

Marian BŁACHUTA

METODY IDENTYFIKACJI STACJONARNYCH I NIESTACJONARNYCH MODELI ARMA *

Streszczenie. W artykule przedstawiono modele ARMA stacjonarnych szeregów czasowych, ich reprezentacje w przestrzeni stanu, równania filtru Kalmana, równania predyktora optymalnego oraz estymację parametrów modelu za pomocą metody największej wiarygodności. Dokonano szerokiego przeglądu metod identyfikacji rzędu modelu $ARMA(p, q)$ oraz zaproponowano oryginalną modyfikację jednej z nich. Przedstawiono również modele ARIMA niestacjonarnych szeregów czasowych z wielomianowym trendem deterministycznym. Podano oryginalne algorytmy identyfikacji, estymacji i predykcji dla modeli niestacjonarnych.

IDENTIFICATION METHODS OF STATIONARY AND NONSTATIONARY TIME SERIES

Summary. In the paper, ARMA models of stationary time series are presented along with their state-space representations, Kalman filter equations, the optimal predictor, and the maximum-likelihood estimator of model parameters. Based on an extensive survey of the identification methods of the $ARMA(p, q)$ model order, an original modification of one of them has been proposed. ARIMA models of nonstationary time series with a polynomial deterministic trend are addressed, and original algorithms of their identification, parameter estimation and prediction are presented.

* Niniejsza praca została wykonana w ramach Projektu Badawczego KBN nr 3 P403 010 06 w czasie pobytu autora w The University of Birmingham.

1. Wprowadzenie

Dane ekonomiczne, inżynierskie, środowiskowe lub inne w naukach przyrodniczych są często podawane w zazwyczaj równo odległych odcinkach czasu, np. godzinach, dniach, miesiącach lub latach. Dane takie nazywa się *szeregiami czasowymi*.

W wielu zagadnieniach musimy rozważać zjawiska, w których występuje duża liczba nieznanych czynników i znalezienie modelu deterministycznego jest niemożliwe. Niemniej jednak na ogół można w takiej sytuacji znaleźć model pozwalający określić prawdopodobieństwo, że przyszła wartość zawarta będzie w określonych granicach. Model taki nazywamy *modelem probabilistycznym* lub *modelem stochastycznym*. Konstrukcja modeli stochastycznych na podstawie praw fizyki lub ekonomii na ogół nie jest możliwa i buduje się ją korzystając z procedur identyfikacji, bazujących na obserwacjach modelowanych zjawisk. W dalszym ciągu konieczne jest rozróżnienie modelu probabilistycznego lub *procesu stochastycznego* i obserwowanego szeregu czasowego. Tak więc szereg czasowy z_1, z_2, \dots, z_N , złożony z N kolejnych obserwacji, uważa się za *realizację próbkową* z nieskończonej populacji takich szeregów czasowych, które mogą być generowane przez proces stochastyczny.

Ze względu na inercję systemów dane szeregów czasowych są często skorelowane między sobą. Na przykład temperatura o danej godzinie ma na ogół wartość skorelowaną z temperaturą mierzoną godzinę wcześniej, wskaźnik zanieczyszczenia powietrza w południe może być silnie uzależniony od warunków pogodowych i natężenia ruchu rano, zaś wydatki ludzi w danym miesiącu mogą być silnie skorelowane z ich dochodami i wydatkami w poprzednich miesiącach.

Metody analizy takich ciągów obserwacji zależnych oraz tworzenie ich modeli matematycznych nazywa się *analizą szeregów czasowych*.

Możliwości optymalnego prognozowania, zrozumienia związków dynamicznych pomiędzy zmiennymi i optymalnego sterowania odcrowane przez modele szeregów czasowych mają doniosłe znaczenie w praktyce. Niniejsze opracowanie jest poświęcone typom modeli matematycznych szeregów czasowych, ich identyfikacji oraz wykorzystaniu do celów predykcji.

2. Model autoregresji i średniej ruchomej ARMA(p, q)

Szereg czasowy $z(t)$, będący realizacją procesu losowego, w którym kolejne wartości są silnie zależne, można rozpatrywać jako szereg generowany przez ciąg niezależnych zmiennych losowych $a(t)$ o rozkładzie normalnym z zerową wartością średnią i kowariancją σ^2 (Box, Jenkins, 1983).

Proces $a(t)$ oznacza się skrótowo jako n.i.d $(0, \sigma^2)$ (*normally independently distributed*).

Modelem zapewniającym elastyczność w dopasowaniu do rzeczywistego szeregu czasowego jest model ARMA(p, q):

$$\Phi(B)z(t) = \Theta(B)a(t), \quad (2.1)$$

w którym $\Phi(B)$ oraz $\Theta(B)$ są wielomianami

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \dots - \phi_i B^i - \dots - \phi_p B^p, \quad (2.2)$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \dots - \theta_i B^i - \dots - \theta_q B^q, \quad (2.3)$$

zaś B jest operatorem opóźnienia, tzn. $z(t-1) = Bz(t)$.

Wzór (2.1) można zapisać w równoważnej postaci:

$$z(t) = \sum_{i=1}^p \phi_i z(t-i) + a(t) - \sum_{i=1}^q \theta_i a(t-i). \quad (2.4)$$

W modelu tym bieżąca wartość procesu wyrażona jest przez skończoną kombinację liniową poprzednich wartości procesu (regresja względem swoich poprzednich wartości) oraz skończoną kombinacją liniową $a(t)$ i wartości poprzednich (średnia ruchoma).

Modelem stochastycznym szczególnie użytecznym przy opisywaniu pewnych praktycznie występujących szeregów jest model autoregresji AR(p) rzędu p (z ang. *AutoRegressive Process*):

$$\Phi(B)z(t) = a(t), \quad (2.5)$$

$$z(t) = \phi_1 z(t-1) + \phi_2 z(t-2) + \dots + \phi_p z(t-p) + a(t). \quad (2.6)$$

Procesy autoregresji mogą być stacjonarne lub niestacjonarne. Warunkiem koniecznym i wystarczającym stacjonarności procesu AR jest, aby wszystkie pierwiastki λ_i , ($i = 1, 2, \dots, p$) wielomianu $\Phi(\lambda)$ leżały na zewnątrz okręgu jednostkowego na płaszczyźnie zespolonej λ . Wielomian taki nazywa się *wielomianem stabilnym*.

Jeżeli wzór (2.5) zapiszemy w następujący sposób:

$$z(t) = \Psi(B)a(t), \quad (2.7)$$

to otrzymamy przedstawienie za pomocą nieskończonego wielomianu

$$\Psi(B) = \Phi^{-1}(B), \quad (2.8)$$

przy czym

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots + \psi_i B^i + \dots \quad (2.9)$$

Jeżeli wielomian $\Phi(B)$ jest stabilny, to ciąg wag $\psi_1, \psi_2, \psi_3, \dots$ jest zbieżny do zera, a proces $z(t)$ jest procesem stacjonarnym. Oznacza to, że wpływ poprzednich wartości $a(t)$ na aktualną wartość $z(t)$ maleje wraz z odległością od aktualnej chwili czasowej t .

W modelu średniej ruchomej $MA(q)$ (z ang. *Moving Average Process*) $z(t)$ zależy liniowo od $a(t)$ oraz skończonej liczby q poprzednich wartości $a(t-j)$, ($j = 1, 2, \dots, q$):

$$z(t) = \Theta(B)a(t). \quad (2.10)$$

Procesy średniej ruchomej są zawsze stacjonarne. Pojęciem dualnym do stacjonarności jest pojęcie *odwracalności*, ważne dla konstrukcji predyktora. Przedstawmy model (2.10) w postaci:

$$a(t) = \Pi(B)z(t), \quad (2.11)$$

gdzie

$$\Pi(B) = \Theta^{-1}(B) = 1 + \pi_1 B + \pi_2 B^2 + \pi_3 B^3 + \dots \quad (2.12)$$

Odpowiada to odwróceniu problemu. Obecnie na podstawie aktualnej i uprzednich wartości obserwowanego szeregu czasowego $z(t)$ staramy się odtworzyć wartości pobudzenia $a(t)$. Jeżeli ciąg wag $\pi_1, \pi_2, \pi_3, \dots$ jest zbieżny do zera, to model MA nazywamy modelem odwracalnym. Oznacza to, że wpływ poprzednich wartości $z(t)$ na aktualną wartość $a(t)$ maleje wraz z odległością od aktualnej chwili czasowej t .

