

Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

IO 67/5

Oslo, 21. august 1967

FAKTORER SOM PÅVIRKER INNTEKTENS STØRRELSE FOR INDIVIDER

I NORGE

a v

Per A. Vattle

I N N H O L D

	Side
I. Innledning	1
II. Generelt om problemet	1
III. Modell	2
IV. Egenskaper ved restleddet	3
V. Data	4
1. Brukbarheten av tallmaterialet	5
VI. Overgang til en funksjonsform som kan benyttes ved det videre arbeid	6
1. Geografisk område	6
2. Næring	7
3. Kjønn	8
4. Fellesliknet	9
5. Stilling	9
6. Alder	9
7. Sluttmerknader	11
VII. Variabelspesifikasjon	13
VIII. Estimeringsmetode og statistiske tester	14
IX. Enkelte resultater og tolkninger	16
Litteraturliste	22
Resultattabell	23

Dette arbeid er opprinnelig skrevet som seminaroppgave ved det sosialøkonomiske studium. Forfatteren har stått fritt i valg av opplegg og undersøkelsesmetoder. Arbeidet gjengis her med mindre endringer som forfatteren har ønsket å foreta. Synspunkter og konklusjoner står for forfatterens regning.

Ikke for offentliggjøring. Dette notat er et arbeidsdokument og kan siteres eller refereres bare etter spesiell tillatelse i hvert enkelt tilfelle. Synspunkter og konklusjoner kan ikke uten videre tas som uttrykk for Statistisk Sentralbyrås oppfatning.

I. Innledning

I denne undersøkelsen skal vi forsøke å belyse hvilke faktorer som påvirker inntektens størrelse for individer i Norge. Dette skal vi gjøre ut fra betraktninger i den økonomiske teori og så ved hjelp av et regresjonsopplegg teste resultatene når vi benytter tallmaterialet fra inntektsstatistikken for 1962.

Noen forskere som i de senere år har benyttet seg av denne type opplegg, men dog uten å foreta noen teoretiske betraktninger, er i U.S.A. F.G. Adams [1]¹⁾ og i England T.P. Hill [2]. En kort oversikt over disse undersøkelsene samt et liknende opplegg finnes i en bok av R. L. Klein [3]. Cand. oecon. Ivar Østby [4] har også benyttet seg av dette opplegget i en undersøkelse av norske data.

II. Generelt om problemet

Av primærmaterialet til inntektsstatistikken får man ikke bare opplysninger om forskjellige former for inntekt, formue, gjeld og fradragsposter, men også en rekke tilleggsopplysninger om hver telleenhet. Disse opplysningene er for eksempel alder, bosted, stilling, kjønn osv. Når så materialet fra undersøkelsen skal bearbeides og publiseres, blir det gjerne foretatt grupperinger i tabeller der flere av disse kjennetegnene blir benyttet som grupperingsstørrelser.

I denne undersøkelsen skal vi ta for oss problemet med hvilke faktorer som påvirker inntektens størrelse. Når vi skal forsøke å forklare den, kan vi gruppere den i en rekke tabeller. Som forklaringsvariable kan vi benytte de ovenfornevnte tilleggsopplysningene. Dette blir gjort ut fra den antakelse at disse størrelsene har betydning for forklaringen av inntektens størrelse. Denne påstand er imidlertid vanskelig å få undersøkt riktigheten av med denne vanlige bearbeidingsmetoden. De variasjonene som vi vil kunne observere fra en forklaringsvariabel til en annen og også innenfor en og samme variabel som er oppsplittet, kan komme av tilfeldige variasjoner i materialet. En viktig ting som kommer i tillegg er at vi ikke er i stand til å undersøke om vi bevisst eller ubevisst har utelatt eller mangler opplysninger om faktorer som signifikant kunne ha forklart en del av variasjonen i inntekten.

For å kunne undersøke de ovenfor nevnte problemer må man ha en teori om hvilke faktorer som påvirker inntekten til et individ. Selv om vår teoretiske modell gir opplysninger om en rekke faktorer som skal forklare

1) [] viser til litteraturlisten side 22.

inntektens størrelse, kan det tenkes at vi i vår teori ikke har tatt hensyn til alle de viktigste forklaringsvariable. Videre kan en rekke forklarende faktorer som vi finner i vår modell vise seg ikke å ha noen signifikant forklarende effekt, og/eller at den oppdeling som vi foretar av disse faktorene (ofte ut fra tidligere praksis) er uten interesse eller bør utføres på en annen måte.

En del av disse problemene som er skissert ovenfor, skal vi forsøke å belyse ved hjelp av en regresjonsanalyse sammen med statistiske tester av observasjonsmaterialet.

III. Modell²⁾

Lønnen til et individ blir bestemt av likevekten mellom tilbud og etterspørsel etter dets arbeidskraft. Nå er det slik at for mange mennesker blir deres inntekt fastsatt ved tariff-forhandlinger mellom forskjellige organisasjoner. Det bør vel i denne forbindelse nevnes at også bedriftseiernes inntekter kan bli bestemt av forhandlinger både med arbeidernes organisasjoner og med andre, f.eks. myndighetene, jfr. jordbruksavtalen. Av det som er sagt ovenfor, kan en få det inntrykket at individenes inntekt er helt fastlåst gjennom disse forhandlingene. Dette er ikke tilfellet. I mange tilfelle vil en række andre faktorer virke inn på lønnen. Vi kan bare nevne lønnsglidning og personlige tillegg som kan skyldes faktorer som dyktighet, evner og interesse.

Vi kan se på alle de punktene som representerer likevektspunktene mellom tilbudt og etterspurt lønn for de forskjellige individene vi har observert i et L,c diagram som vist i fig. 1.

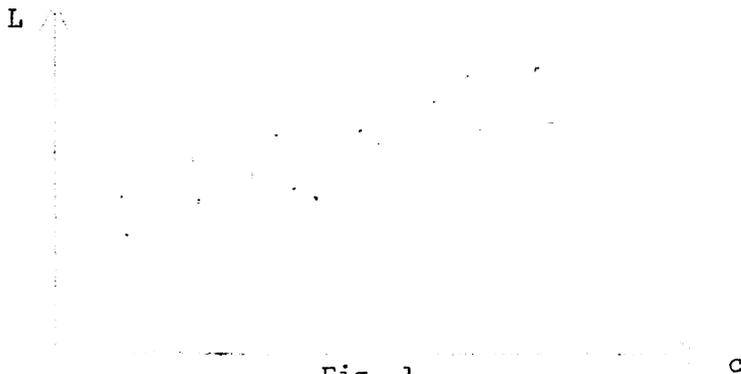


Fig. 1.

Her betegner c alle faktorer som er bestemmende for inntekten L.

Hvert punkt i fig. 1 er altså markedslikevektspunkter for individer

2) Dette kapittel er gjengitt sterkt forkortet.

med forskjellig c . Lar vi c være lik den ene størrelsen alder, vil vi finne at inntektens størrelse endres med alderen.

