Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 64/1 Oslo, 12. mars 1964

Opplegg til en konsumfordelingsmodell for Norge

Av Ivar Østby

1. Bakgrunn for undersøkelsene

Et resultat av forbruksundersøkelsen 1958 var bl.a. publikasjonen NOS A 41 "Forbruksundersøkelsen 1958. 3. hefte - regresjonsberegninger".

Et av hovedformålene med forbruksundersøkelsen er å klarlegge hvorledes forbrukssammensetningen endres med familiestørrelse, familieinntekt, alder etc. Den siterte publikasjonen benytter en analysemetode som avviker vesentlig fra den tradisjonelle grupperingsmetode, idet en har gjort bruk av regresjonsmetode for å vise de nevnte sammenhenger. Det er forutsatt at den typiske utgiftsvariasjonen beskrives ved relasjonen:

$$(1.1) \quad x = \alpha_o + \alpha_1 y + \alpha_2 z + u$$

hvor x = utgift pr. husholdning til en bestemt vare
y = familiestørrelsen målt ved antall forbruksenheter
z = familieinntekten, evt. utgiften
u = stokastisk restledd

Ved hjelp av de innsamlede data, ble koeffisientene i (1.1) anslått for de ulike forbruksposter på Byråets elektroniske regnemaskin. Materialet ble splittet opp i en rekke ulike sosialgrupper.

Det siterte hefte 3 gir en oversikt over disse resultatene. Bl.a. er det tabulert elastisiteter m.h.p. inntekt for de ulike grupper, utgiftsendringer som følge av endring i familiestørrelse etc. for de ulike forbruksposter og for de ulike sosialgrupper.

Disse størrelsene som her er angitt, refererer seg til gjennomsnittspunktene i materialet. Dvs. når utgiftselastisiteten til matvarer er oppgitt

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.

til 0,37 for høyere funksjonærer, så kan en si at denne elastisitet refererer seg til den "gjennomsnittlige familie".

Det er m.a.o. ikke gjort noe forsøk på å vise om - og i tilfelle hvorledes elastisitetene varierer med familiestørrelse, familieinntekt m.v.

Rent generelt er det klart at det vil være av interesse å klarlegge hvorledes utgiftselastisitetene varierer med disse størrelsene. A priori er det god grunn til å anta at elastisitetene neppe er konstante uansett forbruksutgift og familiestørrelse.

I Artikler nr. 6 er det forutsatt at konsumelastisitetene varierer med inntekt og familiestørrelse. I dette hefte er det gitt eksempler på hvorledes denne variasjonen kan tenkes å være, idet en bygger på en spesifisert forbruksrelasjon (se Artikler nr. 6, p. 15).

Dette sistnevnte opplegget vil være av betydelig interesse når en skal belyse de karakteristiske trekk ved konsumets fordeling etter inntekt og familiestørrelse for ulike sosialgrupper. Den videre bearbeiding av forbruksundersøkelsen 1958, som det vil bli gjort rede for her, bygger på dette opplegget.

Resultatene av disse undersøkelsene vil kunne gi oss en konsumfordelingsmodell som kan være et viktig hjelpemiddel til å vurdere fordelings-politiske tiltak. (Eksempelvis: endringer i skattesatser, endring i alderstrygd, stønad til barnerike familier m.v.).

Ved et slikt opplegg kan en forutsetning om konstante elastisiteter, uavhengig av familieinntekt og familiestørrelse, neppe brukes.

Nå kan det bemerkes at de første regresjonsberegningene som ble foretatt på forbruksundersøkelsen 1958, strengt tatt bygger på at elastisitetene varierer med disse størrelsene. Men da relasjonene er rent lineære, blir dette variasjonsområdet svært snevert.

En annen ting er også at de tidligere beregninger neppe er helt gode for de ekstreme inntektsgrupper og familiestørrelsegrupper, og derfor er det neppe relevant å bygge en konsumfordelingsmodell på et slikt "stivt" opplegg, idet vi kan være spesielt interessert i hvorledes de ekstreme inntekts- og størrelsesgrupper (som f.eks. alderstrygdede) reagerer overfor endringer i de forklaringsvariable.

Vi bygger derfor beregningene på samme regresjonsmodell som i Artikler nr. 6, dvs. at logaritmen til utgiftspostene er et annengrads-polynom i logaritmen til totalutgiften og i familiestørrelsen.

I det følgende skal vi redegjøre for de beregninger og forutsetninger som er gjort. Videre skal vi se på de enkelte problemstillinger som fordelingsmodellen kan hjelpe oss å løse. Til slutt vil vi gå nærmere inn på de rent statistiske forutsetninger, og dessuten presentere endel numeriske resultater i tabellform.

2. Gangen i arbeidet med en konsumfordelingsmodell

2.a. Materiale og forutsetninger

Foruten problemet med selve formen på regresjonslikningen, møter en for det første problemet: "Hvilke variable kan sies å forklare konsumet".

Inntekt, priser, familiestørrelse, alder m.v. er her vesentlige faktorer. I dette spesielle tilfelle råder vi over et tverrsnittsmateriale, dvs. et materiale som er samlet inn over en begrenset tidsperiode, slik at prisfaktoren godt kan utelates i en regresjonsmodell som omfatter dette materialet under estimeringen av strukturkonstantene. De vesentligste forklaringsvariable blir da familieutgiften eller familieinntekten, familiestørrelsen og alderssammensetningen.

Vi vil her bruke familieutgiften som forklaringsvariabel, slik at det som beregnes ikke er inntektselastisiteter, men totalutgiftselastisiteter. Forskjell mellom utgift og inntekt er familiens sparing.

Familiestørrelsen regnes i antall forbruksenheter, omregnet ved hjelp av den vanlig brukte skala. Her tilsvarer f.eks. en voksen mann 1,00 forbruksenheter, et barn 0-1 år 0,20 forbruksenheter osv.[1)]

Familieoverhodets alder kan også trekkes inn i beregningene. Her er gjort den forutsetning at totalutgiftene til de ulike forbruksgoder avhenger av alderen, men at relative tilvekstgrader (elastisiteter) ikke avhenger av alderen. Dette innebærer at den prosentvise øking i f.eks. matvareforbruket når inntekten øker er uavhengig av alderen, derimot er totalnivået av matvareforbruket avhengig av alderen.

Et annet problem som må løses ved spesifiseringen av konsumrelasjonen, er hvorledes sesongvariasjonen i materialet skal elimineres. Forbruket av de ulike varegrupper viser ofte en viss sesongbevegelse. F.eks. kan vi nevne at ølforbruket er på topp om sommeren, skiutstyr og vinterklær selges mest om vinteren etc. Dette kan delvis løses slik at familiene bare fører regnskap for 1 måned hver, slik som det er gjort i forbruksundersøkelsen 1958. Likevel bør kanskje materialet "renses" for den resterende sesongeffekt i konsumet.

Problemet er her løst på den måten at vi har innført en rekke "dummy" variable i modellen. En "dummy" variabel antar verdien 1 i den måned observasjonen er foretatt og verdien 0 ellers. Innføringen av disse variable gjør oss i stand til å anslå sesongeffekten for de ulike vareposter.

Som det går fram av det foregående, kan vi formulere en generell modell for konsumatferden slik:

---

1) Se NOS A 41, p. 12.

$$(2.1) \quad \log x_i = g_i\,(c, f, Q_1 \ldots Q_m, Z_1 \ldots Z_k, u)$$

x_i = utgift til gode nr. i pr. familie
c = totalutgift $\left[c = \sum_i x_i\right]$
f = familiestørrelse
Q_j = sesongvariabel
Z_k = alder
u = stokastisk restledd

(Her er $m = 11$ og $k = 2$.)

Den eksplisitte formen på (2.1) skal presenteres senere i dette notatet. I (2.1) inngår det en rekke parametre som karakteriserer funksjonsformen $g_i(\ldots\ldots)$. Disse skal anslås på grunnlag av observasjonene; når det er gjort, kan vi finne anslag for hvorledes f.eks. utgiftselastisitetene varierer med de forklaringsvariable. Utgiftselastisitetene er definert som:

$$(2.2) \quad \frac{\partial \log x_i}{\partial \log c} = E_i = h_i\,(c, f \ldots\ldots)$$

2.b. Grupperingsproblemet

Etter å ha fastlagt forbruksrelasjonene, og dermed de forklarende variable i konsumet, kommer spørsmålet om grupperingen.

Hvor mange spesifiserte varegrupper skal vi regne med inngår i et husholds budsjett? Det kan være fristende å ta med så mange vareposter "som mulig". Det ville være meget interessant om vi kunne si med hvor mange prosent utgiftene til f.eks. sigarettmerket "South State" avtok (tiltok) om familiestørrelsen økte med en forbruksenhet. Imidlertid er det en stor svakhet ved å la varegrupperingen bli altfor spesifisert. Månedsregnskapene gir et nokså utilfredsstillende grunnlag for registrering av forbruket av en rekke enkeltposter, og derfor gir ikke materialet utsagnskraftige resultater for utgiftspostenes variasjon med inntekt og familiestørrelse. En har også ofte svært få observasjoner i enkelte grupper, og det gir utslag i at anslagene som foretas ved regresjonsberegningene blir "dårlige" i statistisk forstand.

Det er foretatt regresjonsberegninger etter dette opplegget for en 14-gruppering av utgiftspostene. Selv denne aggregeringen er kanskje noe for vid for vårt formål. Vi vil derfor samle utgiftspostene i 4 grupper, nemlig: (1) matvarer, drikkevarer og tobakk, (2) bolig, lys og brensel, (3) klær og skotøy og (4) annet konsum.

Denne aggregeringen er mer enn tilstrekkelig til å komme med relativt "solide" konklusjoner ifølge de foreliggende regresjonsberegningene. (Det går bl.a. fram av tabell 5.1 bak i dette notatet, jfr. spredningsmålene for regresjonskoeffisientene.)

En slik 4-gruppering av utgiftspostene har også en viss interesse, idet en lett kan jamføre disse resultatene med tidligere undersøkelser som f.eks. undersøkelsen i Artikler nr. 7 (pp. 23, 36 m.v.).

Skal en også trekke generelle konklusjoner om virkningene av den økonomiske politikk, er en heller ikke så svært interessert i å betrakte <u>enkelt</u>postenes endringer, som endringene i større totalgrupper.

2.c. Nærmere om prisenes rolle i modellen

Som nevnt bygger beregningene på et tverrsnittsmateriale, og prisene trekkes derfor ikke eksplisitt inn under estimeringen av modellens konstanter.

Det er klart at for å få en mer relevant konsumfordelingsmodell, må prisene "innarbeides" i modellen. Selv om vi, som sagt, ikke kan estimere priselastisiteter på grunnlag av <u>vårt</u> materiale, kan vi med visse forbehold bruke tidligere beregninger, f.eks. beregningene i Artikler nr. 7. En meget enkel måte å føye inn prisene er følgende:

$$(2.3) \qquad \log x_i = g_i(c, f....) + \sum_{j=1}^{n} e_{ij} \log p_j \qquad (i = 1, 2...n)$$

Elastisiteten av gode i m.h.p. prisen på gode j, blir derfor

$$(2.4) \qquad \underline{\frac{\partial \log x_i}{\partial \log p_j} = e_{ij}}$$

Relasjonen (2.4) forteller med hvor mange prosent utgiften til gode i øker (avtar) når prisen på gode j øker med en prosent.

Størrelsene e_{ij} er anslått tidligere ved hjelp av tidsrekkedata i Artikler nr. 7.

Når vi på denne måten har trukket inn prisene eksplisitt, er det klart at vi får en konsumfordelingsmodell i noe videre forstand. En avgiftsøking som vesentlig slår ut i detaljprisene vil vi da kunne beregne effekten av for ulike inntekter, familiestørrelser og sosialgrupper.

— — —

Det som foreløpig er gjort med konsumfordelingsmodellen, er at vi for sosialgruppen funksjonærer har estimert konstantene i relasjonene (2.1) først for 14 vareposter og deretter for 4 vareposter. Materialet består av 765 husholdninger. Problemet har i alt 22 variable, hvorav 4 er avhengige variable (utgiftspostene, se s. 5). For DEUCE finnes et standardprogram for slike regresjonsberegninger. Programmet er temmelig komplisert, og det tar ca. 2 1|2 time å få regnet ut resultatene når datakortene er ferdige for de 765 husholdene. Foruten estimater på regresjonskoeffisientene blir det regnet ut standardavvik for disse, residualvarianser, multiple korrelasjonskoeffisienter m.v.

