OM ESTIMERINGSPROBLEMER FOR MAKROMODELLER[1])

Af P. NØRREGAARD RASMUSSEN.

I sin bog: »Finansprocessen i det økonomiske kredsløb« (1948) har Jørgen Gelting i kapitel 3 påbegyndt et arbejde, som man må håbe i fremtiden vil blive fortsat. Der opstilles her en matematisk makromodel for dansk økonomi i form af et simultant ligningssystem. I økonometrisk forskning fra de sidste 10–15 år har man gang på gang set sådanne modeller opstillet. Arbejdet er ikke mindst blevet intensiveret efter Keynes' »General Theory..«.

I dansk forskning har økonometriske forskningsmetoder imidlertid hidtil kun vunnet ringe indpas. Man må derfor være Jørgen Gelting taknemmelig, fordi han med omtalte arbejde gør det første forsøg på at opstille og estimere parametrene i en model, som eventuelt kan have gyldighed for danske forhold. Der kan næppe være tvivl om, at arbejdet med sådanne modeller vil blive noget af det centrale i den kommende forskning.

Statistisk Departements udsendelse af undersøgelsen over nationalprodukt og nationalindkomst i Danmark 1930–46[2]) må i denne forbindelse nævnes, idet man først hermed har fået nogenlunde tilfredsstillende grundlag for arbejdet med makroteorien. Kun må det beklages, at tallene ikke går længere tilbage end til 1930. Det vil naturligvis være ønskeligt at få så lange tidsrækker som muligt.

Der har i de senere år i Amerika – og ganske særligt ved Cowles Commission i Chicago – været arbejdet overordentlig meget med udredning af

---

[1]) Denne artikel pretenderer ikke at give noget egentlig nyt. Imidlertid synes de her antydede problemer ikke at være blevet diskuteret så meget i Danmark som i andre lande. M. h. t. litteraturen kan bl. a. henvises til: Trygve Haavelmo: »The Statistical Implications of a System of Simultaneous Equations«, *Econometrica*, Vol. 11, 1943; videre af Trygve Haavelmo: »The Probability Approach in Econometrics,« *Econometrica*, Vol. 12, Supplement, 1944 og »Methods of Measuring the Marginal Propensity to Consume«, *Journal of the American Statistical Association*, Vol. 42, 1947; videre kan nævnes Tjalling Koopmans, »Statistical Estimation of Simultaneous Economic Relations«, *Journal of the American Statistical Association*, Vol. 40, 1945. – Der er i det følgende lagt vægt på at gøre fremstillingen elementær.

[2]) Statistiske Meddelelser, 4. række, 129. bind, 5. hæfte.

OM ESTIMERINGSPROBLEMER FOR MAKROMODELLER[1])

Af P. NØRREGAARD RASMUSSEN.

I sin bog: »Finansprocessen i det økonomiske kredsløb« (1948) har Jørgen Gelting i kapitel 3 påbegyndt et arbejde, som man må håbe i fremtiden vil blive fortsat. Der opstilles her en matematisk makromodel for dansk økonomi i form af et simultant ligningssystem. I økonometrisk forskning fra de sidste 10–15 år har man gang på gang set sådanne modeller opstillet. Arbejdet er ikke mindst blevet intensiveret efter Keynes' »General Theory..«.

I dansk forskning har økonometriske forskningsmetoder imidlertid hidtil kun vunnet ringe indpas. Man må derfor være Jørgen Gelting taknemmelig, fordi han med omtalte arbejde gør det første forsøg på at opstille og estimere parametrene i en model, som eventuelt kan have gyldighed for danske forhold. Der kan næppe være tvivl om, at arbejdet med sådanne modeller vil blive noget af det centrale i den kommende forskning.

Statistisk Departements udsendelse af undersøgelsen over nationalprodukt og nationalindkomst i Danmark 1930–46[2]) må i denne forbindelse nævnes, idet man først hermed har fået nogenlunde tilfredsstillende grundlag for arbejdet med makroteorien. Kun må det beklages, at tallene ikke går længere tilbage end til 1930. Det vil naturligvis være ønskeligt at få så lange tidsrækker som muligt.

