The autocorrelation and autocovariance functions - helpful tools in the modeling problem

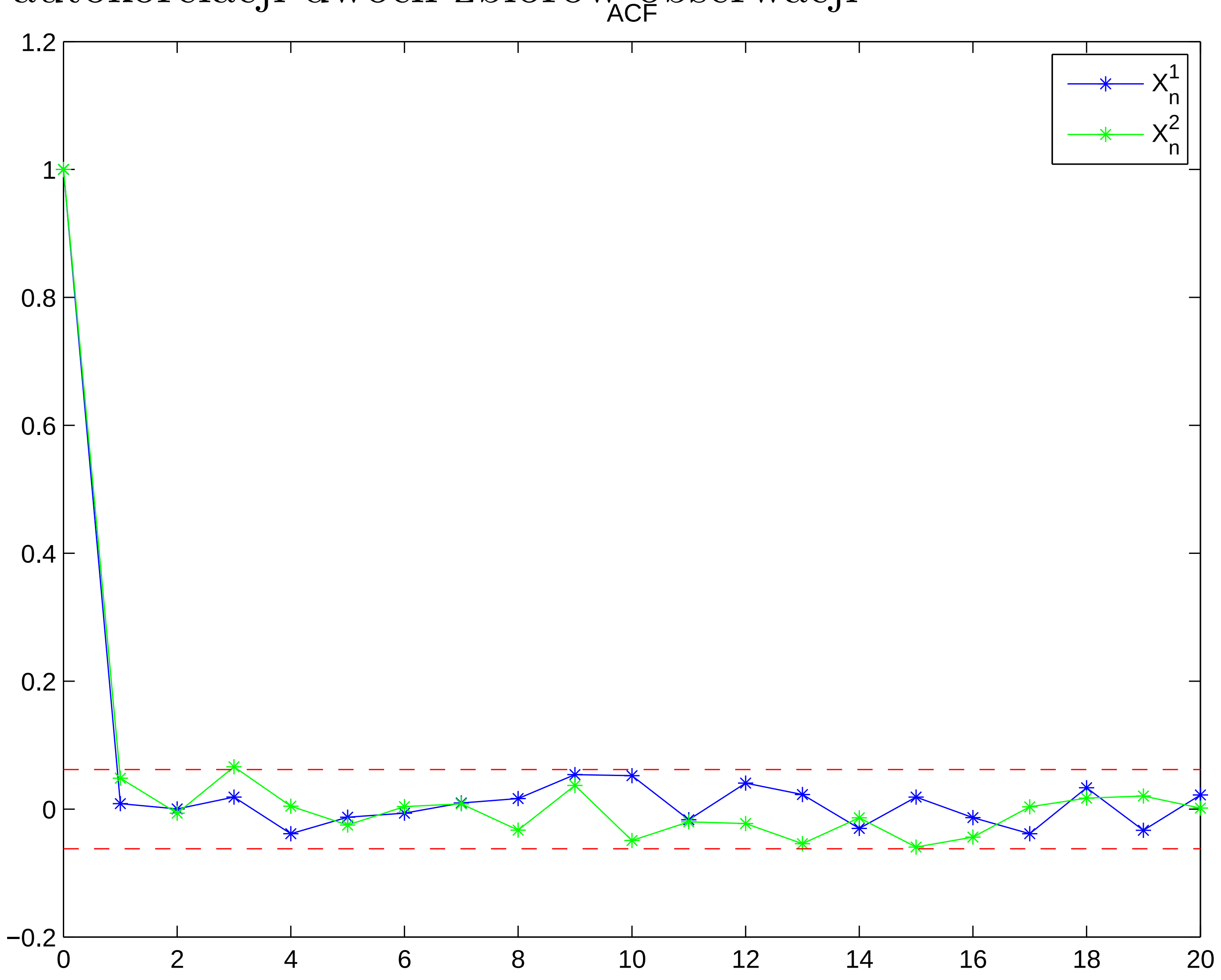
Joanna Nowicka-Zagrajek

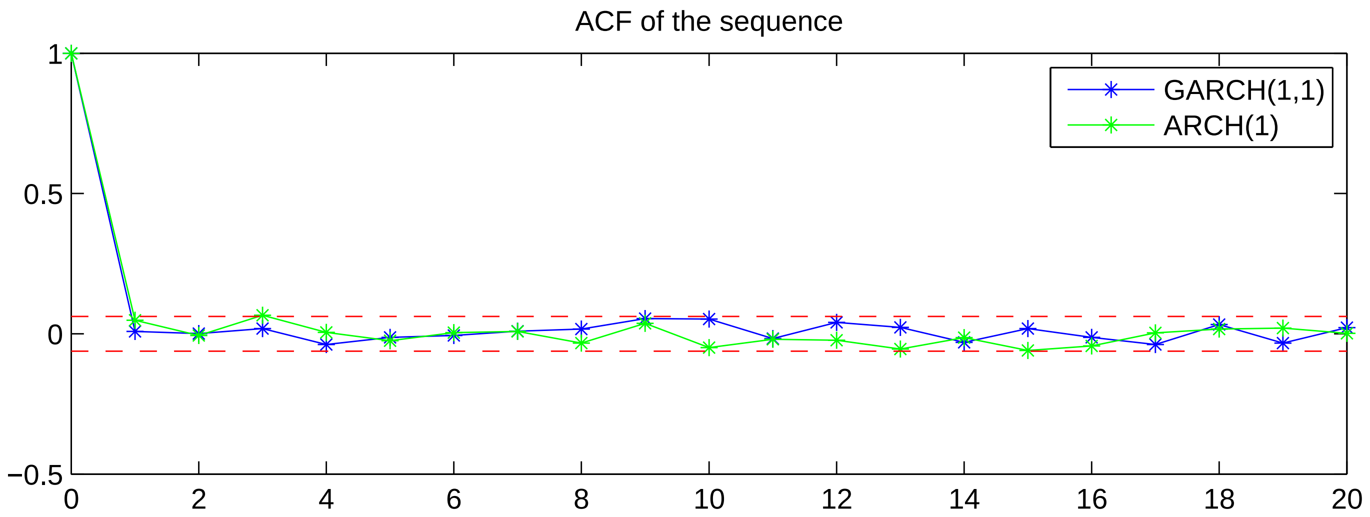
Agnieszka Wyłomańska

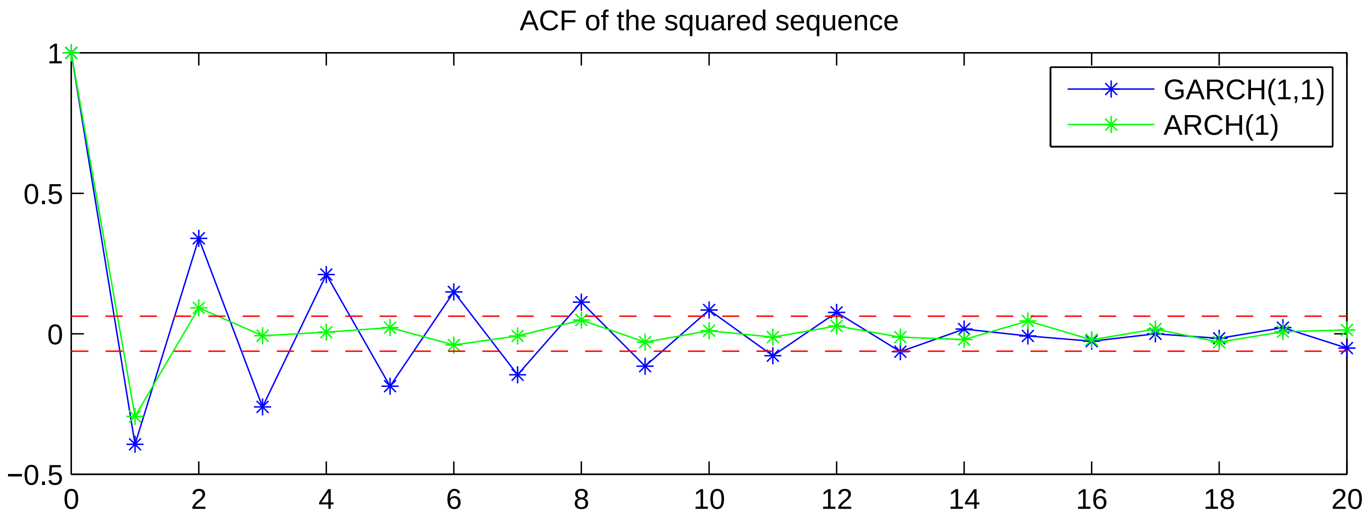
Institute of Mathematics and Computer Science
Hugo-Steinhaus Center for Stochastic Processes
Wrocław University of Technology
Poland

Spis treści

Funkcja autokorelacji dwóch zbiorów obserwacji




ACF of the sequence
GARCH(1,1)
ARCH(1)
1
0.5
0
−0.5
0
2
4
6
8
10
12
14
16
18
20


ACF of the squared sequence
GARCH(1,1)
ARCH(1)
1
0.5
0
−0.5
0
2
4
6
8
10
12
14
16
18
20

Podstawowe definicje i założenia

- X, Y - zmienne losowe o skończonych wartościach oczekiwanych $\mathbb{E}X, \mathbb{E}Y$ i skończonych wariancjach $\mathbb{V}ar(X), \mathbb{V}ar(Y)$
- Najbardziej popularną miarą zależności X i Y jest kowariancja:

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$$

- W wielu zastosowaniach zamiast kowariancji używa się tzw. przeskalowanej kowariancji (korelacji):

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathbb{V}ar(X)}\sqrt{\mathbb{V}ar(Y)}}$$

