Analiza danych rzeczywistych przy pomocy modelu ARMA

Autorzy

Kacper Budnik, 262286
Maciej Karczewski, 262282

Wydział Matematyki 8 lutego 2023

Spis treści

1 Wprowadzenie

2 Przygotowanie danych do analizy

3 Dekompozycja szeregu czasowego

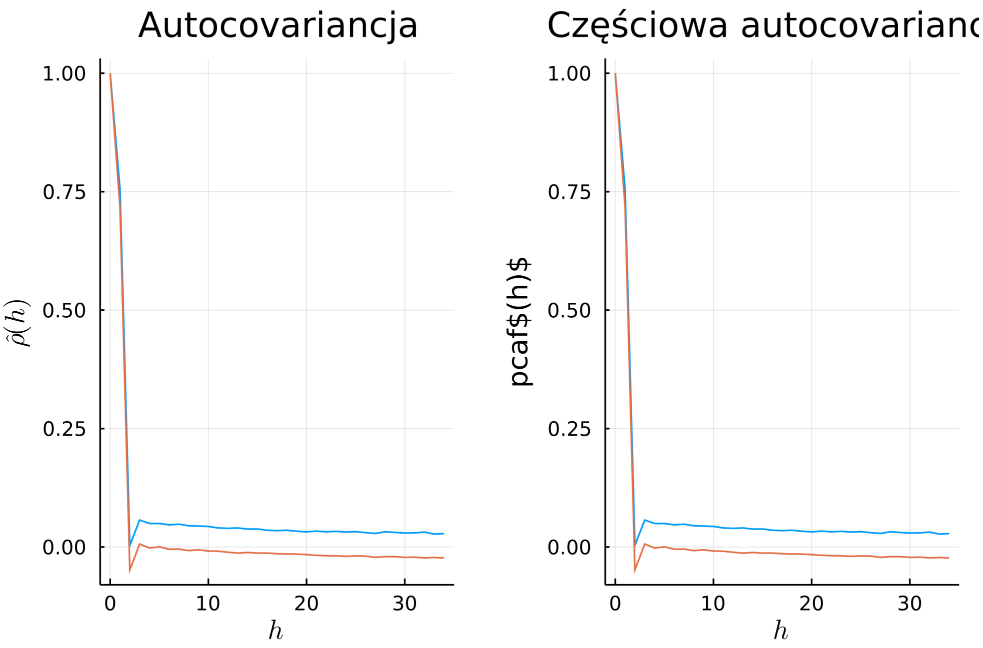
3.1 Wykresy dla surowych danych

3.2 Transformacja Boxa-Coxa

3.3 Różnicowanie sezonowe

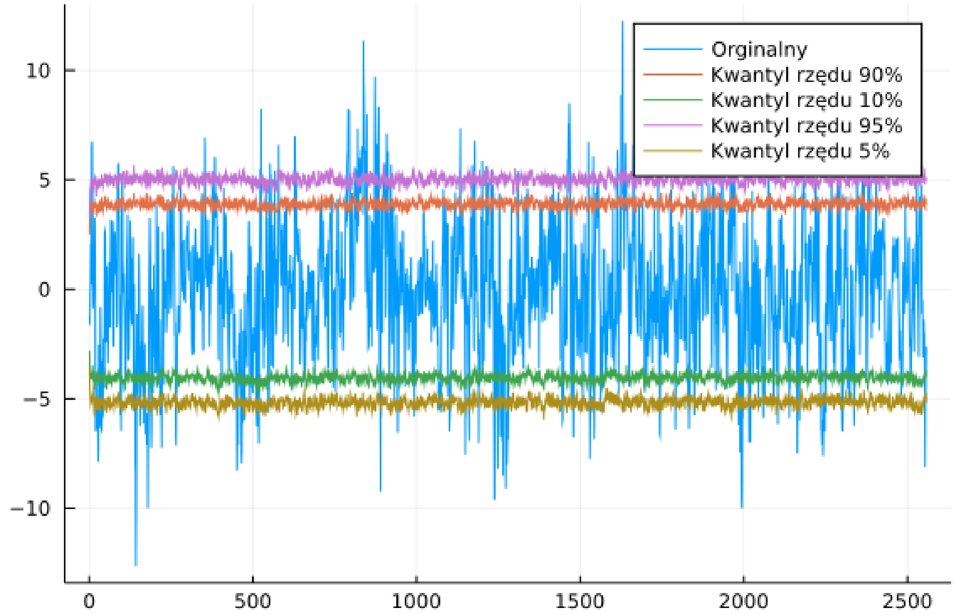
4 Ocena dopasowania modelu

W celu oceny dopasowania modelu do danych zobaczymy jak wyglądają przedziały ufności na poziomie $\alpha = 5\%$ dla ACF i PACF



Rysunek 1: Przedziały ufności dla ACF i PACF dla naszego modelu

Jak możemy zobaczyć na wykresie 1 istnieje tylko okresowa zależność pomiędzy danymi w modelu. Następnie porównamy trajektorię z liniami kwantylowymi dla naszego modelu



