Projekt Egzaminacyjny

Maria Koren

Listopad 2023 - Styczeń 2024

Spis treści

1 Wstęp

W tym projekcie reprezentowana jest analiza danych spółek KMR.UK (Kenmare Resources Plc) oraz JJB (Jujubee S.A).

Kenmare Resources plc to uznana firma wydobywcza, która zarządza kopalnią minerałów tytanu Moma położoną na północno-wschodnim wybrzeżu Mozambiku. Kopalnia prowadzi produkcję komercyjną od 2009 roku i jest uznawana za głównego dostawcę produktów z piasku mineralnego dla klientów na całym świecie, działających w ponad 15 krajach.

Jujubee S.A. to studio deweloperskie zajmujące się tworzeniem gier wideo, które ma swoim koncie takie tytuły jak: „FLASHOUT 3D", „Suspect in Sight", „Take Off – The Flight Simulator", strategię czasu rzeczywistego „Realpolitiks", grę przygodowo-dokumentalną „KURSK" oraz „Deep Diving Simulator". Studio zostało założone przez byłych pracowników CD Projekt RED, Traveller's Tales oraz Inifinite Dreams. Celem firmy jest tworzenie niesamowicie grywalnych i doskonale wyglądających gier na wszystkie istotne platformy sprzętowe, takie jak iOS (iPhone, iPod, iPad), Android, Mac, PC i konsole. Jujubee jest spółką notowaną na rynku NewConnect (JJB).

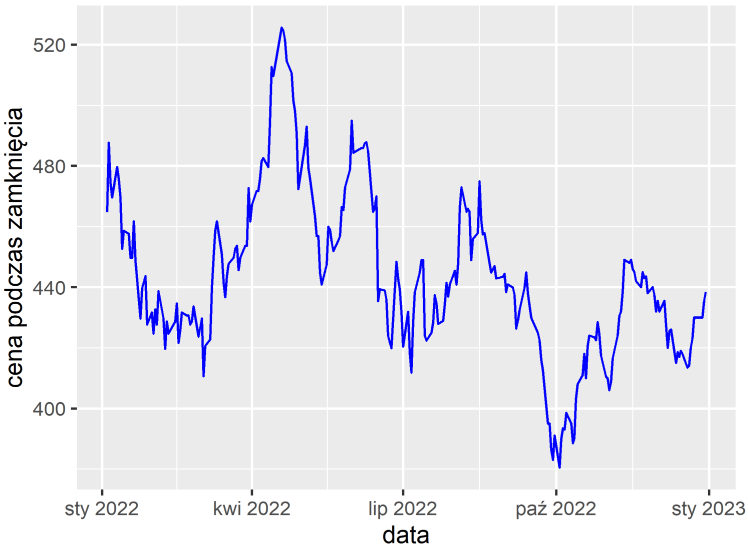
W pierwszej części projektu analizowano kursy zamknięcia spółki KMR za 2022 rok. Badano jakie ma rozkłady za pomocą wykresów diagnostycznych oraz statystyk opisowych. W drugiej części projektu badano zależność między log-zwrotami kursów powyższych spółek. W trzeciej części projektu zrobiona regresja liniowa dla log-zwrotów

