

Projekt Egzaminacyjny

Jan Milewczyk, Maciej Wojciechowski, Kajetan Lach

December 5, 2024

1 Analizy log-zwrotów spółek

1.1 Vigo Photonics

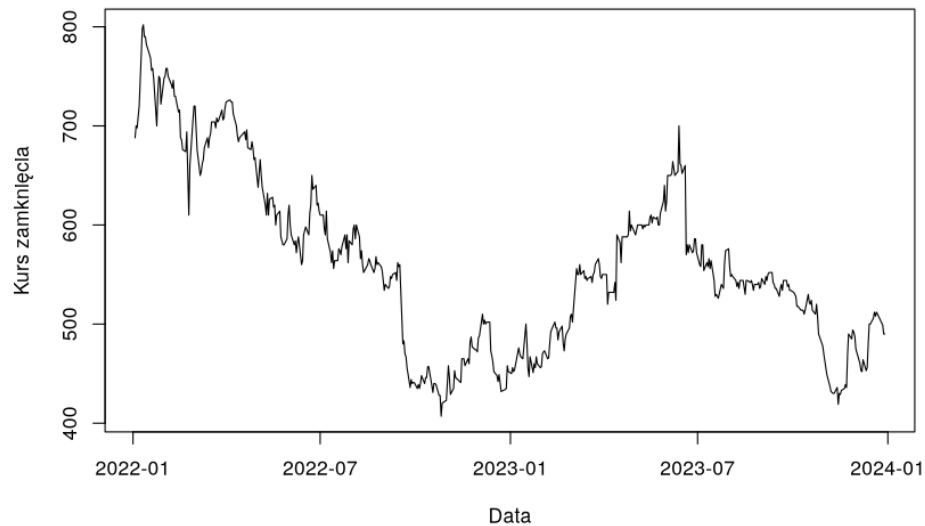
1.1.1 Wstęp

Vigo Photonics to polskie przedsiębiorstwo specjalizujące się w wytwarzaniu materiałów i przyrządów półprzewodnikowych do zastosowań fotonicznych i mikroelektronicznych. Spółka jest liderem na światowym rynku fotonowych detektorów średniej podczerwieni, a wszystkie produkty opiera na własnej, unikalnej technologii.

1.1.2 Analiza log-zwrotów spółki (Vigo Photonics)

Pierwszy rozdział zawiera analizę log-zwrotów spółki.

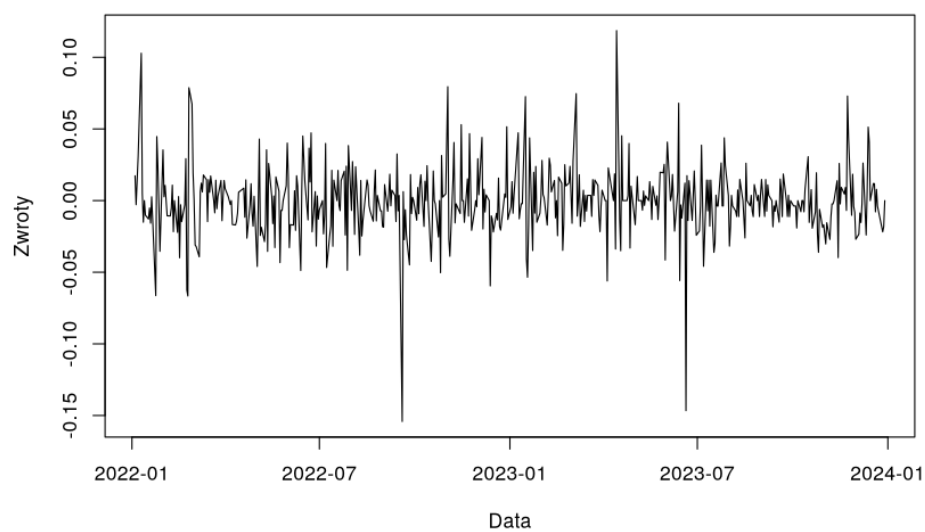
Wykresy kursów zamknięcia oraz log-zwrotów: Poniższy wykres ilustruje zmianę cen zamknięcia akcji w czasie.



Zmianę log-zwrotów, wyliczonych według wzoru

$$r_1 = \ln \frac{S_0}{S_1}, r_2 = \ln \frac{S_2}{S_1}, \dots, r_n = \ln \frac{S_n}{S_{n-1}}$$

gdzie s_0, s_1, \dots, s_n są kursami zamknięcia z kolejnych dni, na osi czasu ilustruje wykres poniżej.



Wartość oczekiwana: Zakładając, że log-zwroty r_1, r_2, \dots, r_n są niezależnymi realizacjami zmiennej losowej X , wyliczamy μ przy użyciu nieobciążonego estymatora wartości oczekiwanej:

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i = \mu$$

w wyniku czego otrzymujemy wynik

$$\mu = -0.0006787669$$

Wariancja i odchylenie standardowe: Korzystając ze wzoru na nieobciążony estymator wariancji

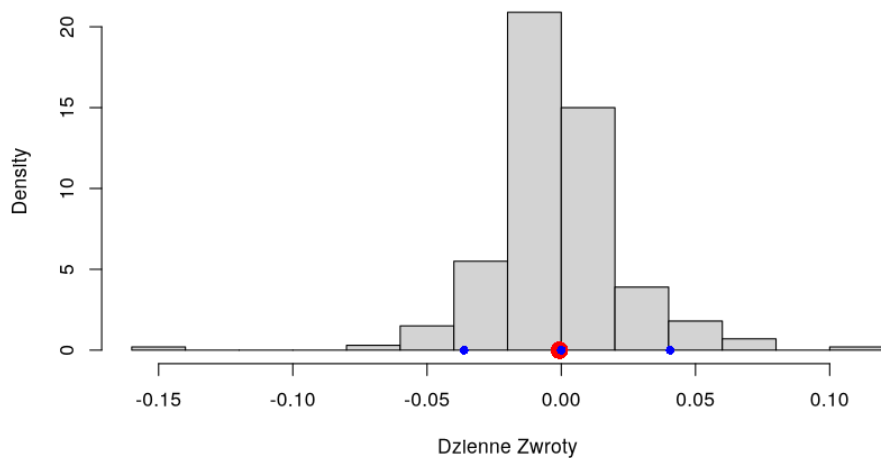
$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

otrzymujemy $\sigma^2 = 0.0006189792$ oraz $\sigma = 0.02487929$.