Warunkiem koniecznym i wystarczającym odwracalności modelu MA jest, aby wszystkie pierwiastki λ_i , ($i = 1, 2, \dots, q$) wielomianu $\Theta(\lambda)$ leżały na zewnątrz okręgu jednostkowego na płaszczyźnie zespolonej λ .

Model $ARMA(p, q)$ można przedstawić również w postaci rozwinięcia nieskończonego, tzn.:

$$z(t) = \Psi(B)a(t) = \Phi^{-1}(B)\Theta(B)a(t) = (1 + \psi_1 B + \psi_2 B^2 + \dots)a(t), \quad (2.13)$$

gdzie:

$$\psi_0 = 1, \quad \psi_1 = \phi_1 - \theta_1, \quad \dots \quad \psi_s = \sum_{i=1}^{\min(p, s-1)} \phi_i \psi_{s-i} + \phi_s - \theta_s, \quad (s = 2, 3, \dots) \quad (2.14)$$

Podobnie można wyznaczyć rozwinięcie nieskończone

$$a(t) = \Pi(B)z(t) = \Phi(B)\Theta^{-1}(B)z(t) = (1 + \pi_1 B + \pi_2 B^2 + \dots)z(t). \quad (2.15)$$

Warunkiem koniecznym i wystarczającym stacjonarności modelu $ARMA$ jest stabilność wielomianu $\Phi(B)$. Podobnie warunkiem koniecznym i wystarczającym odwracalności modelu $ARMA$ jest stabilność wielomianu $\Theta(B)$.

3. Reprezentacja modelu ARMA(p, q) w przestrzeni stanu

Równaniom (2.1) modelu ARMA(p, q) można przypisać model w przestrzeni stanów o postaci:

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{g}a(t) \\ z(t) &= \mathbf{d}'\mathbf{x}(t) + a(t), \end{aligned} \quad (3.1)$$

gdzie na przykład wg Shea (1989):

$$\mathbf{A} = \begin{bmatrix} \phi_1 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\ \phi_2 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_r & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} \phi_1 - \theta_1 \\ \phi_2 - \theta_2 \\ \vdots \\ \phi_r - \theta_r \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (3.2)$$

przy czym $r = \max(p, q)$, zaś $\mathbf{x}(t)$ jest r -wymiarowym wektorem stanu. Model ten jest przydatny do konstrukcji równań predyktora optymalnego. Równania (3.1) można, eliminując z równania stanu wielkość $a(t)$, zapisać w postaci:

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{F}\mathbf{x}(t) + \mathbf{g}z(t), \\ a(t) &= z(t) - \mathbf{d}'\mathbf{x}(t), \end{aligned} \quad (3.3)$$

gdzie

$$\mathbf{F} = \mathbf{A} - \mathbf{g}\mathbf{d}'. \quad (3.4)$$

Łatwo sprawdzić, że wartości własne macierzy \mathbf{F} są pierwiastkami wielomianu średniej ruchomej $\lambda^q \Theta(\lambda^{-1})$.

Innym, alternatywnym modelem w przestrzeni stanu odpowiadającym modelowi ARMA(p, q) jest:

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{g}a(t+1) \\ z(t) &= \mathbf{d}'\mathbf{x}(t), \end{aligned} \quad (3.5)$$

gdzie (Mélard, 1984):

$$\mathbf{A} = \begin{bmatrix} \phi_1 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\ \phi_2 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{r-1} & 0 & \cdot & \cdot & \cdot & \cdot & 1 \\ \phi_r & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 1 \\ -\theta_1 \\ -\theta_2 \\ \vdots \\ -\theta_{r-1} \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (3.6)$$

przy czym $r = \max(p, q + 1)$, zaś $\mathbf{x}(t)$ jest r -wymiarowym wektorem stanu. Model ten, aczkolwiek o rzędzie wyższym o jeden od modelu poprzedniego, jest również przydatny do konstrukcji równań predyktora optymalnego.

4. Predykcja szeregu stacjonarnego

Podamy obecnie równania predyktora optymalnego, bazujące na opisie (3.5)-(3.6) w przestrzeni stanu. Punktem wyjścia dla predyktora k -krokowego jest optymalny predyktor 1-krokowy, określony przez filtr Kalmana. Założmy, że macierz kowariancji stanu w chwili początkowej $t = 0$ jest określona następująco:

$$\text{cov}(\mathbf{x}_0) = \mathbf{Q}_0 \sigma^2, \quad (4.1)$$

przy czym \mathbf{Q}_0 jest rozwiązaniem algebraicznego równania Lapunowa

$$\mathbf{Q}_0 = \mathbf{A} \mathbf{Q}_0 \mathbf{A}' + \mathbf{g} \mathbf{g}'. \quad (4.2)$$

Oznaczmy przez $\hat{\mathbf{x}}_{t|t-1}$ liniową ocenę stanu, zapewniającą minimum błędu średniokwadratowego stanu \mathbf{x}_t przy danych pomiarach do chwili $t - 1$, zaś kowariancję błędu oceny stanu $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}$ przez

$$\text{cov}(\tilde{\mathbf{x}}_t) = \Sigma_{t|t-1} \sigma^2. \quad (4.3)$$

Wówczas jednokrokowy predyktor stanu jest określony poprzez reprezentację innowacyjną. Dla $t = 1, 2, \dots, n$ błędy predykcji jednokrokowej ϵ_t oraz ich wariancja $\text{var}(\epsilon_t) = \sigma^2 f_t$ są określone zależnościami:

$$\epsilon_t = z_t - \mathbf{d}' \hat{\mathbf{x}}_{t|t-1}, \quad (4.4)$$

gdzie

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{A} \hat{\mathbf{x}}_{t|t-1} + \mathbf{k}_t \epsilon_t, \quad \hat{\mathbf{x}}_{0|-1} = \mathbf{0}, \quad (4.5)$$

$$\mathbf{k}_t = \frac{\mathbf{A} \Sigma_{t|t-1} \mathbf{d}'}{\mathbf{d}' \Sigma_{t|t-1} \mathbf{d}}, \quad (4.6)$$

$$\Sigma_{t+1|t} = \mathbf{A} \left(\Sigma_{t|t-1} - \frac{\Sigma_{t|t-1} \mathbf{d} \mathbf{d}' \Sigma_{t|t-1}}{\mathbf{d}' \Sigma_{t|t-1} \mathbf{d}} \right) \mathbf{A}' + \mathbf{g} \mathbf{g}', \quad \Sigma_{0|-1} = \mathbf{Q}_0 \quad (4.7)$$

oraz

$$f_t = \mathbf{d}' \Sigma_{t|t-1} \mathbf{d}. \quad (4.8)$$

Korzystając z powyższych zależności można określić k -krokową predykcję wyjścia $\hat{z}_{t+k|t}$ i wariancję jej błędu $\tilde{z}_{t+k|t} = z_{t+k} - \hat{z}_{t+k|t}$ następująco:

$$\hat{z}_{t+k|t} = \mathbf{d}' \hat{\mathbf{x}}_{t+k|t} = \mathbf{d}' \mathbf{A}^{k-1} \hat{\mathbf{x}}_{t+1|t} = \mathbf{d}'_k \hat{\mathbf{x}}_{t+1|t} \quad (4.9)$$

$$\sigma_{t+k|t}^2 = (\mathbf{d}'_k \Sigma_{t+1|t} \mathbf{d}_k + \sum_{j=0}^{k-1} e_j^2) \sigma^2, \quad e_0 = 1, \quad e_j = \mathbf{d}'_j \mathbf{g}, \quad j = 1, 2, \dots \quad (4.10)$$

Dla $k = r + 1, r + 2 \dots k$ -krokowe predykcje spełniają równanie rekurencyjne:

$$\hat{z}_{t+k|t} = \phi_1 \hat{z}_{t+k-1|t} + \phi_2 \hat{z}_{t+k-2|t} + \dots + \phi_r \hat{z}_{t+k-r|t} \quad (4.11)$$

Przy $t \rightarrow \infty$ zachodzi $\Sigma_{t|t-1} \rightarrow gg'$, $k_t \rightarrow g^* = Ag$ i równania predyktora przyjmują postać asymptotyczną:

$$\hat{x}_{t+1|t} = F^* \hat{x}_{t|t-1} + g^* z_t, \hat{x}_{0|t-1} = 0, \quad (4.12)$$

gdzie:

$$F^* = \begin{bmatrix} \theta_1 & 1 & \dots & \dots & \dots & 0 \\ \theta_2 & 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{r-1} & 0 & \dots & \dots & \dots & 1 \\ \theta_r & 0 & \dots & \dots & \dots & 0 \end{bmatrix}, g^* = \begin{bmatrix} \phi_1 - \theta_1 \\ \phi_2 - \theta_2 \\ \vdots \\ \phi_{r-1} - \theta_{r-1} \\ 0 \end{bmatrix}, d = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (4.13)$$

Predyktor ten, aczkolwiek często używany, nie jest predyktorem optymalnym i jest stabilny tylko dla modeli odwracalnych, w przeciwieństwie do predyktora Kalmana, pracującego poprawnie również dla modeli nieodwracalnych.

