MATHEMATISCH CENTRUM

Statistische Afdeling
Rapport S 142

1 Juni 1954.

Verslag van een colloquium over:
"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" van G.H. Orcutt 1)

(Rapport M 9)

Inhoud

§ 1. Inleiding

Het probleem voorspellingen te doen op grond van korte tijdreeksen wordt benaderd door te onderzoeken in hoeverre men terecht kan onderstellen, dat een groot aantal van deze reeksen afkomstig is uit één enkele populatie. Het onderzochte materiaal wordt gevormd door de reeksen, die voorkomen in het door Tinbergen opgestelde model voor de Verenigde Staten 2). Een theoretische beschouwing leidt tot de conclusie, dat aan de genoemde onderstelling onder bepaalde voorwaarden inderdaad is voldaan, waarna voor het autoregressieve schema van de reeksen gevonden wordt:

$$\underline{x}_t = \underline{x}_{t-1} + 0,3(\underline{x}_{t-1}-\underline{x}_{t-2}) + \underline{\varepsilon}_t \quad .$$

§ 2. Theoretische beschouwingen

Om het model van Tinbergen te beschrijven wordt een operator D ingevoerd, die - op een variabele toegepast - deze met één tijdseenheid vertraagt, bv. $D\ X_t = X_{t-1}$.
D^2 is per definitie D.D, d.w.z. dat de operator D twee keer wordt toegepast; evenzo is $D^3\ X_t = D^2(D\ X_t) = X_{t-3}$; enz. Een polynomium in D, werkend op X_t transformeert deze grootheid dus als volgt:

$$(a_1+a_2\ D + a_3\ D^2)X_t = a_1\ X_t + a_2\ X_{t-1} + a_3\ X_{t-2}\ .$$

Het optellen van polynomen in D is op dezelfde wijze gedefinieerd als bij de gewone polynomen. Ook voor de vermenigvuldiging is dit het geval. Immers als $\beta_1 . \beta_2$, waarin bv. $\beta_1 = a_1+a_2D+a_3D^2$ en $\beta_2 = b_1+b_2D$ betekent dat op de variabele X_t eerst de operator β_2 en daarna β_1 moet worden toegepast, dan kan men in

1) J.R.S.S.Series B 10 (1948) 1-53; eveneens: University of Cambridge. Department of Applied Economics, Reprint Series No. 7.
2) J.Tinbergen, Statistical Testing of Business Cycle Theory, Vol. II. Business Cycles in the United States of American, 1919-1932, League of Nations, Geneva (1939).

plaats hiervan ook alleen de operator $(a_1+a_2D+a_3D^2)(b_1+b_2D) = a_1b_1+(a_1b_2+a_2b_1)D+(a_3b_1+a_2b_2)D^2+a_3b_2D^3$ op X_t laten werken.

Het genoemde model bestaat nu uit een stelsel lineaire differentievergelijkingen van de vorm

$$\begin{array}{l} \beta_{11}X_{1t}+\beta_{12}X_{2t}+\dots\dots+\beta_{1n}X_{nt}=Y_{1t} \\ \beta_{21}X_{1t}+\dots\dots\dots\dots+\beta_{2n}X_{nt}=Y_{2t} \\ \vdots \qquad\qquad\qquad\qquad\qquad \vdots \\ \beta_{n1}X_{1t}+\dots\dots\dots\dots+\beta_{nn}X_{nt}=Y_{nt}\,. \end{array} \tag{1}$$

Hierin zijn de β_{ij} polynomen in D, $X_{1t},\dots,X_{nt}$ endogene en $Y_{1t},\dots,Y_{nt}$ exogene variabelen; de laatste kunnen functies zijn van verschillende, al dan niet vertraagde exogene variabelen en bovendien kunnen ze storingen bevatten.

