

Przegląd metod określania rzędu wektorowego modelu autoregresji

Jan Koziara

Instytut Zastosowań Matematyki, Zakład Matematyki i Statystyki
Matematycznej, Akademia Rolnicza w Lublinie

Streszczenie

Jednym z podstawowych problemów w analizie szeregów czasowych jest dopasowanie odpowiedniego modelu do danego szeregu czasowego (konkretnej jego realizacji). Jeśli dopasowujemy model autoregresji, to powstaje pytanie jakiego rzędu ma to być model. Istnieje szereg kryteriów pozwalających dobrać odpowiedni rząd modelu. Niniejszy artykuł omawia kryteria *FPE*, *AIC*, *HQ*, *SC*, kryterium oparte na metodzie Bayesa oraz test ilorazu wiarygodności w odniesieniu do wektorowych szeregów czasowych. Podano własności omawianych kryteriów oraz ich porównanie, zamieszczono również wyniki symulacji porównawczej. W oparciu o prezentowane rezultaty opracowany został program w języku *Turbo Pascal 7.0* realizujący dobór rzędu modelu autoregresji oraz estymację parametrów dla modeli skalarnych i wektorowych.

1. Symbolika i założenia

Rozpatrujemy wektorowy model autoregresji p -tego rzędu postaci

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (1.1)$$

gdzie:

$\mathbf{y}_t = [y_{1t}, y_{2t}, \dots, y_{Kt}]'$, $t = 1, 2, \dots, T$, jest ciągiem wektorów losowych o wymiarach $K \times 1$,

$\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ są macierzami parametrów ($K \times K$),

\mathbf{v} jest wektorem parametrów ($K \times 1$),

\mathbf{u}_t jest ciągiem niezależnych wektorów losowych ($K \times 1$) takich, że $E[\mathbf{u}_t] = \mathbf{0}$,

$E[\mathbf{u}_t \mathbf{u}_{t'}'] = \mathbf{0}$, dla $t \neq t'$, $\mathbf{C}_u = E[\mathbf{u}_t \mathbf{u}_t']$, dla $t = 1, 2, \dots, T$.

Słowa kluczowe: model autoregresji, rząd modelu, kryteria *FPE*, *AIC*, *HQ*, *SC*, metoda Bayesa'a, test ilorazu wiarygodności

Zakładając, że rząd modelu p jest znany, przeprowadza się estymację nieznanych parametrów \mathbf{v} , $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ oraz macierzy \mathbf{C}_u . Sprawą pierwszoplanową staje się więc określenie rzędu modelu p . Zanim zajmiemy się bliżej problemem wyznaczania rzędu modelu, wprowadzimy następujące oznaczenia:

$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ macierz o wymiarach $K \times T$,

$\mathbf{Y}^o = (\mathbf{y}_1 - \boldsymbol{\mu}, \mathbf{y}_2 - \boldsymbol{\mu}, \dots, \mathbf{y}_T - \boldsymbol{\mu})$, wektor o wymiarach $K \times 1$,

$\boldsymbol{\mu} = (\mathbf{I}_K - \mathbf{A}_1 - \mathbf{A}_2 - \dots - \mathbf{A}_p)^{-1} \mathbf{v}$, wektor o wymiarach $K \times 1$,

$\mathbf{Y}_t^o = [(\mathbf{y}_t - \boldsymbol{\mu})', (\mathbf{y}_{t-1} - \boldsymbol{\mu})', \dots, (\mathbf{y}_{t-p+1} - \boldsymbol{\mu})']'$, wektor o wymiarach $Kp \times 1$,

$\mathbf{X} = (\mathbf{Y}_0^o, \mathbf{Y}_1^o, \dots, \mathbf{Y}_{T-1}^o)$, macierz o wymiarach $Kp \times T$,

$\mathbf{B} = (\mathbf{v}, \mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p)$ macierz o wymiarach $K \times (Kp + 1)$,

$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p)$, macierz o wymiarach $K \times Kp$,

$\mathbf{Z}_t = (1, \mathbf{y}_t', \mathbf{y}_{t-1}', \dots, \mathbf{y}_{t-p+1}')'$ wektor o wymiarach $(Kp + 1) \times 1$,

$\mathbf{Z} = (\mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_{T-1})$ macierz o wymiarach $(Kp + 1) \times T$,

$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T)$ macierz o wymiarach $K \times T$,

$\mathbf{u} = \text{vec}(\mathbf{U})$ wektor o wymiarach $KT \times 1$,

$\mathbf{y} = \text{vec}(\mathbf{Y})$ wektor o wymiarach $KT \times 1$,

$\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ wektor o wymiarach $(K^2p + K) \times 1$,

$\mathbf{b} = \text{vec}(\mathbf{B}')$ wektor o wymiarach $(K^2p + K) \times 1$,

gdzie $\text{vec}(\cdot)$ jest operatorem przekształcającym macierz do postaci wektora kolumnowego, tzn. $\text{vec}(\mathbf{A}) = \text{vec}[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = [\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_n]'$ gdzie \mathbf{a}_i ($i = 1, 2, \dots, n$) oznaczają kolumny macierzy \mathbf{A} .