Der har i de senere år i Amerika – og ganske særligt ved Cowles Commission i Chicago – været arbejdet overordentlig meget med udredning af

---

[1]) Denne artikel pretenderer ikke at give noget egentlig nyt. Imidlertid synes de her antydede problemer ikke at være blevet diskuteret så meget i Danmark som i andre lande. M. h. t. litteraturen kan bl. a. henvises til: Trygve Haavelmo: »The Statistical Implications of a System of Simultaneous Equations«, *Econometrica*, Vol. 11, 1943; videre af Trygve Haavelmo: »The Probability Approach in Econometrics,« *Econometrica*, Vol. 12, Supplement, 1944 og »Methods of Measuring the Marginal Propensity to Consume«, *Journal of the American Statistical Association*, Vol. 42, 1947; videre kan nævnes Tjalling Koopmans, »Statistical Estimation of Simultaneous Economic Relations«, *Journal of the American Statistical Association*, Vol. 40, 1945. – Der er i det følgende lagt vægt på at gøre fremstillingen elementær.

[2]) Statistiske Meddelelser, 4. række, 129. bind, 5. hæfte.

de metodologiske problemer, som rejser sig, når man forsøger at konfrontere en makromodel med virkeligheden. De resultater, som herved er nået, må afgjort siges at være af ganske betydelig interesse. Jeg skal i det følgende forsøge at skitsere visse problemstillinger ved sådanne undersøgelser.

En makromodel vil bestå af et vist antal ligninger, som kræves tilfredsstillet samtidig. I relationerne indgår for det første forskellige makrostørrelser: samlet nationalindkomst, samlet forbrug, samlet import o. s. v. Desuden vil der indgå en række karakteriserende parametre: forbrugstilbøjeligheden, importtilbøjeligheden o. s. v. De relationer, som tilsammen danner det simultane ligningssystem kan inddeles i to afgørende forskellige hovedgrupper.

For det første er der *definitionssammenhængene*, som f. eks.

$$R = C + I. \tag{1}$$

R betegner total indkomst, C forbrug og I investering. Disse relationer er rent bogholderimæssige. Man får dem »gratis«, fordi de er en logisk følge af vore definitioner. Heraf følger imidlertid også, at det er rent tautologiske udtryk, som i og for sig ikke udsiger noget som helst »nyt«. I denne gruppe vil også indgå »betingelsesligninger«.

For det andet vil der indgå, hvad man kunne kalde *tekniske relationer* (i vid forstand), som f. eks.

$$X = f\,(N, K). \tag{2}$$

X betegner her samlet produktion, N betegner samlet beskæftigelse og K betegner kapitalapparatets størrelse. (2) udtrykker altså, at den samlede produktion er en eller anden funktion af anvendt mængde arbejdskraft samt kapitalapparatet — al produktion foregår jo ved at anvende en vis mængde arbejdskraft i forbindelse med en vis mængde kapital (herunder »jord«).

Ligesom der kan opstilles en lang række definitionssammenhænge i lighed med (1), kan der opstilles en lang række tekniske relationer (strukturrelationer). Hvor mange relationer man skal opstille, d. v. s. hvor stor og detailleret man skal gøre sin model, må afgøres ud fra spørgsmålet om, hvortil man vil anvende modellen.

Forskellen mellem definitionsrelationerne og de tekniske relationer svarer vist ganske til nypositivismens skelnen mellem analytiske og syntetiske sætninger[1]). Dette kan være praktisk at have in mente. Thi det vil vides, at

[1]) Jfr. f. eks. G. H. von Wright: »Den logiska empirismen«, Stockholm 1943. — En analytisk sætning er en sætning, hvis sandhed eller falskhed følger af de i sætningen indgående ords betydning, mens en påstand, hvis sandhed eller falskhed ikke er en følge af de i sætningen indgående ord, kaldes syntetisk. Eksempel på en analytisk sætning er: »Økonomi er en socialvidenskab« — denne sætnings sandhed afhænger alene af de i sætningen indgående ord. Eksempel på syntetisk sætning: »I dag skinner solen« — denne sætning er ikke sand alene i kraft af de indgående ord.

mens man kan tillægge analytiske sætninger sandhedsværdi a priori, så er dette udelukket for syntetiske sætninger.

