Szeregi czasowe, analiza zależności krótko- i długozasięgowych

Rafał Weron
rweron@im.pwr.wroc.pl

Definicje

- Mając dany proces $\{X_t\}$ **autokowariancję** definiujemy jako :
  $\gamma(t_1, t_2) = \mathrm{cov}(X_{t_1}, X_{t_2}) =$
  $= E[(X_{t_1} - EX_{t_1})(X_{t_2} - EX_{t_2})]$
- Proces $\{X_t\}$ nazywamy **stacjonarnym** (w szerszym sensie), jeśli:
- $E(X_t) = \mu$ oraz $\mathrm{cov}(X_t, X_{t+\tau}) = \gamma(\tau)$, dla każdego t, τ.

Definicje cd.

- Niech $\{X_t\}$ będzie procesem stacjonarnym, wówczas **autokorelacją** nazywamy funkcję
- $\rho(\tau) = \gamma(\tau) / \gamma(0)$
- **Białym szumem** nazywamy
- proces $\{Z_t\}$ taki, że Z_t są niezależnymi zmiennymi losowymi oraz $E(Z_t) = 0$, $\mathrm{Var}(Z_t) = \sigma^2 < \infty$

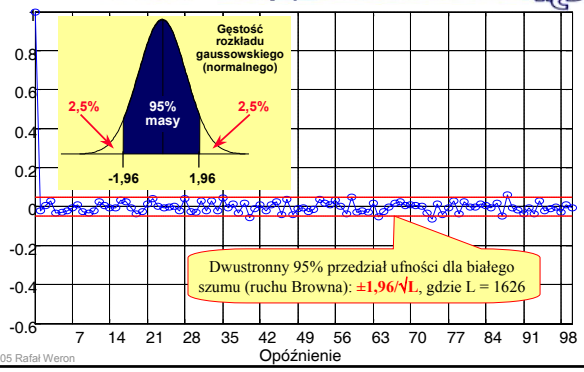
Funkcja (opóźnionej) korelacji
(AutoCorrelation Function)

- Dla szeregu czasowego Z_t wykreślamy funkcję autokorelacji **acf**(Z,k) jako funkcję opóźnienia k

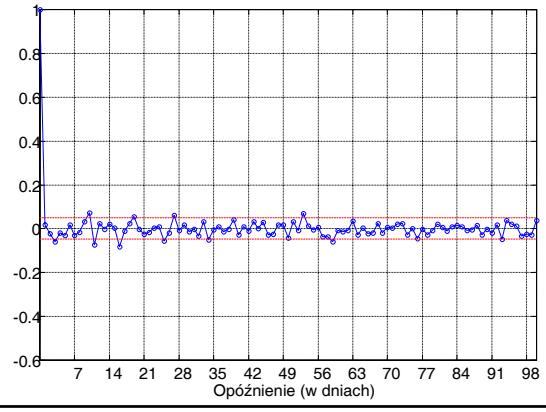
$$\mathbf{acf}(Z,k) = \frac{\sum_{t=k+1}^{N}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{N}(Z_t - \bar{Z})^2},$$

$$\bar{Z} = \frac{1}{N}\sum_{t=1}^{N} Z_t.$$

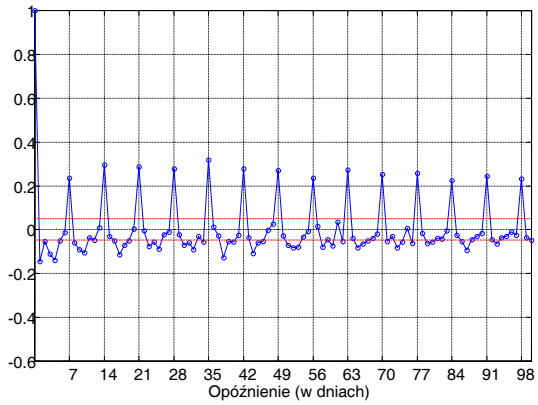
Funkcja acf dla 1626 wartości (gaussowskiego) białego szumu



Funkcja acf 1626 dziennych zwrotów (% zmian cen) r_t indeksu DJIA (93/07/23-99/12/30)