2. Podstawowe definicje i twierdzenia

Definicja 2.1. Proces \mathbf{y}_t nazywać będziemy procesem autoregresji p -tego rzędu, jeśli $\mathbf{A}_p \neq \mathbf{0}$ i $\mathbf{A}_i = \mathbf{0}$, dla wszystkich $i > p$.

Mówiąc o dopasowaniu modelu należy wprowadzić przede wszystkim pewną miarę dopasowania. Wprowadzimy więc następujące określenie.

Definicja 2.2. Prognozę $\mathbf{y}(t+h)$ wartości \mathbf{y}_{T+h} procesu opartą na wartościach $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$ nazywać będziemy prognozą średniokwadratową z horyzontem prognozowania h , jeśli

$$\mathbf{y}(t+h) = E[\mathbf{y}_{t+h} | \mathbf{y}_s, s \leq t] .$$

Tak zdefiniowana prognoza minimalizuje macierz kowariancji błędu prognozy w tym sensie, że jeśli $\tilde{\mathbf{y}}(t+h)$ jest dowolną inną prognozą, wówczas

$$\begin{aligned} E[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}(t+h)] [\mathbf{y}_{t+h} - \tilde{\mathbf{y}}(t+h)]' - E[\mathbf{y}_{t+h} - \mathbf{y}(t+h)] [\mathbf{y}_{t+h} - \mathbf{y}(t+h)]' = \\ = P_{\tilde{\mathbf{y}}}(h) - P_{\mathbf{y}}(h) \geq 0 . \end{aligned}$$

$P_{\hat{y}}(h)$ oraz $P_y(h)$ oznaczają odpowiednie macierze kowariancji błędów prognozy, zaś zapis ≥ 0 oznacza, że macierz stojąca po lewej stronie nierówności jest macierzą nieujemnie określoną. Można pokazać, że macierz kowariancji błędu prognozy dla horyzontu prognozowania $h=1$ ma postać (por. Lütkepohl (1991), wzór (3.5.13), str. 88)

$$P_{\hat{y}}(1) = \frac{T + Kp + 1}{T} C_u \quad (2.1)$$

Ze wzoru tego wynika, że $P_{\hat{y}}(1)$ jest funkcją rosnącą rzędu modelu p ; oznacza to, że prognoza wyznaczana na podstawie modelu z zawyżonym rzędem będzie gorsza od prognozy opartej na modelu z właściwie dobranym rzędem.

Jedną z możliwości określenia rzędu modelu autoregresji jest zastosowanie sekwencyjnego testu istotności. Jeśli przyjmiemy, że rząd modelu nie przekracza pewnej liczby naturalnej M wówczas testowanie rozpoczynamy od weryfikacji hipotezy $H_0^{(1)}: \mathbf{A}_M = \mathbf{0}$, następnie w sytuacji gdy hipoteza ta nie może być odrzucona, testujemy hipotezę $H_0^{(2)}: \mathbf{A}_{M-1} = \mathbf{0}$ itd. aż do momentu, gdy będziemy mogli odrzucić hipotezę zerową. Do tego celu możemy wykorzystać test ilorazu wiarygodności. Jak wiadomo idea tego testu polega na porównaniu maksimum funkcji wiarygodności na całej przestrzeni parametrów i pewnej podprzestrzeni wyznaczonej przez określone restrykcje nałożone na parametry. Odpowiednia statystyka testowa ma postać:

$$\lambda_{LR} = 2 \ln \frac{L(\tilde{\theta})}{L(\tilde{\theta}_r)} = 2[\ln L(\tilde{\theta}) - \ln L(\tilde{\theta}_r)] \quad (2.2)$$

gdzie $\tilde{\theta}$ jest estymatorem największej wiarygodności wektora parametrów $\mathbf{0}$ otrzymanym przez maksymalizację funkcji wiarygodności po całej przestrzeni parametrów, zaś $\tilde{\theta}_r$ jest estymatorem "restrykcyjnym" otrzymanym przez maksymalizację funkcji wiarygodności po pewnej podprzestrzeni wyznaczonej restrykcjami nałożonymi na wektor parametrów.

Wiadomo, że statystyka λ_{LR} ma asymptotyczny rozkład χ^2 (por. Lütkepohl (1991), str 123) z tyloma stopniami swobody, ile jest różnych restrykcji liniowych nałożonych na wektor parametrów.

Jeśli zainteresowani jesteśmy dopasowaniem dokładnego rzędu modelu AR to powinniśmy wybierać estymator rzędu mający określone własności statystyczne np. estymator asymptotycznie zgodny.

Definicja 2.3. Estymator \hat{p} rzędu modelu autoregresji nazywać będziemy zgodnym jeśli

$$\text{plim}_{T \rightarrow \infty} \hat{p} = p \quad \text{tzn.} \quad \lim_{T \rightarrow \infty} P(\hat{p} = p) = 1.$$

Definicja 2.4. Estymator \hat{p} rzędu modelu autoregresji nazywać będziemy mocno zgodnym jeśli

$$P [\lim_{T \rightarrow \infty} \hat{p} = p] = 1.$$

Definicja 2.5. Kryterium wyboru rzędu modelu autoregresji nazywać będziemy zgodnym lub mocno zgodnym jeśli estymator ma odpowiednią własność.