- Obydwie miary są symetryczne.

- Przy studiowaniu struktury zależności szeregu czasowego $\{X_n\}$ wykorzystuje się tzw. funkcję autokowariancji

$$\mathrm{cov}(X_n, X_m) = \mathbb{E}(X_n X_m) - \mathbb{E}X_n \mathbb{E}X_m$$

oraz funkcję autokorelacji (ACF)

$$\mathrm{corr}(X_n, Y_m) = \frac{\mathrm{cov}(X_n, X_m)}{\sqrt{\mathbb{V}ar(X_n)}\sqrt{\mathbb{V}ar(X_m)}}$$

- Dla stacjonarnego szeregu czasowego funkcje te przybierają postać

$$\mathrm{cov}(X_n, X_m) = \mathrm{cov}(X_0 X_{m-n}) = \mathbb{E}(X_0 X_{m-n}) - (\mathbb{E}X_0)^2 = \mathrm{cov}_X(m-n),$$

$$\mathrm{corr}(X_n, Y_m) = \mathrm{corr}(X_0 X_{m-n}) = \frac{\mathrm{cov}_X(m-n)}{\mathbb{V}ar(X_0)} = \mathrm{corr}_X(m-n).$$

- Estymatorami tych funkcji są autokowariancja i autokorelacja próbkowa

$$\hat{\operatorname{cov}_X}(h) = \hat{\operatorname{cov}}(X_0, X_h) = \frac{1}{n}\sum_{i=1}^{n-h}(x_i - \bar{x})(x_{i+h} - \bar{x}),$$

$$\hat{\operatorname{corr}_X}(h) = \frac{\hat{\operatorname{cov}_X}(h)}{\hat{\operatorname{cov}_X}(0)},$$

gdzie $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ oraz $x_1, x_2, \ldots, x_n$ są zaobserwowanymi wartościami szeregu $\{X_n\}$.