Aangezien de β_{ij} zich ten opzichte van vermenigvuldiging en optelling gedragen als de reële getallen, is het stelsel (1) op de bekende wijze oplosbaar, wanneer men aanneemt, dat de determinant $|B|$ van de coëfficiënten $\beta_{ij} \neq 0$ is. Zij $|E_j|$ de determinant, die ontstaat wanneer de elementen $\beta_{1j},\dots,\beta_{nj}$ uit $|B|$ vervangen worden door $Y_{1t},\dots,Y_{nt}$, dan vindt men

$$\begin{array}{l} |B|\,.\,X_{1t}=|E_1| \\ |B|\,.\,X_{2t}=|E_2| \\ \qquad\vdots \\ |B|\,.\,X_{nt}=|E_n|\,. \end{array} \tag{2}$$

De linkerleden van (2) bevatten alle de operator $|B|$; de rechterleden kunnen voor de verschillende vergelijkingen uiteen lopen. Werkt men de $|E_j|$ uit, dan ontstaan lineaire combinaties van de exogene variabelen, die ook vertraagd hierin kunnen voorkomen. Voor enige vormen van de Y_{it} wordt nu de structuur van de $|E_j|$ onderzocht.

Hypothese 1. De Y_i zijn constant in de tijd. Daar een constante niet verandert wanneer de operator D erop wordt toegepast zijn de $|E_j|$ hier lineaire combinaties van constanten, zodat (2) overgaat in

$$|B|X_{1t}=K_1\,,\quad |B|X_{2t}=K_2,\dots,|B|X_{nt}=K_n\,,$$

waarin $K_1, \ldots, K_n$ constanten zijn.

Worden in een stationnair systeem de X_i gegeven als afwijkingen van hun gemiddelden, dan zijn alle K_j nul, zodat alle variabelen zouden voldoen aan dezelfde autoregressieve vergelijking, die bovendien geen stochastisch element bevat. Uit een beperkt aantal waarnemingen van één der grootheden $X_1, \ldots, X_n$ zouden voor alle X_i de exacte toekomstige waarden te geven zijn!

Hypothese 2. De Y_i zijn lineaire functies in t. Uit $D^r t = t-r$ volgt, dat de $|E_h|$ van de vorm $K_j + L_j . t$ worden, waarin K_j en L_j constanten zijn.
Nu geldt

$$|B| \, X_{j,t} - |B| X_{j,t-1} = (K_j + L_j \; t) - [K_j + L_j (t-1)] \; ,$$

of, anders geschreven

$$|B| \; (1-D) X_{j,t} = L_j \qquad\qquad j = 1, \ldots, n \; ;$$

zodat in dit geval de verschillen van de eerste orde der variabelen alle voldoen aan vergelijkingen, waarvan de linker leden dezelfde coëfficiënten bevatten.

Hypothese 3. De Y_i zijn polynomen van de rde graad in t. Op dezelfde wijze als bij de tweede hypothese blijken hier de verschillen van de rde orde alle aan een vergelijking met hetzelfde linker lid te voldoen. Voor de tweede orde wordt dit bv. $|B| \; (1-D)^2 \; X_{jt} = M_j$, waarin M_j een constante is.

Zijn alle variabelen gegeven als afwijkingen van hun gemiddelden, dan voldoen de verschillen van de rde orde aan dezelfde vergelijking. Worden sommige variabelen waargenomen met meetfouten, dan kan men verwachten betere schattingen te vinden voor de coëfficiënten van de autoregressieve vergelijking, naarmate men meer reeksen beschouwt.

Hypothese 4. De Y_i zijn stochastisch en bezitten de vorm $\underline{Y}_{i,t} = k_i + \underline{\varepsilon}_{i,t}$, waarin de k_i constanten zijn en de $\underline{\varepsilon}_{i,t}$ voor iedere i en t homogeen zijn verdeeld met verwachting nul, terwijl voor iedere t en $i \neq j$. $\underline{\varepsilon}_{i,t}$ en $\underline{\varepsilon}_{j,t}$ onderling onafhankelijk zijn en bovendien alle autocorrelaties 1) de waarde nul hebben.

---

1) zie volgende blz.

