

Komputerowa analiza szeregów czasowych, Raport II

Katarzyna Macioszek

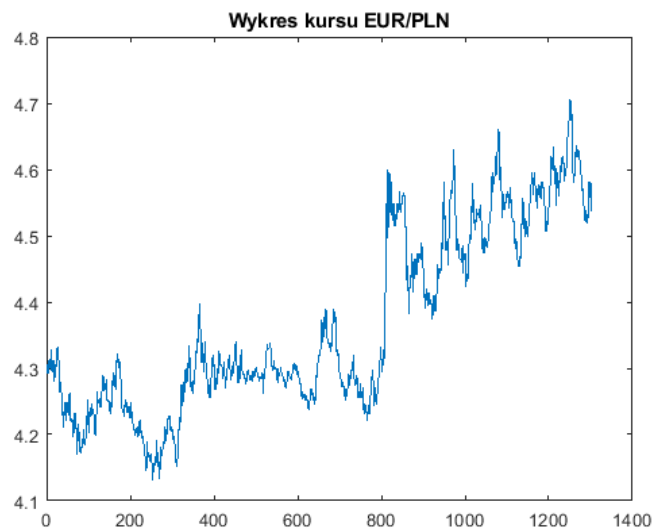
7 lutego 2022

1 Opis wybranych danych

Wybrane dane pochodzą ze strony <https://finance.yahoo.com> i dotyczą kursu waluty Euro w porównaniu do złotówki. Dane pochodzą z okresu 06.02.2017 - 04.02.2022 i zawierają 1305 obserwacji, po usunięciu brakujących obserwacji otrzymujemy próbkę o długości 1303. W raporcie wykorzystujemy cenę open, czyli cenę otwarcia.

1.1 Analiza danych

Na początku tworzymy wykres danych oraz obliczamy podstawowe statystyki.



Rysunek 1: Wykres danych

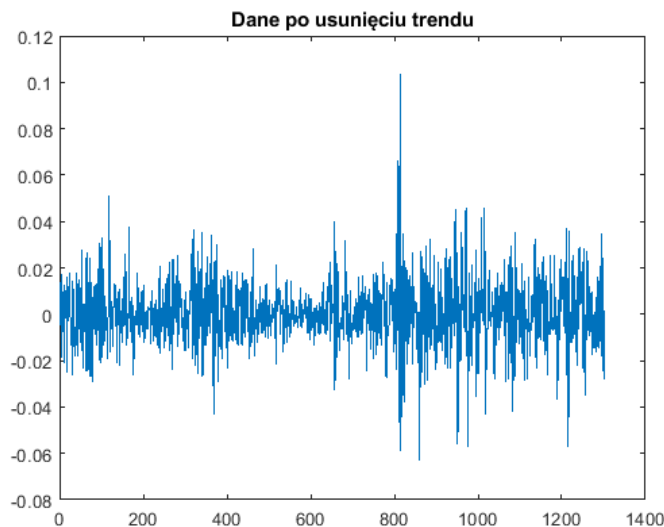
- Średnia: $\bar{X} = \sum_{i=1}^n X_i = 4,3662$
- Wariancja: $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = 0,0191$
- Odchylenie standardowe: $S = \sqrt{S^2} = 0,1384$
- Kwartyle: $q_1 = 4,2627$, $q_2 = 4,3052$, $q_3 = 4,4994$
- Rozstęp międzykwartyłowy: $IQR = q_3 - q_1 = 0,2367$
- Rozstęp z próby: $\max X - \min X = 0,5737$

Widzimy, że ceny euro rosną z tendencją liniową oraz obserwujemy niewielkie wahania.

2 Przygotowani danych

Jako że wybrane dane posiadają trend liniowy. Do jego usunięcia wykorzystujemy metodę różnicowania danych. W tym celu, w następujący sposób tworzymy wektor Y :

$$Y_i = X_i - X_{i-1}$$



Rysunek 2: Wykres danych po zastosowaniu jednokrotnego różnicowania

Podstawowe statystyki dla przetransformowanych danych prezentują się następująco:

- Średnia: $\bar{X} = \sum_{i=1}^n X_i = 1,8629 \cdot 10^{-4}$
- Wariancja: $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = 2,0819 \cdot 10^{-4}$
- Odchylenie standardowe: $S = \sqrt{S^2} = 0,0144$
- Kwartyle: $q_1 = -0,0080$, $q_2 = -2,3 \cdot 10^{-4}$, $q_3 = 0,0076$
- Rozstęp międzykwartylowy: $IQR = q_3 - q_1 = 0,0155$
- Rozstęp z próby: $\max X - \min X = 0,1661$

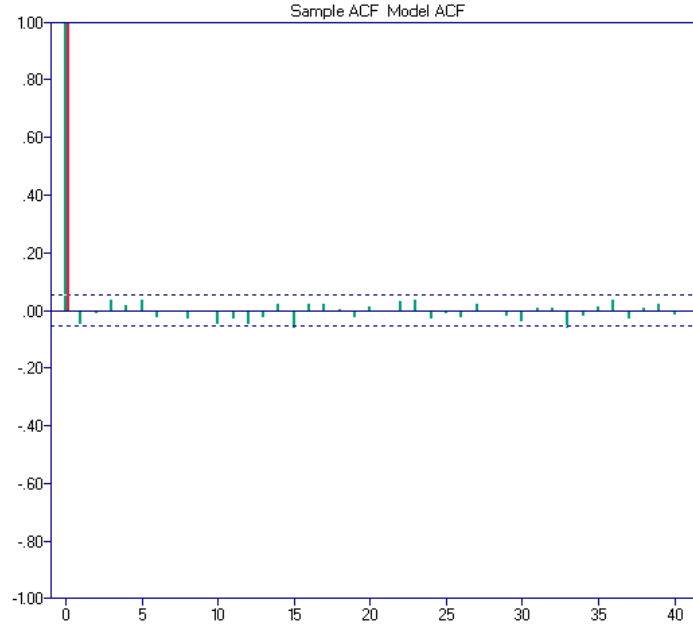
Po usunięciu trendu dane oscylują wokół zera, a ich średnia μ jest w przybliżeniu równa zero.

Następnie sprawdzamy empiryczne funkcje ACF i PACF dla naszych danych.

- ACF:

$$\rho(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{1}{n} \sum_{i=1}^{n-|h|} (x_{i+|h|} - \bar{x})(x_i - \bar{x})$$

gdzie: n - długość próby, $\hat{\gamma}(h)$ - empiryczna funkcja autokowariancji.

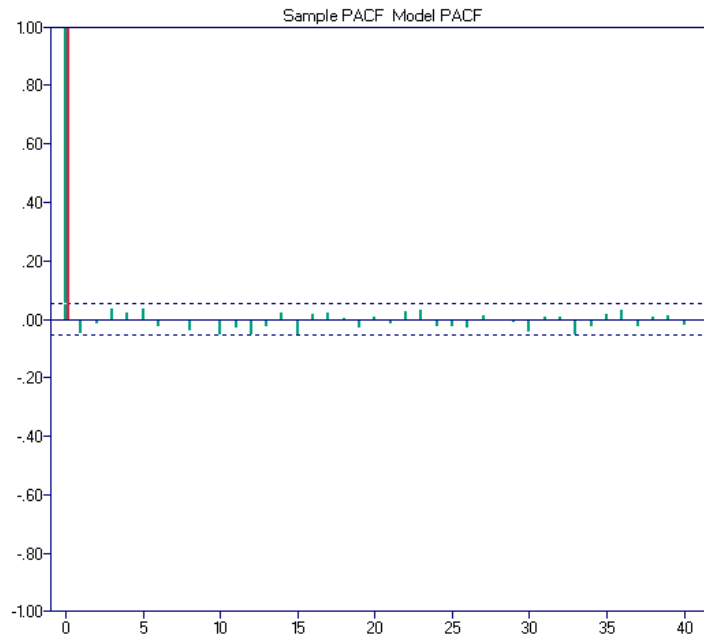


Rysunek 3: ACF dla danych po usunięciu trendu

- PACF:

$$\alpha(h) = \phi_{hh}$$

gdzie: ϕ_{hh} to ostatni element wektora $\phi_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h$, $\hat{\Gamma}_h = [\hat{\gamma}(i-j)]_{i,j=1}^h$



Rysunek 4: PACF dla danych po usunięciu trendu

Widzimy, że zarówno wartości funkcji ACF jak i PACF są bliskie zera dla wszystkich argumentów poza 0 oraz nie są zależne od czasu. Ta obserwacja oraz średnia przetransformowanych danych pokazują, że otrzymaliśmy szereg stacjonarny w słabym sensie, zatem możemy przejść do doboru modelu ARMA(p, q).