Rysunek 2: Porównanie orginalnych danych z liniami kwantylowymi modelu

Analizując linie kwantylowe naszego modelu pokazane na wykresie 2 możemy zauważyć, że model w miarę dobrze się sprawdza dla naszych danych.

Patrząc na wykres zależności w modelu 1 oraz na wykres z liniamy kwantylowymi 2 dochodzimy do wniosku, że model dobrze się sprawdza dla naszych danych.

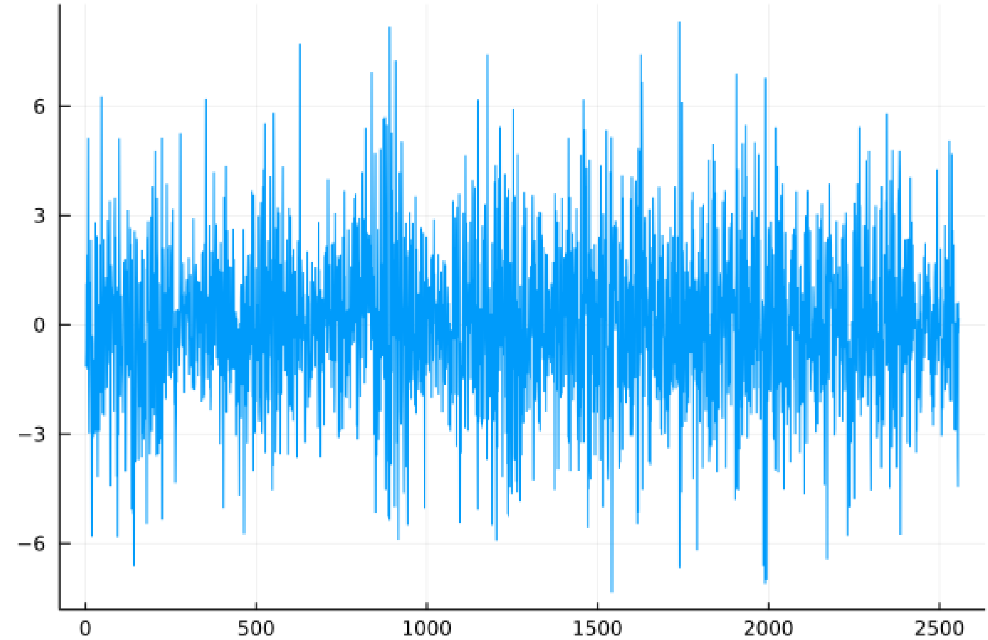
5 Analiza szumu

Podczas tworzenia modelu ARMA zakładaliśmy następujące warunki odnośnie szumu

1. $\mathbb{E}\xi_i = 0 \, \forall i$,
2. $Var\xi_i = \sigma^2 < \infty \quad \forall i,$
3. $\xi_i \perp\!\!\!\perp \xi_j$ dla $i \neq j,$

5.1 Stała średnia równa 0

Sprawdzimy czy rezidua naszego modelu mają stałą średnią równą 0. W tym celu zobaczymy jak wyglądają nasze wartości resztowe.