Zij $|\underline{E}'_j|$ de determinant, die uit $|B|$ ontstaat, wanneer $\beta_{1j},\dots,\beta_{nj}$ wordt vervangen door $\underline{\varepsilon}_{1,t}, \dots, \underline{\varepsilon}_{n,t}$, dan gaat de oplossing van (1) over in

$$|B| \cdot \underline{X}_{j,t} = K_j + |\underline{E}'_j| \qquad j = 1,\dots,n,$$

of indien $|\underline{E}'_j|$ na uitwerking de stochastische variabele $\underline{\nu}_{j,t}$ oplevert

$$|B| \cdot \underline{X}_{j,t} = K_j + \underline{\nu}_{j,t} .$$

De verwachting van $\underline{\nu}_{j,t}$ is voor iedere j nul, doch voor verschillende j kunnen de $\underline{\nu}_{j,t}$ verschillende autocorrelatiecoëfficiënten bezitten. Het autoregressieve patroon is dus niet door $|B|$ en K_j volledig vastgelegd en men kan de reeksen in het algemeen niet beschouwen als te zijn voortgebracht door één bepaald universum.

In het bijzondere geval, dat $|B|$ gelijk is aan 1-aD en de variabelen gemeten zijn als afwijkingen van hun verwachtingen voldoen de $\underline{X}_{j,t}$ aan

$$\underline{X}_{j,t} = a\,\underline{X}_{j,t-1} + \underline{\nu}_{j,t} . \qquad (3)$$

Zij nu $\rho'_{j,s}$ de sde autocorrelatie van de reeks $\underline{\nu}_{j,t}$, dan concludeert Orcutt uit een vergelijking van $|B|$ en $|\underline{E}'_j|$, dat het redelijk is aan te nemen, dat $\rho'_{j,s}$ nul is voor waarden van j, groter dan het aantal vertragingen in B . Voldoen de $\underline{X}_{j,t}$ dus aan (3), dan zouden alle $\rho'_{j,s}$ voor s>1 nul zijn en alle processen bepaald worden door a en $\rho'_{j,1}$.

Geeft men de sde autocorrelatie van de reeks $\underline{X}_{j,t}$ aan met $\rho_{j,s}$, dan is

$$\rho_{j,1} = \frac{a + \rho'_{j,1}(1+a^2)}{1 + 2a\,\rho'_{j,1}}$$

en

$$\rho_{j,s} = a\,\rho_{j,s-1} .$$

1) Onder de sde autocorrelatie van de tijdreeks $\underline{X}_t$ (t=1,2,...) zal in deze nota verstaan worden

$$\rho_s = \frac{E(\underline{X}_t - E\,\underline{X}_t)(\underline{X}_{t+s} - E\underline{X}_t)}{E(\underline{X}_t - E\underline{X}_t)^2} = \qquad (s=1,2,\dots),$$ terwijl de overeenkomstige grootheid in de waargenomen reeks de sde kettingcorrelatie genoemd wordt (in het Engels meestal serial-correlation).

Aannemende, dat a voor de reeksen van Tinbergen 0,9 moet zijn kan men $\rho_{j,1}$ schetsen als functie van $\rho'_{j,1}$. Het blijkt dan, dat $\rho_{j,1}$ weinig meer verandert voor $\rho'_{j,1} > -0,1$. Daar voor de korte reeksen, die worden gebruikt de steekproefspreidingen in de schattingen voor $\rho_{j,1}$ groot zijn, zal men indien $\rho'_{j,1}$ inderdaad $\geq -0,1$ is geen onderscheid meer kunnen maken tussen de variaties in de schattingen tengevolge van deze twee oorzaken, zodat men dan van de onderstelling uit kan gaan, dat de reeksen uit één universum afkomstig zijn.

Hypothese 5. De Y_i bestaan uit de som van een polynoom in t van de graad r en een term ε_{it}, die aan dezelfde voorwaarden voldoet als onder de 4de hypothese. In dit geval geldt voor de verschillende reeksen van de rde orde hetzelfde als boven gold voor de tijdreeksen zelf.

