(u)) \xrightarrow{D} \frac{1}{\mu} (Y(u) - L(u)Y(1))$ . Dermed følgjer resultatet i Setning 3 frå Setning 2.

I Korollar 2 prova vi at

$$(5) \quad \sum_{i=1}^{\infty} q_i(u) q_i(v) = \tau^2(u) + \lambda(u,v), \quad 0 < u \leq v < 1,$$

der  $\tau^2(u)$  og  $\lambda(u,v)$  er gitt ved respektive (3) og (4). Vi skal blant anna gjere bruk av denne identiteten for å vise neste resultat.

Korollar 4. Under føresetnadene i Setning 1 blir

$$\text{cov}(\hat{L}(u), \hat{L}(v)) \approx \frac{1}{n\mu^2} [\tau^2(u) + \lambda(u,v) - L(u)(\tau^2(v) + \lambda(v,1)) - L(v)(\tau^2(u) + \lambda(u,1)) + L(u)L(v)\tau^2(1)]$$

for  $0 \leq u \leq v \leq 1$  når  $n$  er stor.

Prov. Setning 3 gir for store  $n$  når  $0 < u \leq v < 1$ ,

$$\begin{aligned} \text{cov}(\hat{L}(u), \hat{L}(v)) &\approx \frac{1}{n} \text{cov}\left(\sum_{i=1}^{\infty} h_i(u)Z_i, \sum_{j=1}^{\infty} h_j(v)Z_j\right) = \\ \frac{1}{n} \sum_{i=1}^{\infty} h_i(u)h_i(v) &= \frac{1}{n\mu^2} \left[ \sum_{i=1}^{\infty} q_i(u)q_i(v) - L(v) \sum_{i=1}^{\infty} q_i(u)q_i(1) - \right. \\ &\left. L(u) \sum_{i=1}^{\infty} q_i(v)q_i(1) + L(u)L(v) \sum_{i=1}^{\infty} q_i^2(1) \right]. \end{aligned}$$

Frå (5) følgjer da resultatet.

Merknad 4: Korollar 4 gir at  $\sqrt{n} \hat{L}(u)$  har asymptotisk varians  $\gamma^2(u)$  der  $\gamma^2(u)$  er definert ved

$$(6) \quad \gamma^2(u) = \frac{1}{\mu^2} [\tau^2(u) - 2L(u)(\tau^2(u) + \lambda(u,1)) + L^2(u)\tau^2(1)] \text{ for } 0 \leq u \leq 1.$$

Viss vi ønskjer å teste ein hypotese om eit gitt punkt på Lorenz-kurva, vil ein aktuell test følgje direkte frå neste korollar.

Korollar 5. Føreset at  $\gamma^2(u) < \infty$  for  $0 < u < 1$ . Under føresetnadene i Setning 1 er

$$\sqrt{n} \frac{\hat{L}(u) - L(u)}{\gamma(u)}$$

asymptotisk  $N(0,1)$  for fast  $u$ ,  $0 < u < 1$ .

Korollar 5 følgjer direkte frå Setning 3 og Korollar 4.

På grunn av samanhengane mellom  $M(\cdot)$ ,  $N(\cdot)$ ,  $O(\cdot)$ ,  $P(\cdot)$  og  $W(\cdot, \cdot)$  (sjå kapittel 2), kan vi avgrense oss til å studere asymptotiske eigenskapar for den empiriske prosessen  $\sqrt{n}(\hat{W}(u,v) - W(u,v))$ .

Setning 4. Under føresetnadene i Setning 1 blir

$$\lim \Pr(\sqrt{n} (\hat{W}(u,v) - W(u,v)) \leq k) = \Pr \left( \sum_{j=1}^{\infty} t_j(u,v) Z_j \leq k \right) \text{ for } 0 \leq u, v < 1$$

der

$$t_j(u,v) = \frac{1-v}{u(1-L(v))} \left[ h_j(u) + \frac{L(u)}{1-L(v)} h_j(v) \right]$$

og  $Z_1, Z_2, \dots$  er uavhengige og standard normalfordelte variable.

Prov. La  $0 \leq u, v < 1$ . Følgeleg blir  $L(v) < 1$ . Vi har at

$$\begin{aligned} \sqrt{n} (\hat{W}(u,v) - W(u,v)) &= \sqrt{n} \frac{1-v}{u} \left( \frac{\hat{L}(u)}{1-L(v)} - \frac{L(u)}{1-L(v)} \right) = \\ &= \frac{1-v}{u(1-L(v))} (X_n(u) + \frac{L(u)}{1-L(v)} X_n(v)) \end{aligned}$$

der

$$X_n(u) = \sqrt{n} (\hat{L}(u) - L(u)).$$

I provet for Setning 3 viste vi at  $X_n(u) \xrightarrow{D} X(u)$ , der

$$X(u) = \frac{1}{\mu} \left[ \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt - L(u) \int_0^1 \frac{W_0(t)}{f(F^{-1}(t))} dt \right].$$

Dermed har vi at

$$X_n(u) + \frac{L(u)}{1-L(v)} X_n(v) \xrightarrow{D} X(u) + \frac{L(u)}{1-L(v)} X(v).$$

Sidan  $\hat{L}(v) \xrightarrow{P} L(v)$  (følgjer frå Setning 3) gir Cramer-Slutskeys setning at

$$\sqrt{n} (\hat{W}(u,v) - W(u,v)) \xrightarrow{D} \frac{1-v}{u(1-L(v))} (X(u) + \frac{L(u)}{1-L(v)} X(v)).$$

Ved å nytte Setning 2 finn vi at  $X(u) \stackrel{D}{=} \sum_{j=1}^{\infty} h_j(u) Z_j$ . Herav følgjer resultatet i Setning 4.