Det vi ønsker, er å finne en L -funksjon som vil uttrykke inntekten L som en funksjon av c og som vi kan betegne med

$$(1) \quad L = L(c)$$

IV. Egenskaper ved restleddet

Uten at vi har sagt noe om formen av L -funksjonen, skal vi komme med en rekke forutsetninger om det stokastiske restleddet u' . Dette vil ta vare på stokastiske variasjoner av inntekten, eventuelle målefeil i de variable og de ikke-observerbare restvariable, da det vil være uråd å få opplysning om alle størrelsene som vi har behandlet ovenfor og som vi antar er bestemmende for inntektens størrelse.

Vi har at relasjon (1) kan uttrykkes slik:

$$(2) \quad L = L(e, u')$$

Vi vil anta at de opplysningene som vi får om det enkelte individ og som inngår i c , er ikke-stokastiske størrelser uten målefeil. Lønnen L vil derimot være stokastisk, og med målefeil. Restleddet vil vi anta er uavhengig av c .

Det kan a priori være vanskelig å tillegge restleddet bestemte egenskaper. Vi skal her benytte oss av resultatene i de ovenfor nevnte artiklene³⁾. Her er det utført analyser av restleddene. Det viser seg at fordelingen av restleddene er skjev og at variansen av restleddene endrer seg med c når man benytter en additiv modell. Benyttes derimot en multiplikativ modell, er residualvariansen betydelig mer stabil ved endringer i c .

Ut fra det som er funnet ved liknende undersøkelser, vil jeg anta at i en additiv modell vil:

$$(3) \quad \text{var } u' = \sigma^2 L^2(c)$$

der σ er en konstant og $\text{var } u'$ betegner variansen av restleddet u' . Dette medfører at restleddet er innført multiplikativt i relasjon (2), slik at vi kan sette

$$(4) \quad L = L(c) \cdot (1 + u')$$

3) Se side 374 i [2]. Se også [1].

Tar vi den naturlige logaritmen av uttrykket i (4), får vi:

$$(5) \quad \ln L = \ln L(c) + \ln(1 + u') = \ln L(c) + u$$

Det stokastiske restleddet i (5) vil da bli uavhengig av L og c og alminnelig minste kvadraters regresjonsestimater vil bli forventningsrette og ha minimum varians⁴⁾.

Man kan så spørre seg om hva som vil skje dersom våre forutsetninger ovenfor ikke er tilfredsstilt med hensyn på betingelse (3). Vi vil i alle tilfelle få forventningsrette estimater, men variansene på estimatene vil bli større enn når vi tar hensyn til heteroscedastisity⁵⁾.

I det følgende skal vi altså utvikle en multiplikativ modell. For å estimere regresjonskoeffisientene skal vi benytte minstekvadraters metode. Vi forutsetter at det stokastiske restleddet u er identisk fordelt og uavhengig av c. E u = 0 og i den multiplikative modell vil vi forutsette var u = σ^2 . (σ her er en annen enn den som er benyttet ovenfor).

V. Data

Tallmaterialet er hentet fra Inntektsstatistikken 1962⁶⁾. Telleenheten i undersøkelsen for den del som omfattet personlige skattytere, var husholdningen. Husholdningen ble definert som personer som hadde samme bolig og som ble liknet sammen, samt ektefelle som krever særskilt likning. Hjemmeverende barn som ble liknet for seg selv, ble regnet som egen husholdning. Utvalgsplanen var satt opp slik at hver husholdning skulle ha samme sannsynlighet for å komme med i utvalget. Innsamlingen for 1962 ble basert på to utvalg. Det ene ble trukket slik at det dekket vel 0,5 prosent av alle husholdningene. Dette omfattet husholdninger med under kr. 80 000 i antatt inntekt ved statskattelikningen i 1961. Det andre utvalget var totaltelling av de med over kr. 80 000 i antatt inntekt i 1961.

Tallmaterialet ble innhentet gjennom likningskontorene fra selvangivelsene til de utvalgte husholdningene. Direktoratet for sjømenn gav opplysninger om inntekt opptjent i handelsflåten. I Statistisk Sentralbyrå har enkelte inntektsposter blitt beregnet, så som "soldatlønn" for alle personer som avtjener verneplikt, og alders- og uføretrygd der den manglet.

I tillegg til de økonomiske opplysningene får vi også en rekke kjennetegn for hovedinntektstakeren i husholdningen. Dette er fødselsår, stilling, bransje eller næring, bostedskommune, om vedkommende var fellesliknet, kjønn, gift eller ugift og forsørgelsesbyrde

4) Se [5] og [8].

5) Se [5] kap. 8-2.

6) Se [9].

1. Brukbarheten av tallmaterialet

I denne undersøkelsen har vi bare benyttet den delen av materialet som omfatter personlige skattytere hvis hushold skal ha mindre enn kr. 80 000 i antatt inntekt til statskattelikningen. Dette utvalget er på 9 035 husholdninger.

Det inntektsbegrepet som vi kommer til å benytte er:

Antatt inntekt ombord (i skip)
 + Netto inntekt ved statsskattelikningen
 + Inntekt som var for lav til å bli iliknet skatt
 = Netto inntekt til husholdningen

Det inntektsbegrepet som vi skulle ha hatt, var brutto inntekt til hvert individ. Grunnen til at vi benytter netto inntekt, er at grunnmaterialet forelå på magnetbånd og et program som kunne hente ut nettoinntekt, var tilgjengelig. Da materialet var innhentet for husholdningen, vil inntekten ofte skrive seg fra ett individ, mens den i noen tilfelle vil komme fra flere som har blitt liknet sammen. Disse feilkildene vil medføre at inntektene blir for lave når det i husholdet er bare en inntektstaker. Der hvor det er flere individer som har inntekt, vil hovedpersonen i husholdningen av og til få for stor, av og til for liten inntekt i forhold til det inntektsbegrepet vi ønsket å benytte. I tillegg til disse feilkildene kan vi få et tredje, nemlig bevisst feilaktige opplysninger om inntektens størrelse. Hvordan disse feilkildene vil virke på den inntektsstørrelsen som vi skulle hatt, er vanskelig å uttale seg om. Det er vel grunn til å anta at som oftest vil den netto inntekten til husholdningene som vi vil benytte i beregningene, være mindre enn den virkelige nettoinntekten. Denne vil igjen være mindre enn bruttoinntekten til hovedpersonen.

Opplysningene om hovedpersonen antar vi er uten feil. Unntatt fra denne regel kan være bransje eller næring og stilling. Særlig i fiske og jordbruk vil det være mange som er sysselsatt i andre næringer og som også vil få en annen stilling. Som eksempel kan tas at en bonde er selvstendig næringsdrivende i en del av året, men anleggsarbeider i bygg og anlegg den resterende del. Han er da lønnstaker. Det kan i slike tilfelle være vanskelig å avgjøre hvilken næring og stilling vedkommende skal tilhøre. Ofte blir den som gir størst inntekt ført opp, men både den man har lengst tid eller den som gir størst sosial anseelse, kan også bli ført opp.

