RAPPORTER 79/7

ESTIMERING AV INNTEKTSDERIVERTE PÅ TVERRSNITTSDATA MED MÅLEFEIL

AV
ODD SKARSTAD

STATISTISK SENTRALBYRÅ
OSLO

RAPPORTER FRA STATISTISK SENTRALBYRÅ 79/7

ESTIMERING AV INNTEKTSDERIVERTE PÅ TVERRSNITTSDATA MED MÅLEFEIL

AV
ODD SKARSTAD

OSLO 1979
ISBN 82-537-0976-5

FORORD

Statistisk Sentralbyrås Forbruksundersøkelser gir tverrsnittsdata over private husholdningers forbruk. Byrået gjennomfører også inntekts- og formuesundersøkelser, som delvis omfatter de samme husholdninger som deltar i forbruksundersøkelsene. De empiriske resultatene i denne rapporten bygger på det kombinerte materialet fra undersøkelsene i 1973.

Ved de fleste statistiske undersøkelser må en dessverre regne med feilkilder i datamaterialet. I så måte danner ikke forbruks- og inntekts- og formuesundersøkelsene noe unntak. Formålet er her å beskrive et opplegg for estimering av inntektsderiverte på datamateriale som er beheftet med målefeil av forskjellige slag.

Statistisk Sentralbyrå, Oslo, 27. juni 1979

Odd Aukrust

Sverre Hove

INNHOLD

Side

1. Innledning

Tverrsnittsdata over private husholdningers inntekt og forbruksutgifter blir ofte nyttet til å estimere inntektsderiverte og -elastisiteter (også kalt Engelelastisiteter). Formålet med dette arbeidet er å beskrive en estimeringsmetode som er spesielt beregnet på data som må antas å være beheftet med målefeil.

I mange analyser av inntektsderiverte mv. forutsettes det at det ikke er målefeil i datamaterialet, eller at målefeilene er relativt små eller av en slik art at de ikke influerer på estimeringsresultatene. Slike forutsetninger forenkler gjerne estimeringen. Vi mener imidlertid at det er høyst betenkelig å ignorere målefeilene i inntekts- og forbruksdataene under estimeringen.

Forbruksoppgavene er hentet fra forbruksundersøkelsen 1973. For samme år ble det gjennomført en inntekts- og formuesundersøkelse for det samme utvalget av husholdninger. På grunnlag av opplysninger om inntekt, fradrag mv. blir det definert et inntektsbegrep, som blir nyttet som indikator på disponibel inntekt for husholdningene. Den vanlige måten å definere privat sparing er som differansen mellom privat disponibel inntekt og privat konsum. Dette er nok også den vanligste måten å måle sparingen på. Som vi skal komme tilbake til, er det grunn til å tro at målefeil, særlig i inntektsdata, gjør at denne måten å måle sparingen på er betenkelig. Som et forhåpentlig bedre alternativ vil vi la husholdningens sparing være indikert ved formuesendringen i 1973, beregnet ut fra skattedata for 1972 og 1973.

Utgangspunktet for analysen er at man har inntekts-, forbruks- og spareoppgaver som alle inneholder målefeil. Vi skal drøfte og anvende en spesiell metode ved estimeringen for å oppnå konsistente estimatorer. Metoden, eller en nokså analog metode, er tidligere foreslått av Nissan Liviatan i en artikkel i Econometrica i 1961 (se [1]).

Vi blir nødt for å forutsette at såvel forbruks- som sparedataene bare inneholder tilfeldige målefeil, dvs. at feilene varierer tilfeldig (positivt eller negativt) fra husholdning til husholdning. For på forhånd å eliminere systematiske feil vil vi foreta såkalt "trimming" av forbruksdataene, i et visst omfang. Når det gjelder inntektsoppgavene, gjøres det derimot ingen særlig sterke forutsetninger om målefeilens karakter.

Under estimeringen nyttes den postulerte definisjonssammenheng at inntekt = forbruk + sparing. Denne betingelsen er i praksis oppfylt bare ved datamateriale uten målefeil. Estimeringen skjer under den bibetingelse at den marginale konsumtilbøyelighet for de forskjellige utgiftsgrupper og sparetilbøyeligheten tilsammen skal være lik 1. Ved å pålegge denne restriksjonen viser det seg at konsumtilbøyelighetene og sparetilbøyeligheten under visse betingelser blir identifiserbare.

Konsumfunksjonen som nyttes er en lineær utforming. En slik forenkling synes nødvendig på grunn av målefeilene. I kapittel 7 er muligheten for alternativ utforming kort vurdert.

Resultatene er gitt i kapittel 6. De må rimelig nok tas med visse forbehold. Dette skyldes både data og metode som er nyttet. Det er vanskelig å vurdere om forutsetningene om målefeilene er holdbare. Særlig gjelder vel dette data om sparingen, som er identisk med formuesøkningen. Formuestallene er i større eller mindre grad basert på ligningskontorenes skjønn. Dermed vil selvsagt også sparingen - formuesendringen - i betydelig grad være bygget på skattevesenets skjønn.

2. Data

2.1. Forbruk

Ved forbruksundersøkelsen 1973 ble det hentet inn oppgaver fra et tilfeldig utvalg på ca. 3 400 husholdninger. Hver husholdning førte regnskap over forbruket i to uker. Husholdningene ble tilfeldig fordelt over året i 26 delperioder, slik at 1/26 av utvalget førte regnskap i tidsrommet 1.1. - 14.1., neste 1/26 førte i perioden 15.1 - 28.1. osv. slik at hele året ble dekket[1]. Kjøp av

---

1) Ved siden av kjøp regnes varer fra egen produksjon og mottatte gaver med i forbruket.

visse større varige goder i løpet av en 12-månedersperiode ble registrert ved et intervju. Dette gjaldt bl.a. kjøp av privat bil, husholdningsmaskiner, fjernsynsmottaker mv.

Regnskapsføring gir som vi ser, informasjon om forbruksutgifter i en meget avgrenset periode av året. På grunn av sesongvariasjonene vil det gjennomsnittlige forbruket varierer fra periode til periode (for gitt årsinntekt). Som vi skal komme tilbake til vil vi søke å "ta hensyn til" sesongvariasjonene ved spesifikasjon av konsumfunksjonen.

2.2. Inntekt

Det ble gjennomført en inntekts- og forbruksundersøkelse i 1973 som bl.a. omfattet husholdningsutvalget i forbruksundersøkelsen av samme år. I undersøkelsen er følgende opplysninger registrert:

- ligningsdata (bl.a. bruttoinntekt, nettoinntekt, pensjonsgivende inntekt, fradrag, direkte skatter og avgifter)
- barnetrygd (beregnet ut fra barnetall i husholdningene)
- offentlig støtte etter lov om sosial omsorg

På grunnlag av data fra inntekts- og formuesundersøkelsen er det etablert et begrep som her blir kalt "disponibel inntekt". Dette omfatter begrepet "inntekt" i inntektsundersøkelsen 1973 tillagt fradragsposter på selvangivelsen som er definert som privat forbruk i forbruksundersøkelsene, eller som sparing. Nærmere om dette i kapittel 4.3.

2.3. Sparing

I sosialøkonomisk tankegang er det vanlig å oppfatte inntekten som den løpende tilgang på materielle verdier. Disse verdier kan enten forbrukes (konsumeres), eller de kan spares. Sparing blir følgelig definert som den delen av inntekten som ikke går til forbruk, altså

sparing = inntekt - forbruk.

De sparte midler (verdier) kan legges til formuen; sparingen i en periode kan derfra også defineres som

sparing = økning i formue (netto)

Fra skattebåndet har vi formuesdata for husholdningene både for 1972 og 1973. I kapittel 4.4. vil vi diskutere om og hvordan man i dette kan måle sparingen i husholdningene ved endringen i formuen.

3. Modell

3.1. Konsumfunksjon og variabelspesifikasjon

Utgangspunktet er den statiske konsumentteorien anvendt på tverrsnittsdata (husholdninger i 1973). Vi innfører følgende symboler:

X'_{ji}	betegner utgift til konsumgruppe j for husholdning nr. i	$(j = 1,2,\ldots, m)$ $(i = 1,2,\ldots, n)$
$X'_i = \sum_{j=1}^{m} X'_{ji}$	betegner total forbruksutgift for husholdning nr. i	$(i = 1,2,\ldots, n)$
Y'_i	betegner disponibel inntekt for husholdning nr. i	$(i = 1,2,\ldots, n)$
S'_i	betegner sparing (= formuesøkning) for husholdning nr. i	$(i = 1,2,\ldots, n)$

Symbolene ovenfor betegner de sanne størrelsene, dvs. de oppgavene en ville få dersom observasjonene ikke var beheftet med målefeil. Følgende sammenheng antas å gjelde:

$$Y'_i = X'_i + S'_i \qquad (i = 1,2,\ldots, n)$$

dvs. at inntekt = forbruk + sparing.

Inntekt og sparing (formuesendring) gjelder året 1973. Forbruket er relatert til forskjellig periode i 1973. Husholdningene ble fordelt på grupper som førte regnskaper mv. i forskjellige perioder i året.

Vi antar, som vanlig er, at forbruk og sparing først og fremst avhenger av inntekten. Videre antar vi at husholdningenes størrelse og sammensetning spiller en rolle. Vi skal nytte antall personer som en forklaringsvariabel, og skiller mellom voksne og barn.

Vi nytter bineærvariable for å fange opp nivåforskjeller i forbruket mellom kvartalene (sesongene).

Følgende symboler innføres i tillegg til de foran:

V_{1i} betegner antall voksne personer ($\geq$ 16 år) i husholdning nr. i $(i = 1,2,\ldots, n)$

V_{2i} betegner antall barn (< 16 år) i husholdning nr. i $(i = 1,2,\ldots, n)$

W_{1i} er lik 1 dersom husholdning nr. i har ført regnskap i 2. kvartal, null ellers $(i = 1,2,\ldots, n)$

W_{2i} betegner tilsvarende for 3. kvartal

W_{3i} betegner tilsvarende for 4. kvartal

(Referansegruppe er 1. kvartal)

U_{ji} betegner stokatiske restledd $(j = 1,2,\ldots, m)$ $(i = 1,2,\ldots, n)$

Variablene V_1, V_2, W_1, W_2 og W_3 antas å være observert uten målefeil. Konsumfunksjonene antar vi kan approksimeres til en lineær utforming:

(3.1.1.) $$X'_{ji} = \alpha_j + \beta_j Y'_i + \gamma_{ji}V_{1i} + \gamma_{j2} V_{2i} + \lambda_{j1} W_{1i} + \lambda_{j2} W_{2i} + \lambda_{j3} W_{3i} + U_{ji} \quad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

(3.1.2.) $$X'_i = \alpha + \beta Y_i' + \gamma_1 V_{1i} + \gamma_2 V_{2i} + \gamma_1 W_{1i} + \lambda_2 W_{2i} + \lambda_3 W_{3i} + U_i \quad (i = 1,2,\ldots, n)$$ (hvor α, β,

γ-ene og λ-ene betegner konstanter) og tilsvarende for sparefunksjonen:

(3.1.3.) $$S'_i = \alpha_s + \beta_s Y'_i + \gamma_{s1} V_{1i} + \gamma_{S2} V_{2i} + \lambda_{s1} W_{1i} + \lambda_{S2} W_{2i} + \lambda_{S3} W_{3i} + U_{Si} \quad (i = 1,2,\ldots, n)$$

Relasjonen (3.1.1.) beskriver altså hvordan utgiften til den enkelte vare- og tjenestegruppe (konsumgruppe) avhenger av inntekten, antall voksne og barn og av årstid for regnskapsføringen. (3.1.2.) og (3.1.3.) viser tilsvarende for henholdsvis totalforbruket og sparingen.