La  $v(u,v)$  vere definert ved

$$(7) \quad v(u,v) = \frac{1}{\mu^2} [\tau^2(u) + \lambda(u,v) - L(u)(\tau^2(v) + \lambda(v,1)) - L(v)(\tau^2(u) + \lambda(u,1)) + L(u)L(v)\tau^2(1)], \quad 0 \leq u \leq v \leq 1.$$



Setning 4 gir no følgjande resultat,

Korollar 6. Under føresetnadene i Setning 1 blir

$$\text{var } (\sqrt{n} \hat{W}(u,v)) \approx \left( \frac{1-v}{u(1-L(v))} \right)^2 [v(u,u) + 2 \frac{L(u)}{1-L(v)} \sum_{i=1}^{\infty} h_i(u) h_i(v) + \left( \frac{L(u)}{1-L(u)} \right)^2 v(v,v)]$$

for  $0 < u < 1$  og  $0 \leq v \leq 1$  når  $n$  er stor.

I provet for Korollar 4 viste vi at

$$(8) \quad \sum_{i=1}^{\infty} h_i(u) h_i(v) = v(u,v), \quad 0 \leq u \leq v \leq 1.$$

Ved å bruke (8) får vi frå Korollar 6 følgjande asymptotiske varianser for  $\sqrt{n} \hat{M}(u)$ ,  $\sqrt{n} \hat{N}(u)$ ,  $\sqrt{n} \hat{P}_1(u)$  og  $\sqrt{n} \hat{P}_2(u)$ ,

$$(9) \quad \text{var } (\sqrt{n} \hat{M}(u)) \approx \frac{1}{u^2} v(u,u), \quad 0 < u < 1$$

$$(10) \quad \text{var } (\sqrt{n} \hat{N}(u)) \approx \frac{1}{(1-L(1-u))^2} [v(u,u) + 2 \frac{L(u)}{1-L(1-u)} v(u,1-u) + \left( \frac{L(u)}{1-L(1-u)} \right)^2 v(1-u, 1-u)], \quad 0 < u \leq \frac{1}{2}$$

$$\text{var } (\sqrt{n} \hat{N}(u)) \approx \frac{1}{(1-L(1-u))^2} [v(u,u) + 2 \frac{L(u)}{1-L(1-u)} v(1-u,u) + \left( \frac{L(u)}{1-L(1-u)} \right)^2 v(1-u, 1-u)], \quad \frac{1}{2} < u < 1.$$

$$(11) \quad \text{var } (\sqrt{n} \hat{P}_1(u)) \approx \left( \frac{u}{1-u} \right)^2 v(u,u), \quad 0 < u < 1$$

$$(12) \quad \text{var } (\sqrt{n} \hat{P}_2(u)) \approx \left[ \left( \frac{1-u}{u(1-L(u))} \right) \left( 1 + \frac{L(u)}{1-L(u)} \right) \right]^2 v(u,u), \quad 0 < u < 1.$$

Resultata i Setning 4 og Korollar 6 gir opphav til slutttningsreglar (testar) for hypotesar om faste punkt på ulikskapskurvene  $M(\cdot)$ ,  $N(\cdot)$ ,  $P_1(\cdot)$  og  $P_2(\cdot)$ . I praksis må vi derfor estimere  $\mu$ ,  $L(u)$ ,  $\tau^2(u)$  og  $\lambda(u,v)$ .

Vi har at

$$\hat{\mu} = \bar{X}$$

og

$$\hat{L}(u) = \frac{\hat{uH}(u)}{\bar{X}}$$

der  $\hat{H}(u)$  er gitt ved (5.1), er konsistente estimatorar for respektive  $\mu$  og  $L(u)$ . Ved å erstatte  $F$  med  $F_n$  i uttrykka for  $\tau^2(u)$  og  $\lambda(u,v)$  får vi konsistente estimatorar for respektive  $\tau^2(u)$  og  $\lambda(u,v)$ :

$$(13) \quad \hat{\tau}^2(u) = 2 \int_0^{F_n^{-1}(u)} [(1-F_n(y)) \int_0^y F_n(x) dx] dy =$$

$$\frac{2}{n} \sum_{j=1}^{i-1} \left[ \left(1 - \frac{j}{n}\right) (X_{(j+1)} - X_{(j)}) \left( \sum_{\ell=1}^j (X_{(\ell+1)} - X_{(\ell)}) \right) \right]$$

når

$$u \in \left[ \frac{i}{n}, \frac{i+1}{n} \right), \quad 2 \leq i.$$

$$(14) \quad \hat{\lambda}(u,v) = \int_{F_n^{-1}(u)}^{F_n^{-1}(v)} \int_0^{F_n^{-1}(u)} F_n(x) (1-F_n(y)) dx dy =$$

$$\frac{1}{n} \left[ \sum_{\ell=1}^i (X_{(\ell+1)} - X_{(\ell)}) \right] \left[ \sum_{\ell=i}^{j-1} \left(1 - \frac{\ell}{n}\right) (X_{(\ell+1)} - X_{(\ell)}) \right]$$

når

$$u \in \left[ \frac{i}{n}, \frac{i+1}{n} \right), \quad v \in \left[ \frac{j}{n}, \frac{j+1}{n} \right) \quad \text{og} \quad i \leq j-1.$$

Ved å bruke Setning 4 og Cramer-Slutsksys setning får vi at

$$\sqrt{n} \frac{\hat{M}(u) - M(u)}{\hat{\sigma}_M(u)}, \quad \sqrt{n} \frac{\hat{N}(u) - N(u)}{\hat{\sigma}_N(u)},$$

$$\sqrt{n} \frac{\hat{P}_1(u) - P_1(u)}{\hat{\sigma}_{P_1}(u)} \quad \text{og} \quad \sqrt{n} \frac{P_2(u) - P_2(u)}{\hat{\sigma}_{P_2}(u)}$$

er asymptotisk  $N(0,1)$  for fast  $u$ , der

$$(15) \quad \hat{\sigma}_M^2(u) = \frac{1}{u^2} \hat{v}(u,u), \quad 0 < u < 1$$

$$(16) \quad \hat{\sigma}_N^2(u) = \frac{1}{(1-\hat{L}(1-u))^2} [\hat{v}(u,u) + 2 \frac{\hat{L}(u)}{1-\hat{L}(1-u)} \hat{v}(u,1-u) +$$