§ 3. Methode en resultaten van onderzoek

Uit het voorgaande concludeert Orcutt: "it is not completely unreasonable to hope that the rth differences of the economic series involved will behave very nearly as though they were drawings from some population of linear stochastic autoregressive series." Voor korte reeksen kan dit reeds het geval zijn met de reeksen zelf, daar men dan de exogene variabelen op de wijze genoemd onder de 4de hypothese kan voorstellen.

Het onderzochte materiaal bestaat uit de genoemde reeksen van Tinbergen, welke zich uitstrekken over 14 jaar. Voor alle reeksen worden de eerste vier kettingcorrelaties berekend, waarna de onderzoekingen zich richten op het correlogram, dat ontstaat wanneer men voor iedere vertraging het gemiddelde van deze kettingcorrelaties uitzet.

De gebruikte kettingcorrelatie wordt gedefinieerd door

$$\underline{r}_s = \frac{n \sum_{t=1}^{n-s} (\underline{X}_t - \bar{\underline{X}})(\underline{X}_{t+s} - \bar{\underline{X}})}{(n-s)\sum_{t=1}^{n} (\underline{X}_t - \bar{\underline{X}})^2} ,$$

waarin $\bar{\underline{X}} = \frac{1}{n} \sum_{t=1}^{n} \underline{X}_t$.

Experimenteel blijkt $\underline{r}_s$ bij kleine waarden van n gemiddeld een onderschatting te zijn van de autocorrelatiecoëfficiënt

$$\rho_s = \frac{E(\underline{X}_t - E\underline{X}_t)(\underline{X}_{t+s} - E\underline{X}_t)}{E(\underline{X}_t \; E X_t)^2} .$$

Daar er theoretisch weinig bekend is over de mate van onzuiverheid van $\underline{r}_s$ als schatting van ρ_s moet men overgaan op een experimentele methode om conclusies te kunnen trekken omtrent het universum, waaruit de reeksen zijn voortgekomen.

In de eerste plaats onderzocht Orcutt drie processen welke voorkomen in een publicatie van Kendall[1]. Het eerste hiervan voldeed aan de vergelijking

$$\underline{X}_{t+2} = 1,1\,\underline{X}_{t+1} - 0,5\,\underline{X}_t + \underline{\varepsilon}_{t+2} , \qquad (4)$$

waarin de ε's aselecte getallen zijn, getrokken uit een homogene verdeling; die van -49,5 met één opklimt tot 49,5. De reeks, die 480 elementen bevatte werd gesplitst in groepen van 14 en voor elke groep bepaalde Orcutt de coëfficiënten r_1 t/m r_4, de gemiddelden en standaarddeviaties hiervan. Dezelfde waarden werden berekend voor reeksen voortgebracht door de modellen

$$\underline{X}_{t+2} = 1,2\,\underline{X}_{t+1} - 0,4\,\underline{X}_t + \underline{\varepsilon}_{t+2} \qquad (5)$$

en

$$\underline{X}_{t+2} = 1,1\,\underline{X}_{t+1} - 0,8\,\underline{X}_t + \underline{\varepsilon}_{t+2} \qquad (6)$$

Een vergelijking van de correlogrammen van de gemiddelde kettingcorrelaties van Tinbergen's reeksen met die van de reeksen (4), (5) en (6) leidde ertoe ook het model

$$\underline{X}_{t+1} = 0,9\,\underline{X}_t + \underline{\varepsilon}_{t+1} \qquad (7)$$

te onderzoeken. Uitgedrukt in de ε's heeft deze reeks de vorm

$$\underline{X}_{t+1} = \sum_{i=0}^{t} (0,9)^i \; \underline{\varepsilon}_{t+1-i} \qquad (8)$$

Teneinde de gevoeligheid van de correlogrammen na te gaan voor wijzigingen in de coëfficiënten van (8) bestudeerde Orcutt ook

1) M.G.Kendall, Contributions to the Study of Oscillatory Time Series. National Institute of Economic and Social Research, Occasional Papers no. 9 (1946).

het process

$$\underline{x}_t = 5\ \underline{\varepsilon}_t + 4\ \underline{\varepsilon}_{t-1} + 3\ \underline{\varepsilon}_{t-2} + 2\ \underline{\varepsilon}_{t-3} + \underline{\varepsilon}_{t-4} \qquad (9)$$

waarin de ε's weer aselecte trekkingen uit een homogene verdeling zijn.