Som konklusjon kan vi si at vårt materiale skulle være brukbart til denne undersøkelsen, men vi må være oppmerksomme på dets svakheter og mangler.

VI. Overgang til en funksjonsform som kan benyttes ved det videre arbeid

Ovenfor har vi utledet en teoretisk modell for lønnen til et individ når det befinner seg i likevektspunktet mellom tilbud og etterspørsel. Noen av de faktorene som er bestemmende for lønnen legger bedriften særlig vekt på, mens andre påvirkes hovedsakelig av det enkelte individ. Som vi har nevnt ovenfor, er vi interessert i å teste ut fra vårt observasjonsmateriale hvilke faktorer som påvirker inntekten. Slik som relasjon (5) står, kan den ikke benyttes til dette formålet. Vi må omforme uttrykket på en slik måte at vi kan benytte det ved de numeriske beregningene.

Som vi har sett av gjennomgåelsen av tallmaterialet, vil inntekten og alderen være kvantitative størrelser. Alle de andre opplysningene om hovedpersonen i husholdningene vil være kvalitative størrelser. Dette vil medføre at vi må benytte oss av dummy-variable når de kvalitative kjennetegnene skal innføres i funksjonen. Vi innfører også en funksjon av alderen. Modellen vil få følgende form:

$$(6) \quad \ln L = \alpha + \alpha' f(A) + \sum_j \gamma_j Q_j + u$$

der α , α' og γ_j er konstanter. $f(A)$ er en funksjon av alderen, og Q_j er enten 1 eller 0 alt ettersom individet har en bestemt kvalitativ størrelse j eller ikke. u er det stokastiske restleddet.

Modellen er relativt enkel, idet vi ikke har tatt hensyn til samspillet mellom de forskjellige variabelgruppene. Dette medfører at vi regner en bestemt effekt på inntekten som følge av at individet har en bestemt stilling og en effekt som skyldes at det bor i et bestemt område. Vi har altså ikke regnet med noen separat lønnseffekt som skyldes at individet både har en bestemt stilling og bor i et bestemt område av landet.

I det følgende skal vi ta for oss hver enkelt faktor som vi mener påvirker inntektens størrelse og se hvilken gruppering vi vil foreta i modellen.

1. Geografisk område

På grunn av de geografiske forhold i vårt land, med store avstander og befolkningen på sett og vis samlet i områder som er avgrenset av naturlige hindringer som fjell og fjorder, kan man lett tenke seg landet oppdelt i områder som naturlig hører sammen. Som eksempel kan vi tar den grove inndelingen i landsdeler: Øst-, Sør- og Vestlandet, Trøndelag og Nord-Norge. En liknende oppdeling er den som følger Standard for handelsområder. Denne

oppdelingen tar mer hensyn til de geografiske områdene som hører naturlig sammen når det gjelder handel og transport. Dette vil også ha innflytelse på lønnen til et individ. Et individ som bor i et bestemt geografisk område, vil ofte naturlig søke til næringslivet på vedkommende sted. Mer sjelden vil det søke til et annet område dersom avstandene er store og kommunikasjonsforholdene dårlige. Selv om grunnlønnen til et individ er bestemt ved forhandlinger mellom organisasjoner, vil det kunne oppstå tilbuds- eller etterspørselspress etter henholdsvis arbeidstakere eller arbeidsplasser innenfor disse områdene. Dette vil kunne medføre at lønnen til et individ i hvertfall kan øke utover den avtalte minstelønn. Videre kan nevnes at bedriftseierne kan synes at i visse deler av landet er det mer lønnsomt å investere selv om lønnsutgiftene blir større. Det kan på denne måte oppstå forskjeller i lønn i forskjellige deler av landet. Grunnen til at menneskene ikke søker arbeid der lønnen er høyest, kan skyldes flere faktorer, f.eks. at det liker seg godt i det arbeid det har, likeromgivelsene, naturen, klimaet, omgangskrets osv.

Etter dette skulle det være naturlig å følge Standarden for handelsfelt. Det viser seg imidlertid at etter denne oppdelingen vil mange kommuner i enkelte fylker bli overført til et annet handelsfelt enn de gjenværende. Dette vil medføre en del vansker når materialet skal bearbeides. Vi har derfor valgt å benytte den vanlige oppdelingen i landsdeler som er nevnt ovenfor, da denne inndelingen medfører at fylkene ikke blir delt (de to første siffer i herreds-, bykoden angir hvilket fylke, herredet eller byen tilhører). Denne oppdelingen vil ikke skille seg så mye fra handelsfeltinndelingen. To områder har vi skilt ut i tillegg til landsdelene. Det gjelder fylkene Oslo og Bergen. Disse er skilt ut, idet vi mener at de har en innflytelse på inntekten som skiller dem ut fra landsdelen de tilhører. For å få en enda bedre inndeling burde også omegnskommunene til disse byene vært tilført byene og ikke til resten av vedkommende landsdel. Det kunne også ha vært interessant å ha fått med en viss gruppering av byer etter en størrelsesgruppering, men dette har ikke blitt gjort i denne undersøkelsen.

2. Næring

Som vi har nevnt ovenfor, bestemmes lønningene for arbeidstakerne ved tarifforhandlinger. Avtalene er gjerne sluttet mellom arbeidere og arbeidsgivere etter regler som innebærer at endringene i lønningene kan bli forskjellige i de enkelte næringene. Dette vil da føre til at det kan oppstå ulikheter i inntekten mellom de enkelte næringene. Disse avtalene vil igjen

føre til at bedriftseiernes inntekter blir avhengige av forhandlingsresultatene. For noen næringer er det avtaler mellom staten og næringens organisasjon, som består av bedriftseierne (jordbruksavtale, fiskeriavtale).

Oppgavene over næringen eller bransjen som hovedinntektstakeren har oppgitt at han arbeider i, har blitt oppgitt meget spesifisert. Denne grupperingen har vi imidlertid slått sammen til et mindre antall grupper, idet vi ellers ville ha fått for mange variable slik at vi ikke kunne benytte regnemaskinens regresjonsprogram. Den grupperingen som vi har benyttet, har vi forsøkt å foreta slik at næringer som har noenlunde lik struktur har kommet i samme gruppe.

Da vi gjennomgikk tallmaterialet ovenfor, pekte vi på det forhold at det forekommer en rekke tilfelle der inntekten til et individ blir opptjent i flere næringer. For at vi skal få et visst innblikk i hvilken størrelsesorden dette kan dreie seg om, skal vi se på de næringene hvor det er vanlig at utøveren i større eller mindre grad har beskjeftigelse også i en annen næring.

	Eneyrke	Hovedyrke	Biyrke
Jordbruk (1959)	77 257	44 620	75 655
Fiskere (1960)	21 405	23 832	15 660

Hvordan forholdene er i andre yrker er det vanskelig å uttale seg om, men det er å anta at det er en del blanding. Vi vil imidlertid anta at dette forhold ikke vil gi seg så store utslag i resultatene at de vil endre våre konklusjoner.