Vi skal ikke her søke å gi noen særlig begrunnelse for å velge lineær funksjon. Vår viktigste "begrunnelse" for å anta en lineær utforming er at våre inntekts-, forbruks- og sparedata er beheftet med målefeil. Ved slike målefeil vil en mer komplisert funksjonsutforming forringe estimeringsmulighetene radikalt. Dette er behandlet nærmere i kapittel 7.

3.2. Om restledd mv.

Vi forutsetter at restleddene har forventning null:

$$E\,(U_{ji}) = 0 \qquad (j = 1,2,\ldots,\, m)$$

$$E\,(U_i) = 0 \qquad (i = 1,2,\ldots,\, n)$$

$$E\,(U_{Si}) = 0$$

for alle verdier på de uavhengige variablene. Videre forutsettes at

$$\text{Kovar}\,(U_{ki},\, U_{hl}) = 0 \quad \text{for alle } i \neq l, \quad (k,\, h = 1,2,\ldots,\, m,\, S)$$

dvs. at restleddene i funksjonene til forskjellige husholdninger ikke er korrelert med hverandre. Dette synes å være en nokså rimelig forutsetning tatt i betraktning den "tilfeldighet" som ligger bak trekkingen av husholdningene via Byråets utvalgsplan.

Derimot er det rimelig å tro at

$$\text{Kovar}\,(U_{ki},\, U_{hi}) \neq 0 \quad (k,\, h = 1,2,\ldots,\, m,\, S)$$

dvs. at restleddene i de forskjellige funksjonene fra <u>en og samme</u> husholdning kan være korrelert med hverandre. Definisjonsmessig gjelder

$$\sum_{j=1}^{m} U_{ji} = U_i$$

og $\qquad (i = 1,2,\ldots,\, n)$

$$U_i + U_{si} = o$$

Vi skal også forutsette at restleddene har konstant varians. Det er imidlertid verdt å merke at konsistensegenskapene til de estimatorene vi skal nytte ikke er avhengige av denne forutsetningen.

Det gjelder også at

$$\sum_{j=1}^{m} \beta_j = \beta$$

og

$$\sum_{j=1}^{m} \beta_j + \beta_s = \beta + \beta_s = 1{,}0$$

dvs. at enhver inntektsøkning enten blir konsumert eller spart.

4. Målefeil i materialet. Trimming av data

4.1. Generelt

Det kan tenkes mange typer målefeil og mange måter å karakterisere målefeil på. Vi vil her forutsette en todeling av feilene, i heholdsvis systematiske og tilfeldige feil. Denne todelingen er også nyttet av Malinvaud (se [2], Ch. 10). Den samlede målefeilen antas å bestå av en systematisk og en tilfeldig komponent. For å forklare begrepene nærmere innfører vi følgende symboler:

Z betegner observert variabelverdi
Z' betegner sann variabelverdi
Z'' betegner samlet målefeil
'Z' betegner tilfeldig målefeil

Vi forutsetter at den observerte verdien kan uttrykkes som en eller annen eksakt funksjon (her symbolisert ved g (·)) av den sanne verdien og eventuelt andre variable, og av den tilfeldige målefeilen

$$Z = g\ (Z',\ \text{eventuelt andre variable}) + 'Z'$$

Her representerer g (·) - funksjonen den systematiske målefeilen - som gjerne antas å være ikkestokatisk - og 'Z' det stokatiske elementet i målefeilen. Det gjelder at

$$Z = Z' + Z''$$

og

$$Z = g(Z') + 'Z'$$

hvor

$$E('Z') = o$$

Den samlede målefeilen kan uttrykkes som summen av systematiske og tilfeldige feil:

$$Z'' = [g(Z') - Z'] + 'Z'$$

Vi mener denne betraktningsmåten kan være hensiktsmessig. Den gir en viss presisering av begrepet målefeil, samtidig som kanskje de fleste typer av målefeil i praksis kan omfattes av modellen.

Det er velkjent at målefeil i et datamateriale kan umiliggjøre identifikasjon av strukturparametrene i en relasjon. Vi skal her kort omtale problemet.

Vi antar en relasjon mellom to variable, Z'_1 og Z'_2:

$$Z'_{1i} = a + b\ Z'_{2i} + \varepsilon_i \qquad (i = 1,2,\dots,\ n)$$

hvor a og b betegner konstanter og ε_i et restledd med forventning

$$E(\varepsilon_i) = o \qquad (i = 1,2,\dots,\ n)$$

Vi setter inn de observerte i stedet for de sanne variabelverdiene og skriver relasjonen:

$$Z_{1i} = a + b\ Z_{2i} + (\varepsilon_i + Z''_{1i} - b\ Z''_{2i}) \quad (i = 1,2,\dots,\ n)$$

Det er en generell betingelse på leddet $(\varepsilon_i + X''_{1i} - b\ Z''_{2i})$ som må være oppfylt for at parametrene a og b skal være identifiserbare. Leddet må ha forventning null for alle verdier på den uavhengige variablen Z_{2i}:

$$E[(\varepsilon_i + Z''_{1i} - b\,Z''_{2i}) \mid Z_{2i}] = 0$$

Forventningen vil normalt ikke være null dersom det er systematiske målefeil enten i Z_1 eller Z_2. En får heller ikke identifikasjon selv om det bare er tilfeldige feil, dersom feilene er i forklaringsvariablen Z_2. I praksis er det bare feilleddet Z''_1, som under visse forutsetninger ikke ødelegger mulighetene for identifikasjon, nemlig hvis Z''_1 har samme generelle egenskaper som restleddet ε i f.eks. konsumfunksjonen. (Dette er nærmere behandlet av E. Malinvaud, se [2], Ch. 18.)

Det som her er sagt om identifikasjonsmulighetene gjelder dersom man ikke har noen tilleggsinformasjon/-beskrankninger utenom de som her er forutsatt. Vi skal senere vise at slik tilleggsinformasjon kan være til stor nytte når det gjelder å få identifikasjon og å finne konsistente estimatorer.

I dette kapittel skal vi ta for oss datamaterialet nokså detaljert og vurdere kvaliteten på det. Vi skal også gjennomføre såkalt "trimming" av data, dvs. at vi "justerer" materialet for å eliminere visse typer målefeil før materialet nyttes til estimering. Som det vil gå fram av kapittel 5 er det særlig viktig at det ikke er systematiske målefeil i de avhengige variablene, forbruk og sparing, med den estimeringsmetoden vi skal nytte.

4.2. Forbruksutgifter

a. Systematiske målefeil

Det kan være grunn til å tro at husholdningene gjennomsnittlig ikke får med alle utgiftspostene ved forbruksundersøkelser. Det kan være en av årsakene til at forbruksundersøkelsens forbrukstall ligger noe under nasjonalregnskapets tall (regnet pr. husholdning eller pr. person). Men det er heller ikke så lett å si hvor pålitelig nasjonalregnskapets tall er. For én varegruppe er det imidlertid sikkert at forbruksundersøkelsen klart underestimerer forbruket, nemlig for hovedgruppen drikkevarer og tobakk. For denne varegruppen opptrer det systematiske målefeil i forbruksundersøkelsen i mye større grad enn for de andre vare- og tjenestegruppene.

Vi så i avsnittet foran at systematiske målefeil normalt hindrer at det oppnås identifikasjon av parametrene. Dette gjelder som vi så også den avhengige variablen. Det kan derfor være ønskelig å søke å "eliminere" den systematiske feilen på én eller annen måte før dataene nyttes. Dette omtales gjerne som "trimming" av data.

Det er ofte vanskelig å ta stilling til konkrete forslag til hvordan trimmingen skal gjennomføres. Den antatte systematiske målefeilen i varegruppen drikkevarer og tobakk antar vi vesentlig skyldes to forhold:

1. Husholdningene som deltar i undersøkelsen fører gjennomgående opp mindre utgifter enn de har hatt særlig til drikkevarer. Dette kan skyldes at husholdningene prøver å vise omverdenen et "fornuftigere" eller "bedre" forbruk enn sitt sanne. Det kan også tenkes at de enkelte medlemmene i husholdningen ønsker å holde sitt personlige forbruk skjult overfor hverandre og derfor unnlater å føre utgiftene (selv om alle over 15 år fikk tildelt eget regnskapshefte, og som de ikke behøvde å vise fram til andre)
2. Deltakelse i forbruksundersøkelsen er frivillig; vi regner med at frafallet i undersøkelsen er særlig stort blant storforbrukere av alkoholholdige drikkevarer, noe som bidrar til at estimert gjennomsnittsforbruk blir lavt.

Det synes vanskelig å "korrigere" dataene for den feilkomponent som er nevnt under pkt. 2. Derimot skal vi prøve å trimme data for å "rette opp" feil som skyldes forhold nevnt i pkt. 1.

Vi forutsetter at alle de systematiske målefeilene skyldes en systematisk underrapportering av utgifter. Videre antar vi at nasjonalregnskapets budsjettandel for drikkevarer og tobakk (av totalt privat konsum) er den "riktige".

Vi spesifiserer nå oppgaven som å korrigere de individuelle husholdningenes oppgitte forbruksutgifter - på en eller annen måte - slik at budsjettandelene for alle husholdningene i undersøkelsen sett under ett tilsvarer nasjonalregnskapets budsjettandel for drikkevarer og tobakk.

Problemet er å avgjøre hvordan korreksjonen bør gjennomføres. Vi har valgt å bygge på følgende antakelse: Det er nærliggende å tro at underrapporteringen er avhengig av nivået på forbruket. Hvis f.eks. en familie har høye utgifter og mange utgiftsposter til drikkevarer og tobakk, vil vi tro at - absolutt sett - mange poster blir utelatt under bokføringen. Vi vil her forutsette at den systematiske underrapporteringen er proporsjonal med utgiftsbeløpet. Vi korrigerer for dette ved å blåse opp utgiftsbeløpene for alle husholdningene med en konstant faktor slik at utgifter til drikkevarer og tobakk i gjennomsnitt for husholdningene blir den samme som i nasjonalregnskapet for 1973. Vi har på forhånd korrigert nasjonalregnskapets tall på visse poster (først og fremst utgifter til helsepleie) som er definert annerledes enn i forbruksundersøkelsen. Vi finner da følgende tall (pr. person):

	Forbruksundersøkelsen 1973	Nasjonalregnskapet 1973[1)]
Drikkevarer og tobakk	kr 630	kr 1 155
Total forbruksutgift	" 12 789	" 15 024

Med utgangspunkt i disse tallene korrigerer vi forbruksundersøkelsens gjennomsnittlige utgifter til drikkevarer og tobakk til et beløp 630 + T, hvor T er bestemt slik at

$$\frac{630 + T}{12\ 789 + T} = \frac{1\ 155}{15\ 024}$$

Dette gir T = 383 og 630 + T = 1 013. Utgiftsbeløpene til drikkevarer og tobakk blir altså blåst opp med en faktor

$$1\ 013/630 = 1{,}61.$$

Med dette vil vi gå ut fra at vi har renset forbruksutgiftstallene for de systematiske målefeilene.

b. Tilfeldige målefeil

Før vi definerer tilfeldige målefeil, er det naturlig å si hva vi prøver å måle. Vi ønsker egentlig forbrukstall for de deltakende husholdningene for hele undersøkelsesperioden, som er året 1973 (som vi også har inntektsoppgaver for). Spørsmålet er nå: hva slags oppgaver får vi når husholdningene fører regnskap bare i 14 dager? (Vi ser bort fra sesongsvingninger som vi "tar hensyn til" ved utformingen av konsumfunksjonen.) Vi må regne med visse tilfeldige variasjoner i utgiftene over tiden. Disse variasjoneer skyldes ikke svingninger i inntekten over året; de skyldes bare tilfeldige variasjoner mellom 2-ukersperiodene. Ve betrakter disse som tilfeldige målefeil.