Det følger nu umiddelbart heraf, at hvad ovenfor er kaldt definitionssammenhænge ikke giver anledning til statistiske problemer overhovedet. Derimod er det ved de tekniske relationer de statistiske problemer findes. *Dette kan siges at være en simpel følge af, at disse sætninger er de eneste, som giver udsagn om et eller andet forhold.* Sådan set kan man sige, at det er herom al statistisk teori handler.

Ud af de tekniske relationer skiller man nu en særlig gruppe, som man – ifølge det mere eller mindre tilfældige system, som består i opdeling i forskellige videnskaber – beskæftiger sig med qua økonomer. Det drejer sig f. eks. om relationen

$$C = C\,(R,\ P,\ \alpha_o,\ \alpha_1,\ \ldots\ \alpha_n) \tag{3}$$

Her er P »prisniveauet« og α'erne karakteriserende parametre. Iblandt α 'erne kan f. eks. også være saadanne, som bestemmer fordelingsfunktionen for nationalindkomsten (R) på individerne.

Man kan nu præcisere sin hypotese nærmere ved at sige, at forbruget er den og den bestemte funktion (f. eks. lineær) af R, P og α 'erne. Ved valget af hypotese må man have to hensyn for øje. Dels skal hypotesen »passe« til observationerne, og dels skal hypotesen alt andet lige være så simpel som mulig.

Efter at have opstillet sin hypotese kan man søge at estimere – numerisk bestemme – de karakteriserende parametre.

Lad os antage, at man efter et eller andet princip har bestemt sine α'er i (3). Man kan da finde, hvor stort forbruget er ifølge den opstillede relation og sammenligne dette forbrug med det faktiske i den betragtede periode. Man vil da altid i praksis opdage, at der er en »lille« forskel mellem beregnet og faktisk værdi. Denne forskel opstår af tre grunde: dels fordi der kan være målefejl, dels fordi vor parameter er »sample« bestemt og dels – og frem for alt – fordi vi ikke har draget hele omverdenen ind i vor funktion, men kun har taget hensyn til visse »relevante« størrelser.

Dette betyder, at (3) egentlig burde skrives:

$$C = C\,(R,\ P,\ \alpha_o,\ \alpha_1,\ \ldots\ \alpha_n) + u \tag{4}$$

hvor u er en størrelse, som netop angiver afvigelserne mellem beregnet og faktisk værdi af C. Hvis vi nu har opstillet en »god« hypotese, så betyder det, at de mest »væsentlige« forklaringsfaktorer er med i vor relation. Det vil imidlertid betyde, at u vil være uden *systematiske* variationer. Vi kan følgelig opfatte u som en stokastisk variabel, som vil være fuldstændig beskrevet, når frekvensfunktionens form og parametre er angivet. Da der så vidt muligt ikke skal forekomme systematiske variationer i u, kan man sætte forventningen (E) og variansen (var) til:

$$E(u) = o \text{ og} \quad var(u) = E(u^2) = \sigma_u^2. \tag{5}$$

Det kan bemærkes, at ofte vil det ikke være nødvendigt at sætte den stærke begrænsning på u, som ville ligge i en forudsætning om normal fordeling.

Til antagelsen om, at vi har fået de »væsentlige« forklaringsfaktorer med i vor relation, hører antagelsen om, at u ikke er autokorreleret, d. v. s. man forudsætter:

$$E(u_t u_{t-\tau}) = o. \tag{6}$$

t angiver her tiden og (6) skal være opfyldt for alle værdier af $\tau \neq o$.

Problemerne vil måske blive lidt mere anskuelige, om vi f. eks. angiver følgende form for (4):

$$C_t = a_o + a_1 R_t + u_t \tag{7}$$

Betegnelserne er som før, C_t og R_t er forbrug og indkomst i periode t, mens α'erne er karakteriserende parametre.

Man vil nu bemærke, at C_t er ikke determineret blot fordi vi angiver størrelsen af a_o og a_1. Man må tillige angive variansen for den stokastisk variable u_t. Dette problem skal jeg senere komme nærmere ind på.

Det er klart, at der opstår en lang række problemer omkring valget af hypotese. Hvorfor således netop vælge (7)? Jeg skal ikke her gå i detailler med disse problemer, men blot berøre et par spørgsmål.