3. Kjønn

Man har inntrykk av at kvinner er lavere betalt enn menn, men dette er det vanskelig å uttale seg om dersom vi ikke kan holde alle andre faktorer like. Det å holde alle andre faktorer like er ofte vanskelig dersom vi ikke får en meget sterk oppdeling av så godt som alle forklarende faktorer, samt tar hensyn til mange forhold som vi ikke vil få opplysninger om i den type tallmateriale som vi skal benytte. Ofte kan stillingene skilles i flere grupper så som kvinnearbeid, mannsarbeid og stillinger som kan være (eller ønskes) besatt av begge kjønn. Det skulle her være nok å nevne avisenes "stilling-ledig-annonser". Videre kan det være forskjell i muligheter for avansement for kvinner og menn. I enkelte yrker og hos enkelte arbeidsgivere (staten) er det nå likestilling mellom kjønnene, idet det er lov- eller tariff-festet lik lønn for likt arbeid.

I tallmaterialet er det oppgitt kjønnnet til vedkommende som har fylt ut selvangivelsen. I de fleste tilfellene vil det stemme overens med

hovedinntektstakerens kjønn, og det vil ofte være slik at denne er den eneste som har hatt inntekt. Et forstyrrende tilfelle er når ektefellene har hver sin inntekt og de er fellesliknet. Dette skal vi ta for oss i neste avsnitt.

4. Fellesliknet

Vi har nå kommet til en størrelse som vi ikke har behandlet ovenfor. Grunnen til at vi vil ta den med når vi skal foreta beregningene, er den forstyrrende effekt som den har ved, som vi har nevnt ovenfor, at en rekke ektepar er liknet sammen. Dette gjør at hovedpersonen vil få registrert en inntekt i tillegg til sin egen, og det er denne effekten som er søkt tatt vare på ved hjelp av denne variabelen.

Det er imidlertid slik at ektefeller blir betegnet som fellesliknet dersom de ikke har sendt egne selvangivelser. Dette medfører at det totale antall ektepar minus antall ektepar som er særskilt liknet er lik antall ektepar som er fellesliknet. Lønningene vi kommer fram til for fellesliknede ektepar, vil i gjennomsnitt bli mindre enn dem vi ville ha fått dersom bare ekteskap der begge ektefeller hadde inntekt, var med som fellesliknet.

5. Stilling

Som vi har omtalt ovenfor, vil individets stilling (gjøremål) i forhold til produksjonsprosessen ha betydning for den inntekt det vil få for sine ytelser. Dette er en faktor i tillegg til de faktorene som vi har nevnt tidligere, som ved lønnsforhandlingene er bestemmende for inntekten.

Fra tallmaterialet har vi 9 stillingsgrupper. Disse har vi slått sammen til 3. De gruppene som vi har omtalt ovenfor, er lønnstakere og selvstendig næringsdrivende. Disse utgjør to av gruppene. I tillegg har vi en rekke stillingsgrupper så som pensjonister, rentenister, studenter og skoleelever og andre. Disse kunne vi ha sortert ut, men da dette ville ha krevd et større regneprogram, skilte vi dem ut ved hjelp av dummy-variable. Vi dannet en gruppe: Pensjonister og rentenister. De resterende, studenter, skoleelever og andre, ble inkludert i stillingsgruppen selvstendig næringsdrivende.

Som vi har nevnt ovenfor, kan inntekten være tjent i forskjellige stillinger. Dette er det ikke mulig å ta hensyn til.

6. Alder

Alderen er den eneste av de uavhengige variable som er en kvantitativ størrelse. Holder vi alle andre størrelser konstante, så kan vi uttrykke den

naturlige logaritmen til lønnen L som en funksjon av alderen A :

$$(7) \quad \ln L = f(A)$$

Formen på f -funksjonen er ukjent, men vi har flere måter å gå fram når den skal bestemmes, og som vi skal se på nedenfor.

Alderen, A , er en kontinuerlig størrelse selv om vi bare observerer størrelser som er heltallige. Lønnen, L , kan vi anta er kontinuerlig (selv om det alltid vil være minimum 1 øre i forskjell på to lønninger). Avsetter vi observasjonspaar av A og L i et A, L -diagram for individer hvor de andre faktorene som påvirker inntekten er omlag de samme, vil forløpet kunne være som i fig. 2.



Fig. 2

Søker vi å tilpasse en kurve i denne punktmassen, vil den få et forløp som antydnet. Fra flere undersøkelser viser det seg at:

$$(8) \quad 0 < f' \quad \text{og} \quad f'' < 0$$

Grunnene til at f -funksjonen vil ha en form som i figur 2, kan være så mange, blant annet er det ikke mulig å få tilfredsstillt forutsetningen om "like vilkår ellers". Dette vil særlig gjelde ved så grov inndeling av de forklarende variable som vi har foretatt. Dersom vi skulle komme med en del grunner som vi kan tenke oss å ha betydning, kan vi kanskje med fordel dele dem i to grupper:

1. Økende inntekt med alder:
 - a) Deltidsarbeid i ung alder, fast i høyere alder
 - b) Etter at barna har blitt gamle nok, har hustruen gått ut i arbeidslivet, men hun blir liknet sammen med mannen.
 - c) Alderstillegg
 - d) Økt utdannelse
 - e) Avansement
 - f) Økt tempo ved erfaring

2. Avtakende inntekt med alder:

- a) Pensjonister fører opp sitt tidligere yrke
- b) De unge har bedre utdanning enn de eldre, slik at man får målt effekten under ulike utdannelsesforhold
- c) Sykdom og nedsatt arbeidsevne av andre årsaker

Benytter vi oss av Taylor's formel ⁷⁾ vil vi få et tilnærmet uttrykk for $f(A)$

$$(9) \quad f(A) = \alpha'_0 + \alpha'_1 A + \alpha'_2 A^2 + \dots + K$$

der α'_i $i = 0, 1, \dots$ er konstanter og K er restleddet. Dersom betingelsene i (8) er de som vi i første omgang ønsker oppfylt, vil vi kunne få en god tilnærming til kurven når vi benytter oss av to ledd i Taylor-utviklingen,

$$(10) \quad f(A) = \alpha'_0 + \alpha'_1 A + \alpha'_2 A^2 + K$$

En annen framgangsmåte er å gruppere alderen i intervaller, f.eks.

$$|\underline{0} - 20 > , |\underline{20} - 30 > , \dots, |\underline{70} \rightarrow >$$

Disse gruppene kan vi så enten innføre i analysen ved hjelp av dummy-variable ⁸⁾ eller ved hjelp av en nummerering av intervallene ⁹⁾. Fordelene ved å bruke dummy-variable er at vi ikke behøver å uttale oss på forhånd om f -funksjonens form.

Videre antar vi, når vi bruker en grov inndeling i aldersgrupper, at en liten endring av alderen ikke spiller noen stor rolle, slik at man får med det vesentlige ved å se på utviklingen over flere aldersår. Dette siste argumentet kan også benyttes når vi nummererer intervallene, men vi må da si noe om funksjonsformen.