Stort sett har det vel vært vanlig å anta at tilpassingsmekanismen kan tenkes å foregå på to prinsippielt forskjellige måter (jfr. f.eks. R. Summers (1959), se [5]):

- forbrukerne fastlegger sin totale forbruksutgift når de har en gitt inntekt til disposisjon. Innenfor denne rammen fordeler de så midlene mellom kjøp av de forskjellige vare- og tjenestegruppene.
- forbrukerne fastlegger forbruket av de forskjellige vare- og tjenestegrippene ut fra en gitt inntekt. Totalutgiftene blir så summen av de forskjellige vare- og tjenestegruppene.

Det er ikke lett å finne holdepunkter for hva som er den typiske konsumentatferden. Det kan selvsagt [også] tenkes kombinasjoner av de to måtene, f.eks. at husholdningen først setter av mer eller mindre på forhånd bestemte beløp til "nødvendige" og "faste" utgifter, f.eks. matvare- og boligutgifter, og at det som så blir igjen av inntekten (etter evt. sparing) deretter fordeles mellom de øvrige vare- og tjenestegruppene.

---

1) NOS A 969.

Det er vel imidlertid nærliggende at husholdningen må ha en total ramme for forbruksutgiftene in mente når den foretar de forskjellige anskaffelser, eller foretar handlinger som medfører betalingsforpliktelser. Dette gjelder når man betrakter et tidsrom av en viss lengde. Og nettopp tidshorisonten er vesentlig ved studium av forbruksatferden. På svært kort sikt setter ikke inntekten noen streng grense for forbruket, f.eks. som følge av evt. tidligere oppsparte midler eller opptak av lån.

Dersom man antar at husholdningene først fastlegger totalutgiften og deretter fordeler midlene mellom de forskjellige vare- og tjenestegruppene, burde det stort sett være en viss negativ korrelasjon mellom de tilfeldige avvikene for de forskjellige gruppene: Dette er tidligere forsøkt undersøkt (jfr. [3]). Resultatene der tyder ikke på negative korrelasjoner mellom leddene. Legg merke til den begrensning som ligger i utsagnet; man finner ikke negative korrelasjoner når utgiftene er registrert som i forbruksundersøkelsene, med blant annet en kort regnskapsperiode. Dette sier imidleritd ikke så mye om hvordan husholdningens innkjøpsatferd kan tenkes å arte seg betraktet over et lengre tidsrom. Det kunne her ha vært relevant å skille mellom "forbruksteori" og "kjøpsteori".

Det er neppe aktuelt å forsøke å korrigere forbruksdataene for de tilfeldige målefeilene. Det er vanskelig å se hvordan dette eventuelt skulle gjøres. Med den korreksjon som er foretatt for å rette opp de systematiske målefeilene i konsumgruppen drikkevarer og tobakk, antar vi nå at forbruksdataene er renset for systematiske målefeil. Med utgangspunkt i omtalen av identifikasjonsproblemet i avsnitt 4.1. er det rimelig å konkludere med at forbruksdataene er "brukbare" i relasjoner hvor forbruksutgifter inngår som avhengig (venstresidig) variabel. Derimot vil den tilfeldige målefeilen normalt hindre identifikasjon av parametrene i en relasjon hvor variablen nyttes som forklaringsvariabel (høyresidig variabel).

4.3. Inntekt

Vårt inntektsbegrep "disponibel inntekt" omfatter begrepet "inntekt" i inntektsundersøkelsen 1973 tillagt de fradragsposter (på selvangivelsen) som er definert som privat forbruk i forbruksundersøkelsene eller som sparing. Begrepet består litt grovt sagt av bruttoinntekt minus direkte skatter og renteutgifter - unntatt renter på boliglån - pluss beregnet barnetrygd og bidrag etter lov om sosial omsorg. Gjennomsnittlig disponibel inntekt for alle husholdningene i utvalget er kr 37 600.

Begrepet omfatter på langt nær alle inntektskategorier. Skattefrie ytelser fra folketrygden er ikke med. Det er heller ikke stipend i forbindelse med utdanning, arbeidsløshetsstønad, skattefrie sykepenger eller skattefri lønn under sykdom, bostøtte fra Husbanken foruten en del andre ytelser. Dessuten vil sikkert materialet være preget av inntektsunndragelser i ikke ubetydelig grad.

Poenget er at en må regne med betydelige feil i inntektsoppgavene. For det første vil det trolig være en systematisk underregistrering av inntekten, dvs. en systematisk komponent av ett eller annet slag. Videre må en regne med store individuelle variasjoner i målefeilen, altså også store tilfeldige målefeil. Vi vet følgelig at bruk av dette inntektsbegrepet i en relasjon generelt ikke vil gi identifikasjon av parametrene ved estimeringen. Det blir et hovedpunkt i kapittel 5 å behandle identifikasjons- og konsistensproblemene ved bruk av inntektsdata med målefeil som forklaringsvariabel.

4.4. Sparing

Som nevnt i kapittel 2.3. er det vanlig å definere sparing som differansen mellom inntekt og forbruk. Etter at vi har "trimmet" forbruksdataene for systematiske målefeil i varegruppen drikkevarer og tobakk, har vi antatt at forbruksdataene ikke lenger er beheftet med generende systematiske feil. Derimot er det systematiske feil i inntektsdataene. Det er da åpenbart at bruk av differansen mellom observert inntekt og forbruk som indikator på sparing medfører både tilfeldige og systematiske feil i indikatorvariablen. Som sagt vil systematiske målefeil i den avhengige variable (her sparingen) medføre at parametrene i sparefunksjonen ikke blir identifiserbare.

Alternativt til å anslå sparingen som differansen mellom inntekt og konsum vil vi undersøke mulighetene for å måle sparingen "direkte" - som økningen i formuen. Sparingen defineres - foreløpig forsøksvis - som økningen i netto formue fra inngangen til utgangen av året 1973.

Det er et problem at inntekts- og formuesundersøkelsen fra 1973 bare inneholder formuestall for utgangen av året. Undersøkelsen gir altså ikke data om formuesøkningen i løpet av året.

Statistisk Sentralbyrå får hvert år inn visse summariske inntekts-, formues- og skatteopplysninger fra ligningsetater via det såkalte Skattebåndet. På dette båndet finnes bl.a. netto formue for inntektstakerne, også inntektstakerne i de husholdningene som deltok i forbruksundersøkelsen og i inntekts- og formuesundersøkelsen 1973. I mangel av noe bedre kan det være nærliggende på én eller annen måte å nytte formuesendringen fra utgangen av 1972 til utgangen av 1973 som indikerer på sparingen i husholdningene i 1973. Tilføyelsen "på en eller annen måte" indikerer at det kanskje må "gjøres noe" med formuesendringstallene før de nyttes.

Vi skal imidlertid først kommentere formuesbegrepet, før vi seinere ser på formuesendringen, dvs. sparingen.

a. Formuesbegrepet

Formuesansettelser er generelt et problem. Det er selvsagt ikke mulig å si noe generelt om hva som er et rimelig eller "riktig" prinsipp å verdsette formuen etter. Det mest rimelige ville kanskje vært å nytte markedsverdien, eller en anslått markedsverdi, definert på én eller annen måte. Vi har imidlertid ingen slike opplysninger. Det er forøvrig antakelig også vanskelig å skaffe opplysninger om den sannsynlige markedsverdi på husholdningenes formue, først og fremst på realformuen.

En teoretisk mulighet ville vært å intervjue husholdningene om hva de tror er markedsverdien på deres bolig og annen realformue. Det er vel imidlertid høyst usikkert hvilke opplysninger man ville få inn på den måten. Problemet er ikke bare at de innhentede opplysningene ville være usikre. Det må man regne med at data innhentet på annen måte også kan være. Ulempen er særlig at det ville bli svært vanskelig å vurdere egenskapene ved materialet; er det f.eks. sannsynlig at husholdningene selv ville verdsette sin egen formue spesielt høyt eller spesielt lavt? Ett poeng er at formuen kunne komme til å bli vurdert nokså forskjellig, avhengig av de enkelte husholdningenes innsikt i prisnivå mv.

Av praktiske grunner må vi her i stor grad bygge på skattedata.

Det er allment antatt at skattevesenet verdsetter realformue "lavt". I forbruksundersøkelsen 1973, som omfatter de samme husholdningene som inntekts- og formuesundersøkelsen, ble husholdninger med prosentlignede selveide boliger bedt om å oppgi både branntakst og skattetakst/ligningsverdi på boligen. Det viser seg at nivået på branntaksten gjennomsnittlig lå over tre ganger så høyt som skattetakst/formuesverdi. Allerede et slikt resultat antyder følgende: Det er åpenbart svært avgjørende for de beregningsresultatene vi kommer fram til hvilke prinsipper vi legger til grunn når vi indikerer nivået på realformuen.

Et forhold av betydning for beregningen av en husholdnings formue er eierforholdet til boligen. Av alle husholdningene i undersøkelsen var det i følge oppgavene 63 prosent som bodde i selveid bolig, og av disse igjen 78 prosent som bodde i prosentlignet bolig. For alle med prosentlignet selveid bolig kan vi ved formuesberegningen bytte ut skattetakst/ ligningsverdi med branntakster. Dette fører til en økning i samlet netto formue for denne gruppen fra kr 96 000 til kr 178 000, dvs. en økning på ca. 85 prosent. For alle husholdninger sett under ett bevirker dette en økning fra kr 104 000 til kr 155 100, eller 49 prosent.

For husholdninger som bor i leide boliger inneholder forbruksundersøkelsen 1973 opplysninger om hva som ble betalt i eventuelt innskott/andel/aksje/obligasjon da boligen ble kjøpt. Dette sier imidlertid svært lite om hva boligen "egentlig er verdt" på undersøkelsestidspunktet.

Blant annen fast eiendom i private husholdninger kan særlig nevnes fritidshus (hytte, landsted). Det var i følge undersøkelsen ca. 18 prosent av husholdningene som hadde fritidshus. Vi har imidlertid ikke opplysninger som gjør det mulig å korrigere ligningsverdien på fritidshus på noen som helst måte.

b. Sparing

I det foregående er det antydet hvordan man kan beregne formuestall. Vi er imidlertid primært ute etter tall for sparingen, altså formuesendringen i løpet av året. På det omtalte årlige skattebåndet fra skattedirketoratet er det visse summariske opplysninger fra ligningen, bl.a. nettoformue. Vi har hentet fram disse formuesopplysningene fra husholdningene som deltok i inntekts- og formuestellingen 1973 og forbruksundersøkelsen av samme år. I alt var det 3 363 husholdninger som deltok i nevnte forbruksundersøkelse. Av disse var det 3 271 som ble "funnet igjen" i inntekts- og formuesundersøkelsen og 3 071 på skattebåndene for både 1972 og 1973. Denne "uoverensstemmelsen" skyldes særlig at nye skatteobjekter er kommet til, mens andre er falt fra i løpet av 1973.

Skattebåndene er imidlertid mangelfulle på et viktig punkt: Formuestall foreligger bare for husholdninger med så høy formue at de betaler formuesskatt. Blant de 3 071 husholdningene er det bare for 1 038 vi har formuesopplysninger i 1973, og i 905 tilfelle vi har opplysninger både for 1972 og 1973. Det er derfor bare for disse 905 husholdninger i materialet det har noen særlig mening å prøve å beregne sparingen med utgangspunkt i formuesendringen.

For de 905 husholdningene finner vi følgende tall for inntekt, forbruk og formuesøkning:

	Gjennomsnitt pr. husholdning (905 husholdninger)
Observert inntekt ($\bar{y}$)	44 700
Observert forbruksutgift ($\bar{x}$)	42 400
Observert forbruksøkning = observert sparing ($\bar{S}$) ...	8 200

Vi har allerede forkastet å nytte y-x som indikatorer på sparingen, p.g.a. antatt systematiske målefeil i inntekten (y). Vi ser her at y-x i gjennomsnitt ville gi $\overline{y-x}$ = 2 300, et tall som forekommer noe lavt.

