

Kryteria wyboru modelu dla wielowymiarowych szeregów czasowych o zredukowanym rzędzie – studium symulacyjne

Marcin Hławka

Instytut Matematyki i Informatyki
Politechnika Wroclawska

XXXV Konferencja
„Statystyka Matematyczna Wiśła 2009”

7-11 grudnia 2009

Plan prezentacji

- 1 Opis problemu
- 2 Specyfikacja modelu RRVAR
- 3 Estymacja macierzy współczynników
- 4 Kryteria wyboru modelu RRVAR(p, r)
- 5 Symulacyjne porównanie metod wyboru parametrów p i r
- 6 Podsumowanie

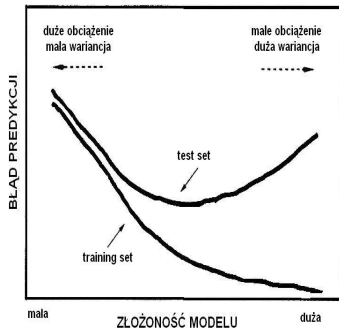


Jakimi szeregami zajmujemy się?

- Długość szeregu czasowego $T \in (50, 300)$
- Liczba składowych $K \in (3, 20)$
- Występuje korelacja pomiędzy poszczególnymi składnikami

Cel: wybór parametrów modelu

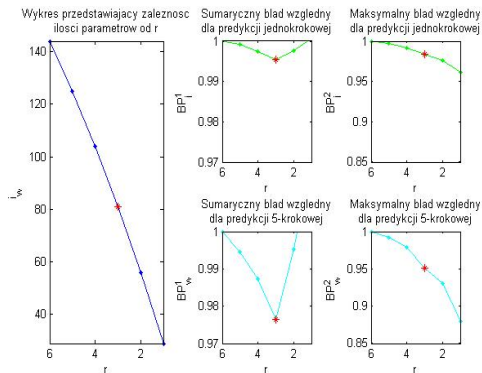
Chcemy wybrać optymalne parametry modelu $RRVAR(p, r)$:
 p – rząd autoregresji
 r – rząd macierzy współczynników



Rysunek: Porównanie błędu predykcji dla „training set” i „test set” w zależności od stopnia złożoności modelu

Dlaczego warto stosować modele o zredukowanym rzędzie?

- Lepsza interpretacja zależności pomiędzy składowymi szeregu czasowego
- Poprawa zdolności predykcyjnych
- Mniejsza liczba estymowanych współczynników
- Pozwalają wyznaczyć profile kanoniczne (wizualizacja)



Rysunek: Porównanie modelu RRVAR z modelem VAR dla $K = 15$ oraz $r = 5$

SPECYFIKACJA MODELU RRVAR

Definicja (Model VAR(p))

Szereg czasowy $\{Y_t\}$ jest szeregiem czasowym z modelu VAR(p) (ang. Vector Autoregression) jeżeli:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \epsilon_t,$$

gdzie:

- $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{Kt})^T$
- Φ_i dla $i = 1, \dots, p$ jest macierzą współczynników wymiaru $(K \times K)$
- $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{Kt})^T$

Definicja (Model RRVAR(p, r))

Szereg czasowy $\{Y_t\}$ jest szeregiem czasowym z modelu RRVAR(p, r) (ang. *Reduced Rank Vector Autoregression*) jeżeli:

$$Y_t = ABX_t + \epsilon_t \quad (t = 1, \dots, T)$$

gdzie:

- $X_t = [Y'_{t-1}, \dots, Y'_{t-p}]'$ — wektor kolumnowy wymiaru: $Kp \times 1$
- A — macierz współczynników wymiaru $K \times r$
- $B = [B_1, B_2, \dots, B_p]$ — macierz wymiaru $r \times Kp$
- B_i — macierz współczynników wymiaru $r \times K$
- ϵ_t — wektor błędu o nieosobliwej macierzy kowariancji

Zauważmy, że macierze współczynników z wyjściowego procesu VAR(p) wyrażają się przy pomocy współczynników RRVAR(p, r) w następujący sposób $\Phi_i = AB_i$.