Selv om disse sistnevnte metodene har sine fordeler, vil vi benytte relasjon (10) og A =alder målt i år. Grunnen er, som nevnt flere ganger ovenfor, at all forenkling vil spare oss for vansker ved bearbeidningen.

7. Sluttmerknader

Da vi behandlet modellen, nevnte vi en rekke andre faktorer som vi mente kunne ha innvirkning på inntekten. Vi har ikke noen mulighet for å få opplysninger om disse, og derfor er det selvsagt uråd å ta dem med i beregningene. Disse vil derfor i prinsippet være med blant de faktorer som påvirker restleddet i relasjonen.

7) Se [6] kap. 7

8) Se [3]

9) Se [1] og [2]

Vi skal her se litt på hvilken rolle arbeidstiden vil ha, selvom vi ikke har noen opplysninger om den. Rent intuitivt kan vi anta at inntekten til et individ vil være nærmest en lineær funksjon av tiden det er i arbeid. Vi har regnet med kontinuitetsfaktorer også for arbeidskraft, og da et individ ikke kan deles, vil det si at dersom det ikke er i arbeid full arbeidsdag, må vedkommende individ betraktes som om det arbeider deltid.

I undersøkelsen av F. G. Adams er det vist at arbeidstiden er av stor betydning, selv om han bare benyttet to grupper av arbeidstid, nemlig mindre enn 11 - 12 måneder og et helt år. T. P. Hill hadde bare med personer som arbeidet i heldagsstilling et helt år. Ivar Østby benyttet et materiale for fast ansatte heldagsbeskjeftigede. Vi skal senere se om dette forhold kan ha hatt noen innvirkning på våre resultater.

Slik relasjon (5) eller (6) står, med c eller $f(A)$ og Q_i som de faktorene som skal forklare endringene i $\ln L$, kan en stille spørsmålet om det kan tenkes at enkelte av disse faktorene blir bestemt av $\ln L$ og ikke omvendt. Dersom dette er tilfelle, vil vi ikke få den ønskede form på L -funksjonen.

Av faktorer som sikkert kan påvirke lønnen, er alder og kjønn. Størrelser som en kan være usikker på, er f.eks.: stilling, næring, utdanning og geografisk område. Dersom lønnen påvirker disse faktorene kan dette skyldes henholdsvis etterspørsels- og/eller tilbudssiden. Både på etterspørsels- og tilbudssiden kan vi finne eksempler der det synes som om disse faktorene bestemmes av inntekten. Bedriftene kan ha et gitt beløp til disposisjon som lønn til en eller flere stillinger som de ønsker besatt, og de søker så å tilpasse de ovenfor nevnte faktorene slik at de får størst mulig lønnsomhet. Også på tilbudssiden kan vi finne eksempler der det ser ut som inntekten er den forklarende faktor. Som eksempel kan vi nevne ønsket om en høy inntekt kan forårsake flytting til bestemte områder av landet. Et annet er at ønsket om en høy inntekt kan medføre at man får seg en bestemt utdanning. Det som kan gi en bedre forklaring på disse nevnte forhold, er faktorer som evner, interesser og miljø for arbeidstakeren, samt liknende faktorer for eventuelle barn og ektefelle. En bedrift og et individ kan ha visse ønsker om å la inntekten bestemme enkelte faktorer, men dersom ikke evner, interesser o.l. er til stede vil de ikke få gjennomført sine ønsker. Vi vil derfor anta at inntekten blir bestemt av de ovenfor nevnte faktorer sammen med stilling, næring osv.

VII. Variabelspesifikasjon

Av det som vi har gjennomgått ovenfor, vil modellen vår få følgende form:

$$I \quad \ln L = \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \sum_{i=1}^{16} \gamma_i Q_i + u$$

De variable Q_i ; $i = 1, \dots, 16$ er innført som dummy-variable (binære hjelpevariable) som antar verdien 1 når observasjonsenheten (individet) har det nedenfor oppførte kvalitative kjennetegn, og antar verdien 0 når observasjonsenheten ikke har dette kjennetegn.

Vi har følgende variable:

Uavhengige variable:

A	Alder, målt i år	
A^2	Alder kvadrert	
Q_K	Mann	Referansegruppe
Q_1	Kvinne	
Q_S	Lønnstakere	Referansegruppe
Q_2	Selvstendig næringsdrivende	
Q_3	Trygdede og pensjonister	
Q_N	Bergverk og industri	Referansegruppe
Q_4	Jordbruk	
Q_5	Skogbruk	
Q_6	Fiske	
Q_7	Bygg og anlegg	
Q_8	Varehandel	
Q_9	Andre næringer	
Q_G	Østlandet (-Oslo)	Referansegruppe
Q_{10}	Oslo	
Q_{11}	Sørlandet	
Q_{12}	Vestlandet (-Bergen)	
Q_{13}	Bergen	
Q_{14}	Trøndelag	
Q_{15}	Nord-Norge	
Q_F	Ikke fellesliknet	Referansegruppe
Q_{16}	Fellesliknet	

Avhengig variabel:

$\ln L$ Naturlig logaritme til inntekt før skatt

Skokastisk restledd:

u

Konstanter:

$\alpha_0, \alpha_1, \alpha_2$ og γ_i ; $i = 1, \dots, 16$

Når variable Q_i ; $i = 1, \dots, 16$ er innført som dummy-variable, vil γ_i ; $i = 1, \dots, 16$ representere tillegg eller fradrag ut fra de tilhørende referansegruppene som brukes som utgangspunkt.

VIII. Estimeringsmetode og statistiske tester

Vi har forutsatt at restleddet u er stokastisk og identisk fordelt med forventning 0 og varians σ^2 . Under denne forutsetning vil minste kvadraters metode gi effisiente estimater. Testene som blir beregnet, vil gjelde tilnærmet dersom vi ikke har normalfordeling da observasjonsmaterialet er på hele 9 035 observasjoner (jfr. sentralgrenseteoremet).