$$\left( \frac{\hat{L}(u)}{1-\hat{L}(1-u)} \right)^2 \hat{v}(1-u,1-u)], \quad 0 < u \leq \frac{1}{2}$$

$$\hat{\sigma}_N^2(u) = \frac{1}{(1-\hat{L}(1-u))^2} [\hat{v}(u,u) +$$

$$2 \left( \frac{\hat{L}(u)}{1-\hat{L}(1-u)} \right) \hat{v}(1-u,u) + \left( \frac{\hat{L}(u)}{1-\hat{L}(1-u)} \right)^2 \hat{v}(1-u,1-u)], \quad \frac{1}{2} < u < 1,$$

$$(17) \quad \hat{\sigma}_{P_1}^2(u) = \left( \frac{u}{1-u} \right)^2 \hat{v}(u,u), \quad 0 < u < 1,$$

$$(18) \quad \hat{\sigma}_{P_2}^2(u) = \left( \frac{1-u}{u(1-\hat{L}(u))} \right)^2 \hat{v}(u,u), \quad 0 < u < 1,$$

der

$$(19) \quad \hat{v}(u,v) = \left( \frac{1}{X} \right)^2 [\hat{\tau}^2(u) + \hat{\lambda}(u,v) - \hat{L}(u) (\hat{\tau}^2(v) + \hat{\lambda}(v,1)) -$$

$$\hat{L}(v) (\hat{\tau}^2(u) + \hat{\lambda}(u,1)) + \hat{L}(u)\hat{L}(v) \hat{\tau}^2(1)], \quad 0 \leq u \leq v < 1.$$

Merknad 5: Merknad 3 og (13) gir at

$$\frac{2}{n} \sum_{j=1}^{n-1} \left[ \left( 1 - \frac{j}{n} \right) (X_{(j+1)} - X_{(j)}) \left( \sum_{\ell=1}^j \ell (X_{(\ell+1)} - X_{(\ell)}) \right) \right]$$

er ein konsistent estimator for den asymptotiske variansen  $(\tau^2(1))$  til  $\sqrt{n}$  ganger gjennomsnittet av  $n$  (ikkje-negative) uavhengige og identisk fordelte variable.

Diskusjon: Resultata ovafor gir oss grunnlag for å vurdere kor nøyaktig og påliteleg ei empirisk ulikskapskurve er i eit gitt punkt. Ved å konstruere konfidensbandfor ulikskapskurvene, vil vi oppnå ei simultan vurdering

av kor påliteleg informasjonen frå ei empirisk ulikskapskurve blir. For å utvikle konfidensband for ulikskapskurvene i kapittel 5.2, må vi etablere asymptotiske fordelingar for observatorar av typen

$$K_n = \sqrt{n} \sup_{0 \leq u \leq 1} \frac{|\hat{L}(u) - L(u)|}{\Psi(\hat{L}(u))}$$

der  $\Psi$  er ein kontinuerleg vektfunksjon.  $K_n$  vil gi opphav til konfidensband for Lorenz-kurva  $L(\cdot)$ .

Setning 3 og Teorem 5.1, Billingsley (1968) gir at

$$\lim \Pr(K_n \leq k) = \Pr \left( \sup_{0 \leq u \leq 1} \left| \sum_{j=1}^{\infty} \frac{h_j(u)}{\Psi(L(u))} Z_j \right| \leq k \right).$$

For å kunne konstruere konfidensband for Lorenz-kurva på grunnlag av  $K_n$ , må vi derfor finne fordelinga til

$$\sup_{0 \leq u \leq 1} \left| \sum_{j=1}^{\infty} \frac{h_j(u)}{\Psi(L(u))} Z_j \right|$$

for ein passende vektfunksjon  $\Psi$ .

#### b. Asymptotiske resultat for estimatorene til ulikskapsmåla A, B, C, D og G.

Vi skal først vise at ei vega utgåve av den naturlege estimatoren  $\hat{G}$  for Gini-koeffisienten  $G$  er asymptotisk normalfordelt.  $\hat{G}$  er definert ved (5.14). La  $\gamma^2$  vere ein parameter definert ved

$$(20) \quad \gamma^2 = \frac{4}{u^2} \left[ 2 \int_0^1 \int_0^v (\tau^2(u) + \lambda(u,v)) du dv - \right. \\ \left. (1-G) \int_0^1 (\tau^2(u) + \lambda(u,1)) du + \frac{1}{4}(1-G)^2 \tau^2(1) \right]$$

der  $\tau^2(u)$  og  $\lambda(u,v)$  er definert ved respektive (3) og (4).

Setning 5. Føreset at  $\gamma^2 < \infty$ . Under føresetnadene i Setning 1 blir da

$$\sqrt{n} \frac{\hat{G} - G}{\gamma}$$

asymptotisk  $N(0,1)$ .

Prov. Vi har at

$$\sqrt{n} (\hat{G}-G) = -2 \int_0^1 \sqrt{n} (\hat{L}(u) - L(u)) du.$$

I Setning 3 viste vi at

$$\sqrt{n} (\hat{L}(u) - L(u)) \xrightarrow{D} X(u)$$

der

$$X(u) = (Y(u) - L(u) Y(1)) / \mu$$

og

$$Y(u) = \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt.$$

Det følger nå direkte fra argumentasjonen i kap. 4.4, Durbin (1973) at

$\int_0^1 X(u) du$  som funksjon av  $X(u)$  er kontinuerleg i Skorokhod-topologien.