Postać macierzy kowariancji

Model RRVAR jest w pełni określony przy pomocy następujących macierzy kowariancji:

- $\Sigma_{YY} = \text{Cov}(Y_t, Y_t) = E(YY')$
- $\Sigma_{XY} = \text{Cov}(X_t, Y_t) = E(XY')$
- $\Sigma_{XX} = \text{Cov}(X_t, X_t) = E(XX')$
- $\Sigma_{\epsilon} = \text{Cov}(\epsilon_t, \epsilon_t) = E(\epsilon\epsilon') = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$

Struktura macierzy kowariancji w modelu VAR, a tym samym też w modelu RRVAR jest strukturą blokową, gdzie budującymi je elementami są odpowiednie macierze autokowariancji zdefiniowane, jak poniżej:

Definicja

Macierz autokowariancji z opóźnieniem j określana jest w następujący sposób:

$$\Gamma(j) = E[Y_t Y'_{t+j}].$$

Postać macierzy kowariancji

Korzystając z tych oznaczeń odpowiadające macierze kowariancji przyjmują postać: $\Sigma_{YY} = \Gamma(0)$,

$$\Sigma_{XX} = \begin{bmatrix} \Gamma(0) & \Gamma(1)' & \dots & \Gamma(p-1)' \\ \Gamma(1) & \Gamma(0) & \dots & \Gamma(p-2)' \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(p-1) & \Gamma(p-2) & \dots & \Gamma(0) \end{bmatrix}, \Sigma_{XY} = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(p) \end{bmatrix}$$

Co gdy $\det(\Sigma_{XX})$ jest bliskie zera?

Gdy $\det(\Sigma_{XX})$ jest bliskie zera dokonujemy regularyzacji, to znaczy dodajemy macierz λI . W ten sposób otrzymujemy model RRRVAR.

Postać macierzy współczynników

Przy znanych parametrach p i r macierze współczynników wyrażają się następująco:

$$\begin{aligned}A(p, r) &= \Omega^{-\frac{1}{2}}[V_1, \dots, V_r] = \Omega^{-\frac{1}{2}}V = [\alpha_1, \dots, \alpha_r], \\B(p, r) &= V'\Omega^{\frac{1}{2}}\Sigma_{YX}\Sigma_{XX}^{-1} = [\beta_1, \dots, \beta_r],\end{aligned}$$

gdzie $V = [V_1, \dots, V_r]$ i V_j jest wektorem własnym (kolumnowym) odpowiadającym j -tej największej wartości własnej λ_j^2 macierzy $W = \Omega^{\frac{1}{2}}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Omega^{\frac{1}{2}}$ ($j = 1, \dots, r$). W praktyce należy zamienić autokowariancję teoretyczną autokowariancją próbkową.

Funkcje wagowe

Najbardziej powszechnymi i użytecznymi przykładami macierzy wagowej Ω w zagadnieniu minimalizacji są $\Omega = \Sigma_{YY}^{-1}$ oraz $\Omega = \Sigma_{\epsilon}^{-1}$.

KRYTERIA WYBORU MODELU RRVAR(p, r)

W jaki sposób wybieramy parametry modelu?

Dwa podejścia

Rozpatrujemy dwie zasadnicze kategorie metod służących do wyboru parametrów p i r :

- Dla pierwszego z nich, przy pomocy kryteriów wybierane są równocześnie rząd autokorelacji p oraz rząd macierzy współczynników r . Do metod jednoczesnego wyboru parametrów należą kryteria oparte na funkcji wiarygodności oraz kryteria oparte na błędzie predykcji.
- Drugi typ metod najpierw wybiera rząd p przy pomocy kryteriów wybierających zwykły model VAR. Następnie znając już parametr p stosujemy testowanie sekwencyjne, przy pomocy którego wyznaczamy rząd macierzy współczynników modelu RRVAR r .

Wybór parametrów p oraz r

Wyróżniamy następujące kryteria służące do jednoczesnego wyboru parametrów p oraz r :

- kryterium Akaika (AIC)
- kryterium Schwarza (SC)
- kryterium Hannana-Quinna (HQ)
- kryterium FPE
- kryteria oparte na metodzie cross-validation (CV)
- kryterium oparte na metodzie bootstrap (TRERR)