De nulhypothese, dat de economische tijdreeksen uit één van deze modellen afkomstig zijn, werd op de volgende drie wijzen getoetst:

1) de tweezijdige toets van Student, toegepast op de gemiddelde kettingcorrelaties $\underline{r}_1$ t/m $\underline{r}_4$;

2) daar de toets, genoemd onder 1) voor een bepaald model bij bv. $\underline{r}_1$ en $\underline{r}_2$ wèl en tegelijkertijd bij $\underline{r}_3$ en $\underline{r}_4$ niet tot verwerping kan leiden, werd als tweede toets genomen de toets voor het combineren van onafhankelijke toetsen, welke werd toegepast op de vier uitkomsten van 1);

3) de F-toets voor het quotient van de varianties der kettingcorrelaties, toegepast op de varianties van $\underline{r}_1$, $\underline{r}_2$, $\underline{r}_3$ en $\underline{r}_4$ afzonderlijk.

Aan 1) en 3) ligt de onderstelling ten grondslag, dat $\underline{r}_1$, $\underline{r}_2$, $\underline{r}_3$ en $\underline{r}_4$ normaal verdeeld zijn, wat asymptotisch inderdaad het geval is. 1) De voor 2) vereiste onafhankelijkheid is niet aanwezig, maar wanneer het niet mogelijk is een model te verwerpen op grond van de uitgevoerde toets, zou dit zeker niet mogelijk geweest zijn, wanneer met de correlatie tussen de toetsingsgrootheden rekening was gehouden.

Onderstaande tabel geeft een overzicht van de gevonden overschrijdingskansen:

1) A.M.Walker, The asymptotic distribution of secial correlation coefficients for autoregressive processes with dependent residuals. Proceedings of the Cambridge Philosophical Society 50 (1954) 60-64.

		t-toets	combinatietoets	F-toets
model (4)	r_1	0,246		0,10
	r_2	0,242	$< 10^{-6}$	0,42
	r_3	0,00042		0,04
	r_4	0,000006		0,001
model (5)	r_1	0,063		0,09
	r_2	0,190	0,03	0,13
	r_3	0,858		0,25
	r_4	0,019		0,05
model (6)	r_1	0,390		0,007
	r_2	0,000000	$<10^{-6}$	0,05
	r_3	0,000000		0,40
	r_4	0,000000		0,003
model (7)	r_1	0,212		0,06
	r_2	0,624	0,70	0,40
	r_3	1,000		0,08
	r_4	0,460		0,02
model (9)	r_1	0,358		0,48
	r_2	0,660	0,62	0,16
	r_3	0,364		0,18
	r_4	0,516		0,24

Op grond van de 2de toets zal men de modellen (7) en (9) niet verwerpen, op grond van de 3de (5), (7) en (9) niet. Het is daarom niet onredelijk aan te nemen, dat alle reeksen uit éën universum komen, dat wellicht voorgesteld kan worden door

$$\underline{x}_{t+1} = 0{,}9\, \underline{x}_t + \underline{\varepsilon}_{t+1} ,$$

hoewel dit proces weinig voorkeur heeft boven (9) en gezien vanuit de varianties ook boven (5).

Vooral uit een vergelijking van de modellen (7) en (9) blijkt dus dat de correlogrammen tamelijk ongevoelig zijn voor wijzigingen in de autoregressieve parameters van het model. Daar bovendien de standaardfouten in de schattingen $\underline{r}_1, \ldots, \underline{r}_4$ groot zijn kan een nauwkeurige bepaling van het model niet worden verwacht.

