

G. van den Dijk

DUPLICAAT

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953 - 30(1)

Enkele statistische problemen in
de econometrie

Eerste Voordracht, gehouden op de Werkbespreking
van 10 April 1953 door

J.A.Hartog.

MATHEMATISCH CENTRUM
Statistische Afdeling

1. Inleiding.

Voorbeeld van een econometrisch model. [1]

Geven we de geconsumeerde hoeveelheid aan met x_1 , de prijs met x_2 , de loonvoet (= het arbeidersinkomen) met x_3 en het consumenteninkomen met x_4 , dan kunnen b.v. de volgende relaties bestaan:

$$\underline{x}_1 + \alpha_{s2} \underline{x}_2 + \alpha_{s3} \underline{x}_3 + \beta_s = \underline{u}_s \quad (1)$$

$$\underline{x}_1 + \alpha_{d2} \underline{x}_2 + \alpha_{d4} \underline{x}_4 + \beta_d = \underline{u}_d \quad (2)$$

(1) is een aanbodvergelijking, (2) een vraagvergelijking: de constanten in (1) geven we daarom de index s (= supply), die in (2) de index d (= demand). We onderstellen, dat x_3 en x_4 uit andere vergelijkingen, buiten dit stelsel kunnen worden bepaald en hier dus bekend zijn. $\underline{u}_s$ en $\underline{u}_d$ zijn storingstermen, b.v. onderling onafhankelijk normaal verdeeld met gemiddelde nul en spreiding σ_s resp. σ_d . β_s en β_d zijn constanten.

Doen we een aantal o. o. waarnemingen, dan is het mogelijk de α 's te schatten. Uit overwegingen van economische aard is het plausibel, dat geldt:

$$\alpha_{s2} < 0, \alpha_{d4} < 0, \alpha_{s3} > 0, \alpha_{d2} > 0 \quad (3)$$

Hierin ligt een controlemiddel voor het al of niet bruikbaar zijn van ons model. Het is echter niet bevredigend, dat van (3) slechts op deze indirecte wijze gebruik kan worden gemaakt.

Doel van econometrische modellen. [2], [3].

De vergelijkingen, die het verband aangeven tussen beschouwde variabelen, samen met de a priori veronderstellingen omtrent de verdelingen van $\underline{u}_s$ en $\underline{u}_d$ en ongelijkheden van de vorm (3) noemt men in de economie het model. Geven we de α 's en de parameters in de verdelingsfuncties van $\underline{u}_s$ en $\underline{u}_d$ bepaalde waarden, dan ontstaat een structuur. Het model bestaat dus uit een klasse van structuren; we veronderstellen, dat deze de "ware" structuur bevat (de "maintained hypothesis"). Slagen we er in deze te vinden, dan kunnen we voorspellingen doen omtrent de ontwikkeling in de toekomst, en wel in twee vormen:

- routinevoorspellingen, b.v. over de ontwikkeling van de prijzen x_2 op basis van het verloop van x_3 en x_4 , als de structuur gehandhaafd blijft,
- structurele voorspellingen, dat zijn prognoses omtrent de gevolgen van wijzigingen in de bestaande structuur.

2. Enkele kenmerken van econometrische modellen.

1. In de meeste gevallen worden ze geformuleerd in de vorm van geaggregeerde variabelen. [4], [5], [6].

Niet het verband tussen de verschillende grootheden voor een bepaald individu en een bepaald artikel wordt beschreven, doch het gedrag van bevolkingen en veelal het verband tussen vraag en prijs voor grote groepen artikelen, of de totale vraag van alle consumenten naar alle artikelen tezamen. Gemiddelde prijzen voor groepen consumptiegoederen vinden we door het toekennen van gewichten (= indexcijfers) aan de afzonderlijke prijzen; helaas is het meestal niet gerechtvaardigd in verschillende vergelijkingen dezelfde indexcijfers te gebruiken. De vergelijkingen geldend voor de individuen afzonderlijk trachten we samen te nemen om een a priori inzicht te krijgen in het gedrag van de gemeenschap. Een voor ons doel bruikbare theorie van de massa-psychologie ontbreekt.

De beginvergelijkingen voor de individuen ontleent men aan de wiskundige economie, waar men er van uitgaat, dat de mensen rationeel handelen. Daar dit allerm minst zeker is wil Prof. v. Dantzig de geaggregeerde vergelijkingen rechtstreeks formuleren op grond van statistisch materiaal. De Hartog meent echter dat het aantal waarnemingen beperkt is (in Nederland b.v. is het uiterste, wat men op het ogenblik bereiken kan, kwartaalcijfers over een zeer beperkte periode). Hier komt bij, dat de reeksen in de economie gebruikt, meestal een cyclisch karakter hebben, zodat hoge correlaties geen aanwijzing behoeven te zijn voor de juistheid van de veronderstelde relatie. Als gevolg hiervan is men niet in staat op basis van de waarnemingen alleen een structuur te formuleren, waarvan de relaties een min of meer permanent karakter dragen.