Materialet for de resterende sosialgrupper skal bearbeides på tilsvarende måte. Dataene foreligger her i en 30-gruppering av utgiftspostene, og arbeidsgangen videre med konsumfordelingsmodellen blir å aggregere dette materialet opp i 4 varegrupper, og å kjøre det tilsvarende program på DEUCE som for gruppen funksjonærer som allerede er ferdig bearbeidd. Regner vi i alt med ytterligere 4 sosialgrupper, skulle selve maskintiden på DEUCE bli ca. 10 timer. Til dette kommer ytterligere tid til tilrettelegging av datakortene og listing av resultatene.

3. Nærmere om anvendelsen av konsumfordelingsmodellen

Det skulle gå fram av det foregående at konsumfordelingsmodellen skulle være vel egnet til å vise hvorledes konsumets sammensetning endres med størrelser som familieinntekt, evt. utgift, familiestørrelse, aldersfordeling og priser.

Fra en rent sosialpolitisk synsvinkel bør modellen ha interesse, idet en her har muligheter for å angi virkningene av de sosialpolitiske virkemidlene. For det andre kan modellen være til hjelp i prognosearbeidet. Gitt f.eks. at priser og inntekt øker i samsvar med en generell målsetting, hvorledes vil da konsumsammensetningen se ut ved slutten av planleggingsperioden?

I det følgende skal vi behandle endel spesifiserte enkle problemstillinger som kan belyse anvendelsen av modellen, idet vi bygger på de numeriske resultatene som allerede foreligger for gruppen funksjonærer.

Problem a: Ikke sjelden drøftes problemet: "Hva betyr det å få et barn til?", eller "Hva koster det å øke familiestørrelsen?" Dette spørsmålet kan f.eks. tolkes slik "Hvor meget endres sammensetningen av konsumet når antall forbruksenheter øker med 0,20 (tilsvarer et nyfødt barn) og inntekten er konstant?" Vi innfører

x_i = utgift til varepost nr. i (i = 1, 2 ... 4)

E_i = utgiftselastisitet for varepost nr. i (i = 1, 2 ... 4)

$F_i = \frac{\partial x_i}{\partial f} \cdot \frac{1}{x_i}$ er den relative øking i x_i når f = familiestørrelsen øker med en enhet

c = totalutgift

Vi har forbindelsen:

$$(3.1) \quad \left(\frac{\Delta x_i}{x_i} \right) = E_i \left(\frac{\Delta c}{c} \right) + F_i \, \Delta f$$

fordi vi har at $x_i = g_i\,(c, f \ldots)$

x_1 står for "matvarer, drikkevarer og tobakk"

x_2 står for "bolig, lys og brensel"

x_3 står for "klær og skotøy"

x_4 står for "annet konsum"

$100\left(\frac{\Delta x_i}{x_i}\right)$ står for den prosentvise øking i godegruppe i

Problem a spesifiserer at $\Delta c = 0$, og $\Delta f = 0{,}20$ i (3.1). Hvis den opprinnelige familiestørrelse er ca. 2 voksne og 1 barn (f = 2,48) og månedsinntekten ca. 1 300,- kroner, vil vi ha at:[1)]

$$(3.2) \quad \begin{cases} 100\left(\frac{\Delta x_1}{x_1}\right) = +\,4\ \% \\ 100\left(\frac{\Delta x_2}{x_2}\right) = -\,6\ \% \\ 100\left(\frac{\Delta x_3}{x_3}\right) = +\,3\ \% \\ 100\left(\frac{\Delta x_4}{x_4}\right) = +\,3\ \% \end{cases}$$

Disse resultatene kan tolkes slik: "Hvis en familie på 2 voksne og 1 barn får et barn til, og når månedsinntekten er gitt ca. kr. 1 300,-, vil matvareutgiftene øke med ca. 4 %, utgiftene til bolig, lys og brensel <u>avtar</u> med ca. 6 %, og utgiftene til klær og skotøy same annet konsum øker hver med ca. 3 %.

---

1) Jfr. tabellene (5.10) - (5.13).

Det betyr altså at en slik familie vil øke matvare-, klær- og skotøy- og utgiftene til annet konsum på bekostning av boligutgiftene. En bruker altså mindre til bolig etc. fordi barnetallet øker med 1, under gitt inntekt.

Problem b: La oss anta at et indekstillegg medfører at totalutgiften øker med 3 °|o for en familie med 2 voksne og 1 barn, samt opprinnelig månedsinntekt ca. kr. 1 300,-. Hvorledes endret da forbrukssammensetning. Her er $(\frac{\Delta c}{c}) = 0{,}03$ og $\Delta f = 0$ i 3.1. Vårt tallmateriale gir oss følgende tallserier:

$$
(3.3) \quad \begin{cases} 100\,(\frac{\Delta x_1}{x_1}) = +\,1{,}4\ ^{\circ}|\mathrm{o} \\ 100\,(\frac{\Delta x_2}{x_2}) = +\,5{,}1\ ^{\circ}|\mathrm{o} \\ 100\,(\frac{\Delta x_3}{x_3}) = +\,3{,}8\ ^{\circ}|\mathrm{o} \\ 100\,(\frac{\Delta x_4}{x_4}) = +\,4{,}4\ ^{\circ}|\mathrm{o} \end{cases}
$$

En partiell utgiftsøking på 3 °|o vil altså medføre at utgifter til matvarer, drikkevarer og tobakk øker med bare 1,4 °|o, utgifter til bolig, lys og brensel med hele 5,1 °|o, utgifter til klær og skotøy med 3,8 °|o, og utgifter til andre varer med 4,4 °|o.

Problemene a og b kan under våre forutsetninger regnes ut for ulike familiestørrelser og ulike totalinntekter. Når f og c endres, vil også E_i og F_i endres, og de søkte $(\frac{\Delta x_i}{x_i})$ vil også være annerledes i eksemplene a og b.

Tabellene (5.6) - (5.13) bak i notatet gir verdier av E_i og F_i for ulike verdier av f og c, og gir derfor et inntrykk av hvorledes konsumsammensetningen vil endres ved endringer av disse variable.

Problem c: Anta at en langtidsmålsetting går ut på å realisere en gitt inntektsfordeling og en gitt familiestørrelse. Hvorledes vil konsumsammensetningen se ut under disse forutsetninger i motsetning til den eksisterende fordeling?

Et anslag på dette kan en få ved å bruke

$$
(3.4) \quad \log x_i = g_i\,(c,\ f\ \ldots.)
$$

og sette inn de ønskede verdier for c og f, slik at vi får generert de ønskede verdier av x_i ved hjelp av (3.4).

Problem d: En øking i den alminnelige omsetningsavgift medfører at alle priser øker med 3 %. Samtidig vil vi anta at inntekten øker med 2 % som følge av prisøkingen. Hvorledes endres konsumsammensetningen for de ulike familiestørrelser og for de ulike inntektsgrupper?

Når prisvirkningene tas hensyn til i modellen, vil vi ha at den relative tilveksten i utgiften til varegruppe nr. i, som følge av endringer i de eksogene variable, blir

$$(3.5) \quad \left(\frac{\Delta x_i}{x_i}\right) = E_i \left(\frac{\Delta c}{c}\right) + F_i \, (\Delta f) + \sum_{j=1}^{4} e_{ij} \left(\frac{\Delta p_j}{p_j}\right)$$

Vår modell gir oss E_i og F_i som funksjoner av c og f. Altså

$$(3.6) \quad \begin{cases} E_i = \theta_{oi} + \theta_{1i} \cdot \log c + \theta_{2i} \cdot f \\ F_i = \gamma_{oi} + \gamma_{1i} \cdot \log c + \gamma_{2i} \cdot f \end{cases}$$

Settes (3.6) inn i (3.5), får vi en modell som spesifiserer problemet d fullstendig. Vi får

$$(3.7) \quad \left(\frac{\Delta x_i}{x_i}\right) = (\theta_{oi} + \theta_{1i} \log c + \theta_{2i} \cdot f) \left(\frac{\Delta c}{c}\right) + (\gamma_{oi} + \gamma_{1i} \log c + \gamma 2_i \cdot f)\,(\Delta f) + \sum_{j=1}^{4} e_{ij} \frac{\Delta p_j}{p_j}$$

Problem d spesifiserer at $\left(\frac{\Delta c}{c}\right) = 0{,}02$ og $\left(\frac{\Delta p_j}{p_j}\right) = 0{,}03$ for alle j. Dessuten tenker vi oss at $\Delta f = 0$ under den betraktede perioden. (3.7) blir da:

$$(3.8) \quad \left(\frac{\Delta x_i}{x_i}\right) = (\theta_{oi} + \theta_{1i} \log c + \theta_{2i} \cdot f)\, 0{,}02 + 0{,}03 \sum_{j=1}^{4} e_{ij}$$

Vår konsumfordelingsmodell gir oss estimater for bl.a. θ_{oi}, θ_{1i} og θ_{2i}. Størrelsen e_{ij} kan vi regne er som i tabell 7 i Artikler nr. 7. Ved hjelp av (3.8) kan vi derfor regne ut $\left(\frac{\Delta x_i}{x_i}\right)$ for ulike f- og c-verdier. Dette vil gi et svar på problem d.

Eksempelvis kan vi sette $\log c \approx 2{,}45$ (tilsvarer $c \sim 1\,300$ kr. pr. mnd.) og $f \approx 2{,}48$ (tilsvarer en familie med 2 voksne og 1 barn). For å løse dette problemet, trenger vi ikke å kjenne matriksen $\{e_{ij}\}$, idet vi har at:

$$(3.9) \quad \sum_j e_{ij} = -E_i \,,$$

når forbruksrelasjonene forutsettes homogene av grad 0 i priser og forbruksutgift; dvs. (3.8) reduserer seg til

$$(3.10) \quad \left(\frac{\Delta x_i}{x_i}\right) = (\theta_{0i} + \theta_{1i} \log c + \theta_{2i} \cdot f) \cdot (-0{,}01)$$

I vårt eksempel får vi

$$(3.11) \quad \begin{cases} 100 \left(\frac{\Delta x_1}{x_1}\right) = -\ 0{,}5\ \% \\ 100 \left(\frac{\Delta x_2}{x_2}\right) = =\ 1{,}7\ \% \\ 100 \left(\frac{\Delta x_3}{x_3}\right) = -\ 1{,}3\ \% \\ 100 \left(\frac{\Delta x_4}{x_4}\right) = -\ 1{,}4\ \% \end{cases}$$

Tolkingen av (3.11) må bli at en familie på 2 voksne og 1 barn med månedsinntekt ca. kr. 1 300,-, som i en periode får et inntektstillegg på ca. 2 % samtidig som alle priser øker med 3 %, vil redusere matvareforbruket med 0,5 %, bolig, lys og brensel med 1,7 %, klær og skotøy med 1,3 %, og endelig reduserer familien annet konsum med 1,4 %. Nettopp <u>denne</u> familien ville ifølge vårt materiale redusere alt forbruk ved en slik pris- og inntektsendring.

Hvis vi derimot antar at månedsinntekten er ca. kr. 5 000,- og familiestørrelsen den samme, vil vi få følgende endringer etter vårt tallmateriale:

$$(3.12) \quad \begin{cases} 100 \left(\frac{\Delta x_1}{x_1}\right) = +\ 0{,}05\ \% \\ 100 \left(\frac{\Delta x_2}{x_2}\right) = -\ 0{,}10\ \% \\ 100 \left(\frac{\Delta x_3}{x_3}\right) = -\ 0{,}90\ \% \\ 100 \left(\frac{\Delta x_4}{x_4}\right) = -\ 1{,}39\ \% \end{cases}$$

(3.11) og (3.12) viser at virkningene av en kombinert pris- og inntektsøking slår meget forskjellig ut for de ulike inntektsgrupper. Hadde vi også latt familiestørrelsen, f, variert, hadde vi fått andre resultater.

Problemene a-d gir et bilde av modellens anvendelsesmuligheter. Våre numeriske løsninger av problemene gjelder sosialgruppen funksjonærer. Når det resterende tallmateriale er bearbeidd, kan vi på tilsvarende måte nytte modellen for de resterende sosialgrupper. Det er a priori grunn til å anta at resultatene vil endres når vi går over fra en sosialgruppe til en annen. En bekreftelse på denne antakelsen har vi allerede i det nevnte hefte 3 av forbruksundersøkelsen 1958.