Funkcja acf 1626 dziennych zwrotów r_t indeksu DJ Palo Verde (96/05/06-00/10/25)



Szeregi autoregresji AR
(*AutoRegression*)

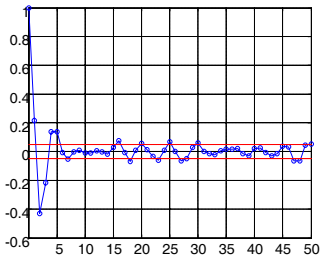
- Niech $\{Z_t\}$ będzie białym szumem, wówczas $\{X_t\}$ jest **szeregiem AR rzędu p** AR(p), jeśli:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t,$$

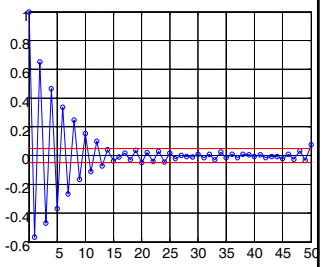
gdzie
- $\alpha_1,\ldots,\alpha_p$ – stałe
- Z_t traktuje się jako błąd losowy

Funkcja acf dla procesu AR

$X_t = 0{,}3\, X_{t-1} - 0{,}5\, X_{t-2} + Z_t$



$X_t = -0{,}3\, X_{t-1} + 0{,}5\, X_{t-2} + Z_t$

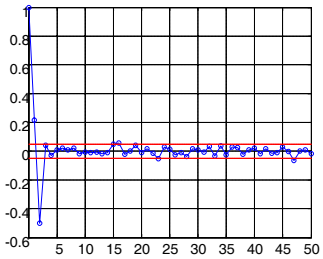


Funkcja częściowej korelacji
(*Partial AutoCorrelation Function*)

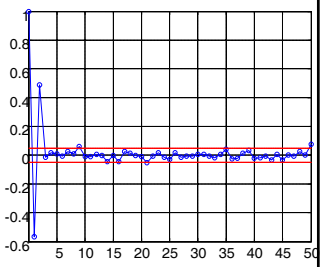
- Dla opóźnienia k funkcja częściowej autokorelacji **pacf**(X,k) mierzy liniową zależność między X_t i X_{t-k} „poprawioną" o wpływ $X_{t-1}, \ldots, X_{t-k+1}$
- Tzn. **pacf**(X,k) to współczynnik ϕ_{kk} regresji liniowej w modelu: $X_t = \phi_0 + \phi_{k1}X_{t-1} + \ldots + \phi_{kk}X_{t-k} + e_t$
- **pacf**$(X,1)$ = **acf**$(X,1)$
- Dla procesu AR pacf wskazuje na rząd p modelu

Funkcja pacf dla procesu AR

$X_t = 0{,}3\, X_{t-1} - 0{,}5\, X_{t-2} + Z_t$



$X_t = -0{,}3\, X_{t-1} + 0{,}5\, X_{t-2} + Z_t$



Szeregi ARMA
(AutoRegressive Moving Average)

- Bazując na szeregu AR, ale uwzględniając autokorelację błędów losowych otrzymujemy **szereg ARMA(p,q)**:

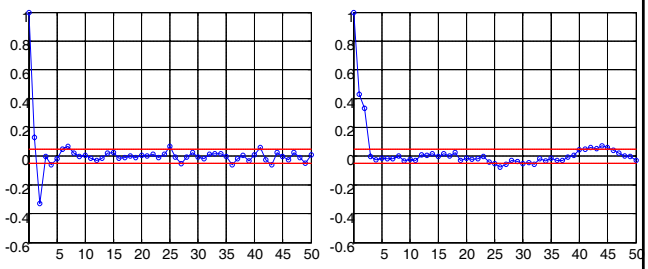
$$X_t = \underbrace{\alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p}}_{\text{AR(p)}} + \underbrace{Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}}_{\text{MA(q)}}$$

AR(p)