Kwantyle: Z wykorzystaniem klasycznego estymatora kwantyli wyestymowano kwantyle rzędu $\alpha = 5\%, 50\%$ i 95% . Wyniki przedstawione zostały w tabeli poniżej:

| \bar{x}_n | s_n^2 | s_n | $q(5\%)$ | $q(50\%)$ | $q(95\%)$ |
|---------------|--------------|------------|-------------|-----------|------------|
| -0.0006787669 | 0.0006189792 | 0.02487929 | -0.03620640 | 0 | 0.04055989 |

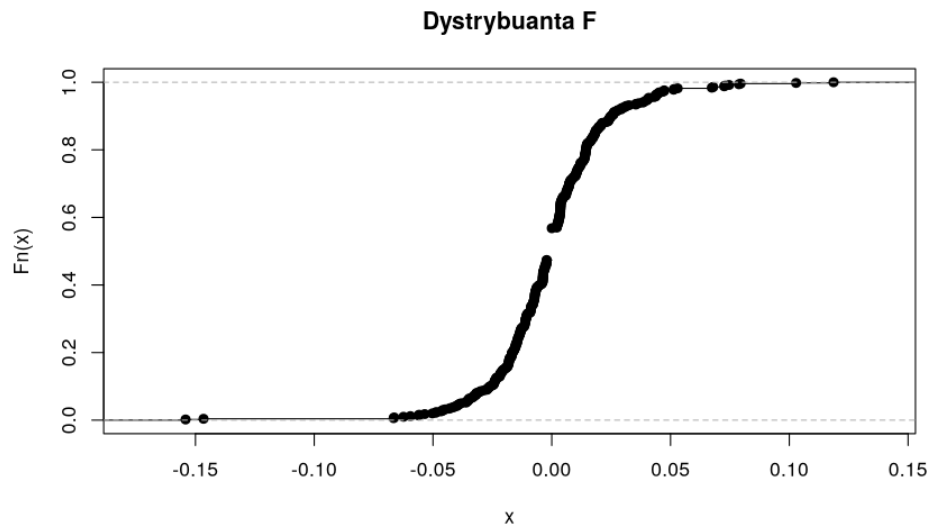
Histogram log-zwrotów: Na histogramie dziennych log-zwrotów cen akcji na czerwono i niebiesko oznaczono odpowiednio wartość wyestymowanej średniej oraz wartości kwantyli przedstawionych wcześniej.



Dystrybuanta: Z wykorzystaniem dystrybuanty empirycznej jako nieobciążonego estymatora

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}} = \left(\frac{\text{liczba elementów w próbie} \leq x}{n} \right)$$

estymujemy dystrybuantę F zaprezentowaną na wykresie poniżej

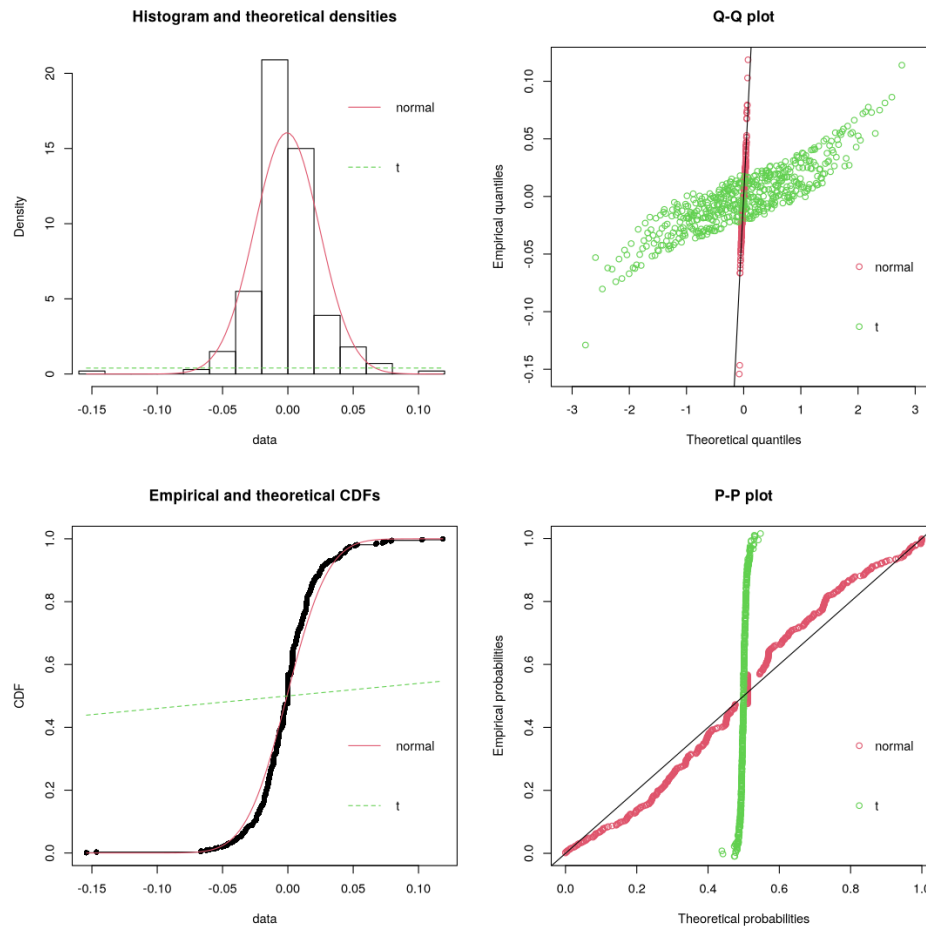


Analiza dobroci dopasowania rozkładu normalnego i t-Studenta:
 Korzystając z estymatora największej wiarygodności (MLE) parametru estymujemy parametry rozkładu normalnego i t-Studenta.

| m | s | df |
|---------------|-----------|----------|
| -0.0006787669 | 0.0248544 | 314.3698 |

1.1.3 Wykresy diagnostyczne

Poniższe wykresy prezentują kolejno porównania: histogram-gęstość wybranych rozkładów, kwantyl-kwantyl, dystrybuanta empiryczna-dystrybuanta teoretyczna wybranych rozkładów oraz prawdopodobieństwo teoretyczne wybranych rozkładów-prawdopodobieństwo empiryczne.



Na ich podstawie można stwierdzić, że rozkład normalny lepiej dopasowuje się do danych.

Weryfikacja wyboru z wykorzystaniem statystyk oraz kryteriów informacyjnych: Poniższe tabele prezentują wyniki testów statystycznych oraz kryteriów informacyjnych dla rozkładów normalnego i t-Studenta.

| Statystyki | Rozkład normalny | Rozkład t-Studenta |
|--------------------|------------------|--------------------|
| Kolmogorov-Smirnov | 0.08251589 | 0.469477 |
| Cramer-von Mises | 1.18780300 | 39.177553 |
| Anderson-Darling | 6.95700925 | 183.184800 |

| Kryteria | Rozkład normalny | Rozkład t-Studenta |
|----------|------------------|--------------------|
| AIC | -2271.782 | 922.0439 |
| BIC | -2263.353 | 926.2585 |

Powyższe wyniki potwierdzają, że rozkład normalny lepiej dopasowuje się do danych.

Test hipotezy równości rozkładów metodą Monte Carlo: Testujemy hipotezę zerową o równości dystrybuant

$$H_0 : F = F_0$$

przeciwko hipotezie alternatywnej (kontrhipotezie)

$$H_1 : F \neq F_0$$

gdzie F_0 jest dystrybuantą rozkładu normalnego wybranego w poprzedniej podsekcji, a F dystrybuantą nieznaną.

Do przetestowania hipotezy zerowej o równości dystrybuant wykorzystamy metodę Monte Carlo przy użyciu statystyki Kołmogorowa-Smirnowa.

W tym celu generujemy $N = 10000$ prób z rozkładu normalnego o parametrach

$$m = -0.0006787669 \text{ oraz } s = 0.0248544$$

wyestymowanych wcześniej oraz obliczamy wartość statystyki Kołmogorowa-Smirnowa dla każdej z nich.

Następnie obliczamy prawdopodobieństwo

$$p = P(D_n > d_n)$$

gdzie D_n to wartość statystyki dla wygenerowanych prób a d_n to jej wartość dla log-zwrotów spółki.

Ustalamy poziom istotności $\alpha = 0.05$ i sprawdzamy czy $p < \alpha$. W naszym przypadku

$$p = 0.0025 < \alpha = 0.05$$

co pozwala nam odrzucić hipotezę zerową na rzecz hipotezy alternatywnej.