2 Analiza cen spółek

2.1 Analiza cen spółki KMR.UK

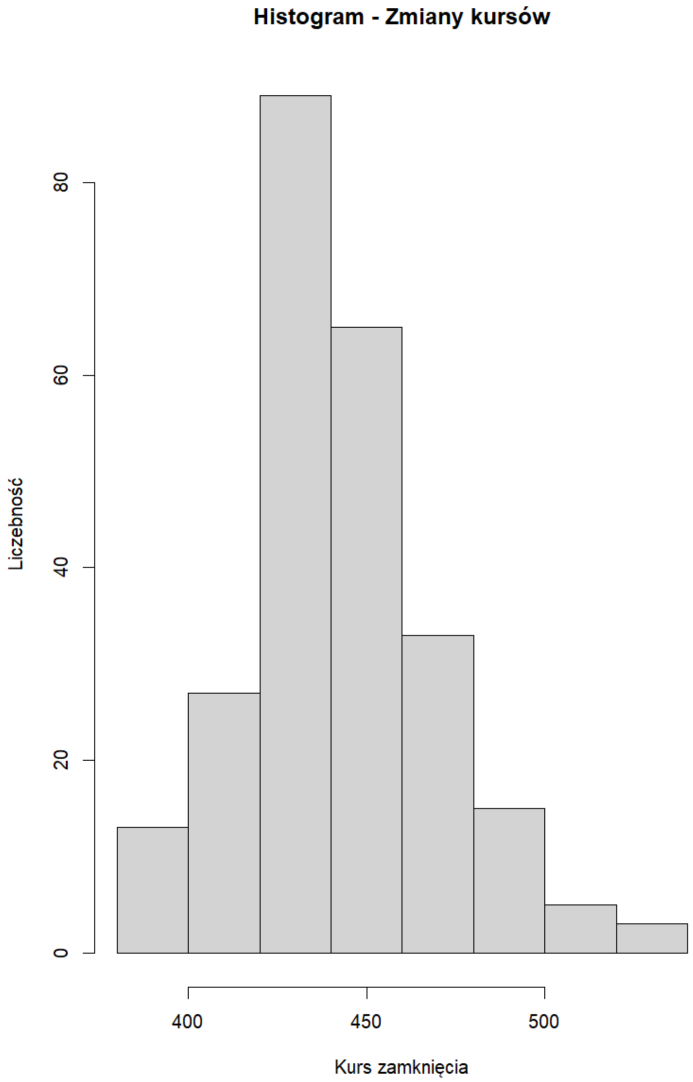
2.1.1 Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram

Zrobiono wykres kursu zamknięcia pokazujący zmiany w czasie, rysunek 1



Rysunek 1: Cena podczas zamknięcia

Oraz histogram, pokazujacy liczebność danych, rysunek 2



Rysunek 2: Histogram danych

2.1.2 Statystyki opisowe

Zostały obliczone nastepujące statystyki opisowe: średnia, odchylenie standardowe, skośność oraz kurtoza

Wyniki statystyk znajdują się w poniższej tabeli 1

	$\bar{x}$	odch. st.	skośność	kurtoza
akcje	442.8741	26.7262	3.574323	0.52783

Tabela 1: Statystyki opisowe

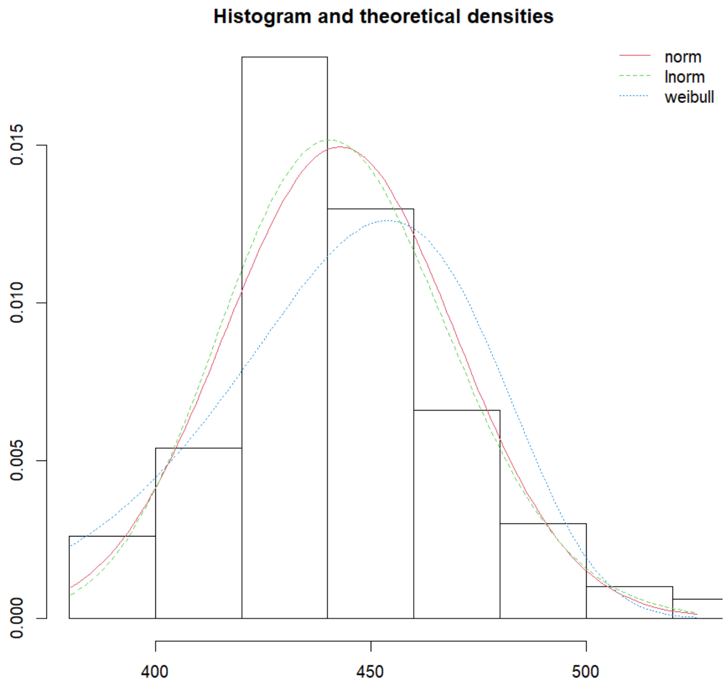
Interpretacja wyników

- Otrzymana skośność mówi o przewadzę wartości wyższych (wartość skośności powyżej zera)