De oorzaak ligt gedeeltelijk in het geringe aantal waarnemingen waar iedere reeks uit bestaat. Echter niet alleen hierin. Schrijft men het model in de vorm

$$\underline{X}_{t+1} = a\,\underline{X}_t + b(\underline{X}_t - \underline{X}_{t-1}) + \underline{\varepsilon}_{t+1} \tag{10}$$

dan zijn de autocorrelaties wel gevoelig voor a, maar wanneer a ongeveer 1 is, niet voor b. Dit volgt direct uit de eenvoudig af te leiden formules

$$\rho_1 = \frac{a+b}{1+b} \tag{11}$$

en

$$\rho_s = (a+b)\rho_{s-1} - b\rho_{s-2} \tag{12}$$

De gevoeligheid voor b kan worden opgevoerd door correlogrammen van de eerste orde verschillen van de reeksen te nemen. Aangezien de sde autocorrelatie hiervan gelijk is aan

$$\frac{E(\underline{X}_{t+s+1}-\underline{X}_{t+s})(\underline{X}_{t+1}-\underline{X}_t)}{E(\underline{X}_{t+1}-\underline{X}_t)^2} \quad \text{p.d.} \quad \rho_{ds}$$

bestaat tussen ρ_{ds} en de autocorrelaties van de oorspronkelijke reeks het volgende verband

$$\rho_{ds} = \frac{2\rho_s - \rho_{s+1} - \rho_{s-1}}{2(1-\rho_1)} .$$

Met behulp van de formules (11) en (12) blijkt, dat althans de ρ_{ds} gevoeliger zijn voor variaties in b, dan de ρ_s (bv. $\rho_1 = b + \frac{1}{2}(a-1)$. Voor de in groepjes van 14 gesplitste reeksen (5) en (7) schat Orcutt m.b.v. de formule

$$\underline{r}_s = \frac{(n-s)\sum_{t=1}^{n-s-1}(\underline{X}_{t+s-1}-\underline{X}_{t+s})(\underline{X}_{t+1}-\underline{X}_t)}{\sum_{t=1}^{n-s}(\underline{X}_{t+1}-\underline{X}_t)^2}$$

ρ_{ds} $(s=1,\dots,4)$. Hierbij blijken de steekproefvarianties niet veel groter te zijn dan die bij de schattingen $\underline{r}_s$ van ρ_s, terwijl er geen duidelijke aanwijzing is voor onzuiverheid van de schattingen. Men kan daarom verwachten b beter te kunnen schatten, wanneer gebruik wordt gemaakt van de correlogrammen van de eerste orde verschillen.

Orcutt vergelijkt het correlogram, dat ontstaat door de gemiddelden te nemen van de $\underline{r}_{ds}$ ($s = 1,\ldots,4$) van de 52 reeksen van Tinbergen met de autocorrelogrammen van de verschillen van reeksen voortgebracht door het schema (10), waarin voor a en b gekozen zijn de waarden

1) $a = 1,0$ $\quad b = 0,4$
2) $a = 1,0$ $\quad b = 0,3$
3) $a = 0,9$ $\quad b = 0,0$
4) $a = 0,9$ $\quad b = 0,4$.

Figuur 1 leidt dan tot de conclusie, dat de differentievergelijking

$$\underline{x}_{t+1} = \underline{x}_t + 0,3\,(\underline{x}_t - \underline{x}_{t-1}) + \underline{\varepsilon}_{t+1} \qquad (13)$$

een tamelijk nauwkeurige benadering vormt van het proces waaruit Tinbergen's reeksen afkomstig zijn.

Hoewel dit onderzoek er dus op wijst, dat de door Tinbergen gemaakte onderstelling, dat een economisch systeem door een stelsel lineaire differentievergelijkingen voorgesteld kan worden niet onredelijk is, bestaat er groot verschil tussen het model (13) en het model

$$z_t^c = 0,398\, z_{t-1}^c - 0,220\, z_{t-2}^c + 0,013\, z_{t-3}^c + 0,027\, z_{t-4}^c\,,$$

dat Tinbergen vond voor de winsten van maatschappijen.