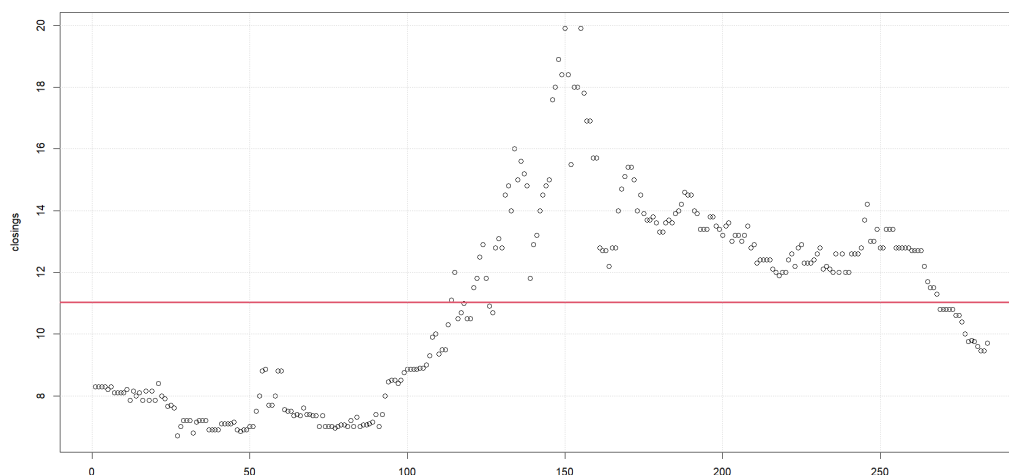
1.2 Digitree

1.2.1 Wstęp

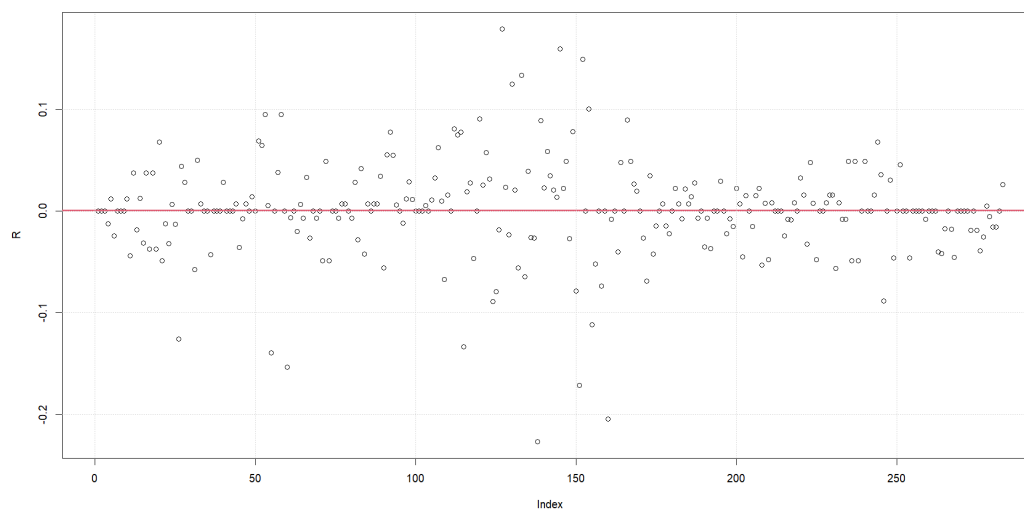
Digitree to polska spółka technologiczna specjalizująca się w kompleksowych rozwiązaniach z zakresu digital marketingu i wsparcia sprzedaży online.

1.2.2 Analiza log-zwrotów spółki DIGITREE

Spółka DIGITREE: Poniżej przedstawiam wykres kursów zamknięcia wybranej spółki



Poniżej przedstawiam wykres log-zwrotów wybranej spółki



Podstawowa analiza statystyczna log-zwrotów: Wartości log-zwrotów zostały obliczone jako różnice logarytmiczne dziennych kursów zamknięcia. Poniżej przedstawiono podstawowe statystyki log-zwrotów:

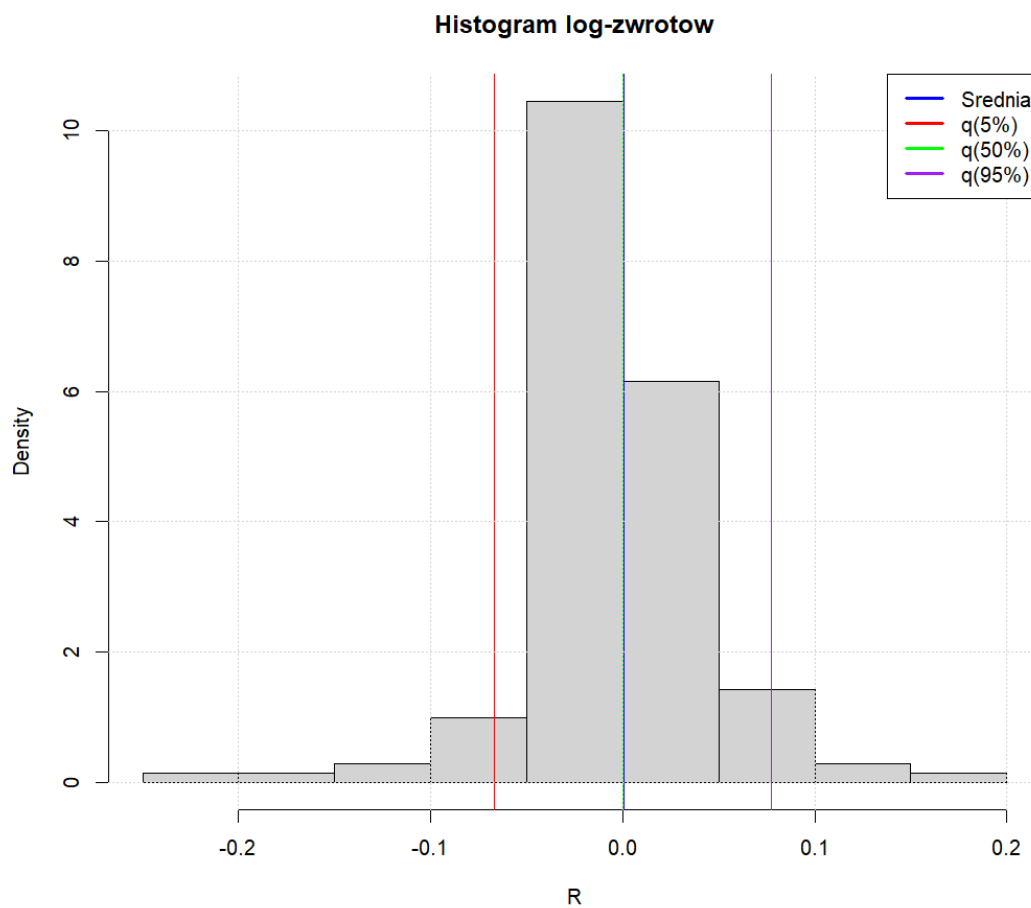
- Średnia log-zwrotów: $\bar{x}_n = 0.0005507787$
- Wariancja log-zwrotów: $s_n^2 = 0.0022445$
- Odchylenie standardowe log-zwrotów: $s_n = 0.04737616$