Også formuesøkningen (S) vil selvsagt være beheftet med målefeil. Håpet er at den systematiske delen av feilen ikke er verre enn at variablen kan nyttes om avhengig variable (spareindikator). I kapittel 6.1. er parametrene i relasjonen for disse 905 husholdningene estimert.

c. Generert sparing

Det er ønskelig å kunne estimere parametre i relasjonen hvor alle 3 071 husholdninger er med. Det er imidlertid problematisk å finne en variabel som kan nyttes som spareindikator for de øvrige (3 071 - 905 = 2 166) husholdningene.

En tenkelig utvei er først å estimere parametrene i en relasjon mellom sparing (s) og (y-x) ved hjelp av data fra de 905 husholdningene. Deretter genereres formuesdata (s) for de øvrige 2 166 husholdningene. Vi skal, delvis som en illustrasjon, gjennomføre dette.

Det er selvsagt et problem hvordan funksjonssammenhenger skal utformes og hvilke variable som skal inngå i en modell hvor sparingen "forklares".

Den mest nærliggende variable å trekke inn er vel yrkesstatus, og da særlig for hovedinntektstakeren i husholdningen. En kan kanskje regne med at sammenhengen mellom s og (y-x) delvis er annerledes blant selvstendig næringsdrivende enn blant f.eks. lønnstakere. Dette kan tenkes å ha sin bakgrunn i f.eks. forskjellig målefeil i inntekten (y) for selvstendige og for andre.

Et annet forhold av betydning for en antatt sammenheng mellom (y-x) og s berører selve vurderingen av formuen. På grunn av problemer med verdsetting av fast eiendom vil vi la eierforholdet til boligen inngå som en variabel, og skiller mellom selveid prosentlignet, selveid regnskapslignet og ikke selveid bolig.

Vi definerer følgende variable og symboler:

L_{1i} = 1 hvis hovedinntektstakeren i husholdningen er selvstendig næringsdrivende, null ellers

L_{2i} = 1 hvis hovedinntektstakeren er lønnstaker, null ellers
(husholdninger med hovedinntektstaker ikke yrkesaktiv er referansegruppe)

M_{1i} = 1 hvis husholdningen bor i selveid prosentlignet bolig, null ellers

M_{2i} = 1 hvis husholdningen bor i selveid regnskapslignet bolig, null ellers
(ikke selveid bolig er referansegruppe)

Vi danner relasjonen:

$$S_i = k_0 + k_1 (Y_i - X_i) + k_2 L_{1i} + k_3 L_{2i} + k_4 M_{1i} + k_5 M_{2i} + \text{restledd} \qquad (i = 1,2,\ldots, 905)$$

hvor k_0, k_1, ... k_5 betegner kontanter. Restleddet antas å ha forventning null for alle verdier på de eksogene variablene. Parametrene estimeres ved minste kvadraters metode.

Det kan selvsagt reises en rekke spørsmål ved denne fremgangsmåten. Ett av disse gjelder målefeilen i Y-X. Lar det seg gjøre å nytte relasjonen som en slags prediktor for S når det er målefeil i Y-X? Er ikke dette et tilfelle som er nokså analogt med et som frarådes foran ved estimering av parametrene i konsumfunksjonene?

Det er kjent at "konsistensegenskapene" til estimatoren for bl.a. k_1 påvirkes av nevnte målefeil. Dette hindrer likevel ikke (nødvendigvis) at relasjonen kan nyttes til å simulere brukbare sparetall. Jfr. at en relasjon med estimerte parametre kan være nyttig til prediksjon selv om parametrene er nokså forventningsskjeve.

Det er imidlertid vanskelig å si sikkert i hvilken grad dette har lyktes her.

Estimeringen ga følgende estimater og standardavvik (avrundede tall):

Parameter	Estimat	Standardavvik
k_o	$\hat{k}_o$ = 16 371	
k_1	$\hat{k}_1$ = 0,08719	(0,078)
k_2	$\hat{k}_2$ = 3 678	(5 680)
k_3	$\hat{k}_3$ = - 2 357	(5 201)
k_4	$\hat{k}_4$ = -11 149	(5 711)
k_5	$\hat{k}_5$ = - 8 659	(6 892)

Ved hjelp av estimatene på parametrene er det så generert sparedata for de øvrige 2 166 husholdningene. Multippel korrelasjon er 0,07902. Det er grunn til å bemerke den lave korrelasjonen, dvs. at svært lite blir forklart ved relasjonen. (Vi ser også at standardavvikene på $\hat{k}$-ene er store).

Den lave korrelasjonen behøver ikke bety at relasjonen fører til dårlig genererte spareoppgaver. Store tilfeldige målefeil i sparetallene fører vel til lav korrelasjon, uten at dette ødelegger konsistensegenskapene på estimatoren for sparetilbøyeligheten.

Ved hjelp av estimatene er det generert sparedata (S_i) for de øvrige 2 166 husholdningene

$$S_i = \hat{k}_0 + \hat{k}_1 (Y_i - X_i) + \hat{k}_2 L_{1i} + \hat{k}_3 L_{2i} + \hat{k}_4 M_{1i} + \hat{k}_5 M_{2i} \qquad (i = 906, 907,\ldots, 3\ 071)$$

Det er alltid vanskelig å vurdere om generering av et datamateriale har vært vellykket, dvs. om de genererte observasjonene ligger noenlunde i nærheten av det vi ville mene med de "sanne" observasjonene. Vi kan heller aldri få særlig sikre indikasjoner på hvor "godt" det genererte datamaterialet er. Vi skal imidlertid se litt på noen gjennomsnittstall for derved å prøve å vurdere materialet.

For alle 3 071 husholdningene får vi følgende gjennomsnittstall for inntekt, forbruk, sparing (= generert formuesøkning):

	Alle hus
Inntekt $(\bar{y})$	40 300
Forbruk $(\bar{x})$	40 500
Sparing $(\bar{s})$	8 900

Vi får altså en gjennomsnittlig sparing/formuesøkning på ca. 9 000 kroner pr. år. Dette skal i prinsippet også omfatte (netto) økning i realformue.

Det kan også ha interesse å se på inntekts-, forbruks- og sparetall for enkelte grupper. Vi betrakter nå tall for selvstendig næringsdrivende, lønnstakere og ikke yrkesaktive hver for seg:

	Selvstendige	Lønnstakere	Ikke yrkesaktive
Inntekt	39 200	45 000	30 300
Forbruk	40 900	45 500	28 800
Sparing	11 400	7 900	9 700

Her går det fram at tallene viser størst sparing blant selvstendige. Dette er vel ikke altfor uventet ut fra hva man vanligvis antar om de forskjellige sosialgruppers konsum- og spareforhold. Videre ser vi at spareandelen, dvs. i prosent av forbruk + sparing, er enda større blant ikke yrkesaktive. Mens altså sparingen gjennomsnittlig er lavest blant lønnstakere, både absolutt og relativt.

Følgende oppstilling er en inndeling i forskjellige aldersgrupper for hovedinntektstakeren:

	Under 40 år	40 - 66 år	67 år og over
	Kr	Kr	Kr
Inntekt	39 700	43 800	29 400
Forbruk	43 500	43 500	24 600
Sparing	11 100	7 500	9 500

Vi ser at sparingen er størst i husholdninger hvor hovedinntektstakeren er under 40 år. Den gjennomsnittlige sparetilbøyelighet, regnet i prosent av summen av forbruk + sparing, er imidlertid størst blant de eldste.

Det er ikke lett å vurdere rimeligheten av sparetallene, dvs. hvor "vellykket" genereringen har vært.

I kapittel 6.2. er det gitt emgiriske resultater fra estimering av parametrene i konsum- og sparefunksjonen for de 3 071 husholdningene.

5. Estimeringsmetode

5.1. Innledning

Vi innfører følgende symboler:

X_{ji} betegner observert utgift til konsumgruppe j for husholdning nr. i $(j = 1,2,\dots, m)$ $(i = 1,2,\dots, n)$

$X_i = \sum_{j=1}^{m} X_{ji}$ betegner observert total forbruksutgift for husholdning nr. i $(i = 1,2,\dots, n)$

Y_i betegner observert inntekt for husholdning nr. i $(i = 1,2,\dots, n)$

S_i betegner observert formuesøkning for husholdning nr. i $(i = 1,2,\dots, n)$

Målefeilene er definert som differansen mellom observert og sann variabelverdi:

$$X''_{ji} = X_{ji} - X'_{ji} \qquad (j = 1,2,\ldots, m)$$

$$X''_i = X_i - X'_i$$

$$Y''_i = Y_i - Y'_i \qquad (i = 1,2,\ldots, n)$$

$$S''_i = S_i - S'_i$$

Ifølge forutsetningene i kapittel 4 er forbruks- og sparedataene beheftet utelukkende med tilfeldige målefeil.

Forventningen til disse målefeilene antas lik null, dvs. at

$$E(X_{ji}) = E(X'_{ji}) \qquad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

$$E(X_i) = E(X'_i) \qquad (i = 1,2,\ldots, n)$$

$$E(S_i) = E(S'_i) \qquad (i = 1,2,\ldots, n)$$

Inntektsdataene antas å være beheftet både med systematiske og tilfeldige feil,

$$Y_i = g(Y'_i) + {}'Y'_i \qquad (i = 1,2,\ldots,n)$$

Her gjelder at

$$Y''_i = (g(Y'_i) - Y'_i) + {}'Y'_i \quad (i = 1,2,\ldots, n)$$

Vi skal i dette kapittel behandle en del problemer i forbindelse med estimering ved målefeil. For å unngå at framstillingen blir unødig tungvint vil vi først gjennomføre våre ressonnementer på en forenklet versjon av konsum-/sparefunksjonene. Vi vil derfor foreløpig anta at inntekten er eneste forklaringsvariabel til konsum og sparing (sanne verdier):

(5.1.1.) $$X'_{ji} = \alpha_j + \beta_j Y'_i + U_{ji} \qquad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

(5.1.2.) $$X'_i = \alpha + \beta Y'_i + U_i \qquad (i = 1,2,\ldots, n)$$

(5.1.3.) $$S'_i = \alpha_s + \beta_s Y'_i + U_{si} \qquad (i = 1,2,\ldots, n)$$

hvor det antas at forutsetningene (om restledd mv.) i kapittel 3.2. fortsatt gjelder. Siden vi har at

$$U_i + U_{si} = 0$$

$$Y'_i = X'_i + S'_i$$

$$\beta + \beta_s = 1$$

følger det av denne forenklede versjonen også at

$$\alpha + \alpha_S = 0$$

Hovedspørsmålet er nå følgende: Hvordan kan vi på beste måte utnytte de observerte størrelsene (X-ene, Y og S) for å finne konsistente estimater for parametrene i konsum- og sparerelasjonene, dvs. α-ene og β-ene i den forenklede funksjonsversjonen - i den mer generelle versjonen (i kapittel 3) også γ-ene og λ-ene?

De observerbare variablene gir i prinsippet minst tre tenkelige alternativer for valg av indikatorvariable ved estimeringen av parametrene i relasjonene ($i = 1,2,\ldots, n$):

a) Forbruksutgift representeres ved X_i
Sparing representeres ved S_i
Inntekt representeres ved Y_i

b) Forbruksutgift representeres ved X_i
Sparing representeres ved S_i
Inntekt representeres ved $X_i + S_i$

c) Forbruksutgift representeres ved X_i
Sparing representeres ved $Y_i - X_i$
Inntekt representeres ved Y_i

Vi skal starte med å utelukke alternativ c). Dette er det eneste av disse alternativene som innebærer at en avhengig variabel i følge forutsetningene inneholder systematiske målefeil - nemlig sparingen, som ved dette alternativ indikeres ved (y-x). Og systematiske målefeil i en avhengig variabel synes å være særlig besværlig (jfr. kapittel 4.1. foran).