Wybór parametrów p oraz r

$$AIC(p, r) = \ln |\hat{\Sigma}_\epsilon(p, r)| + \left[\frac{2}{T-p} (rKp + rK - r^2) \right]$$

$$SC(p, r) = \ln |\hat{\Sigma}_\epsilon(p, r)| + \left[\frac{\ln T}{T-p} (rKp + rK - r^2) \right]$$

$$HQ(p, r) = \ln |\hat{\Sigma}_\epsilon(p, r)| + \left[\frac{2 \ln \ln T}{T-p} (rKp + rK - r^2) \right]$$

$$FPE(p, r) = |\hat{\Sigma}_\epsilon(p, r)| \left[\frac{T-p+(1+p)r}{T-p-(1+p)r} \right]^K$$

gdzie:

$$\hat{\Sigma}_\epsilon(p, r) = \frac{1}{T-p} (Y_{p-} - C(p, r)X_{p-})(Y_{p-} - C(p, r)X_{p-})'$$

Krok pierwszy – wybór rzędu autoregresji p

$$AIC_p(p) = \log|\hat{\Sigma}_p| + \frac{2pK^2}{T-p}$$

$$AIC_{C_p}(p) = \log|\hat{\Sigma}_p| + \frac{2pK^2 + K(K+1)}{T-p-Kp-K-1}$$

$$SC_p(p) = \log|\hat{\Sigma}_p| + \frac{\log(T-p)pK^2}{T-p}$$

$$HQ_p(p) = \log|\hat{\Sigma}_p| + \frac{2\log\log(T-p)pK^2}{T-p}$$

$$HQ_{C_p}(p) = \log|\hat{\Sigma}_p| + \frac{2\log\log(T-p)pK^2}{T-p-Kp-K-1}$$

$$FPE_p(p) = |\hat{\Sigma}_p| \left(\frac{T-p+Kp}{T-p-Kp} \right)^K$$

Krok drugi – wybór rzędu macierzy współczynników r

W sposób sekwencyjny, do momentu pierwszego przyjęcia hipotezy zerowej, będziemy dokonywać weryfikacji następującej hipotezy na poziomie istotności α (przez $rank$ oznaczamy rząd macierzy):

$$H_0 : rank(C) = r^* \text{ przeciwko } H_1 : rank(C) > r^*,$$

gdzie $r^* = 1, 2, \dots, K - 1$, a C jest macierzą współczynników. Jeżeli nie zatrzymamy się po $(K - 1)$ krokach, to przyjmujemy, że macierz C jest pełnego rzędu, tzn. $r^* = K$.

Wyróżniamy:

- test Bartletta (B)
- test Bartletta-Lawley'a (BC)
- test oparty na eliminacji Gaussa
- bootstrapowe wersje testów

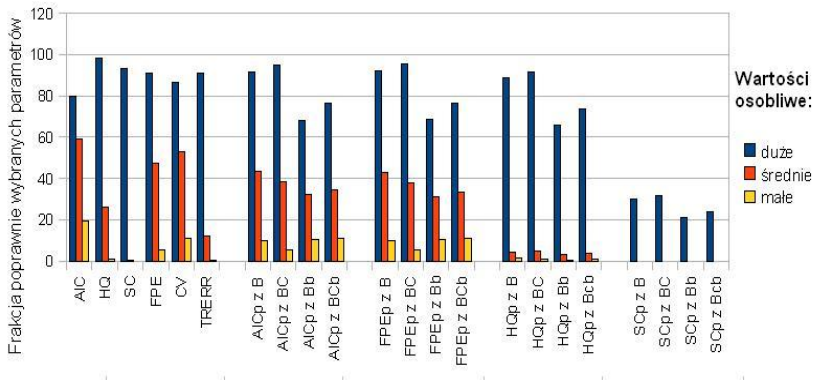
SYMULACYJNE PORÓWNIANIE METOD

Scenariusz symulacyjny

- Długość szeregu czasowego: $T_1 = 100$, $T_2 = 250$
- Wymiar szeregu czasowego: $K_1 = 5$, $K_2 = 9$
- Rząd autoregresji: $p_1 = 2$, $p_2 = 4$
- Rząd macierzy współczynników: dla $K_1 = 5$: $r_1 = 2$, $r_2 = 4$ oraz dla $K_2 = 9$: $r_3 = 3$ oraz $r_4 = 6$
- Powiązania pomiędzy poszczególnymi komponentami (realizowane przez wartości osobliwe):
 - duże – $\theta_i^{(1)} = 1$ dla $i = 1, \dots, r$; $\theta_i^{(1)} = 0$ dla $i = r + 1, \dots, K$
 - średnie – $\theta_i^{(2)} = 0.5$ dla $i = 1, \dots, r$; $\theta_i^{(2)} = 0$ dla $i = r + 1, \dots, K$
 - małe – $\theta_i^{(2)} = 0.3$ dla $i = 1, \dots, r$; $\theta_i^{(3)} = 0$ dla $i = r + 1, \dots, K$
- Macierz kowariancji błędów: $\Sigma_\epsilon^{(1)} = I$,