MA(q)
tzw. średnia ruchoma
moving average

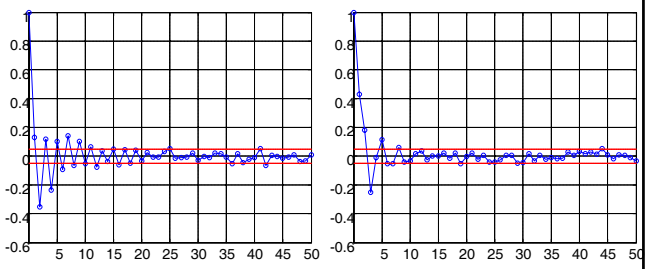
Funkcja acf dla procesu MA

$X_t = Z_t + 0{,}4\, Z_{t-1} - 0{,}5\, Z_{t-2}$ $\qquad$ $X_t = Z_t + 0{,}4\, Z_{t-1} + 0{,}5\, Z_{t-2}$



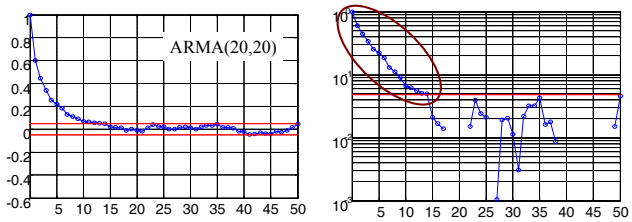
Funkcja pacf dla procesu MA

$X_t = Z_t + 0{,}4\, Z_{t-1} - 0{,}5\, Z_{t-2}$ $\qquad$ $X_t = Z_t + 0{,}4\, Z_{t-1} + 0{,}5\, Z_{t-2}$



Własności szeregów ARMA

- Modele ARMA charakteryzują się **wykładniczo** (szybko) zanikającą funkcją autokorelacji
- Modele ARMA „dobrze" opisują mikrostrukturę rynku



Szeregi ARIMA

- W praktyce mamy zazwyczaj do czynienia z niestacjonarnymi szeregami czasowymi.
- Eliminacja niestacjonarności poprzez różnicowanie zmiennych.

Różnicowanie zmiennych

- Polega ono na zastąpieniu zmiennej X_t przez zmienną $Y_t = X_{t+1} - X_t$
- oznaczenie : $Y_t = \nabla X_{t+1}$
- Zabieg ten pozwala pozbyć się niestacjonarności.

Różnicowanie - przykład

- Niech:
  $X_t = \beta_0 + \beta_1 t + Z_t$; t - czas, Z_t- biały szum
- $E(X_t) = E(\beta_0 + \beta_1 t + Z_t) = \beta_0 + \beta_1 t$
  - zależność od czasu
- Weźmy teraz $Y_t = \nabla X_{t+1}$, wówczas:
  $Y_t = X_{t+1} - X_t = \beta_1 + Z_{t+1} - Z_t$
- $E(Y_t) = \beta_1$
  - stacjonarność

Różnicowanie wielokrotne

- Definicja operatora B :

$$B^k(X_t) = X_{t-k}$$

- Różnicowanie n-krotne :

$$Y_t = \nabla^n X_{t+1} = (1 - B)^n X_{t+1}$$

- Przykład dla n = 3:

$$Y_t = (1 - B)^3 (X_{t+1}) =$$
$$= (1 - 3B + 3B^2 - B^3) X_{t+1} =$$
$$= X_{t+1} - 3X_t + 3X_{t-1} - X_{t-2}$$

Szeregi ARIMA

- **Szereg ARIMA (p, d, q):**

$$(1 - \alpha_1 B - \ldots - \alpha_p B^p)(1 - B)^d X_t =$$
$$= (1 + \beta_1 B + \ldots + \beta_q B^q) Z_t$$

gdzie α_i, β_j - stałe; $1 \le i \le p$, $1 \le j \le q$; $\{Z_t\}$ - szum.

- Zauważmy, że wprowadzając $Y_t = \nabla^d X_{t+1}$ z szeregu ARIMA(p, d, q) otrzymujemy szereg ARMA(p,q):

$$(1 - \alpha_1 B - \ldots - \alpha_p B^p) Y_t = (1 + \beta_1 B + \ldots + \beta_q B^q) Z_t$$

Szeregi ARCH

- W modelu ARIMA uwzględnia się zmienność wartości oczekiwanej w czasie. Co robić, gdy zmienia się wariancja ?
- Taki przypadek opisuje model ARCH - pierwszy szereg czasowy uwzględniający zależność wariancji procesu od jego poprzednich wartości.