Dermed gir Teorem 5.1. Billingsley (1968) at

$$\sqrt{n} (\hat{G}-G) \xrightarrow{D} -2 \int_0^1 X(u) du.$$

Ved å bruke Setning 2 får vi videre at  $X(u) \stackrel{D}{=} \sum_{j=1}^{\infty} h_j(u) Z_j$ , der  $Z_1, Z_2, \dots$

er uavhengige og standard normalfordelte variable,  $h_j(u) = \frac{1}{\mu} (q_j(u) - q_j(1)L(u))$  og  $q_j(u)$  er definert ved (1). Dette gir

$$\lim \Pr(\sqrt{n} (\hat{G}-G) \leq k) = \Pr(-2 \int_0^1 (\sum_{j=1}^{\infty} h_j(u) Z_j) du \leq k).$$

Da får vi ved å bruke Fubinis teorem,

$$\lim \Pr(\sqrt{n} (\hat{G}-G) \leq k) = \Pr(-2 \sum_{j=1}^{\infty} (\int_0^1 h_j(u) du) Z_j \leq k).$$

Dermed har vi vist at  $\sqrt{n} (\hat{G}-G)$  er asymptotisk normalfordelt med forventning

0 og varians  $4 \sum_{j=1}^{\infty} [\int_0^1 h_j(u) du]^2$ . Vi skal nå vise at  $4 \sum_{j=1}^{\infty} [\int_0^1 h_j(u) du]^2 = \gamma^2$ .

$$\sum_{j=1}^{\infty} \left[ \int_0^1 h_j(u) du \right]^2 = \frac{1}{\mu^2} \sum_{j=1}^{\infty} \left[ \left( \int_0^1 q_j(u) du \right)^2 - 2q_j(1) \int_0^1 q_j(u) du \int_0^1 L(u) du + q_j^2(1) \left( \int_0^1 L(u) du \right)^2 \right].$$

Vidare får vi

$$\begin{aligned} \sum_{j=1}^{\infty} \left( \int_0^1 q_j(u) du \right)^2 &= \sum_{j=1}^{\infty} \iint_{00}^{11} q_j(u) q_j(v) du dv = \iiint_{0000}^{11vu} \frac{2}{f(F^{-1}(t))f(F^{-1}(s))} \\ &\sum_{j=1}^{\infty} \frac{\sin(j\pi t) \sin(j\pi s)}{(j\pi)^2} dt ds du dv = 2 \iint_{00}^{1v} \left[ 2 \iint_{00}^{us} \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds \right. \\ &+ \left. \iint_{u0}^{vu} \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds \right] du dv = \\ &2 \iint_{00}^{1v} \left[ 2 \int_0^{F^{-1}(u)} \int_0^y (F(x)(1-F(y))) dx dy + \right. \\ &\left. \int_{F^{-1}(u)}^{F^{-1}(v)} \int_0^{F^{-1}(u)} F(x)(1-F(y)) dx dy \right] du dv = \\ &2 \iint_{00}^{1v} (\tau^2(u) + \lambda(u,v)) du dv . \end{aligned}$$

Under utviklinga av uttrykket ovafor har vi nytta Fubinis teorem og (2).

Fubinis teorem og (5) gir at

$$\begin{aligned} \sum_{j=1}^{\infty} q_j(1) \int_0^1 q_j(u) du &= \int_0^1 \left( \sum_{j=1}^{\infty} q_j(1) q_j(u) \right) du = \\ &\int_0^1 (\tau^2(u) + \lambda(u,1)) du . \end{aligned}$$

Frå (5) har vi vidare at  $\sum_{j=1}^{\infty} q_j^2(1) = \tau^2(1)$ .

Sidan  $\int_0^1 L(u) du = \frac{1}{2}(1-G)$  får vi derfor

$$2 \int_0^1 L(u) du \sum_{j=1}^{\infty} (q_j(1) \int_0^1 q_j(u) du) = (1-G) \int_0^1 (\tau^2(u) + \lambda(u,1)) du$$

og

$$\sum_{j=1}^{\infty} [q_j(1) \int_0^1 L(u) du]^2 = \frac{\tau^2(1)}{4} (1-G)^2.$$

Vi har dermed vist at  $4 \sum_{j=1}^{\infty} \left[ \int_0^1 h_j(u) du \right]^2 = \gamma^2.$

La

$$(21) \quad \hat{\gamma}^2 = \frac{4}{X^2} \left[ 2 \iint_{00}^{1v} (\hat{\tau}^2(u) + \hat{\lambda}(u,v)) du dv - (1-\hat{G}) \int_0^1 (\hat{\tau}^2(u) + \hat{\lambda}(u,1)) du + \frac{1}{4}(1-\hat{G})^2 \hat{\tau}^2(1) \right].$$

$\hat{\gamma}^2$  er ein konsistent estimator for den asymptotiske variansen til  $\sqrt{n} \hat{G}.$

Dersom vi ønskjer å vurdere Gini-koeffisienten  $G$  i forhold til ein bestemt verdi  $G_0$ , kan vi på grunnlag av eit sett uavhengige og identisk fordelte observasjonar nytte observatoren  $\sqrt{n} (\hat{G}-G_0)/\hat{\gamma}$  for å konstruere nødvendige sluttningsreglar. Det naturlege i slike situasjonar er å formulere problemet som eit tre-desisjonsproblem. Vi ønskjer å avgjere om  $G < G_0$  eller  $G > G_0$ , dvs. vi har val mellom tre avgjerder:

$a_1 =$  påstå  $G < G_0$ ,  $a_2 =$  påstå  $G > G_0$ ,  $a_3 =$  ikkje seie noko.

La observasjonsmaterialet  $X$  ha fordeling  $F$ . Vi definerer no den statistiske metode  $\Psi$  ved

$$\Psi(X) = \text{betinga sannsynet for å velje } a_i \text{ gitt } X, \text{ der } \sum_{i=1}^3 \Psi_i(X) = 1.$$

Vi vil kreve at metoden har nivå  $\varepsilon$ , dvs. sannsyna for feilaktig å påstå  $G < G_0$  og feilaktig å påstå  $G > G_0$  skal vere høgst  $\varepsilon$ . I tillegg vil vi kreve at metoden skal vere prestasjonsrett, dvs. sannsyna for rett utsagn skal vere minst  $\varepsilon$ . Desisjonsregelen som gir grunnlag for å påstå

$$G < G_0 \text{ når } \sqrt{n} \frac{\hat{G}-G_0}{\hat{\gamma}} \leq -k_{1-\varepsilon}$$

og

$$G > G_0 \text{ når } \sqrt{n} \frac{\hat{G}-G_0}{\hat{\gamma}} \geq k_{1-\varepsilon}$$

har asymptotisk nivå  $\varepsilon$  og er asymptotisk prestasjonsrett.  $k_{1-\varepsilon}$  er  $1-\varepsilon$  frakti-  
len frå standard normalfordelinga.