Rysunek 3: Residua naszego modelu

Jak możemy zobaczyć na 3 średnia jest stała w czasie i wynosi w przybliżeniu 0. W celu upewnienia się wykonamy też t test dla jednej zmiennej.

```
One sample t-test
-----------------
Population details:
    parameter of interest:   Mean
    value under h_0:         0
    point estimate:          0.000791916
    95% confidence interval: (-0.08009, 0.08167)

Test summary:
    outcome with 95% confidence: fail to reject h_0
    two-sided p-value:           0.9847

Details:
    number of observations:   2557
    t-statistic:              0.019200134473658807
    degrees of freedom:       2556
    empirical standard error: 0.04124531616232871
```

Rysunek 4: Test t dla naszych wartości resztowych

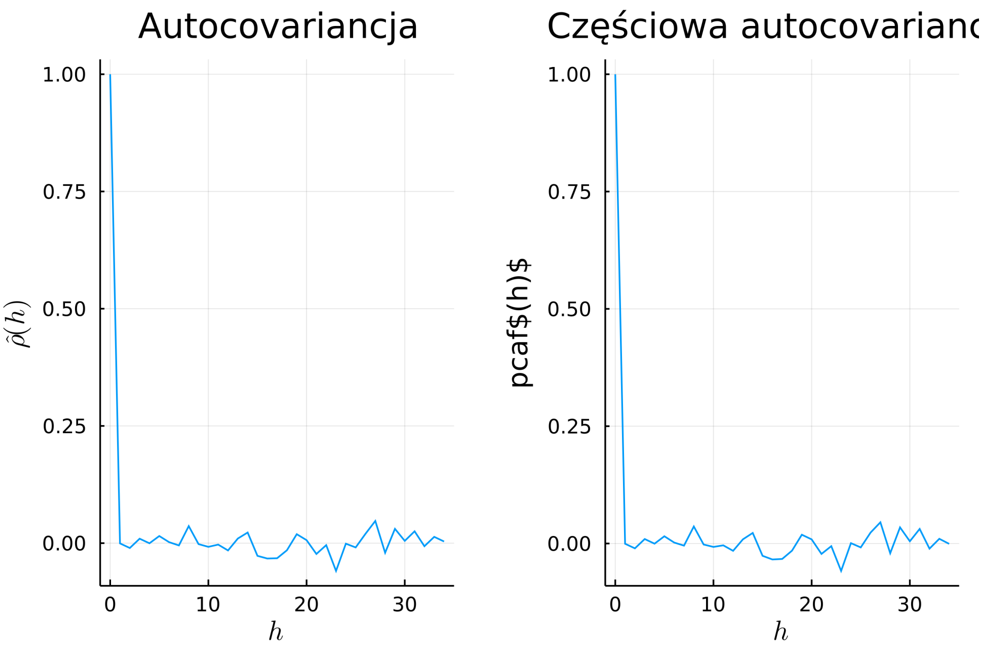
Jak możemy zobaczyć test t 4 nie miał podstaw do odrzucenia hipotezy zerowej, która mówiła ,że średnia wynosi 0. Test zwrócił p-value równe 0.9847. Tak więc możemy przyjąć, że średnia jest stała i równa 0

5.2 Stała wariancja

Z wykresu 3 możemy odczytać, że wariancja jest stała.Dodatkowo możemy obliczyć ,że należy ona do przedziału $[4.141; 4.554]$ na poziomie ufności $\alpha = 5\%$