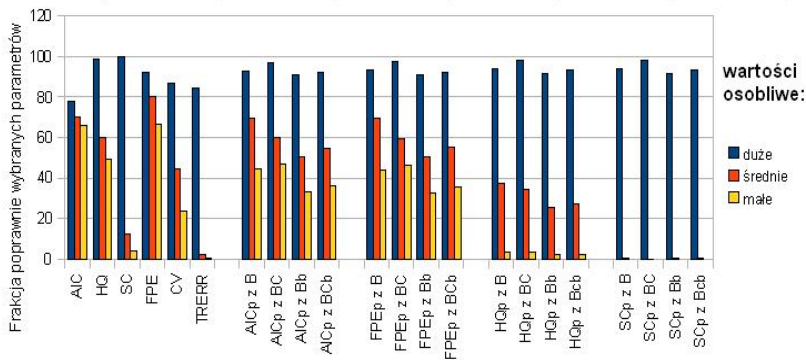
$$\Sigma_\epsilon^{(2)} = \begin{bmatrix} 1 & 0.3 & \dots & 0.3 \\ 0.3 & 1 & \dots & 0.3 \\ \vdots & \vdots & \ddots & \vdots \\ 0.3 & 0.3 & \dots & 1 \end{bmatrix}, \quad \Sigma_\epsilon^{(2)} = \begin{bmatrix} 1 & 0.7 & \dots & 0.7 \\ 0.7 & 1 & \dots & 0.7 \\ \vdots & \vdots & \ddots & \vdots \\ 0.7 & 0.7 & \dots & 1 \end{bmatrix}$$

Porównanie kryteriów – wpływ wartości osobliwych



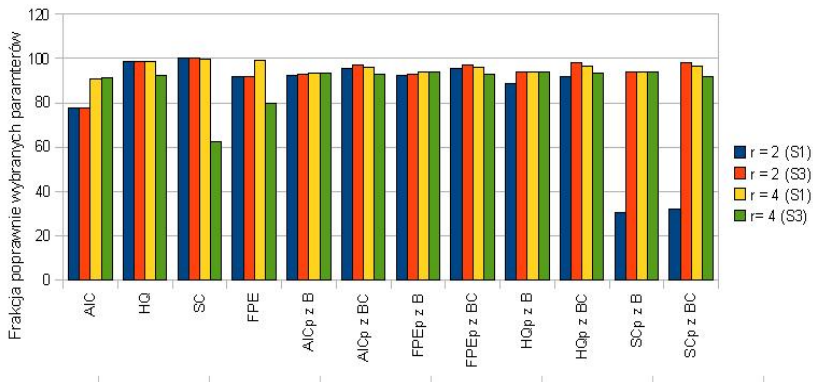
Rysunek: Porównanie kryteriów dla $K = 5$, $T = 100$, $p = 2$, $r = 2$ oraz macierzy kowariancji błęd $\Sigma_{\epsilon}^{(1)}$

Porównanie kryteriów – wpływ macierzy kowariancji błędu



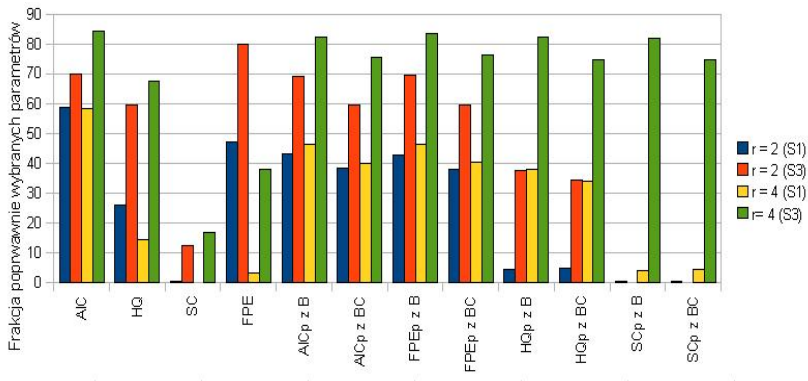
Rysunek: Porównanie kryteriów dla $K = 5$, $T = 100$, $p = 2$, $r = 2$ oraz macierzy kowariancji błędu $\Sigma_{\epsilon}^{(3)}$