Viss vi ønskjer å samanlikne Gini-koeffisientane for to populasjonar oppnår vi ein tilsvarande desisjonsregel ved å basere oss på observatoren

$$\sqrt{\frac{mn}{N}} \frac{\hat{G}_{1,m} - \hat{G}_{2,n}}{((1-\lambda)\hat{\gamma}_1^2 + \lambda\hat{\gamma}_2^2)^{\frac{1}{2}}}$$

der  $\hat{G}_{1,m}$  og  $\hat{G}_{2,n}$  er estimatorane for Gini-koeffisienten basert på utval frå respektive populasjon 1 (n observasjonar) og populasjon 2 (m observasjonar),  $N = m+n$  og  $m/n \rightarrow \lambda$ .

La  $\alpha^2$  vere ein parameter definert ved

$$(22) \quad \alpha^2 = \frac{1}{\mu^2} \left[ 2 \int_0^1 \int_0^1 \frac{1}{uv} (\tau^2(u) + \lambda(u,v)) du dv - \right. \\ \left. 2(1-A) \int_0^1 \frac{1}{u} (\tau^2(u) + \lambda(u,1)) du + (1+A)^2 \tau^2(1) \right]$$

der A definert ved (3.4) er ulikskapsmålet for ulikskapskurva  $M(\cdot)$  og  $\hat{A}$  er definert ved (5.16).

Setning 6. Føreset at  $\alpha^2 < \infty$ . Under føresetnadene i Setning 1 er

$$\sqrt{n} \frac{\hat{A} - A}{\alpha}$$

asymptotisk  $N(0,1)$ .

Prov. Vi har at

$$\sqrt{n} (\hat{A} - A) = - \int_0^1 \sqrt{n} (\hat{M}(u) - M(u)) du$$

I provet av Setning 4 viste vi at

$$\sqrt{n} (\hat{M}(u) - M(u)) = \sqrt{n} (\hat{W}(u,0) - W(u,0)) \stackrel{D}{\rightarrow} Z_1(u) \text{ for } 0 \leq u < 1,$$

der

$$Z_1(u) = \frac{1}{u\mu} [Y(u) - L(u)Y(1)]$$

og

$$Y(u) = \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt.$$

Dette resultatet gjeld også for  $u=1$  sidan  $\hat{M}(1) - M(1) = Z_1(1) = 0$ .



Da følger direkte frå grunngevinga av Setning 5 og frå Setning 2 at

$$\lim \Pr(\sqrt{n}(\hat{A}-A) \leq k) = \Pr(-\int_0^1 Z_1(u)du \leq k) = \Pr(-\sum_{j=1}^{\infty} (\int_0^1 \frac{h_j(u)}{u} du) Z_j \leq k).$$

For å finne  $\sum_{j=1}^{\infty} [\int_0^1 (h_j(u)/u)du]^2$  kan vi bruke den teknikken vi brukte for å

finne  $\sum_{j=1}^{\infty} [\int_0^1 h_j(u)du]^2$  i provet av Setning 5. Dette gir

$$\sum_{j=1}^{\infty} [\int_0^1 \frac{h_j(u)}{u} du]^2 = \alpha^2.$$

Dermed har vi vist at  $\sqrt{n}(\hat{A}-A)$  er asymptotisk normalfordelt med forventning 0 og varians  $\alpha^2$ .

La  $\beta^2$  vere ein parameter definert ved

$$(23) \quad \beta^2 = \frac{1}{\mu^2} [2 \iint_{00}^{1v} p(u)p(v) (\tau^2(u) + \lambda(u,v))du dv -$$

$$2(\int_0^1 p(u) L(u) du)(\int_0^1 p(u)(\tau^2(u) + \lambda(u,1))du) +$$

$$\tau^2(1) (\int_0^1 p(u) L(u)du)^2]$$

der

$$p(u) = \frac{1}{1-L(1-u)} - \frac{L(1-u)}{(1-L(u))^2}.$$

Vi skal no utvikle den asymptotiske fordelinga til  $\sqrt{n}(\hat{B}-B)$  der B definert ved (3.5) er ulikskapsmålet for ulikskapskurva  $N(\cdot)$ .  $\hat{B}$  er definert ved (5.17).

Setning 7. Føreset at  $\beta^2 < \infty$ . Under føresetnadene i Setning 1 er

$$\sqrt{n} \frac{\hat{B}-B}{\beta}$$

asymptotisk  $N(0,1)$ .

Prov: Vi har at

$$\sqrt{n}(\hat{B}-B) = -\int_0^1 \sqrt{n}(\hat{N}(u) - N(u))du = -\int_0^1 \sqrt{n}(\hat{W}(u,1-u) - W(u,1-u))du.$$

I provet av Setning 4 viste vi at

$$\sqrt{n} (\widehat{W}(u, 1-u) - W(u, 1-u)) \xrightarrow{D} Z_2(u) \quad \text{for } 0 \leq u < 1$$

der

$$Z_2(u) = \frac{1}{1-L(1-u)} (X(u) + \frac{L(u)}{1-L(1-u)} X(1-u))$$

og

$$X(u) = \frac{1}{\mu} \left[ \int_0^u \frac{W_0(t)}{f(F^{-1}(t))} dt - L(u) \int_0^1 \frac{W_0(t)}{f(F^{-1}(t))} dt \right].$$

Dette resultatet gjeld også for  $u=1$  sidan  $\lim_{u \rightarrow 1} (\widehat{N}(u) - N(u)) = \lim_{u \rightarrow 1} Z_2(u) = 0$ .