For det første vil det være klart, at man ofte vil være tilbøjelig til at give sin hypotese en dynamisk formulering. En dynamisk variant af (4) vilde f. eks. være:

$$C_t = a_o + a_1 R_t + a_2 R_{t-1} + a_3 \cdot \frac{dP}{dt} + u_t. \tag{8}$$

Naturligvis er der mange andre muligheder. Man kunne f. eks. lade renten, indkomstfordelingen o. s. v. indgå. Det angivne er blot at tage som et eksempel.

For det andet kan der være grund til at gøre et par bemærkninger om den valgte funktionsform. I de hidtil gennemførte undersøgelser har man i reglen – så vidt mig bekendt – ladet sig nøje med lineære ligninger. Problemerne mangedobles, hvis man inddrager mere komplicerede funktioner i billedet. Om fejlen herved er særlig »stor« er også et spørgsmål.

Tinbergen anfører[1]), at »The use of linear relations means much less loss of

[1]) »Statistical Testing of Business Cycle Theories, II, Business Cycles in The United States of America 1919–32,« Geneva 1939, p. 11.

generality than is sometime believed. In the case of small variations in variables, . . . it can even be proved mathematically that there is no loss of generality at all.«

Der kan måske være grund til at nævne, at så snart man lader tidsforskudte størrelser indgå, vil man dermed have mulighed for at få svingninger ind i sit system. Matematisk er dette et simpelt resultat at det forhold, at f. eks. differensligningen

$$a_o y_t + a_1 y_{t-1} + \ldots + a_n y_{t-n} = o$$

under visse omstændigheder vil kunne tilfredsstilles af funktionen:

$$y_t = A\lambda^t \sin (a + \alpha \cdot t)$$

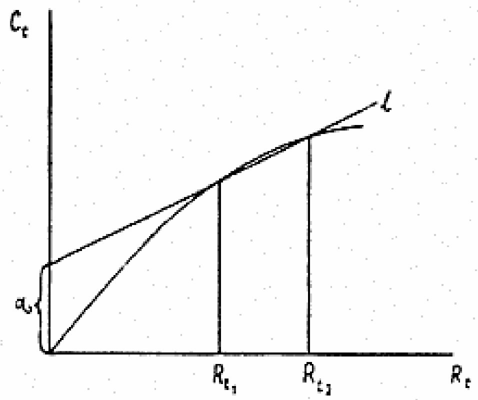
som er en funktion med bølgekomponent.

Det er således en misforståelse, når der indvendes mod brugen af lineære ligningssystemer, at man da ikke kan få f. eks. konjunkturbevægelser med i modellen. Så snart man lader størrelser med »lag« indgå i modellen, har man mulighed for at få svingninger frem.

I den ovenfor nævnte afhandling opstiller Gelting sin forbrugsfunktion på formen (se p. 49, formel (10) – her er anvendt de ovenfor indførte symboler med tilføjelse af t, s og d som er de marginale kvoter for henholdsvis direkte skat, opsparing og indirekte skat):

$$C = R\,(1 - t)\,(1 - s)\,(1 - d). \tag{9}$$

Det vil være åbenbart, at dette er en særlig simpel relation. Naturligvis er man også nødt til og må foretrække simple relationer, og på forhånd kan man ikke afvise en sådan simpel relation. Der kan vist imidlertid være grund til at henlede opmærksomheden på en særlig abstraktion i (9). I denne relation er det konstante led fra de tidligere foreslåede relationer (a_o) åbenbart ikke medtaget – sat lig o. Man fristes også let til at anlægge den betragtning, at hvis indkomsten er lig o, så må forbruget også være o, hvorfor det er ulogisk at operere med et konstant led forskelligt fra o. Her må man imidlertid huske, at f. eks. relation (7) ikke er opsat med krav på gyldighed for alle værdier af R. Den »gælder« kun indenfor visse grænser. Der er vel ingen tvivl om, at den »rigtige« forbrugsfunktion er ikke-lineær m. h. t. R, men det centrale er netop, at *indenfor »moderate« intervaller* kan vi betragte denne funktion som en ret linie. Hvis i følgende figur den krumme kurve er den »rigtige« forbrugsfunktion (m. h. t. R), så ser man, at $R_t = o$ giver $C_t = o$. Hvis vi imidlertid i intervallet fra R_{t1} til R_{t2} substituerer kurven med den rette linie l, så må vi lade l få et analytisk udtryk, hvor det konstante led (a_o) er $\neq o$. I al almindelighed vil det sikkert være betænkeligt at kræve $a_o = o$. Det betyder nemlig, at vi forlanger, at den linie, som vi lader substituere et stykke af kurven, skal gå gennem origo, hvilket krav i reglen vil virke usmidigt.