Szeregi ARCH

- Szereg $\{X_t\}$ jest szeregiem ARCH(p), jeśli:

$$\mathbf{X_t = (h_t)^{1/2} Z_t},$$

gdzie :

$h_t = \alpha_0 + \alpha_1 X^2_{t-1} + \ldots + \alpha_p X^2_{t-p}$ - wariancja ;

$\alpha_0, \alpha_p > 0, \alpha_1, \ldots, \alpha_{p-1} \ge 0$; $\{Z_t\}$ - biały szum

- Zauważmy, że szumy w modelu ARCH są mnożone, a nie dodawane.

Szeregi GARCH

- Uogólnieniem modelu ARCH jest model GARCH (generalized ARCH)
- W przypadku szeregów GARCH wariancja h_t zależy nie tylko od poprzednich wartości szeregu, ale także od poprzednich wariancji.

Szeregi GARCH - definicja

- Szereg $\{X_t\}$ jest szeregiem GARCH(p, q), jeśli:

$$X_t = (h_t)^{1/2} Z_t,$$

gdzie:

$h_t = \alpha_0 + \alpha_1 X^2_{t-1} + \ldots + \alpha_p X^2_{t-p} + \beta_1 h_{t-1} + \ldots + \beta_q h_{t-q}$

$\alpha_0, \alpha_p, \beta_q > 0$,

a pozostałe współczynniki są nieujemne

Estymacja parametrów

- We wszystkich przedstawionych modelach występowały stałe, których często nie znamy.
- Estymacja stałych na przykładzie szeregu ARCH:
  - niech $\alpha = (\alpha_0,..., \alpha_p)$,
  - wtedy dysponując próbą $X = (X_1,..., X_n)$ możemy estymować wektor α
  - jeśli $X_t \sim N(0, h_t)$, to stosując metodę NW wystarczy znaleźć maksimum funkcji:
    $\ln L = - (n/2) \ln 2\pi - (1/2)\sum^n_{t=1}[\ln h_t(\alpha) + (Z^2_t / h_t(\alpha))]$

Estymacja parametrów cd.

- Im większa wartość parametru p dla szeregu ARCH(p), tym lepsze dopasowanie modelu do danych.
- Z drugiej strony estymacja większej liczby parametrów wiąże się z większymi błędami.
- Kompromis - kryterium Akaikego.

Kryterium Akaikego

- Do wyznaczania optymalnego modelu można stosować kryterium Akaikego:
  $$AIC(p) = -2 \ln L + 2p$$
- Najmniejsza wartość AIC(p) wskazuje na optymalną wielkość modelu.
- Najczęściej dobre dopasowanie uzyskujemy dopiero dla bardzo dużych wartości p.

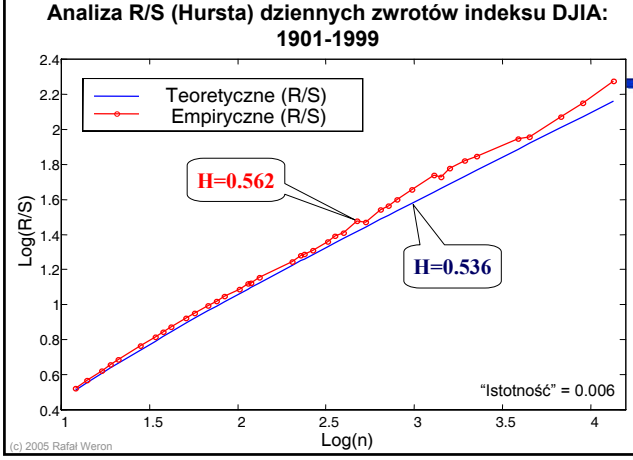
Analiza zależności długoterminowych

Analizy R/S (Hursta) oraz DFA, dają odpowiedź na pytanie czy dzisiejszy wzrost ceny spowoduje kolejny wzrost, spadek czy też nie ma wpływu na przyszłe ruchy cen