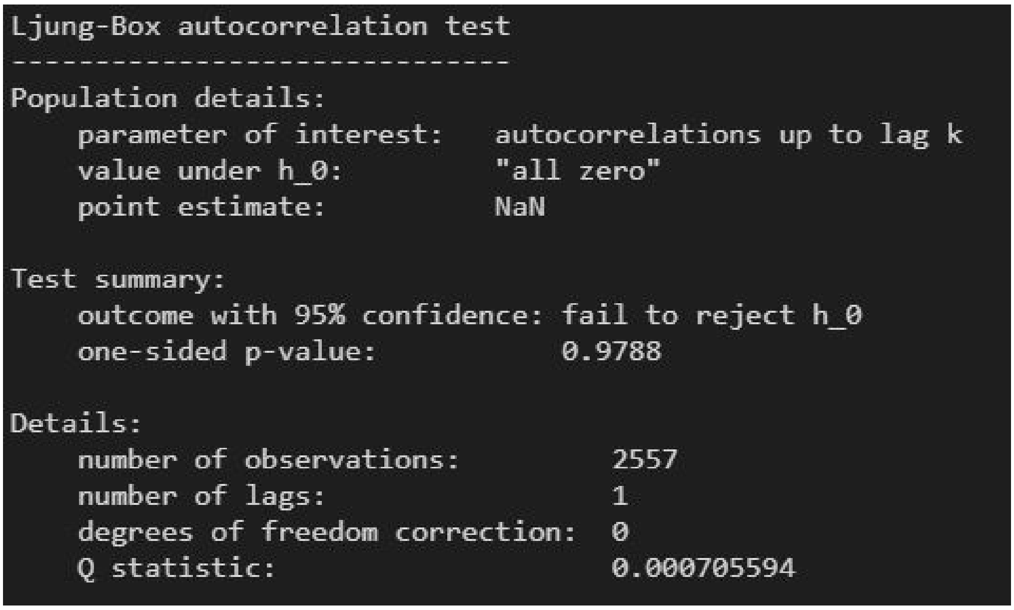
5.3 Niezależność szumu

Sprawdzimy teraz założenie dotyczące niezależności szumu w czasie. Sprawdzimy na początek wykres autokowariancji i częściowej autokowariancji.



Rysunek 5: ACF i PACF dla wartości resztkowych modelu

Na wykresie 5 możemy zobaczyć że mamy szum jest zależny tylko od siebie w tym samym momencie. Dodatkowo wykonamy test Ljunga-Boxa by upewnić się że szum jest niezależny od siebie.



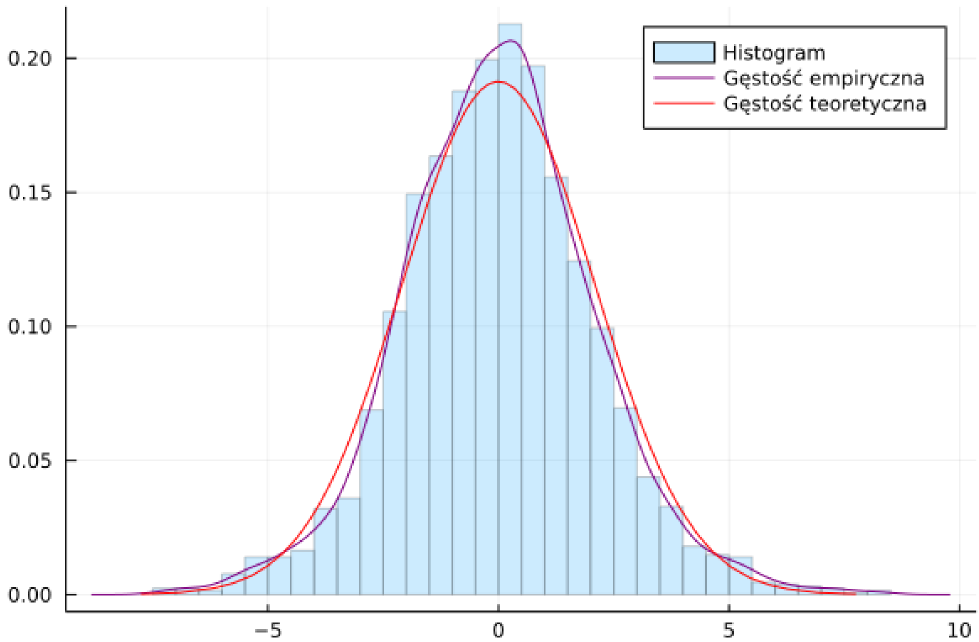
```
Ljung-Box autocorrelation test
------------------------------
Population details:
    parameter of interest:   autocorrelations up to lag k
    value under h_0:         "all zero"
    point estimate:          NaN

Test summary:
    outcome with 95% confidence: fail to reject h_0
    one-sided p-value:           0.9788

Details:
    number of observations:         2557
    number of lags:                 1
    degrees of freedom correction:  0
    Q statistic:                    0.000705594
```

Rysunek 6: Test Ljunga Boxa dla naszych wartości resztowych

Test nam potwierdza, że nie mamy podstaw do odrzucenia hipotezy mówiącej, że szum jest niezależny od siebie.