Przy badaniu własności kryteriów ważną rolę odgrywa następujące twierdzenie Hannan'a i Quinn'a (1979) podające warunek konieczny i dostateczny zgodności i mocnej zgodności kryterium. Dowód podaje Quinn (1980) oraz Paulsen (1984).

Twierdzenie 2.1.

Niech

1. \mathbf{y}_t będzie K -wymiarowym stacjonarnym procesem autoregresji z białym szumem \mathbf{u}_t , o skończonych czwartych momentach.
2. Rząd procesu autoregresji nie przekracza pewnej stałej M , tzn. $p \leq M$.
3. Estymator rzędu \hat{p} został wybrany przez minimalizację

$$Cr(m) = \ln |\tilde{\mathbf{C}}_u(m)| + \frac{mc_T}{T} \quad (2.3)$$

względem $m = 0, 1, \dots, M$, zaś $\tilde{\mathbf{C}}_u(m)$ jest pseudo-estymatorem największej wiarogodności macierzy $\mathbf{C}_u = \mathbf{E}[\mathbf{u}_t \mathbf{u}_t']$, c_T jest niemalejącym ciągiem liczb rzeczywistych zależnym od wielkości próby T .

Przy powyższych założeniach \hat{p} jest estymatorem zgodnym wtedy i tylko wtedy, gdy

$$c_T \rightarrow \infty \quad \text{i} \quad \frac{c_T}{T} \rightarrow 0, \quad \text{przy} \quad T \rightarrow \infty, \quad (2.4)$$

i mocno zgodnym gdy oprócz warunku (2.4), ma miejsce

$$\frac{c_T}{2} \ln \ln T > 1 \quad (\text{być może przy } T \rightarrow \infty). \quad (2.5)$$

3. Test ilorazu wiarogodności

Przy oznaczeniach wprowadzonych w poprzednim rozdziale model (1.1) może być zapisany w jednej z równoważnych postaci:

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{U} \quad (3.1)$$

lub

$$\mathbf{y} = (\mathbf{Z}' \otimes \mathbf{I}_K) \boldsymbol{\beta} + \mathbf{u}$$

Macierz kowariancji wektora $\mathbf{u} = \text{vec}(\mathbf{U})$ ma wtedy postać $\boldsymbol{\Gamma}_u = \mathbf{I}_T \otimes \mathbf{C}_u$, zaś $\det \boldsymbol{\Gamma}_u = T \det \mathbf{C}_u$. Przy założeniu, że $\mathbf{u}_t \sim N(0, \mathbf{C}_u)$, logarymiczna funkcja wiarygodności będzie miała postać:

$$\ln L(\boldsymbol{\beta}, \mathbf{C}_u) = -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\mathbf{C}_u| - \frac{1}{2} [\mathbf{y} - (\mathbf{Z}' \otimes \mathbf{I}_K) \boldsymbol{\beta}]' (\mathbf{I}_T \otimes \mathbf{C}_u)^{-1} [\mathbf{y} - (\mathbf{Z}' \otimes \mathbf{I}_K) \boldsymbol{\beta}].$$

Estymatory największej wiarygodności parametrów $\boldsymbol{\beta}$ i \mathbf{C}_u uzyskane przez maksymalizację funkcji wiarygodności po całej przestrzeni parametrów są następujące:

$$\hat{\boldsymbol{\beta}} = [(\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z} \otimes \mathbf{I}_K] \mathbf{y},$$

$$\hat{\mathbf{C}}_u = \frac{1}{T} (\mathbf{Y} - \hat{\mathbf{B}}\mathbf{Z})(\mathbf{Y} - \hat{\mathbf{B}}\mathbf{Z})'.$$

Jeśli na wektor parametrów $\boldsymbol{\beta}$ nałożymy restrykcje

$$\mathbf{W}\boldsymbol{\beta} = \mathbf{w}, \quad (3.2)$$

gdzie \mathbf{W} jest znaną macierzą o wymiarach $N \times (K^2p + K)$ rzędu N , zaś \mathbf{w} znanym wektorem o wymiarach $N \times 1$, to estymatory największej wiarygodności dla $\boldsymbol{\beta}$ i \mathbf{C}_u przy warunku (3.2) mogą być znalezione metodą mnożników Lagrange'a. Funkcja Lagrange'a może być zapisana jako

$$G(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{C}_u) = \ln L(\boldsymbol{\beta}, \mathbf{C}_u) + \boldsymbol{\gamma}' (\mathbf{W}\boldsymbol{\beta} - \mathbf{w}),$$

gdzie $\boldsymbol{\gamma}$ ($N \times 1$) jest wektorem mnożników Lagrange'a.

