Raport 2

Autor: Kostiantyn Skopych, 255916

Przedstawienie danych

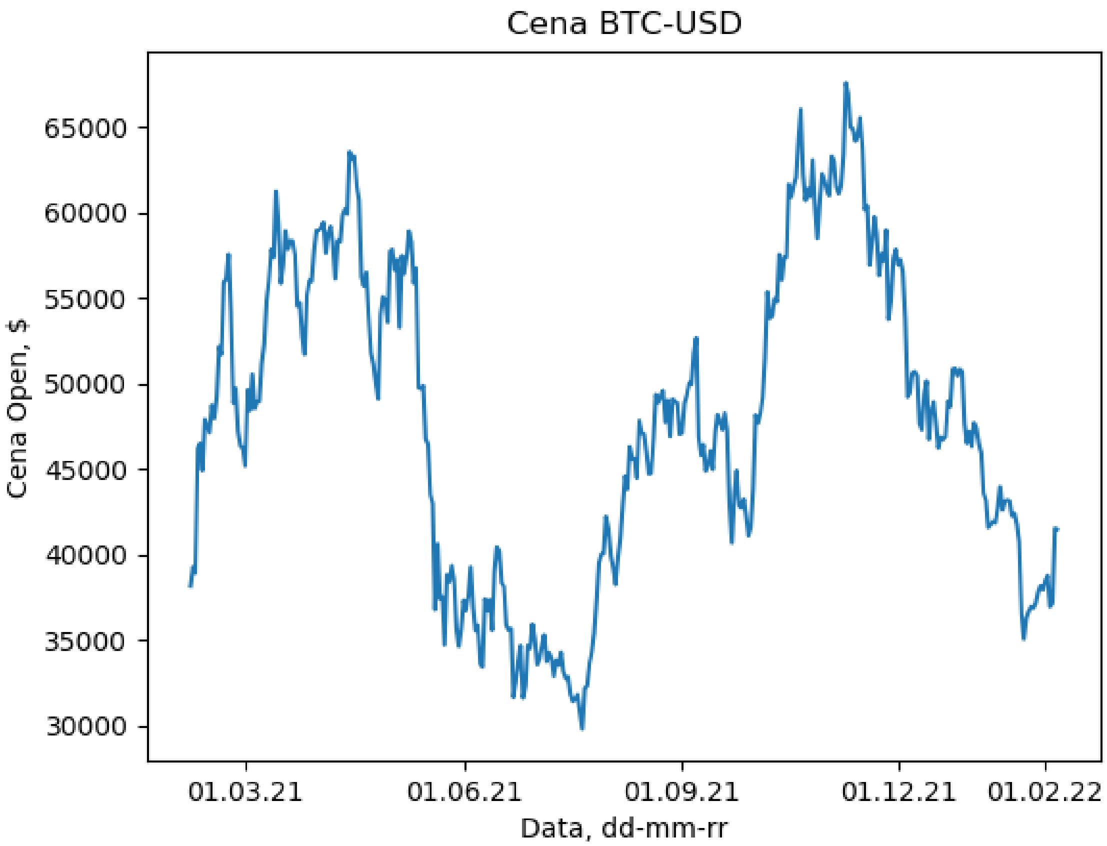
Opis: Codzienne dane ceny Open pary walutowej BTC-USD