- W pracy studiujemy nie tylko strukturę zależności obserwowanego szeregu czasowego $\{X_n\}$, ale również szeregu utworzonego z kwadratów wyjsciowych danych

$$Y_n = X_n^2$$

- Zakładamy, że $\mathbb{E}X_n^4 = \mathbb{E}Y_n^2$ istnieje dla każdego n
- Szereg $\{\xi_n\}$ - niezależne zmienne losowe o jednakowym rozkładzie (i.i.d.) o średniej zero i skończonym czwartym momencie z następującymi parametrami

$$\sigma^2 = \mathbb{E}\xi_n^2, \quad \gamma = \mathbb{E}\xi_n^4.$$

Niezależne zmienne losowe o jednakowym rozkładzie

Rozpatrujemy szereg czasowy o średniej zero niezależnych zmiennych losowych o jednakowym rozkładzie o skończonym czwartym momencie, tzn. $X_n = \xi_n$.

- Szereg ten jest stacjonarny
- Funkcja autokowariancji

$$\mathrm{cov}_X(k) = \mathrm{cov}(X_n, X_{n+k}) = \sigma^2 \mathbb{I}_{k=0}$$

- Funkcja autokorelacji

$$\mathrm{corr}_X(k) = \mathbb{I}_{k=0}.$$

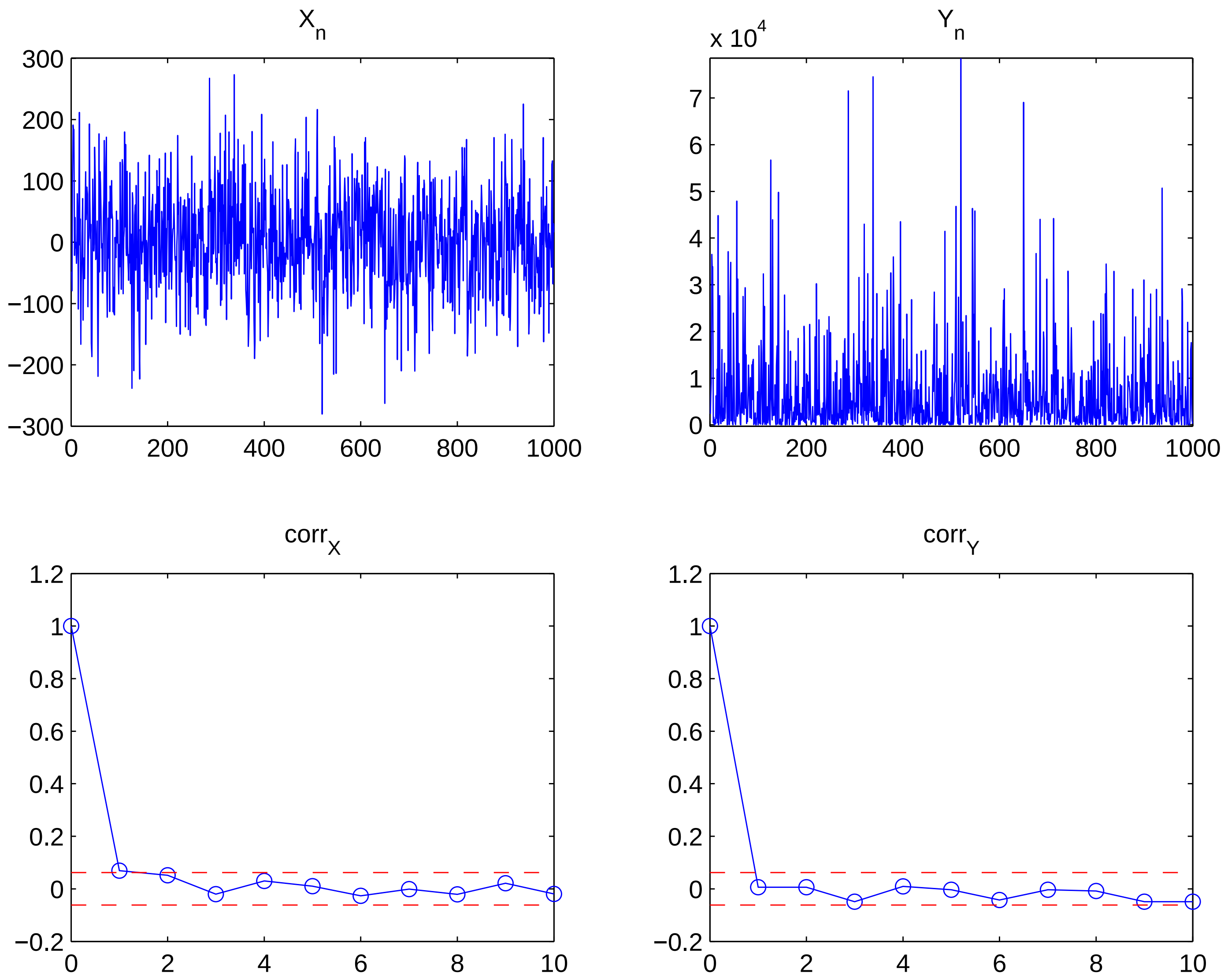
- Funkcja autokowariancji dla szeregu kwadratów $Y_n = X_n^2$

$$\mathrm{cov}_Y(k) = (\gamma - \sigma^4)\mathbb{I}_{k=0},$$

- Funkcja autokorelacji dla szeregu kwadratów

$$\mathrm{corr}_Y(k) = \mathbb{I}_{k=0}.$$

- Jedynie dla $k = 0$ rozpatrywane miary zależności są niezerowe dla $\{X_n\}$ oraz $\{Y_n\}$.