2. Onderscheid tussen verschillende soorten variabelen.

[7], [8].

Variabelen, die buiten het stelsel vergelijkingen worden bepaald heten exogeen (in (1) en (2) x_3 en x_4), de andere endogeen. Om het stelsel op te kunnen lossen, moet het aantal vergelijkingen gelijk zijn aan het aantal endogene variabelen; soms worden daarom variabelen te lichtvaardig als exogeen beschouwd [7].

Een variabele, die men in sterke mate in de hand heeft heet beheersbaar (controlable), b.v. de loonvoet; een grootheid waarop men geen invloed kan uitoefenen heet niet-beheersbaar (uncontrolable), b.v. de invloed van het weer.

De beheersbare exogene variabelen noemt men instrumental, de niet-beheersbare exogene en de endogene non-instrumental.

3. Het model is niet constant in de tijd.

Zowel uit waarnemingen als uit theoretische overwegingen blijkt, dat de factoren, die invloed uitoefenen, niet steeds dezelfde zijn. Het model moet daarom steeds aangepast worden, teneinde bruikbare voorspellingen te kunnen doen, zodat het aantal waarnemingen betreffende een model steeds beperkt moet blijven.

4. Het naar voren komen van het steekproefelement. [9].

T.g.v. het niet steeds rationele gedrag van de mensen, de aanwezigheid van meetfouten en het niet opnemen van alle variabelen, die invloed uitoefenen, kunnen we de constanten uit de regressievergelijkingen niet zonder meer oplossen, maar moeten we ze schatten uit verschillende waarnemingen.

Dit komt tot uitdrukking in

- α) het opnemen van een "error term" u , die geïnterpreteerd wordt, hetzij als een aggregaat van verwaarloosde variabelen (opvatting van de Cowles-Commission), hetzij als een meetfout in de te verklaren variabele (opvatting van R.A.Fisher).
- β) het postuleren van meetfouten in alle variabelen, onafhankelijke en afhankelijke (methode van Frisch).

5. Bijzonder karakter van het waarnemingsmateriaal in de econometrie. [10].

Terwijl men bij biologische en agrarische onderzoeken vaak willekeurig kan experimenteren, is dit in de economie niet mogelijk. De econoom beschikt slechts over een beperkt aantal waarnemingen, waarin bovendien de verschillende variabelen niet onderling onafhankelijk gevarieerd zijn. Hij beschouwt zijn materiaal meestal als een steekproef uit een bepaald universum (in 4^{α} b.v. een stabiel universum van "errors"). De idee van "herhaalde steekproeven" wordt onder deze omstandigheden uiteraard zeer hypothetisch, evenals de verwante begrippen als steekproefverdeling en standaardfouten van de steekproeffuncties.

3. Schattingsmethoden.

I. Methode van de kleinste kwadraten, toegepast op iedere vergelijking afzonderlijk.

a) Behandeling volgens R.A.Fisher.

Stel in de regressievergelijking komen k variabelen voor, waarvan we σ, σ waarnemingen hebben:
 $X_1^{(t)}, \dots, X_k^{(t)}$ ($t = 1, \dots, T$). Alleen in $X_1^{(t)}$ bevinden zich meetfouten $u^{(t)}$, die we $N(0, \sigma)$ verdeeld onderstellen. We vinden:

$$X_1^{(t)} = \xi^{(t)} + u^{(t)}$$

$$X_{k'}^{(t)} = \xi_{k'}^{(t)} \quad (k' = 2, \dots, k)$$

$$\sum_{t=1}^T \alpha_j \xi_j^{(t)} - \alpha = 0 \quad (\alpha_j \neq 0)$$

Met de maximum-likelihood methode vinden we schattingen $\hat{\alpha}_j, \hat{\alpha}$ en $\hat{\sigma}$ voor resp. α_j, α en σ . Eenvoudige functies van $\hat{\alpha}_j$ en $\hat{\alpha}$ zijn exact verdeeld volgens Student; alle schattingen zijn asymptotisch normaal verdeeld.

b) Behandeling volgens Markoff [11].

Het gebruik van de methode der kleinste kwadraten kan ook worden verdedigd op grond van het Markoff-theorema. Volgens dit theorema zijn minder beperkende veronderstellingen nodig dan in a) werden gemaakt.

II. Tegen Fisher's behandeling zijn drie bezwaren aangevoerd.

- er kan autocorrelatie bestaan, d.w.z. dat de reeksen waarnemingen van de variabelen die in de regressievergelijking voorkomen, geen toevallige (random) reeksen zijn.
- de aanwezigheid van meerdere relaties tussen de variabelen
- meetfouten in de onafhankelijke variabelen.