Alle alternativene medfører at det er målefeil i den høyresidige variable (systematiske eller tilfeldige målefeil). Og vi vet at parametre i en relasjon generelt ikke er identifiserbare dersom det overhodet er målefeil i noen av de høyresidige variable. Dette gjelder imidlertid under forutsetning av at man ikke har tilleggsinformasjon som eventuelt kan gi identifikasjon. Identifikasjon- og estimeringsmuligheter er behandlet i de følgende avsnittene.

5.2. Inntekten som forklaringsvariabel

I dette og neste avsnitt blir alternativ a) behandlet. Vi nytter de forenklede funksjonsformene i avsnittet foran (5.1.1.) - (5.1.3.) og etablerer tilsvarende sammenhenger mellom de observerte størrelsene:

(5.2.1.) $$X_{ji} = a_j + b_jY_i + \text{restledd} \qquad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

(5.2.2.) $$X_i = a + bY_i + \text{restledd} \qquad (i = 1,2,\ldots, m)$$

(5.2.3.) $$S_i = a_s + b_s Y_i + \text{restledd} \qquad (i = 1,2,\ldots, n)$$

hvor a-ene og b-ene betegner konstanter. Bruk av vanlige minste kvadraters metode gir følgende estimatorer for b-ene[1]:

1) M betegner sentralmomentene.

(5.2.4.) $$\hat{b}_j = \frac{M_{(x_j)(y)}}{M^2_{(y)}} \qquad (j = 1,2,\ldots, m)$$

(5.2.5.) $$\hat{b} = \frac{M_{(x)(y)}}{M^2_{(y)}}$$

(5.2.6.) $$\hat{b}_s = \frac{M_{(s)(y)}}{M^2_{(y)}}$$

Poenget er her å undersøke om $\hat{b}$-ene kan brukes, på én eller annen måte, til å finne konsistente estimatorer på β-ene. Vi erstatter de observerte variabelverdiene for forbruk og sparing i tellerne med de sanne verdiene og målefeil, og skriver (når det ikke er systematiske feil i X og S):

(5.2.7.) $$\hat{b}_j = \frac{M_{(x_j')(y)} + M_{('x_j')(y)}}{M^2_{(y)}} \qquad (j = 1,2,\ldots, m)$$

(5.2.8.) $$\hat{b} = \frac{M_{(x')(y)} + M_{('x')(y)}}{M^2_{(y)}}$$

(5.2.9.) $$\hat{b}_s = \frac{M_{(s')(y)} + M_{('s')(y)}}{M^2_{(y)}}$$

Vi skal betrakte de asymtotiske verdiene på tellerne, og studere først det siste leddet i hver teller.

Leddet $M_{('x_j')(y)}$ i (5.2.7.) er altså sentralmomentet for målefeilen i forbruket av varegruppe j og observert inntekt. Vi vil anta at dette konvergerer mot null ved økende antall observasjoner. Vi bygger denne antakelsen på den rutinen som ligger bak innsamlingen av henholdsvis forbruks- og inntektsdataene. Tilsvarende gjelder totalforbruket, $M_{('x')(y)}$ (i 5.2.8.). Vi skal gjøre tilsvarende for sparingen, dvs. at $M_{('s')(y)}$ (i 5.2.9.) også konvergerer mot null. Dette siste er nok en noe mer tvilsom forutsetning, tatt i betraktning at både inntekts- og spareopplysningene bygger på ligningsdata. Siden det er samme informasjonskilde til både inntekts- og spareopplysningene, kan det lett hende at målefeilene i de to variablene er korrelert med hverandre og her skal det skytes inn at feilestimering av sparetilbøyeligheten også fører til feilestimering av konsumtilbøyeligheten - med den estimeringsmetode vi her skal benytte.

Det første leddet i hver av tellerne kan splittes opp i to ledd:

$$M_{(x_j')(y)} = \beta_j M_{(y')(y)} + M_{(u_j)(y)} \qquad (j = 1,2,\ldots, m)$$

$$M_{(x')(y)} = \beta\, M_{(y')(y)} + M_{(u)(y)}$$

$$M_{(s')(y)} = \beta_s\, M_{(y')(y)} + M_{(u_s)(y)}$$

Leddet $M_{(u_j)(y)}$ vil vi anta går asymtotisk mot null. Det synes i hvertfall ikke å være noen åpenbare grunner til hvorfor eller hvordan den skulle kunne tenkes å avvike fra null. Tilsvarende gjelder $M_{(u)(y)}$ og $M_{(u_s)(y)}$.

Vi innfører

$$p \lim_{n \to \infty} M = m$$

og

$$p \lim_{n \to \infty} \frac{M_{(y')(y)}}{M^2_{(y)}} = \frac{m_{(y')(y)}}{m^2_{(y)}} = k$$

og skriver

(5.2.10) $\quad p \lim \hat{b}_j = k\,\beta_j \qquad (j = 1,2,\ldots, m)$

(5.2.11) $\quad p \lim \hat{b} = k\,\beta$

(5.2.12) $\quad p \lim \hat{b}_s = k\,\beta_s$

Det går av dette fram at $\hat{b}$-ene generelt ikke vil være konsistente estimatorer for β-ene. (Dersom det nå ikke hadde vært målefeil i inntekten ($y = y'$), ville selvsagt $\hat{b}$-ene ha vært konsistente, i det k i så fall er lik 1,0.) Derimot vil følgende estimatorer være konsistente:

(5.2.13.) $\quad \hat{\hat{b}}_j = \hat{b}_j / (\hat{b} + \hat{b}_s) \qquad (j = 1,2,\ldots, m)$

(5.2.14.) $\quad \hat{\hat{b}} = \hat{b} / (\hat{b} + \hat{b}_s)$

(5.2.15.) $\quad \hat{\hat{b}}_s = \hat{b}_s / (\hat{b} + \hat{b}_s)$

fordi vi ser at

$$p \lim \hat{\hat{b}}_j = \beta_j \qquad (j = 1,2,\ldots, m)$$

$$p \lim \hat{b} = \beta$$

$$p \lim \hat{b}_s = \beta_s$$

Vi forstår av (5.2.10.) - (5.2.12.) at den vanlige minste kvadratersestimatoren ($\hat{b}$) gir forventningsskjeve estimatorer, men at forventningsskjevheten relativt sett er den samme for alle vare- og tjenestegrupper og for sparingen. Ved å bygge på den tilleggsbetingelsen at summen av konsum- og sparetilbøyeligheten skal være lik 1,0, har vi tilstrekkelig informasjon til å finne (estimere) den felles skjevheten. Dermed blir parametrene (β-ene) identifiserbare, og det kan finnes konsistente estimatorer for dem. Vi har nettopp vist at $\hat{\hat{b}}$-ene er konsistente under de forutsetninger vi har spesifisert.

Estimatorene er analoge med estimatorer som ble foreslått av N. Liviatan i en artikkel i Econometrica i 1961 (se [1]). Hans måte å ressonnere omkring problemet med målefeil var imidlertid basert på den såkalte instrumentvariabelmetoden, og han betraktet estimatoren $\hat{\hat{b}}_j$ som en instrumentvariabelestimator. Videre gjaldt hans framstilling bare estimering av utgiftsderiverte, som med våre symboler tilsvarer $\hat{\hat{b}}_j = b_j/b$. I vedlegg 1 er Liviatans opplegg kort gjengitt.

Det viktige resultatet er at det lar seg gjøre å finne konsistente estimatorer på de inntektsderiverte selv med målefeil i inntektsdataene. Det spiller ingen rolle om målefeilene er systematiske eller tilfeldige, eller evt. begge deler. Kravet er begrenset til at målefeil i forbruk og sparing skal være ukorrelert med observert inntekt, og restleddene i konsum- og sparefunksjonen heller ikke korrelert med den samme inntekten. Det er også en forutsetning at målefeilene i de avhengige variablene, forbruk og sparing, ikke må være systematiske.

Det kan lett finnes konsistente estimatorer for δ-ene. Dette kan ikke gjøres ved den "vanlige" måten,

$$\hat{a}_j = \overline{X}_j - \hat{\hat{b}}_j \overline{Y} \qquad (j = 1,2,\ldots, m)$$

fordi det er antatt å være systematiske feil i $\overline{Y}$. En må derfor nytte estimatorene

$$\hat{\hat{a}}_j = \overline{X}_j - \hat{\hat{b}}_j(\overline{X} + \overline{S}) \qquad (j = 1,2,\ldots, m)$$

og

$$\hat{\hat{a}} = \overline{X} - \hat{\hat{b}}(\overline{X} + \overline{S})$$

$$\hat{\hat{a}}_s = \overline{S} - \hat{\hat{b}}_s(\overline{X} + \overline{S})$$

som vil være konsistente under de spesifiserte forutsetningene.

Estimeringen av parametrene foregår på en måte i to etapper. Først estimeres β-ene ved $\hat{\hat{b}}$-ene, hvor en nytter den observerte inntekten som en forklaringsvariabel. Deretter estimeres konstantleddet ved bruk av totalt forbruk + sparing som "forklaringsvariabel".

I vedlegg 2 er standardavvik på $\hat{\hat{b}}$-ene beregnet tilnærmet. Et interessant resultat der er at det kan lønne seg å nytte $\hat{\hat{b}}$-ene som estimatorer fremfor de vanlige minstekvadratersestimatorene ($\hat{b}$-ene) selv om det ikke er målefeil i inntekten, når det er tilfeldig målefeil i forbruk og sparing. Dette synes å ha vært ukjent for Liviatan i hans artikkel i Econometrica. Hans angrepsmåte førte til at $\hat{\hat{b}}$-ene ble betraktet som instrumentvariabelestimatorer. Og for det vanlige tilfellet av instrumentvariabelmetoden gjelder det at det aldri lønner seg å nytte en instrumentvariabel dersom det ikke er målefeil i forklaringsvariabelen. (Den instrumentvariabelen som gir minst varians er forklaringsvariabelen selv.) I vedlegg 2 er det vist at dette ikke generelt gjelder $\hat{\hat{b}}$-ene. Dette tyder for øvrig kanskje på at $\hat{\hat{b}}$-ene ikke bør betraktes som instrumentvariabelstimatorer.

5.3. Estimering på den generelle modellen

Vi tar igjen utgangspunkt i funksjonene som ble spesifisert i kapittel 3.1. (dvs. (3.1.1.) - (3.1.3.). En antar altså funksjoner med også andre høyresidige variable enn inntekten. Det er her viktig å merke at disse andre variablene forutsettes observert uten målefeil.

Vi setter inn observert forbruk, inntekt og sparing i funksjonene i stedet for de sanne verdiene og estimerer ved minste kvadraters metode, og finner størrelsene tilsvarende $\hat{b}_j$, $\hat{b}$ og $\hat{b}_s$ foran. Heller ikke i dette mer generelle tilfellet kan disse uttrykkene tas om konsistente estimatorer for h.h.v. β_j, β og β_s i funksjonene. Videre fører målefeilen i inntekten, den ene av forklaringsvariablene, til at minste kvadraters metode heller ikke gir konsistente estimatorer for konstantene foran de andre høyresidevariablene.

Også her vil imidlertid uttrykkene $\hat{\hat{b}}_j = \hat{b}_j/(\hat{b} + \hat{b}_s)$, $\hat{\hat{b}} = \hat{b}/(\hat{b} + \hat{b}_s)$ og $\hat{\hat{b}}_s = \hat{b}_s/(\hat{b} + \hat{b}_s)$ være konsistente som estimatorer for henholdsvis β_j, β og β_s. Dette er lett å innse ved å studere momentmatrisene. Forholdet blir forøvrig påvist av Liviatan når det gjelder <u>utgifts</u>deriverte i hans artikkel fra 1961 (se [1]).