X_n
300
200
100
0
−100
−200
−300
0
200
400
600
800
1000
Y_n
x 10^4
7
6
5
4
3
2
1
0
0
200
400
600
800
1000
$corr_X$
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
0
2
4
6
8
10
$corr_Y$
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
0
2
4
6
8
10

Stacjonarne modele ARMA

DEFINICJA 1 *Proces* $\{X_n\} = \{X_n, n = 0, \pm 1, \pm 2, ...\}$ *jest szeregiem ARMA(p,q)* $(p, q \in N)$, *jeśli* $\{X_n\}$ *jest stacjonarny oraz dla każdego* $n \in N$ *zachodzi*

$$X_n - b_1 X_{n-1} - ... - b_p X_{n-p} = \xi_n + a_1 \xi_{n-1} + ... + a_q \xi_{n-q}, \qquad (1.1)$$

gdzie innowacje ξ_n *są i.i.d. ze średnią zero oraz* $E\xi_n^2 = \sigma^2$.

Równanie (1.1) może być zapisane następująco

$$B(L)X_n = A(L)\xi_n, \quad n = 0, \pm 1, \pm 2, ...,$$

gdzie $A(z)$ i $B(z)$ są wielomianami odpowienio stopni q i p

$$\begin{aligned} A(z) &= 1 + a_1 z + ... + a_q z^q, \\ B(z) &= 1 - b_1 z - ... - b_p z^p, \end{aligned}$$

oraz L jest operatorem przesunięcia

$$L^j X_n = X_{n-j}, \quad j = 0, \pm 1, \pm 2, ... \quad .$$

Jeśli wielomiany $A(z)$ i $B(z)$ nie mają wspólnych pierwiastków oraz pierwiastki $B(z)$ są poza dyskiem jednostkowym $\{z : |z| \leq 1\}$, wtedy $\{X_n\}$ jest szeregiem przyczynowym (causal) oraz

$$X_n = \sum_{j=0}^{\infty} \phi_j \xi_{n-j}, \quad \text{a.s.} \quad n = 0, \pm 1, \pm 2, \dots \ ,$$

gdzie

$$\sum_{j=0}^{\infty} \phi_j z^j = \frac{A(z)}{B(z)}.$$

Dla przyczynowego modelu ARMA(p,q) otrzymujemy

$$
\begin{aligned}
\mathrm{cov}_X(k) &= \mathrm{cov}(X_{n+k}, X_n) = \sigma^2 \sum_{j=0}^{\infty} \phi_j \phi_{j+|k|}, \\
\mathrm{corr}_X(k) &= \mathrm{corr}(X_{n+k}, X_n) = \frac{\sum_{j=0}^{\infty} \phi_j \phi_{j+|k|}}{\sum_{j=0}^{\infty} \phi_j^2}, \\
\mathrm{cov}_Y(k) &= \gamma \sum_{j=0}^{\infty} (\phi_j \phi_{j+|k|})^2 + \sigma^4 \left(\sum_{j=0}^{\infty} \phi_j \phi_{j+|k|} \right)^2 \\
\mathrm{corr}_Y(k) &= \frac{\gamma \sum_{j=0}^{\infty} (\phi_j \phi_{j+|k|})^2 + \sigma^4 \left(\sum_{j=0}^{\infty} \phi_j \phi_{j+|k|} \right)^2}{\gamma \sum_{j=0}^{\infty} {\phi_j}^4 + \sigma^4 \left(\sum_{j=0}^{\infty} {\phi_j}^2 \right)^2}.
\end{aligned}
$$

Model AR(1)