Frisch heeft critiek uitgeoefend t.o.v. de punten b) en c); hij construeerde een model, met meetfouten in alle grootheden en meerdere relaties en paste hierop toe de methode van Fisher, d.w.z. met een regressievergelijking, alsof er geen andere relaties bestonden. Dit leidde niet tot verwerping van deze vergelijking, m.a.w. Fisher's standaardfout is geen bruikbare toetssten ter beoordeling van de betrouwbaarheid der regressiecoëfficiënten als

aanwijzing voor een systematisch verband.

Wanneer we zowel b) als c) onderstellen, ontstaat de volgende moeilijkheid. Indien tussen k variabelen een lineair verband bestaat zullen de gevonden waarnemingen een puntenwolk vormen in de omgeving van een $(k-1)$ -dimensionaal hypervlak (s_{k-1}). Bestaat er nog een tweede lineaire relatie, dan vormen de punten een wolk in de omgeving van de doorsnee s_{k-2} van de twee varieteiten s_{k-1} . Bij een gegeven puntenwolk is het echter moeilijk te bepalen, in hoeverre de dimensie het gevolg is van meetfouten en in hoeverre van het bestaan van simultane lineaire relaties.

Prof.v.Dantzig vestigt in verband hiermee, de aandacht op een artikel [12], waarin het begrip rang van een matrix wordt gegeneraliseerd, als volgt:

zij $A \gg 0$ en $\varepsilon > 0$, dan is het gehele getal $r(A, \varepsilon)$ rang van een matrix, indien deze een vierkante matrix bevat van de orde r , waarvan de determinant een absolute waarde heeft $\gg A$, en indien iedere vierkante matrix van de orde $r+1$ een determinant heeft met absolute waarde $\leq \varepsilon$.

Verder wordt een benaderde oplossing voor een stelsel van homogene vergelijkingen gedefinieerd, die wellicht van nut is bij het bestuderen van vergelijkingen, waarin de coëfficiënten niet exact bepaald zijn.

Literatuurlijst.

1. Bibliography of Applications of Mathematical Statistics to Economics, 1943-1949, J.R.S.S. A III, 1951.
2. J.Tinbergen, On the Theory of economic Policy.
3. H. Theil, De betrouwbaarheid van model-voorspellingen. C.P.B.afdeling II, notitie no.27.
4. W.Dresch, Index numbers and the general economic equilibrium, (Bull.Math.Soc.1938).
5. Cowles Commission Paper met artikelen van Klein, Kenneth May and Pu.
6. H.Theil, Remarks on Optimal Aggregation (W.E.K.no.20). A statistical Approach to the Problem of Aggregation.
7. Orcut, Koopmans, Tinbergen e.a., Toward Partial Re-direction of Econometrics. (Review of Economics and Statistics (Aug.1952)).
8. H.A.Simon, The causal Principle and the Identification Problem (Cowles Commission Discussion Paper; Statistics no.353).
9. R.Frisch, Statistical Confluence Analysis by means of complete regression systems.
10. Tj.Koopmans, Linear Regression Analysis of Economic Time Series.
11. F.N.David and J.Neyman, Extension of the Markoff Theorem on least squares (Stat.Research Memoirs, vol.II, 1938).
12. N.G.de Bruyn and D.van Dantzig, Inequalities concerning determinants and systems of linear equations.(Kon.Ned.Ac.van Wetenschappen, Proceedings LV, no.3, A).

G. Hemelrijk

DUPLICAAT

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953 - 30(1)

Enkele statistische problemen in de econometrie.

Tweede voordracht, gehouden op de werkbijeenkomst

van 1 Mei 1953

door

J. A. Hartog

Naar aanleiding van de behandelde critiek op het model van Fisher informeert Prof. Van Dantzig naar de relatieve grootte van de meetfouten in de verschillende variabelen. In het algemeen zijn de fouten zeer groot, zelfs bij grootheden als b.v. de productie van steenkool in een bepaald land. Hierbij komt dat de definities niet vast staan. Men moet kiezen uit verschillende indices; indien deze zich t.g.v. wijzigingen in de exogene variabelen niet gelijk gedragen is het niet mogelijk conclusies te trekken omtrent de invloed van de wijzigingen. Een voorbeeld hiervan is de invloed van de wisselkoers op de import; bij enige Zuid-Amerikaanse landen bleek het teken van de betreffende regressiecoëfficiënt af te hangen van de gebruikte indexformule. Bovendien is vaak de ontwikkeling in de afzonderlijke sectoren belangrijker dan die van het totaal. Een nadere toelichting van de orde van grootte van meetfouten in economische gegevens vindt men in een boekje van MORGENSTERN [13].