Table 1: Estymacja parametrów log-zwrotów

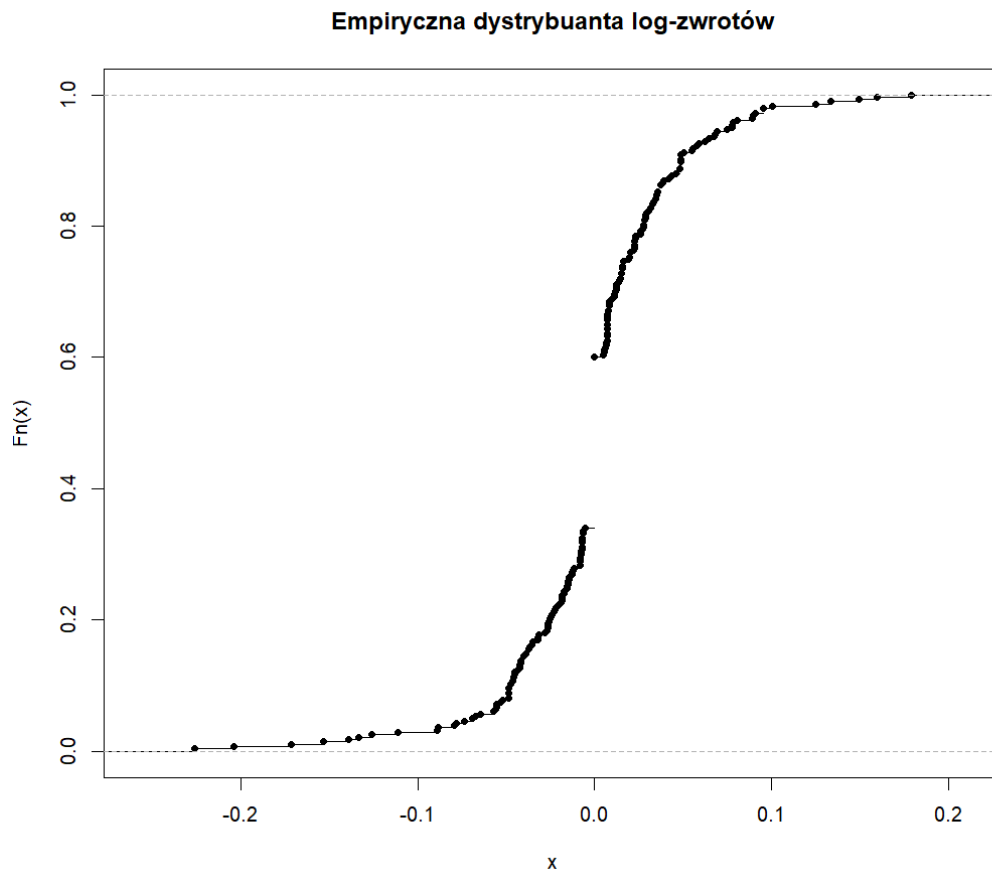
| \bar{x}_n | s_n^2 | s_n | $q(5\%)$ | $q(50\%)$ | $q(95\%)$ |
|--------------|-----------|------------|-------------|-------------|------------|
| 0.0005507787 | 0.0022445 | 0.04737616 | -0.06694173 | 0.000000000 | 0.07764551 |

Kwantyle 5

Histogram log-zwrotów z zaznaczoną średnią i kwantylami: Poniżej zamieszczono histogram log-zwrotów z zaznaczonymi wartościami średniej oraz kwantyli 5



Dystrybuanta empiryczna log-zwrotów: Estymacja dystrybuanty empirycznej dla log-zwrotów została przeprowadzona przy użyciu wzoru empirycznej dystrybuanty $F_n(x)$. Wykres dystrybuanty empirycznej log-zwrotów znajduje się poniżej.



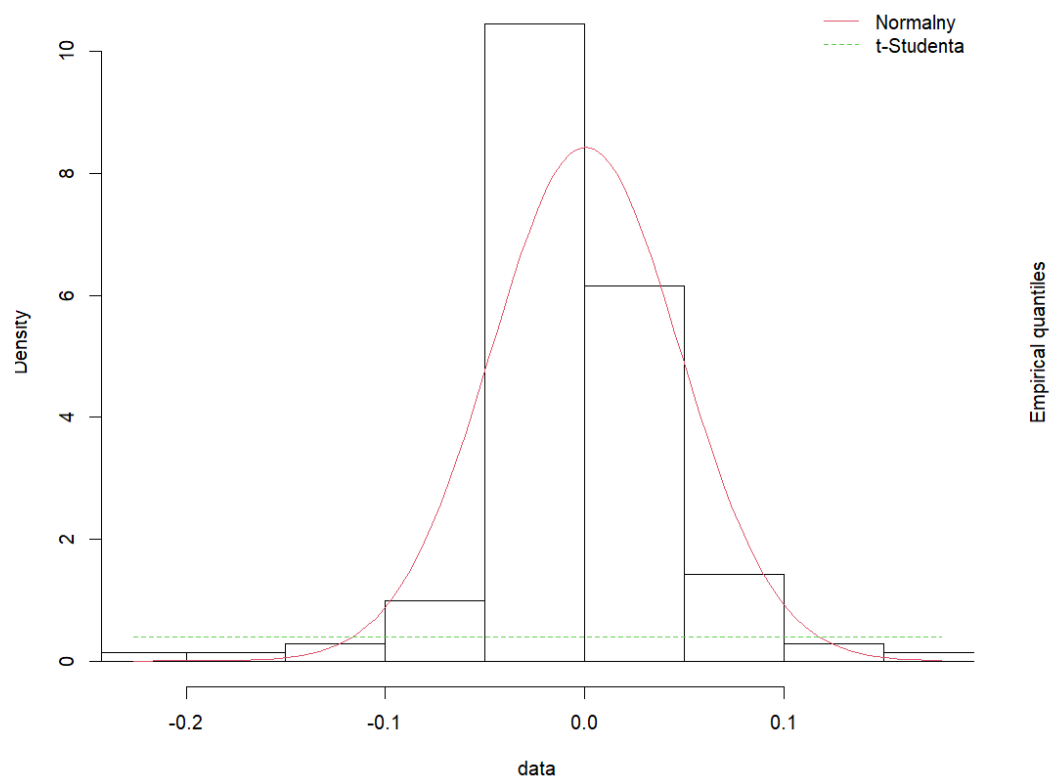
1.2.3 Analiza dobroci dopasowania rozkładów

Estymacja parametrów rozkładu normalnego i t-Studenta: Parametry rozkładu normalnego i t-Studenta zostały wyestymowane przy użyciu estymatora największej wiarygodności (MLE):