Då følgjer frå den grunngevinga vi ga i provet av Setning 5 at

$$\lim \Pr(\sqrt{n} (\widehat{B} - B) \leq k) = \Pr\left(-\int_0^1 Z_2(u) du \leq k\right).$$

Setning 2 gir at

$$Z_2(u) \stackrel{D}{=} \frac{1}{1-L(1-u)} \left[ \sum_{j=1}^{\infty} \left( h_j(u) + \frac{L(u)}{1-L(1-u)} h_j(1-u) \right) Z_j \right].$$

Ved å bruke dette resultatet saman med Fubinis teorem får vi derfor

$$\begin{aligned} \lim \Pr(\sqrt{n} (\widehat{B} - B) \leq k) = \\ \Pr\left( \sum_{j=1}^{\infty} \left( \int_0^1 \left( \frac{L(1-u)}{(1-L(u))^2} - \frac{1}{1-L(1-u)} \right) h_j(u) du \right) Z_j \leq k \right). \end{aligned}$$

Da står det att å vise at  $\sum_{j=1}^{\infty} \left[ \int_0^1 p(u) h_j(u) du \right]^2 = \beta^2$ . Dette resultatet kan

vi etablere ved å gjere bruk av den framgangsmåten som blei skissert i provet av Setning 5.

La  $\omega^2$  og  $\delta^2$  vere parametrar definert ved respektive

$$(24) \quad \omega^2 = \sum_{j=1}^{\infty} \left[ \int_0^1 \frac{(1-u) h_j(u)}{u(1-L(u))^2} du \right]^2$$

og

$$(25) \quad \delta^2 = \sum_{j=1}^{\infty} \left[ \int_0^1 \frac{h_j(u)}{1-u} du \right]^2$$

der

$$h_j(u) = \frac{1}{\mu} (q_j(u) - L(u) q_j(1)) \text{ og } q_j(u) \text{ er definert ved (1).}$$

Setning 8. Føreset at  $\omega^2 < \infty$ . Under føresetnadene i Setning 1 er

$$\sqrt{n} \frac{\hat{C} - C}{\omega}$$

asymptotisk  $N(0,1)$ .

Setning 9. Føreset at  $\delta^2 < \infty$ . Under føresetnadene i Setning 1 er

$$\sqrt{n} \frac{\hat{D} - D}{\delta}$$

asymptotisk  $N(0,1)$ .

Prova av Setning 8 og Setning 9 går langs dei same linjer som prova for Setning 6 og Setning 7.

På grunnlag av resultatata i Setning 6, Setning 7, Setning 8 og Setning 9 kan vi på tilsvarende måte som for Gini-koeffisienten  $G$  etablere desisjonsreglar for ulikskapsmåla  $A$ ,  $B$ ,  $C$  og  $D$ .

### c. Asymptotiske resultat for den empiriske fraktildifferansefunksjonen

Den empiriske fraktildifferansefunksjonen  $\hat{\Lambda}(t)$  er definert ved (5.18). Vi skal no studere asymptotiske eigenskapar for den empiriske prosessen

$$\sqrt{N} (\hat{\Lambda}(t) - \Lambda(t)) = \sqrt{\frac{N}{m}} \sqrt{m} (G_m^{-1}(t) - G^{-1}(t)) - \sqrt{\frac{N}{n}} \sqrt{n} (F_n^{-1}(t) - F^{-1}(t))$$

der  $N = m+n$ .

La  $m$  og  $n$  variere med  $N$  slik at  $m/n \rightarrow \lambda$  der  $0 < \lambda < 1$ .

I provet av Setning 1 viste vi at

$$\sqrt{n} (F_n^{-1}(t) - F^{-1}(t)) \xrightarrow{D} \frac{W_1(t)}{f(F^{-1}(t))}$$

og dermed at

$$\sqrt{m} (G_m^{-1}(t) - G^{-1}(t)) \xrightarrow{D} \frac{W_2(t)}{g(G^{-1}(t))}$$

der  $W_1(t)$  og  $W_2(t)$  er uavhengige Brownske bruer på  $[0,1]$ . Dette gir derfor følgjande resultat:

Setning 10. Føreset at definisjonsområda til dei kumulative fordelingsfunksjonane  $F$  og  $G$  er endelege intervall. Føreset vidare at  $F(x)$  og  $G(x)$  har kontinuerleg derivert  $f(x)$  og  $g(x)$  som er slik at  $0 < f(x) < \infty$  og  $0 < g(x) < \infty$  på definisjonsområda til respektive  $F$  og  $G$ . Da konvergerer den empiriske prosessen

$$\sqrt{N} (\hat{\Lambda}(t) - \Lambda(t))$$

i fordeling mot den Gaussiske prosessen

$$\frac{W_2(t)}{\lambda^{\frac{1}{2}} g(G^{-1}(t))} - \frac{W_1(t)}{(1-\lambda)^{\frac{1}{2}} f(F^{-1}(t))}$$

Korollar 7. Under føresetnadene i Setning 10 har  $\sqrt{N} (\hat{\Lambda}(t) - \Lambda(t))$  asymptotisk varians  $\kappa^2(t)$  der  $\kappa(t)^2$  er definert ved

$$(26) \quad \kappa^2(t) = \left( \frac{1}{\lambda g^2(G^{-1}(t))} + \frac{1}{(1-\lambda) f^2(F^{-1}(t))} \right) t(1-t)$$

Korollar 8. Føreset at  $\kappa^2(t) < \infty$ . Under føresetnadene i Setning 10 blir

$$\sqrt{N} \frac{\hat{\Lambda}(t) - \Lambda(t)}{\kappa(t)}$$

asymptotisk  $N(0,1)$  for fast  $t$ ,  $0 < t < 1$ .

Viss vi ønskjer å teste ein hypotese om eit fast punkt på fraktil-differansefunksjonen  $\Lambda(t)$ , kan vi etablere ein test eller ein desisjonsregel på grunnlag av resultatet i Korollar 8.