Obliczając pochodne cząstkowe $\partial G / \partial \boldsymbol{\beta}$, $\partial G / \partial \boldsymbol{\gamma}$ oraz przyrównując je do zera uzyskujemy estymatory parametrów $\boldsymbol{\beta}$ i \mathbf{C}_u przy warunku (3.2):

$$\begin{aligned} \hat{\boldsymbol{\beta}}_r &= \hat{\boldsymbol{\beta}} + [(\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\mathbf{C}}_u] \mathbf{W}' [\mathbf{W}(\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\mathbf{C}}_u \mathbf{W}]^{-1} (\mathbf{w} - \mathbf{W}\hat{\boldsymbol{\beta}}) \\ \hat{\mathbf{C}}_u^r &= \frac{1}{T} (\mathbf{Y} - \hat{\mathbf{B}}_r \mathbf{Z})(\mathbf{Y} - \hat{\mathbf{B}}_r \mathbf{Z})' \end{aligned}, \quad (3.3)$$

gdzie $\hat{\mathbf{B}}_r$ jest macierzą współczynników modelu odpowiadającą wektorowi $\hat{\boldsymbol{\beta}}_r$, tzn. $\hat{\boldsymbol{\beta}}_r = \text{vec}(\hat{\mathbf{B}}_r)$. Statystyka (2.2) λ_{LR} przyjmuje więc postać:

$$\lambda_{LR} = 2[\ln L(\hat{\boldsymbol{\beta}}, \hat{\mathbf{C}}_u) - \ln L(\hat{\boldsymbol{\beta}}_r, \hat{\mathbf{C}}_u^r)] = T[\ln |\hat{\mathbf{C}}_u^r| - \ln |\hat{\mathbf{C}}_u|].$$

Jak już było wspomniane wcześniej, statystyka ta ma asymptotyczny rozkład $\chi^2(N)$. Rezultat ten pozostaje prawdziwy również w przypadku, gdy szereg czasowy \mathbf{y}_t nie jest gausowski lecz należy do szerszej rodziny rozkładów. Estymatory

największej wiarygodności są w takim przypadku nazywane pseudo-estymatorami największej wiarygodności.

Zastanówmy się jakie jest prawdopodobieństwo błędu pierwszego rodzaju w takiej procedurze testowania. Niech D_j oznacza zdarzenie polegające na odrzuceniu hipotezy H_0^j w j -tym teście podczas gdy hipoteza ta jest prawdziwa. Przypuśćmy, że prawdziwa jest hipoteza $H_0^i: \mathbf{A}_{M-i+1} = \mathbf{0}$, tzn., że rząd modelu $p \leq M - i$. Prawdopodobieństwo błędu pierwszego rodzaju dla i -tej hipotezy w tej procedurze testowania może być zapisane jako

$$\varepsilon_i = P(D_1 \cup D_2 \cup \dots \cup D_i) ,$$

przy czym $P(D_i) = \alpha_i$ jest poziomem istotności i -tego indywidualnego testu. Paulsen i Tjøstheim (1985) pokazali, że dla $m \neq j$ oraz $j, m \leq i$, wartości statystyki λ_{LR} dla m -tej i j -tej hipotezy ($\lambda_{LR}(m)$ i $\lambda_{LR}(j)$) są asymptotycznie niezależne, jeśli hipotezy H_0^1, \dots, H_0^i są prawdziwe. Wobec tego przy dużych próbach zdarzenia D_m i D_j są niezależne, tak więc

$$\varepsilon_i = P(D_1 \cup D_2 \cup \dots \cup D_{i-1}) + P(D_i) - P[(D_1 \cup D_2 \cup \dots \cup D_{i-1}) \cap D_i] ,$$

dla $i = 1, 2, \dots, M$. Stosując zasadę indukcji matematycznej można pokazać, że

$$\varepsilon_i = 1 - (1 - \alpha_1)(1 - \alpha_2) \dots (1 - \alpha_i)$$

dla $i = 1, 2, \dots, M$.

4. Kryteria FPE i AIC

Niekiedy model autoregresji jest konstruowany do celów prognozowania i wobec tego dobór odpowiedniego rzędu modelu powinien opierać się na kryterium uzyskania jak najlepszej prognozy. Nie jest wówczas rzeczą najważniejszą uzyskanie właściwego rzędu modelu, a raczej otrzymanie modelu użytecznego dla celów prognozowania. Sensownym jest więc dobranie takiego rzędu modelu, przy którym błąd prognozy jest minimalny. Wychodząc z takich przesłanek Akaike (1969, 1971) zaproponował, aby oprzeć wybór rzędu modelu AR na ocenie macierzy kowariancji błędu prognozy średniokwadratowej. Ogólna postać takiej macierzy jest dość skomplikowana (por. Koziara, 1986), natomiast dla horyzontu prognozowania $h = 1$ (prognoza 1-krokowa) ma prostą strukturę (2.1). Praktyczne wykorzystanie wzoru (2.1) polega na tym, że macierz kowariancji \mathbf{C}_u białego szumu \mathbf{u}_t jest zastąpiona estymatorem z odpowiednio dobranymi stopniami swobody. Akaike zasugerował wykorzystanie następującej oceny:

$$\hat{\mathbf{C}}_u(m) = \frac{TK}{TK - K^2m - K} \tilde{\mathbf{C}}_u(m) = \frac{T}{T - Km - 1} \tilde{\mathbf{C}}_u(m) , \quad (4.1)$$

gdzie T jest liczebnością próbki, m rzędem modelu, $\tilde{C}_u(m)$ estymatorem największej wiarygodności dla C_u otrzymanym przez dopasowanie do danych z próby modelu AR(m). Ponadto, aby kryterium miało postać skalarną, brany jest wyznacznik. Zaproponowane przez Akaike kryterium jest nazywane kryterium *FPE* (*final prediction error*) i ma postać:

$$FPE(m) = \det \left[\frac{T + Km + 1}{T} \frac{T}{T - Km - 1} \tilde{C}_u(m) \right] = \left[\frac{T + Km + 1}{T - Km - 1} \right]^K \det \tilde{C}_u(m). \quad (4.2)$$