4. Statistisk presisering av modellen

Konsumfordelingsmodellen bygger på følgende atferdsrelasjoner:

$$E(\log x_{ij}) = \beta_{oi} + \beta_{1i} \log c_j + \beta_{2i} (\log c_j)^2 + \beta_{3i} f_j + \beta_{4i} f_j^2 + \beta_{5i} f_j (\log c_j) + \sum_{s=1}^{11} \alpha_{si} Q_{sj} + \gamma_{1i} Z_{1j} + \gamma_{2i} Z_{2j} \tag{4.1}$$

x_{ij}	= utgiften til varegruppe i for familie j	$(i = 1, 2, 3, 4)$ $(j = 1, \dots 765)$
c_j	= totalutgiften til familie j	$(j = 1, 2.. 765)$
f_j	= familiestørrelsen til familie j	$(j = 1, 2.. 765)$
Q_{sj}	= sesongvariabel	$(s = 1, 2... 11)$ $(j = 1, 2.. 765)$
Z_{hj}	= aldersvariabel	$(h = 1, 2)$ $(j = 1, 2.. 765)$

x_{ij} og c_j måles i absolutte kroner. f_j måles i antall forbruksenheter omregnet etter den skotske skala. Q_{sj} er dummy-variable (som bare kan anta verdiene 0 og 1). Z_{hj} er et annet sett av dummy-variable som tar vare på aldereffekten i konsumet. Alderen er delt inn i 3 grupper: fra 0 til 35 år, fra og med 35 år til 55 år, og mer enn 55 år.

Som jeg tidligere har vært inne på, avhenger nivået av konsumet av alderen, men selve elastisitetsbegrepene er uavhengige av alderen innen vår modell.

Vi har i alt 22 variable, hvorav 4 er "avhengige" variable (x_i, $i = 1, 2, 3, 4$). De resterende variable er regnet som uavhengige.

Koeffisientene $\beta_{oi} \dots \beta_{5i}$, $\alpha_{1i} \dots \alpha_{11i}$ og γ_{1i}, γ_{2i} er beregnet ved minste kvadraters metode på DEUCE etter et standardprogram.

Filosoferingen om minste kvadraters metode er relevant under estimeringen av (4.1) kan diskuteres, men diskusjonen er neppe fruktbar.

Jeg skal nevne enkelte momenter som taler mot minste kvadraters metode. For det første har vi betingelsen at $c_j = \sum_{i=1}^{4} x_{ij}$; denne restriksjonen er det ikke tatt hensyn til under estimeringen, og den kan kanskje medføre at våre estimater blir "biassed". Noen særlig sterk innvendelse er ikke dette, da vi har en rekke kontroller på at estimatene er rimelige. For det første tidligere beregninger (tidsrekkedata f.eks.), og for det andre at f.eks. $\sum_{i=1}^{4} \alpha_i E_i = 1$, hvor α_i er budsjettprosenten for varegruppe i. Denne sistnevnte relasjonen følger av den tradisjonelle teorien for konsumentenes tilpassing.

Andre kontroller er også anvendt, og det ser ut til at våre estimater er rimelige, selv med så få grupper som 4, da en må anta at budsjettrestriksjonen $c_j = \sum_{i=1}^{4} x_{ij}$ er nokså sterk.

For det andre kan innvendes at f.eks. c_j er en stokastisk størrelse, og ikke et gitt datum. Denne innvendingen er ikke særlig relevant, idet vi i dette tilfelle opererer med betingede begreper når vi refererer til estimatene. (Vi har undersøkt fordelingen for c spesielt, og en χ^2-føyningstest viser at vi ikke kan forkaste hypotesen om at c er log-normalt fordelt. Se appendiks I i dette notatet.)

En annen metode som kunne vært relevant under estimeringen er "instrumentvariabelteknikken". Da antall variable er nokså stort, og da denne metoden ikke er programmert for DEUCE, måtte vi se bort fra denne muligheten. (En artikkel om instrumentvariabelmetoden i forbruksundersøkelsen kan en finne i Econometrica, vol. 29, No. 3, 1962, pp. 336-362 - av N. Livitian.)

For de 765 funksjonærhusholdene ble det i alt kjørt ut 16 regresjonslikninger, idet vi ville prøve effekten av å utelate visse variable på høyresiden i (4.1). Dette gir oss muligheten av å teste om disse variable skal være med i (4.1) når vi forutsetter at restleddene i (4.1) er N $(0, \sigma)$.

Hefte nr. 22 i Statistisk Sentralbyrås håndbøker beskriver disse testene.

Resultatene av testingen er at det er relevant å ha med de sesongvariable i (4.1). Under tvil skal leddet (log c) · f være med. De aldersvariable bør være med.

5. Enkelte numeriske resultater for konsumfordelingsmodellen - funksjonærer

Følgende tabeller vil gi estimatene på koeffisientene i (4.1) for de 4 varepostene. I parentes under koeffisientene er standardavvikene for disse tilføyd med et pluss-minus tegn. Videre finner en de multiple korrelasjons-koeffisienter (R) og residualvariansene (σ_u^2) i tabellene.

Tabellen (5.1) gir en oversikt over de estimerte koeffisientene i relasjonen (4.1). I kolonne 1 i tabellen (merket "koeffisienter") står symbolene som tilsvarer denne relasjonen. Annen kolonne gir eksempelvis de estimerte koeffisientene i (4.1) for gruppen "matvarer, drikkevarer og tobakk". Det første tallet i denne kolonnen gir estimatet på konstantleddet β_{oi} i relasjon (4.1).

I parentes under estimatene har vi som nevnt tilføyd spredningene for koeffisientene tilføyd et pluss-minus tegn.

En koeffisient er skarpt bestemt om spredningen er "moderat" i forhold til størrelsen på koeffisienten.

Av relasjon (4.1) ser vi at koeffisientene $\alpha_{1i} \ldots \alpha_{11i}$ gir uttrykk for betydningen av sesongbevegelsen i materialet.

Tar vi f.eks. kolonne 2 i tabell (5.1), ser vi at α_{11} er - 0,1048. Det kan tolkes som om sesongkomponenten i <u>februar</u>-forbruket er ca. 10 °|o mindre enn sesongkomponenten i januar. Likeledes er α_{11i} + 0,1041. I dette tilfellet er det i desember et 10 °|o tillegg.

Koeffisientene γ_{ij} i tabell (5.1) sier noe om <u>alderens</u> innflytelse på konsumet. Vi har her brukt familieoverhodets alder i 1958.

Eksempelvis er γ_{23} lik - 0,3194. Det betyr at utgiftene til klær og skotøy avtar med ca. 32 °|o når en går over til å bli mer eller lik 55 år, i forhold til den tid en var i aldersgruppen 0 til 35 år.

Som en ser av spredningsmålene for koeffisientene α_{ij} og γ_{ij}, så er disse ikke alltid skarpt bestemt. Imidlertid er koeffisientene testet ved hjelp av delspesifiserte testemetoder, og det er ikke mulig å forkaste f.eks. den hypotese at alle $\alpha_{ij} = 0$ (testnivå 0,95).

Koeffisientene β_{ij} er rimelig godt bestemte unntaken kanskje β_{5i} som er koeffisienten foran leddet $f \cdot \log c$ i relasjon (4.1).

Tabell (5.1)

Koeffi-sienter	Matvarer, drikkevarer og tobakk	Bolig, lys og brensel	Klær og skotøy	Annet konsum
β_{oi}	- 1,2928 (± 0,0102)	- 4,5915 (± 0,0460)	- 4,0148 (± 0,0473)	- 1,4482 (± 0,0158)
β_{1i}	1,2457 (± 0,1193)	3,5650 (± 0,5371)	1,8945 (± 0,5540)	1,3561 (± 0,1848)
β_{2i}	- 0,1785 (± 0,0313)	- 0,5437 (± 0,1409)	- 0,1205 (± 0,1453)	- 0,0178 (± 0,0485)
β_{3i}	0,3571 (± 0,0648)	- 0,7713 (± 0,2916)	- 0,0262 (± 0,3007)	- 0,4646 (± 0,1003)
β_{4i}	- 0,0511 (± 0,0080)	- 0,0515 (± 0,0360)	0,0438 (± 0,0371)	0,0311 (± 0,0124)
β_{5i}	0,0348 (± 0,0269)	0,3034 (± 0,1212)	- 0,0215 (± 0,1250)	0,0715 (± 0,0417)
α_{1i}	- 0,1048 (± 0,0502)	- 0,0559 (± 0,2262)	- 0,1412 (± 0,2333)	0,1383 (± 0,0778)
α_{2i}	0,0397 (± 0,0482)	- 0,0429 (± 0,2171)	- 0,4028 (± 0,2239)	0,0703 (± 0,0747)
α_{3i}	0,0117 (± 0,0491)	0,0551 (± 0,2212)	0,0076 (± 0,2282	- 0,1296 (± 0,0761)
α_{4i}	0,0535 (± 0,0500)	- 0,2940 (± 0,2253)	0,0013 (± 0,2324)	- 0,0208 (± 0,0775)
α_{5i}	0,0214 (± 0,0551)	- 0,4005 (± 0,2480)	0,2803 (± 0,2558)	- 0,1289 (± 0,0854)
α_{6i}	- 0,0412 (± 0,0567)	- 0,5504 (± 0,2554	- 0,3552 (± 0,2634)	0,1446 (± 0,0879)
α_{7i}	0,0250 (± 0,0495)	- 0,4179 (± 0,2229)	- 0,0290 (± 0,2298)	0,0620 (± 0,0767)
α_{8i}	0,0698 (± 0,0509)	- 0,4618 (± 0,2290)	- 0,3233 (± 0,2362)	0,0608 (± 0,0788)
α_{9i}	0,0648 (± 0,0493)	- 0,1942 (± 0,2221)	0,1749 (± 0,2291)	- 0,1624 (± 0,0764)
α_{10i}	0,0154 (± 0,0477)	- 0,3009 (± 0,2148)	0,3158 (± 0,2215)	- 0,0902 (± 0,0739)
α_{11i}	0,1041 (± 0,0488)	- 0,8233 (± 0,2197)	0,1323 (± 0,2266)	0,0826 (± 0,0756)
γ_{1i}	0,0633 (± 0,0299)	0,0059 (± 0,1341)	- 0,1543 (± 0,1389)	- 0,0064 (± 0,0463)
γ_{2i}	0,1329 (± 0,0331)	- 0,1655 (± 0,1489)	- 0,3194 (± 0,1535)	- 0,0482 (± 0,0512)

De multiple korrelasjonskoeffisienter og residualvariansene finner en i tabell (5.2):

Tabell (5.2)

Varegruppe	R	σ_u^2
Matvarer, drikkevarer og tobakk	0,82	0,079
Bolig, lys og brensel	0,54	1,611
Klær og skotøy	0,53	1,711
Annet konsum	0,85	0,191

Gjennomsnittet av log c i materialet og gjennomsnittet av f er henholdsvis 2,4163 og 2,4795. Engel-elastisitetene for de 4 varegruppene blir da $(E_i = \frac{\partial \log x_i}{\partial \log c})$:

$$
(5.3) \quad \begin{cases} E_1 = 0{,}4694 - 0{,}3570\,(\log c - 2{,}4163) + 0{,}0348\,(f - 2{,}4795) \\ E_2 = 1{,}6898 - 1{,}0874\,(\log c - 2{,}4163) + 0{,}3034\,(f - 2{,}4795) \\ E_3 = 1{,}2589 - 0{,}2410\,(\log c - 2{,}4163) - 0{,}0215\,(f - 2{,}4795) \\ E_4 = 1{,}4474 - 0{,}0356\,(\log c - 2{,}4163) + 0{,}0715\,(f - 2{,}4795) \end{cases}
$$

Størrelsen $F_i = \frac{\partial \log x_i}{\partial f}$ er gitt ved:

$$
(5.4) \quad \begin{cases} F_1 = 0{,}1878 - 0{,}1022\,(f - 2{,}4795) + 0{,}0348\,(\log c - 2{,}4163) \\ F_2 = -0{,}2936 - 0{,}1030\,(f - 2{,}4795) + 0{,}3034\,(\log c - 2{,}4163) \\ F_3 = 0{,}1390 + 0{,}0876\,(f - 2{,}4795) - 0{,}0215\,(\log c - 2{,}4163) \\ F_4 = -0{,}1376 + 0{,}0622\,(f - 2{,}4795) + 0{,}0715\,(\log c - 2{,}4163) \end{cases}
$$

Likningene (5.3) og (5.4) er skrevet på formen "om sitt gjennomsnitt" for å lette oversikten.

Som nevnt er

$$
(5.5) \quad \frac{\Delta x_i}{x_i} = E_i\left(\frac{\Delta c}{c}\right) + F_i\,\Delta f \quad \Delta \qquad \text{(når Q og Z er gitte tall)}
$$

og tabeller over E_i og F_i for ulike verdier av f og log c vil være av nytte når en skal regne ut de relative utgiftsendringer som følge av endringer i c og f.

