Programmatuur-sectie

Een computerprogramma voor de analyse van enkelvoudige en multiple tijdseries

door

W. Immink

Vakgroep Psychometrie, Statistiek en Modelvorming

Subfaculteit Psychologie

Rijksuniversiteit Utrecht

St.Jacobsstraat 14,

3511 BS Utrecht.

Tel.: 030-328711.

I. Inleiding

Algemeen.

i. Een tijdserie is een verzameling observaties geordend in een of andere ruimte (meestal de tijd), de tijds-as genoemd. Deze observaties worden echter niet, zoals gebruikelijk in de klassieke statistiek, stochastisch onafhankelijk verondersteld. Wanneer de tijds-as uit een rij equidistante punten (tijdstippen genoemd) bestaat, is er sprake van een discrete tijdserie. Voorbeelden van discrete tijdseries zijn:

- de gemiddelde maandelijkse exportprijzen van Nederlandse kaas;
- het gemiddelde jaarlijkse zoutgehalte van de Rijn gedurende de jaren 1920 tot 1978;
- de hartslagfrequentie van ratten (in een psychologisch experiment) om het uur gemeten (eventueel onder verschillende experimentele condities).

Analyse van een tijdserie met betrekking tot een bepaald fenomeen, heeft tot doel bepaalde eigenschappen hiervan op te sporen; bovendien kan het van belang zijn een voorspelling te doen over het toekomstig gedrag van dit fenomeen.

De methode van Box en Jenkins voor de analyse van tijdseries is gebaseerd op:

a) de tijdserie wordt beschouwd als realisatie van een proces;
b) vervolgens wordt een lineair stochastisch model van dit proces gezocht, waarbij verondersteld wordt dat betreffende tijdserie als een realisatie van dit model kan worden beschouwd;
c) de veronderstellingen welke hierbij gehanteerd worden zijn:
   - elke realisatie z_t op tijdstip t is afhankelijk van realisaties op vorige tijdstippen $z_{t-1}, z_{t-2}, \ldots$
   - elke realisatie z_t wordt tevens bepaald door een stochastisch proces, witte ruis.

Box en Jenkins modellen.

ii. Een stochastisch proces $z_t (t \in Z)$ heet stationair indien

$$\rho(z_{t_1},\ldots,z_{t_n}) = \rho(z_{t_1+\tau}, z_{t_2+\tau}, \ldots, z_{t_n+\tau}) \quad (1)$$

voor iedere deelverzameling $\{t_1, t_2, \ldots, t_n, \tau\}$ uit Z en iedere $n \in N$; hier stelt ρ de verdelingsfunctie van $(z_{t_1}, \ldots, z_{t_n})$ voor.

Een stationair stochastisch proces a_t heet een "witte ruis" proces met variantie σ_a^2 indien:

$E(a_t) = 0$, $Var(a_t) = \sigma_a^2$ en $Cov(a_t, a_{t+k}) = 0$ voor alle $k \in Z \setminus \{0\}$

Box en Jenkins (1970, pag. 46) definiëren het algemene lineaire stochastische proces z_t als

$$\tilde{z}_t \overset{def.}{===} \Psi(B)\, a_t \quad (2)$$

hierin is a_t het "witte ruis" proces

$$\tilde{z}_t \overset{def.}{===} z_t - c \text{ (c een constante)}$$

$$\Psi(B) \overset{def.}{===} 1 + \Sigma_{i=1}^{\infty} \Psi_j B^j$$

$$B\, a_t \overset{def.}{===} a_{t-1}$$

Het lineaire proces $\tilde{z}_t$ wordt dus gedefinieerd door een transformatie van het "witte ruis" proces a_t door een lineair filter met transferfunctie $\Psi(B)$ (zie I-iv).

iii. Het lineaire stochastische proces z_t dat voldoet aan

$$\phi(B)\, \nabla^d\, \tilde{z}_t = \theta(B)\, a_t + \theta_o \quad (3)$$

waarin:

$$\phi(B) \overset{def.}{===} 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta(B) \overset{def.}{===} 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

$\nabla \overset{\text{def.}}{===} 1 - B$

θ_0 de trendparameter is

heet een ARIMA (p,d,q)-proces (Auto Regressive, Integrated, Moving Average).

Als z_t voldoet aan de relatie

$$\phi_p(B) \quad \Phi_P(B^s) \nabla^d \nabla_s^D \tilde{z}_t = \theta_q(B) \, \Theta_Q(B^s) \, a_t \tag{4}$$

waarin:

$\nabla_s = 1 - B^s \; (s \in N)$

a_t: "witte ruis" proces

ϕ_p:	een	polynoom	in	B	van	de graad	p
Φ_P:	"	"	"	B^s	"	"	P
θ_q:	"	"	"	B	"	"	q
Θ_Q:	"	"	"	B^s	"	"	Q

is er sprake van een proces met seizoeneffect van periode s.
In dit model kan men het proces z_t als volgt opgebouwd denken:

- $\Phi_P(B^s) \nabla_s^D z_t = \Theta_Q(B^s) \, d_t$ geeft een relatie tussen

  z_t, z_{t-s}, z_{t-2s}, ... voor tijstippen t; hierin is d_t een ruisproces;

  een proces waarvoor geldt dat $\mathrm{Cov}(d_t, d_{t+k}) \neq 0 \; (k \neq 0)$

- Vervolgens geldt voor het ruisproces d_t de relatie

  $\phi_p(B) \nabla^d d_t = \theta_q(B) \, a_t$

  met a_t het "witte ruis" proces.