Porównanie kryteriów – wpływ rzędu macierzy współczynników



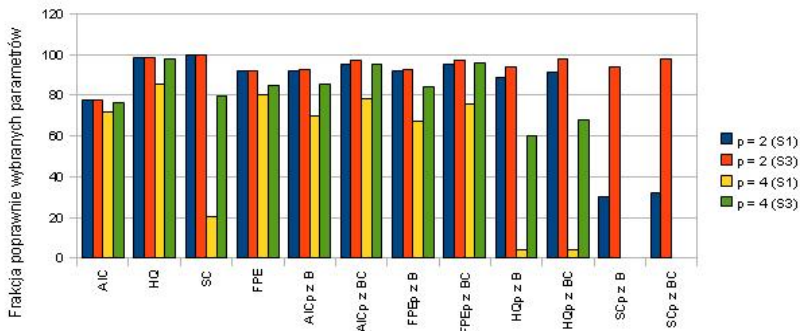
Rysunek: Porównanie kryteriów dla $r = 2$ oraz $r = 4$ oraz macierzy kowariancji błędów $\Sigma_{\epsilon}^{(1)}$ i $\Sigma_{\epsilon}^{(3)}$ (pozostałe parametry $K = 5$, $T = 100$, $\rho = 2$, duże wartości osoblive)

Porównanie kryteriów – wpływ rzędu macierzy współczynników



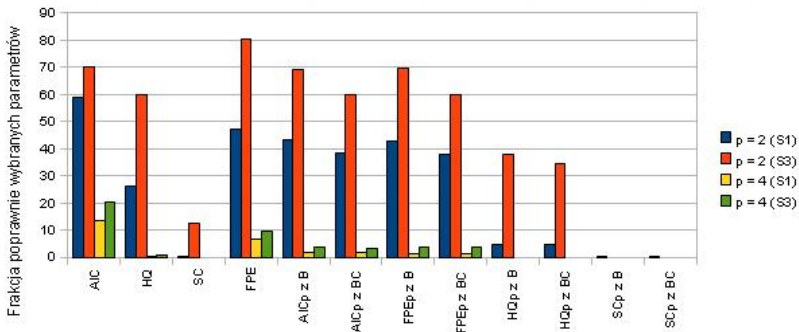
Rysunek: Porównanie kryteriów dla $r = 2$ oraz $r = 4$ oraz macierzy kowariancji błęd $\Sigma_{\epsilon}^{(1)}$ i $\Sigma_{\epsilon}^{(3)}$ (pozostałe parametry $K = 5$, $T = 100$, $p = 2$, średnie wartości osobiwe)

Porównanie kryteriów – wpływ rzędu autoregresji



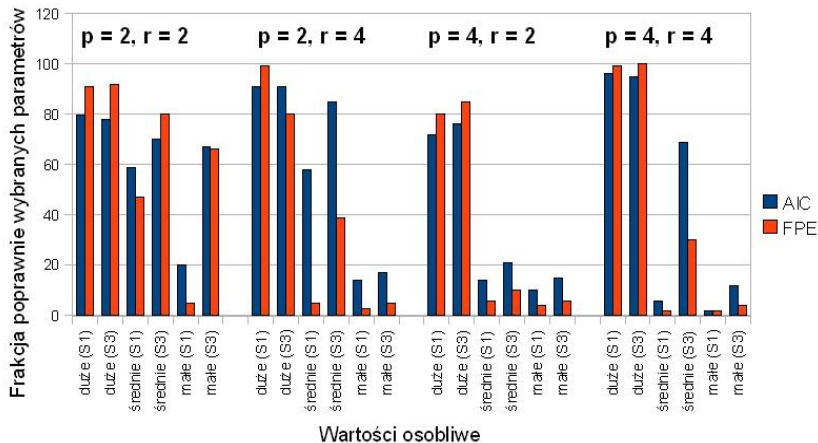
Rysunek: Porównanie kryteriów dla $p = 2$ oraz $p = 4$ oraz macierzy kowariancji błędów $\Sigma_{\epsilon}^{(1)}$ i $\Sigma_{\epsilon}^{(3)}$ (pozostałe parametry $K = 5$, $T = 100$, $r = 2$, duże wartości osobliwe)