5. Estymator ML parametrów modeli ARMA

Najefektywniejszym estymatorem parametrów jest estymator największej wiarygodności.

Ponieważ innowacje uzyskiwane z filtru Kalmana są ciągiem zmiennych losowych niezależnych o wariancji f_t , więc funkcja wiarygodności dla szeregu czasowego $z = \{z_t, t = 1..T\}$ generowanego przez model ARMA jest określona wzorem:

$$\ln L(z, \theta, \phi, \sigma^2) = \text{const} - T/2 \ln \sigma^2 - 1/2 \sum_{t=1}^T \ln f_t - 1/(2\sigma^2) \sum_{t=1}^T \epsilon_t^2 / f_t \quad (5.1)$$

Oceny największej wiarygodności parametrów $\hat{\theta}, \hat{\phi}, \hat{\sigma}^2$ znajduje się maksymalizując $\ln L(z, \theta, \phi, \sigma^2)$.

Zależność (5.1) jest ogólną postacią funkcji wiarygodności, niezależną od metody jej obliczenia. Praktyczne algorytmy obliczania funkcji wiarygodności korzystają na ogół z innych algorytmów filtracji, mniej pracochłonnych od rekurencji równań (4.6)-(4.8). Jeden z takich algorytmów jest zamieszczony w pracy Méléarda (1984). Osobnym istotnym problemem jest metodyka minimalizacji (5.1) i składające się na nią elementy: wyznaczenie wstępnych ocen parametrów oraz obliczanie gradientu i hesjanu funkcji wiarygodności względem estymowanych parametrów. W literaturze (np. Burshtein, 1993; Kohn, Ansley, 1982) istnieją analityczne algorytmy wyznaczania gradientu funkcji wiarygodności. Jedną z metod wyznaczania wstępnych ocen parametrów jest opisana w punktach 6.5 i 6.6. Pozostałe problemy wykraczają jednak poza zakres niniejszego opracowania.

6. Identyfikacja modeli ARMA

Identyfikacja modelu ARMA polega na ocenie stopnia p wielomianu autoregresyjnego $\Phi(B)$, stopnia q wielomianu średniej ruchomej $\Theta(B)$ oraz wartości współczynników $\phi_1 \dots \phi_p$, $\theta_1 \dots \theta_q$ na podstawie realizacji $z(1) \dots z(T)$.

Określenie rzędu procesu (p, q) jest ważną, trudną i delikatną częścią analizy szeregu czasowego. Analitycy szeregów czasowych często stosują podejście zaproponowane przez Boxa i Jenkinsa (1983). Metoda ta bazuje na wyciąganiu wniosków z charakterystycznych kształtów funkcji autokorelacji i korelacji cząstkowych estymowanych na podstawie danego szeregu czasowego. Podejście to w znacznym stopniu polega na subiektywnych ocenach analityka. W tym rozdziale dokonamy przeglądu najważniejszych alternatywnych, mniej heurystycznych, metod określania rzędu modelu ARMA. Jednym z kryteriów klasyfikacji metod jest przyporządkowanie im określenia "subiektywna" lub "obiektywna".

W przypadku metody subiektywnej decyzja jest zawsze pozostawiona analitykowi. Może ona polegać na wyborze poziomu istotności testu lub wizualnym zbadaniu wykresów lub tablic w celu uchwycenia charakterystycznego wzoru w zachowaniu analizowanej charakterystyki.

Metody obiektywne mogą być scharakteryzowane przez fakt, że udział analityka w procesie modelowania nie jest niezbędny, zaś oceny rzędu (p, q) modelu ARMA (p, q) dokonuje się zazwyczaj poszukując ekstremum pewnej funkcji.

Klasa metod subiektywnych może być podzielona na metody bazujące na testowaniu hipotez statystycznych oraz na metody bazujące na teorii realizacji stochastycznych. Wśród klasy metod obiektywnych należy wymienić metody bazujące na błędzie jednokrokowej predykcji, miarach informacyjnych i metodach bayesowskich.

6.1. Metody bazujące na teorii testowania hipotez statystycznych

Jednym z najczęstszych problemów wnioskowania statystycznego jest zdecydowanie, na bazie skończonego zbioru obserwacji, czy zbiór parametrów β spełnia s niezależnych ograniczeń $h_i(\beta) = 0$, $(i = 1, 2, \dots, s)$.

Załóżmy, że wszystkie ograniczenia polegają na zerowaniu się części parametrów. Przyjmijmy, że wektor parametrów β jest podzielony na dwie części: $(\beta'_1, \beta'_2)'$, a hipotezą zerową jest $H_0 : \beta_2 = (0, 0, \dots, 0)'$. Wówczas, pod warunkiem H_0 , dowolna ze statystyk: LR (*Likelihood Ratio*), W (*Walda*) czy też LM (*Lagrange Multiplier*) posiada asymptotyczny rozkład χ^2 o s stopniach swobody (Harvey, 1981b).

Jest jasne, że H_0 jest zagnieżdżona w hipotezie alternatywnej poprzez ograniczenie $\beta_2 = (0, 0, \dots, 0)'$. Hipotezy tworzą zbiór jednoznacznie uporządkowany, co pozwala na

przeprowadzenie sekwencyjnej procedury testującej. T.W. Anderson (1971) przedstawił procedurę sekwencyjną do ustalania rzędu procesu AR. Zakłada ona określenie liczby L takiej, że prawdziwy rząd jest mniejszy lub równy L . Następnie są testowane sekwencyjnie wzajemnie wykluczające się hipotezy:

$$\begin{aligned} H_1 &: \Phi_L = 0, \\ H_2 &: \Phi_L = \Phi_{L-1} = 0, \\ &\vdots \\ H_L &: \Phi_L = \Phi_{L-1} = \dots = \Phi_1 = 0. \end{aligned} \quad (6.1)$$

Jeśli jakakolwiek hipoteza jest prawdziwa, poprzedzające hipotezy są też prawdziwe, jeśli jakakolwiek z tych hipotez jest fałszywa, następne są również fałszywe. W ostatnim przypadku procedura kończy się.

Whittle (1954) pokazał, że dla testowania dwu zagnieżdżonych hipotez - H_0 : założony model jest autoregresją rzędu s , ($s = 0, 1, 2, \dots, L-1$) przeciwko H_1 : dana realizacja jest generowana przez proces $AR(L)$ - statystyka testowa LR przyjmie w przybliżeniu postać:

$$LR = -(T-L) \log(\hat{\sigma}_a^2 / \hat{\sigma}_L^2), \quad (6.2)$$

gdzie $\hat{\sigma}_a^2$ i $\hat{\sigma}_L^2$ są ocenami największej wiarygodności odpowiednio przy H_0 i H_1 . Statystyka ta posiada asymptotycznie rozkład χ^2 o $L-s$ stopniach swobody.