Należy zauważyć, że $\det \tilde{C}_u(m)$ jest nierosnącą funkcją rzędu modelu m . Wynika to stąd, że maksimum logarytmicznej funkcji wiarygodności zawiera składnik $-\ln |\tilde{C}_u(m)|$ i dla $m < n$ model AR(m) może być interpretowany jako model AR(n) z restrykcyjami. Wobec tego

$$-\ln |\tilde{C}_u(m)| \leq -\ln |\tilde{C}_u(n)| \Leftrightarrow \det[\tilde{C}_u(m)] \geq \det[\tilde{C}_u(n)]$$

(maksimum lokalne nie może przekraczać globalnego). Z drugiej strony czynnik $[(T + Km + 1)/(T - Km - 1)]^K$ jest funkcją rosnącą m . Rząd modelu \hat{p}_{FPE} dobierany jest jako wartość dla której te dwa "przeciwstawne" czynniki są zrównoważone tzn.

$$FPE[\hat{p}_{FPE}] = \min\{ FPE(m) \mid m = 0, 1, \dots, M \}.$$

W praktyce oznacza to, że należy znaleźć estymatory parametrów modeli rzędów $m = 0, 1, \dots, M$ a następnie obliczyć $FPE(m)$, dla $m = 0, 1, \dots, M$. Wartość m minimalizująca $FPE(m)$ jest poszukiwanym rzędem modelu oznaczonym przez \hat{p}_{FPE} .

Akaike (1973, 1974) jest ponadto autorem podobnego kryterium oznaczonego skrótowo jako *AIC* (*Akaike's Information Criterion*). Kryterium to ma postać:

$$AIC(m) = \ln |\tilde{C}_u(m)| + \frac{2}{T} m K^2. \quad (4.3)$$

Zgodnie z nim, za rząd modelu \hat{p}_{AIC} przyjmuje się taką wartość m , dla której $AIC(m)$ osiąga minimum. Pomiedzy tymi dwoma kryteriami zachodzi następująca zależność:

$$\ln FPE(m) = AIC(m) + \frac{2K}{T} + O(T^{-2}) \quad (4.4)$$

Wynika to stąd, że dla stałej N

$$\frac{T+N}{T-N} = 1 + \frac{2N}{T} \left[\frac{1}{1-N/T} \right] = 1 + \frac{2N}{T} + O(T^{-2}).$$

Wobec tego

$$\begin{aligned} \ln FPE(m) &= \ln |\tilde{C}_u(m)| + K \ln \frac{T+Km+1}{T-Km-1} = \ln |\tilde{C}(m)| + \\ &+ K \ln \left[1 + \frac{2(Km+1)}{T} + O(T^{-2}) \right] = \ln |\tilde{C}_u(m)| + K \frac{2(Km+1)}{T} + O(T^{-2}) = \\ &= \ln |\tilde{C}_u(m)| + \frac{2mK^2}{T} + \frac{2K}{T} + O(T^{-2}) = AIC(m) + \frac{2K}{T} + O(T^{-2}). \end{aligned}$$

Ponieważ składnik $2K/T$ nie zależy od m , wobec tego zarówno $AIC(m)$ jak i $AIC(m) + 2K/T$ osiągają minimum dla tej samej wartości m . Widać stąd, że obydwa kryteria są sobie równoważne asymptotycznie (dla dużych wartości T).

Wniosek 4.1. Jeśli spełnione są założenia twierdzenia 2.1, wówczas \hat{p}_{FPE} i \hat{p}_{AIC} nie są estymatorami zgodnymi.

Ponieważ kryteria FPE i AIC są asymptotycznie równoważne, wystarczy pokazać, że \hat{p}_{AIC} nie jest estymatorem zgodnym. Porównując kryteria $AIC(m)$ i $Cr(m)$ dostajemy $mc_T/T = 2mK^2/T$ tzn. $c_T = 2K^2$. Jak łatwo zauważyć ciąg ten nie spełnia warunków (2.4). Prawdopodobieństwo graniczne zaniżenia rzędu modelu w kryteriach FPE i AIC jest równe zero na co wskazują zależności (7.4) z paragrafu 7. Skoro więc kryteria te nie są zgodne oznacza to, że zawyżają rząd modelu z dodatnim prawdopodobieństwem. Paulsen i Tjøstheim (1985) pokazali, że prawdopodobieństwo graniczne zawyżenia rzędu modelu zmniejsza się ze wzrostem wymiaru K i staje się nieistotne dla $K \geq 5$. Tak więc kryteria te wyznaczają właściwy rząd z prawdopodobieństwem 1, jeśli szereg czasowy ma wysoki wymiar K .

5. Kryteria HQ i SC

Obecnie rozpatrzmy dwa kryteria zgodne. Pierwsze z nich HQ (*Hannan-Quinn criterion*) (por. Hannan i Quinn, 1979; Quinn, 1980) dobiera rząd modelu \hat{p}_{HQ} przez minimalizację

$$HQ(m) = \ln |\tilde{C}_u(m)| + \frac{2 \ln \ln T}{T} mK^2 \quad (5.1)$$

względem $m = 0, 1, \dots, M$.