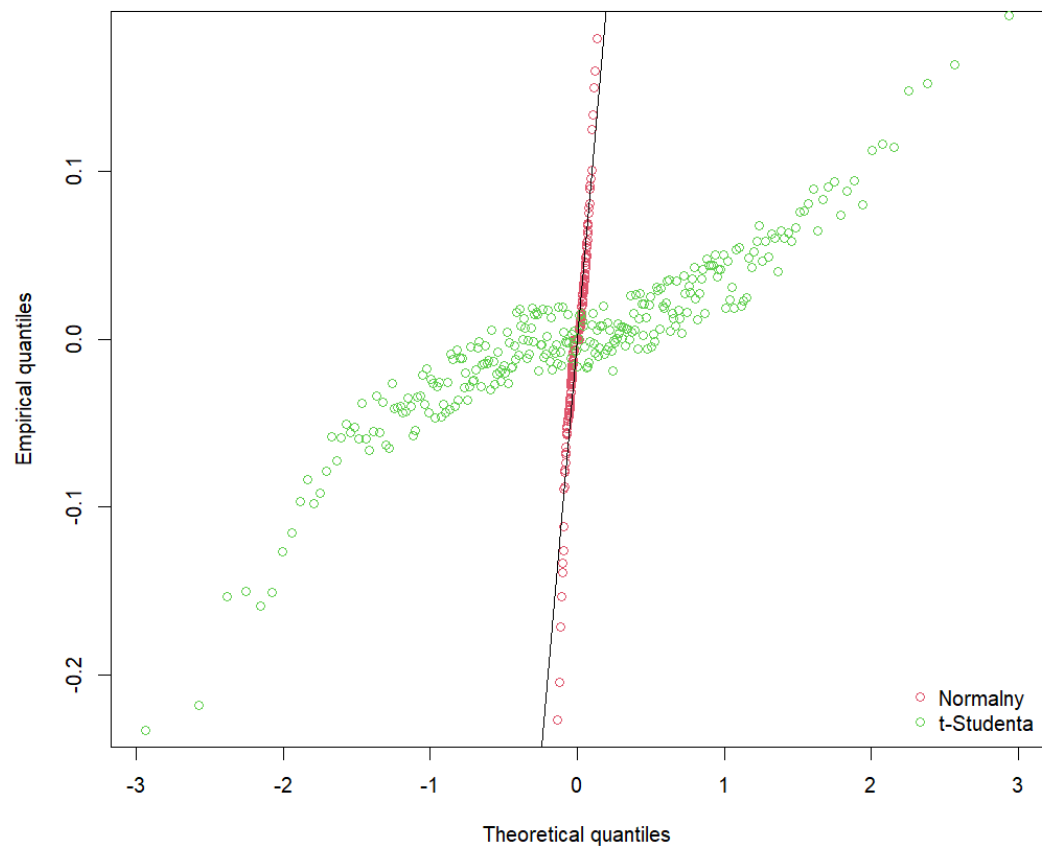
- Rozkład normalny: średnia = 0.0005507787, odchylenie standardowe = 0.0472923802
- Rozkład t-Studenta: liczba stopni swobody = 307.2041

Wykresy diagnostyczne: Poniżej przedstawiono wykresy diagnostyczne dla dopasowania rozkładów normalnego i t-Studenta do danych log-zwrotów, umożliwiające ocenę jakości dopasowania.

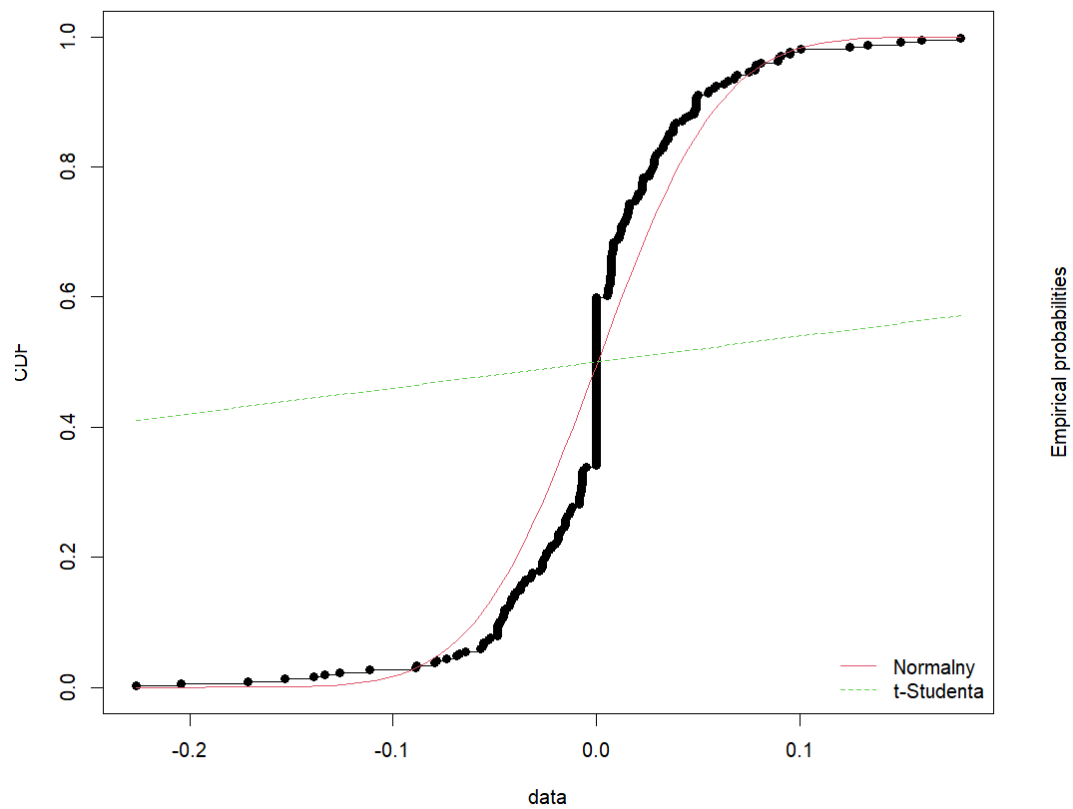
Porównanie dopasowania - gestosc



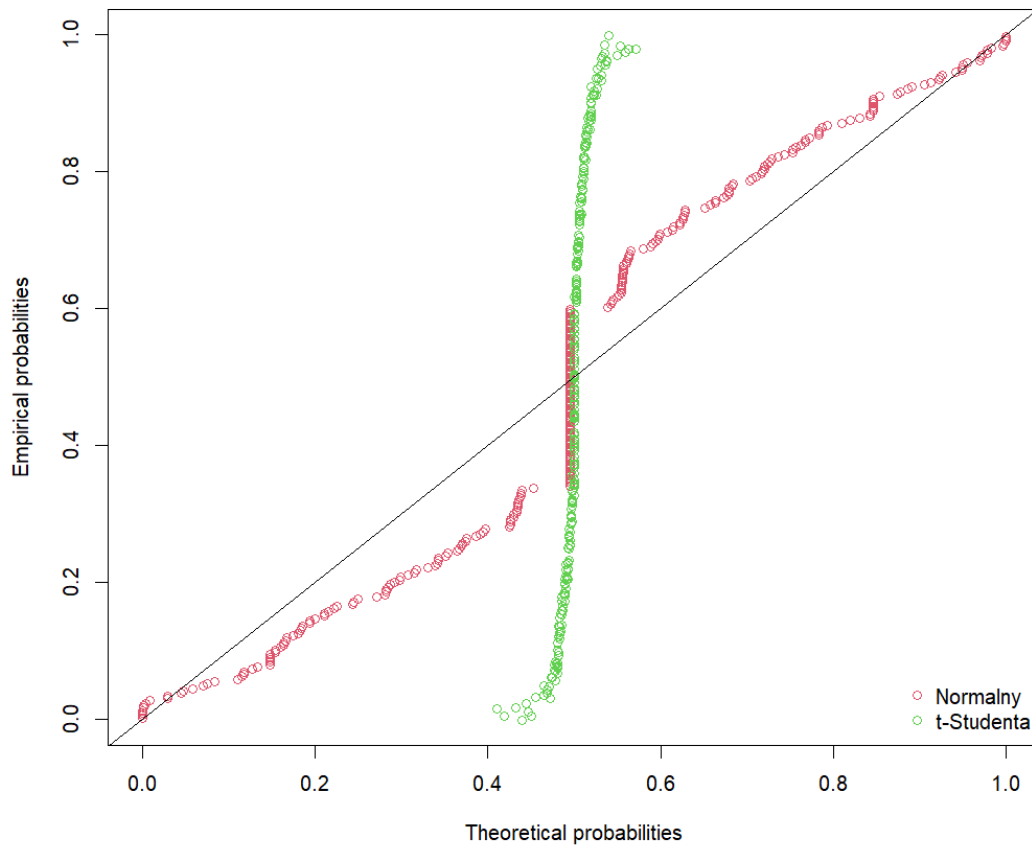
Porównanie dopasowania - kwantyle



Porównanie dopasowania - dystrybuanta



Porównanie dopasowania - prawdopodobieństwo



1.2.4 Ocena dopasowania rozkładów

Na podstawie wykresów diagnostycznych oraz statystyk dopasowania:

- Statystyki dla rozkładu normalnego: $KS = 0.1561$, $CM = 1.8659$, $AD = 9.3673$, $AIC = -919.9766$, $BIC = -912.6857$
- Statystyki dla rozkładu t-Studenta: $KS = 0.4424$, $CM = 21.0334$, $AD = 99.0999$, $AIC = 523.2149$, $BIC = 526.8603$

Wyniki sugerują, że rozkład normalny lepiej dopasowuje się do danych log-zwrotów niż rozkład t-Studenta. Wybór rozkładu normalnego uzasadniają niższe wartości AIC i BIC oraz lepsze dopasowanie na wykresach diagnostycznych.

Test hipotezy o równości rozkładów: Przeprowadzono test hipotezy o równości rozkładów dla wybranego rozkładu normalnego, wykorzystując statystykę Kolmogorova-Smirnova (KS). Wyniki testu przedstawiono poniżej:

- **Statystyka testowa D :** Obliczona wartość wynosi $D = 0.1561313$. Statystyka D mierzy maksymalną różnicę pomiędzy dystrybucją empiryczną danych a teoretyczną dystrybucją rozkładu normalnego. Wysoka wartość D wskazuje na większą różnicę, co może sugerować, że dane nie pochodzą z badanego rozkładu.
- **P-wartość (p):** Obliczona p-wartość wynosi $p = 0$. Oznacza to, że przy założeniu prawdziwości hipotezy zerowej, prawdopodobieństwo uzyskania tak dużej lub większej różnicy D wynosi praktycznie zero. W rezultacie, hipoteza zerowa o zgodności rozkładu danych z rozkładem normalnym zostaje odrzucona na dowolnym poziomie istotności.

Interpretacja:

- Obliczona wartość statystyki D sugeruje, że istnieją istotne różnice pomiędzy rozkładem danych a teoretycznym rozkładem normalnym.
- Niska p-wartość ($p = 0$) wskazuje, że te różnice są statystycznie istotne. Oznacza to, że dane najprawdopodobniej nie pochodzą z badanego rozkładu normalnego.

1.3 Wawel

1.3.1 Wprowadzenie

Wawel SA (WWL) jest polskim producentem słodczy, znanym z produkcji czekolad, cukierków, wafli oraz innych wyrobów cukierniczych. Firma cieszy się długą tradycją na rynku i dostarcza wysokiej jakości produkty konsumentom zarówno w Polsce, jak i za granicą.

Przeprowadzona analiza stanowi pierwszą część projektu zaliczeniowego, w której skupimy się na analizie log-zwrotów akcji spółki Wawel SA.

1.3.2 Wykresy kursów zamknięcia oraz log-zwrotów

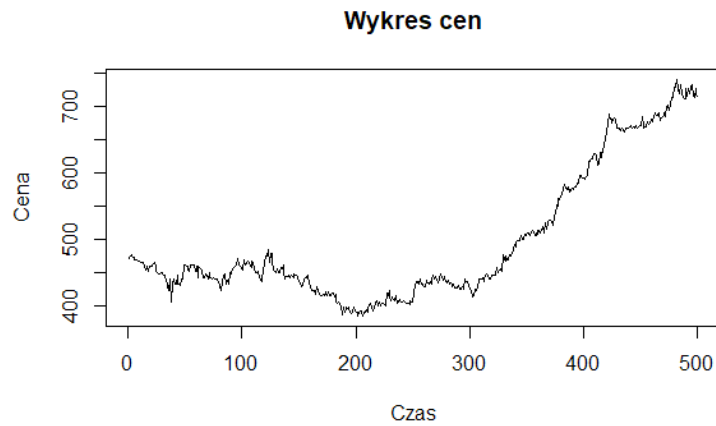


Figure 1: Wykres cen akcji spółki Wawel SA (WWL)

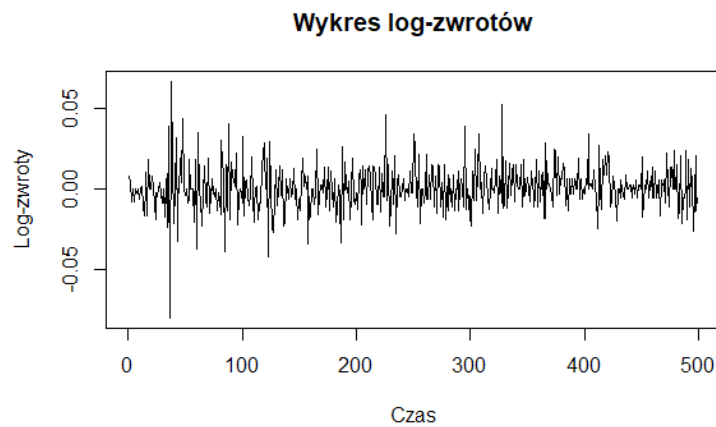


Figure 2: Wykres log-zwrotów akcji spółki Wawel SA (WWL)

Wykresy powyżej ilustrują zmiany kursów zamknięcia oraz log-zwrotów akcji spółki Wawel SA w czasie. Widać na nich zmienność cen akcji oraz ich wpływ na log-zwroty.

1.3.3 Analiza log-zwrotów

Zakładamy, że log-zwroty r_1, r_2, \dots, r_n są niezależnymi realizacjami zmiennej losowej X o dystrybucji F , wartości oczekiwanej μ i wariancji σ^2 .

Odpowiednie estymatory użyte w analizie to:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n r_i \quad (1)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (r_i - \hat{\mu})^2 \quad (2)$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \quad (3)$$

Korzystając z klasycznego estymatora kwantyli, wyestymowano kwantyle rzędu $\alpha = 5\%, 50\%, 95\%$. Wyniki zostały przedstawione w tabeli poniżej:

| Estymator | Wartość |
|---|-----------|
| Liczność próby (n) | 499 |
| Wartość oczekiwana ($\hat{\mu}$) | 0,000837 |
| Wariancja ($\hat{\sigma}^2$) | 0,000202 |
| Odchylenie standardowe ($\hat{\sigma}$) | 0,014220 |
| Kwantyl 5% ($q_{5\%}$) | -0,020416 |
| Kwantyl 50% ($q_{50\%}$) | 0,000000 |
| Kwantyl 95% ($q_{95\%}$) | 0,023786 |

Table 2: Podstawowe statystyki log-zwrotów akcji spółki Wawel SA (WWL)



Figure 3: Histogram log-zwrotów z zaznaczoną średnią i kwantylami

1.3.4 Interpretacja kwantyli

Każdy z kwantyli ma swoją specyficzną interpretację w kontekście log-zwrotów akcji:

- **Kwantyl 5%:** Wartość log-zwrotu, poniżej której znajduje się 5% najmniejszych obserwacji. Oznacza to, że istnieje 5% szans, że log-zwrot będzie niższy niż -0,020416.
- **Kwantyl 50% (mediana):** Wartość środkowa log-zwrotów. Oznacza to, że połowa log-zwrotów jest większa, a połowa mniejsza niż 0,000000.
- **Kwantyl 95%:** Wartość log-zwrotu, poniżej której znajduje się 95% obserwacji. Oznacza to, że tylko 5% log-zwrotów jest wyższych niż 0,023786.