Het verschil is volgens Orcutt ondermeer ontstaan, doordat bij Tinbergen het eliminatieproces moest leiden tot grote standaardfouten in de schattingen van de coëfficiënten en doordat de kettingcorrelaties de overeenkomstige autocorrelaties ernstig onderschatten; het laatste heeft tengevolge, dat de mate van continuïteit in de reeksen wordt onderschat.

In principe kan worden getoetst of de gevonden tijdreeksen alle uit model (13) zijn ontstaan. Men moet hiervoor de verdelingsfuncties bepalen van de correlatiecoëfficiënten tussen twee reeksen, ontstaan uit (13) met verschillende verzamelingen aselecte getallen voor de $\underline{\varepsilon}_t$. Aangezien hier theoretisch niets van bekend is, zou men op experimentele wijze de vorm van deze verdeling moeten bepalen.

§ 4. Enige opmerkingen en conclusie

1. De redenering, die Orcutt gebruikt om aan te tonen, dat $\underline{r}_s$ een onderschatting vormt van ρ_s, is niet juist. De schatting $\underline{r}_s$ wordt goed benaderd door

$$1 - \frac{n \sum_{t=1}^{n-s} (\underline{X}_t - \underline{X}_{t+s})^2}{2(n-s) \sum_{t=1}^{n} (\underline{X}_t - \underline{\bar{X}})^2} ,$$

waarin $\underline{\bar{X}} = \frac{1}{n} \sum_{t=1}^{n} . \underline{X}_t$. Hierin is $\underline{\bar{X}}$ die schatting van $E\underline{X}_t$, die de som van kwadraten $(\underline{X}_t - \underline{\bar{X}})^2$ minimaliseert, zodat volgens Orcutt bij iedere fout in de schatting van $E\underline{X}_t$ de noemer in werkelijkheid groter is, waardoor $\underline{r}_s$ een onderschatting vormt van ρ_s. Dat deze van n onafhankelijke redenering niet goed kan zijn, blijkt reeds hieruit, dat Kendall voor grote n steeds overschattingen vindt voor ρ_s. 1)

Een bewijs is eerst dan geleverd, wanneer kan worden aangetoond, dat bij een klein aantal waarnemingen geldt $E\underline{r}_s < \rho_s$.

2. Wanneer men de differentievergelijking (13) oplost vindt men:

$$\underline{X}_t = c_1 + c_2(0{,}3)^t + \sum_{j=1}^{t} \frac{1-(0{,}3)^j}{0{,}7} \underline{\varepsilon}_{t-j+1} \qquad (14)$$

waarin c_1 en c_2 vrij te kiezen constanten zijn. Voor grote waarden van t gaat (14) bij benadering over in

$$\underline{X}_t = c_1 + \sum_{j=1}^{t} \frac{1-(0{,}3)^j}{0{,}7} \underline{\varepsilon}_{t-j+1} \qquad (15)$$