Porównanie z (2.3) daje $c_T = 2K^2 \ln \ln T$ i wobec tego warunek (2.4) twierdzenia 2.1 jest spełniony, ponadto dla $K > 1$ spełniony jest również warunek (2.5). Kry-

terium HQ jest więc zgodne dla procesów skalarnych i mocno zgodne dla wektorowych (przy warunkach twierdzenia 2.1).

Drugim kryterium zgodnym jest kryterium SC (por. Schwartz, 1978). Podobnie jak w poprzednich kryteriach rząd modelu wyznaczany jest przez minimalizację odpowiedniej funkcji, w tym przypadku postaci

$$SC(m) = \ln |\tilde{C}_u(m)| + \frac{\ln T}{T} mK^2 \quad (5.2)$$

względem $m = 0, 1, \dots, M$.

Łatwo sprawdzić, że obydwa warunki (2.4) i (2.5) twierdzenia 2.1 są spełnione i wobec tego kryterium to jest mocno zgodnym dla dowolnego wymiaru K .

6. Kryterium oparte na metodzie Bayesa

Krzyśko i Smoczyński (1984) zaproponowali metodę doboru rzędu modelu autoregresji opartą na metodzie Bayesa. Dysponując jedynie pojedynczą realizacją szeregu czasowego, wyboru rzędu modelu możemy dokonać według następującej procedury.

Zakładając, że rząd modelu p jest zmienną losową o znanej funkcji gęstości *a priori* $h(p)$, oraz biorąc funkcję straty postaci

$$S(p, r) = \begin{cases} p - r & \text{dla } r \leq p \\ c(r - p) & \text{dla } r > p \end{cases} \quad (6.1)$$

($c > 1$ jest pewną stałą przyjętą arbitralnie, zaś $r \in \{1, 2, \dots, M\}$), doboru rzędu modelu dokonuje się przez minimalizację ryzyka *a posteriori*

$$R(r) = \sum_{p=1}^M S(p, r) g(p | \mathbf{Y}) \quad (6.2)$$

po $r \in \{1, 2, \dots, M\}$. Gęstość *a posteriori* $g(p | \mathbf{Y})$ zmiennej losowej p ma postać

$$g(p | \mathbf{Y}) \propto h(p) (2\pi)^{-KT/2} [\det \mathbf{C}_u]^{-T/2} \exp \left[-\frac{1}{2} \text{tr} (\mathbf{Y}^\circ - \mathbf{A}\mathbf{X})' \mathbf{C}_u^{-1} (\mathbf{Y}^\circ - \mathbf{A}\mathbf{X}) \right] .$$

Tak więc

$$\hat{p} = [r : r \in \{1, 2, \dots, M\} \wedge R(r) = \min] .$$

Aby efektywnie wykorzystać tą metodę należałoby znać gęstość *a priori* $h(p)$ oraz stałą c . Autorzy metody sugerują by w przypadku braku informacji o $h(p)$ przyjąć $h(p) = 1/M$. Nie podają zaś żadnych sugestii co do wyboru stałej c . W Tabeli 1 przedstawione są rezultaty doboru rzędu modelu z wykorzystaniem tej metody

przy różnych wartościach stałej c . Liczby w kolumnach od (a) do (g) podają w ilu przypadkach na 100 możliwych wybrany został rząd modelu $0,1,\dots,5$, przy różnych długościach T realizacji szeregu czasowego oraz wartościach stałej c równych odpowiednio: $c = 2$, $c = [(M + 1)K + 1]KT$, $c = 2MK^2T$, $c = 2(MK + 1)KT$, $c = 2[(M + 1)K + 1]KT$, $c = MK^2T^2$, $c = (MK + 1)KT^2$. Łatwo zauważyć, że decydujący wpływ na efektywność tego kryterium ma wartość stałej c . Wartość c zależy od długości T realizacji szeregu czasowego, wymiaru K wektora obserwacji, oraz górnego ograniczenia M rzędu modelu. Najkorzystniejsze rezultaty zostały osiągnięte przy $c = MK^2T^2$, (kolumna (f) Tabeli 1). Przy wartości $c = 2$ (kolumna (a) Tabeli 1) widać wyraźnie tendencję do zawyżania rzędu modelu.

7. Porównanie kryteriów

Własności estymatorów rzędu modelu, o których była mowa dotychczas, dotyczą dużych prób. W małych próbach kryteria AIC i FPE mogą mieć lepsze własności (wybierają właściwy rząd częściej) niż kryteria HQ i SC . Trzeba też pamiętać, że kryteria AIC oraz FPE są skonstruowane z myślą o minimalizacji błędu prognozy i wobec tego (szczególnie przy małych próbach) modele oparte na tych kryteriach mogą dawać lepsze prognozy mimo, że kryteria nie dobierają poprawnie rzędu modelu. Shibata (1980) otrzymał pewne asymptotycznie optymalne własności AIC oraz FPE . Następujący lemat i twierdzenie (por. Lütkepohl, 1991, str.133) pozwalają ustalić relacje pomiędzy niektórymi kryteriami.