Anderson (1971) zaproponował statystykę bazującą na testowaniu hipotezy zerowej $\Phi_{s+1} = \dots = \Phi_L = 0$, ($s = 0, 1, 2, \dots, L-1$) w modelu $AR(L)$. Niech R_L będzie macierzą $L \times L$ wymiarową, podzieloną na s i $L-s$ wierszy i kolumn odpowiednio do wektora parametrów:

$$R_L = \begin{bmatrix} R_1 & R_{12} \\ R'_{12} & R_2 \end{bmatrix}, \quad \hat{\Phi}_L = \begin{bmatrix} \hat{\Phi}^{(1)} \\ \hat{\Phi}^{(2)} \end{bmatrix}, \quad (6.3)$$

gdzie (i, j) -ty element R_L , r_{i-j} jest funkcją autokorelacji z próby, zdefiniowaną jako:

$$r_k = \frac{\sum_{t=1}^{T-k} z_t z_{t+k}}{\sum_{t=1}^{T-k} z_t^2} \quad (6.4)$$

i gdzie $\hat{\Phi}$ jest wektorem ocen parametrów modelu $AR(L)$. Wówczas statystyka

$$T \hat{\Phi}^{(2)'} [R_2 - R'_{12} R_1^{-1} R_{12}] \hat{\Phi}^{(2)} / \sigma_L^2 \quad (6.5)$$

jest statystyką testu Walda.

Po zastąpieniu parametrów ϕ_i przez θ_i ; sekwencyjna procedura testująca może być także użyta dla testowania hipotez w czystym modelu MA.

Procedura testowania sekwencyjnego ma niestety dwie wady. Pierwszą jest to, że rząd modelu AR lub MA konieczny do właściwego dopasowania się do szeregu czasowego może być duży. Drugą wadą jest trudność znalezienia właściwego poziomu istotności na bazie kompromisu pomiędzy prawdopodobieństwami błędu I i II rodzaju, ponieważ to drugie nie jest znane dla rozważanego testu.

6.2. Metody bazujące na teorii realizacji stochastycznych

Rozpatrzmy tu procedury nie wymagające, w przeciwieństwie do poprzednich, uprzedniej estymacji parametrów modelu.

6.2.1. Odwrotne autokorelacje i korelacje cząstkowe

Chatfield (1979) i Cleveland (1972) przedstawili argumenty na rzecz używania w procedurze Boxa-Jenkinsa odwrotnych autokorelacji IAC (*Inverse Autocorrelations*) zamiast autokorelacji.

Odwrotna autokowariancja dla k -tego opóźnienia, oznaczana jako $\gamma_i(k)$, może być zdefiniowana jako współczynnik przy B^k , w funkcji generującej $\Gamma_i(B)$, spełniającej warunek $\Gamma_i(B)\Gamma(B) = 1$. Jest zatem oczywiste, że dla procesu ARMA(p, q):

$$\Gamma_i(B) = \frac{\Phi_p(B)\Phi_p(B^{-1})}{\Theta_q(B)\Theta_q(B^{-1})} \frac{1}{\sigma_a^2}. \quad (6.6)$$

Z kolei odwrotne autokorelacje są zdefiniowane jako $\rho_i(k) = \gamma_i(k)/\gamma_i(0)$. Implikuje to, że odwrotne autokorelacje dla procesu AR(p) są określone następującym wzorem:

$$\rho_i(k) = (-\Phi_k + \sum_{j=1}^{p-k} \Phi_j \Phi_{j+k}) / (1 + \sum_{j=1}^p \Phi_j^2), \quad k \leq p \quad (6.7)$$

oraz

$$\rho_i(k) = 0, \quad \text{dla } k > p. \quad (6.8)$$

W związku z tym w przypadku autoregresyjnym $\rho_i(k)$ jest ucięte dla $k > p$. Własność ta czyni $\rho_i(k)$ konkurencyjnym w stosunku do PACF (*Partial Autocorrelation Function*). Alternatywną funkcję, znaną jako odwrotna funkcja autokorelacji cząstkowych IPACF (*Inverse Partial Autocorrelation Function*) omawia Chatfield (1979). Można ją zdefiniować jako PACF dla modelu odwrotnego $\Theta_q(B)z_t = \Phi_p(B)a_t$ i obliczyć z równań Youle'a - Walkera poprzez zastąpienie autokorelacji autokorelacjami odwrotnymi. Zachowanie się IPACF jest podobne do zachowania się funkcji autokorelacji.

6.2.2. Metoda narożnikowa

Metodą wykorzystującą autokorelacje do wygenerowania tablicy dla identyfikacji rzędu modelu ARMA jest tzw. metoda kątowna zaproponowana przez Beguina, Gourieroux i Monforta (1980). Bazą dla tej metody jest wyznacznik $\Delta(i, j)$ macierzy $j \times j$ -wymiarowej, zdefiniowany jako

$$\Delta(i, j) = \begin{vmatrix} \rho_i & \rho_{i-1} & \cdots & \rho_{i-j+1} \\ \rho_{i+1} & \rho_i & \cdots & \rho_{i-j+2} \\ \cdot & \cdot & \cdots & \cdot \\ \rho_{i+j-1} & \rho_{i+j-2} & \cdots & \rho_i \end{vmatrix}, \quad (i, j = 1, 2, \dots, L). \quad (6.9)$$

Beguina i inni (1980) dowodzą, że proces stacjonarny ma minimalną reprezentację ARMA(p, q) wtedy i tylko wtedy, gdy $\Delta(i, j) = 0$ dla wszystkich wartości $i \geq q + 1$ oraz $j \geq p + 1$, podczas gdy $\Delta(i, p) \neq 0$ dla wszystkich wartości $i \geq q$ oraz $\Delta(q, j) \neq 0$ dla wszystkich wartości $j \geq p$. Tablica wartości $\Delta(i, j)$ nazywa się tablicą Δ . Identyfikacja p i q jest równoważna identyfikacji narożnika wartości zerowych w tablicy Δ .

6.2.3. Uogólnione autokorelacje

Uogólnioną funkcję autokorelacji użyteczną dla określenia rzędu procesu ARMA zaproponował Takemura (1984). Oznaczając: $\gamma(j, i) = (\gamma_j, \gamma_{j+1}, \dots, \gamma_{j-i+1})'$, dla $i > 0$, oraz

$$\Gamma(i, j) = \begin{vmatrix} \gamma_i & \gamma_{i-1} & \cdots & \gamma_{i-j+1} \\ \gamma_{i+1} & \gamma_i & \cdots & \gamma_{i-j+2} \\ \cdot & \cdot & \cdots & \cdot \\ \gamma_{i+j-1} & \gamma_{i+j-2} & \cdots & \gamma_i \end{vmatrix}, \quad (i, j = 1, 2, \dots, L) \quad (6.10)$$

gdzie γ_j - autokowariancja teoretyczna, uogólnione autokorelacje wyrażą się wzorami:

$$\rho(i + 1, j + 1) = 0,$$

dla $\det \Gamma(j, i) = 0$ oraz

$$\begin{aligned} \rho(i + 1, j + 1) &= \gamma(i + j + 1) - \gamma(j + 1, i)'\Gamma(j, i)^{-1}\gamma(j + 1, i) \\ &\times [\gamma_0 - 2\gamma(j, i)'\Gamma(j, i)^{-1}\gamma(j + 1, i) \\ &+ \gamma(j + 1, i)\Gamma(j, i)^{-1}\Gamma(0, i)\Gamma(j, i)^{-1}\gamma(i + j, i)]^{-1} \end{aligned} \quad (6.11)$$

dla $\det \Gamma(j, i) \neq 0$.

Takemura (1984) dowodzi, że $\rho(i, j)$ ma następujące własności: $\rho(1, j) = \rho_j$ dla $j \geq 1$, $\rho(i, 1) = \phi_{ii}^{(0)}$ dla $i \geq 1$; $-1 \leq \rho(i, j) \leq 1$ oraz dla procesu ARMA(p, q) $\rho(i, j) = 0$ dla $i > p$ oraz $j > q$. Jest oczywiste, że własność ta może być wykorzystana do identyfikacji rzędu (p, q) procesu. Wówczas autokowariancje teoretyczne γ_i są zastępowane

przez autokowariancje z próby. Takemura pokazał, że jeśli przez $r(i+1, j+1)$ oznaczymy uogólniony współczynnik autokorelacji z próby, to dla procesu ARMA(p, q) o $\phi_p \neq 0$ i $\theta_q \neq 0$ statystyka:

$$n^{1/2}r(p+1, q+1)/[1 + 2 \sum_{i=1}^q (r_i^2)^{1/2}]^{1/2} \quad (6.12)$$

posiada asymptotycznie standardowy rozkład normalny, przy czym r_i^2 oznacza zgodną ocenę $\gamma_i^2 = \text{cov}(x_t, x_{t-1})$, gdzie $x_t = \Phi_p(B)z_t = \Theta_q(B)a_t$.