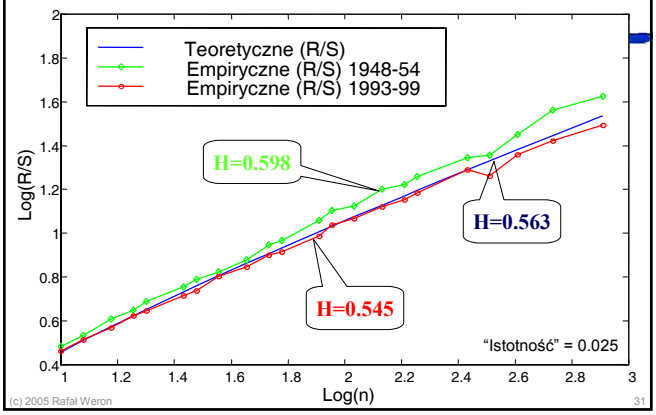
Analiza R/S (Hursta)

- Podziel szereg danych o długości N na d podprzedziałów o długości n
- Dla każdego podprzedziału:
  - oblicz średnią i odchylenie std. (S) zwrotów
  - odejmij średnią od zwrotów
  - zbuduj szereg skumulowanych zwrotów Y_i
  - oblicz zasięg (*range*) R=max$\{Y_i\}$-min$\{Y_i\}$
  - przeskaluj (*rescale*) zasięg: R/S

$$(R/S)_n = \frac{1}{d}\sum_{m=1}^{d}\frac{R_m}{S_m}$$

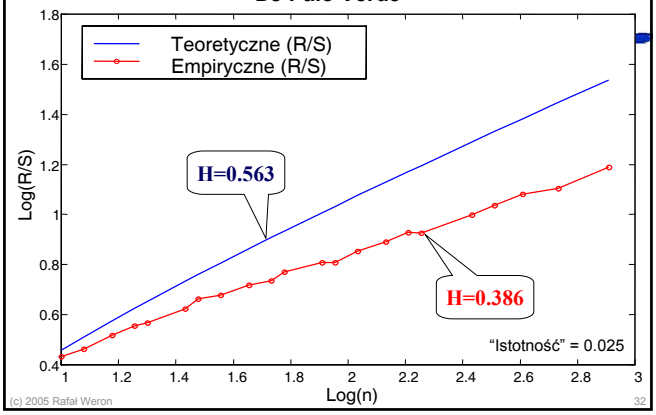


Analiza R/S (Hursta) 1620 dziennych zwrotów indeksu DJIA



(c) 2005 Rafał Weron

Analiza R/S (Hursta) 1620 dziennych zwrotów indeksu DJ Palo Verde



(c) 2005 Rafał Weron

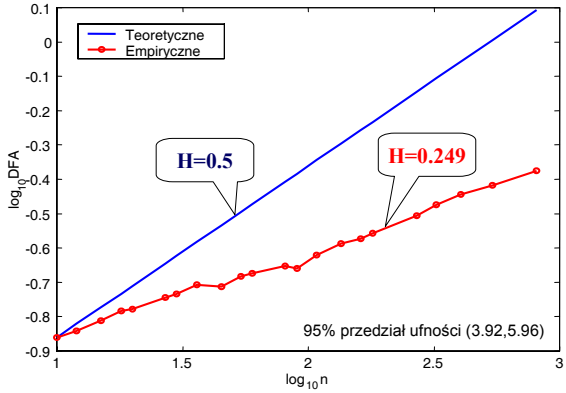
Analiza DFA

(Detrended Fluctuation Analysis)

- Podziel szereg danych (zwrotów) o długości N na d podprzedziałów o długości n
- Dla każdego podprzedziału:
  - wyznacz trend liniowy $z(t)=at+b$
  - odejmij trend liniowy i policz odchylenie std.:

$$F(n)=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y(t)-z(t))^2}$$

- Policz średnią wartość $F_{śr}$ funkcji $F(n)$
- Wykreśl $F_{śr}$ jako funkcję n

(c) 2005 Rafał Weron

Analiza DFA 1620 dziennych zwrotów indeksu DJ Palo Verde