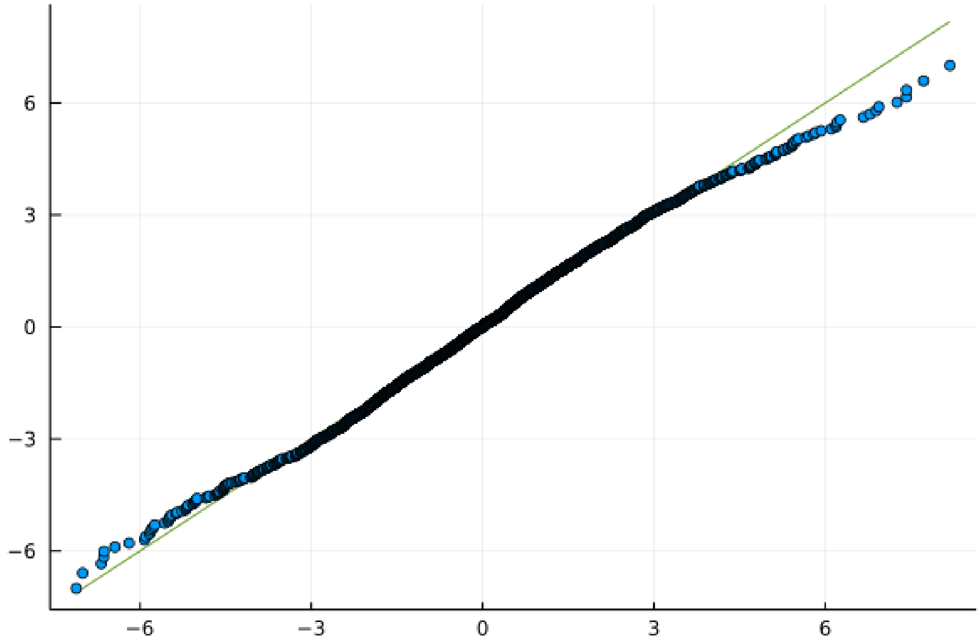
5.4 Założenie o normalności rozkładu

Sprawdzimy teraz dodatkowe założenie mówiące o tym że szum ma rozkład normalny. Na początku sprawdzimy histogram wraz z gęstością empiryczną i teoretyczną rozkładu normalnego.



Rysunek 7: Histogram, gęstość rozkładu szumu wraz z rozkładem normalnym

Patrząc na histogram wraz z gęstościami 7 możemy załuważyć, że rozkład wartości resztowych może być normalny ale nie musi. W celu dalszego sprawdzenia normalności zobaczymy jak będzie wyglądał wykres kwantylowy.



Rysunek 8: Qqplot dla wartości resztowych

Analizując wykres kwantylowy 8 widzimy, że potencjalnie ogony rozkładu niezf=gadzają się z rozkładem normalnym. W celu ostatecznej weryfikacji normalności wykonamy test Andersona Darlinga.

```
One sample Anderson-Darling test
--------------------------------
Population details:
    parameter of interest:   not implemented yet
    value under h_0:         NaN
    point estimate:          NaN

Test summary:
    outcome with 95% confidence: reject h_0
    one-sided p-value:           0.0334

Details:
    number of observations:   2557
    sample mean:              0.0007919156167252843
    sample SD:                2.0856431385378396
    A² statistic:             2.830020981780576
```

Rysunek 9: Test Andersona Darlinga dla naszych wartości resztowych

Test Andersona Darlinga 9 odrzuca hipotezę o normalności rozkładu. Dla naszych danych p-value wynosi 3.3%.

Podsumowując rozkład szumu nie ma rozkładu normalnego.

6 Wnioski autorów