Okres: 06.02.2021 – 06.02.2022

Źródło: https://finance.yahoo.com/quote/BTC-USD/history?p=BTC-USD

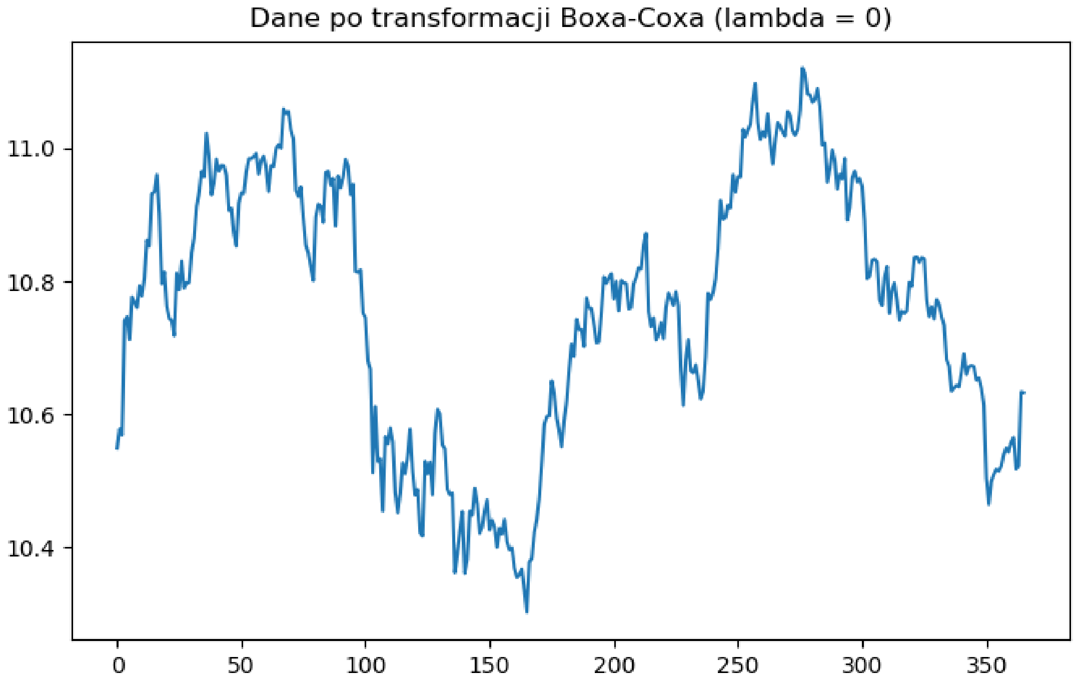
Liczba obserwacji: 366

Wykres danych:

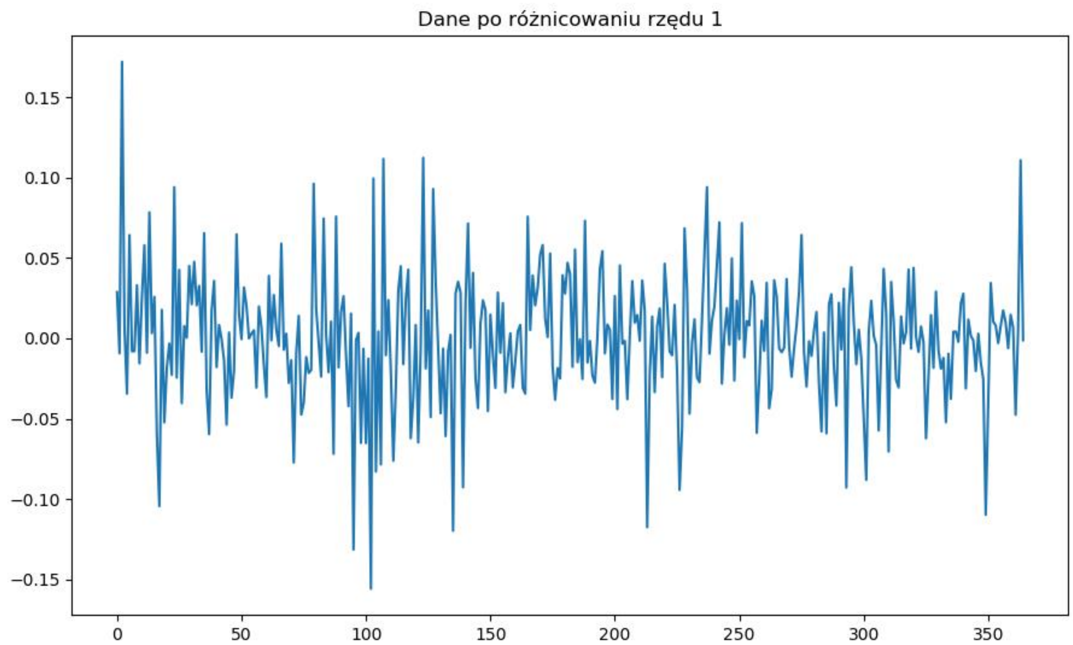


Przygotowanie danych

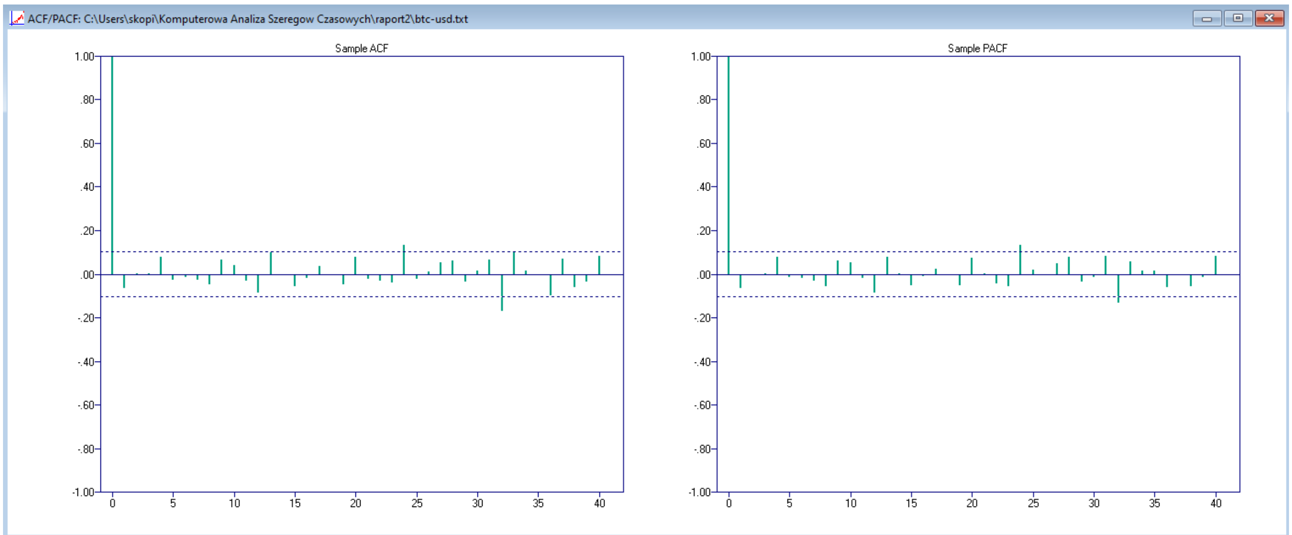
Używamy transformacji Boxa-Coxa z parametrem $\lambda = 0$ (czyli log(x)) dla stabilizacji wariancji. Po transformacji dane wyglądają jak na wykresie poniżej.



Następnie stosujemy różnicowanie rzędu 1, żeby pozbyć się trendu. Patrząc na wykres danych zakładamy, że nie ma tutaj żadnej sezonowości, dlatego nie stosujemy żadnej metody do pozbycia się jej. Dane po różnicowaniu wyglądają następująco:



Na oko widzimy, że ten szereg raczej jest stacjonarny. Potwierdzają to „mniej więcej" stałe średnia (blisko 0) oraz wariancja (blisko 0.0015). Nie widzimy tutaj żadnego trendu lub sezonowości. Zobaczmy wykresy ACF i PACF. Widać, że natychmiast zbiegają w okolice zera. Stąd wnioskujemy, że szereg jest stacjonarny.



Zaproponowanie modelu ARMA(p,q) oraz estymacja parametrów

Dla doboru modelu ARMA zarówno jak i dla estymacji jego parametrów używamy wbudowanego w ITSMa narzędzia. Dobór modelu odbywa się na podstawie kryterium AICC. Parametry są estymowane metodą największej wiarygodności. Wynik jest przedstawiony na poniższym zdjęciu.

ML estimates: C:\Users\skopi\Komputerowa Analiza Szeregow Czasowych\raport2\btc-usd.txt