Het werk van FRISCH leidde tot een publicatie van T.HAAVELMO [14], welk artikel de basis werd van de studies van de Cowles Commission for Research in Economics (leden o.a. A.COWLES, T.C.KOOPMANS, W.B.SIMPSON). Men tracht de onzuiverheid van de maximum likelihood schattingen der regressiecoëfficiënten te verwijderen door de aanwezigheid van meer lineaire relaties te onderstellen ([15], [16]). De andere tegen de theorie van Fisher ingebrachte bezwaren worden echter niet ondervangen, terwijl men door het grote aantal onderstellingen (b.v. neemt men voor de afleiding van de formules steeds aan dat alle storingen normaal verdeeld zijn met gemiddelde nul en constante spreiding) modellen krijgt, die slechts zeer beperkte geldigheid bezitten. Bovendien is het rekenwerk bij deze methode zeer omvangrijk.

III. Methode van H.Wold.

WOLD merkte op dat niet steeds onzuivere schattingen ontstaan, wanneer we de onbekende grootheden niet simultaan schatten. Hij ging er van uit dat het economisch proces te beschrijven was door een model van de volgende vorm:

$$\alpha_{11} \underline{X}_1 + \sum_i \beta_{1i} z_i = \underline{u}_1 \quad (1)$$

$$\alpha_{21} \underline{X}_1 + \alpha_{22} \underline{X}_2 + \sum_i \beta_{2i} z_i = \underline{u}_2 \quad (2)$$

$$\alpha_{31} \underline{X}_1 + \alpha_{32} \underline{X}_2 + \alpha_{33} \underline{X}_3 + \sum_i \beta_{3i} z_i = \underline{u}_3 \quad (3)$$

enz.

Hierin zijn $\underline{X}_1, \underline{X}_2, \dots$ exact gemeten grootheden en de z_i exogene variabelen. De $\underline{X}_i$ zijn stochastisch t.g.v. de $\underline{u}_i$, die

de verwaarloosde variabelen voorstellen. De fluctuaties in X_1 worden alleen veroorzaakt door die van u_1 , die in X_2 door de variaties in X_1 en u_2 , enz. We schatten nu uit (1) m.b.v. de methode der kleinste kwadraten de β_{12} ; vervolgens uit (2) op dezelfde wijze de β_{21} ; enz. Indien we onderstellen:

- a) de u_i zijn onderling onafhankelijk verdeeld,
- b) u_j is onafhankelijk van $X_1, \dots, X_{j-1}$ en van de exogene variabelen,

dan zijn deze schattingen van de β 's zuiver. Zodra we aannemen, dat de u_i gecorreleerd zijn met de verklarende variabelen worden de schattingen van de regressiecoëfficiënten onzuiver. Autocorrelatie in de storingen beïnvloedt de zuiverheid van de schattingen niet, maar wel de standaardfout in de regressiecoëfficiënten. Wold neemt hiervoor dan ook correcties op in zijn formules.

Indien $u_1, u_2, \dots$ in het algemene model met simultane vergelijkingen simultaan normaal verdeeld zijn, dan kan men dit inderdaad in de vorm (1), (2), (3) gieten met onderling onafhankelijke u_i .

Het volgende voorbeeld verduidelijkt dit. Zijn de aanbod- en de vraagvergelijking betreffende consumptiejs resp.

$$\underline{X}_1 + \alpha_{s2} \underline{X}_2 + \alpha_{j3} X_3 = \underline{u}_1 \quad (4)$$

$$\underline{X}_1 + \alpha_{p2} \underline{X}_2 + \alpha_{p4} X_4 = \underline{u}_2, \quad (5)$$

dan bevatten zowel u_1 , als u_2 de invloed van het weer en zijn dus afhankelijk. Door (4) vermenigvuldigd met een constante op te tellen bij (5) kunnen we de afhankelijkheid in u_1 en u_2 elimineren, terwijl (5) toch zijn oorspronkelijke vorm behoudt.

IV. Methode van T.KOOPMANS [10].

De critiek van Frisch (II c) was voor KOOPMANS aanleiding om een methode te ontwikkelen, waarin rekening wordt gehouden met meetfouten in alle variabelen en die gebaseerd was op VAN UVENS gewogen regressie.

We bepalen in de k-dimensionale waarnemingsruimte een hypervlak, z.d.d. de som van de kwadraten van de afstanden der waarnemingspunten tot dat vlak minimaal wordt (orthogonale regressie). In de behandeling van M.J.VAN UVEN wordt de afstanden een a priori gegeven gewicht ϵ toegekend, waarvoor hier de variantie van de willekeurige components u_i in de waarneming X_i wordt gekozen. Koopmans berekent een benadering van de steekproefverdeling van de gewogen regressiecoëfficiënten en geeft grenzen aan voor de gevolgen van fouten in de toegekende gewichten.

V. In alle voorafgaande methoden is met de belangrijkste moeilijkheid nog geen rekening gehouden, n.l. met het tijdsprobleem: de opeenvolgende reeksen waarnemingen vertonen autocorrelatie [10, § 6].