Lemat 7.1. Jeśli ciągi skończone $a_i, b_i, c_i, i = 0,1,\dots,M$ spełniają następujące warunki:

1. a_i, b_i są ciągami rosnącymi,
2. c_i jest ciągiem nierosnącym,
3. $b_{i+1} - b_i < a_{i+1} - a_i$ dla $i = 0,1,\dots,M - 1$,
4. Liczby całkowite n i k są tak dobrane, że $n, k \in \langle 0, M \rangle$
i $c_n + a_n = \min[c_i + a_i \mid i = 0,1,\dots,M]$, $c_k + b_k = \min[c_i + b_i \mid i = 0,1,\dots,M]$,
wówczas $k \geq n$.

Twierdzenie 7.1. Jeśli $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ jest realizacją szeregu czasowego do której dopasowano modele utoregresji rzędu $m = 0,1,2,\dots,M$, wówczas zachodzą następujące relacje

$$\hat{p}_{SC} \leq \hat{p}_{AIC}, \text{ jeśli } T \geq 8, \quad (7.1)$$

$$\hat{p}_{SC} \leq \hat{p}_{HQ}, \text{ dla wszystkich } T, \quad (7.2)$$

$$\hat{p}_{HQ} \leq \hat{p}_{AIC}, \text{ jeśli } T \geq 16. \quad (7.3)$$

Dowód: Niech $c_m = \ln |\tilde{C}_u(m)|$, $b_m = 2mK^2/T$, $a_m = (mK^2 \ln T)/T$. Aby warunki lematu 7.1 były spełnione musi zachodzić zależność $b_{m+1} - b_m < a_{m+1} - a_m$. Ponieważ $b_{m+1} - b_m = 2K^2/T$, $a_{m+1} - a_m = (K^2 \ln T)/T$, wobec tego powinna zachodzić zależność $2K^2/T < (K^2 \ln T)/T$, czyli $T > e^2 \approx 7.39$. Tak więc warunki lematu są spełnione dla $T \geq 8$. W ten sposób nierówność (7.1) została udowodniona. W analogiczny sposób można udowodnić nierówności (7.2) i (7.3). Należy zaznaczyć, że twierdzenie to pozostaje prawdziwe również w przypadku, gdy szereg czasowy y_t nie jest stacjonarny lub nie jest generowany przez proces autoregresji. \square

Jako wniosek z tego twierdzenia dostajemy następujące zależności słuszne przy założeniach twierdzenia (2.1),

$$\lim_{T \rightarrow \infty} P[\hat{p}_{AIC} < p] = 0 \quad \text{i} \quad \lim_{T \rightarrow \infty} P[\hat{p}_{AIC} > p] > 0. \quad (7.4)$$

(podobne zależności zachodzą dla FPE). Uzasadnienie podanych związków wynika z faktu, że $\hat{p}_{SC} \leq \hat{p}_{AIC}$ dla $T \geq 8$ i wobec tego $P[\hat{p}_{AIC} < p] \leq P[\hat{p}_{SC} < p]$. Ze zgodności zaś SC wynika, że $P[\hat{p}_{SC} < p] \rightarrow 0$, tak więc $P[\hat{p}_{AIC} < p] = 0$. Ponieważ AIC nie jest kryterium zgodnym więc $P[\hat{p}_{AIC} = p] < 1$ i w rezultacie $P[\hat{p}_{AIC} > p] > 0$.

Należy podkreślić, że aby estymatory rzędu miały określone własności asymptotyczne niekiedy należy wyznaczać je na podstawie bardzo licznej próby. Sytuację taką ilustruje następujący przykład. Przyjmując we wzorze (2.3) twierdzenia 2.1 $c_T = 2 \ln \ln T$ otrzymamy kryterium postaci:

$$K_T(m) = \ln |\tilde{C}_u(m)| + \frac{2m \ln \ln T}{T}.$$

Korzystając z twierdzenia 2.1 łatwo się przekonać, że jest to kryterium zgodne, a ponadto na podstawie lematu 7.1 $\hat{p}_{AIC} \leq \hat{p}_{K_T}$ dla $T \leq \exp[\exp(K^2)]$. Na przykład dla $K = 2$ dostajemy $T \leq 5.14 \times 10^{23}$. Oznacza to, że dla takich wartości T kryterium zgodne K_T wybiera rząd wyższy lub równy \hat{p}_{AIC} , a wiadomo, że AIC ma dodatnie prawdopodobieństwo zawyżania rzędu modelu.

8. Wyniki obliczeń symulacyjnych

W celu porównania omawianych tu metod wyznaczania rzędu modelu autoregresji, zastosowano metodę symulacji. Przy użyciu komputera wygenerowano po 100 realizacji dwuwymiarowego szeregu czasowego o 20, 30, 50, 100 oraz 200 obserwacjach. Wszystkie realizacje zostały wygenerowane w oparciu o model drugiego rzędu

$$\mathbf{y}_t = \begin{bmatrix} 1.2 \\ -0.9 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.1 \\ 0 & 0.3 \end{bmatrix} \mathbf{y}_{t-1} + \begin{bmatrix} -0.1 & 0.2 \\ 0.4 & 0.1 \end{bmatrix} \mathbf{y}_{t-2} + \mathbf{u}_t \quad (8.1)$$