Følgende tabeller gir E_i- og F_i-verdier for ulike verdier av log c og f :

f = 0,75	betyr	et	hushold	med	en	enslig	over 65 år (mann eller kvinne)				
= 0,83	"	"	"	"	"	"	kvinne 14 til 65 år				
= 1,00	"	"	"	"	"	"	herre 14 til 65 år				
= 1,50	"	"	"	"	ektepar uten barn over 65 år						
= 1,83	"	"	"	"	"	"	" 14 til 65 år				
= 2,43	"	"	"	"	"	med 1 barn (6 - 8 år)					
= 3,73	"	"	"	"	"	" 3 " under 15 år					
= 4,83	"	"	"	"	"	" 4 "					
= 6,03	"	"	"	"	"	" 7 "					

Tabellene (5.6) til (5.13) gir verdier av de tidligere definerte størrelsene E_i og F_i for ulike verdier av f (= familiestørrelsen) og c (= månedsutgiften).

Eksempelvis gir tabell (5.6) utgiftselastisiteten, E_1, for gruppen "matvarer, drikkevarer og tobakk".

Vi ser at for svært store månedsutgifter (4 000 til 5 000 kroner) er E_i negativ. Disse resultatene må tas "cum grano salis", da våre relasjoner neppe er helt tilfredsstillende for så ekstreme utgiftsgrupper.

Det er prinsipielt den "midtre del" av tabellene som er av vesentlig interesse, og som er best bestemt sett fra et statistisk synspunkt.

Tabuleringen av F_i og E_i er av interesse når vi skal regne ut

$$\underline{\left(\frac{\Delta x}{x}\right)_i = E_i \frac{\Delta c}{c} + F_i \, \Delta f}$$

for ulike verdier av c og f, og viser dessuten at elastisitetene varierer betydelig med c og f i de fleste tilfelle.

En forutsetning om konstante elastisiteter i vårt materiale ville derfor neppe være relevant. En av de vesentlige forutsetninger for konsumfordelingsmodellen var at en slik avhengighet var til stede.

Bak i dette notatet er tabellene illustrert grafisk for to verdier av f og ulike verdier av c. De grafiske figurene gir en grei oversikt over elastisitetenes variasjonsområde når de jamføres med tabellene (5.6) til (5.13).

Tabell 5.6

Matvarer, drikkevarer og tobakk. $E_l = 1,2457 - 0,3570 \cdot \log c + 0,0348 \cdot f$

f	Mnd. inntekt c og log c	Forbruksutgift												
	c	300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
	log c	1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
Antall forbruksenheter	0,75	0,8795	0,7952	0,5981	0,4428	0,3347	0,2522	0,1847	0,1283	0,0798	0,0369	0,0027	-0,0452	-0,1248
	0,83	0,8823	0,7980	0,6009	0,4456	0,3375	0,2550	0,1875	0,1311	0,0826	0,0397	0,0055	-0,0424	-0,1220
	1,00	0,8882	0,8039	0,6068	0,4515	0,3434	0,2609	0,1934	0,1370	0,0085	0,0456	0,0114	-0,0365	-0,1161
	1,50	0,9056	0,8213	0,6242	0,4689	0,3608	0,2783	0,2108	0,1544	0,1059	0,0630	0,0288	-0,0191	-0,0987
	1,83	0,9171	0,8328	0,6357	0,4804	0,3723	0,2898	0,2223	0,1659	0,1174	0,0745	0,0403	-0,0076	-0,0872
	2,43	0,9380	0,8537	0,6566	0,5013	0,3932	0,3107	0,2432	0,1868	0,1383	0,0954	0,0612	0,0133	-0,0663
	2,83	0,9519	0,8676	0,6705	0,5152	0,4071	0,3246	0,2571	0,2007	0,1522	0,1093	0,0751	0,0272	-0,0524
	3,73	0,9832	0,8989	0,7018	0,5465	0,4384	0,3559	0,2884	0,2320	0,1835	0,1406	0,1064	0,0585	-0,0211
	4,83	1,0215	0,9372	0,7401	0,5848	0,4767	0,3942	0,3267	0,2703	0,2218	0,1789	0,1447	0,0968	0,0172
	6,03	1,0632	0,9789	0,7818	0,6265	0,5184	0,4359	0,3684	0,3120	0,2635	0,2206	0,1864	0,1385	0,0589

Tabell 5.7

Bolig, lys, brensel. $E_{\ell} = 3,5650 - 1,0874 \cdot \log c + 0,3034 \cdot f$

f	Mnd.inntekt C og log c	300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
		1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
Antall forbruksenheter	0,75	2,5975	2,3409	1,7407	1,2677	0,9382	0,6670	0,4815	0,3097	0,1618	0,0313	-0,0731	-0,2188	-0,4613
	0,83	2,6217	2,3651	1,7649	1,2919	0,9624	0,7112	0,5057	0,3339	0,1860	0,0555	-0,0489	-0,1946	-0,4371
	1,00	2,6733	2,4167	1,8165	1,3435	1,0140	0,7628	0,5573	0,3855	0,2376	0,1071	0,0027	-0,1430	-0,3855
	1,50	2,8250	2,5684	1,9682	1,4952	1,1657	0,9145	0,7090	0,5372	0,3893	0,2568	0,1544	0,0087	-0,2338
	1,83	2,9251	2,6685	2,0683	1,5953	1,2658	1,0146	0,8091	0,6373	0,4894	0,3589	0,2545	,1088	-0,1337
	2,43	3,1072	2,8506	2,2504	1,7774	1,[illegible]479	1,1967	0,9912	0,8194	0,6715	0,5410	0,4366	0,2909	0,0484
	2,83	3,2285	2,9719	2,3717	1,8987	1,5692	1,3180	1,1125	0,9407	0,7928	0,6623	0,5579	0,4122	0,1697
	3,73	3,5016	3,2450	2,6448	2,1718	1,8423	1,5911	1,3856	1,2138	1,0659	0,9354	0,8310	0,6853	0,4428
	4,83	3,8353	3,5787	2,9785	2,5055	2,1760	1,9248	1,7193	1,5475	1,3996	1,2691	1,1647	1,0190	0,7765
	6,03	4,1994	3,9428	3,3426	2,8696	2,5401	2,2889	2,0834	1,9116	1,7637	1,6332	1,5288	1,3831	1,1406

Tabell 5.8

Klær og sko. $E_3 = 1{,}8945 - 0{,}2410 \cdot \log c - 0{,}0215 \cdot f$

	Mnd.inntekt C og log c / f	Forbruksutgift												
		300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
		1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
Antall forbruksenheter	0,75	1,6135	1,5567	1,4236	1,3188	1,2458	1,1901	1,1446	1,1065	1,0737	1,0448	1,0216	0,9894	0,9356
	0,83	1,6118	1,5550	1,4219	1,3171	1,2441	1,1884	1,1429	1,1048	1,0720	1,0431	1,0199	0,9877	0,9339
	1,00	1,6081	1,5513	1,4182	1,3134	1,2404	1,1847	1,1392	1,1011	1,0683	1,0394	1,0162	0,9840	0,93[illegible]
	1,50	1,5973	1,5405	1,4074	1,3026	1,2296	1,1739	1,1284	1,0903	1,0575	1,0286	1,0054	0,9732	0,9194
	1,83	1,5903	1,5335	1,4004	1,2956	1,2226	1,1669	1,1214	1,0833	1,0505	1,0216	0,9984	0,9662	0,9124
	2,43	1,5774	1,5206	1,3875	1,2827	1,2097	1,1540	1,1085	1,0704	1,0376	1,0087	0,9855	0,9533	0,8995
	2,83	1,5688	1,5120	1,3789	1,2741	1,2011	1,1454	1,0999	1,0618	1,0290	1,0001	0,9769	0,9447	0,8909
	3,73	1,5494	1,4926	1,3595	1,2547	1,1817	1,1260	1,0805	1,0424	1,0096	0,9807	0,9575	0,9253	0,8715
	4,83	1,5258	1,4690	1,3359	1,2311	1,1581	1,1024	1,0569	1,0188	0,9860	0,9571	0,9339	0,9017	0,8479
	6,03	1,5000	1,4432	1,3101	1,2053	1,1323	1,0766	1,0311	0,9930	0,9602	0,9313	0,9081	0,8759	0,8221

Tabell 5.9

Annet konsum. $E_4 = 1{,}3561 - 0{,}0356 \cdot \log c + 0{,}0715 \cdot f$

	f \ Mnd. inntekt C og log c	300 1,099	380 1,335	660 1,887	1 020 2,322	1 380 2,625	1 740 2,856	2 100 3,045	2 460 3,203	2 820 3,339	3 180 3,459	3 500 3,555	4 000 3,689	5 000 3,912
		Forbruksutgift												
Antall forbruksenheter	0,75	1,3706	1,3622	1,3425	1,3270	1,3162	1,3080	1,3013	1,2957	1,2908	1,2866	1,2831	1,2784	1,2704
	0,83	1,3763	1,3679	1,3482	1,3327	1,3219	1,3137	1,3070	1,3014	1,2965	1,2923	1,2888	1,2841	1,2761
	1,00	1,3885	1,3801	1,3604	1,3449	1,3341	1,3259	1,3192	1,3136	1,3087	1,3045	1,3010	1,2963	1,2883
	1,50	1,4243	1,4159	1,3962	1,3807	1,3699	1,3617	1,3550	1,3494	1,3445	1,3403	1,3368	1,3321	1,3241
	1,83	1,4478	1,4394	1,4197	1,4042	1,3934	1,3852	1,3785	1,3729	1,3680	1,3638	1,3603	1,3556	1,3476
	2,43	1,4907	1,4823	1,4626	1,4471	1,4363	1,4281	1,4214	1,4158	1,4109	1,4067	1,4032	1,3985	1,3905
	2,83	1,5193	1,5109	1,4912	1,4757	1,4649	1,4567	1,4500	1,4444	1,4395	1,4353	1,4318	1,4271	1,4191
	3,73	1,5837	1,5753	1,5556	1,5401	1,5293	1,5211	1,5144	1,5088	1,5039	1,4997	1,4962	1,4915	1,4835
	4,83	1,6623	1,6539	1,6342	1,6187	1,6079	1,5997	1,5930	1,5874	1,5825	1,5783	1,5748	1,5701	1,5621
	6,03	1,7481	1,7397	1,7200	1,7045	1,6937	1,6855	1,6788	1,6732	1,6683	1,6641	1,6606	1,6559	1,6479

Tabell 5.10

Matvarer, drikkevarer og tobakk. $F_1 = 0{,}3571 - 0{,}1022 \cdot f + 0{,}0348 \cdot \log c$

f — Antall forbruksenheter	Mnd.inntekt c og log c	Forbruksutgift												
	c	300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
	log c	1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
0,75		0,3186	0,3269	0,3461	0,3612	0,3718	0,3798	0,3864	0,3919	0,3966	0,4008	0,4041	0,4088	0,4165
0,83		0,3105	0,3188	0,3380	0,3531	0,3637	0,3717	0,3783	0,3838	0,3885	0,3927	0,3960	0,4007	0,4084
1,00		0,2931	0,3014	0,3206	0,3357	0,3463	0,3543	0,3609	0,3664	0,3711	0,3753	0,3786	0,3833	0,3910
1,50		0,2420	0,2503	0,2695	0,2846	0,2952	0,3032	0,3098	0,3153	0,3200	0,3242	0,3275	0,3322	0,3399
1,83		0,2083	0,2166	0,2358	0,2509	0,2615	0,2695	0,2761	0,2816	0,2863	0,2905	0,2938	0,2985	0,3062
2,43		0,1470	0,1553	0,1745	0,1896	0,2002	0,2082	0,2148	0,2203	0,2250	0,2292	0,2325	0,2372	0,2449
2,83		0,1061	0,1144	0,1336	0,1487	0,1593	0,1673	0,1739	0,1794	0,1841	0,1883	0,1916	0,1963	0,2040
3,73		0,0141	0,0224	0,0416	0,0567	0,0673	0,0753	0,0819	0,0874	0,0921	0,0963	0,0996	0,1043	0,1120
4,83		-0,0983	-0,0900	-0,0708	- 0,0557	-0,0451	-0,0371	-0,0305	-0,0250	-0,0203	-0,0161	-0,0128	-0,0081	-0,0004
6,03		-0,2210	-0,2127	-0,1935	-0,1784	-0,1678	-0,1598	-0,1532	-0,1477	-0,1430	-0,1388	-0,1355	-0,1308	-0,1231