J. TINBERGEN heeft modellen geconstrueerd, waarin de autoregressie betrokken werd. G.H. ORCUTT onderzocht in hoeverre alle door Tinbergen gebruikte reeksen voorgesteld konden worden als trekkingen uit één universum van een stochastische grootheid met een autoregressieve structuur, n.l.

$$\underline{Y}_t = \underline{Y}_{t-1} + 0,3 (\underline{Y}_{t-1} - \underline{Y}_{t-2}) + \underline{\varepsilon}_t ;$$

hierin is $\underline{\varepsilon}_t$ een stochastische variabele met gemiddelde nul. Hij slaagde er niet in deze structuur te verwerpen [17].

Het is moeilijk te bepalen voor hoeveel onafhankelijke waarnemingen we een tijdreeks kunnen tellen, zolang we het cyclisch karakter van de reeks niet nauwkeurig kennen. In de economie werkt men veelal met reeksen van ongeveer twee cycli.

Ook in de storingstermen kan autocorrelatie aanwezig zijn ([18], [19]). Hoewel FISHER deze zonder meer verwaarloost, is dit zeer plausibel als we bedenken dat in de storingstermen ook verwaarloosde economische factoren opgenomen zijn. In dergelijke gevallen kunnen we de storingen zeker niet opvatten als onderling onafhankelijke trekkingen uit een universum van "errors". WOLD tracht daarom conclusies te trekken uit correlogrammen, dat zijn grafieken, die het verloop van de autocorrelatie-coëfficiënt in de tijd aangeven. In deze richting is de laatste tijd veel onderzoek verricht ([20], [21]).

De voordelen van de methode van TINBERGEN bestaan uit:

- a) het kleinste aantal onderstellingen dat hij maakt omtrent zijn model; door het beperkte empirische materiaal is er van de eigenschappen der storingstermen weinig bekend, zodat alle onderstellingen daarover in de lucht hangen.
- b) de gemakkelijke wijze waarop er mee kan worden geëxperimenteerd; het is niet moeilijk om allerlei variabelen in de vergelijking op te nemen teneinde na te gaan in hoeverre deze de uitkomsten beïnvloeden.

Doordat niet simultaan wordt geschat en niet steeds dezelfde grootheden als afhankelijke resp. onafhankelijke veranderliken worden genomen vormt zijn techniek geen logisch gesloten geheel; soms worden verschillende antwoorden gevonden.

Helaas bestaat er nog geen statistische methode welke met alle drie onder III genoemde bezwaren rekening houdt. Het toetsen van de resultaten heeft dan ook tot droevige conclusies aanleiding gegeven; zie b.v. publicaties van CARL CHRIST [22] en

D.J.MORGAN en W.J.CORLETT [23].

Vervolg literatuurlijst.

13. O.MORGENSTERN, On the accuracy of economic observations, Princeton University Press, 1950.
14. T.HAAVELMO, The probability approach in econometrics, Suppl. Econometrica 12 (1944) 1-115.
15. T.KOOPMANS, Statistical estimation of simultaneous economic relations, J.A.S.A. 40 (1945) 448-466 en Cowles Commission Papers, New Series, no 11.
16. C.P.B. W.C.K. no 21, Limited information; Maximum likelihood methods.
17. G.H.ORCUTT, A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919-1932, J.R.S.S. B X (1948) 1-53.
18. D.COCHRANE and G.H.ORCUTT, Application of least squares regression to relationships containing autocorrelated error terms, J.A.S.A. 44 (1949) 32-61.
19. D.COCHRANE and G.H.ORCUTT, A sampling study of the merits of autoregressive and reduced form transformations in regression analysis, J.A.S.A. 44 (1949) 356-372.
20. H.WOLD, A study in the analysis of stationary time series, Uppsala, Almqvist and Wiksells (1938).
21. H.WOLD and L.JURÉEN, Demand analysis, John Wiley and Sons (1953).
22. C.CHRIST, A test of an econometric model for the United States, Cowles Commission Papers, New Series, no 49 (1952).
23. D.J.MORGAN and W.J.CORLETT, The influence of price in international trade, J.R.S.S. CXIV, 3 (1951) 307-358.

DUPLICAAT

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953 - 30(1)

Enkele statistische problemen in de econometrie.

Derde voordracht, gehouden op de werkbijeenkomst
van 15 Mei 1953

door

Prof. Dr J. Tinbergen

Knelpunten in de econometrie.

De leidraad van deze voordracht is de vraag wat de mathe-
maten kunnen doen in de econometrie naast het werk dat de
economisten bezig houdt.

Gedurende de laatste decennia zijn er door het statistisch
toetsen van verschillende hypothesen via de wiskundige eco-
nomie veel vermoedens naar voren gekomen omtrent de structuur
van het economisch gebeuren. Na een aanvankelijk enthousiasme
ontstond echter kritiek van twee zijden, n.l. zowel van de kant
der statistici als van die der economisten.