Met model in (4) wordt het multiplicatieve model van de orde (p,d,q) x $(P,D,Q)_s$ genoemd.

iv. Een stochastisch proces y_t heet het outputproces van een lineairfilter (Box en Jenkins, 1970) met transferfunctie $v(B)$ van de inputserie x_t indien:

$$y_t = v(B)\, x_t \tag{5}$$

met $v(B) \overset{\text{def.}}{===} v_0 + v_1 B + v_2 B^2 + \ldots$

De verzameling $\{v_0, v_1, v_2, \ldots\}$ heet de impulse-response functie van het dynamische systeem $y_t = v(B)\, x_t$.

Box en Jenkins (1970, pagina 345) beperken zich nu tot transferfunctie modellen met outputserie y_t en inputserie x_t waarvoor:

$$\delta(B)\, y_t = \omega(B)\, x_{t-b}. \tag{6}$$

waarin:

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 \ldots - \delta_r B^r$$

$$\omega(B) = 1 - \omega_1 B - \omega_2 B^2 \ldots - \omega_s B^s$$

Indien het dynamische systeem $y_t = v(B)\, x_t$ verstoord wordt door een ARIMA (p,d,q)-proces n_t, heeft het transferfunctie model de vorm:

$$y_t = v(B)\, x_t + n_t \tag{7}$$

Hierin is $v(B)$ meestal gelijk aan $\delta^{-1}(B)\, \omega(B)\, B^b$

Fitten van het model

v. Het computer programma van D.J. Pack stelt de gebruiker in staat een tijdserieanalyse uit te voeren.
Deze analyse bestaat uit de identificatie-, schattings- en verificatie-procedure.
Onder de identificatie van een model wordt verstaan, het bepalen van de onbekenden:

- p, d en q in het model van (3)
- p, P, s, d, D, q en Q in het model van (4) òf
- r, s en b in het model van (6).

Deze identificatieprocedure vindt plaats in het eerste hoofdprogramma van Pack.

Na de identificatie worden de onbekende parameters van het model geschat. Deze schattingsprocedure vindt plaats in het tweede hoofdprogramma van Pack.

Na identificatie en schatting kan met behulp van het eerste hoofdprogramma getoetst worden of het gevonden model adequaat is.

Voor een beknopte behandeling van:
- de Box Jenkins modellen
- de definities van stationariteit, in verteerbaarheid, stabiliteit, etc.
- de identificatie-, schattings- en verificatieprocedures
- de gebruikte rekenformules in het computerprogramma van D.J. Pack,

wordt verwezen naar het PSM-bulletin,

jaargang: 1979; no. 1
auteur : W. Immink.

Voor een uitgebreide uiteenzetting over de structuur van het computerprogramma wordt verwezen naar de ontwerper daarvan: Pack (1977).

II. Het computerprogramma van D.J. Pack.

Bij het uitvoeren van een tijdserieanalyse met behulp van dit programma, zijn de onderstaande gegeneraliseerde Box-Jenkins modellen toegestaan.

a Het gegeneraliseerde univariate Box-Jenkins model.

Dit model heeft de vorm:

$$\prod_{i=1}^{a} \phi_i(B) \prod_{j=1}^{b} (1-B^{s_j})^{d_j} \tilde{z}_t = \theta_0 + \prod_{k=1}^{c} \theta_k(B)\, a_t$$

($a \leq 3$, $b \leq 3$ en $c \leq 3$)

Hierin is:

i). $\phi_i(B) \overset{\text{def.}}{=\!=\!=} 1 - \phi_{i1} B - \phi_{i2} B^2 - \ldots - \phi_{iP_i} B^{P_i}$

de i^e "autoregressive factor" met "autoregressive parameter" $\phi_{ij} (1 \leq j \leq p_i)$; j heet de orde van ϕ_{ij}

ii). $\theta_k(B) \overset{\text{def.}}{=\!=\!=} 1 - \theta_{k1} B - \theta_{k2} B^2 - \ldots - \theta_{kq_k} B^{q_k}$

de k^e "moving average factor" met "moving average parameter" $\theta_{kl} (1 \leq l \leq q_k)$; l heet de orde van θ_{kl}

iii). de term $(1-B^{s_j})^{d_j}$ de j^e differentie factor (of differentie-type) ; s_j is per definitie de orde van de differentie en d_j is het aantal differenties van de vorm $1-B^{s_j}$.

iv). θ_0 de trendparameter.

b Het gegeneraliseerde multiple Box-Jenkins model.