3 Dobranie modelu

3.1 Definicja

Szereg X_t jest szeregiem ARMA(p, q) jeśli jest stacjonarny w słabym sensie oraz spełnia następujące równanie:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

gdzie: $Z_t \sim WN(0, \sigma^2)$ oraz wielomiany $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ nie mają wspólnych pierwiastków.

3.2 Model ARMA(p, q)

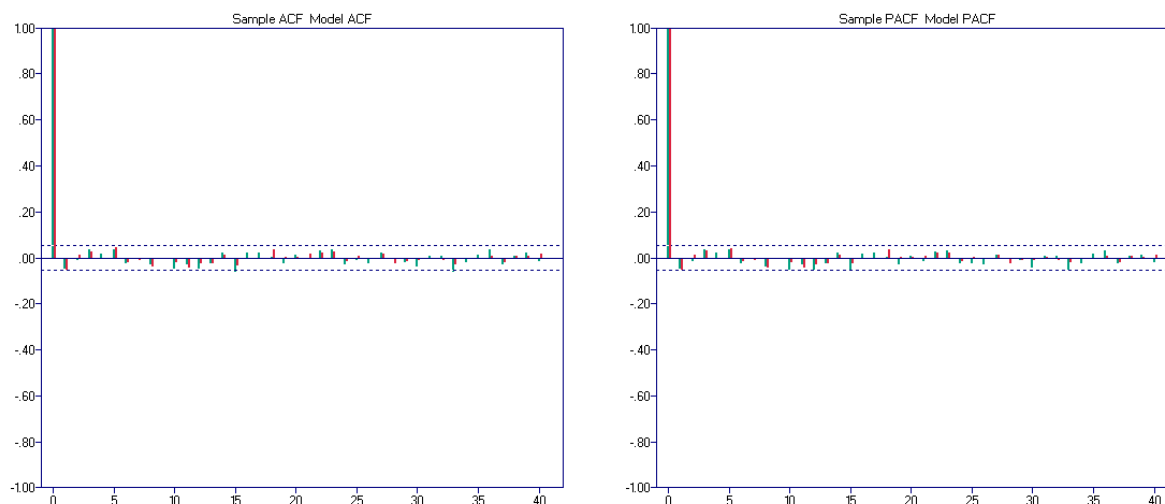
Model ARMA wybieramy wykorzystując kryterium informacyjne Akaikiego (AIC), które mówi nam jak bardzo wartości dobranego modelu różnią się od rzeczywistych. Do doboru modelu ARMA(p, q) wykorzystujemy program *ITSM*, który sprawdza kombinacje parametrów p i q , a następnie wybiera ten, dla którego wartość AIC jest najmniejsza, zatem ten, dla którego model najmniej różni się od danych rzeczywistych.

Najlepszym modelem do estymacji naszych danych jest model ARMA(6, 7) z parametrami:

ϕ_1	0,178134	θ_1	0,178134
ϕ_2	0,569468	θ_2	0,569468
ϕ_3	-0,327550	θ_3	-0,327550
ϕ_4	0,628713	θ_4	0,628713
ϕ_5	0,212080	θ_5	0,212080
ϕ_6	-0,883555	θ_6	-0,883555
		θ_7	-0,883555

Wariancja białego szumu czyli szeregu Z_t wynosi $\sigma^2 = 0,000202$

Aby sprawdzić dopasowanie modelu, tworzymy wykresy porównujące empiryczne i teoretyczne wartości funkcji ACF i PACF.