C.Christ toetste de economische fluctuaties in de Verenig-
de Staten tussen 1945 en 1950 aan een door L.R.Klein opgesteld
model [24], [25]. Hierbij bleek echter dat voorspellingen in de
vorm "alles blijft gelijk" of "we trekken alle variabelen
zonder meer recht door", d.w.z. $x_{t+1} = x_t + (x_t - x_{t-1})$ zeker
niet slechter met de waarnemingen overeenstemden.

Aan de fundamentele relatie

$$C = \frac{1}{2} Y$$

waarin C voorstelt de consumptieve uitgaven en Y het nationaal
inkomen, beide gemeten als verschil met hun gemiddelde, was in
1951 plotseling niet meer voldaan, terwijl er lange tijd, ook
gedurende depressies, geen belangrijke afwijkingen gevonden
waren. De voor de hand liggende oorzaak is het uitbreken van de
vijandelijkheden in Korea, doch geheel verklaard is de vershui-
ving zeker niet. In ieder geval illustreert het dat de "psycho-
logische" factoren die een rol spelen niet in voldoende mate
bekend zijn.

De onderzoekingen van de dynamische modellen die de econo-
mische beweging in de tijd moeten beschrijven zijn na 1940 niet
voortgezet. Enerzijds omdat de oorlog de normale functionering
van het economische leven uitschakelde en anderzijds t.g.v.
een zeker scepticisme richtte men zich meer op theoretisch dan
op empirisch onderzoek. Eerst nu de volkshuishoudingen weer min
of meer regelmatig functioneren en men zich zorg gaat maken
over de conjunctuur-ontwikkeling in de toekomst, wil men trach-
ten dit werk weer op gang te brengen. Hierbij is de dienst van
de wiskundige wel nodig, doch niet in de eerste plaats.

Bezit men modellen dan kunnen voorspellingen worden gedaan
over de toekomstige ontwikkeling. Doordat niemand kan preten-
deren alle van belang zijnde factoren te kennen en in het han-
delen van de mensen een vrij element aanwezig is wordt de zin
hiervan door sommigen betwijfeld. In elk geval is er een vrij
grote toevallige components in alle historische economische
reeksen. Belangrijker is dan ook het nagaan van de gevolgen, die
bepaalde wijzigingen in de economische politiek kunnen hebben;

de toevallige component valt hier grotendeels weg.

De knelpunten bij het onderzoek kunnen in de volgende vijf groepen worden verdeeld:

1. Ontwikkeling van de economische theorie.

Er is twijfel gerezen over de mogelijkheid de economie door een bepaald gedetermineerd proces te kunnen voorstellen. Prof. Goudriaan wil daarom drie soorten verbanden toelaten, n.l.

- a) volkomen gedetermineerde,
- b) stochastische,
- c) onbepaalde, die dus zelfs m.b.v. de waarschijnlijkheidsrekening niet te beschrijven zijn.

Bij deze studie van de economische theorie is enige steun van wiskundige zijde zeker nodig.

2. Verzamelen van concrete statistische gegevens.

Tot nu toe zijn vrijwel geen gegevens beschikbaar over voorraden, orderportefeuilles, kosten en productiecapaciteit van de ondernemingen. Ook aangaan de reacties op belastingverlaging en inkomenwijziging, toekomstverwachtingen, e.d. tast men ondanks de enquêtes van het N.I.P.O. nog in het duister. Bovendien is hierover niet veel meer te achterhalen voor de jaren vóór 1940.

3. Concretiseren van residuen.

De van voor 1940 aanwezige gegevens zijn nog niet voldoende onderzocht. Het is zeer goed mogelijk dat de residuen, die men toen overhield nog relevante grootheden bevatten. Men kan dit nagaan door aan de gebruikte verklarende grootheden nog enige toe te voegen, vervolgens de regressiecoëfficiënten opnieuw te bepalen en dan de overblijvende residuen te vergelijken met de oorspronkelijke. Deze studie vereist een uitgebreide zoek- en rekenarbeid.

4. Aggregatieproblemen.

Doordat men meestal werkt met gemiddelde grootheden ontstaat het probleem van het toekennen van gewichten. Een deel van de residuen is vermoedelijk het gevolg van onjuiste aggregatie van b.v. prijzen en consumpties. Theil tracht dit probleem via nieuwe methoden op te lossen [6].

5. Onderzoek van de mathematisch statistische methoden.

Hiervoor wordt naar de voordrachten van J.A.Hartog verwezen.

In de discussie informeert Prof. Van Dantzig naar de mogelijkheid de vertragende intervallen in de autoregressieve reeksen als stochastische variabelen op te vatten. Wellicht zijn bepaalde residuen niet te elimineren, zolang de intervallen con-

stant gehouden worden. Prof. Tinbergen wijst er op dat vergelijkingen in deze vorm onder zekere voorwaarden te herleiden zijn tot regressievergelijkingen met stochastische coëfficiënten, waardoor ze mathematisch beter hanteerbaar worden.