Porównanie kryteriów – wpływ rzędu autoregresji



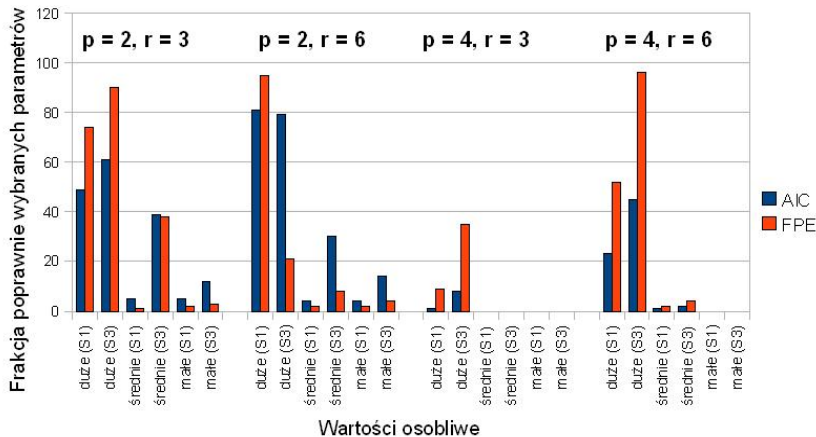
Rysunek: Porównanie kryteriów dla $p = 2$ oraz $p = 4$ oraz macierzy kowariancji błędów $\Sigma_{\epsilon}^{(1)}$ i $\Sigma_{\epsilon}^{(3)}$ (pozostałe parametry $K = 5$, $T = 100$, $r = 2$, średnie wartości osobliwe)

Porównanie dwóch najlepszych kryteriów



Rysunek: Porównanie kryterium AIC oraz FPE dla $K = 5$ oraz $T = 100$

Porównanie dwóch najlepszych kryteriów



Rysunek: Porównanie kryterium AIC oraz FPE dla $K = 9$ oraz $T = 100$

- Zarówno zwiększenie liczby składowych, jak i zwiększenie rzeczywistego rzędu autoregresji ma bardzo istotny wpływ na skuteczność kryteriów. W praktyce nie dysponujemy przeważnie szeregami czasowymi dłuższymi niż $T = 100$. Dla takiego przypadku jedynie w sytuacji dużych wartości szczególnych kryteria pozostają jeszcze skuteczne, gdy zwiększymy liczbę składowych z $K = 5$ do $K = 9$ lub gdy zwiększymy rząd autoregresji z $p = 2$ na $p = 4$. Dla pozostałych przypadków skuteczność kryteriów jest mniejsza niż 40%.
- Wielkość wartości szczególnych macierzy współczynników ma największe znaczenie na skuteczność kryteriów. Gdy wartości te są zbyt małe kryteria nie wychwytyją ich, co w praktyce utrudnia wybór rzeczywistego modelu. Z drugiej strony jednak wybranie modelu o niższym rzędzie, w tym przypadku, powoduje zmniejszenie błędu predykcji.

- Zwiększenie długości szeregu czasowego skutkuje poprawieniem skuteczności kryteriów. Nie zawsze jednak poprawa ta jest znaczna. W przypadku, gdy niektóre wartości szczególne są bardzo małe (mniejsze od 0.3), to wówczas nawet przy zwiększeniu długości szeregu czasowego do $T = 250$ kryteria wybierają poprawny model w około 10 – 30% przypadków. Natomiast, gdy wartości szczególne są już odpowiednio duże (większe niż 0.5) to zwiększenie szeregu czasowego do $T = 250$ powoduje, że skuteczność większości kryteriów waha się pomiędzy 75 – 100%.
- Zmniejszenie rzeczywistego rzędu macierzy współczynników powoduje spadek skuteczności kryteriów.
- Wielkość korelacji pomiędzy składowymi wektora błędu ma zasadnicze znaczenie na skuteczność kryteriów. Zwiększanie tej wartości poprawia skuteczność kryteriów w prawie każdej sytuacji.

Wyznaczenie profili kanonicznych, dzięki którym możemy wyznaczyć główne trendy występujące w szeregach, a tym samym możemy np.

- identyfikować składowe okresowe (klasyfikacja genów)
- identyfikacja zachowań klientów, co pozwala dostosować oferty bankowe

Odkrywanie powiązań pomiędzy składowymi szeregu czasowego.

Dziękuję za uwagę

