

WYKŁAD: Szeregi czasowe II

Zaawansowane Metody Uczenia Maszynowego

Zwroty indeksów finansowych

Y_t : indeks finansowy w momencie t (wartość waloru, kurs walutowy itp).

Określimy zwrot indeksu finansowego jako jego relatywną zmianę.

Z reguły rozpatruje się dwa rodzaje zwrotów:

Zwrot prosty Y_t w momencie t (simple return) jest zdefiniowany jako

$$\tilde{R}_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

Zwrot prosty jest zmianą Y_t , $\Delta Y_t = Y_t - Y_{t-1}$ względem wartości Y_{t-1} .

Znacznie częściej używanym zwrotem jest **zwrot logarytmiczny** Y_t w momencie t (dla okresu czasu $h = 1$) zdefiniowany jako

$$R_t = \log(Y_t/Y_{t-1}) = (1-B) \log Y_t = \Delta \log Y_t = \log\left(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) \approx \tilde{R}_t,$$

gdzie przybliżona równość zachodzi, gdy wartość \tilde{R}_t jest mała.

($BY_t = Y_{t-1}$.)

Zwroty indeksów finansowych cd

Analogicznie definiuje się zwrot logarytmiczny dla okresu czasu h :

$$R_{t,h} = \log(Y_t/Y_{t-h}).$$

Kilka użytecznych własności powoduje, że zwroty logarytmiczne są częściej używane. Suma zwrotów logarytmicznych dla h kolejnych jednostek czasu daje zwrot logarytmiczny dla okresu czasu h

$$R_t + \dots + \dots R_{t-h+1} = \log(Y_t/Y_{t-h}).$$

Zwroty logarytmiczne są symetryczne: ujemna wartość zwrotu logarytmicznego równa co do wartości bezwzględnej dodatniej wartości w poprzednim momencie oznacza powrót do tej samej wartości indeksu. Np.

$$R_t = \log\left(\frac{Y_t}{Y_{t-1}}\right) = 0,5$$

$$R_{t+1} = \log\left(\frac{Y_{t+1}}{Y_t}\right) = -0,5$$

wtedy $Y_{t+1} = \exp(-0,5) Y_t = \exp(-0,5) \exp(0,5) Y_{t-1} = Y_{t-1}$.

Inaczej dla zwrotów prostych

$$\tilde{R}_t = \frac{(Y_t - Y_{t-1})}{Y_{t-1}} = 0.5$$

$$\tilde{R}_{t+1} = \frac{(Y_{t+1} - Y_t)}{Y_t} = -0.5,$$

then $Y_{t+1} = 0.5Y_t = 0,5 \times 1,5Y_{t-1} = 0.75Y_{t-1}$.

Historycznie pierwszym modelem dla zwrotów (R_t) była hipoteza błędzenia przypadkowego:

Hipoteza błędzenia przypadkowego

R_t jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie.

Niezgodna z faktami empirycznymi, które później sformułowano w postaci praw (stylizowanych faktów) zwrotów finansowych.

Stylizowane fakty o zwrotach indeksów finansowych

Analiza danych finansowych prowadzi do sformułowania trzech faktów o zwrotach:

- Ogony rozkładu zwrotów R_t maleją wolniej niż rozkładu $N(0, 1)$;
- R_t są nieskorelowane, ale R_t^2 są skorelowane,
- Duże zmiany kolejnych wartości R_t często następują po poprzednich dużych zmianach indeksu.

Zauważmy, że drugi i trzeci z faktów przeczy hipotezie błędzenia przypadkowego.

Zdefiniujmy **zmienność** (*volatility*) indeksu jako pewną miarę rozproszenia R_t , z reguły odchylenie standardowe, ale, w zależności od kontekstu, rozumiane bezwarunkowo lub warunkowo pod warunkiem przeszłości procesu.

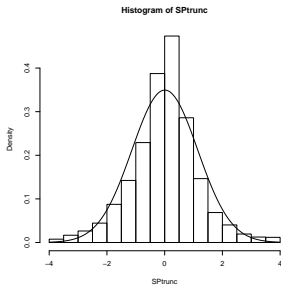
Pierwszy fakt stylizowany

Rozpatrzmy pierwszy fakt stylizowany dla indeksu *S&P500*.

Dla tego szeregu $\min R_t = -23\%$. Przy założeniu normalności prawdopodobieństwo, że otrzymamy taką wartość lub większą wynosi $2,23 \times 10^{-97}$, to powinno zdarzyć się nie częściej raz na 10^{96} lat (wiek ziemi oceniany jest na 10^7 lat). Przeczy to w oczywisty sposób normalności rozkładu zwrotów.