Tabell 5.11

Bolig, lys og brensel. $F_2 = -0,7713 - 0,1030 \cdot f + 0,3034 \cdot \log c$

f \ Mnd. inntekt (c) og log c		Forbruksutgift												
		300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
		1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
Antall forbruksenheter	0,75	-0,5152	-0,4436	-0,2761	-0,1441	-0,0522	0,0179	0,0753	0,1232	0,1645	0,2009	0,2300	0,2706	0,3383
	0,83	-0,5234	-0,4518	-0,2843	-0,1523	-0,0604	0,0097	0,0671	0,1150	0,1563	0,1927	0,2218	0,2624	0,3301
	1,00	-0,5409	-0,4693	-0,3018	-0,1698	-0,0779	-0,0078	0,0496	0,0975	0,1388	0,1752	0,2043	0,2449	0,3126
	1,50	-0,5924	-0,5208	-0,3533	-0,2213	-0,1294	-0,0593	-0,0019	0,0460	0,0873	0,1237	0,1528	0,1934	0,2611
	1,83	-0,6264	-0,5548	-0,3873	-0,2553	-0,1634	-0,0933	-0,0359	0,0120	0,0533	0,0897	0,1188	0,1594	0,2271
	2,43	-0,6882	-0,6166	-0,4491	-0,3171	-0,2252	-0,1551	-0,0977	-0,0498	-0,0085	0,0279	0,0570	0,0976	0,1653
	2,83	-0,7294	-0,6578	-0,4903	-0,3583	-0,2664	-0,1963	-0,1389	-0,0910	-0,0497	-0,0133	0,0158	0,0564	0,1241
	3,73	-0,8221	-0,7505	-0,5830	-0,4510	-0,3591	-0,2890	-0,2316	-0,1837	-0,1424	-0,1060	-0,0769	-0,0363	0,0314
	4,83	-0,9354	-0,8638	-0,6963	-0,5643	-0,4724	-0,4023	-0,3449	-0,2970	-0,2557	-0,2193	-0,1902	-0,1496	-0,0819
	6,03	-1,0[illegible]0	-0,9874	-0,8199	-0,6879	-0,5960	-0,5259	-0,4685	-0,4206	-0,3793	-0,3429	-0,3138	-0,2732	-0,2055

Tabell 5.12

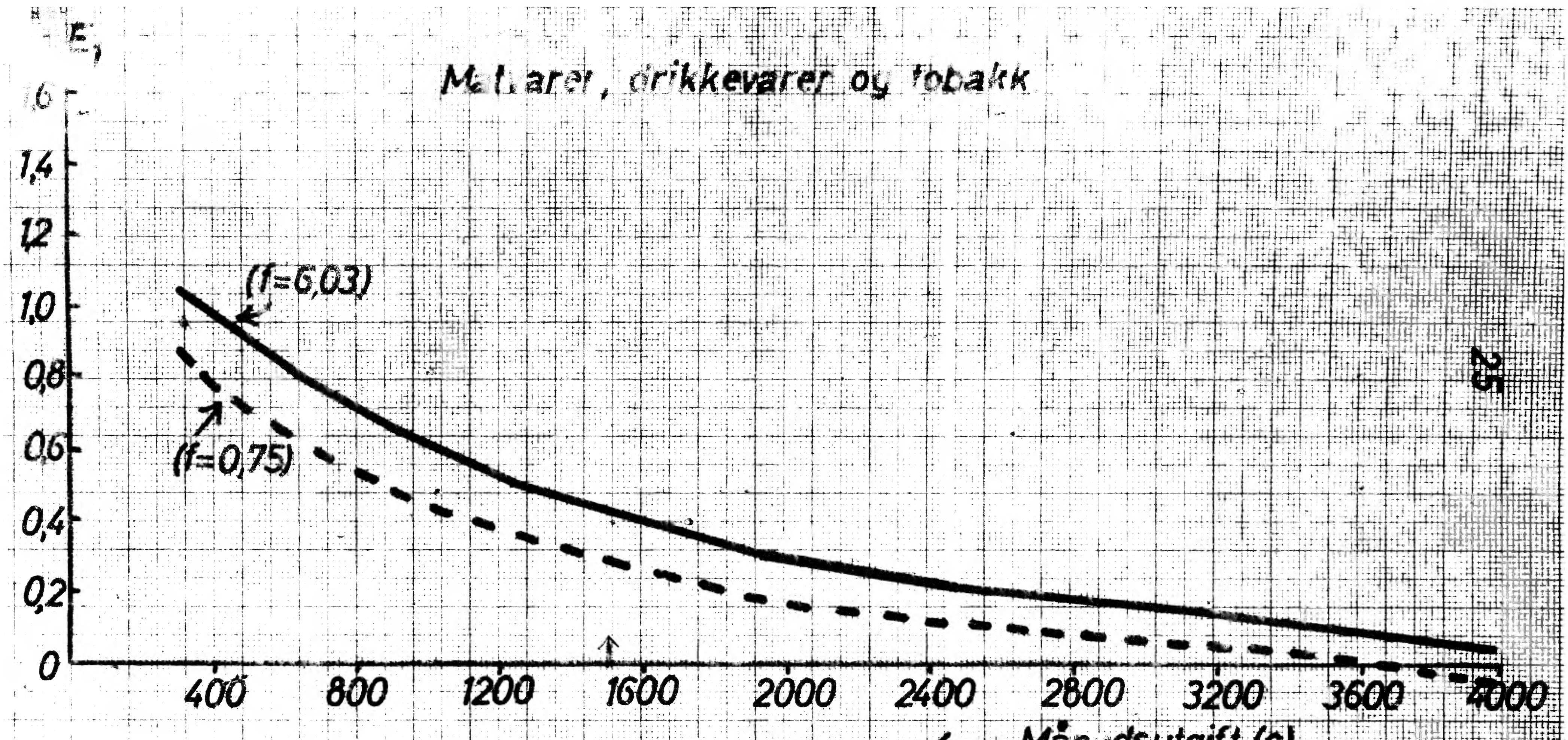
Klær og sko. $F_3 = -0,0262 + 0,0876 \cdot f - 0,0215 \cdot \log c$

f \ Mnd.inntekt C og log c		Forbruksutgift												
		300	380	660	1 020	1 380	1 740	2 100	2 460	2 820	3 180	3 500	4 000	5 000
		1,099	1,335	1,887	2,322	2,625	2,856	3,045	3,203	3,339	3,459	3,555	3,689	3,912
Antall forbruksenheter	0,75	0,0159	0,0108	-0,0011	-0,0104	-0,0169	-0,0219	-0,0260	-0,0294	-0,0323	-0,0349	-0,0369	-0,0398	-0,0446
	0,83	0,0229	0,0178	0,0059	-0,0034	-0,0099	-0,0149	-0,0190	-0,0224	-0,0253	-0,0279	-0,0299	-0,0328	-0,0376
	1,00	0,0378	0,0327	0,0208	0,0115	0,0050	0,0000	-0,0041	-0,0075	-0,0104	-0,0130	-0,0150	-0,0179	-0,0227
	1,50	0,0816	0,0765	0,0646	0,0553	0,0488	0,0438	0,0397	0,0363	0,0334	0,0308	0,0288	0,0259	0,0211
	1,83	0,1105	0,1054	0,0935	0,0842	0,0777	0,0727	0,0686	0,0652	0,0623	0,0597	0,0577	0,0548	0,0500
	2,43	0,1631	0,1580	0,1461	0,1368	0,1303	0,1253	0,1212	0,1178	0,1149	0,1123	0,1103	0,1074	0,1026
	2,83	0,1981	0,1930	0,1811	0,1718	0,1653	0,1603	0,1562	0,1528	0,1499	0,1473	0,1453	0,1424	0,1376
	3,73	0,2769	0,2718	0,2599	0,2506	0,2441	0,2391	0,2350	0,2316	0,2287	0,2261	0,2241	0,2212	0,2164
	4,83	0,3733	0,3682	0,3563	0,3470	0,3405	0,3355	0,3314	0,3280	0,3251	0,3225	0,3205	0,3176	0,3128
	6,03	0,4784	0,4733	0,4614	0,4521	0,4456	0,4406	0,4365	0,4331	0,4302	0,4276	0,4256	0,4227	0,4179

Tabell 5.13

Annet konsum. $F_4 = -0,4646 + 0,0622 \cdot f + 0,0715 \cdot \log c$

f \ Mnd. inntekt C og $\log_e$	Forbruksutgift 300 / 1,099	380 / 1,335	660 / 1,887	1 020 / 2,322	1 380 / 2,625	1 740 / 2,856	2 100 / 3,045	2 460 / 3,203	2 820 / 3,339	3 180 / 3,459	3 500 / 3,555	4 000 / 3,689	5 000 / 3,912
Antall forbruksenheter 0,75	-0,3393	-0,3224	-0,2830	-0,2519	-0,2302	-0,2137	-0,2002	-0,1889	-0,1792	-0,1706	-0,1637	-0,1541	-0,1382
0,83	-0,3344	-0,3175	-0,2781	-0,2470	-0,2253	-0,2088	-0,1953	-0,1840	-0,1743	-0,1657	-0,1588	-0,1492	-0,1333
1,00	-0,3238	-0,3069	-0,2675	-0,2364	-0,2147	-0,1982	-0,1847	-0,1734	-0,1637	-0,1551	-0,1482	-0,1386	-0,1227
1,50	-0,2927	-0,2758	-0,2364	-0,2053	-0,1836	-0,1671	-0,1536	-0,1423	-0,1326	-0,1240	-0,1171	-0,1075	-0,0916
1,83	-0,2722	-0,2553	-0,2159	-0,1848	-0,1631	-0,1466	-0,1331	-0,1218	-0,1121	-0,1035	-0,0966	-0,0870	-0,0711
2,43	-0,2349	-0,2180	-0,1786	-0,1475	-0,1258	-0,1093	-0,0958	-0,0845	-0,0748	-0,0662	-0,0593	-0,0497	-0,0338
2,83	-0,2100	-0,1931	-0,1537	-0,1226	-0,1009	-0,0844	-0,0709	-0,0596	-0,0499	-0,0413	-0,0344	-0,0248	-0,0089
3,73	-0,1540	-0,1371	-0,0977	-0,0666	-0,0449	-0,0284	-0,0149	-0,0036	0,0061	0,0147	0,0216	0,0312	0,0471
4,83	-0,0856	-0,0687	-0,0293	0,0018	0,0235	0,0400	0,0535	0,0648	0,0745	0,0831	0,0900	0,0996	0,1155
6,03	-0,0109	0,0060	0,0454	0,0765	0,0982	0,1147	0,1282	0,1395	0,1492	0,1578	0,1647	0,1743	0,1902



E_i
Matvarer, drikkevarer og tobakk
1,6
1,4
1,2
1,0
0,8
0,6
0,4
0,2
0
(f=6,03)
(f=0,75)
400
800
1200
1600
2000
2400
2800
3200
3600
4000
← Månedsutgift (c)

Fig.(5.14).