Konstantleddet og de andre parametrene kan nå lett estimeres ved å danne følgende relasjoner:

$$X_{ji} - \hat{\hat{b}}_j(X_i + S_i) = \alpha_j + \gamma_{j1}V_{1i} + \gamma_{j2}V_{2i} + \lambda_{j1}W_{1i} + \lambda_{j2}W_{2i} + \lambda_{j3}W_{si} + U_{ji}$$

$$(j = 1,2,\ldots, m)$$

$$(i = 1,2,\ldots, n)$$

$$X_i - \hat{\hat{b}}(X_i + S_i) = \alpha + \gamma_1V_{1i} + \gamma_2V_{2i} + \lambda_1W_i + \lambda_2W_{2i} + \lambda_3W_{3i} + U_i$$

$$S_i - \hat{\hat{b}}_s(X_i + S_i) = \alpha_s + \gamma_{s1}W_{1i} + \gamma_{s2}V_{2i} + \lambda_{s1}W_{1i} + \lambda_{s2}W_{2i} + \lambda_{s3}W_{3i} + U_{si}$$

For å slippe å få noen høyresidig variabel med målefeil har vi altså flyttet inntektsleddet over på venstre side. Og for å eliminere systematisk målefeil i den venstresidige variable, har vi byttet ut Y med (X + S). Sidet det er forutsatt ikke å være målefeil i noen av variablene som nå står igjen på høyre side, vil vanlig minste kvadraters metode gi konsistente estimatorer på disse parametrene (i kapittel 6. symbolisert ved $\hat{\alpha}_j$, $\hat{\gamma}_{j1}$, $\hat{\gamma}_{j2}$, $\hat{\lambda}_{j1}$, $\hat{\lambda}_{j2}$, $\hat{\lambda}_{j3}$).

<u>5.4. Summen av observert forbruk og sparing som inntektsindikator</u>

Vi tar utgangspunkt i alternativ b) i kapittel 5.1. og de forenklede funksjonsutformingene (5.1.1.) - (5.1.3.), og setter inn de observerte størrelsene og skriver:

$$X_{ji} = a_j + b_j(X_i + S_i) + \text{restledd} \qquad (j = 1,2,\ldots, m)$$

$$(i = 1,2,\ldots, u)$$

$$X_i = a + b(X_i + S_i) + \text{restledd}$$

$$S_i = a_s + b_s(X_i + S_i) + \text{restledd}$$

hvor observert inntekt er byttet ut med forbruk + sparing. Minste kvadraters metode gir uttrykkene

$$b_j^* = \frac{M_{(x_j)(x+s)}}{M^2_{(x+s)}} \qquad (j = 1,2,\ldots, m)$$

$$b^* \quad \frac{M_{(x)(x+s)}}{M^2_{(x+s)}}$$

$$b_s^* \quad \frac{M_{(s)(x+s)}}{M^2_{(x+s)}}$$

Det første vi kan påpeke er at der er lett å vise at følgende "riktige" sammenhenger mellom estimatorene gjelder

$$\sum_{j=1}^{m} b_j^* = b^*$$

og

$$b^* + b_s^* = 1,$$

altså at summen av de inntektsderiverte er lik 1.

Dette sikrer imidlertid naturligvis ikke at estimatorene er "gode". Tvert imot, vi har allerede i kapittel 5.1. med henvisning til den generelle teorien om estimering påpekt at parametrene β_j, β og β_s ikke vil være identifiserbare, medmindre vi har nødvendig tilleggsinformasjon. Det er imidlertid vanskelig å innse hvilken tilleggsinformasjon dette skulle være. Vi kan ikke - nokså analogt med det vi gjorde i kapittel 5.2. - pålegge den "ekstrabetingelse" at summen av estimatorene for de inntektsderiverte skal være lik 1. Estimeringsmetoden har allerede denne betingelsen innebygget ($b^* + b_s^* = 1$), slik at dette ikke vil være noen ekstrabetingelse

---

Denne estimeringsmetoden er analog med å nytte totalt observert forbruk som forklaringsvariabel ved estimering av utgiftsderiverte for de forskjellige vare- og tjenestegrupper. Dette gir som nevnt ikke identifikasjon, noe som er fremhevet av bl.a. R Summers i en artikkel i Econometrica i 1959 (se [5]). Problemet er også behandlet i arbeidsnotat IO 77/44, se [3].

6. Resultater

6.1. Sparing målt ved formuesendring

Vi skal her gi empiriske resultater fra estimering av parametrene i konsum- og sparefunksjonen for 905 husholdninger. Dette er de 905 husholdningene vi har observert formuesendringstall for (jfr. kapittel 4.4. b)). Standardavvikene på $\hat{b}_j$ er ikke regnet ut etter variansformelen i vedlegg 2. Dette fordi regresjonsprogrammet ikke er lagt opp med dette for øye. For å gi en slags antydning av størrelsesordnenen på variansen, oppgis her standardavviket på $\hat{b}_j$.

Standardavvikene på $\hat{\gamma}_{j1}$, $\hat{\gamma}_{j2}$, $\hat{\lambda}_{j1}$, $\hat{\lambda}_{j2}$ og $\hat{\lambda}_{j3}$ er kommet fram ved vanlige minste kvadratersestimeringen. Ovennevnte program gir ikke estimater på standardavvik på konstantledd, og slike estimater er følgelig ikke beregnet.

Det første som springer i øynene er vel at den marginale konsumtilbøyeligheten er lavere enn sparetilbøyeligheten, henholdsvis ca. 0,48 og 0,52. Dette er det grunn til å kommentere.

Man regner vel med at den marginale konsumtilbøyelighet vanligvis ikke er særlig lavere enn ca. 0,7. Hva kan så årsaken(e) være til at man får et slikt resultat?

Det første som kan nevnes er tolkningen av inntektsderiverte som er beregnet på denne måten. Modellen bygger på at all inntekt er likeverdig, dvs. at den fritt kan disponeres enten til konsum eller sparing. Det skal altså ikke spille noen rolle for konsumtilbøyeligheten i hvilken form inntekten er, om den er i form av løpende lønnsinntekt eller f.eks. ligningsvesenets oppvurdering av fast eiendom. Det kan rimeligvis reises innvendinger mot et slikt utgangspunkt. Det er ikke uvanlig å skille mellom forskjellige typer inntekt når konsum- og sparetilbøyeligheten skal estimeres. Det er imidlertid vanskelig å tenke seg hvordan estimeringen skal kunne gjennomføres med de målefeil vi antar det er i materialet. I opplegget som er nyttet her har vi søkt å bøte på ulemper med svakheter i data ved å postulere en definisjonsmessig sammenheng mellom inntekt, forbruk og sparing.

Poenget i det ovenstående er bare å antyde at den lave konsumderiverte kanskje delvis kan skyldes at inntekt og sparing er nokså vidt definert, til å omfatte en nær sagt "usynlig" størrelse som verdistigning på eiendom.

Tabell 1. Estimerte parameterverdier med anslåtte standardavvik. 905 husholdninger

	Parameter med estimat												
		β_j		γ_{j1}		γ_{j2}		λ_{j1}		λ_{j2}		λ_{j3}	
	$\hat{\alpha}_j$	$\hat{b}_j$	St. av.1)	$\hat{\gamma}_{j1}$	St. av	$\hat{\gamma}_{j2}$	St. av.	$\hat{\lambda}_{j1}$	St. av	$\hat{\lambda}_{j2}$	St. av.	$\hat{\lambda}_{j3}$	St. av.
TOTAL FORBRUKSUTGIFT ..	-7 274	0,483	0,038	7 359	1 102	1 317	931	7 095	2 926	6 841	3 193	10 562	2 943
Matvarer	-262	0,034	0,008	2 402	172	1 066	145	1 359	456	1 978	497	2 973	458
Drikkevarer og tobakk	-1 043	0,034	0,006	606	147	-149	124	1 187	389	682	425	695	392
Klær og skotøy	-2 340	0,048	0,008	1 280	191	539	161	1 436	508	388	554	1 774	511
Bolig, lys og brensel ...	1 768	0,080	0,012	-776	299	-89	252	-72	794	973	866	2 170	799
Møbler og hushold.-artikler	-69	0,057	0,008	342	201	-112	170	-101	535	151	584	1 003	538
Helsepleie ...	85	0,008	0,004	100	90	-3	76	-75	239	418	261	295	240
Reiser og transport	-4 297	0,137	0,020	1 830	445	-170	376	2 884	1 182	2 324	1 290	2 123	1 189
Fritidssysler og utdanning .	-60	0,034	0,008	1 002	183	508	155	-338	486	-153	530	-564	489
Andre varer og tjenester .	-1 041	0,050	0,008	583	192	-265	162	820	510	87	557	101	513
SPARING	7 274	0,517	0,09	-7 359	1 102	-1 317	931	-7 095	2 926	-6 841	3 193	-10 562	2 943

1) Standardavvik på $\hat{b}_j$.

Det neste man vil tenke på er vel målefeil i observert sparing. Denne delmassen av husholdninger (905 av i alt 3 071) utgjør i hovedsak husholdninger med stor positiv observert sparing - slik denne er målt. I forutsetningene for modellen (kapittel 3) ligger det at målefeilene på observert sparing er tilfeldige, og kan være enten positive eller negative, varierende fra husholdning til husholdning. Blant disse 905, vesentlig med stor positiv observert sparing, er det rimeligvis en viss overvekt av husholdninger hvor observert sparing er beheftet med <u>positive</u> tilfeldige målefeil, dvs. at

$$S_i > S'_i \qquad (i = 1,2,\ldots, 905)$$

eller

$$'S'_i > 0$$

Dette moment kan konkretiseres på følgende måte: Vi rangordner husholdningene etter fallende observert sparing; de med høyest sparing (S_i) først. Vi har

$$S_i = S'_i + 'S'_i \qquad (i = 1,2,\ldots, 3\ 071)$$

Blant de første husholdningen, med høyest S, er det en tendens til at 'S' gjennomsnittlig er <u>positiv</u>. Blant de 905 kan man derfor frykte at målefeilen i S <u>ikke</u> er tilfeldig. Selv om S kan betraktes som observert med <u>tilfeldige</u> målefeil når man ser alle 3 071 husholdninger som én masse, gjelder ikke det for delmasser som er plukket ut etter nivået på S.

Det ovennevnte kan ha ført til at den marginale sparetilbøyeligheten er blitt overestimert, og konsumtilbøyeligheten tilsvarende underestimert.

Det kan også delvis være andre forklaringer på resultatene, som vel berører spesifikasjoner av modellen. Modellen forutsetter at forbruk og sparing kan forklares ved hjelp av inntekten, tallet på voksne og barn og sesongsvingninger. Men man kan vel tenke seg at det også er andre faktorer som influerer vesentlig på fordelingen mellom forbruk og sparing, faktorer som her altså er omfattet av restleddet. Det kan være så mange grunner til at folk velger å fordele inntekten mellom forbruk og sparing, alt fra relativt enkle og målbare grunner til de mer "usynlige". Når man plukker ut en delmasse av husholdningen etter nivået på sparingen, vil man lett kunne finne uventede resultater, som følge av at restleddet for en slik utplukket delmasse ikke har de samme gjennomsnittlige gunstige egenskaper som for et mer tilfeldig utvalg, dvs. et utvalg som ikke er "håndplukket".