En lang række andre problemer opstår i tilslutning til spørgsmålet om valg af hypotese. Jeg må imidlertid lade disse ligge og vender derfor tilbage til problemet om estimering af de indgående parametre.

Lad os sige, vi har opstillet en model i en eller anden form for det økonomiske samfund, som vi ønsker at beskrive. Modellen består af et antal simultane ligninger:

(10) $f_i(x_1, x_2 \ldots x_n, a_o, a_1, \ldots a_m) + u_i = o \ (i = 1, 2, \ldots n)$

hvoriblandt findes en forbrugsfunktion, en investeringsfunktion, en eksportfunktion, en importfunktion, en produktionsfunktion o. s. v. samt en række definitionsligninger.

Har vi først valgt vore funktioner, er opgaven på grundlag af en række observerede data (f. eks. tidsrækker) at bestemme de indgående karakteriserende parametre. Den hidtil almindeligt anvendte metode har bestået deri, at man har taget sine funktioner *en for en* og efter et eller andet princip – i reglen mindste kvadraters metode – forsøgt at estimere de ukendte parametre.

Lad os sige, at der blandt relationerne i modellen findes en forbrugsfunktion som (7). Man har da »klippet modellen i stykker« og taget (7) ud og analyseret den særskilt uden hensyntagen til de andre funktioner. Sætter man – idet p er antal observationer –

$$\sum_{t=1}^{p} (C_t - a_o - a_1 R_t)^2 = Q,$$

har man krævet, at Q skal minimaliseres. Den nødvendige betingelse herfor er, at

$$\frac{\delta Q}{\delta a_o} = \frac{\delta Q}{\delta a_1} = o.$$

herved finder man normalligningerne til (7), hvorved a_o og a_1 er blevet estimeret.

Nu er forholdet imidlertid det, at denne fremgangsmåde ofte vil føre til »skæve« estimater for parametrene. Fejlen kan vel siges at være den, at man glemmer, at formulere sine relationer stokastisk. Alle de foran omtalte tekniske relationer er i virkeligheden relationer, som forsøger at beskrive en eller flere variable ved en eller flere andre variable. *Imidlertid afhænger alt af alt.* Hvis man derfor skulde beskrive fuldstændigt, måtte man have hele »verden« med i hver relation. I stedet nøjes vi med at tage de »vigtigste« og tillader så en »lille« afvigelse mellem den observerede og den beregnede størrelse.

Tager vi som eksempel atter relation (7), må man huske, at her indgår R_t som uafhængig variable. Men R_t er *ikke* en uafhængig variabel, som leveres »udefra«. R_t indgår selv som afhængig variabel i modellen. Det betyder, at R_t selv er en stokastisk variabel, hvorfor resultatet bliver, at u_t – som jo ifølge (7) er forskellen mellem observeret og beregnet C_t – kommer til at indeholde både stokastiske variationer fra R_t og fra C_t.

Således kan man måske udtrykke den centrale fejl. Det betyder nemlig, at man ikke kan gå direkte løs på (7), men må tage hensyn til, at (7) er en enkelt relation i et simultant ligningssystem. *Man må derfor ved den empiriske bestemmelse af sine parametre ikke klippe sit system (modellen) i stykker, men må tage hensyn til de andre relationer.*

Dette kan gøres på den måde, at man substituerer R_t med andre størrelser hentet fra den øvrige del af ligningssystemet. De størrelser, man herved når frem til som uafhængige i relationen, må være størrelser, som kan betragtes som autonom variable f. eks. simpelthen som en række udefra givne tal. Man når herved frem til det såkaldte reducerede ligningssystem.

Et konkret gennemregnet eksempel vil måske gøre disse forhold mere anskuelige.