For å teste hypotesen

$$H_0 : \gamma_i = 0$$

mot

$$H_1 : \gamma_i \neq 0$$

må vi benytte testobservatoren

$$T = \frac{\hat{\gamma}_i}{\hat{\sigma}_{\hat{\gamma}_i}}$$

som er t-fordelt med $n - k$ frihetsgrader, der k er antall parametre i modellen inklusive konstantleddet. Tegnet $\hat{}$ over en koeffisient betyr her og nedenfor at denne størrelsen er estimert. Vi får å forkaste H_0 dersom $|T| > t_{\epsilon, n-k}$ der $t_{\epsilon, n-k}$ er ϵ -fraktilen i t-fordelingen med $n - k$ frihetsgrader. Har vi $\epsilon = 0,05$, vil $t_{0,05, 9016} = 1,96$. Beregninger for hver variabel viser at i bare fire tilfelle vil vi ikke få forkastning av H_0 . Dette vil gjelde koeffisientene for de variable:

Q_5 Skogbruk

Q_7 Bygg og anlegg

Q_{11} Sørlandet

Q_{13} Bergen

Ved hjelp av denne t-testen kan vi altså teste om de enkelte parametrene er lik null. Det kan i mange tilfelle også være av interesse å teste om s e t t av forklaringsvariable har betydning for inntektens størrelse¹⁰⁾. I denne undersøkelsen kan dette være av interesse for de kvalitative størrelsene som er innført ved hjelp av dummy-variable. Framgangsmåten er at man tar utgangspunkt i modell I som man estimerer koeffisientene for, samt at man beregner kvadratet av den multiple korrelasjonskoeffisienten, R^2 . Deretter utelates ett sett av forklaringsvariable og denne nye modellen får estimert koeffisientene, og får beregnet den multiple korrelasjonskoeffisient R_i^2 . Slik fortsetter man til en har utelatt alle unntatt et visst minimalt sett av forklaringsvariable. Testen består i å påstå signifikantsnår:

$$F_{n,n-k} = \frac{R^2 - R_i^2}{1 - R^2} \cdot \frac{n - k}{h} \geq f_{h,n-k} (1 - \epsilon)$$

der ϵ er testnivået, n er sampelstørrelsen, k er antall parametre i den første modellen inklusiv konstantledd og h er antall parametre som er utelatt i den reduserte modellen. $F_{h,n-k}$ er da den beregnede verdi som er Fisherfordelt med h og $n-k$ frihetsgrader. $f_{h,n-k}(1 - \epsilon)$ er $(1 - \epsilon)$ prosentfraktilen i F-fordelingen med h og $n - k$ frihetsgrader.

De modellene som vi estimerer koeffisientene i, blir:

$$\begin{aligned} \text{I} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \sum_{i=1}^{16} \gamma_i Q_i + u_1 \\ \text{II} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \sum_{i=1}^{15} \gamma_i Q_i + u_2 \\ \text{III} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \sum_{i=1}^9 \gamma_i Q_i + u_3 \\ \text{IV} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \sum_{i=1}^3 \gamma_i Q_i + u_4 \\ \text{V} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + \gamma_1 Q_1 + u_5 \\ \text{VI} \quad \ln L &= \alpha_0 + \alpha_1 A + \alpha_2 A^2 + u_6 \end{aligned}$$

Resultatene for disse modellene er gitt i tabelloppstillingen side 23.

Tabellen gir de estimerte regresjonskoeffisientene og standardavvikene for disse estimatene. Videre finner vi de multiple korrelasjonskoeffisientene R^2 og den estimerte residualvariansen $\hat{\sigma}_u^2$ for hvert modellalternativ.

10) Se [7]

Nederst i tabellen er oppført en rekke opplysninger i tilknytning til F-testen.

$n - k$ = antall estimeringsmessige frihetsgrader ($n = 9035$ og k er tallet på koeffisientene i vedkommende modell medregnet konstantleddet).

h = antall koeffisienter som modellen har i tillegg til modellen i nærmeste kolonne til høyre.

$F_{h,n-k}$ = F-verdi (h og $n - k$ frihetsgrader).

$f_{h,n-k}(0,95)$

= 95 prosent-fraktilen i F-fordelingen med h og $n - k$ frihetsgrader.

Signifikans i forhold til modell til høyre : Markert med JA når $F_{h,n-k} \geq f_{h,n-k}(0,95)$. Ellers NEI.

IX. Enkelte resultater og tolkninger

Resultatene av regresjonskjøringene og de testene som vi har benyttet, finner vi side 23. Vi finner at alle de gruppene med forklaringsvariable som vi har med i beregningene er utsagnskraftige.

Går vi over til å se på den oppdelingen som har blitt foretatt av hvert sett av forklaringsvariable, viser det seg at de aller fleste av disse undergruppene er signifikant forskjellig fra den gruppen som er valgt ut som referansegruppe for vedkommende variabel. Dette kan vi sammen med de andre resultatene tolke slik at vår oppdeling har skilt ut en rekke undergrupper som kan gi oss en del forklaring av den variasjonen i inntektens størrelse som er til stede. Vi skal komme nærmere tilbake til dette nedenfor.

På side 10 kom vi med noen a priori antakelser om hvordan alderen virker inn på inntekten, og vi antok at $0 \leq f'(A)$ og $f''(A) < 0$ for forskjellige verdier av A . Av resultatene ser vi at denne hypotese blir bekreftet. Det kan nå være av interesse å undersøke hvilke verdier av A som gir maksimum inntekt når vi hele tiden holder oss til de enkelte modellens referansegrupper.

Maksimum finner vi når vi deriverer modellene partielt med hensyn på A , og førsteordensbetingelsen for maksimum er oppfylt når de er lik null:

$$(11) \quad \frac{\partial \ln L}{\partial A} = \hat{\alpha}_1 + 2\hat{\alpha}_2 A = 0$$

som gir

$$A = - \frac{\hat{\alpha}_1}{2\hat{\alpha}_2}$$

Annenordensbetingelsen er at

$$\frac{\partial^2 \ln L}{\partial A^2} = 2\hat{\alpha}_2 < 0$$

Av resultatene ser vi at denne er oppfylt.

Den naturlige logaritme til inntekten et individ i referansegruppen kan forvente å få, blir når vi befinner oss i det punktet der alderen gir maksimum inntekt:

$$\ln L^* = \hat{\alpha}_0 - \frac{\hat{\alpha}_1^2}{4\hat{\alpha}_2}$$

Resultatene av disse beregningene for de forskjellige modellene er oppført i tabell 1.

Tabell 1.

Modell	Alder som gir maksimum inntekt, $A = -\frac{\hat{\alpha}_1}{2\hat{\alpha}_2}$	Estimert inntekt, gitt at $A = -\frac{\hat{\alpha}_1}{2\hat{\alpha}_2}$
VI	49,6	9 363
V	51,2	11 660
IV	57,5	18 910
III	57,4	19 980
II	56,8	21 780
I	57,7	15 870

Ved overgang fra modell VI til V øker alderen som gir maksimum inntekt litt. Når vi så kommer over til de påfølgende modellene ligger alderen ca. 8 år over den for modell VI. Det kan se ut som om vi får tre nivåer for alderen og hvor det særlig er et stort og markert sprang fra de to første til de fire siste.

Endringen når vi går fra modell VI til V er at vi får kjønn med som forklaringsvariabel. Grunnen til spranget i alderen er nok at kvinnene som er ute i arbeidslivet, er av de yngre årsklassene. Det andre nivået i alderen som viser seg for modell V, antas å skyldes studenter og skoleelever som med sin lave alder påvirker alderen som gir maksimum inntekt. I de følgende modellene er de skilt ut fra referansegruppen som omfatter menn som er lønns-
takere, slik at det er ikke så forbausende at alderen som gir maksimum inntekt er nærmest konstant og betydelig høyere enn for de to første modellene.