AR(1) model (model autoregressji rzędu 1) dany jest wzorem

$$X_n - bX_{n-1} = \xi_n,$$

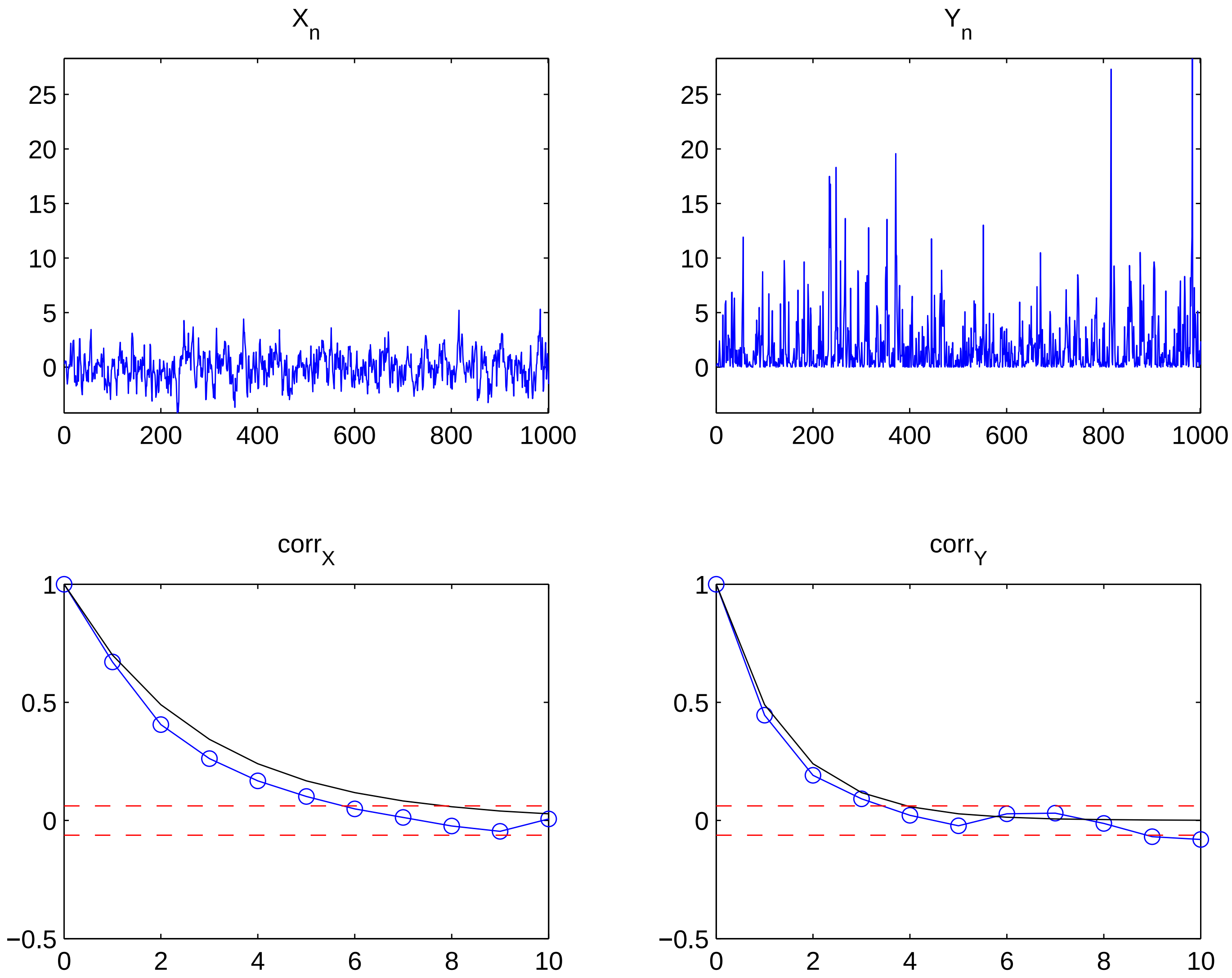
gdzie ξ_n i.i.d. ze średnią zero. Przy założeniu $0 < |b| < 1$ model AR(1) jest przyczynowy $X_n = \sum_{j=0}^{\infty} b^j \xi_{n-j}$.

Dla modelu AR(1) otrzymujemy

$$\begin{aligned} \mathrm{cov}_X(k) = \mathrm{cov}(X_{n+k}, X_n) &= \frac{b^k \sigma^2}{1-b^2}, \\ \mathrm{corr}_X(k) = \mathrm{corr}(X_{n+k}, X_n) &= b^k. \end{aligned}$$

Obie miary exponencjalne dążą do zera dla wzrastających wielkości k. Ponadto

$$\begin{aligned} \mathrm{cov}_Y(k) = \mathrm{cov}(Y_{n+k}, Y_n) &= b^{2k} \left(\frac{\gamma}{1-b^4} + \frac{\sigma^4}{(1-b^2)^2} \right), \\ \mathrm{corr}_Y(k) = \mathrm{corr}(Y_{n+k}, Y_n) &= b^{2k}. \end{aligned}$$


X_n
Y_n
$corr_X$
$corr_Y$
25
20
15
10
5
0
0
200
400
600
800
1000
1
0.5
−0.5
2
4
6
8
10

Model MA(1)

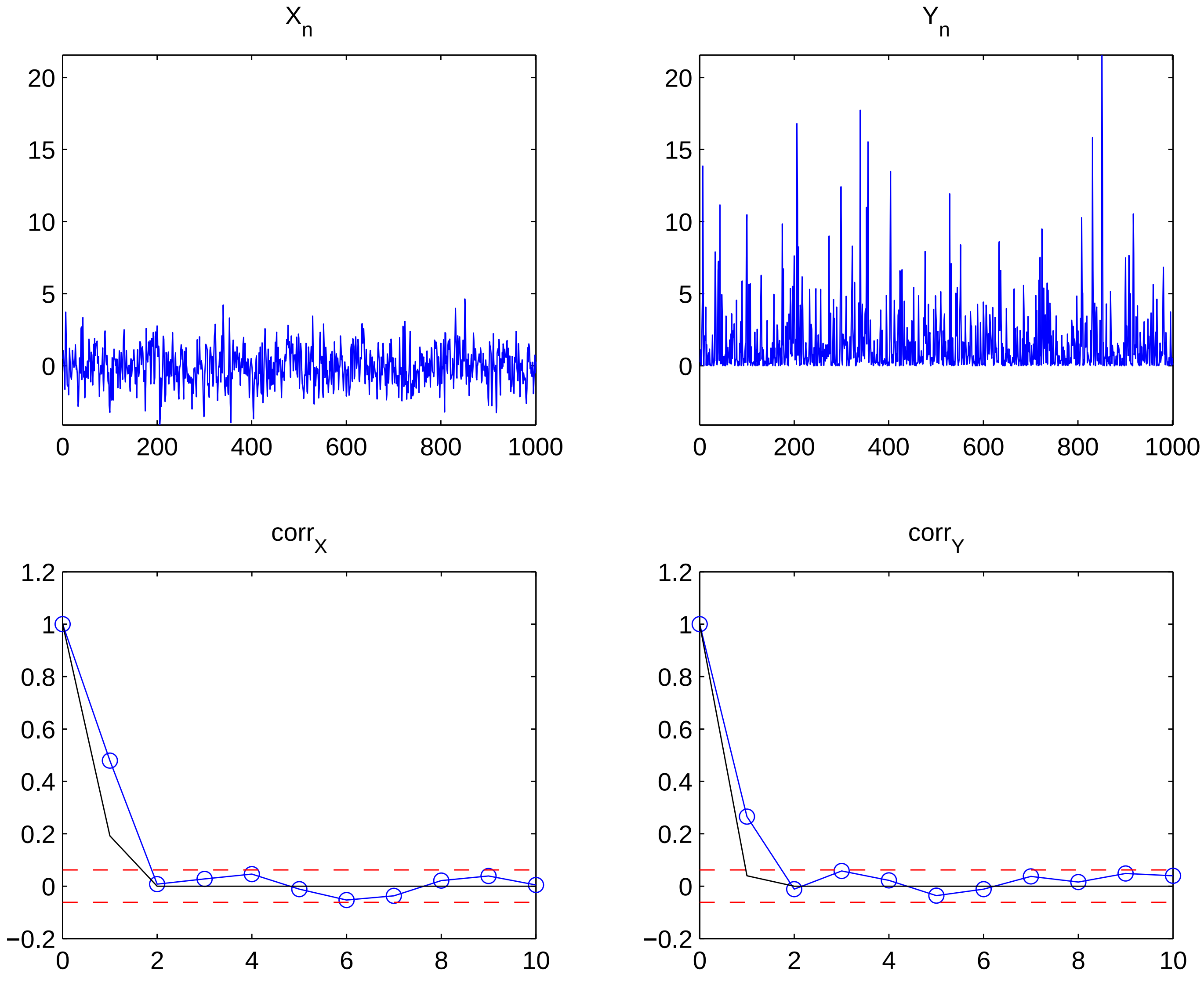
Model MA(1) (moving average) dany jest wzorem:

$$X_n = \xi_n + a\xi_{n-1},$$

gdzie innowacje ξ_n są i.i.d. ze średnią zero.

Dla modelu MA(1) otrzymujemy

$$
\begin{aligned}
\mathrm{cov}_X(k) = \mathrm{cov}(X_{n+k}, X_n) &= \sigma^2(\mathbb{I}_{k=0} + a^2\mathbb{I}_{k=0} + a\mathbb{I}_{k=1}) \\
\mathrm{corr}_X(k) = \mathrm{corr}(X_{n+k}, X_n) &= \frac{\mathbb{I}_{k=0} + a^2\mathbb{I}_{k=0} + a\mathbb{I}_{k=1}}{1+a^2} \\
\mathrm{cov}_Y(k) = \mathrm{cov}(Y_{n+k}, Y_n) &= \begin{cases} \gamma(1+a^4) + \sigma^4(4a^2 - 1 - a^4) & \text{dla } k = 0, \\ a^2\gamma - \sigma^4 a^2 & \text{dla } k = 1, \\ 0 & \text{dla } k > 1. \end{cases} \\
\mathrm{corr}_Y(k) = \mathrm{corr}(Y_{n+k}, Y_n) &= \begin{cases} 1 & \text{dla } k = 0, \\ \frac{a^2\gamma - \sigma^4 a^2}{\gamma(1+a^4) + \sigma^4(4a^2-1-a^4)} & \text{dla } k = 1, \\ 0 & \text{dla } k > 1. \end{cases}
\end{aligned}
$$


X_n
Y_n
20
15
10
5
0
0
200
400
600
800
1000
$corr_X$
$corr_Y$
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
0
2
4
6
8
10

Modele ARCH(p)

DEFINICJA 2 *Proces ARCH(q) ($q \in N$) z gaussowskimi innowacjami jest zdefiniowany następującym równaniem*

$$\begin{aligned} X_n &= \sqrt{h_n}\xi_n, \quad n = 0, 1, 2, ..., && (1.2) \\ h_n &= \theta + \sum_{j=1}^{q} \psi_j X_{n-j}^2, && (1.3) \end{aligned}$$

gdzie innowacje ξ_n są i.i.d. o standardowym rozkładzie normalnym oraz $\theta, \psi_q > 0$, $\psi_j \geq 0$, $j = 1, ..., q-1$.

Model ARCH(2)