1.3.5 Estymacja dystrybuanty

W celu estymacji dystrybuanty F wykorzystano empiryczny estymator dystrybuanty, którego wzór jest następujący:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(r_i \leq x), \quad (4)$$

gdzie $I(r_i \leq x)$ jest funkcją indykatora, a n oznacza liczbę obserwacji.

Dystrybuanta empiryczna została przedstawiona na wykresie poniżej:

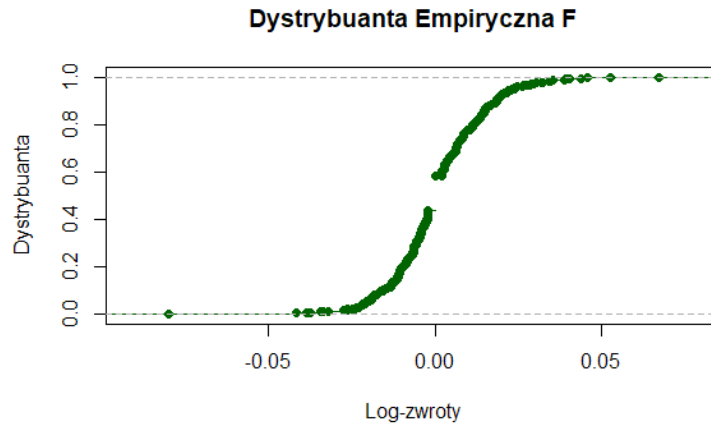


Figure 4: Dystrybuanta empiryczna $\hat{F}(x)$ dla log-zwrotów

1.3.6 Analiza dobroci dopasowania rozkładu normalnego i t-Studenta

Estymacja parametrów: Wykorzystując estymator największej wiarygodności (MLE) oraz bibliotekę `fitdistrplus` w R, wyestymowano parametry rozkładu normalnego i t-Studenta dla danych log-zwrotów.

| | Estymata | Błąd standardowy |
|---|----------|------------------|
| Średnia ($\hat{\mu}$) | 0,000837 | 0,000636 |
| Odchylenie standardowe ($\hat{\sigma}$) | 0,014205 | 0,000440 |

Table 3: Estymowane parametry rozkładu normalnego

Rozkład normalny

| | Estymata | Błąd standardowy |
|---------------------------------------|----------|------------------|
| Liczba stopni swobody ($\hat{\nu}$) | 4,9764 | 1,1504 |
| Średnia ($\hat{\mu}$) | 0,000344 | 0,000578 |
| Skala ($\hat{\sigma}$) | 0,011042 | 0,000611 |

Table 4: Estymowane parametry rozkładu t-Studenta

Rozkład t-Studenta

1.3.7 Wykresy diagnostyczne

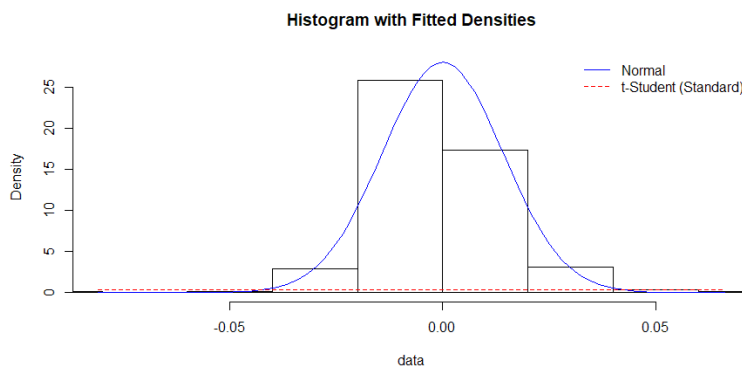


Figure 5: Histogram log-zwrotów z nałożonymi funkcjami gęstości rozkładów normalnego i t-Studenta

Histogram z nałożonymi gęstościami **Opis:** Wykres histogramu przedstawia empiryczny rozkład danych, na który nałożono dopasowane gęstości rozkładów normalnego (niebieska linia) i t-Studenta (czerwona linia przerywana). Analiza wykazuje, że:

- Rozkład normalny (niebieski) lepiej odwzorowuje dane empiryczne, szczególnie w centralnym zakresie wartości.
- Rozkład t-Studenta (czerwony) jest znacząco niedopasowany.

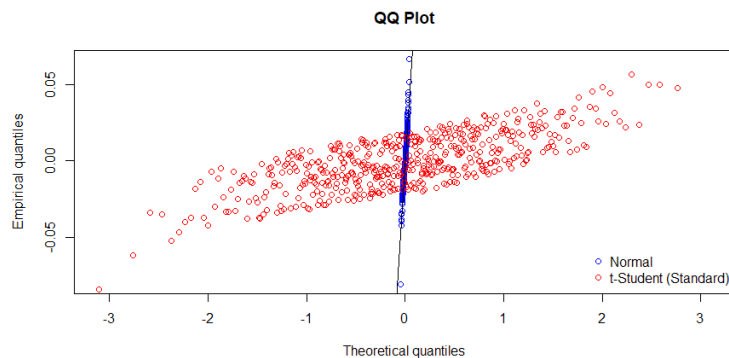


Figure 6: Wykres Q-Q dla rozkładów normalnego i t-Studenta

Wykres kwantyl-kwantyl (Q-Q plot) Opis: Wykres QQ porównuje kwantyle empiryczne danych z kwantylami teoretycznymi rozkładów normalnego i t-Studenta. Na wykresie zauważalne są następujące cechy:

- Punkty dla rozkładu normalnego (niebieskie) układają się niemal idealnie wzdłuż linii prostej, co sugeruje wysoką zgodność z danymi empirycznymi.
- Rozkład t-Studenta (czerwony) znacznie odstaje od linii prostej, wskazując na niedopasowanie tego rozkładu do danych.

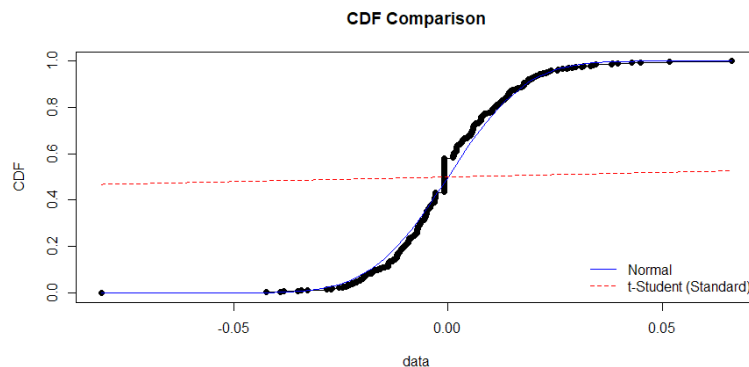


Figure 7: Porównanie dystrybuant empirycznej z teoretycznymi dla rozkładów normalnego i t-Studenta

Porównanie dystrybuant (CDF) Opis: Wykres porównuje empiryczną dystrybuantę danych z teoretycznymi dystrybuantami rozkładów normalnego i t-Studenta. Główne obserwacje:

- Dystrybuenta rozkładu normalnego (niebieska linia) doskonale pokrywa się z empiryczną dystrybuantą danych.