- Otrzymana kurtoza mówi cieńszych ogonach niż rozkład normalny (bardziej płaski) (wartość kurtozy mniej niż 3)

2.1.3 Estymacja parametrów trzech rozkładów korzystając z estymatora największej wiarygodności (MLE)

Wyestymowano wyniki trzech rozkładów: normalnego, log-normalnego oraz rozkładu Weibulla za pomocą estymatora MLE. Wyżej wymienione wykresy dodano do wcześniejszego histogramu, co widać za rysynku 3



Rysunek 3: Histogram wraz z estymowanami rozkładami

Wyniki estymacji parametrów są przedstawione w tabeli 2

	μ	σ
normalny	442.8741	26.67269
	μ	σ
log-normalny	6.091499	0.05957788
	a	σ
weibulla	15.61307	455.8914

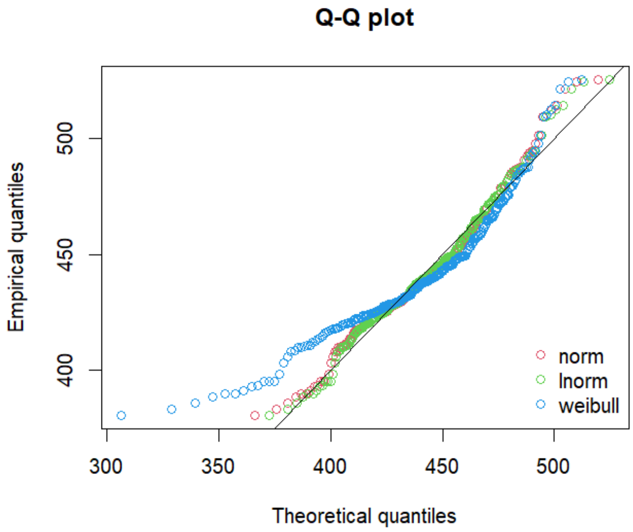
Tabela 2: Wyniki estymacji parametrów

Te wyniki oznaczają, że zostały dopasowane następujące rozkłady:

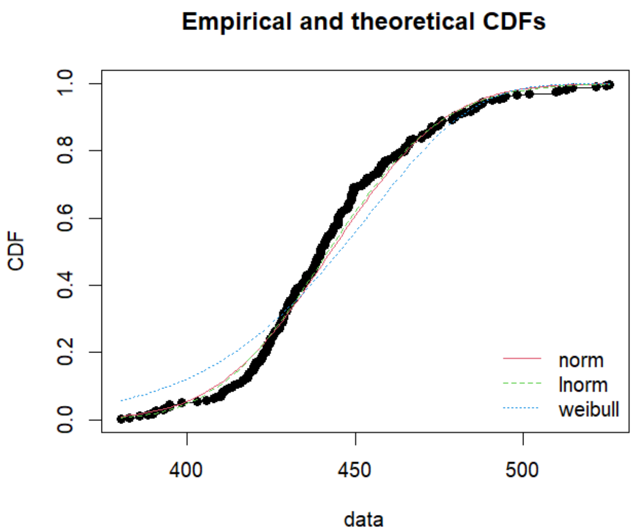
- $X \sim N(442.87, 26.67)$
- $X \sim LN(6.09, 0.059)$
- $X \sim W(15.61, 455.89)$

2.1.4 Wykresy diagnostyczne

Zostały zrobione wykresy diagnostyczne qq-plot (rysunek 4) oraz cdf (rysunek 5)



Rysunek 4: Wykres qq-plot



Rysunek 5: Wykres cdf

- Wykres qq-plot

  Jest to wykres kwantyl-kwantyl, na osi pionowej są kwantyle teoretyczne, na osi poziomowej są kwantyle empiryczne. Kwantyl rzędu $\alpha \in (0, 1)$ zmiennej losowej ciągłej X to taka liczba q, dla której prawdopodobieństwo, że zmienna X przyjmuje wartości mniejsze lub równe q jest równe α.