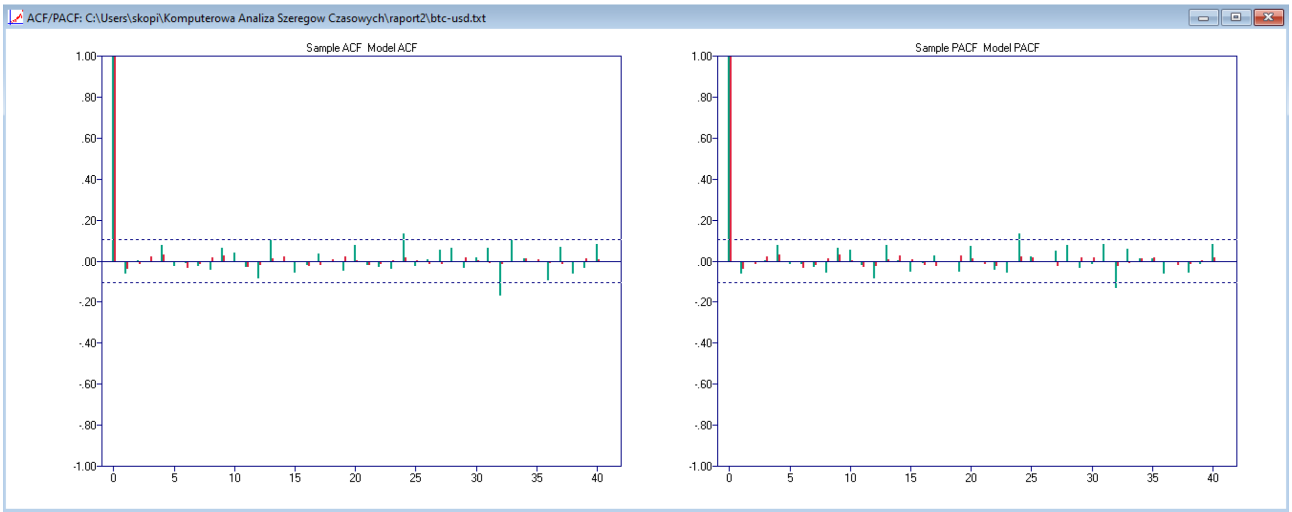
```
========================================
ITSM::(Maximum likelihood estimates)
========================================

Method: Maximum Likelihood

ARMA Model:
X(t) = .6478 X(t-1) - .9510 X(t-2)
     + Z(t) - .7011 Z(t-1) + .9999 Z(t-2)

WN Variance = .001543

AR Coefficients
       .647804       -.951043

Standard Error of AR Coefficients
       .016225        .016225

MA Coefficients
      -.701092        .999879

Standard Error of MA Coefficients
       .000817        .000817

(Residual SS)/N = .00154304

AICC = -.131178E+04
BIC  = -.131777E+04

-2Log(Likelihood) = -.132194E+04

Accuracy parameter = .100000E-08

Number of iterations = 1

Number of function evaluations = 58482

Uncertain minimum.
```

Czyli naszym dopasowanym modelem jest ARMA (2, 2). $\sigma^2 = 0.001543$ (wariancja białego szumu). Parametry AR: $\varphi_1 = 0.647804$, $\varphi_2 = -0.951043$. Parametry MA: $\theta_1 = -0.701092$, $\theta_2 = 0.999879$.

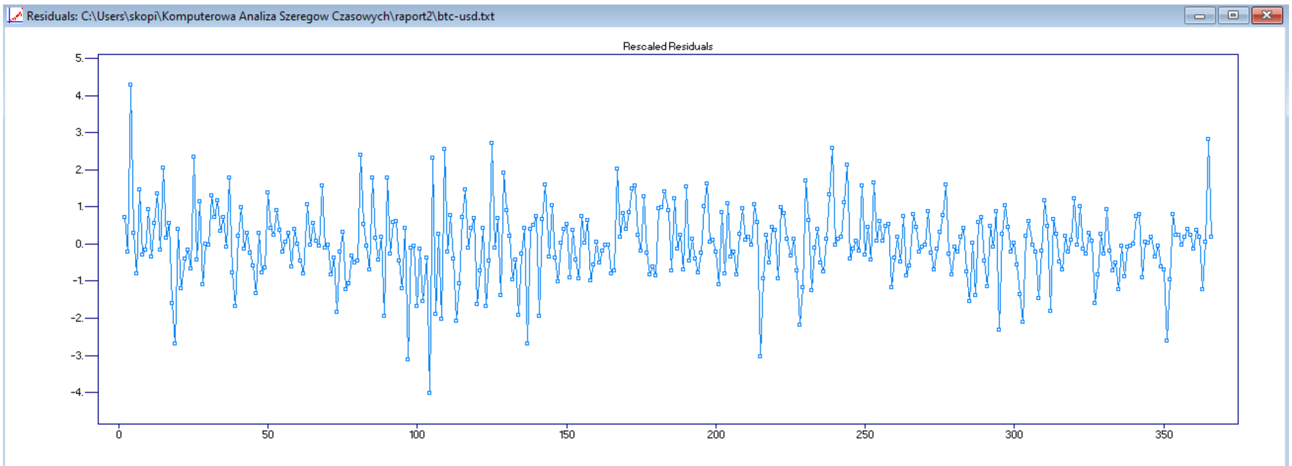
Zwróćmy uwagę na to, że znaki parametrów zależą od tego, w jakiej postaci przedstawimy równanie ARMA. Na przykład ITSM zostawia po lewej stronie tylko X(t), a resztę przerzuca na prawą stronę.