La  $F$  vere ei kjent fordeling. Eit  $(1-\varepsilon)$ -nivå, simultant, fordelingsfritt konfidensband for  $\Lambda(t)$ ,  $0 < t < 1$  er da gitt ved

$$\left[ G_n^{-1}\left(t - \frac{k}{\sqrt{n}}\right) - F^{-1}(t), G_n^{-I}\left(t + \frac{k}{\sqrt{n}}\right) - F^{-1}(t) \right]$$

der  $G_n^{-1}$  og  $G_n^{-I}$  representerer respektive venstre og høgre inversen og  $k$  er valt slik at

$$\Pr(\sqrt{n} \sup | G_n (\Lambda(t) + F^{-1}(t)) - t | \leq k) = 1 - \varepsilon.$$

$$\text{Sidan } \sqrt{n} \sup (G_n (\Lambda(t) + F^{-1}(t)) - t) = \sqrt{n} \sup (G_n (G^{-1}(t)) - t)$$

blir dette Kolmogorov-Smirnov observatoren. Vi kjenner derfor både eksakt og asymptotisk fordeling for observatoren

$\sqrt{n} \sup | G_n (\Lambda(t) + F^{-1}(t)) - t |$  og kan føygeleg bestemme  $k$  både for småe og store  $n$ .

## APPENDIKS 2

## FORMELL DRØFTING AV EIGENSKAPAR TIL ULIKSKAPSMÅLA A, B, C, D og G

Sidan ulikskapsmåla A, B, C, D og G tilfredsstillar overføringsprinsippet, er dei vare ovafor inntektsoverføringar på alle inntektsnivå. Det er derfor viktig å undersøke den vekta kvart av desse ulikskapsmåla gir til overføringar på forskjellige inntektsnivå.

Det er lett å vise at ulikskapsmåla A, B, C, D og G kan skrivast på forma

$$K = \int_0^{\infty} K(x) dF(x) = \int_0^1 K(F^{-1}(u)) du$$

der  $F$  er ein kumulativ fordelingsfunksjon og  $K(F^{-1}(\cdot))$  er ei ulikskapskurve.

Viss vi rangerer eit sett inntektsfordelingar etter verdiane på ulikskapsmålet  $K$ , vil effekten av ei lita overføring frå ein person med inntekt  $x_1$  til person med inntekt  $x_1-h$  ( $h$  liten) vere gitt ved  $K'(x_1) - K'(x_1-h)$  (sjå Atkinson (1970)). Ved å variere  $x_1$  får vi dermed eit bilde på kva slags vekt målet  $K$  gir inntektsoverføringar i forskjellige deler av fordelinga. På denne måten skal vi vise at Gini-koeffisienten gir størst vekt til inntektsoverføringar i toppdelen av fordelinga (når fordelinga er eintoppa).

Ved delvis integrasjon finn vi

$$\int_0^1 u dL(u) = 1 - \int_0^1 L(u) du,$$

som innsett i (3.1) gir

$$\begin{aligned} G &= \int_0^1 u dL(u) - \int_0^1 L(u) du = \frac{1}{\mu} \int_0^{\infty} xF(x) dF(x) - \int_0^{\infty} L(F(x)) dF(x) \\ &= \frac{1}{\mu} \int_0^{\infty} [xF(x) - \mu L(F(x))] dF(x). \end{aligned}$$

Da blir

$$K(x) = \frac{1}{\mu} [xF(x) - \mu L(F(x))]$$

og

$$K'(x) = \frac{1}{\mu} F(x).$$

Dermed blir effekten av ei lita overføring mellom to inntekter som er nesten like store ( $x_1$  og  $x_1-h$ ) proporsjonal med  $F(x_1) - F(x_1-h)$ . Dette vil seie at Gini-koeffisienten gir størst vekt til inntektsoverføringer i toppdelen av fordelinga.

Ved å drøfte A, B, C og D i forhold til G skal vi, med utgangspunkt i kunnskapen om G, etablere kunnskap om eigenskapar for A, B, C og D.

Sidan

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

og

$$A = 1 - \int_0^1 \frac{L(u)}{u} du,$$

ser vi at forskjellen på integrandane i dei to uttrykka er at  $L(u)$  får vekta 2 i G og vekta  $1/u$  i A.  $L(u)$  i integranden for A har altså ei vekt som gir større effekt til mindre  $u$  er. Sidan G gir størst vekt til overføringer i den sentrale delen (toppen) av fordelinga, vil derfor A gi størst vekt til overføringer i den nedre delen av fordelinga. Av same grunn finn vi at B legg størst vekt på overføringer i den nedre delen av fordelinga ( $0 \leq 1 - L(1-u) \leq 1$  og  $1 - L(1-u)$  er veksande som funksjon av  $u$ ). I tillegg har vi at A gir større vekt enn B til overføringer i den nedre delen av fordelinga. Dette følgjer av at  $1 - L(1-u) \geq u$  ( $L(u) \leq u$ ).

Legg merke til at ulikskapskurva  $P_1(\cdot)$  kan skrivast på forma

$$P_1(u) = \frac{EX}{uE(X|X \geq F^{-1}(u))} L(u).$$

Følgeleg har  $L(u)$  i integranden for C ei vekt som gir større vekt til mindre  $u$  er. Vidare har vi at  $[uE(X|X \geq F^{-1}(u))]/EX \geq u$ . Av betraktningane ovafor følgjer det derfor at C gir størst vekt til inntektsoverføringer i den nedre delen av fordelinga, men at også C gir mindre vekt enn A til inntektsoverføringer i denne delen av fordelinga.

I ulikskapsmålet D får  $L(u)$  vekta  $1/(1-u)$ , som er ein veksande funksjon av  $u$ . Følgjeleg gir D størst vekt til inntektsoverføringer i den øvre delen av fordelinga.

## LITTERATUR

- Atkinson, A.B. (1970). "On the Measurement of inequality". *Journal of Economic Theory*, Vol. 2.
- Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.
- Dalton, H. (1920). "The Measurement of the Inequality of Incomes". *Economic Journal*, Vol. 30.
- Doksum, K. (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case". *The Annals of Statistics*, Vol. 2, No. 2, 267-277.
- Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics, Philadelphia.
- Gastwirth, J.L. (1972). "The Estimation of the Lorenz curve and the Gini index". *The Review of Economics and Statistics*, 54, 306-316.
- Hajek, J & Sidak, Z. (1967). *Theory of Rank Tests*. Academic Press. New York.
- Lorenz, M.O. (1905). "Methods for Measuring Concentration of Wealth. *JASA*, Vol. 9.
- Royden, H. (1963). *Real Analysis*. Macmillan. New York.
- Sen, A. (1972). *On Economic Inequality*. Clarendon Press. Oxford.
- Shorack, G.R. (1972). "Function of Order Statistics". *Ann. Math. Statist.*, Vol. 43, No.2, 412-427.