Maksimumsinntekten som man oppnår i 50-årsalderen, blir kr. 9 363 når vi betrakter modell VI. Dette beløpet viser seg å ligge betydelig under det aritmetriske gjennomsnitt av inntekten som er kr. 12 111. Dette er et resultat som er uventet, da en ville ha ventet at maksimumsinntekten beregnet ved hjelp av modellen skulle bli større enn $\frac{1}{n} \sum_i L_i$.

Hva som kan være årsaken til dette uventede resultatet av våre beregninger, skal vi forsøke å klarlegge da det kan være forårsaket av flere faktorer.

Vår modell benytter som avhengig variabel $\ln L$. Observasjoner av denne inngår både i konstantleddet og i regresjonskoeffisienter blant annet ved sin gjennomsnittsverdi, som er den naturlige logaritmen til det geometriske gjennomsnittet av L . For de L -verdiene som vi har, vil vi ha:

$$\ln \left(\frac{1}{n} \sum_i L_i \right) > \frac{1}{n} \sum_i \ln L_i = \ln \sqrt[n]{\prod_i L_i}$$

Hadde vi benyttet en additiv modell der L hadde vært den avhengige variabelen, ville maksimumsinntekten, L^{XX} , man oppnådde ved den alderen som gav størst inntekt ved tilsvarende modell som VI bli større enn i den multiplikative modell

$$L^{XX} > L^X$$

på grunn av de faktorer vi har omtalt ovenfor. Vi ville nå få at

$$L^{XX} > \frac{1}{n} \sum_i L_i$$

L kan anta verdier opp mot 80 000 og dette vil føre til at ulikheten

$$\ln \left(\frac{1}{n} \sum_i L_i \right) > \frac{1}{n} \sum_i \ln L_i$$

kan bli betydelig og dermed kan L^X bli meget mindre enn L^{XX} , slik at vi kan få et resultat slik som her at

$$\frac{1}{n} \sum_i L_i > L^X$$

Et annet forhold som vi bør nevne i denne sammenhengen, er spredningen på de estimerte koeffisientene og restleddet. Særlig er spredningen på restleddet av betydelig størrelsesorden.

Funksjonsformen av $f(A)$ som vi har benyttet, kan også få virkning på resultatet. Den vil med den store variasjonen i materialet som vi har, vanskelig kunne føye seg godt i materialet. Dette viser seg også å holde

stikk, da $R^2 = 0,024$ for modell VI. Det kan derfor tenkes at vi ville fått bedre tilpassing ved å benytte flere ledd i Taylor's formel ved rekkeutvikling av $f(A)$.

Tilveksten eller reduksjonen av inntekten (ΔL) på de forskjellige alderstrinn i prosent av inntekten på vedkommende trinn, blir for grunnmodellen I som vist i tabell 2 (beregnet på grunnlag av (11)).

Tabell 2.

Alder	20	30	40	50	60	70
ΔL i prosent	4,6	3,4	2,2	0,9	-0,3	-1,5

For å få et bedre bilde av formen på funksjonen

$$(10) \quad y=f(A)\ln L = \alpha_0 + \alpha_1 A + \alpha_2 A^2 + u_6$$

kan vi beregne krumningen¹¹⁾:

$$T = \frac{y''}{(1 + y'^2)^{3/2}}$$

Det viser seg at den er om lag konstant, $T \approx 0,00122$ i det aktuelle området for alderen. Formen på funksjonen i et $(\ln L, A)$ diagram kan betraktes som en del av en sirkel. Tallverdien av de prosentvise endringene og inntektens størrelse vil altså være like store når vi befinner oss like langt borte fra den alderen som gir maksimum inntekt. Formen på $f(A)$ i (10) kan være årsak til at T er tilnærmet konstant. Det kunne derfor ha vært interessant å ha undersøkt en annen form på $f(A)$.

Som vi har omtalt ovenfor, er sysselsatte kvinner gjennomgående betydelig yngre enn menn. Videre er lønnen til en kvinne under halvparten av det en mann får. Dette kan se mye ut dersom vi har oppfylt betingelsen om at de to kjønnene blir betraktet under like forhold. Det kan være berettiget tvil om dette, da vi mangler - som vi tidligere har omtalt - opplysninger om tiden vedkommende har vært sysselsatt. For kvinner vil det ofte være tilfelle at de har deltidsarbeid. Enda et forhold som kan komme til, er at stillings- og næringsgrupperingene er for grove. Som en sammenlikning kan nevnes at mennenes inntekt lå 20 - 25 prosent over kvinnenes i den ovenfor nevnte undersøkelse av Ivar Østby. Denne undersøkelsen omfattet ikke deltidsansatte og rengjøringshjelp. I de andre undersøkelsene som vi har vist til, omfattet materialet bare menn, slik at vi ikke kan få noen sammenlikning med våre resultater.

11) Se [6] side 226.

Går vi over til å se på stillingsgrupperingen, samtidig som vi betrakter referansegruppene for de andre forklaringsvariablene, viser det seg at en person i gruppen selvstendig næringsdrivende tjener om lag 1/6 av det en lønnstaker får. Dette synes å være en meget stor forskjell når vi ser på samme geografiske område. Vi må imidlertid ta hensyn til den grupperingen som vi har foretatt for denne stillingen. Den omfatter skoleungdom og studenter. Dette vil medføre at denne store gruppen som arbeider bare en del av året, vil senke inntekten i denne gruppen betydelig. En antakeligvis bedre inndeling ville være å dele den i to: Selvstendig næringsdrivende og andre. Gruppen "Andre" vil da omfatte studenter, skoleungdom, rentenister og andre stillingsgrupper.

Gruppen trygdede har ca. 20 - 30 prosent lavere inntekt enn referansegruppen. Det er vanskelig å sammenlikne to så ulike grupper da aldersfordelingen er helt forskjellig. Forsøker vi å ta den med i beregningene vil de trygdedes inntekter bli betraktelig lavere. For en 80-årig mann vil inntekten bli om lag halvparten av hva han tjente som maksimum, dersom han tilhørte referansegruppen.

Næringsgrupperingen inneholder to grupper: Skogbruk og bygg og anlegg, som har koeffisienter som ikke er signifikant forskjellige fra null. Det kan allikevel være en fordel å skille ut skogbruk, da den ellers vil bli slått sammen med jordbruk. Bygg og anlegg kunne vi derimot ha slått sammen med referansegruppen, bergverk og industri. Gruppene jordbruk og fiske skiller seg ut med sine meget store tillegg i inntekten. Her må vi imidlertid huske at de alt vesentlige som er beskjeftiget i disse næringene, er selvstendig næringsdrivende, og de vil derfor netto få en lavere inntekt enn referansegruppen.

I den geografiske inndelingen er det to grupper, der koeffisientene ikke er signifikant forskjellige fra null. Dette gjelder Sørlandet og Bergen. Det vil også her gjøre seg gjeldende at selv om de ikke skiller seg ut fra referansegruppen, Østlandet (minus Oslo), så vil de skille seg ut fra sine omgivelser. Det kunne i denne forbindelse ha vært foretatt en enda sterkere oppdeling, f.eks. ved utskillelse av de største byene i landet fra de omliggende landdistrikter eller av et par grupper av byer og kommuner etter folketall kombinert med landsdelsinndeling¹²⁾.