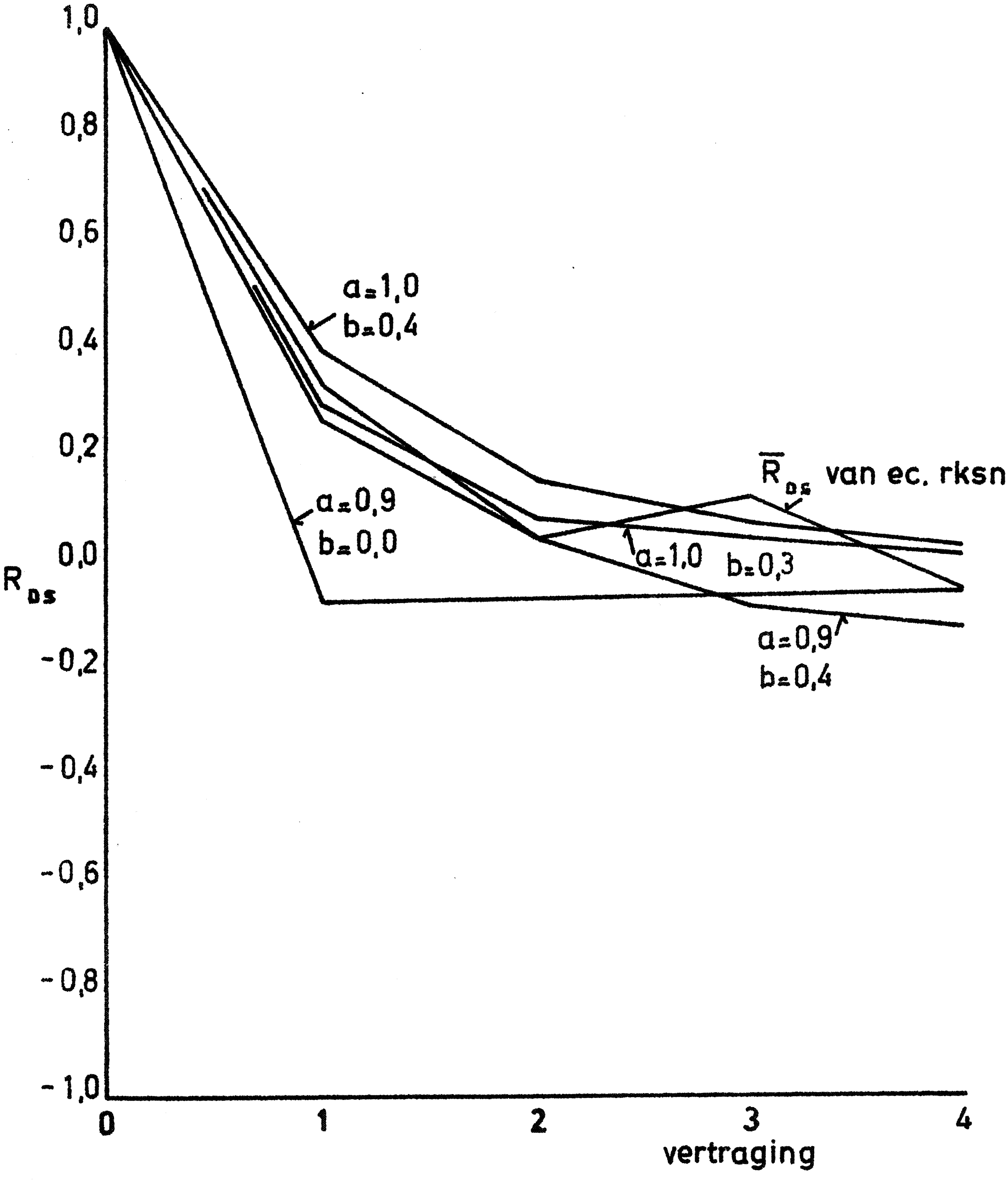
Het proces is dus niet stationnair, terwijl de variantie van de schatting $\underline{X}_t$ willekeurig groot wordt voor toenemende t.

3. De methode aangegeven door Orcutt kan in principe goede resultaten opleveren. De verkregen vergelijking (13) is echter

---

1) Men vergelijke bv. M.G.Kendall, On the Analysis of Oscillatory Time Series, J.R.S.S. 108(1945) 93-141, waar Kendall voor de reeksen (4), (5) en (6) een groot aantal kettingcorrelaties berekent. De eningszins andere schatting, die wordt gebruikt kan voor grote n geen verschil met $\underline{r}_s$ opleveren.

weinig betrouwbaar, daar zeker niet overtuigend is aangetoond, dat aan de gestelde eisen, in het bijzonder de hypothese, dat alle reeksen afkomstig zijn uit hetzelfde universum, in voldoende mate is voldaan. Men had dan ook mogen verwachten, dat de vergelijking (13) op dezelfde wijze zou zijn getoetst als de vergelijkingen (4) t/m (7) en (9). Bovendien is het niet vanzelfsprekend, dat $\underline{r}_{ds}$ in het gekozen model een goede schatting is van ρ_{ds}.



1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
R DS
0
1
2
3
4
vertraging
a=1,0
b=0,4
a=0,9
b=0,0
a=1,0
b=0,3
R̄ DS van ec. rksn
a=0,9
b=0,4

FIG. 1

8727 NL