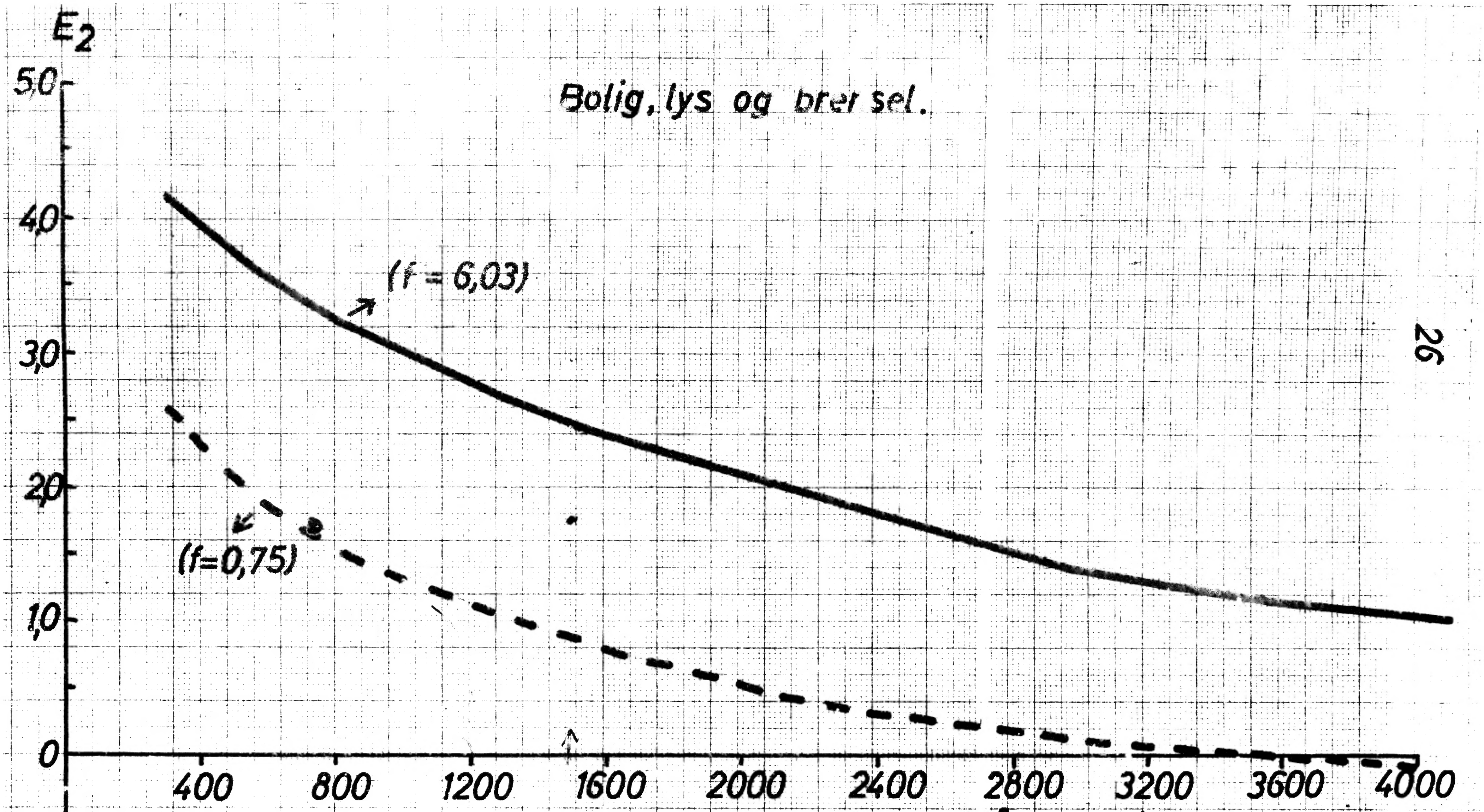
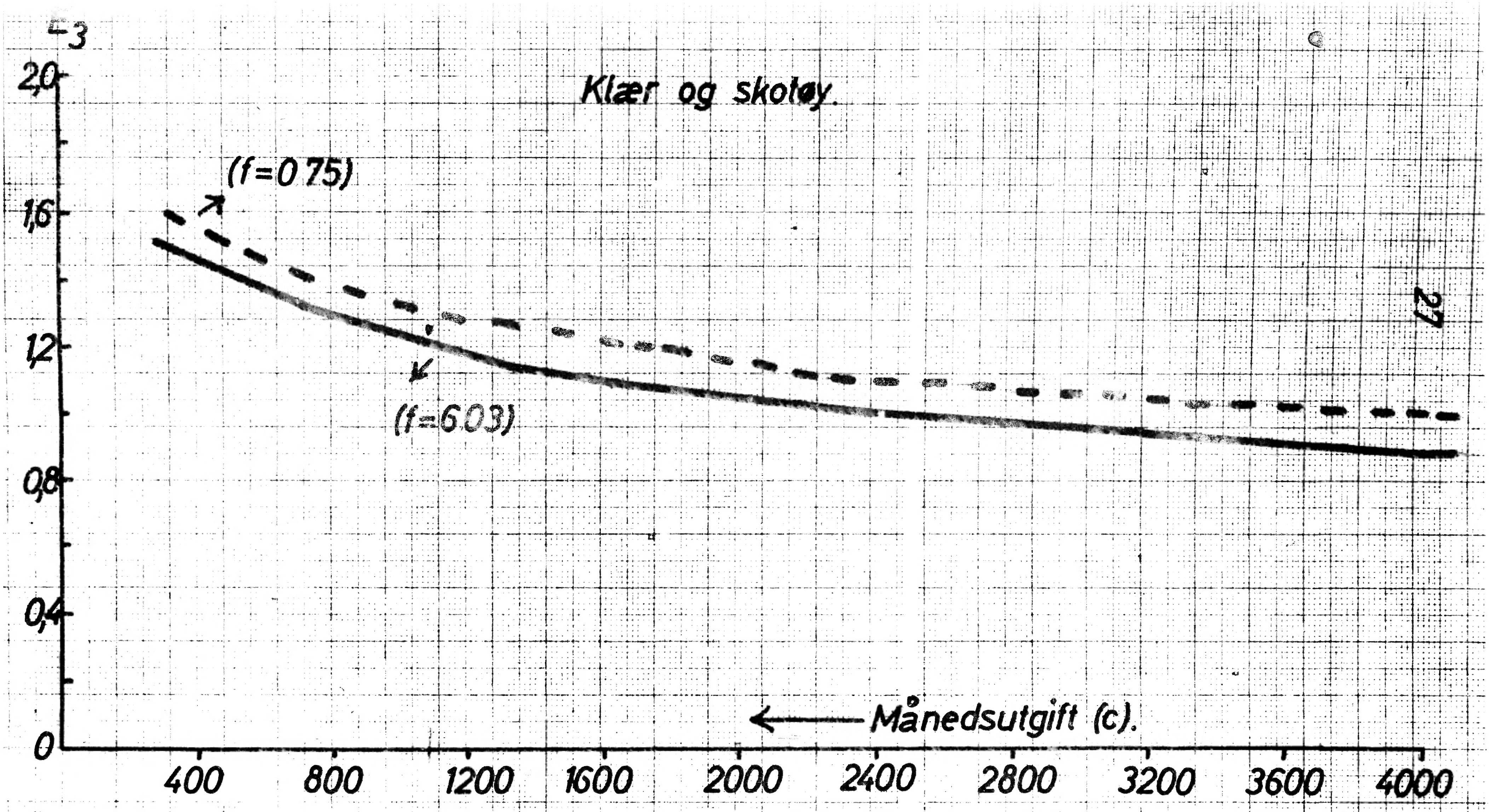

E2
Bolig, lys og brensel.
(f = 6,03)
(f=0,75)
5,0
4,0
3,0
2,0
1,0
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Månedsutgift (c).

Fig.(5.15)



E_3
2,0
1,6
1,2
0,8
0,4
0
Klær og skotøy.
(f=0.75)
(f=6.03)
Månedsutgift (c).
400
800
1200
1600
2000
2400
2800
3200
3600
4000

Fig.(5.16).

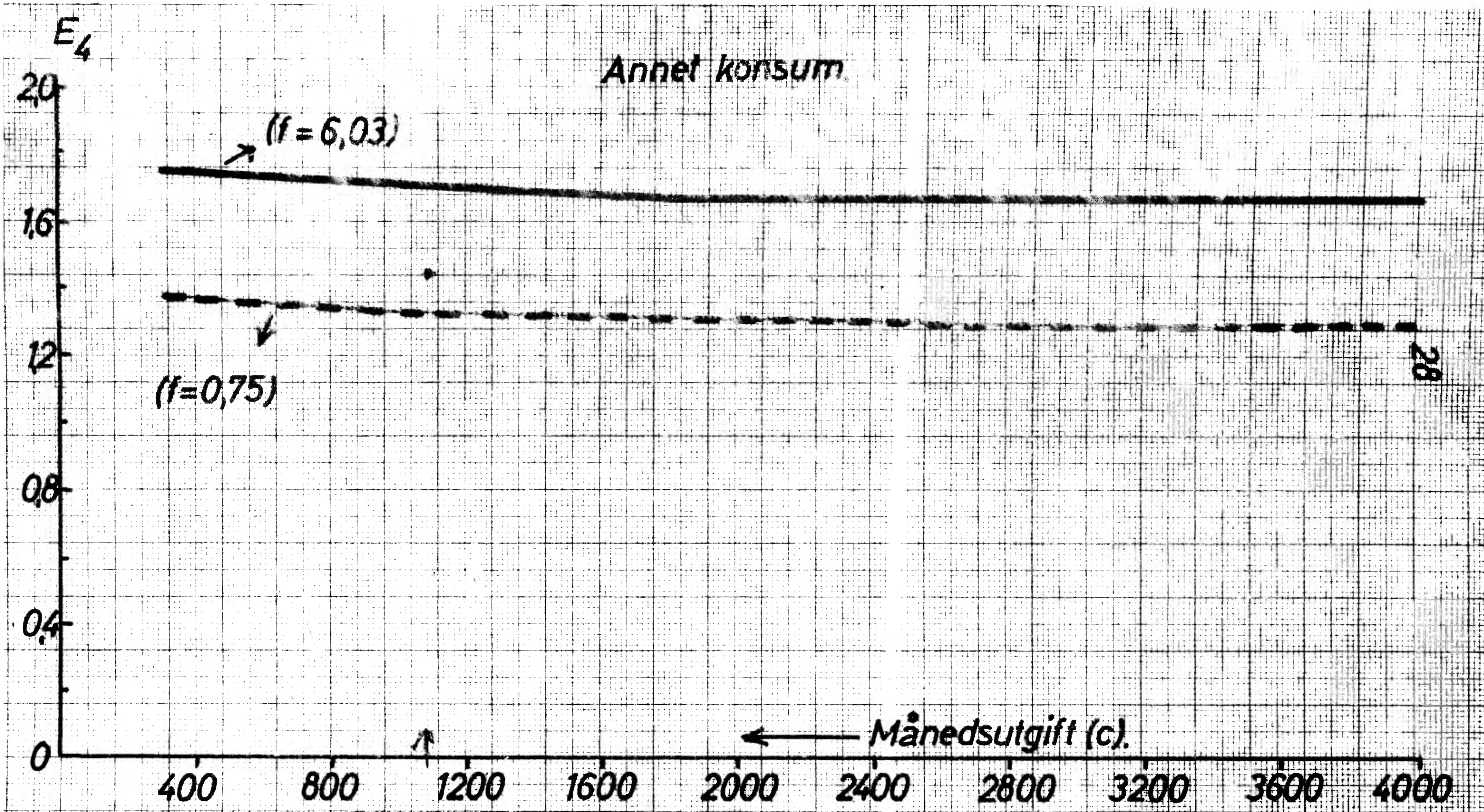

E_4
Annet konsum
(f = 6,03)
(f=0,75)
2,0
1,6
1,2
0,8
0,4
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Månedsutgift (c).

Fig.(5. 17).