Verder komt nog een artikel van M.G.Kendal ter sprake [26].

Vervolg literatuurlijst.

24. L.R.Klein, Economic fluctuations in the United States, 1921-1941, Cowles Commission for research in economics, Monograph no 11.
25. C.Christ, A test of an econometric model for the United States, 1921-1947, Cowles Commission Papers, New Series, no 49 (1952).
26. M.G.Kendall, The analysis of economic time series, Part I, Prices J.R.S.S. serie A, CXVI 1 (1953) 11-34.

D. 30 10
DUPLICAAT 10

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953 - 30(1) D

Enkele statistische problemen in de econometrie

Vierde voordracht, gehouden op de werkbijeenkomst
van 29 Mei 1953 door

J.A.Hartog.

Vergelijking van de methoden gebruikt in Chicago, in Rotterdam en in Uppsala ¹⁾.

We gaan uit van het volgende model:

$$C y(t) + B x(t) + z(t) = 0 \quad (1)$$

Hierin zijn C en B matrices van onbekende coëfficiënten, C vierkant, B rechthoekig; $y(t)$, $x(t)$ en $z(t)$ zijn kolomvectoren: $y(t)$ bestaat uit endogene, $x(t)$ uit exogene en vertraagde endogene variabelen; de $z(t)$ zijn storingstermen, veroorzaakt door verwaarloosde factoren. We onderstellen, dat alle variabelen zonder meetfouten zijn. De verdelingsfuncties van de z_i zijn functies van de tijd, die min of meer gespecificeerd zijn. De verdelingen van de y_i zijn vastgelegd door het stelsel (1) en de verdelingen van de z_i .

De Cowles-Commission tracht een dergelijk stelsel simultaan te schatten. Daarentegen licht men er in Rotterdam veelal één vergelijking uit, waarmee men verder werkt onder verwaarlozing van de andere bestaande relaties. Verschillende van de endogene variabelen kunnen dan afzonderlijk als afhankelijke grootheden beschouwd worden, waarbij de vergelijking de volgende vorm krijgt:

$$y_n(t) = \sum_{i \neq n} c_i y_i(t) + \sum_j b_j x_j(t) + z_n(t) \quad (2)$$

Op het tijdstip t schat men de regressiecoëfficiënten; Chicago in principe de matrices C en B uit (1), Rotterdam de c_i en b_j uit (2). De x_j zijn dan parameters, bekend uit een ander model als ze exogeen zijn, uit vroegere waarnemingen, indien ze vertraagd endogeen zijn. Aangezien we van alle variabelen y_i en x_j voor $t = t_1$ slechts één waarneming bezitten en we hieruit alleen niet kunnen schatten laten we een steekproef bestaan uit de waarnemingen op de tijdstippen $t = 1, \dots, T$ van de grootheden y_i en x_j . Deze behoren dan wel bij verschillende modellen, maar indien we het stochastische proces ergodisch onderstellen, d.w.z. dat de verwachtingen van de y_i constant zijn in de tijd, kunnen we ze toch gebruiken om de verwachtingen van de y_i op het tijdstip $t = t_1$ te schatten.

In de meeste gevallen is het schatten van de hele matrices C en B te gecompliceerd om uit te voeren. Ook de Cowles-Commission beperkt zich dan tot één vergelijking, maar blijft in tegenstelling tot Rotterdam rekening houden met de a priori

1) Stochastische grootheden worden door onderstreping aangegeven; vaak gebruikt met hiervoor Griekse letters.

restricties, die opgesloten liggen in de overige vergelijkingen van het model.

Rotterdam schat nu met de kleinste kwadraten methode zonder na te gaan aan welke voorwaarden de variabelen moeten voldoen opdat dit gerechtvaardigd is. Wold heeft hiervoor de eisen geformuleerd.

Zowel hij als de Cowles-Commission onderstellen ²⁾:

- 1) de storingstermen $\underline{z}_i$ bezitten een waarschijnlijkheidsverdeling, die een functie is van de tijd.
- 2) de grootheden x zijn bekende functies van de tijd, dus in de vergelijkingen parameters.
- 3) $\sum_{t=1}^n \underline{z}_t = 0$
 $\sum_{t=1}^n \underline{z}_t^2$ ($t=1, \dots, n$) is eindig en gelijk voor alle waarden van n .
 $\sum_{t=1}^n \underline{z}_t (t=1, \dots, n) \cdot \underline{z}_t (t=1, \dots, n) = 0$.
- 4) in iedere vergelijking is de storing ongecorreleerd met de vertraagde endogene variabelen.

Verder neemt de Cowles Commission aan:

- 5) de $\underline{z}_i$ zijn normaal verdeeld met gemiddelde nul en constante spreiding.