Sprawdzenie dopasowania modelu do danych

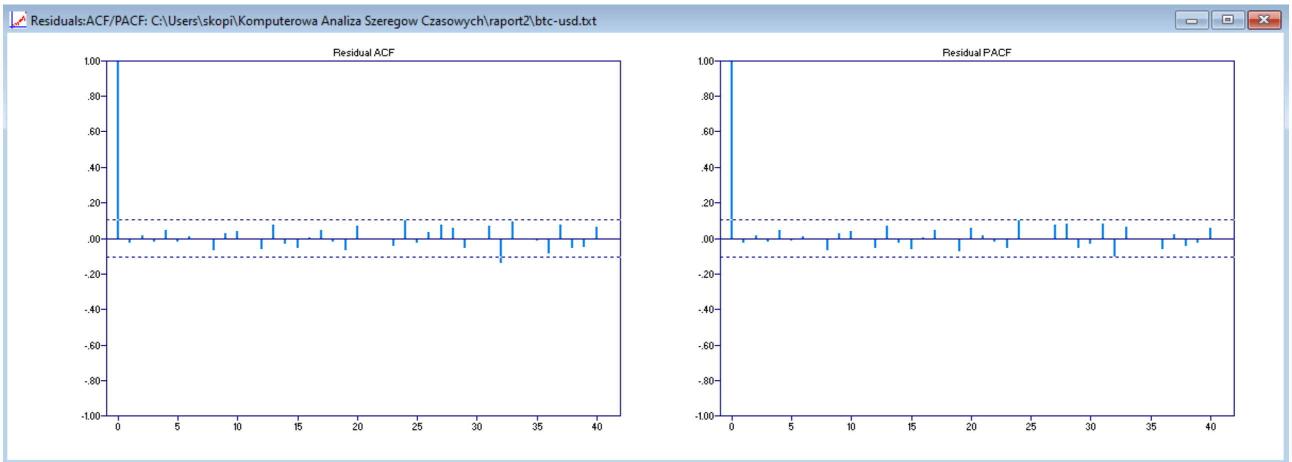
Na poniższym wykresie porównujemy empiryczne ACF i PACF z teoretycznymi dla danego modelu. Wyniki tego porównania będą omówione później we wnioskach.



Następnie przeanalizujemy residua. Poniżej znajduje się wykres residuów w czasie.



Wykresy ACF i PACF residuów są przedstawione poniżej.