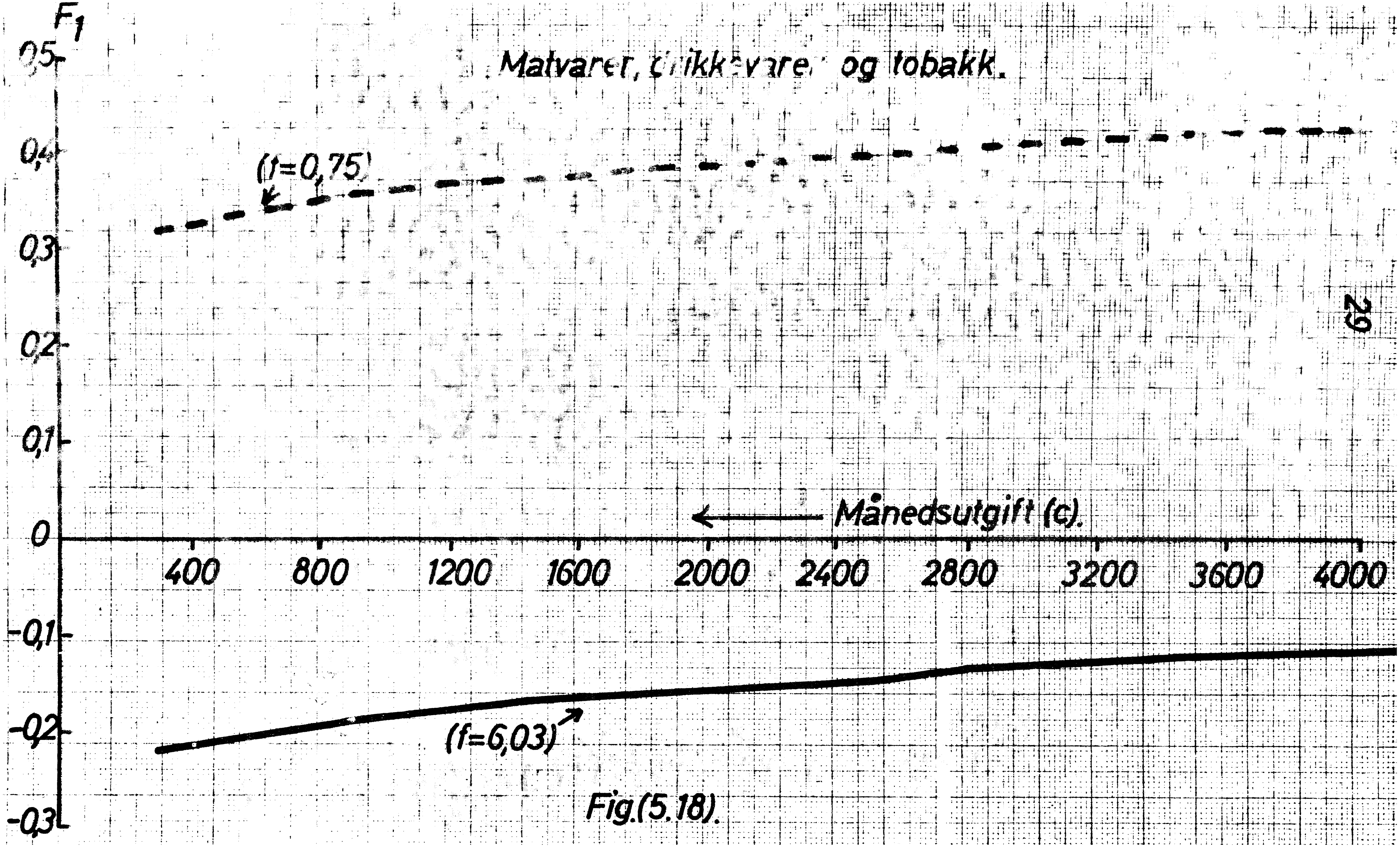
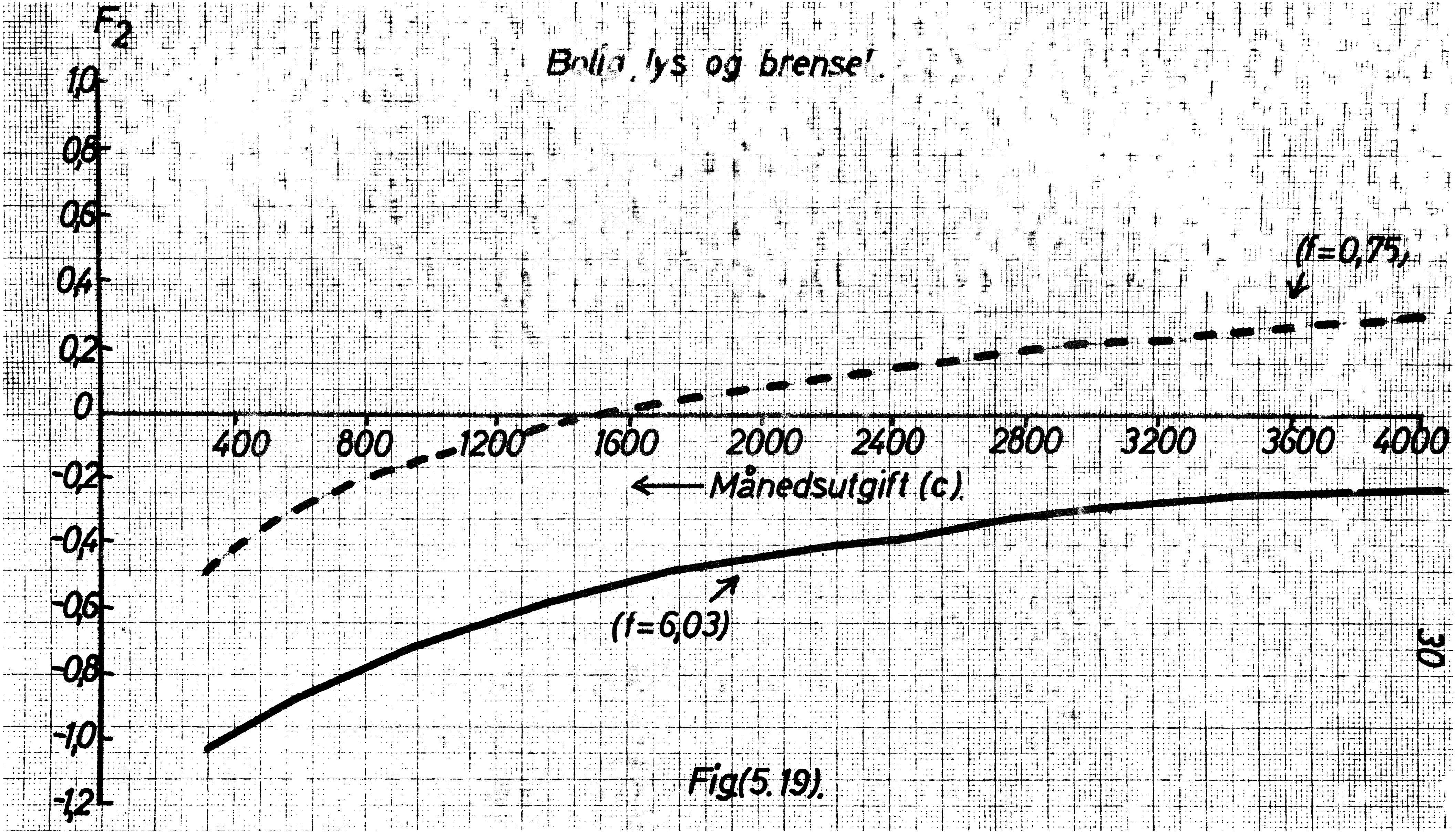
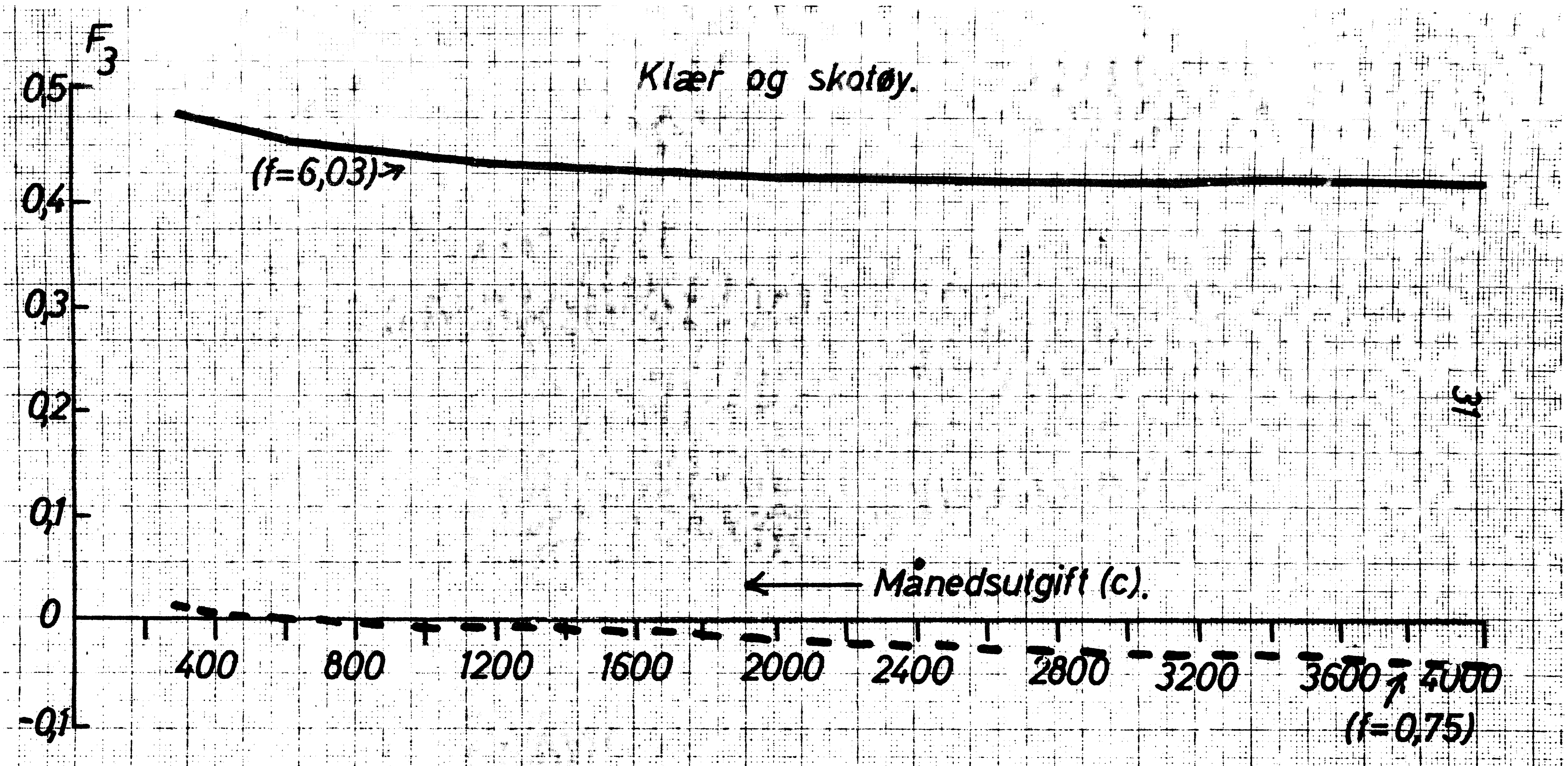

Matvarer, drikkevarer og tobakk.
F_1
0,5
0,4
0,3
0,2
0,1
0
-0,1
-0,2
-0,3
(f=0,75)
(f=6,03)
Månedsutgift (c).
400
800
1200
1600
2000
2400
2800
3200
3600
4000

Fig.(5.18).



F_2
Bolig, lys og brensel.
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
-1,2
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Månedsutgift (c)
(f=0,75)
(f=6,03)

Fig(5.19).



F_3
Klær og skotøy.
(f=6,03)
Månedsutgift (c).
(f=0,75)
0,5
0,4
0,3
0,2
0,1
0
-0,1
400
800
1200
1600
2000
2400
2800
3200
3600
4000

Fig.(5.20).

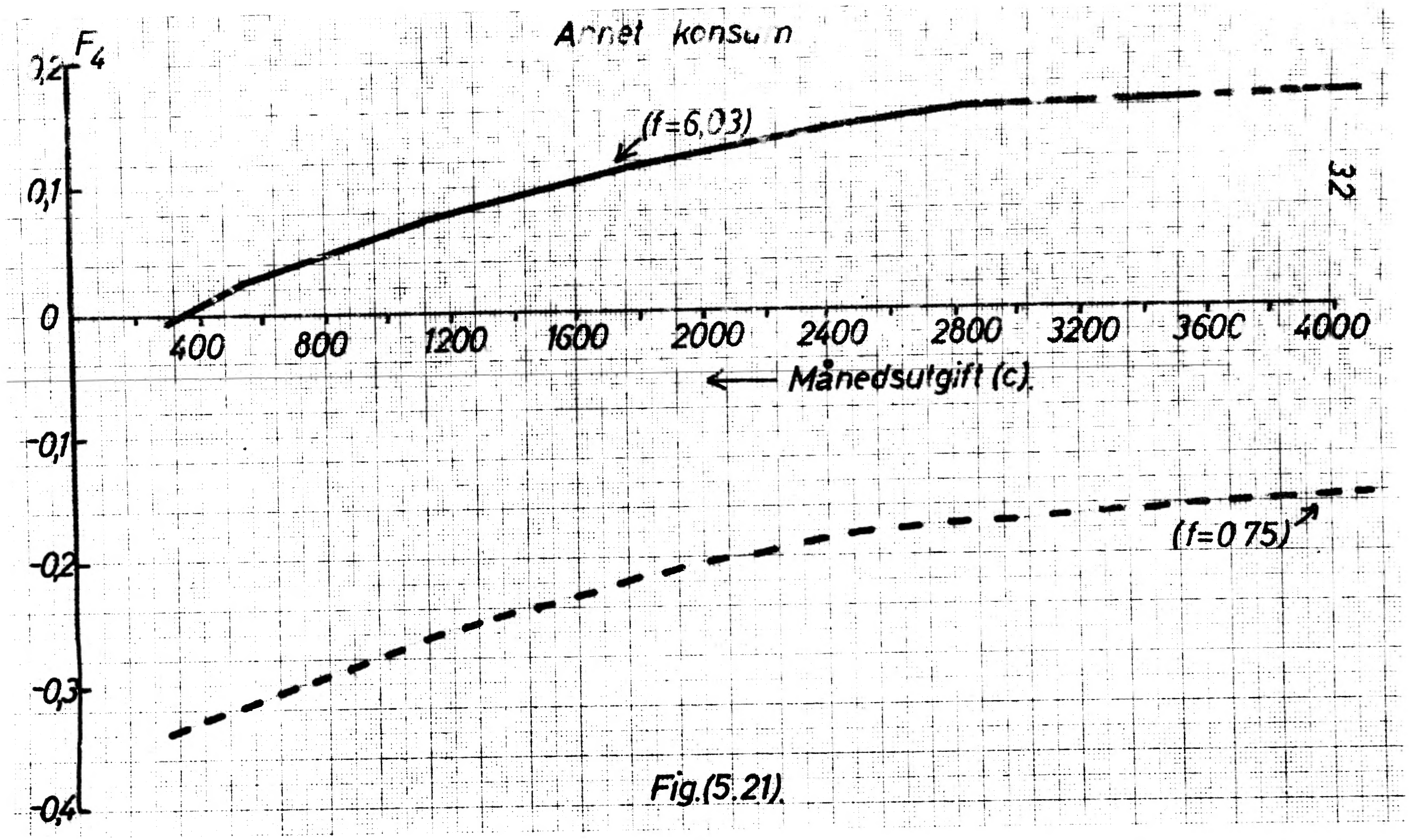

Annet konsum
F_4
0,2
0,1
0
-0,1
-0,2
-0,3
-0,4
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Månedsutgift (c)
(f=6,03)
(f=0 75)

Fig.(5.21).