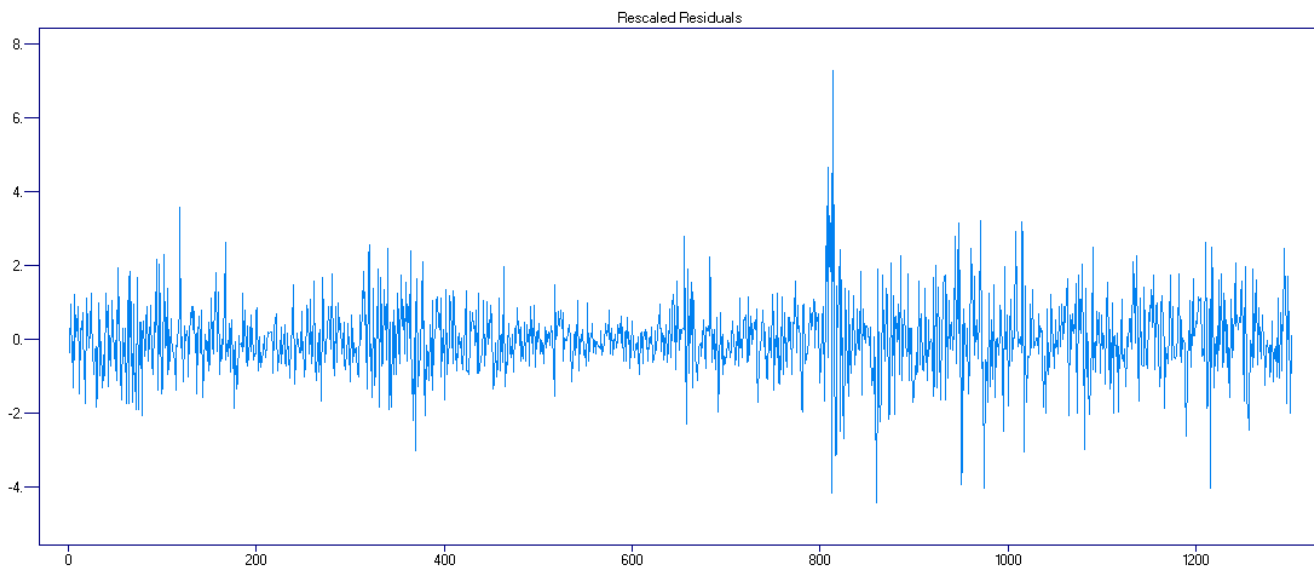


Rysunek 5: Porównanie ACF i PACF dla dobranego modelu

Widzimy, że empiryczne wartości zarówno funkcji ACF jak i PACF są bardzo zbliżone do teoretycznych. Możemy więc wnioskować, że przedstawiony wcześniej model jest dobrze dobrany.