Når man er fellesliknet, vil inntekten bli 30 - 40 prosent høyere enn ellers. Tilleggsinntekten er altså av relativ beskjedent størrelse. Dette kan komme av at det er kvinner på deltid som får denne ekstraintekten.

12) Se [4] side 4.

En variasjon av samtlige variable skal vi ikke drøfte virkningene av her. Ved å benytte resultatene kan vi beregne inntekt for en rekke forskjellige konstellasjoner av de forklaringsvariable. Dette kan så danne grunnlaget for oppstilling av tabeller. Som vi nevnte i innledningen, kan disse resultatene benyttes ved utformingen av tabellene. Alle de forklaringsvariable som er med, har signifikant virkning på inntektens størrelse, og det kan derfor være av interesse å lage tabeller der de blir benyttet til å forklare inntektsvariasjoner. Det vil være av liten interesse å produsere tabeller der en benytter forklaringsvariable som har liten eller ingen forklaringskraft.

----- o 0 o -----

Det som på sett og vis kan virke forbausende er at etter at vi har tatt hensyn til en rekke faktorer som vi har ansett for å ha betydning for forklaringen av inntektens størrelse, så er det enda en betydelig restvariasjon tilbake som vi ikke har fått forklart. Vi har flere ganger omtalt en rekke faktorer som burde ha vært tatt med, så som arbeidstid og utdanning. Dette er i hvert fall opplysninger som det kan være mulig å få tak i. Andre egenskaper ved individene, så som evner, interesse o.l., vil det være så godt som umulig å få oppgave over.

Sammenlikner vi resultatene fra vår undersøkelse med de andre som vi har omtalt tidligere, ser vi at i hovedtrekkene stemmer de overens, men den omtalte restvariasjonen er betydelig større i vår undersøkelse enn i de andre. Dette kan komme av materialet som vi har benyttet, da det er meget uensartet. Dette vil derfor medføre at de forhold som vi ikke har opplysninger om, vil bli inkludert i restleddet. Er de utelatte faktorene av stor betydning, vil restleddet også bli stort. Et mer spesialutvalgt materiale, enten ved at man får innhentet opplysning om flere faktorer eller ved å velge ut en betydelig mer homogen masse, vil derfor være ønskelig når inntektens størrelse skal forklares, men på den annen side viser resultatene av denne undersøkelsen hvor forskjellig vi mennesker er. Med et spesialutvalg vil vi derfor komme til å betrakte et samfunn som tilfredsstillende vår modell, men som det kanskje er av liten praktisk interesse å betrakte.

Litteraturliste

- [1] F. G. Adams, "The Size of Individual Incomes: Socio-Economic Variables and Chance Variation". The Review of Economics and Statistics, Vol. XL (1958). Sidene 390 - 398
- [2] T. P. Hill, "An Analysis of the Distribution of Wages and Salaries in Great Britain". Econometrica, Vol. 27 (1959). Sidene 355 - 381
- [3] Lawrence R. Klein, "An Introduction to Econometrics". Sidene 171 - 178
- [4] Ivar Østby, "Lønnsstrukturundersøkelser i varehandelen", IB 65/1 Arbeidsnotater, Statistisk Sentralbyrå (Ikke offentliggjort)
- [5] J. Johnston, "Econometric Methods"
- [6] R. Tambs Lyche, "Matematisk analyse", bind I
- [7] H. W. Watts, "Statistisk testing av hypoteser ved regresjonsberegninger". Statistisk Sentralbyrås håndbøker nr. 22
- [8] H. T. Amundsen, "Innføring i teoretisk statistikk", bind III
- [9] Petter Myklebust, "Hovedresultater fra inntektsstatistikken 1958 og 1962" IO 66/2. Statistisk Sentralbyrå. (Ikke offentliggjort)

Størrelser som karakteriserer regresjonsmodellene I - VI¹⁾

Uavhengig-variabel; regresjons-koeffisient	Modell	I		II		III		IV		V		VI	
		Koeffisient	Standardavvik	Koeffisient	Standardavvik	Koeffisient	Standardavvik	Koeffisient	Standardavvik	Koeffisient	Standardavvik	Koeffisient	Standardavvik
Konstantledd	α_0	7,64068		7,56732		7,36893		7,36443		7,37478		6,81128	
Alder	α_1	0,07040	.0066	0,08523	.0063	0,08834	.0064	0,08631	.0064	0,07776	.0062	0,09416	.0063
Alderkvad.	α_2	-0,00061	.0001	-0,00075	.0001	-0,00077	.0001	-0,00075	.0001	-0,00076	.0001	-0,00095	.0001
Kvinne	γ_1	-0,7817	.0523	-0,9481	.0479	-0,8971	.0479	-0,9864	.0468	-0,8791	.0481		
Selvst.	γ_2	-1,9646	.0692	-1,9443	.0694	-1,9712	.0698	-1,5442	.0580				
Trygdet	γ_3	-0,2221	.0875	-0,2463	.0877	-0,3080	.0881	-0,5766	.0848				
Jordbruk	γ_4	0,9011	.1045	0,8836	.1049	0,7803	.1050						
Fiske	γ_5	1,3207	.1378	1,2662	.1380	0,9719	.1352						
Skogbruk	γ_6	0,0630	.1580	0,0108	.1584	0,0345	.1583						
Bygg og anlegg	γ_7	0,1180	.0901	0,1143	.0904	0,0544	.0908						
Varehandel	γ_8	0,3367	.0777	0,3424	.0780	0,3604	.0780						
Andre	γ_9	-0,3507	.0585	-0,3726	.0587	-0,4009	.0587						
Oslo	γ_{10}	0,2437	.0645	0,2054	.0646								
Sørlandet	γ_{11}	-0,0809	.0879	-0,0905	.0882								
Vestlandet	γ_{12}	-0,1948	.0579	-0,2028	.0581								
Bergen	γ_{13}	-0,2080	.1194	-0,1758	.1197								
Trøndelag	γ_{14}	-0,3679	.0849	-0,3754	.0852								
Nord-Norge	γ_{15}	-0,6474	.0721	-0,6478	.0723								
Felleslikn.	γ_{16}	0,3845	.0495										
Multipel korrelasjonskoeffisient	R^2	0,17003		0,16447		0,15204		0,12764		0,05917		0,02435	
Estimert residualvar	σ_u^2	3,83592		3,86121		3,91605		4,02603		4,34107		4,50125	
n - k		9016		9017		9023		9029		9031		9032	
h		1		6		6		2		1			
$F_{h,n-k}$		60,40		22,36		43,27		35,43		334,20			
$f_{h,n-k} (0,95)$		3,84		2,10		2,10		3,00		3,84			
Signifikans i forhold til modell til høyre		JA		JA		JA		JA		JA			

1) Koeffisientene til α_0 , α_1 og α_2 er gitt med alle beregnede desimaler.