Programmatuursectie

Tijdreeksen analyse met behulp van de programma's Correl en TSX.

door: W. Immink
vakgroep P.S.M.
St. Jacobsstraat 14, Utrecht.
Tel.: 030-328711.

Inleiding:

Binnen de statistiek hanteert men vaak modellen, waarbij verondersteld wordt, dat de observaties onafhankelijk zijn. Bij veel processen, in bijvoorbeeld de economie en sociale wetenschappen, zijn echter de verkregen observaties afhankelijk. Derhalve is men natuurlijk geïnteresseerd in de aard en de mate van die afhankelijkheid.
Een dergelijke reeks van afhankelijke observaties (op één variabele) kan vaak beschouwd worden als een tijdreeks; een verzameling observaties geordend in een of andere ruimte (meestal de tijd).

Het ARIMA-Model.

Een model geconstrueerd door Box en Jenkins, om een tijdreeks te analyseren, is het ARIMA-model. (Auto Regressive Integrated Moving Average). Dit model is gebaseerd op de idee, dat een tijdreeks, waarin de opeenvolgende waarden sterk afhankelijk zijn, beschouwd kan worden als een reeks gegenereerd uit een rij van onafhankelijk, identiek verdeelde variabelen a_t, waarbij a_t normaal verdeeld is met verwachting 0 en variantie σ_a^2 (t is een discrete tijdparameter). De stochastische variabelen $a_t, a_{t-1}, a_{t-2}, \ldots$ heet het "white noise" proces. Er wordt verondersteld dat dit "white noise" proces a_t wordt getransformeerd, via een lineair filter $\Psi(B)$, tot het algemene lineaire proces z_t (waarbij de tijdreeks een realisatie van dit proces is) volgens de relatie:

$z_1 = \mu + \Psi(B)\, a_t$ met $\Psi(B) = \Sigma_{i=0}^{\infty} \psi_i B^i$ en μ heet gemiddelde van de observaties, waarbij

$$B^k a_t \overset{\text{def.}}{=} a_{t-k} \quad k=0,1,2,\ldots. \text{ en } \psi_0 \overset{\text{def.}}{=} 1.$$

Voor praktische problemen kan het bovengenoemde lineaire proces, onder zwakke condities, geschreven worden in de vorm:

$\Phi(B)\ \nabla^d.\tilde{z}_t = \theta(B)\ a_t$ waarbij:

$\tilde{z}_t \overset{\text{def.}}{=} z_t - \mu$

$\nabla z_t \overset{\text{def.}}{=} (1-B)\ z_t = z_t - z_{t-1}$

$\Phi(B) = 1-\phi_1 B-\phi_2 B^2- \ldots . -\phi_p B^p$

$\Theta(B) = 1-\theta_1 B-\theta_2 B^2 - \ldots . -\theta_q B^q$

In de literatuur wordt dit model meestal aangegeven als het ARIMA (p,d,q)-model, waarin p, d en q in het voorgaande gedefinieerd zijn. Bij een gegeven tijdreeks moet nu een bijpassend ARIMA (p,d,q)-model gefit worden. Het fitten van zo'n model vindt plaats in drie stappen, te weten:

- de identificatie van het model (dat wil zeggen het bepalen van de parameters p,d en q);
- het schatten van de optredende, onbekende coëfficiënten ϕ_i en θ_j $(i=1,2,\ldots.p;\ j=1,2,\ldots.q)$;
- het verifiëren van het gefitte model.

Correel en TSX:

De programma's Correl en TSX voeren nu een tijdreeksenanalyse uit, waarbij een ARIMA-model wordt gefit, welke zo goed mogelijk beantwoordt aan de Tijdreeks (Box en Jenkins (1976)).
Het programma Correl berekent de autocorrelatie- en de partiële- autocorrelatiefuncties (met de bijbehorende chi-kwadraat statistics), op grond waarvan de identificatie uitgevoerd kan worden.