- Dystrybuanta rozkładu t-Studenta (czerwona linia przerywana) wykazuje znaczne rozbieżności, co wskazuje na niedopasowanie.

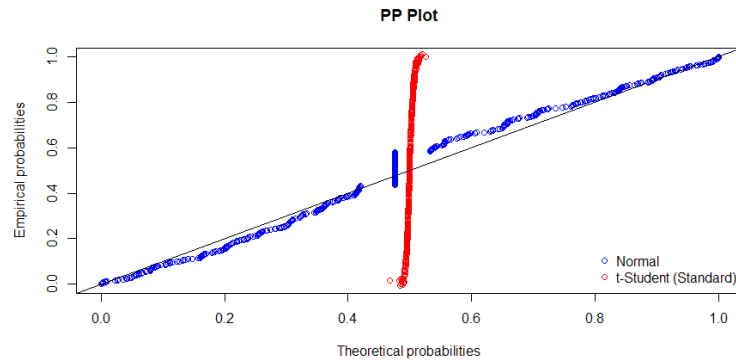


Figure 8: Wykres P-P dla rozkładów normalnego i t-Studenta

Wykres P-P (Probability-Probability plot) Opis: Wykres PP ilustruje zgodność empirycznych prawdopodobieństw z teoretycznymi dla rozkładu normalnego. Natomiast dane rozkładu T-Studenta znacząco odbiegają od danych rzeczywistych. Z analizy wynika:

- Punkty dla rozkładu normalnego (niebieskie) układają się niemal idealnie wzdłuż linii przekątnej, co potwierdza bardzo dobre dopasowanie do danych.
- Rozkład t-Studenta (czerwony) pokazuje znaczne odstępstwa od linii przekątnej, co sugeruje istotne niedopasowanie.

1.3.8 Ocena dopasowania rozkładów

Analiza jednoznacznie wskazuje, że dane empiryczne są znacznie lepiej modelowane przez rozkład **normalny** niż przez rozkład t-Studenta. **Wnioski:**

- Rozkład normalny charakteryzuje się bliskim dopasowaniem zarówno w centralnej części, jak i w ogonach.
- Rozkład t-Studenta znacząco odbiega od danych, co czyni go nieodpowiednim w tym przypadku.

Interpretacja statystyk dopasowania rozkładów
Statystyki dopasowania (*Goodness-of-fit statistics*)

| Statystyka | Normalny | t-Studenta |
|--------------------|----------|------------|
| Kolmogorov-Smirnov | 0.1047 | 0.4811 |
| Cramer-von Mises | 0.6771 | 40.0829 |
| Anderson-Darling | 3.5170 | 186.7583 |

Table 5: Porównanie statystyk dopasowania rozkładów

Interpretacja:

- Rozkład normalny uzyskał znacznie niższe wartości dla wszystkich trzech statystyk, co wskazuje na lepsze dopasowanie do danych w porównaniu z rozkładem t-Studenta.
- Szczególnie wartość statystyki Andersona-Darlinga dla rozkładu t-Studenta (186.7583) podkreśla jego niedopasowanie.

Kryteria dopasowania (*Goodness-of-fit criteria*)

| Kryterium | Normalny | t-Studenta |
|--------------------------------------|-----------|------------|
| Akaike's Information Criterion (AIC) | -2825.534 | 919.8094 |
| Bayesian Information Criterion (BIC) | -2817.109 | 924.0220 |

Table 6: Porównanie kryteriów dopasowania rozkładów

Interpretacja:

- Zarówno kryterium Akaike'go (AIC), jak i kryterium Bayesowskie (BIC) wskazują, że rozkład normalny jest znacznie lepiej dopasowany niż rozkład t-Studenta.
- Niższe wartości dla rozkładu normalnego (-2825.534 dla AIC oraz -2817.109 dla BIC) wskazują na przewagę tego modelu w uwzględnianiu dopasowania i prostoty modelu.

Podsumowanie

Na podstawie przedstawionych statystyk i kryteriów dopasowania:

- Rozkład **normalny** jest znacznie lepszy w modelowaniu danych niż rozkład t-Studenta.
- Rozkład t-Studenta charakteryzuje się dużymi odstępstwami w dopasowaniu, co czyni go nieodpowiednim w tym przypadku.

1.3.9 Test hipotezy metodą Monte Carlo

Dla wybranego rozkładu t-Studenta przeprowadzono test hipotezy o równości rozkładów, wykorzystując statystykę Kolmogorowa-Smirnowa (KS) oraz metodę Monte Carlo. Wyniki testu przedstawiono w tabeli 5.

| Wynik | Wartość |
|-----------------------------------|-----------|
| Zaobserwowana statystyka KS | 0,0893735 |
| P-wartość z symulacji Monte Carlo | 0 |

Table 7: Wyniki testu hipotezy metodą Monte Carlo

Interpretacja: P-wartość równa 0 wskazuje na silne podstawy do odrzucenia hipotezy zerowej, że dane pochodzą z rozkładu t-Studenta z wyestymowanymi parametrami. Oznacza to, że mimo lepszego dopasowania w porównaniu z rozkładem normalnym, rozkład t-Studenta nie jest w stanie w pełni opisać danych log-zwrotów.

Histogram logarytmicznych zwrotów

Opis: Histogram przedstawia empiryczny rozkład logarytmicznych zwrotów danych:

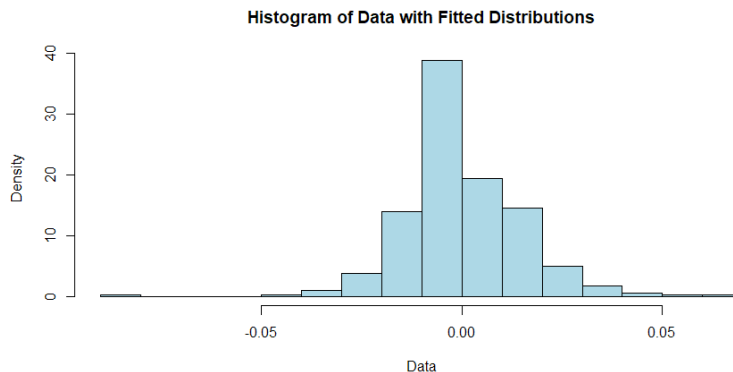


Figure 9: Histogram logarytmicznych zwrotów

Podsumowanie: Przeprowadzona analiza log-zwrotów akcji spółki Wawel SA wykazała, że:

- Średnia log-zwrotów jest bliska zeru, co sugeruje brak wyraźnego trendu wzrostowego lub spadkowego w badanym okresie.
- Odchylenie standardowe wskazuje na umiarkowaną zmienność log-zwrotów.
- Kwantyle pokazują asymetrię w rozkładzie log-zwrotów, co jest typowe dla danych finansowych.
- Rozkład normalny lepiej dopasowuje się do danych niż rozkład T-Studenta, co potwierdzają statystyki dobroci dopasowania i kryteria informacyjne.
- Test hipotezy metodą Monte Carlo sugeruje odrzucenie hipotezy, że dane pochodzą z rozkładu t-Studenta, wskazując na lepsze dopasowanie wyników otrzymanych z rozkładu normalnego.

2 Analiza łącznego rozkładu log-zwrotów