6.2.4. Iterowane regresje i rozszerzona funkcja autokorelacji z próby

Tiao i Tsay (1983a, b) oraz Tsay i Tiao (1984) użyli zgodnych ocen parametrów AR stacjonarnego lub niestacjonarnego procesu ARMA(p, q) dla zdefiniowania *rozszerzonej funkcji autokorelacji z próby* (ESACF - *Extended Sample Autocorrelation Function*), na podstawie której można określić rząd procesu ARMA(p, q). Można pokazać, że $\hat{\phi}_{p(l)}^{(j)} \rightarrow \phi_l$, dla $l = 1, 2, \dots, p$, jeśli $j \geq q$, gdzie $\hat{\phi}_{p(l)}^{(j)}$ jest oceną l -tego parametru wielomianu AR j -tej iterowanej regresji AR(p). Dla ułatwienia obliczeń oceny tych parametrów są otrzymywane rekursywnie dla $j \geq 1$ za pomocą formuły:

$$\hat{\phi}_{p(0)}^{(j-1)} = -1, \quad \hat{\phi}_{p(l)}^{(j)} = \hat{\phi}_{p+1(l)}^{(j-1)} - \hat{\phi}_{p(l-1)}^{(j-1)} \hat{\phi}_{p+1(l+1)}^{(j-1)} / \hat{\phi}_{p(p)}^{(j-1)}. \quad (6.13)$$

Aby obliczyć ESACF konieczne jest obliczenie zwykłych ocen najmniejszych kwadratów $\hat{\phi}_{p(l)}^{(0)}$ dla $p = 1, 2, \dots, p_0 + q_0$ oraz $l = 1, 2, \dots, p$ poprzez sukcesywne dopasowania AR, poczynając od AR(1) do AR($p_0 + q_0$), gdzie p_0 i q_0 są wartościami uprzednio określonymi. $\hat{\phi}_{p(l)}^{(j)}$ oblicza się następnie za pomocą formuły rekurencyjnej dla $p = 1, 2, \dots, p_0$, $l = 1, 2, \dots, p$ oraz $j = 1, 2, \dots, q_0$.

Dla skończonych wartości m , m -ty ESACF jest definiowany następująco:

$$r_{k(m)} = \text{cov}(y_t y_{t+k}) / \text{cov}(y_t y_t), \quad (6.14)$$

gdzie

$$y_t = z_t - \sum_{l=1}^m \hat{\phi}_{m(l)}^{(k)} z_{t-l}. \quad (6.15)$$

Tsay i Tiao (1984) pokazują, że dla procesu ARMA(p, q) ESACF posiada następującą własność asymptotyczną:

$$\begin{aligned} r_{k(j)} &= c(j-p, k-q), \quad (0 \leq k-q \leq j-p) \\ r_{k(j)} &= 0, \quad (k-q > j-p \geq 0), \end{aligned} \quad (6.16)$$

gdzie $c(j-p, k-q)$ jest niezerową stałą lub zmienną losową ograniczoną między -1 a 1 . Własność ta może być użyta do zidentyfikowania rzędu procesu (p, q).

Dla skończonych wartości T , nie wszystkie $r_{k(j)}$ będą zerowe. Jako dodatkowego narzędzia do określania momentu ucięcia $r_{k(j)}$ Tiao i Tsay proponują użycie formuły Bartletta $(n - j - k)^{-1}$ jako asymptotycznej aproksymacji wariancji $r_{k(j)}$. Bazuje to na hipotezie, że szereg przetransformowany y_t jest białym szumem.

6.3. Metody bazujące na błędzie jednokrokowej predykcji

Trudności w określeniu rzędu modelu ARMA pozwalają na stwierdzenie, że na podstawie szeregu czasowego o skończonej długości nie można oczekiwać wyraźnej odpowiedzi na temat prawdziwego rzędu procesu. Ponadto założenie, że proces rzeczywisty jest istotnie procesem ARMA(p, q), jest wygodne tylko koncepcyjnie.

Jeśli celem identyfikacji jest znalezienie modelu najlepiej aproksymującego proces, wówczas rozsądnie jest robić to mając na uwadze cel, któremu model ma służyć. Typowym celem jest predykcja przyszłych wartości procesu.

Akaike (1970) zaproponował metodę wybierającą rząd p autoregresji tak, aby minimalizować wartość oczekiwaną kwadratu błędu jednokrokowej predykcji FPE (*Final Prediction Error*). FPE Akaike jest wówczas wyrażone zależnością:

$$FPE(p) = \hat{\sigma}_a^2 \frac{T+p}{T-p}, \quad (6.17)$$

przy czym wyboru rzędu dokonuje się tak, aby $FPE(p)$, ($p = 1, 2, \dots, L$) osiągnęło wartość minimalną. Jednakże dla dużych wartości T wyrażenie $(T+p)/(T-p)$ jest zależne od p w znikomym stopniu i w konsekwencji trudno jest dobrać właściwy rząd modelu. W istocie dla szeregów dłuższych od 50 kryterium to prowadzi do nadparametryzacji. Söderström (1977) pokazał, że FPE może być użyte niezależnie od struktury modelu poprzez zastąpienie p przez liczbę estymowanych parametrów.

Z kolei Bhansali i Downham (1977) zasugerowali wybór rzędu modelu na bazie minimalizacji funkcji

$$FPE_\delta(p) = \hat{\sigma}_a^2 (1 + \delta p/T), \quad (p = 0, 1, 2, \dots, L), \quad (6.18)$$

gdzie $\delta \geq 2$ jest stałą. Przy zwiększaniu δ zwiększa się kara za nadparametryzację. Bhansali i Downham pokazali, że dla $T \geq 300$ właściwy model jest znajdowany znacznie częściej przy $\delta > 2$.

6.4. Metody bazujące na mierze informacji

Koncepcja informacji wg Fishera i związane z nią pojęcie entropii stało się bazą szeregu metod estymacji rzędu. Akaike (1973, 1974) pierwszy dostrzegł znaczenie entropii dla

estymacji rzędu modelu ARMA. W szczególności pokazał on, że minimalizacja funkcji

$$AIC(p, q) = T \log \hat{\sigma}_a^2 + 2(p + q), \quad (6.19)$$

(Akaike Information Criterion) jest równoważna minimalizacji miary informacji Kullbacka-Leiblera.

Jednakże Hannan (1982) udowodnił, że jeśli (p^*, q^*) jest prawdziwym rzędem modelu ARMA(p, q) oraz $p^* < P$ i $q^* < Q$, wówczas AIC asymptotycznie przecestymowuje (p^*, q^*) z niezerowym prawdopodobieństwem.

6.5. Metody bayesowskie

Metody bayesowskie dla określenia rzędu modelu szeregu czasowego biorą pod uwagę wiedzę a priori w postaci funkcji gęstości prawdopodobieństwa o parametrach modelu. Wychodząc z tej metody, Schwarz (1978) dla wyboru najodpowiedniejszego modelu w zbiorze modeli ARMA(p, q), ($p, q = 0, 1 \dots L$) zaproponował minimalizację funkcji:

$$S(p, q) = T \log \hat{\sigma}_a^2 + (p + q) \log T. \quad (6.20)$$

Z kolei Rissanen (1978) do określania rzędu modelu ARMA zaproponował minimalizację funkcji

$$BIC^* = \log \hat{\sigma}_a^2 + (p + q)(\log T)/T, (p, q = 0, 1 \dots L). \quad (6.21)$$

Jego wyprowadzenie bazuje na zasadzie minimalizacji liczby liczb dwójkowych potrzebnych do odtworzenia obserwowanych danych, gdy każda obserwacja jest dana z pewną precyzją, z zastosowaniem najlepszego kodowania i założonego modelu procesu. Zauważmy, że BIC^* (Bayesian Information Criterion) jest w istocie kryterium Schwarza.

Hannan (1980) oraz Hannan i Rissanen (1982) wykazali, że BIC^* daje zgodne oceny rzędu modelu ARMA.