Lad os tænke os en relation beskrivende en forbindelse af formen:

$$x_1 = a + bx_2 + u. \tag{11}$$

x_1 er en variabel som søges »forklaret« ved bevægelser i en anden variabel, x_2. a og b er konstanter. u er en stokastisk variabel ($u = x_1 - a - bx_2$) med forventning:

$$E\,(u) = 0 \text{ og en varians:}$$
$$E\,(u^2) = \sigma_u^2.$$

Man tænker sig, at det empiriske materiale består i to tidsrækker for x_1 og x_2. Som før antydet må det forudsættes, at u ikke er autokorreleret.

Lad endvidere en tredie tidsrække, x_3, være givet som en række udefra givne tal – en autonom tidsrække. Af denne forudsætning følger:

$$E(x_3 u) = x_3 E(u) = O.$$

Mellem x_3, x_2 og x_1 antages følgende bånd (definitionsrelation) at herske

$$x_1 + x_2 = x_3. \tag{12}$$

Med x_3 som autonomt givet danner (11) og (12) et simultant, determineret system. De to variable x_1 og x_2, er dermed »låst fast« ved de bånd de to relationer lægger på dem.

Statistisk består opgaven nu i at bestemme parametrene i systemet: – a og b.

Som foran er omtalt vil man få et »skævt« estimat for a og b, såfremt man betragter (11) isoleret – altså klipper systemet i stykker – uden hensyntagen til (12). Dette lader sig let vise i det givne meget simple tilfælde[1]).

Lad os tænke os, at man i første omgang vil forsøge at bestemme størrelsen af b. Anvender man det almindeligt brugte princip at tage (11) isoleret – d. v. s. klipper systemet i stykker – får man to normalligninger til (11) (jfr. foran):

$$\Sigma_n x_1 = na + b\,\Sigma_n x_2 \text{ og}$$
$$\Sigma_n x_1 \cdot x_2 = a\,\Sigma_n x_2 + b\,\Sigma_n x_2^2$$

som ved løsning giver[2]):

$$b^0 = \frac{n\,\Sigma x_1 \cdot x_2 - \Sigma x_1 \cdot \Sigma x_2}{n\,\Sigma x_2^2 - (\Sigma x_2)^2} \tag{13}$$

Indsættes heri de fra tidsrækkerne fundne værdier for $\Sigma x_1 x_2$, $\Sigma x_1 \cdot \Sigma x_2$, Σx_2^2 og $(\Sigma x_2)^2$ fås b. b^0 betegner den ved denne metode fundne værdi for b. Dette har været og er stadig – uden reservationer – det mest udbredte estimeringsprincip.

Om man imidlertid betragter systemet under et, d. v. s. *også* tager hensyn til (12) bliver resultatet et andet. Løses (11) og (12) m. h. t. x_1 og x_2 fås det reducerede ligningssystem:

$$x_1 = \frac{a}{1+b} + \frac{b}{1+b} \cdot x_3 + \frac{u}{1+b} \tag{14}$$

[1]) Den anførte model er kun ment som et illustrerende, let tilgængeligt eksempel. Det må nævnes, at så snart man går over til større, mere komplicerede modeller, står man snart overfor meget store problemer. Som eksempel kan antydes det såkaldte identifikationsproblem. Når man løser det oprindelige ligningssystem (f. eks. (11) – (12) ovenfor) og finder frem til det reducerede ligningssystem (f. eks. (14) og (15) nedenfor) vil det i almindelighed gælde, at til hver model svarer eet og kun et reduceret system. Det omvendte behøver imidlertid ikke at være tilfældet. Der kan på denne måde findes »overidentificerede« systemer, eller et system kan mangle identifikation. I sådanne tilfælde må der gribes til meget omfattende og arbejdskrævende estimeringsberegninger. – Alle de herunder hørende problemer vil blive behandlet i en kommende monografi nr. 10 fra Cowles Commission: »Statistical Inference in Dynamic Economic Models«.

[2]) For nemheds skyld betegnes Σ_n i det følgende blot ved Σ; n er antal iagttagelser.