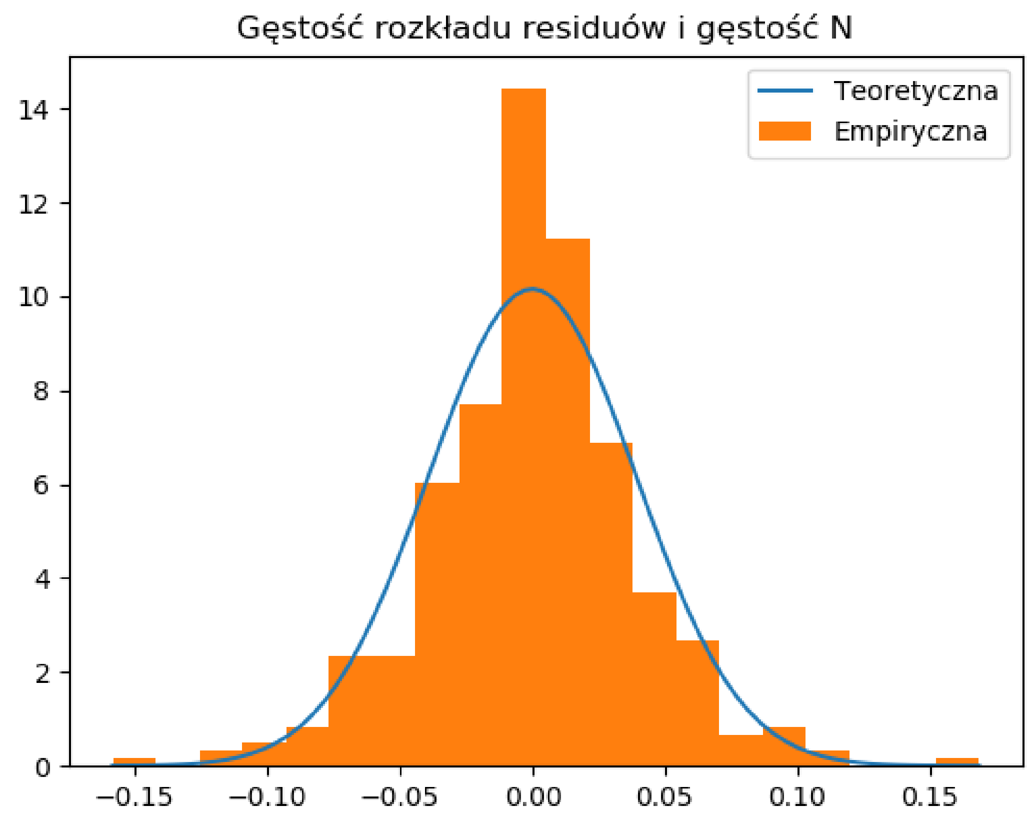


Test losowości residuów oraz test Jarque-Bera na normalność rozkładu.

```
Tests of randomness: C:\Users\skopi\Komputerowa Analiza Szeregow Czasowych\raport2\btc-usd.txt
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 20.618 Chi-Square ( 20 ), p-value = .41993

McLeod - Li statistic = 42.784 Chi-Square ( 24 ), p-value = .01053

# Turning points = .24800E+03~AN(.24200E+03,sd = 8.0353), p-value = .45524

# Diff sign points = .18800E+03~AN(.18200E+03,sd = 5.5227), p-value = .27729

Rank test statistic = .32592E+05~AN(.33215E+05,sd = .11646E+04), p-value = .59268

Jarque-Bera test statistic (for normality) = 51.257 Chi-Square (2), p-value = .00000

Order of Min AICC YW Model for Residuals = 0
```

Wartość średnia residuów = 1.477e-05. Wariancja residuów = 0.00154. Residua układają się wokół zera. Sprawdzając wariancję na różnych przedziałach widzimy pewną różnicę, jednak nie jest ona duża, więc możemy powiedzieć, że wariancja jest „w przybliżeniu” stała.

Na poniższym wykresie widzimy histogram gęstości rozkładu residuów oraz gęstość rozkładu normalnego o średniej i wariancji takich, jakie mają residua.



Boxplot residuów jest przedstawiony poniżej.

Wnioski

Rozpatrywaliśmy dane dotyczące ceny pary walutowej BTC-USD w ciągu ostatniego roku. Potraktowaliśmy te dane jako szereg czasowy i dopasowaliśmy do niego model ARMA (2, 2). Następnie sprawdziliśmy, czy ten model został dobrze dopasowany. Przeanalizujemy teraz, co z tego wyszło.

Przed dopasowaniem modelu ARMA postaraliśmy się doprowadzić nasz szereg do postaci stacjonarnej. Użyliśmy do tego transformacji Boxa-Coxa, która stabilizuje wariancję oraz różnicowania danych dla usunięcia trendu i sezonowości. Dostaliśmy szereg o średniej i wariancji bliskich zeru. Wydają się one być „mniej więcej" stałymi w czasie. Jednak dla stacjonarności potrzebujemy, żeby ACF i PACF szybko zbiegały do zera. Tutaj mieliśmy takie zachowanie, z taką tylko uwagą, że chociaż te dwie funkcje krążyły wokół zera w przedziale „statystycznie zaniedbywalnym", czasem nieco wychodziły z tego przedziału. Te wyjątki były naprawdę małe. Dlatego możemy powiedzieć ogólnie, że dane zostały przetransformowane do postaci stacjonarnej (tak się wydaje też patrząc na ich wykres). Jednak musimy rozumieć, że ta transformacja nie była idealna. Wynikało to z tego, że użyliśmy różnicowania rzędu 1, które dobrze sobie radzi z trendem liniowem oraz założyliśmy, że nie mamy do czynienia z żadną sezonowością. Tak naprawdę mogliśmy tutaj nie zauważyć jakiegoś trendu innej postaci, niż liniowy lub jakiejś słabo rozróżnialnej sezonowości. Zaznaczmy jednak, że były sprawdzone różnicowania różnych rzędów m, które dobrze sobie radzą z sezonowością o okresie m. Okazało się, że z punktu widzenia funkcji ACF i PACF różnicowanie rzędu 1 było najlepszym wyborem.