Dwa inne kryteria oceny rzędu procesu mające charakter bayesowski to: bayesowskie kryterium estymacji BEC (Bayesian Estimation Criterion) Geweke i Meese'a (1981) oraz kryterium HQ Hannana i Quinna (1979) i Hannana (1980). Pierwsze z nich ma postać:

$$BEC(p, q) = \hat{\sigma}_a^2 + (p + q)\hat{\sigma}_L^2 \log T / (T - L), (p, q = 0, 1, \dots L), \quad (6.22)$$

gdzie $\hat{\sigma}_L^2$ jest oceną największej wiarygodności wariancji reszt po dopasowaniu modelu najwyższego rzędu.

Hannan i Quinn sugerują poszukiwanie rzędu modelu poprzez minimalizację wielkości:

$$HQ(p, q) = \log \hat{\sigma}_a^2 + (p + q)c(\log \log T)/T, (p, q = 0, 1 \dots L), \quad (6.23)$$

gdzie c jest stałą, którą należy określić. W pracy Hannana (1980) pokazano, że człon $(\log \log T)$ raczej wykreśla linię pomiędzy estymatorami zgodnymi i niezgodnymi niż jest rezultatem, który należałoby wykorzystywać.

Hannan i Rissanen (1982) zaproponowali praktyczną procedurę określania rzędu modelu ARMA(p, q) bazującą na trzech etapach.

W pierwszym etapie do danych dopasowuje się wielomian AR(L), gdzie L dobiera się z kryterium AIC lub BIC. Oceny parametrów $\hat{\phi}_j$, ($j = 1, \dots, L$) służą do określenia ocen $\hat{a}_t = z_t - \sum \hat{\phi}_j z_{t-j}$ dla $t \geq L + 1$, przy czym suma jest po $j = 1, 2, \dots, L$. W drugim etapie znajduje się oceny parametrów poprzez regresję z_t na z_{t-j} , ($j = 1, 2, \dots, p$) oraz na \hat{a}_{t-j} , ($j = 1, 2, \dots, q$). Aby to wykonać, proponują oni rekurencyjny algorytm estymacji, który można uważać za uogólnienie algorytmu Durбина-Levinsona. Jeśli przez $\hat{\sigma}_{p,q}^2$ oznaczyć wariancję reszt regresji, to ocena (\bar{p}, \bar{q}) rzędu jest określona poprzez minimalizację funkcji:

$$\log \hat{\sigma}_{p,q}^2 + (p + q)(\log T)/T, \quad (p \leq P, Q), \quad (6.24)$$

gdzie P, Q są uprzednio założonymi, odpowiednio dużymi liczbami całkowitymi.

Ponieważ współczynniki $\hat{\phi}_i, \hat{\theta}_i$, dla których (6.24) przyjmuje minimum, nie są asymptotycznie efektywne, Hannan i Rissanen (1982) sugerują estymację parametrów za pomocą dowolnego algorytmu maksymalizacji gaussowskiej funkcji wiarygodności, przyjmując wartości (\bar{p}, \bar{q}) dla rzędu oraz $\hat{\phi}_i, \hat{\theta}_i$ jako początkowe wartości parametrów. Jest to trzeci etap procedury.

Okazuje się, że na drugim etapie procedury może dojść do nadestymacji rzędu. Hannan i Kavalieris (1984) wprowadzili do swej oryginalnej procedury poprawki prowadzące do zgodnych ocen rzędu.

6.6. Zmodyfikowana metoda Hannana-Rissanena

Modyfikacje metody Hannana-Rissanena zaproponowane w tej pracy są związane z modyfikacją drugiego i trzeciego etapu procedury. Polegają one na tym, że metodą najmniejszych kwadratów wyznacza się modele ARMA(r, r) dla $r = 1, 2, \dots, r_{\max}$, co prowadzi do ustalenia $P = Q$; stosuje się pełną metodę najmniejszych kwadratów a nie jej uproszczony wariant rekurencyjny. W celu ostatecznej selekcji najlepszego modelu wyznacza się dokładną funkcję wiarygodności dla kilku modeli wytypowanych w poprzednich etapach.

Pewną modyfikacją techniczną jest korzystanie z eksponent funkcji AIC oraz BIC*, oznaczanych nadal jako AIC oraz BIC. Zastosowane modyfikacje znacznie podwyższają efektywność estymatora kosztem podniesienia złożoności obliczeniowej, co jednak przy

właściwej organizacji obliczeń dla metody najmniejszych kwadratów, modyfikującej jedynie pewne macierze wyznaczone dla modeli niższych rzędów zamiast obliczenia tych macierzy dla modeli o rzędach wyższych oraz efektywnych algorytmach wyznaczania funkcji wiarygodności jest szybka nawet przy obliczeniach prowadzonych na komputerach klasy PC (Błachuta, 1994).

Procedura identyfikacji jest następująca:

- (1) Dopasowanie metodą najmniejszych kwadratów wielomianu autoregresyjnego $\Phi_n(B)$ stopnia n . Liczbę n określa się dopasowując kolejno wielomiany stopnia $n = 0, \dots, n_{\max}$ i wybierając model dający minimalną wartość AIC

$$AIC = \sigma_n^2 \exp \frac{2n}{T}. \quad (6.25)$$

Wyznaczenie ocen innowacji

$$\hat{\epsilon}_t = \Phi_n(B)y_t, t \geq n + 1. \quad (6.26)$$

- (2) Dopasowanie metodą najmniejszych kwadratów na podstawie szeregów y_t i $\hat{\epsilon}_t$ ciągu modeli ARMA(r, r), dla $r = 0, 1, \dots, r_{\max}$.

Znalezienie r_0 minimalizującego BIC

$$BIC = \hat{\sigma}_r^2 \exp \left\{ 2r \frac{\log T}{T} \right\}, r = 0, 1, \dots, r_0 + 1. \quad (6.27)$$

- (3) Wyznaczenie metodą najmniejszych kwadratów modeli ARMA(p, q) dla wszystkich $p \leq r_0 + 1$ oraz $q \leq r_0 + 1$. Selekcja modeli o najmniejszych wartościach BIC

$$BIC = \hat{\sigma}_{p,q}^2 \exp \left\{ (p + q) \frac{\log T}{T} \right\}. \quad (6.28)$$

- (4) Estymacja metodą największej wiarygodności parametrów modeli wyselekcjonowanych w etapie 3. Wybór modelu o najmniejszej wartości BIC.

7. Procesy niestacjonarne

Modele (2.3) i (3.1) są stacjonarne, jeśli wszystkie zera wielomianu $\phi(B)$ leżą na zewnętrznej kółu jednostkowego. Teoretycznie stacjonarność oznacza, że funkcje gęstości prawdopodobieństwa $[z(t_1), z(t_1+1), \dots, z(t_1+k)]$ oraz $[z(t_2), z(t_2+1), \dots, z(t_2+k)]$ posiadają identyczną postać dla dowolnego wyboru parametrów (t_1, t_2, k) . W praktyce oznacza to, że zachowanie się wyników obserwacji pozostaje takie samo w czasie. Jednakże rzeczywiste szeregi czasowe często wykazują zachowanie dryfujące. Takie niestacjonarne szeregi

czasowe mogą być zamodelowane poprzez dopuszczenie niektórych zer $\Phi(B)$ na okręgu jednostkowym. Pisząc zatem $\Phi(B) = \varphi(B)(1 - B)^d$ mamy:

$$\varphi(B)(1 - B)^d z(t) = \mu + \Theta(B)a(t), \quad (7.1)$$

gdzie zera $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ leżą na zewnątrz okręgu jednostkowego.

Model (7.1) jest znany jako model ARIMA($p - d, d, q$) rzędu ($p - d, d, q$). Również niektóre szeregi czasowe z zachowaniem cyklicznym mogą być reprezentowane poprzez $\Phi(B)$, posiadające zera na okręgu jednostkowym. Tak więc (7.1) może być zapisane ogólniej jako:

$$\varphi(B)U(B)z(t) = \mu + \Theta(B)a(t), \quad (7.2)$$

gdzie $U(B) = 1 - U_1 B - U_2 B^2 - \dots - U_s B^s$ posiada wszystkie zera na okręgu jednostkowym.

Specjalna postać $U(B) = (1 - B)^d(1 - B^s)^{d_1}$, gdzie s jest pewną liczbą całkowitą dodatnią, jest szeroko stosowana w praktyce do modelowania sezonowych szeregów czasowych (np. w ekonometrii $s = 12$ dla danych miesięcznych oraz $s = 4$ dla danych kwartalnych).