  Najlepiej jest gdy te kwantyle są takie same bądź bardzo blizkie siebie. Dlatego najlepszym rozkładem jest najbliższy do prostej $y = x$. W rozważanym przykład takim jest wykres log-normalny $X \sim LN(6.09, 0.059)$

- Wykres CDF

  Funkcja rozkładu kumulacyjnego (CDF, Cumulative Distribution Function) to graficzna reprezentacja kumulatywnej dystrybuanty danej zmiennej losowej. CDF dla danej wartości x to prawdopodobieństwo, że zmienna losowa przyjmuje wartość mniejszą lub równą x. Czarnym zaznaczone są dane empiryczne. Najlepszym wykresem jest mający teorytyczne dane najbliższe do danych empirycznych. W rozważanym przykładzie takim wykresem jest log-normalny $X \sim LN(6.091, 0.059)$

Na podstawie wykresów diagnostycznych najlepszym rozkładem jest rozkład logarytmiczno-normalny

Analiza wartości statystyk KS, CM i AD oraz kryteria informacyjne AIC i BIC

Bazując wyłącznie na wykresach diagnostycznych, nie jest możliwe wybranie najlepszego wykresu. Dlatego skorzystano ze statystyk Kołmogorowa-Smirnowa, Cramera-von-Misesa, Andersona-Darlinga, a także z kryteriów informacyjnych AIC (Akaike's Information Criterion) oraz BIC (Bayesian Information Criterion)

Wartości ze statystyk KS, CM, AD są umieszczone w tabeli 3. Wartości kryteriów AIC, BIC w tabeli 4

	normalny	log-normalny	weibull
Kolmogorov-Smirnov	0.09168955	0.0798396	0.138442
Cramer-von Mises	0.3613063	0.2485924	1.257677
Anderson-Darling	2.005469	1.412547	7.416593

Tabela 3: Statystyki

	normalny	log-normalny	weibull
Akaike's Information Criterion	2355.289	2348.983	2415.692
Bayesian Information Criterion	2362.332	2356.026	2422.735

Tabela 4: Kryteria informacyjne

Ponieważ statystyki są oparte na porównaniu odległości dystrybuant, najlepszym rozkładem jest ten, który jest najbliżej do danych teorytycznych (ma najmniejszą odległość), czyli ma najmniejszą wartość statystyki. W kryteriach informacyjnych za najlepszy rozkład również jest uważany rozkład, mający najmniejszą wartosć kryteria. W rozważanym przykładzie takim rozkładem jest rozkład log-normalny $X \sim LN(6.09, 0.059)$

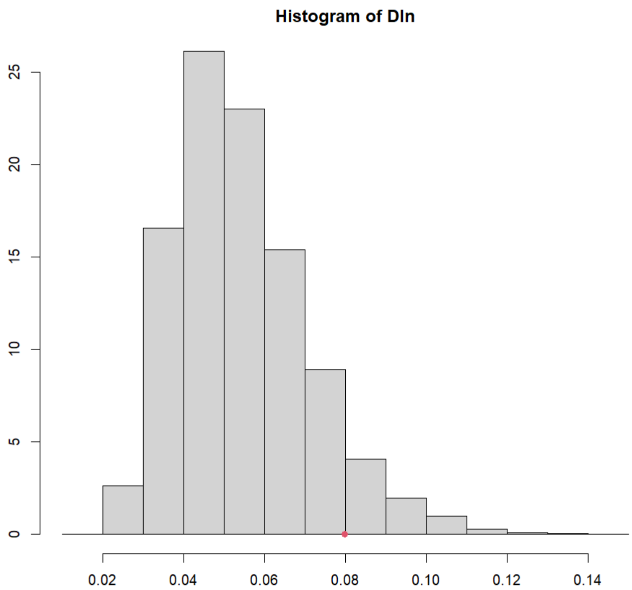
2.1.5 Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS

Zrobiona hipoteza H0: $F = LN(6.09, 0.0595)$ przeciwko hipotezie H1: F nie jest rowny $LN(6.09, 0.059)$