Som allerede påpekt har vi ikke regnet ut standardavvik på $\hat{\hat{b}}$-ene. De anslagene som er gitt på disse standardavvikene er neppe særlig gode. Vi ser f.eks. at anslått standardavvik på $\hat{b}$ er 0,038, mens det samme på $\hat{b}_s$ er 0,098. Estimatorene er imidlertid konstruert slik at $\hat{\hat{b}} + \hat{\hat{b}}_s = 1{,}0$. Hvis vi hadde nyttet den riktige variansformelen (i vedlegg 2), ville standardavviket på $\hat{\hat{b}}$ blitt lik standardavviket på $\hat{\hat{b}}_s$.

Vi har ikke gjort noe forsøk på å beregne den multiple korrelasjonskoeffisienten for ligningene. Problemet med å beregne korrelasjonskoeffisienten er ikke først og fremst at regresjonsprogrammet ikke automatisk foretar de nødvendige beregningene. Beregningene lar seg tross alt gjennomføre. Poenget er at det ikke er gitt hva som skal <u>menes</u> med denne koeffisienten i dette tilfelle. Illustrert ved den forenklede konsumfunksjonen i kapittel 5.1. og 5.2. kan vi tenke oss å velge mellom å beregne korrelasjonen mellom

a) X_j og $(\hat{\hat{a}}_j + \hat{\hat{b}}_j \cdot Y)$ $\quad (j = 1,2,\ldots, m)$

eller

b) X_j og $(\hat{\hat{a}}_j + \hat{\hat{b}}_j (X + S)$

Under estimering av vinkelkoeffisienten benyttet vi nemlig <u>inntekten</u>$^{\circ}$ (Y) som forklaringsvariabel; under estimering av konstantleddet ble forbruk + sparing (X + S) benyttet.

Når vi kjenner $\hat{\hat{b}}_j$ og gjennomsnittlig konsum og sparing (bare med tilfeldige målefeil) kan vi lett finne de gjennomsnittlige inntektselastisiteter. Elastisitetene blir:

	Elastistitet
Totalt forbruk	0,58
Matvarer	0,18
Drikkevarer og tobakk	0,63
Klær og skotøy	0,53
Bolig, lys og brensel	0,88
Møbler og husholdningsartikler	0,75
Helsepleie	0,46
Reiser og transport	0,78
Fritidssysler og utdanning	0,41
Andre varer og tjenester	0,84
Sparing	3,20

Det er iøynefallende hvor lav elastistiteten for totalt forbruk er, og tilsvarende høy for sparingen. Dette har naturligvis sammenheng med nivåene på de deriverte (se foran).

Videre ser vi at elastistiteten for matvarer er meget lav, mens den er relativt høy for bolig, lys og brensel, møbler og husholdningsartikler og reiser og transport. At elastistiteten er lav for matvarer er nokså kjent og vanlig godtatt. Likeledes at den er høy for reiser og transport - som bl.a.

omfatter bilhold. Gruppen andre varer og tjenester inneholder bl.a. hotellopphold og restaurantbesøk, samt såkalte pakketurer.

6.2. Estimering på delvis simulerte sparedata

I dette avsnittet gis empiriske resultater fra estimeringen av parametrene for alle 3 071 husholdningene. Disse omfatter de 905 fra avsnittet foran og 2 166 med simulerte sparedata (jfr. kapittel 4.4. c)).

Tabell 2. Estimerte parameterverdier med anslåtte standardavvik. 3 071 husholdninger

	Parameter med estimat												
		β_j		γ_{j1}		γ_{j2}		λ_{j1}		λ_{j2}		λ_{j3}	
	$\hat{\alpha}_j$	$\hat{b}_j$	St. av.1)	$\hat{\gamma}_{j1}$	St. av.	$\hat{\gamma}_{j2}$	St. av.	$\hat{\lambda}_{j1}$	St. av.	$\hat{\lambda}_{j2}$	St. av.	$\hat{\lambda}_{j3}$	St. av.
TOTAL FORBRUKS-UTGIFT	-8 673	0,695	0,023	4 560	496	1 313	371	3 368	1 228	3 921	1 327	5 573	1 242
Matvarer	-1 009	0,058	0,005	2 351	96	1 190	72	866	239	1 695	258	2 468	242
Drikkevarer og tobakk	-1 088	0,063	0,004	387	84	-195	63	793	208	443	225	295	211
Klær og skotøy	-2 054	0,064	0,005	1 011	95	399	71	550	236	17	255	1 079	239
Bolig, lys og brensel	1 567	0,105	0,007	-961	138	16	104	-561	342	134	370	1 056	347
Møbler og husholdningsartikler	-271	0,078	0,005	-10	97	-16	73	58	240	-67	259	511	243
Helsepleie	-114	0,012	0,003	158	60	24	45	-7	149	43	161	43	151
Reiser og transport	-4 065	0,186	0,012	792	217	-5	162	1 450	537	1 598	581	952	544
Fritidssysler og utdanning ..	-449	0,057	0,005	669	96	194	72	-306	238	74	257	-630	241
Andre varer og tjenester	-1 194	0,074	0,005	163	101	-293	76	526	250	-17	270	-203	253
SPARING	8 668	0,305	0,038	-4 563	496	-1 315	371	-3 369	1 227	-3 923	1 326	-5 575	1 241

1) Standardavvik på $\hat{b}_j$.

Vi finner her at den marginale konsumtilbøyelighet er 0,695, og sparetilbøyeligheten følgelig 0,305. Nedenstående oppstilling viser de gjennomsnittlige inntektselastisitetene:

	Elastisitet
TOTALT FORBRUK	0,85
Matvarer	0,30
Drikkevarer og tobakk	1,00
Klær og skotøy	0,75
Bolig, lys og brensel	1,11
Møbler og husholdningsartikler	1,05
Helsepleie	0,66
Reiser og transport	1,16
Fritidssysler og utdanning	0,73
Andre varer og tjenester	1,38
SPARING	1,70

Av oppstillingen går det fram at sparing, dvs. formuesøkning har den største inntektselastisiteten. Dernest følger gruppen andre varer og tjenester. I denne utgiftsgruppen inngår f.eks. pakketurer til syden og hotell- og restaurantutgifter. På tredjeplass i denne oppstillingen følger gruppen reiser og transport, som omfatter bl.a. anskaffelse, drift og vedlikehold av private transportmidler. Nederst kommer matvarene med en elastisitet på 0,30.

Det kan synes som at resultatene for hele materialet (3 071 husholdninger) passer bedre med våre forestillinger om den typiske konsum- og spareatferd enn hva resultatene fra en "utvalgt" del av materialet gjør (de 905 i avsnittet foran).

7. En ikke-lineær funksjon

Vi har i alt det foregående forutsatt lineær utforming av funksjonene. Det vil selvsagt lett bli reist spørsmål og tvil om rimeligheten i en slik forutsetning. Derfor kan det være ønskelig å tillate at funksjonsformen er noe mer innviklet. Spørsmålet er om det også da lar seg gjøre å identifisere parametrene når det er målefeil i data.

Vi antar nå følgende funksjoner mellom de sanne variablene:

(7.1.1.) $$X'_{ji} = \alpha_j + \beta_j Y'_i + \delta_j Y'^2_i + U_{ji} \qquad (j = 1,2,\dots, m) \quad (i = 1,2,\dots, n)$$

(7.1.2.) $$X'_i = \alpha + \beta Y'_i + \delta Y'^2_i + U_i$$

(7.1.3.) $$S'_i = \alpha_s + \beta_s Y'_i + \delta_s Y'^2_i + U_{si}$$

hvor restleddene innenfor det aktuelle variasjonsområde for variablene er antatt å ha de samme gunstige egenskapene som er forutsatt i kapittel 3. Spørsmålet er nå: Lar det seg gjøre å identifisere og finne konsistente estimatorer for parametrene i funksjonene ved hjelp av de observerbare variable?

Før vi forsøker å svare etablerer vi tilsvarende relasjoner mellom de observerbare variablene - på samme måte som i kapittel 5.2. Vi danner altså relasjonene:

(7.4.) $$X_{ji} = a_j + b_j Y_i + d_j Y_i^2 + \text{restledd} \qquad (j = 1,2,\dots, m) \quad (i = 1,2,\dots, n)$$

$$X_i = a + b Y_i + d Y_i^2 + \text{restledd} \qquad (i = 1,2,\dots, n)$$

$$S_i = a_s + b_s Y_i + d_s Y_i^2 + \text{restledd} \qquad (i = 1,2,\dots, n)$$

Vanlige minste kvadraters metode gir følgende uttrykk

$$\hat{b}_j = \frac{M_{(x_j)(y)} \cdot M^2_{(y^2)} - M_{(x_j)(y^2)} \cdot M_{(y)(y^2)}}{H} \qquad (j = 1,2,\dots, m)$$

$$b = \frac{M_{(x)(y)} \cdot M^2_{(y^2)} - M_{(x)(y^2)} \cdot M_{(y)(y^2)}}{H}$$

$$\hat{b}_s = \frac{M_{(s)(y)} \cdot M^2_{(y^2)} - M_{(s)(y^2)} \cdot M_{(y)(y^2)}}{H}$$

hvor

$$H = M^2_{(y)} \cdot M^2_{(y^2)} - M^2_{(y)(y^2)}$$

Vi ser på de asymtotiske verdiene til noen av leddene i tellerne i uttrykkene - først $\hat{b}_j$. Vi har at

$$M_{(x_j)(y)} = M_{(x_j^!)(y)} + M_{('x_j^!)(y)} = \beta_j \, M_{(y')(y)} + M_{(u_j)(y)} + M_{('x_j^!)(y)} + M_{('x_j^!)(y)}$$

På grunn av forutsetningene i kapittel 5.2. følger det at

$$\underset{n \to \infty}{p \lim} M_{(x_j)(y)} = \beta_j \, m_{(y')(y)}$$

Videre ser vi

$$M_{(x_j)(y^2)} = M_{(x_j^!)(y^2)} + M_{('x'_j)(y^2)} = \beta_j \, M_{(y')(y^2)} + M_{(u_j)(y^2)} + M_{('x_j^!)(y^2)}$$

Det er nærliggende å anta at

$$\underset{n \to \infty}{p \lim} M_{(u_j)(y^2)} = 0$$

og

$$\underset{n \to \infty}{p \lim} M_{('x'_j)(y^2)} = 0$$

og dermed

$$\underset{n \to \infty}{p \lim} M_{(x_j)(y^2)} = \beta_j \, m_{(y')(y^2)}$$

Vi kan nå skrive

$$\underset{n \to \infty}{p \lim} \hat{b}_j = \beta_j \cdot J \qquad (j = 1,2,\ldots, m)$$

og

$$p \lim \hat{b} = \beta \cdot J$$

hvor

$$J = p \cdot \underset{n \to \infty}{\lim} \frac{M_{(y')(y)} \, M^2_{(y^2)} - M_{(y')(y^2)} \, M_{(y)(y^2)}}{H}$$

Ved å gjøre tilsvarende forutsetninger om målefeilene på sparingen kan vi skrive

$$p \lim \hat{b}_s = \beta_s \cdot J$$

Vi ser altså at $\hat{b}$-ene ikke vil være konsistente estimatorer for β-ene. Imidlertid, hvis vi har tilleggsinformasjon her, f.eks. forutsetninger om hva summen $\beta + \beta_s$ skal være, kan vi - analogt med i kapittel 5.2. - finne konsistente estimatorer.

Dersom vi innfører en tilleggsbetingelse på modellen (7.1.1.) - (7.1.3.), nemlig at

$$\sum_{j=1}^{m} \beta_j = \beta \quad \text{og} \quad \beta + \beta_s = 1$$

følger det av framgangsmåten i kapittel 5.2. at

$$\hat{\hat{b}}_j = \hat{b}_j/(\hat{b} + \hat{b}_s) \qquad (j = 1,2,\ldots, m)$$

$$\hat{\hat{b}} = \hat{b}/(\hat{b} + \hat{b}_s)$$

$$\hat{\hat{b}}_s = \hat{b}_s/(\hat{b} + \hat{b}_s)$$

vil være konsistente.