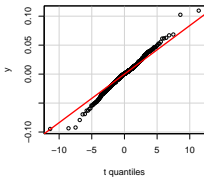
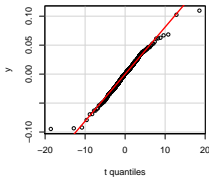
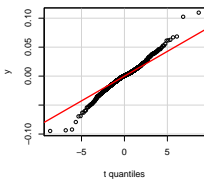
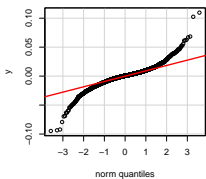
Poniżej histogram indeksu *S&P500* (w procentach i obcięty na 4%) z nałożoną gęstością rozkładu normalnego ze średnią i wariancją takimi, jak parametry empiryczne.

Poniżej histogram indeksu *S&P500* (w procentach i obciety na 4%) z nałożoną gęstością rozkładu normalnego ze średnią i wariancją takimi, jak parametry empiryczne.



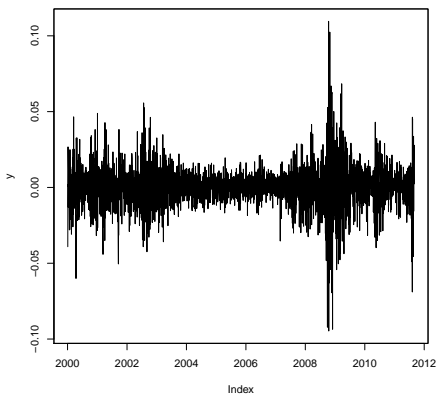
Zauważmy, że pik rozkładu zwrotów jest wyższy niż pik referencyjnego rozkładu normalnego. Jednocześnie jego ogony są grubsze niż rozkładu normalnego. Ponadto ramiona rozkładu empirycznego są niższe niż dla rozkładu normalnego. Dwie pierwsze obserwacje oznaczają, że jest więcej dni z mniejszą i większą zmiennością indeksu w rzeczywistości niż dla dopasowanego rozkładu normalnego.

Spróbujmy dopasować inny rozkład niż rozkład normalny.
Rysunki poniżej przedstawiają wykresy kwantylowe dla rozkładu normalnego vs rozkład empiryczny i analogiczne wykresy dla rozkładu z 3, 4 i 5 stopniami swobody (pokazanymi zgodnie z ruchem zegara)



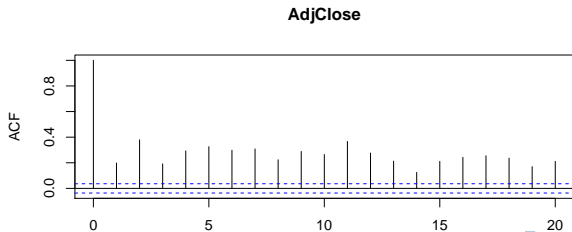
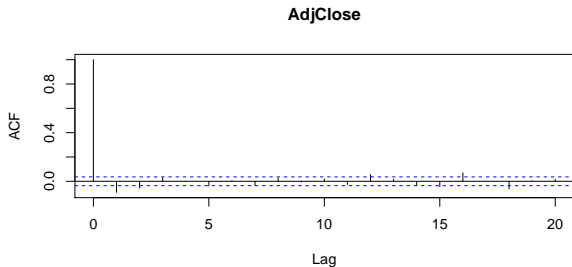
Najlepsze dopasowanie uzyskuje się dla rozkładu t z 4 st. swobody. ▶

Rysunek poniżej pokazuje zwroty dla Y będącego indeksem *S&P500* od początku wieku do 31 sierpnia 2012 roku (na podstawie cen zamknięcia). Skupiska zmienności są bardzo wyraźnie.



Drugi fakt stylizowany stwierdza, że zwroty są nieliniowo zależne. Tak dzieje się również w następującym przykładzie. Jeśli X jest symetryczną zmienną losową i rozpatrzmy parę zmiennych X i $Y = X^2$. Te zmienne są oczywiście zależne, ale nie liniowo zależne: $\rho(X, Y) = 0$, gdyż $\text{Cov}(X, Y) = EX \times X^2 - EXEX^2 = 0$. W takich sytuacjach współczynnik korelacji nie wykryje zależności między zmiennymi. Podobne zjawisko występuje dla zwrotów indeksów finansowych. Rozpatrzmy wykresy ACF dla $Y = S\&P500$ i jego kwadratu.

Wykres ACF sugeruje, że zwroty Y są nieskorelowane. Jednakże, powtarzająca się zależność jest widoczna na drugim rysunku.