(15) $$x_2 = \frac{-a}{1+b} + \frac{1}{1+b} \cdot x_3 - \frac{u}{1+b}$$

(14) og (15) danner det ovenfor omtalte reducerede ligningssystem. Benytter man det reducerede ligningssystem til at finde $\Sigma x_1 x_2$, $\Sigma x_1 \cdot \Sigma x_2$, Σx_2^2 og $(\Sigma x_2)^2$ og indsættes de således fundne værdier i (13) finder man:

$$b^0 = \frac{b\,(n\,\Sigma x_3^2 - (\Sigma x_3)^2) - (n\,\Sigma u^2 - (\Sigma u)^2) + (1-b)(n\,\Sigma x_3 \cdot u - \Sigma x_3 \cdot \Sigma u)}{(n\,\Sigma x_3^2 - (\Sigma x_3)^2) + (n\,\Sigma u^2 - (\Sigma u)^2) - 2(n\,\Sigma x_3 \cdot u - \Sigma x_3 \cdot \Sigma u)}$$

Idet

$$E\,(x_3 \cdot u) = 0$$

får man, at når antal iagttagelser forøges vil b^0 stokastisk[1]) nærme sig til en grænse:

(16) $$\operatorname*{plim}_{n \to \infty} b^0 = \frac{b\,(n\,\Sigma x_3^2 - (\Sigma x_3)^2) - (n\,\Sigma u^2 - (\Sigma u)^2)}{(n\,\Sigma x_3^2 - (\Sigma x_3)^2) + (n\,\Sigma u^2 - (\Sigma u)^2)}$$

Der kan nu mindes om, at idet tidsrækken for x_3 er givet, kan man herudfra beregne det aritmetiske gennemsnit, $\bar{x}_3$, hvorfra kan findes:

$$\sigma_{x3}^2 = \frac{1}{n}\,\Sigma\,(x_3 - \bar{x}_3)^2 \qquad (n \to \infty)$$

hvoraf fås det velkendte udtryk:

$$n \cdot \sigma_{x3}^2 = \Sigma x_3^2 - \bar{x}_3\,\Sigma x_3, \text{ hvoraf igen:}$$
$$n^2 \cdot \sigma_{x3}^2 = n\,\Sigma x_3^2 - (\Sigma x_3)^2$$

som tillige med det tilsvarende udtryk for σ_u kan indsættes i (16), hvorved man får:

(17) $$\operatorname*{plim}_{n \to \infty} b^0 = \frac{b \cdot \sigma_{x3}^2 - \sigma_u^2}{\sigma_{x3}^2 + \sigma_u^2}$$

Skrives (17) på formen.

(18) $$\operatorname*{plim}_{n \to \infty} b^0 = \frac{b - \dfrac{\sigma_u^2}{\sigma_{x3}^2}}{1 + \dfrac{\sigma_u^2}{\sigma_{x3}^2}}$$

ser man let, at i almindelighed vil

(19) $$\operatorname*{plim}_{n \to \infty} b^0 \neq b,$$

hvilket altså betyder, at det fundne estimat (b^0) for b er forskelligt fra det »sande estimat«. Det anvendte estimeringsprincip synes således ikke særlig velegnet. Af (17) ses let, at for $b > -1$ vil $b^0 < b$.

[1]) For at antyde at det her drejer sig om en stokastisk grænseovergang — til forskel fra den i matematikken kendte — skrives »plim b^0« (cfr. »lim b^0«) — probability limit.

Et spørgsmål rejser sig umiddelbart: Hvad skal man forstå ved et »godt« estimat? Om et estimat skal betegnes som »godt« eller »tilfredsstillende« er naturligvis i en vis forstand vilkårligt. Det må bl. a. bestemmes af, hvortil man vil anvende sine estimater. Det vil føre meget vidt, om jeg her skulle gå ind på en diskussion af principper for vurdering af estimeringsmetoder. Men nogle få bemærkninger må det dog være på sin plads at gøre.

Et estimat siges at være konsistent, såfremt den estimerede værdi ligger »tæt på« den sande værdi. Eller mere præcist: Såfremt følgende er opfyldt:

$$\underset{n \to \infty}{plim}\,(b - \varepsilon \leqq b^* \leqq b + \varepsilon) = 1, \tag{20}$$

så siges b^* at være et konsistent estimat for b. (20) læses som følger: Sandsynligheden for, – når antal iagttagelser går mod uendelig – at det fundne estimat (b^*) afviger mindre end et nok så lille tal ε fra den sande værdi (b), går mod 1.