Het programma TSX schat nu de optredende coëfficiënten $\phi_1, \phi_2, \ldots, \phi_p$ alsmede de coëfficiënten $\theta_1, \theta_2, \ldots, \theta_q$; bovendien berekent dit programma de statistics welke bij de design parameters (of interventieparameters) horen. Verder berekent TSX de residuen waarmee een verificatie kan worden uitgevoerd, door deze residuen in het programma Correl in te voeren. Voorts biedt het programma TSX de mogelijkheid het effect van verschillende interventies (zoals bijvoorbeeld het effect bij toepassing van een therapie) te schatten en te toetsen (Glass, Willson and Gottman, 1975)). Voor een uitvoeriger behandeling van het ARIMA-model wordt verwezen naar Box en Jenkins (1976); voor een behandeling van het schatten en toetsen van interventie-effecten naar Glass, Willson and Gottman (1975).

Invoerbeschrijving Correl:

1e kaart:
kol 1-5 : aantal problemen in formaat (1I5)

2e kaart:
kol 1-80: omschrijving problemen X in formaat (8A10)

3e kaart:
kol 1-5 n1 (aantal pre-interventie observaties)
6-10 n2 (aantal post-interventie observaties)
N = n1 + n2 maximaal = 300 observaties
15 0 - ruwe data geanalyseerd
1 - ruwe data + log transformatie
2 - uitsluitend log transformatie
16-20 C: de waarde van de constante in de log transformatie
formule $\ln(Z_t + c)$
defaultwaarde c = 0
F-formaat
21-25 s: de lengte van een periode bij periodieke data

4e kaart:
format kaart voor de in te lezen data (in F-formaat)
vb. (20F4.0)

5e en volgende kaarten:

data volgens opgegeven formaat

Indien meerdere problemen volgen, dan deze setup vanaf 2e kaart herhalen.

Output Correl:

- pre en post-interventie observaties
- op file PUNCH de transformeerde data
- lag k autocorrelaties en standaardfouten
- partiële autocorrelaties en "
- χ^2-waarden.

Invoerbeschrijving TSX.

1e kaart:

kolom 1-5 : aantal problemen formaat (1I5)

2e kaart:

kolom 1-80: omschrijving probleem X formaat (8A10)

3e kaart: (specifikatiekaart)

kolom 1-5	p orde van AR model		$(0 \leq p \leq 3)$
6-10	d orde van verschil ∇^d		$(0 \leq d \leq 4)$
11-15	q orde van MA model		$(0 \leq q \leq 3)$
16-20	n1 indien ICODE = 0 of 1		
	m aantal designparameters indien ICODE = 2		
21-25	n2 indien ICODE = 0 of 1		
	N totaal aantal observaties indien ICODE = 2		
30	ICODE 0	2 design parameters (niveau (L) en niveauverandering (δ))	
	1	4 design parameters (L, δ + drift en driftverandering)	
	2	een door de gebruiker gespecificeerde design matrix van de orde N x m	

31-35 NCYC aantal punten per periode voor periodieke data

36-40 STEP stapgrootte voor ϕ_j en/of θ_j (* 100)

defaultwaarden

p/q	stapgrootte
1	0.02
2	0.10
3	0.25
4	0.50
5	0.50
6	0.50

45 LIMIT 0 ϕ_j en/of θ_j waarden range over het gehele invertibility-stationarity gebied

1 door gebruiker opgegeven waarden range voor ϕ_j en/of θ_j

50 RES 0 geen residuals

1 residuals geprint en op datafile PUCH voor gespecificeerde waarden van ϕ_j en/of θ_j

51-80 indien RES = 1 dan in deze kolommen en (6I5) formaat de waarden van $\phi_1, \phi_2, \phi_3, \theta_1, \theta_2, \theta_3$, waarvan residuen berekend moeten worden.