Model ARCH(2) z gaussowskimi innowacjami dany jest wzorem

$$X_n = \sqrt{h_n}\xi_n$$

$$h_n = \theta + \psi_1 X_{n-1}^2 + \psi_2 X_{n-2}^2.$$

W modelu tym warunkiem gwarantująym istnienie czwartego momentu jest

$$\psi_2 + 3\psi_1^2 + 3\psi_2^2 + 3\psi_1^2\psi_2 - 3\psi_2^3 < 1.$$

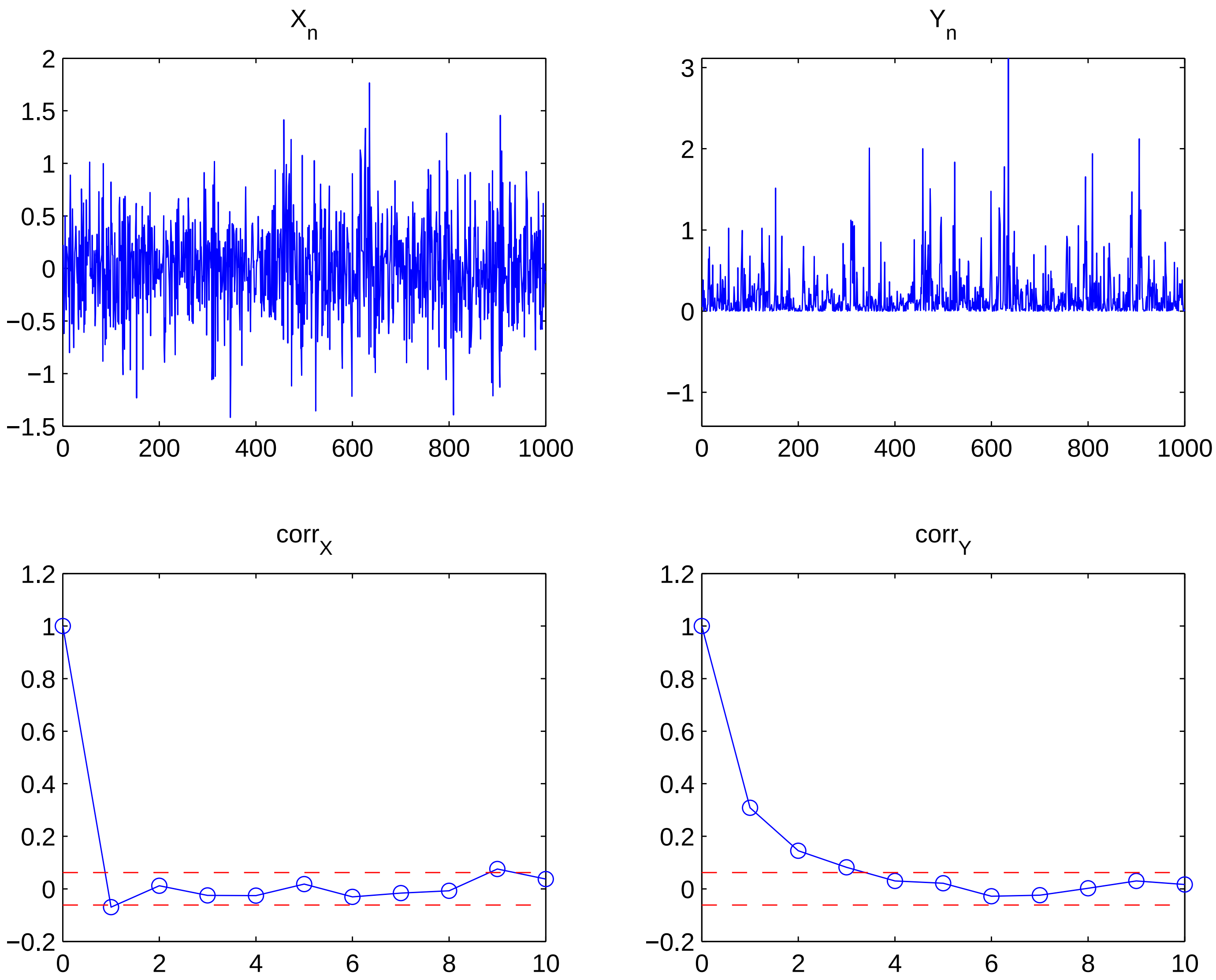
Dla modelu ARCH(2) z gaussowskimi innowacjami otrzymujemy

$$
\begin{aligned}
\mathrm{cov}_X(k) = \mathrm{cov}(X_{n+k}, X_n) &= \frac{\sigma^2\theta}{1-\sigma^2(\psi_1+\psi_2)}\mathbb{I}_{k=0} \\
\mathrm{corr}_X(k) = \mathrm{corr}(X_{n+k}, X_n) &= \mathbb{I}_{k=0} \\
\mathrm{corr}_Y(k) = \mathrm{corr}(Y_{n+k}, Y_n) &= \begin{cases} 1 & \text{dla } k=0, \\ \frac{\psi_1}{1-\psi_2} & \text{dla } k=1, \\ \frac{\psi_2+\psi_1^2-\psi_2^2}{1-\psi_2} & \text{dla } k=2, \\ \psi_1\mathrm{corr}(k-1)+\psi_2\mathrm{corr}(k-2) & \text{dla } k>2. \end{cases}
\end{aligned}
$$

W ogólnym przypadku

$$
\mathrm{corr}_Y(k) = \sum_{i=1}^{max(p,q)} (\psi_i+\phi_i)\mathrm{corr}(k-1), \quad k>p,
$$

gdzie $\psi_i = 0$ for $i > q$ oraz $\phi_i = 0$ dla $j > p$.


X n
2
1.5
1
0.5
0
−0.5
−1
−1.5
0
200
400
600
800
1000
Y n
3
2
1
0
−1
corr X
1.2
0.8
0.6
0.4
0.2
−0.2
2
4
6
8
10
corr Y

Modele GARCH(p,q)

DEFINICJA 3 *Proces GARCH(p,q)* $(p, q \in N)$ *z gaussowskimi innowacjami zdefiniowany jest następująco*

$$\begin{aligned} X_n &= \sqrt{h_n}\xi_n, \quad n = 0, 1, 2, ..., \\ h_n &= \theta + \sum_{j=1}^{p} \phi_j h_{n-j} + \sum_{j=1}^{q} \psi_j X_{n-j}^2, \end{aligned}$$

gdzie innowacje ξ_n *są i.i.d. o standardowym rozkładzie normalnym oraz* $\theta, \phi_p, \psi_q > 0$, $\phi_j \geq 0$, $j = 1, ..., p-1$, $\psi_j \geq 0$, $j = 1, ..., q-1$.