At kræve, at et estimat skal være konsistent, synes åbenbart at være et »rimeligt« krav. – En lang række andre krav kan stilles, hvilket jeg dog ikke skal gå ind på her.

Hvis man går med til at kræve, at et estimat skal være konsistent, vil det være åbenbart, at det foran anvendte estimeringsprincip, kan kaldes uhensigtsmæssigt. At b^0 ikke er noget konsistent estimat for b følger umiddelbart af (19).

Tilbage står blot at vise, at man får et konsistent estimat for b, såfremt man ved estimeringen tager hensyn til hele systemet. Man må da i stedet for at bestemme b ud fra (11) isoleret benytte (14) og (15). Anvender man mindste kvadraters metode på disse to ligninger, får man to normalligninger til bestemmelse af $\frac{b}{1 + b}$ og to til bestemmelse af $\frac{1}{1 + b}$. Løses disse ligninger fås de velkendte formler:

$$\frac{b}{1 + b} = \frac{n \cdot \Sigma\, x_1 \cdot x_3 - \Sigma\, x_1 \cdot \Sigma\, x_3}{n\, \Sigma\, x_3^2 - (\Sigma\, x_3)^2} \tag{21}$$

og

$$\frac{1}{1 + b} = \frac{n\, \Sigma\, x_2 \cdot x_3 - \Sigma\, x_2 \cdot \Sigma\, x_3}{n\, \Sigma\, x_3^2 - (\Sigma\, x_3)^2} \tag{22}$$

Forholdet mellem (21) og (22) giver

$$\bar{b} = \frac{n\, \Sigma\, x_1 \cdot x_3 - \Sigma\, x_1 \cdot \Sigma\, x_3}{n\, \Sigma\, x_2 \cdot x_3 - \Sigma\, x_2 \cdot \Sigma\, x_3} \tag{23}$$

idet $\bar{b}$ betegner det på denne måde fundne estimat for b.

Indsættes i (23) de af (14) og (15) fundne udtryk for $\Sigma\, x_1 \cdot x_3$, $\Sigma\, x_1 \cdot \Sigma\, x_3$, $\Sigma\, x_2 \cdot x_3$ og $\Sigma\, x_2 \cdot \Sigma\, x_3$ får man:

$$\bar{b} = \frac{b\,(n\,\Sigma\,x_3^2 - (\Sigma\,x_3)^2) + (n\,\Sigma\,x_3\,u - \Sigma\,x_3\,\Sigma\,u)}{(n\,\Sigma\,x_3^2 - (\Sigma\,x_3)^2) - (n\,\Sigma\,x_3\,u - \Sigma\,x_3\,\Sigma u)}$$

Da som før $E\,(x_3 \cdot u) = O$, får man:

$$\underset{n \to \infty}{plim}\;\bar{b} = \frac{b\,(n\,\Sigma\,x_3^2 - (\Sigma\,x_3)^2)}{n\,\Sigma\,x_3^2 - (\Sigma\,x_3)^2} = b, \tag{24}$$

hvorved er bevist, at når man estimerer med hele modellen som forudsætning, får man et konsistent estimat.

Hvad her er anført kan måske siges at være et specielt resultat af Markoff's teorem, som helt generelt præciserer under hvilke omstændigheder man kan forvente at få et »rimeligt« estimat ved at anvende mindste kvadraters metode på lineære relationer, hvilket jo kan siges at være det generelle problem i det foregående. Det vil imidlertid føre alt for vidt at gå ind på de mere principielle problemer heri[1]).

Det kan bemærkes, at benyttelsen af det simultane princip ved estimeringer er langt mere arbejdskrævende end det almindeligt anvendte princip med at tage strukturrelationerne en for en direkte. Netop i økonomien, hvor simultane sammenhænge er så helt dominerende, bevirker dette forhold, at det må være berettiget at sige, at socialvidenskaberne støder på usædvanlig store vanskeligheder i forsøgene på at fundere sig empirisk sammenlignet med mange naturvidenskaber, hvor isolering af objektet er mulig. Konsekvensen af dette kan imidlertid kun blive et krav om forøgede anstrengelser.

[1]) Der kan bl. a. henvises til en artikel af F. N. David og J. Neyman i *Statistical Research Memoirs II*, London 1938, »Extension of the Markoff theorem on least squares«.