N.B.: elke waarde ϕ/θ moet met 100 vermenigvuldigd worden (integers!)

4e kaart: (alleen verplicht indien LIMIT = 1)

kolom 1-5 vorige run: stapgrootte voor ϕ_j * 100

6-10 actuele run: " " " "

11-15 min.waarde van ϕ_1 * 100 in vorige run

16-20 " " " ϕ_2 * " " " "

21-25 " " " ϕ_3 * " " " "

26-30 stapgrootte in vorige run voor θ_j * 100

31-35 actuele stapgrootte van θ_j * 100

36-40 min.waarde van θ_1 * 100 in vorige run

41-45 " " " θ_2 * " " " "

46-50 " " " θ_3 * " " " "

De defaultwaarden voor de grenzen zijn

p = 1	$-1 < \phi_1 < 1$
p = 2	$-1 < \phi_2 < 1$
	$\phi_1 + \phi_2 < 1$
	$\phi_2 - \phi_1 < 1$
P = 3	$-1 < \phi < 1$
	$\phi_1 + \phi_2 + \phi_3 < 1$
	$-\phi_1 + \phi_2 - \phi_3 < 1$
	$\phi_3(\phi_3 - \phi_1) - \phi_2 < 1$

voor θ_j en q gelden dezelfde condities.

5e kaart: (indien ICODE = 2)

formaat voor design matrix in F formaat

6e kaart en volgende:

- getransponeerde design matrix $_mX_N$ (m en N zijn via kaart 3 al bekend aan het programma)

N.B! In de design matrix legt de gebruiker de lineaire functie en de parameters vast voor de representatie van de interventie-effecten.

Een voorbeeld: indien een interventie op tijdstip 3 plaatsvindt en het model bevat een beginniveau L en een verandering in niveau δ dan is het model door de volgende lineaire functie weer te geven XB of

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \begin{bmatrix} L \\ \delta \end{bmatrix}$$

Door het vastleggen van de hier (0,1) design matrix X (via ICODE = 2) zijn nu L en δ te schatten.

$(6 + m)^e$ kaart:

F-formaat voor de invoer data
(vb. (8F10.4)
Na deze kaart volgen de data kaarten met de observaties in het opgegeven formaat.

Voor eventuele 2e en volgende problemen vanaf kaart 2 herhalen.

<u>Output TSX</u>

- afdrukken invoergegevens
- voor elke waarde van ϕ_j en/of θ_j

1). residuele error variantie
2). schatting van niveau van de reeks op t_0 + t-statistic (als ICODE $\leq$ 1)
3). schatting niveauverandering δ en t-statistic voor toetsen van significantie van het verschil in niveau (als ICODE $\leq$ 1)
4). schatting van de driftcomponent + t-statistic (als ICODE = 1)
5). schatting van de verandering en drift + t-statistic (als ICODE = 1)
6). schatting van elke design parameter + t-statistic (als ICODE = 2)

Na alle iteraties wordt de minimale errorvariantie geprint met bijbehorende waarden van ϕ_j en θ_j.

- optioneel residuals voor bepaalde waarde van ϕ_j en θ_j in formaat (10F8.4) op file PUNCH.
  N.B.: deze residuals zijn in Correl weer in te voeren; hoge chikwadraat waarden zijn nu een aanwijzing voor onnauwkeurige schattingen van de parameters van het ARIMA-model (immers, in dit geval bestaan de residuals niet uitsluitend uit ruis).

Scope setup.

Voor beide computerprogramma's is een C.M.-parameter in de jobkaart van 60.000 voldoende.

Literatuur:

Bower, C.P.; W.L. Padia en G.V. Glass
TMS: Two Fortran IV programs for Analysis of Time Series Experiments. Laboratory of Educational Research, University of Colorado, 1974.

Box, G.E.P. en G.M. Jenkins
Time Series Analysis: Forecasting and Control, San Fransisco. Holden Day, 1976.

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