Procesy ARCH(p)

Zdefiniujemy teraz klasę procesów ARCH (Autoregressive Conditionally Heteroscedastic) wprowadzonych przez Engle'a w 1982 dla modelowania inflacji w Wielkiej Brytanii.

Niech $p \in \mathbb{N}$. (X_t) jest procesem ARCH(p) jeśli

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2, \quad \alpha_0, \alpha_1, \dots, \alpha_p \geq 0,$$

gdzie Z_t jest silnym WN(0, 1) (ciąg niezależnych zmiennych losowych) i Z_t jest niezależne od X_s , $s < t$.

Zauważmy, że warunkując dostaniemy

$$EX_t = EE(\sigma_t Z_t | X_s, s < t) = E\sigma_t E(Z_t | X_s, s < t) = E\sigma_t E(Z_t) = 0$$

Podobnie uzyskujemy

$$\text{Var}(X_t | X_s, s < t) = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2.$$

Tak więc $\text{Var}(X_t | X_s, s < t)$ jest kombinacją afiniczną poprzednich wartości X_s^2 . Zatem duże wartości $X_{t-1}^2, \dots, X_{t-p}^2$ implikują dużą wartość $\text{Var}(X_t | X_s, s < t)$. Odpowiada to drugiemu faktowi stylizowanemu.

Własności procesu ARCH(p)

- Stacjonarny proces ARCH(p) istnieje gdy $\sum_{j=1}^p \alpha_j < 1$. Wtedy

$$EX_t = 0 \quad \text{Var}X_t = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j}.$$

Własności procesu ARCH(p) cd

- X_t jest słabym białym szumem (ciągami *nieskorelowanych* zmiennych losowych).
- Jeśli $EZ_t^4 < \infty$ to

$$X_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \varepsilon_t, \quad (*)$$

gdzie $\varepsilon_t = (Z_t^2 - 1)(\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2)$ jest słabym białym szumem.

- Kurtoza X_t jest nie mniejsza niż kurtoza Z_t :

$$\kappa_{X_t} \geq \kappa_{Z_t},$$

gdzie $\kappa_X = EX^4/\sigma_X^4$ (miara grubości ogona rozkładu).

Własność (*) jest ważna w identyfikacji: X_t^2 zachowuje się jak proces AR(p) !

Procesy GARCH(p, q)

Proces GARCH(p, q) : uogólniony ARCH(p)

$(X_t)_{t \in Z}$ jest procesem GARCH(p, q) (Generalized ARCH) jeśli

$$X_t = \sigma_t Z_t, \quad (1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (2)$$

gdzie, jak poprzednio Z_t jest niezależny od $X_s, s < t, p, q \in N, p \geq 1$
wszystkie współczynniki $\alpha_0, \alpha_i, \beta_j$ są nieujemne.

Procesy GARCH(p, q)

Własności procesu GARCH(p, q)

- Stacjonarny proces GARCH(p, q) istnieje gdy

$$\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1.$$

Wtedy

$$EX_t = 0 \quad \text{Var}X_t = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j}$$

$$\text{Var}(X_t | X_{t-s}, \sigma_{t-s}^2) = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

zależność wariancji warunkowej od σ_{t-j}^2 .

Własności procesu GARCH(p, q) cd

- Jeśli $EZ_t^4 < \infty$ to

$$X_t^2 = \alpha_0 + \sum_{i=1}^{p \vee q} (\alpha_i + \beta_j) X_{t-j}^2 + \varepsilon_t - \sum_{j=1}^q \beta_j \varepsilon_{t-j},$$

gdzie $\alpha_{p+j} = \beta_{q+j} = 0$, dla $j \geq 1$ i $p \vee q = \max(p, q)$.

Żeby udowodnić ostatnią równość, zauważmy, że

$$X_t^2 = \sigma_t^2 + X_t^2 - \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \varepsilon_t,$$

gdzie

$$\varepsilon_t = X_t^2 - \sigma_t^2$$

Tak więc podstawiając $\varepsilon_{t-j} = X_{t-j}^2 - \sigma_{t-j}^2$, otrzymujemy przedstawienie. Własność ważna w identyfikacji: kwadrat procesu GARCH(p, q) jest procesem ARMA(p, q) !.

Modelowanie procesów niestacjonarnych

Wiele procesów jest niestacjonarnych tzn. trajektorie wykazują istnienie trendu i/lub komponent sezonowych. Jak modelować takie procesy ?