A p p e n d i k s I. Om utgiftsfordelingen (jfr. avsnitt 4)

I. Innledning

I dette appendikset blir fordelingen for c studert nærmere (jfr. (4.1)).

I statistisk litteratur blir ofte inntektsfordelingen til en gruppe personer studert ved hjelp av teoretisk statistikk. En vil m.a.o. ha tak i den "teoretisk underliggende" inntektsfordeling.

Ofte blir Pareto-fordelingen, den log-normale fordeling eller andre mer eller mindre kompliserte fordelinger foreslått. (Se f.eks. P.R. Fisk: "The Graduation of Income Distributions" - Econometrica, Vol. 29, 2 (April 1961).)

Vi skal i det følgende undersøke <u>utgifts</u>fordelingen for 765 funksjonærer ifølge forbruksundersøkelsen 1958. Kaller vi:

r_j = inntekten for familie nr. j

c_j = utgift til forbruksvarer for familie nr. j og

s_j = "sparing" til familie nr. j,

så er s_j definert ved:

$$s_j = r_j - c_j \tag{1.1}$$

Vi vil altså studere c_j's fordelingsegenskaper. Dette er av spesiell interesse fordi c_j oftest blir brukt som forklaringsvariabel i Byråets modeller for estimering av Engelelastisiteter, istedenfor inntekten r_j.

II. Forutsetninger

Som arbeidshypotese går vi ut fra at c_j er log-normalt fordelt, dvs. at følgende frekvensfunksjon gjelder:

$$f(c) = \frac{1}{\sigma\,(2\pi)^{\frac{1}{2}}}\,\frac{1}{c}\,e^{-\frac{1}{2\sigma^2}(\ln C - \mu)^2} \tag{2.1}$$

Hvis $y = \ln C$, vil vi ha at $E(y) = \mu$ og $\mathrm{Var}\,(y) = \sigma^2$, og y er $N(\mu, \sigma)$.

Det er lett å vise at vi har:

$$\begin{aligned} &\underline{E\,(C) = C^{\mu + \frac{1}{2}\sigma^2}} \\ &\underline{\mathrm{Var}\,(C) = C^{2(\mu + \sigma^2)} \cdot \left[1 - C^{-\sigma^2}\right]} \end{aligned} \tag{2.2}$$

En kan spørre seg selv: "Er ikke denne arbeidshypotesen nokså vilkårlig? Hvorfor ikke velge f.eks. en Pareto-fordeling?" Imidlertid kan (2.1) begrunnes nokså stringent. En mer heuristisk begrunnelse er følgende:

Anta at en person har en utgift C_o på tidspunkt $t = 0$. Anta videre at han får et prosentvis tillegg på $100 \cdot r_1$ %o på tidspunkt 1, $100 \cdot r_2$ %o på tidspunkt 2 osv.

På tidspunkt n vil inntekten bli

$$(2.3) \quad C_t = C_o (1 + r_1)(1 + r_2) \cdots (1 + r_n)$$

eller

$$(2.4) \quad \log_c C_t = \log_c C_o + \sum_{j=1}^{n} \log_c (1 + r_j)$$

eller således

$$(2.5) \quad \log_c C_t = A + \sum_{j=1}^{n} X_j$$

Under visse forutsetninger er en sum av stokastiske variable tilnærmet normalfordelt, og p.g.a. (2.5) er derfor C_t tilnærmet log-normalt fordelt når dette gjelder.

III. Numeriske resultater, hypoteseprøving m.v.

Størrelsen C måles i kroner, og det er månedsutgiften vi betrakter. Vi splitter opp materialet i 10 utgiftsintervaller, og kan på grunnlag av data stille opp følgende tabell:

Tabell (3.1). Empirisk hyppighetsfordeling for C

Klasse nr. i	Kjennetegn, klasse	Observert absolutt hyppighet n_i	Relativ hyppighet h_i	Kumulert hyppighet H_i
1	$C < 250$	5	0,0065	0,0065
2	$250 \leqq C < 550$	57	0,0745	0,0810
3	$550 \leqq C < 850$	157	0,2052	0,2862
4	$850 \leqq C < 1150$	184	0,2405	0,5267
5	$1150 \leqq C < 1450$	136	0,1778	0,7045
6	$1450 \leqq C < 1750$	76	0,0993	0,8038
7	$1750 \leqq C < 2050$	60	0,0784	0,8822
8	$2050 \leqq C < 2350$	31	0,0405	0,9227
9	$2350 \leqq C < 2650$	20	0,0261	0,9488
10	$C \geqq 2650$	39	0,0510	1,0000
	Σ	765	1,0000	-

Som en ser av tabell (3.1) er fordelingen tydelig skjev, men er den log-normal ? For å undersøke dette estimerer vi parametrene μ og σ i (2.1) på grunnlag av data. Deretter vil vi teste log-normalitetshypotesen ved en Kjikvadrat-føyningstest.

Vi har at Maximum - Likelihood estimatorene for μ og σ blir

$$(3.2) \qquad \mu^* = \frac{1}{n}\sum_{j=1}^{n} \log_c C_j \qquad \sigma^{2*} = \frac{1}{n}\sum_{j=1}^{n} (\log C_j - \mu^*)^2$$

Vi får av data følgende estimater for μ og σ^2:

$$(3.3) \qquad \mu^* = 2{,}4163 \qquad \sigma^{2*} = 0{,}2778$$

Nå kan vi ved hjelp av (2.1) og (3.3) regne ut sannsynlighetene p_i^* for at observasjonene skal falle i de 10 intervaller i tabell (3.1), kolonne 2. Videre vil da (np_i^*) bli det "teoretiske motstykke" til n_i, hvor n = antall observasjoner (her: 765) og n_i er observert absolutt hyppighet i klasse nr. i.

Vi vet da at for stor n er størrelsen

$$(3.4) \qquad Z^o = \sum_{i=1}^{10} \frac{(n_i - np_i^*)^2}{np_i^*} \quad \text{tilnærmet.}$$

Kjikvadrat-fordelt med 7 frihetsgrader når nullhypotesen gjelder, dvs. når C er log-normalt fordelt. Resultatene av testen finner vi i tabell (3.5).

Tabell (3.5). X^2 - test

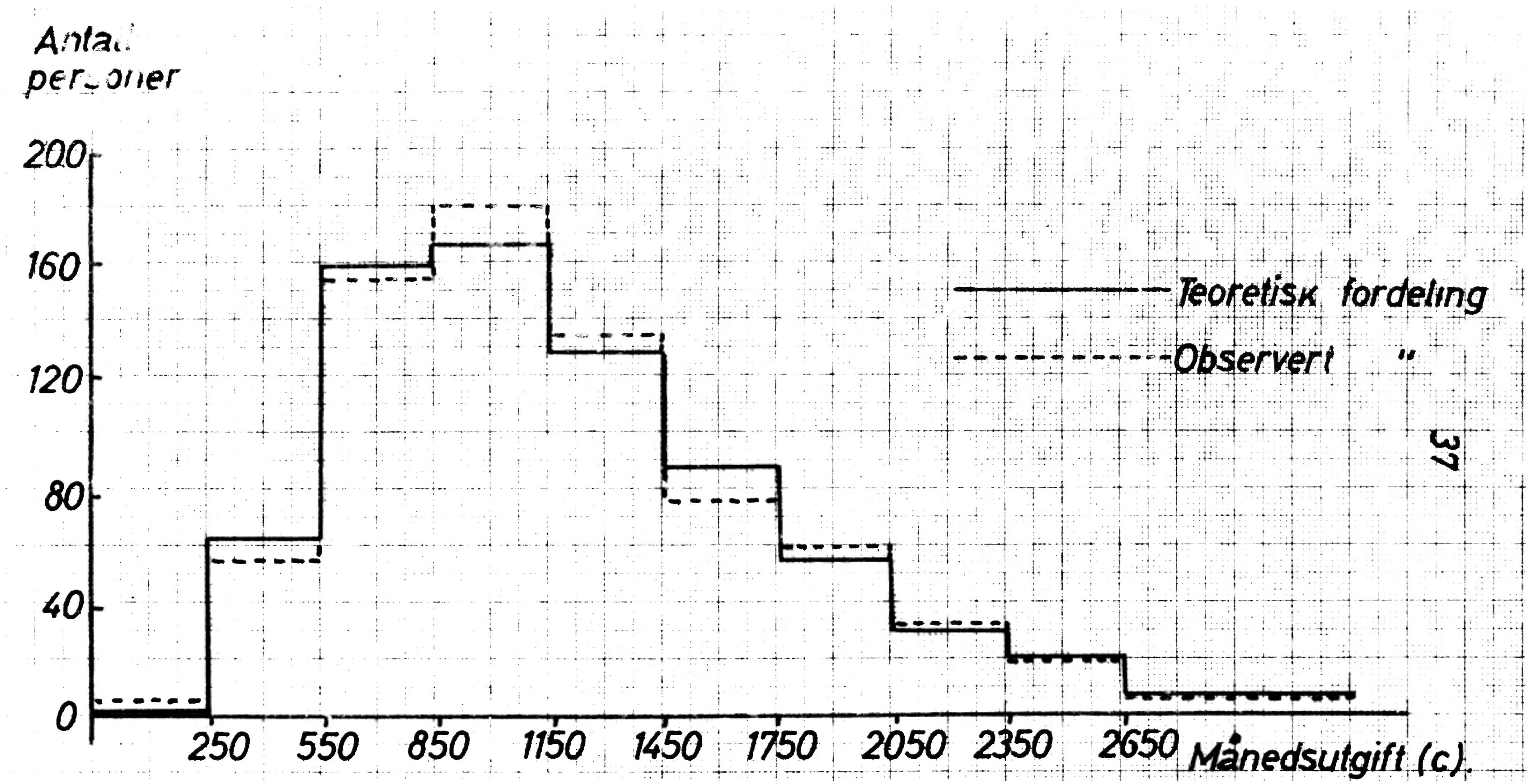
Klasse nr. i	p_i^* (1)	n_i^* (2)	np_i^* (3)	$n_i - np_i^*$ (4)	$\frac{(n_i - np_i^*)^2}{np_i^*}$ (5)
1	0,0022	5	1,683	+ 3,317	6,537
2	0,0863	57	66,020	- 9,020	1,232
3	0,2096	157	160,344	- 3,344	0,071
4	0,2218	184	169,677	+14,323	1,209
5	0,1680	136	128,520	+ 7,480	0,435
6	0,1144	76	87,516	-11,516	1,515
7	0,0726	60	55,539	+ 4,461	0,358
8	0,0458	31	35,037	- 4,037	0,465
9	0,0277	20	21,191	- 1,191	0,067
10	0,0516	39	39,474	- 0,474	0,006
Σ	1,0000	765	765	0	11,895

Summen i femte kolonne i tabell (3.5) er den observerte Kjikvadrat-størrelse (3.4). Vi har at 90, 95 og 99 prosentsfraktilen i $X^2(7)$ fordelingen er

$$X^2(7)_{90} = 12{,}02 \quad (3.6)$$
$$X^2(7)_{95} = 14{,}07$$
$$X^2(7)_{99} = 18{,}48$$

Siden summen $Z^o_{obs.} = 11{,}895$ (se tabell 3.5), kan vi ikke forkaste hypotesen om at C er log-normalfordelt selv med så stort testnivå som 10 $^o|_o$.

Hypotesen er mildt sagt brukbar. En grafisk framstilling av den observerte og den teoretiske fordeling finner en på side 37 i dette notatet.



Antal personer
200
160
120
80
40
0
250
550
850
1150
1450
1750
2050
2350
2650
Månedsutgift (c).
Teoretisk fordeling
Observert "