Wold specificeert de verdeling van de $\underline{z}_i$ niet volledig, maar eist nog wel:

- 6) de verdelingsfuncties van de $\underline{z}_i$ zijn onafhankelijk van de x_j , voor zover dit exogene variabelen zijn (voorwaarde 5) impliceert dit),
- 7) $\underline{z}_n$ is ongecorreleerd met de y_i voor $i < n$.

De schattingen van de Cowles Commission hebben bekende verdelingen, zowel voor grote als voor kleine steekproeven. Die van Rotterdam zijn zuiver, bruikbaar en de asymptotisch doeltreffendste van alle lineaire schattingen; bij kleine steekproeven worden de spreidingen intuïtief gewaardeerd.

Zij $\underline{z}'(t)$ de getransponeerde van de vector $\underline{z}$ uit (1) en definiëren we $\sum_{t=1}^n \underline{z}'(t) \cdot \underline{z}(t) = Z$, dan zijn de uitkomsten van de schattingen in beide gevallen vrijwel dezelfde, indien $|C^{-1} Z C^{-1}|$ klein is.

Een voorbeeld, waarin het verschil tussen de schattingen niet significant is, werd uitgewerkt door een lid van de Cowles Commission.

Hierin werden de regressiecoëfficiënten van de vergelijking

$$\underline{c} = \alpha \underline{y} + \beta + \underline{z} \quad (3)$$

geschat, als bovendien geldt

$$s = \underline{y} - \underline{c} \quad (4)$$

-
- 2) de genoemde voorwaarden vormen geen volledig en uitsluitend stelsel eisen.

Hierin is

y = beschikbaar reëel inkomen per hoofd

c = reële consumptie per hoofd

s = investering plus besparing per hoofd.

s beschouwen we als een exogene, y en c als endogene variabelen. De matrix van de laatste is

$$\begin{vmatrix} -1 & +1 \\ \alpha & -1 \end{vmatrix}.$$

Deze is niet triangulair, zodat de methode van Wold niet kan worden toegepast.

Schatten we nu toch α en β alsof alleen (3) bestond, dan vinden we:

$$\alpha = 0,732, \quad \beta = 84,0$$

Houden we rekening met (4) en elimineren we y , dan gaat (3) over in:

$$c = \frac{\alpha}{1-\alpha} \cdot s + \frac{\beta}{1-\alpha} + \frac{z}{1-\alpha} \quad (5)$$

Schatten we hier de regressiecoëfficiënten, dan zijn de uitkomsten:

$$\alpha = 0,672, \quad \beta = 113,1$$

Bij beide methoden leveren kleinste kwadraten en maximum likelihood schatting dezelfde resultaten. In het laatste geval moet de verdelingsfunctie van z op hoogstens enige parameters na gespecificeerd zijn. Als dit zo is, dan kunnen we ook betrouwbaarheidsintervallen opgeven voor onze schattingen. Bij de Cowles Commission-methode vinden we dan voor α

$$0,57 \leq \alpha \leq 0,73 \quad \text{vs } 0,05,$$

dus een niet significant verschil van de waarde 0,732.

Wold accepteert dit model niet, aangezien hij op economische gronden meent, steeds recursieve modellen te kunnen vinden. Prof. Tinbergen zou hier ook uit (3) en (4) samen schatten, maar dan zo mogelijk met vertraagde y in (3), welke vergelijking dan overgaat in

$$c(t) = \alpha y(t-1) + \beta + z(t) \quad (3')$$

Het model (3'), (4) is nu recursief geworden, waardoor de methode van Wold kan worden gebruikt.

Resumerend kunnen we de verschillende opvattingen als volgt formuleren:

Rotterdam maakt geen onderstellingen omtrent de vorm van de coëfficiëntenmatrix C . Chicago eist, dat zoveel elementen van nul zijn, dat het model identificeerbaar is, terwijl Uppsala aan-

neemt, dat C triangulair is. Heeft C zeker de triangulaire vorm, dan is de oplossing van de Cowles Commission identiek aan die van Wold.

In gevallen waarin naar de mening van Chicago en Rotterdam geen recursief model te vinden is, beschouwt de eerste het gehele stelsel simultaan. Hieraan kleeft het bezwaar, dat vaak moeilijk te meten grootheden opgenomen moeten worden; toch gaat men er van uit, dat de variabelen geen meetfouten hebben, waardoor hun schattingen van de regressiecoëfficiënten onzuiver zijn. Rotterdam licht daarom een bepaalde vergelijking uit het stelsel en schat de coëfficiënten met verwaarlozing van de andere bestaande relaties; met de invloed van sommige factoren wordt dus a priori geen rekening gehouden. De Cowles Commission zoekt naar wegen, waarbij aan de Rotterdamse bezwaren gedeeltelijk tegemoet wordt gekomen.