Klasyczny addytywny model procesu niestacjonarnego

$$X_t = m_t + s_t + Y_t,$$

gdzie Y_t jest procesem stacjonarnym w szerszym sensie o średniej 0, m_t jest *deterministycznym* trendem i s_t jest *deterministycznym* komponentem sezonowym (tzn istnieje takie $d \in \mathbb{N}$, że $s_{t+d} = s_t$)

Jeśli $s_t \equiv 0$, to m_t może być estymowane metodami funkcji regresji, np przy użyciu średniej ruchomej lub estymatora lokalnie liniowego.

Jeśli komponent sezonowy jest niezerowy, to z reguły dokonuje się wstępnej estymacji nieparametrycznej trendu z długością okna d (przy założeniu, że $\sum_{i=1}^d s_i = 0$ i d jest znane), po odjęciu estymatora trendu estymuje się składową sezonową, a później parametryczny model dopasowany do $X_t - \hat{m}_t - \hat{s}_t$. (**procedura pięciostopniowa**).

Procedura Holta-Wintersa

Inne podejście: procedura Holta-Wintersa.

Opiszemy metodę najpierw dla sytuacji $s_t \equiv 0$. Zamiast estymacji tylko trendu m_t będziemy estymowali trend i jego zmianę

$$(m_t, b_t),$$

gdzie b_t oznacza zmianę trendu w momencie t . Procedura jest rekurencyjna. Niech $0 \leq \alpha, \beta \leq 1$ będą parametrami metody. Zdefiniujemy (równania H-W)

$$\begin{cases} \hat{m}_{n+1} &= (\hat{m}_n + \hat{b}_n)(1 - \alpha) + \alpha X_{n+1}, & 0 \leq \alpha \leq 1, \\ \hat{b}_{n+1} &= (\hat{m}_{n+1} - \hat{a}_n)\beta + (1 - \beta)\hat{b}_n, & 0 \leq \beta \leq 1, \end{cases}$$

i kładziemy $\hat{m}_2 = X_2, \hat{b}_2 = (X_2 - X_1)$,

W momencie $n + 1$ tend estymowany jest przez kombinację wypukła trendu w momencie $n + 1$ (tj. $m_n + b_n h$ dla $h = 1$) oraz wartości obserwacji X_{n+1} . Równania H-W są rozwiązywane rekurencyjnie dla $i = 3, \dots, n$.

W przypadku ogólnym chcemy estymować (m_t, b_t, s_t) , gdzie s_t ma okres d . Równania H-W mają postać

$$\begin{cases} \hat{m}_{n+1} &= (\hat{m}_n + \hat{b}_n)(1 - \alpha) + \alpha(X_{n+1} - \hat{s}_{n+1-d}), & 0 \leq \alpha \leq 1, \\ \hat{b}_{n+1} &= (\hat{m}_{n+1} - \hat{m}_n)\beta + (1 - \beta)\hat{b}_n, & 0 \leq \beta \leq 1, \\ \hat{c}_{n+1} &= (1 - \gamma)\hat{s}_{n+1-d} + \gamma(X_{n+1} - \hat{m}_{n+1}), & 0 \leq \gamma \leq 1. \end{cases}$$

W pierwszym równaniu wykorzystaliśmy fakt, że estymator składowej sezonowej s_{n+1} , który nie jest dostępny w momencie $n + 1$, może być zastąpiony przez estymator \hat{s}_{n+1-d} . Ponadto:

$$\hat{m}_{d+1} = X_{d+1},$$

$$\hat{b}_{d+1} = (X_{d+1} - X_1)/d,$$

$$\hat{s}_i = Y_i - (Y_1 + \hat{b}_{d+1}(i - 1)), \quad i = 1, 2, \dots, d.$$

Standardowa metoda wyboru α , β i γ to:

$$(\alpha_0, \beta_0) := \arg \min_{\alpha, \beta, \gamma} \sum_{i=d+1}^n (X_i - \hat{X}_i(\alpha, \beta, \gamma))^2.$$

Estymatory Holta -Wintersa dają prognozę X_{n+h} , mianowicie

$$\hat{X}_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h-d}]$$

Wygładzanie wykładnicze

W przypadku gdy $\beta = \gamma = 0$ i tylko $\alpha \neq 0$ mamy:

$$\hat{m}_{n+1} = \hat{m}_n(1 - \alpha) + \alpha X_{n+1}$$

Rozwiązanie explicite (**wygładzanie wykładnicze**)

$$\hat{m}_{t+1} = \sum_{j=0}^{t-1} \alpha(1 - \alpha)^j X_{t+1-j} + (1 - \alpha)^t X_1.$$

Zauważmy, że dla estymacji m_{t+1} wpływ X_{t+1-j} dany przez współczynnik $\alpha(1 - \alpha)^j$ maleje wykładniczo, stąd nazwa metody (często wprowadzana niezależnie od metody Holta-Wintersa).