8. Procesy niestacjonarne z trendem deterministycznym

Rozważmy proces opisany równaniem:

$$y_t = \frac{\Theta(B)}{\Delta^d \Phi(B)} a_t + \delta_0 + \delta_1 t + \dots + \delta_d t^d + \dots + \delta_{d+m-1} t^{d+m-1}, \quad (8.1)$$

w którym:

B - operator opóźnienia, $\Delta = 1 - B$ - operator różnicy, t - czas dyskretny,

a_t - n.i.d ($0, \sigma^2$) oraz $\Phi(B), \Theta(B)$ - stabilne wielomiany zmiennej B :

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

Weźmy d -krotną różnicę szeregu czasowego y_t . Proces różnic v_t będzie wówczas opisany równaniem:

$$v_t = \Delta^d y_t = \frac{\Theta(B)}{\Phi(B)} a_t + \delta_d^* + \delta_{d+1}^* t + \dots + \delta_{d+m-1}^* t^{m-1}. \quad (8.2)$$

Pierwszy składnik w powyższej zależności opisuje stacjonarny proces ARMA, zaś pozostałe składniki mają charakter trendu deterministycznego.

Współczynniki $\delta_d^* \dots \delta_{d+m-1}^*$ mogą być wyrażone poprzez współczynniki $\delta_d \dots \delta_{d+m-1}$.

Dalsze m -krotne różnicowanie pozwala na całkowite wyeliminowanie trendu, jednakże obecnie proces w_t będzie nieodwracalny:

$$w_t = \Delta^{d+m} y_t = \frac{\Theta^*(B)}{\Phi(B)} a_t = \frac{\Theta(B)\Delta^m}{\Phi(B)} a_t. \quad (8.3)$$

Nieodwracalność procesu w_t wynika z faktu, iż wielomian $\Theta^*(B)$ posiada m -krotny pierwiastek na okręgu jednostkowym.

Model (8.3) może być użyty do prognozowania szeregu czasowego za pomocą specjalnego predyktora skonstruowanego na podstawie filtru Kalmana. Znajomość współczynników $\delta_d \dots \delta_{d+m-1}$ nie jest wówczas konieczna.

9. Predykcja optymalna szeregu niestacjonarnego z trendem

Rozważmy problem predykcji za pomocą modelu ARIMA(p, s, q), (Harvey 1981a, Harvey, McKenzie 1982). W tym celu należy utworzyć reprezentację w przestrzeni stanu procesu stacjonarnego $w_t = \Delta^s z_t$, dla którego można znaleźć predykcję optymalną za pomocą filtru Kalmana.

Następnie tworzy się model rozszerzony poprzez dodanie do wektora stanu $s-1$ opóźnionych wartości z_{T-1}, \dots, z_{T-d} . Niech ten wektor opóźnionych wartości będzie oznaczony przez $\zeta_{t-1} = (z_{t-1}, \dots, z_{t-d})'$ dla wszystkich $t = d+1, \dots, T$. Estymatorem rozszerzonego wektora stanu $\xi_t = (x'_t, \zeta'_{t-1})'$ w chwili T jest wówczas $\hat{\xi}_T = (\hat{x}_T, \hat{\zeta}'_{T-1})'$, najlepszy estymator zapewniający minimum średniokwadratowego błędu predykcji. Ponieważ ζ_{T-1} jest znane, więc macierzą błędu średniokwadratowego jest $\sigma^2 P_T$, gdzie:

$$P_T = \begin{bmatrix} \Sigma_T & 0 \\ 0 & 0 \end{bmatrix}. \quad (9.1)$$

Niech $-\nu_j, j = 1, 2, \dots, s$ będzie współczynnikiem przy B^j w rozwinięciu $\Delta^s = (1-B)^s$ oraz przyjmijmy, że :

$$x(t+1) = Ax(t) + ga(t+1), \quad (9.2)$$

$$w(t) = d'x(t) \quad (9.3)$$

jest układem równań stanu dla procesu $w_t = \{w_t, t = 1, \dots, T\}$. Równanie stanu dla rozszerzonego wektora stanu $\xi_t = [x'_t, \zeta'_{t-1}]'$ będzie miało postać:

$$\begin{bmatrix} x_t \\ \zeta_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & A & 0 & \nu_1 & \dots & \nu_s \\ 0 & & I_{s-1} & & & 0 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ \zeta_{t-2} \end{bmatrix} + \begin{bmatrix} g \\ 0 \end{bmatrix} a_t, \quad (9.4)$$

dla $t = s+1, \dots$

Zastosowanie równań predyktora Kalmana do rozszerzonego równania stanu daje optymalny predyktor \hat{z}_{T+k} , dla $k = 1, 2, \dots$

10. Identyfikacja procesów z trendami wielomianowymi

Założmy, że dana jest realizacja $y(1) \dots y(T)$ procesu opisanego równaniem (8.1).

Zadanie identyfikacji polega na ocenie

- liczb całkowitych d i m charakteryzujących rząd różnicy koniecznej dla stacjonaryzacji składowej stochastycznej oraz stopień trendu deterministycznego,
- liczb całkowitych p i q określających stopnie wielomianów $\Phi(B)$ i $\Theta(B)$,
- współczynników $\phi_1 \dots \phi_p, \theta_1 \dots \theta_q$.

Procedura identyfikacji jest wieloetapowa i polega na selekcjonowaniu modeli na podstawie kryteriów informacyjnych. W pierwszym etapie dokonuje się oceny liczb d i m posługując się modelem autoregresyjnym składowej stochastycznej, określonym wielomianem $\Phi_n(B)$ stopnia n . W drugim etapie na podstawie modelu autoregresyjnego oraz zróżnicowanej realizacji z usuniętym trendem określa się strukturę modelu ARMA i jego parametry. W etapie trzecim, po określeniu struktury i wstępnej ocenie parametrów, dokonuje się ostatecznej łącznej oceny parametrów trendu i parametrów modelu ARMA za pomocą metody największej wiarygodności.

Dokładniejsze omówienie elementów składowych tej procedury znajduje się w punktach 9.1 i 6, zaś w punkcie 9.2 omawia się kompletny algorytm.

10.1. Identyfikacja trendu deterministycznego

Założmy, że dysponujemy realizacją $v(1) \dots v(T)$ procesu

$$v_t = \frac{\Theta(B)}{\Phi(B)} a_t + \delta_d^* + \delta_{d+1}^* t + \dots + \delta_{d+m-1}^* t^{m-1}. \quad (10.1)$$

Zadanie polega na ocenie liczby m oraz współczynników $\delta_d^* \dots \delta_{d+m-1}^*$.

Procedura identyfikacji polega na wykonywaniu dla wartości $m = 0 \dots m_{\max}$ następujących działań:

- Ocena współczynników $\delta_d^* \dots \delta_{d+m-1}^*$ za pomocą metody najmniejszych kwadratów.
- Usunięcie trendu i dopasowanie wielomianu autoregresyjnego $\Phi_n(B)$ stopnia n . Liczbę n określa się dopasowując kolejno wielomiany stopnia $n = 0 \dots n_{\max}$ i wybierając model dający minimalną wartość

$$AIC = \sigma_n^2 \exp \frac{2n}{T}. \quad (10.2)$$

- (c) Korekcja wartości współczynników $\delta_d^* \dots \delta_{d+m-1}^*$ przy danym $\Phi_n(B)$ metodą największej wiarygodności.
- (d) Korekcja współczynników wielomianu $\Phi_n(B)$ przy danych $\delta_d^* \dots \delta_{d+m-1}^*$.
- (e) Powtórzenie punktów c i d aż do uzyskania ekstremum funkcji wiarygodności.
- (f) Ustalenie oceny m z warunku min BIC

$$BIC = \sigma_{n,m}^2 \exp\left\{(n+m) \frac{\log T}{T}\right\}. \quad (10.3)$$

10.2. Identyfikacja modelu z trendem

Procedura identyfikacji jest wieloetapowa.

Etap 1

Oceny liczby różnicowań d oraz stopnia m składowej deterministycznej dokonuje się powtarzając procedurę z punktu 9.2 kolejno dla $d = 0, 1 \dots d_{\max}$. Jako oceny przyjmuje się ten zestaw wartości d i m , przy którym wartość BIC jest najmniejsza.