I denne forbindelse kan det antydes at hvis vi legger på den restriksjon å innføre nevnte betingelse, følger det vel nærmest av seg selv at vi også må innføre betingelsene

$$\delta + \delta_s \approx 0$$

og

$$\alpha + \alpha_s \approx 0$$

for at det skal bli en viss "logisk sammenheng" i modellen.

Vi kan da reise spørsmål om det lar seg gjøre å finne konsistente estimatorer for δ_j, δ og δ_s med våre data. Eller, hvis ikke dette er mulig, lar det seg gjøre å si noe om fortegnene på henholdsvis δ og δ_s? Det siste er heller ikke uinteressant; det sier hvordan krummingen på konsumfunksjonen er (eller på sparefunksjon, om man vil).

Uten å gå nærmere inn på broblemet er det rimelig at graden av krumming ikke kan avsløres medmindre man gjør nokså klare forutsetninger om karakteren av målefeilen på inntekten. Det kan antydes ved et enkelt eksempel. Anta at den systematiske målefeilen er slik at observert inntekt øker når den faktiske avtar. Det er opplagt at man da lett får estimert feil fortegn på parameteren foran 2. ordensleddet. Vi legger her merke til at samme konklusjon ikke gjaldt førsteordensleddet; tilsvarende målefeil vil (under visse betingelser) ikke ødelegge muligheten for å finne konsistente estimatorer på parametrene β og β_s.

LITTERATUR

[1] Liviatan, N.: "Errors in Variables and Engel Curves Analysis", Econometrica 29 (1961).

[2] Malinvaud, E.: "Statistical Methods of Econometrics". NHPC, Amsterdam 1966.

[3] Skarstad, O.: "Estimering av engelderiverte ved manglende inntektsdata", Arbeidsnotat IO 77/44.

[4] Skarstad, O.: "Estimering av engelderiverte på data med målefeil", Arbeidsnotat IO 77/45.

[5] Summers, R.: "A Note on Least Squares Bias in Household Expenditure Analysis", med kommentar av S.J. Prais, Econometrica 27 (1959).

[6] Sverdrup, E.: Lov og tilfeldighet, bind I, Universitetsforlaget (1964).

Liviatans metode

Vi skal her kort gjengi resonnementet i N. Liviatans artikkel [1] fram til estimatoren $\hat{b}_j$ (j = 1,2,..., m).

Vi betrakter den enkelte versjonen av konsumfunksjonen (5.2.1.):

$$X_{ji} = a_j + b_j Y_i + \text{restledd} \qquad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

I stedet for inntektsindikatoren Y satte Liviatan inn observert total forbruksutgift som inntektsindikator

$$X_{ji} = a_j + b_j X_i + \text{restledd} \qquad \begin{matrix}(j = 1,2,\ldots, m)\\(i = 1,2,\ldots, n)\end{matrix}$$

og nyttet

$$Y_i - \bar{Y}$$

som instrumentvariabel vi "mutipliserer igjennom" med den og summerer over i. Det gir

$$\sum_{i=1}^{n} X_{ji}\,(Y_i - \bar{Y}) = b_j \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) + \sum_{i=1}^{n} (\text{restledd})_i(Y_i - \bar{Y}) \qquad (j = 1,2,\ldots, m)$$

Under forutsetning av at $\sum_{i=1}^{n} (\text{restledd})_i\,(Y_i - \bar{Y})$ konvergerer mot null ved økende antall observasjoner, vil estimatoren

$$b_j^* = \left[\sum_{i=1}^{n} X_{ji}(Y_i - \bar{Y})\right] \Big/ \left[\sum_{i=1}^{n} X_i(Y_i - \bar{Y})\right] \qquad (j = 1,2,\ldots, m)$$

være konsistent. Det er imidlertid lett å se at følgende sammenheng gjelder:

$$b_j^* = \hat{b}_j/\hat{b} \qquad (j = 1,2,\ldots, m)$$

(se formel 5.2.4.)). Dette er identisk med estimatoren $\hat{\hat{b}}_j$ bortsett fra at sparingen holdes utenfor, dvs. at man estimerer <u>utgifts</u>deriverte i stedet for <u>inntekts</u>deriverte. I prinsippet er altså b_j^* og $\hat{\hat{b}}_j$ samme estimator.

Vedlegg 2

Variansuttrykk

Vi skal her finne fram til at (omtrentlig) uttrykk for variansen på $\hat{\hat{b}}_j$ ($j = 1,2,\ldots, m$). Variansene på $\hat{b}$ og $\hat{b}_s$ kan selvsagt finnes på samme måte.

$\hat{\hat{b}}_j$ kan skrives

$$\hat{\hat{b}}_j = \hat{b}_j/(\hat{b} + \hat{b}_s) = \frac{M_{x_j y}}{M^2_y} \Big/ \frac{M_{xy} + M_{sy}}{M^2_y} = \frac{M_{x_j y}}{M_{(x+s)y}} = \frac{T}{N} \qquad (j = 1,2,\ldots, m)$$

(T = teller, N = nevner).

Hvis vi generelt har en funksjon

$$w = g\,(T, N)$$

kan variansen på w (under visse betingelser) tilnærmes ved (se [6]):

$$\text{var } w \approx \left[\frac{\partial g(E(T),\, E(N))}{\partial(E(T))}\right]^2 \text{var }(T) + \left[\frac{\partial g(E(T),\, E(N))}{\partial(E(N))}\right]^2 \text{var }(N) +$$

$$2\,\frac{\partial g(E(T),\, E(N))}{\partial(E(T))}\;\frac{\partial g(E(T),\, E(N))}{\partial(E(N))}\;\text{cov }(T, N)$$

Ved funksjonen $\hat{\hat{b}}_j = \frac{T}{N}$ får vi

$$\text{var } \hat{\hat{b}}_j \approx \left(\frac{1}{E(N)}\right)^2 \text{var }(T) + \left(\frac{E(T)}{(E(N))^2}\right)^2 \text{var }(N) - 2\,\frac{1}{E(N)} \cdot \frac{E(T)}{(E(N))^2}\;\text{cov }(T, N)$$

Ved innsetting av de empiriske nomentene lar det seg gjøre å finne omtrentlige anslag på variansene i praktiske tilfelle.

Det lar seg selvsagt også gjøre å uttrykke variansen analytisk via varianser på restledd og de forskjellige målefeilene. Dette har vel imidlertid mindre interesse da slike uttrykk vanskelig kan gi grunnlag for beregninger i praksis, og fordi formlene blir så innfløkte at de heller ikke gir noen analytisk innsikt.

I kapittel 5.2. er det antydet at estimatoren $\hat{\hat{b}}_j$ kan gi mindre varians enn den vanlige minstekvadraters estimatoren. Variansen på den vanlige minstekvadratersestimatoren er identisk med

$$\text{var } \hat{b}_j = \left[\frac{1}{E(N)}\right]^2 \text{var } T$$

Følgelig er

$$\text{var } \hat{\hat{b}}_j \approx \text{var } \hat{b}_j + \left[\frac{E(T)}{(E(N))^2}\right]^2 \text{var }(N) - 2\,\frac{E(T)}{(E(N))^3}\;\text{cov }(T, N)$$

Hvis vi bytter ut ≈ tegnet med =, får vi at var $\hat{\hat{b}}_j < \hat{b}_j$ hvis

$$\left[\frac{E(T)}{(E(N))^2}\right]^2 \text{var }(N) < 2\,\frac{E(T)}{(E(N))^3}\;\text{cov }(T, N)$$

dvs. hvis

$$\text{cov }(T, N) > \tfrac{1}{2}\,\frac{E(T)}{E(N)}\ \text{var }(N)$$

Nå er det selvsagt ikke så interessant å sammenlikne variansene på $\hat{\hat{b}}_j$ og $\hat{b}_j$ siden $\hat{b}_j$ ikke er konsistent (ved målefeil i inntekten). Det interessante ved en slik sammenlikning kommer fram hvis vi et øyeblikk tenker oss at det ikke er målefeil i inntekten. (Vi antar fortsatt tilfeldige målefeil i forbruk og sparing.) Da vil begge estimatorene være konsistente. Men hvem er best?

Man kunne, analogt med Liviatan, ta utgangspunkt i at $\hat{\hat{b}}_j$ er en instrumentvariabelestimator. Det er da nærliggende å tro at estimatoren som sådan gir større varians enn den vanlige minste kvadratersestimatoren ($\hat{b}_j$). Dette resonnementet holder imidlertid ikke her. Det er ingen logisk holdbar grunn til at ikke følgende ulikhet skal kunne være oppfylt (ofte eller sjeldnere):

$$\text{cov }(T, N) > \tfrac{1}{2}\,\frac{E(T)}{E(N)}\ \text{var }(N)$$

Ett spesialtilfelle kan særlig nevnes: vi tenker oss at vi skal "estimere" summen av konsum og sparetilbøyeligheten (en sum vi alt har definert lik 1,0). Vi får da ved de to estimeringsmåtene h.h.v.:

$$\hat{\hat{b}} + \hat{\hat{b}}_s = M_{(x+s)y}/M_{(x+s)y} = 1{,}0$$

og

$$\hat{b} + \hat{b}_s = M_{(x+y)y}/M_y^2$$

Her blir selvsagt variansen på $\hat{\hat{b}} + \hat{\hat{b}}_s$ alltid lik null. Derimot blir ikke variansen på $\hat{b} + \hat{b}_s$ lik null (ved et endelig antall observasjoner). I dette tilfelle ser vi klart at variansen på $\hat{\hat{b}}$-estimatoren er mindre enn på $\hat{b}$-estimatoren.

Man kan selvsagt også reflektere omkring det tilfelle at det heller ikke finnes målefeil i forbruket eller sparingen, hverken systematiske eller tilfeldige feil. Dét er enda en tanke mer urealistisk enn det forannevnte, og har derfor neppe praktisk betydning. Imidlertid, vi ser lett at $\hat{\hat{b}}_j$ og $\hat{b}_j$ ville gi sammenfallende resultater. Dette følger av at

$$X = X'$$
$$S = S'$$
$$Y = Y'$$

og altså at

$$X + S = Y$$

alltid gjelder eksakt. Ved siden av betingelsen på restleddene

$$\sum_{j=1}^{m} U_{ji} = U_i \qquad (i = 1,2,\ldots, n)$$

og

$$U_i + U_{si} = 0$$

følger det da at

$$\hat{\hat{b}}_j = \hat{b}_j$$

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP)

Nr. 79/2 Det norske nasjonalregnskapet. Dokumentasjonsnotat nr. 7 Sektorberegninger for samferdselssektorene og reparasjon av kjøretøyer m.v. ISBN 82-537-0968-4

- 79/3 Undersøkelse av renholdsbedrifter 1977 ISBN 82-537-0969-2

- 79/4 Deltidsundersøkelsen 1978 ISBN 82-537-0970-6

- 79/5 Boligutgiftsbegrepet i forbruksundersøkelsene En metodestudie ISBN 82-537-0971-4

- 79/6 MAFO-Makromodell for folketrygden En skisse av en budsjettmodell ISBN 82-537-0972-2

- 79/7 Estimering av inntektsderiverte på tverrsnittsdata med målefeil ISBN 82-537-0976-5

- 79/8 Det norske nasjonalregnskapet Dokumentasjonsnotat nr. 14 Sektorberegninger for fiske og fangst ISBN 82-537-0977-3

- 79/9 Statsansattes vurdering av arbeidsforholdene i staten ISBN 82-537-0954-4

- 79/13 Forbruksundersøkinga 1967 - 1977 samanlikna med nasjonalreknskapen ISBN 82-537-1001-1

- 79/14 Oppgavebyrden for små bedrifter ISBN 82-537-0995-1

Pris kr 13,00

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

ISBN 82-537-0976-5