Zgenerowano $N = 10000$ probek licznosci n (równej ilości danych) z rozkładu $F0 = LN(6.09, 0.059)$ wybranego wcześniej jako najlepszego rozkładu i obliczono odleglość dystrybuant empirycznych od rozkladu F0 (wartosc statystyki Dn)

Obliczona również wartość statystyki dla rzeczywustych danych

Rysowany jest histogram statystyk testu KS uzyskanych z danych losowych, a także dodany jest punkt dla statystyki testu KS uzyskanej z rzeczywistych danych dla porównania danych rzeczywistych z danymi losowymi tego rozkładu



Rysunek 6: Histogram danych teorytycznych

Z wykresu widać, że wynik z danych rzeczywistych jest umieszczony w miejscu, gdzie dane losowe jeszcze są

Ta statystyka zwraca 2 informacje: odległość (wartość *statistic*) oraz prawdopodobieństwo że wartości statystyki KS są takie same lub większe, gdy hipoteza zerowa jest prawdziwa

Wyniki tego testowania zostaną umieszczone w tabeli 5:

statistic	p-value
0.0798396	0.0827

Tabela 5: Wartość statystyki KS w testowaniu hipotezy

P-wartość informuje o prawdopodobieństwie uzyskania takiej samej lub bardziej ekstremalnej statystyki testu, niż ta, którą otrzymaliśmy z danych rzeczywistych (zakładając że dane pochodzą z tego samego rozkładu)

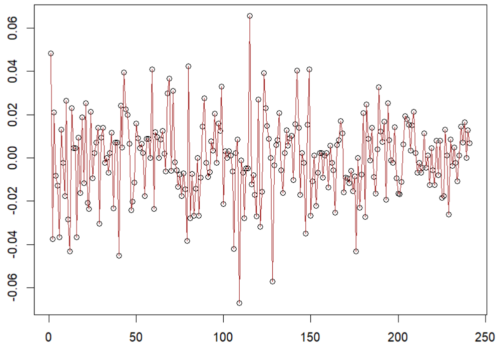
Ponieważ p-wartosć $p = 0.0827 > 0.05$ zatem nie ma powodów odrzucenia hipotezy. Uzyskane wyniki potwierdzają wybraną hipotezę o logarytmiczno-normalnym $LN(6.09, 0.059)$ rozkładzie danych

3 Analiza łącznego rozkładu log-zwrotów

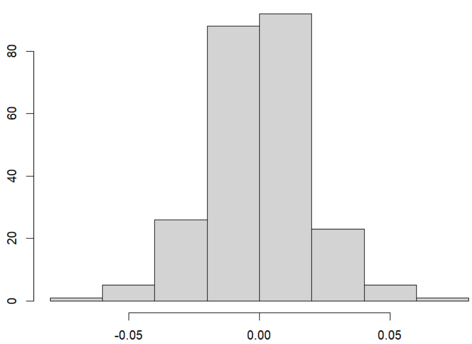
3.1 Analiza rozkładów brzegowych

3.1.1 Analiza rozkładów brzegowych spółki KMR

Na rysynkach 7 oraz 8 są przedstawione odpowiednio wykresy log-zwrotów oraz histogram



Rysunek 7: Wykres log-zwrotów spółki KMR



Rysunek 8: Wykres log-zwrotów spółki KMR

Został dopasowany do tych danych rozkład normalny $X \sim N(-0.0023, 0.018)$, rysunek 9

Zrobiono wykresy diagnostyczne cdf oraz qqplot, rysunki 10, 11

Przeprowadzono również test równości metodą Monte-Carlo. Skorzystano ze statystyki Kolmogorova-Smirnova. Wynik p-value otrzymany w tym teście jest równy 0.5656, co oznacza brak podstaw do odrzucenia hipotezy że rozklad log-zwrotów jest $X \sim N(-0.0023, 0.018)$, rysunek 12

3.1.2 Analiza rozkładów brzegowych spółki JJB

Na rysynkach 13 oraz 14 są przedstawione odpowiednio wykresy log-zwrotów oraz histogram

Został dopasowany do tych danych rozkład normalny $Y \sim N(-0.0025, 0.057)$, rysunek 15