Następnym krokiem było dopasowanie modelu ARMA. Skorzystaliśmy tutaj z ITSMa. On zrobił dopasowanie na podstawie kryterium informacyjnego AICC. Został dopasowany model ARMA (2, 2). Dale ITSM wyestymował parametry tego modelu na podstawie metody największej wiarygodności. Sprawdzenie naszego dopasowania zaczęliśmy od porównania funkcji ACF i PACF. Teoretyczne i empiryczne funkcje pokrywają się, powiedzmy, słabo. Jednak to nie ma znaczenia, ponieważ w obu przypadkach znajdują się one od razu w przydziałach „statystycznie zaniedbywalnych", czyli bliskich zeru. Dlatego wnioskujemy tutaj, że dobrze dobraliśmy model.

Ważną częścią pracy była analiza residuów. Obserwujemy tutaj kilka ciekawych rzeczy. Najpierw zauważmy, że residua układają się wokół zera, co widać z boxplotu, średniej, oraz wykresu residuów w czasie. Wariancja jest „mniej więcej" stała. Wykresy ACF i PACF układają się w odpowiednich przedziałach blisko zera, co świadczy o tym, że residua są nieskorelowane. Jednak przechodząc do testu losowości mamy pewne „dziwne" wyniki. Po pierwsze zawodzi nam jeden z testów, pokazując p-wartość = 0.01053, co jest mniejsze od 0.05. Pozostałe testy dają dobre wyniki. Dalej widzimy wynik testu Jarque-Bera: p-wartość = 0. To miałoby świadczyć o tym, że residua nie pochodzą z rozkładu normalnego. Na dodatek to samo nam mówi test Kołmogorowa-Smirnowa. Chcemy więc zobaczyć histogram residuów w porównaniu do gęstości rozkładu normalnego. Widzimy tutaj, że wyglądają dość „podobno" i na początku nawet pokrywają się. Możliwe, że te residua nie pochodzą z rozkładu normalnego, tylko z jakiegoś rozkładu zbliżającego się ku normalnemu. Moim zdaniem musimy być zadowoleni z takiego rozkładu residuów. Większość testów na losowość też dają pozytywny wynik.

Podsumowując, chociaż analiza residuów może naprowadzać nas na pewne wątpliwości, myślę, że możemy stwierdzić, że model ARMA (2, 2) został dobrze dopasowany.

Użyte definicje i wzory

Model ARMA (p, q) jest modelem szeregu stacjonarnego autoregresyjnego o średniej ruchomej, który zadany jest równaniem ARMA.

Równanie ARMA: $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$,

gdzie X_t to szereg stacjonarny w słabym sensie, Z_t to biały szum.

Szereg stacjonarny w słabym sensie to szereg czasowy, dla którego wartość średnia oraz funkcja autokorelacji nie zależą od czasu.

ACF – funkcja autokorelacji, która pokazuje jak dane w szeregu czasowym zależą między sobą. Mówimy wtedy o danych skorelowanych lub nieskorelowanych.

Funkcja autokorelacji cząstkowej PACF jest podobna do ACF, z tym wyjątkiem, że pokazuje tylko korelację między dwiema obserwacjami, której nie wyjaśniają krótsze opóźnienia między tymi obserwacjami.

Kryterium informacyjne (AIC, BIC, AICC) to kryterium wyboru pomiędzy modelami statystycznymi. Używany jest w następujący sposób: liczymy kryterium dla wszystkich możliwych modeli, a następnie wybieramy ten o najmniejszej wartości kryterium jako najlepsze.