Model GARCH(1,1)

Model GARCH(1,1) z gaussowskimi innowacjami dany jest wzorem

$$X_n = \sqrt{h_n}\xi_n, \quad n = 0, 1, 2, ...,$$

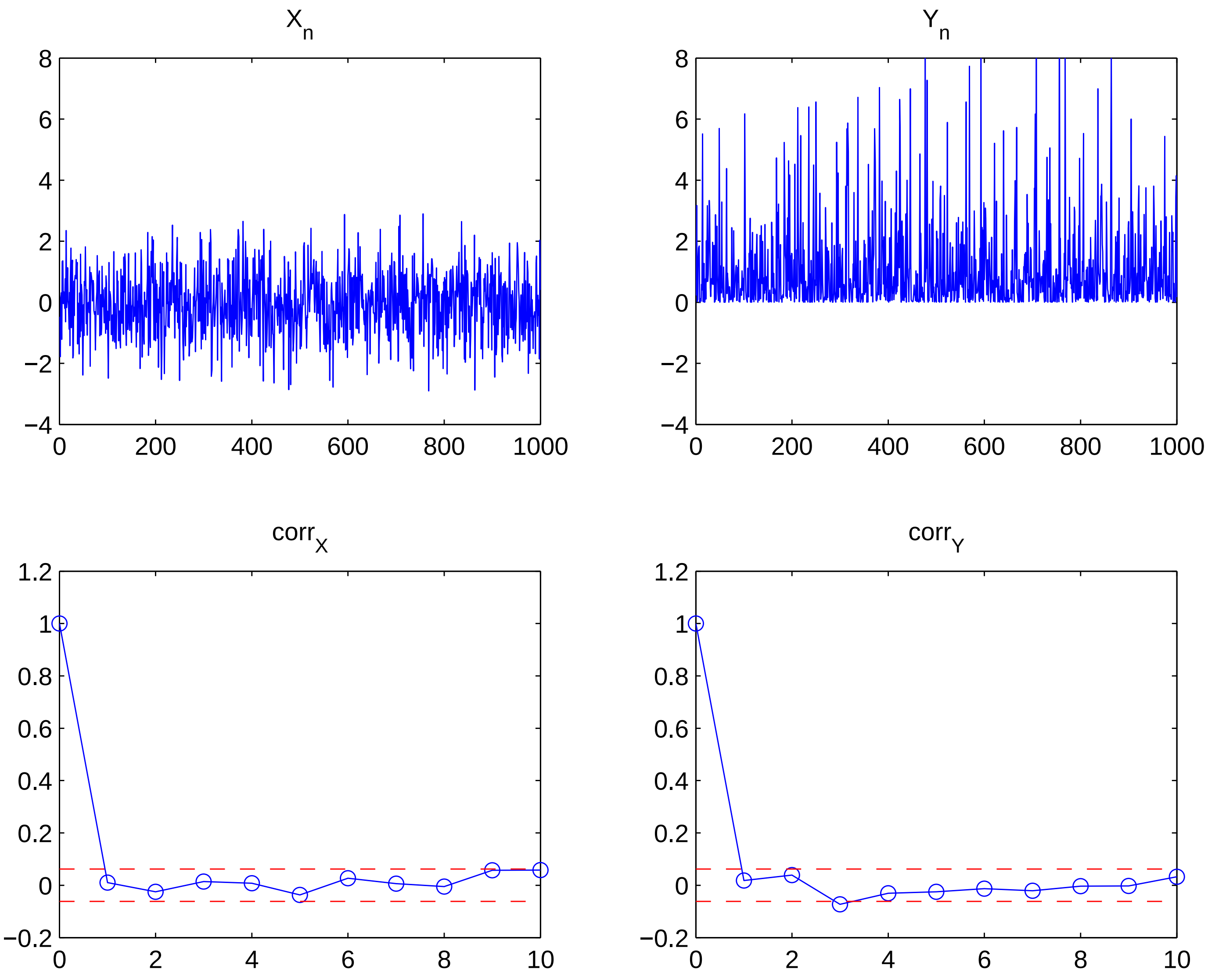
$$h_n = \theta + \phi h_{n-1} + \psi X_{n-1}^2.$$

W modelu tym warunkiem gwarantująym istnienie czwartego momentu jest

$$3\psi^2 + 2\psi\phi + \phi^2 < 1.$$

Dla modelu GARCH(1,1) z gaussowskimi innowacjami otrzymujemy

$$
\begin{aligned}
\mathrm{cov}_X(k) = \mathrm{cov}(X_{n+k}, X_n) &= \frac{\sigma^2\theta}{1-\phi-\sigma^2\psi}\mathbb{I}_{k=0} \\
\mathrm{corr}_X(k) = \mathrm{corr}(X_{n+k}, X_n) &= \mathbb{I}_{k=0} \\
\mathrm{corr}_Y(k) = \mathrm{corr}(Y_{n+k}, Y_n) &= \begin{cases} 1 & \text{dla } k = 0, \\ \frac{\psi(1-\psi\phi-\pi^2)}{1-2\psi\phi-\phi^2} & \text{dla } k = 1, \\ (\psi+\phi)^{k-1}\mathrm{corr}(1) & \text{dla } k > 1. \end{cases}
\end{aligned}
$$


X_n
Y_n
$corr_X$
$corr_Y$

Podsumowanie

- Funkcje autokowariancji i autokorelacji są dobrymi narzędziami wykorzystywanymi w ananlizie danych rzeczywistych
- Ich w miarę prosta postać powoduje że są to narzędzia powszechnie wykorzystywane w praktyce
- Przy ich użyciu można zweryfikować hipotezę odnośnie wyboru odpowiedniego modelu
- Konieczne jest jednak badanie nie tylko miar zależności danego szeregu, ale również szeregu utworzonego z kwadratów obserwowanych wartości.

Literatura

[**1**]T. Bollerslev ”On the correlation structure for the generalized autoregressive conditional heteroskedastic process” Journal of Time Series Analysis, Vol. 9, No 2, 1988.

[**2**]P.J. Brockwell and R.A. Davis ”Introduction to Time Series and Forecasting”, Springer-Verlag, New York,second edition, 2002.

[**3**]A. Milhoj ”The Moment Structure of ARCH processes” Scandinavian Journal of Statistics 12, 281-292, 1985.