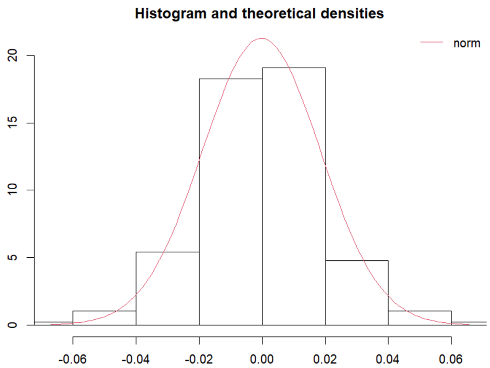
Zrobiono wykresy diagnostyczne cdf oraz qqplot, rysunki 16, 17

Przeprowadzono również test równości metodą Monte-Carlo. Skorzystano ze statystyki Kolmogorova-Smirnova. Wynik p-value otrzymany w tym teście jest równy 0.5709, co nie daje powodów do odrzucenia hipotezy że rozklad log-zwrotów jest $Y \sim N(-0.0025, 0.057)$

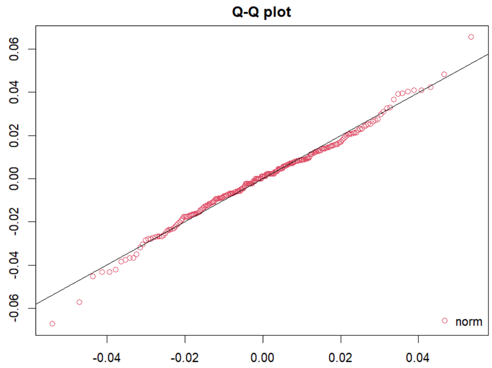
3.2 Estymacja parametrów rozkładu dwuwymiarowego normalnego oraz analiza dobroci dopasowania

3.2.1 Wykres rozrzutu z histogramami rozkładów brzegowych

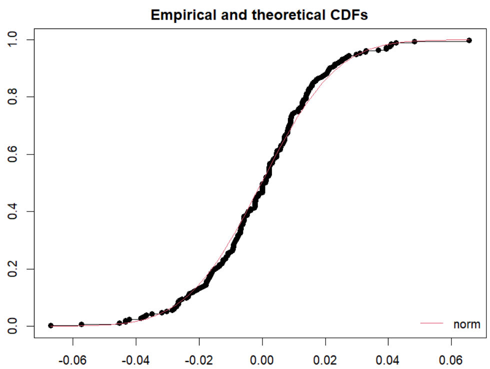
Zrobiono wykres rozrzutu z histogramami brzegowymi, rysunek 19. Wykres rozrzutu wizualizuje zależności między dwiema zmiennymi. Każdy punkt na wykresie reprezentuje log-zwroty spółek jednego dnia. Rozproszenie danych na podanym wykresie sugeruje że log-zwroty rozważanych spółek mogą mieć rozkład dwuwymiarowy normalny



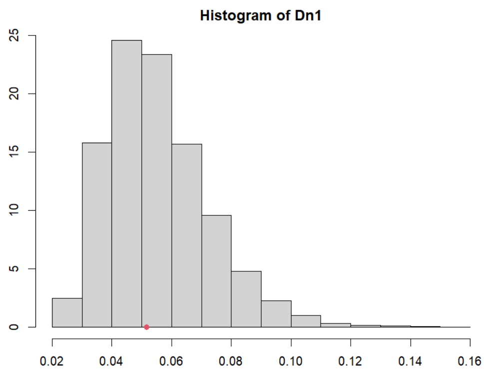
Rysunek 9: Wykres log-zwrotów wraz z dopasowanym rozkładem normalnym spółki KMR