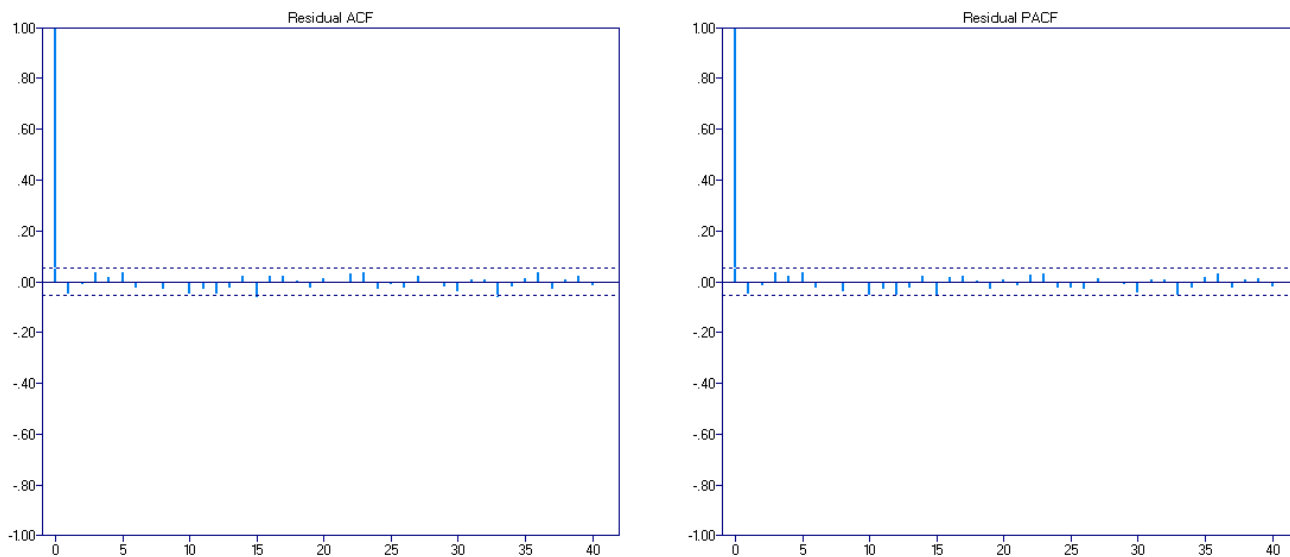
4 Analiza residuów

Pierwszym krokiem w analizie residuów jest wygenerowanie wykresu



Rysunek 6: Wykres residuów

Z wykresu widzimy, że residua oscylują wokół wartości zero (średnia: $\mu = -6,72 \cdot 10^{-19}$), jednak wariancja zdecydowanie nie jest stała. Następnie sprawdzamy, czy residua są skorelowane. W tym celu tworzymy wykresy ACF i PACF dla residuów.



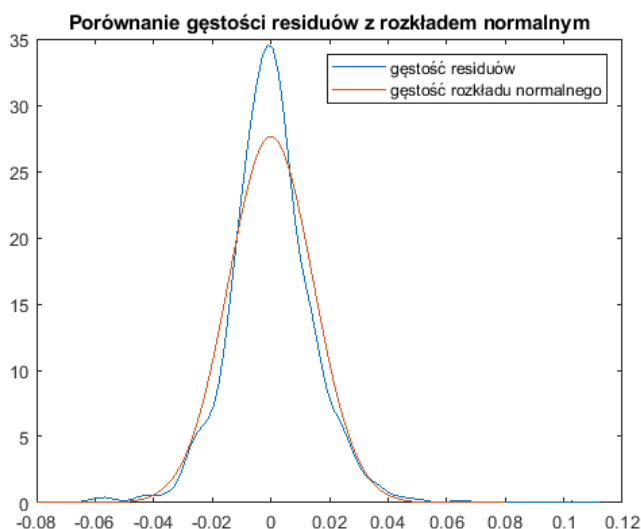
Rysunek 7: ACF i PACF dla residuów

Wartości obu funkcji oscylują wokół zera dla argumentów różnych od 0. Możemy zatem wywnioskować, że residua są nieskorelowane, jednak warto sprawdzić to wykorzystując również testy losowości. Przyjmujemy hipotezę zerową, że residua nie są skorelowane i poziom istotności $\alpha = 0,05$. P-wartości dla kolejnych testów wynoszą:

- Ljung - Box: 0,0013
- McLeod - Li: 0
- Turning Points: 0,72574
- Diff sign points: 0,11332
- Rank test: 0,64107

Spośród wybranych testów trzy prowadzą do przyjęcia hipotezy zerowej, jednak ze względu na inny wynik pozostałych dwóch testów, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej, która prowadzi nas do wniosku, że istnieje korelacja pomiędzy residuami.

Następnie sprawdzamy, czy residua pochodzą z rozkładu normalnego. W tym celu tworzymy wykres porównujący gęstość residuów do gęstości rozkładu normalnego: $N(0, \sigma_1^2)$, gdzie σ_1^2 to empirycznie wyznaczona wariancja residuów.



Rysunek 8: Porównanie gęstości residuów z gęstością rozkładu normalnego

Widzimy, że kształt obu gęstości jest podobny, jednak można zaobserwować sporą różnicę w wartościach. Możemy wywnioskować, że residua nie mają rozkładu normalnego, jednak aby to potwierdzić wykonujemy test Jarque-Bera z hipotezą zerową: H_0 : residua mają rozkład normalny. Otrzymujemy $p\text{-value} = 0$. Zatem zgodnie z tym, co odczytaliśmy z wykresu, odrzucamy hipotezę o normalności residuów.

5 Podsumowanie

Z wykresu kursu euro łatwo zauważyć, że ceny te rosły z tendencją liniową. Po wykorzystaniu jednokrotnego różnicowania otrzymaliśmy szereg stacjonarny w słabym sensie. Następnie za pomocą programu *ITSM* wyznaczyliśmy model ARMA(6, 7), przy czym wartość kryterium AICC jest równa $-0.7376 \cdot 10^4$. Ostatnim krokiem była analiza residuów, które okazały się skorelowane oraz nie pochodzą z rozkładu normalnego.