Etap 2

Na podstawie modelu autoregresyjnego oraz zróżnicowanej realizacji z usuniętym trendem określa się zgodnie z procedurą z punktu 7 strukturę modelu ARMA i jego parametry.

Etap 3

Po określeniu struktury i wstępnej ocenie parametrów dokonuje się ostatecznej łącznej oceny parametrów trendu i parametrów modelu ARMA za pomocą metody największej wiarygodności.

11. Podsumowanie

W artykule przedstawiono modele ARMA stacjonarnych szeregów czasowych, ich reprezentacje w przestrzeni stanu, równania filtru Kalmana, równania predyktora optymalnego oraz estymację parametrów modelu za pomocą metody największej wiarygodności. Dokonano szerokiego przeglądu metod identyfikacji rzędu modelu ARMA(p, q) oraz zaproponowano oryginalną modyfikację jednej z nich. Przedstawiono również modele ARIMA niestacjonarnych szeregów czasowych z wielomianowym trendem deterministycznym. Podano oryginalne algorytmy identyfikacji, estymacji i predykcji dla modeli niestacjonarnych. Ich implementację komputerową przedstawiono w pracy (Błachuta, 1996). Przedstawione tam przykłady wskazują na wysoką efektywność opracowanej metody.

12. Podziękowanie

Autor wyraża wdzięczność p. Profesorowi Mieczysławowi Brdysowi z The University of Birmingham za stworzenie znakomych warunków do pracy w School of Electronic & Electrical Engineering.

LITERATURA

1. Akaike H. (1970), Statistical predictor identification, *Ann. Inst. Statist. Math.*, vol. 21, pp. 243-247
2. Akaike H. (1973), Information theory and an extension of the maximum likelihood principle, in 2nd International Symposium on Information Theory, ed. B. N Petrov and F. Csaki, pp. 267-281, Akademiai Kiado
3. Akaike H. (1974), A new look at the statistical model identification, *IEEE Trans. Auto. Control*, vol. AC-19, pp. 716-723
4. Akaike H. (1977), On entropy maximization principle, in *Applications of Statistics*, ed. P. R. Krisnaiah, pp. 27-41, North-Holland
5. Akaike H. (1978), A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, vol. 30, pp. 9-14
6. Akaike H. (1979), A Bayesian extension of the minimum AIC procedure, *Biometrika*, vol. 66, pp. 237-242
7. Anderson T. W. (1963), Determination of the order of dependence in normally distributed time series, *Proc. of the Symposium on Time Series*, ed. M. Rosenblatt, pp. 425-446, Wiley
8. Anderson T. W. (1971), *The Statistical Analysis of Time Series*, Wiley
9. Beguin J. M., Gourieroux C. and A. Monfort (1980) Identification of a mixed autoregressive-moving average process: the corner method, in *Time Series*, ed. O. D. Anderson, North-Holland, pp. 423-436
10. Bhansali R.J. and D. Y. Downham (1977), Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion, *Biometrika*, vol. 64, pp. 547-551

11. Błachuta M. (1994), PIPS: wersja dydaktyczna programu identyfikacji i predykcji szeregów czasowych, Zeszyty Naukowe Politechniki Śląskiej, seria: Automatyka, z. 120, str. 39-59
12. Box G. E. P., G. M. Jenkins (1983), Analiza szeregów czasowych. Prognozowanie i sterowanie. PWN, Warszawa
13. Burshtein D. (1993), An efficient algorithm for calculating the likelihood and likelihood gradient of ARMA models, IEEE Trans. on Auto. Control, vol. AC-38, pp. 336-340
14. Chatfield C. (1979), Inverse autocorrelations, J. R. Statist. Soc. ser.A, vol. 142, pp. 363-367
15. Cleveland W. S. (1972), The inverse autocorrelations of time series and their applications, Technometrics, vol. 14, pp. 2777-298
16. Geweke J. F. and R. A. Meese (1981), Estimating regression models of finite but unknown orders, Int. Econ. Revue, vol. 22, pp. 55-70
17. Griffiths P. and I. D. Hill (1985), Applied Statistics Algorithms. The Royal Statistical Society/Ellis Horwood Limited
18. Hannan E. J. (1982), Testing for autocorrelation and Akaike's criterion, in Essays in Statistical Science, ed. J.Gani and E. J. Hannan, pp. 403-412, Applied Probability Trust, The University, Sheffield
19. Hannan E. J. (1988), The estimation of the order of an ARMA process, Ann. Statist, vol. 8, pp. 1071-1081
20. Hannan E. J. and L. Kavalieris (1983), Linear estimation of ARMA processes, Automatica, vol. 19, pp. 447-448
21. Hannan E. J. and B. G. Quinn (1979), The determination of the order of an autoregression, J. R. Statist. Soc., ser B, vol. 41, pp. 190-195
22. Hannan E. J. and J. Rissanen (1982), Recursive estimation of mixed autoregressive-moving average order, Biometrika, vol. 69, pp. 81-94
23. Harvey A. C. (1981a), Finite sample prediction and overdifferencing, Journal of Time Series Analysis, vol. 2. No. 4

24. Harvey A. C. (1981b), *The Econometric Analysis of Time Series*, Philip Allan
25. Harvey A. C. and C. R. McKenzie (1982), *Finite sample prediction from ARIMA processes*, Algorithm AS 182, *Applied Statistics*, vol. 31., No. 2
26. Kohn R. and C. F. Ansley (1982), *Computing the likelihood and their derivatives for a Gaussian ARMA model*, *J. Statist. Computn Simuln*, vol. 15, pp. 229-263
27. Mélard G. (1984), *A fast algorithm for the exact likelihood of autoregressive- moving average models*, Algorithm AS 197, *Applied Statistics*, vol. 33, No. 1
28. Otnes R. K., L. Enochson (1978), *Analiza numeryczna szeregów czasowych*. WNT, Warszawa
29. Pate B. and N. Davies (1988), *Computation of Population and Sample Correlation and Partial Correlation Matrices in MARMA(P,Q) Time Series*. *Applied Statistics*. vol. 37, No. 1, pp. 127-138
30. Rissanen J. (1978), *Modeling by the shortest data description*, *Automatica*, vol. 14, pp. 465-471
31. Shea B. L. (1987), *Estimation of Multivariate Time Series*. *Journal of Time Series Analysis*, vol. 8, No. 1
32. Shea B. L. (1989), *The Exact Likelihood of a Vector Autoregressive Moving Average Model*. *Applied Statistics*, vol. 38, No. 1, pp. 161-204
33. Schwarz G. (1978), *Estimating the dimension of a model*, *Ann. Statist*, vol. 6, pp. 461-464
34. Söderström T. (1977), *On model structure testing in systems identification*, *Int. J. Control*. vol. 26, pp. 1-18
35. Takemura A. (1984), *A generalization of autocorrelation and partial autocorrelation functions useful for identification of ARMA(p,q) process*, Technical Report no. 84-16, Department of Statistics, Purdue University
36. Tiao G. C. and R. S. Tsay (1983a), *Consistency properties of least squares estimates of autoregressive parameters in ARMA models*, *Ann. Statist*. vol. 11, pp. 856-871
37. Tiao G. C. and R. S. Tsay (1983b), *Multiple time series modelling and extended sample cross-correlations*, *J. Business Econ. Statist*. vol. 1, pp. 43-56

38. Tsay R. S. and G. C. Tiao (1984), Consistent estimates of autoregressive parameters and extended sample correlation function for stationary and nonstationary ARMA models, *J. Am. Statist. Assoc.* vol. 79, pp. 84-96
39. Whittle P. (1954), Some recent contributions to the theory of stationary time series, 2nd ed., by H. Wold, pp. 196-228, Almqvist and Wicksell

Recenzent: Prof. dr hab. inż. Józef Korbicz
WSI Zielona Góra

Wpłynęło do Redakcji dnia 10.06.1994

Abstract

In the paper, ARMA models of stationary time series are presented along with their state-space representations, Kalman filter equations, the optimal predictor, and the maximum-likelihood estimator of model parameters. The methods of identification of the ARMA(p, q) model order have been surveyed including those based on testing of statistical hypotheses, the realization theory and the information criteria. An original modification of one of them has been proposed. ARIMA models of nonstationary time series with a polynomial deterministic trend are also addressed, and original algorithms of their identification, parameter estimation and prediction are presented.