z macierzą kowariancji białego szumu

$$\mathbf{C}_u = \begin{bmatrix} 0.97 & 0.46 \\ 0.46 & 0.53 \end{bmatrix}.$$

Dla każdej z wygenerowanych realizacji zastosowano omawiane kryteria określania rzędu modelu. Jako górne ograniczenie rzędu przyjęto $M = 5$. W Tabeli 1 zamieszczone zostały wyniki symulacji. Liczby w kolumnach podają w ilu przypadkach na 100 możliwych wybrany został odpowiedni rząd modelu, przy różnych długościach T realizacji szeregu czasowego. Z danych tych wynika, że trudno jest wskazać kryterium zdecydowanie najlepsze, poprawność doboru rzędu modelu zależy w głównej mierze od długości realizacji szeregu. Przy małych próbach częstość wyboru właściwego rzędu modelu metodą Bayesa jest wyższa niż przy innych kryteriach. W przypadku kryteriów FPE i AIC daje się zauważyć tendencję do zawyżania rzędu modelu. W oparciu o omawiane metody został opracowany program w języku *Turbo Pascal 7.0* realizujący dobór rzędu i estymację parametrów dla modeli skalarnych i wektorowych. Program jest dostępny w Instytucie Zastosowań Matematyki Akademii Rolniczej w Lublinie.

Tabela 1

Zestawienie wyników symulacji - liczba realizacji procesu AR rzędu 2, dla których ocena rzędu wynosiła 0,1,...,5, przy różnych długościach szeregu T

T	Ocena rzędu	Kryterium												
		FPE	AIC	HQ	SC	λ_{LR}	Bayesa ¹⁾							
							a	b	c	d	e	f	g	
20	0	20	16	26	58	4								
	1	1	1	1	1	0	3	15	17	18	22	36	38	
	2	45	36	39	31	38	12	59	58	60	57	50	48	
	3	12	11	11	4	8	17	12	14	12	11	9	9	
	4	6	10	5	2	19	32	10	7	6	7	2	2	
	5	16	26	18	4	31	36	4	4	4	3	3	3	
30	0	4	3	18	50	1								
	1	0	0	0	1	0	0	4	6	7	7	19	19	
	2	65	64	65	46	63	5	65	70	69	70	74	74	
	3	14	10	7	3	5	10	23	17	17	16	7	7	
	4	14	15	7	0	15	33	7	7	7	7	0	0	
	5	3	8	3	0	16	52	1	0	0	0	0	0	
50	0	1	0	3	14	0								
	1	0	0	2	5	1	0	0	0	0	0	2	3	
	2	89	85	94	81	84	0	71	76	77	77	94	93	
	3	6	7	0	0	1	1	23	18	17	17	3	3	
	4	3	4	1	0	6	32	5	5	5	5	1	1	
	5	1	4	0	0	8	67	1	1	1	1	0	0	
100	0	0	0	0	0	0								
	1	0	0	0	0	0	0	0	0	0	0	0	0	
	2	88	88	97	100	82	0	81	84	85	86	99	99	
	3	7	7	3	0	5	2	15	13	12	11	1	1	
	4	4	4	0	0	7	23	4	3	3	3	0	0	
	5	1	1	0	0	6	75	0	0	0	0	0	0	
200	0	0	0	0	0	0								
	1	0	0	0	0	0	0	0	0	0	0	0	0	
	2	88	88	100	100	88	0	86	89	90	91	100	100	
	3	8	8	0	0	2	1	12	10	9	9	0	0	
	4	3	3	0	0	4	27	2	1	1	0	0	0	
	5	1	1	0	0	6	72	0	0	0	0	0	0	

¹⁾ Kolumny (a)–(g) odpowiadają różnym wartościom stałej c .

LITERATURA

- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**, 243-247.
- Akaike, H. (1971). Autoregressive model fitting for control. *Annals of the Institute of Statistical Mathematics* **23**, 163-180.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, Budapest: *Académiai Kiadó*, 267-281.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716-723.
- Hannan, E.J., Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, **B41**, 190-195.
- Koziara, J. (1986). Prognozowanie dla wielowymiarowych szeregów czasowych. *XVI Colloquium Metodologiczne z Agro-Biometrii*, PAN Warszawa, 190-199.
- Krzyśko, M., Smoczyński, D. (1984). Parameter estimation and order determination of a multivariate autoregressive process. *COMPSTAT 1984, Proceedings in Computational Statistics, 6th Symposium held at Prague*, 35-40.
- Lütkepohl, H. (1991). *Introduction to multiple time series analysis*. Springer-Verlag.
- Paulsen, J. (1984). Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis* **5**, 115-127.
- Paulsen, J., Tjøstheim, D. (1985). On the estimation of residual variance and order in autoregressive time series. *Journal of the Royal Statistical Society* **B47**, 216-228.
- Quinn, B.G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society* **B42**, 182-185.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* **8**, 147-164.

*Praca wpłynęła 15 stycznia 1995;
w wersji ostatecznej 10 maja 1995*

The order determination methods for a multivariate autoregressive model

Summary

The criterions *FPE*, *AIC*, *HQ*, *SC*, Bayesian approach and sequential test for order selection of the vector AR models are considered. A comparison of methods and some simulation results are presented. The computer program in *Turbo Pascal 7.0* for this criterions has been written.

Key words: autoregressive proces, order of autoregression, criteria *FPE*, *AIC*, *HQ*, *SC*, Bayes method, likelihood ratio test.