Utkome i serien Rapporter frå Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422

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 Nils Håvard Lund: Byggekostnadsindeks for boliger Sidetall 127 Pris kr 15,00 ISBN 82-537-1232-4
- 81/4 Anne Lise Ellingsæter: 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 Bjørn Kjensli: 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 utvalgundersøkelser 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 Morten Reymert: 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: Framskriving av folke­mengden etter ekteskape­lig 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 Foyen: Miljøverninvesteringer i industrien. Problemer ved kartlegging av data Sidetall 34 Pris kr 10,00 ISBN 82-537-1591-9



Utkome i serien Rapportar frå Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422 (framh.)

Trykt 1981

- Nr. 81/22 Petter R. Koren: Etterspørsele 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 Levekårsundersøkingen 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 skattetal 1981 Current Tax Data Sidetall 46 Pris kr 10,00  
ISBN 82-537-1610-9
- 81/27 Tiril Vogt: Planregnskap Ressursregnskap for fysisk planlegging Sidetall 70  
Pris kr 15,00 ISBN 82-537-1614-1
- 81/28 Figurer i publikasjonar 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
- 81/30 Adne Cappelen, Erik Garaas og Svein Longva: MODAG En modell for makroøkonomiske  
analyser Sidetall 70 Pris kr 15,00 ISBN 82-537-1630-3
- 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 fangstnæringa 1977 - 1980 Sidetall 39 Pris kr 10,00  
ISBN 82-537-1634-6
- 81/34 Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1980  
Sidetall 272 Pris kr 20,00 ISBN 82-537-1636-2
- 81/35 Sigurd Tveitereid og Jan Lædre: Markedsindikatorer for norsk eksport Sidetall 47  
Pris kr 15,00 ISBN 82-537-1645-1

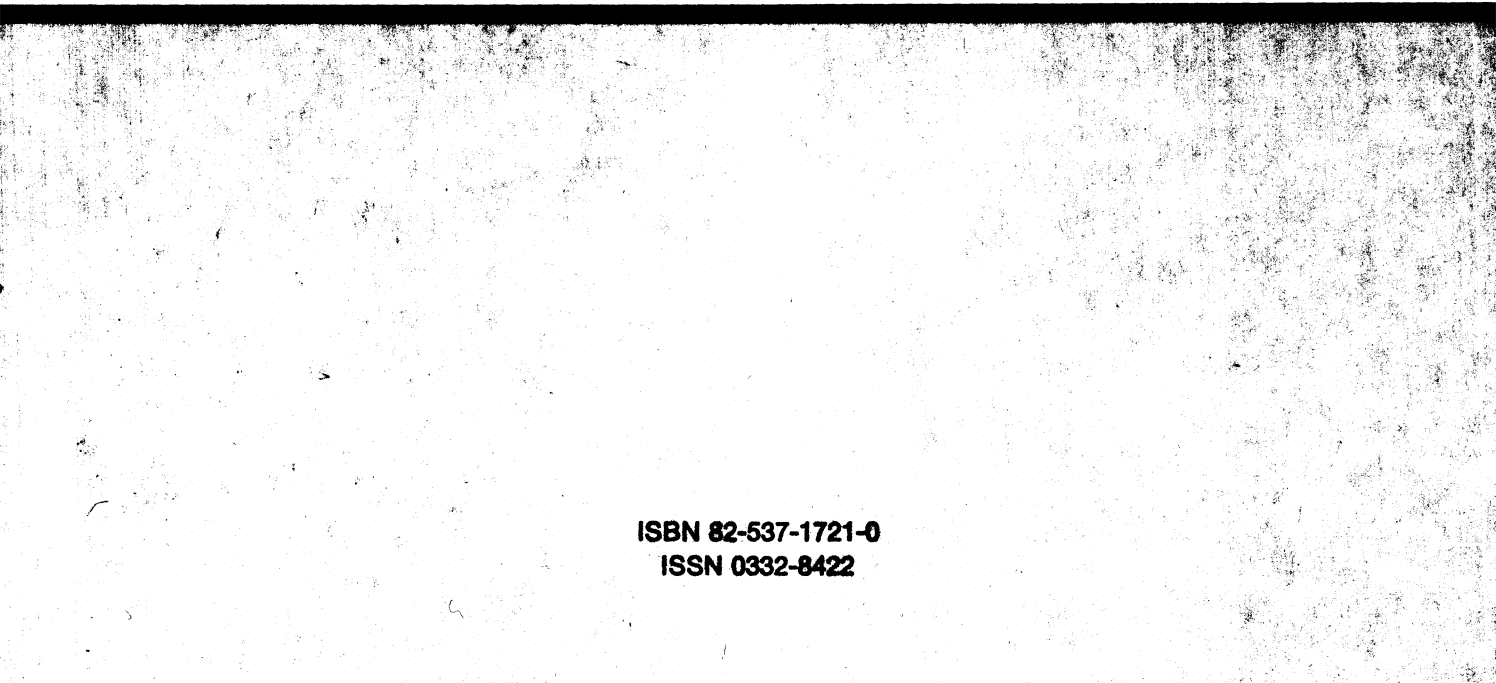
Trykt 1982

- Nr. 82/5 Naturressurser 1981 Sidetall 29 Pris kr 10,00 ISBN 82-537-1651-6
- 82/9 Rolf Aaberge: Om måling av ulikskap Sidetall 80 Pris kr 15,00 ISBN 82-537-1721-0



**Pris kr 15,00**

**Publikasjonen kjem ut i kommisjon hos H. Aschehoug & Co. og  
Universitetsforlaget, Oslo, og er til sals hos alle bokhandlarar.**



**ISBN 82-537-1721-0  
ISSN 0332-8422**