Als y_t de outputserie en x_t de inputserie is van een lineair filter, heeft het gegeneraliseerde model de vorm:

$$\prod_{j=1}^{a} (1-B^{s_{jy}})^{d_{jy}} \tilde{y}_t = \theta_0 + \prod_{i=1}^{b} \delta_i^{-1}(B) \prod_{k=1}^{c} \omega_k(B) \prod_{l=1}^{d} (1-B^{s_{lx}})^{d_{lx}} \tilde{x}_{t-b}$$

$$+ n_t (a \leqslant 3,\ b \leqslant 3,\ c \leqslant 3 \text{ en } d \leqslant 3).$$

Hierin is:

i). $(1-b^{s_{jy}})^{d_{jy}}$ de j^e differentiefactor op de outputserie

ii). $(1-B^{s_{lx}})^{d_{lx}}$ de l^e differentiefactor op de inputserie

iii). $\delta_i(B) \overset{\text{def.}}{===} 1 - \delta_{i1} B - \delta_{i2} B^2 - \ldots - \delta_{ir_i} B^{r_i}$

de i^e "output lag factor" met "output lag parameter" $\delta_{ij} (1 \leqslant j \leqslant r_i)$; j heet de orde van δ_{ij}

iv). $\omega_k(B) \overset{\text{def.}}{===} 1 - \omega_{k1} B - \omega_{k2} B - \ldots - \omega_{ks_k} B^{s_k}$

de k^e "input lag factor" met "input lag parameter" $\omega_{kj} (1 \leqslant j \leqslant s_k)$; j heet de orde van ω_{kj}

v). θ_0 de trendparameter

vi). n_t een ARIMA (p,d,q)-proces.

Opmerking: indien er sprake is van meer dan één inputserie, wordt het rechterlid in deze formule vermeerderd met een term welke analoog is aan de middelste term.

c. De output van het programma.

Het identificatieprogramma kan (afhankelijk van de invoerbeschrijving) het volgende afdrukken:

- listing van de tijdserie(s) met bijbehorende titel(s)
- een plot van de tijdserie(s)
- de autocorrelatie- en partiële autocorrelatiefunctie(s) voor verschillende differentietypen
- een plot van iedere berekende autocorrelatie- en partiële autocorrelatiefunctie
- de cross-correlaties tussen de tijdseries
- plot van iedere berekende cross-correlatiefunctie
- de statistics welke van belang zijn voor de identificatie- en verificatieprocedure.

Het schattingsprogramma kan (afhankelijk van de invoerbeschrijving) het volgende afdrukken:

- listing van de tijdserie(s) met bijbehorende titel(s)
- een plot van de tijdserie(s)
- de iteratiestappen bij de schattingsprocedure
- de uiteindelijke schattingen van de parameters
- de residuen
- een plot van de residuen
- de correlatie matrix van de geschatte parameters
- de autocorrelatie- en partiële autocorrelatiefunctie van de residuen
- de cross-correlatiefunctie tussen de residuen en de inputseries
- plot van de cross-correlatiefuncties
- voorspellingen, met bijbehorende betrouwbaarheidsintervallen
- plot van de voorspellingen
- statistics welke van belang zijn voor de schattingsprocedure, verificatieprocedure en de voorspellingsprocedure

d. De invoerbeschrijving.

Voor de invoerbeschrijving wordt verwezen naar het PSM-bulletin (jaargang 1979 no. 1).

Deze beschrijving bestaat uit 4 onderdelen, te weten:

- identificatie enkelvoudige tijdseries
- identificatie multiple tijdseries
- schattingsprocedure enkelvoudige tijdseries
- schattingsprocedure multiple tijdseries
- twee tabellen met een kort overzicht van de condities behorende bij verschillende leesstatements.

Literatuur

Anderson, O.D., Time Series Analysis and Forecasting: The Box-Jenkins Approach, Butterworth, 1976.

Box, G.E.P. en Jenkins, G.M., Time Series Analysis: Forecasting and Control, Holden Day, 1970.

Box, G.E.P. en Tiao, G.C., Intervention Analysis with Applications to Economic and Environmental Problems, Journal of American Statistical Association, 70, pag. 70-79, 1975.

Glass, G.V.; Wilson, V.L. en Gottman, J.M., Design and Analysis of Time Series Experiments, Colorado Associated University Press, Boulder, 1975.

Pack, D.J., Goodman, M.L. en Miller, R.B. Computer Programs for the Analysis of Univariate Time Series Using the Methods of Box and Jenkins, The Data Center, College of Administrative Science, The Ohio University, 1974.

Pack, D.J., Computer Programs for the Analysis of Single Input Transfer Function Models Using the Methods of Box and Jenkins, The Data Center, College of Administrative Science, The Ohio State University, 1974.

Pack, D.J., Computer Programs for the Analysis of Univariate Time Series Models and Single Input Transfer Function Models Using the Methods of Box and Jenkins, The Data Center, College of Administrative Science, The Ohio State University, 1974.

Pack, D.J., A Computer Program for the Analysis of Time Series Models Using the Box-Jenkins Philosophy, The Data Center, College of Administrative Science, The Ohio State University, 1977.