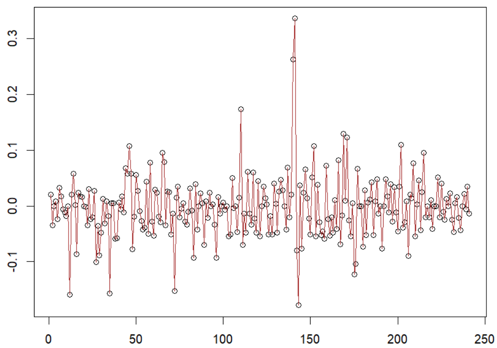
Rysunek 10: Wykres qqplot dla log-zwrotów spółki KMR



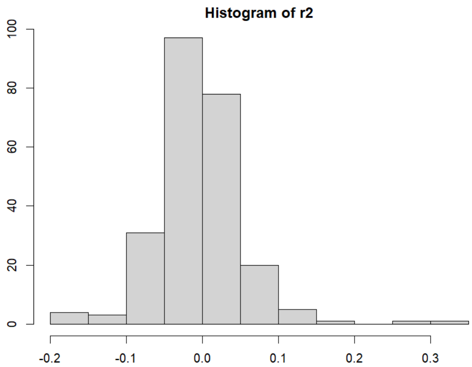
Rysunek 11: Wykres cdf dla log-zwrotów spółki KMR



Rysunek 12: Histogram danych w teście



Rysunek 13: Wykres log-zwrotów spółki JJB



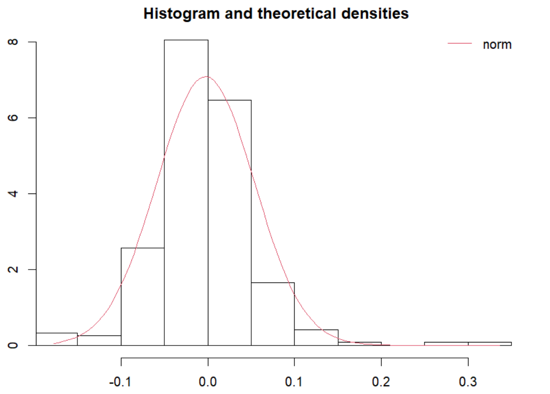
Rysunek 14: Wykres log-zwrotów spółki JJB

3.2.2 Wektor średnich, kowariancji, macierz korelacji, współczynnik korelacji

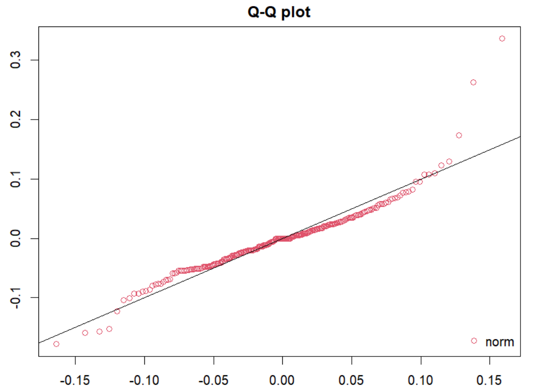
Zostały obliczone następujące estymatory:
Wektor średnich:

Macierz kowariancji (tabela 7), współczynnik kowariancji $cov(X,Y) = 7.719315e - 05$

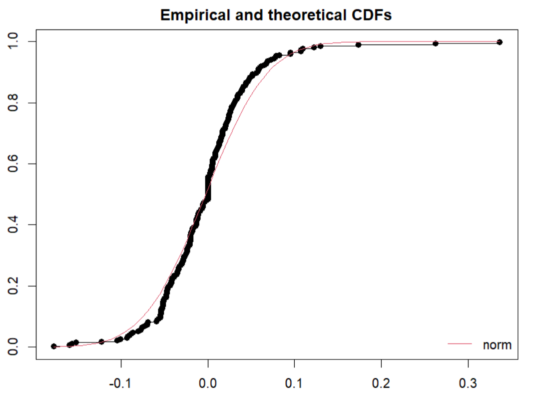
Macierz korelacji (tabela 8). Współczynnik korelacji przyjmuje wartości z przedziału $[-1;1]$ i wskazuje na stopień zależności dwóch zmiennych, im bliżej wartości bezwzględnej 1, tym silniejszy związek. W tym przypadku $\rho = 0.0728$ wskazuje że log-zwroty rozważanych



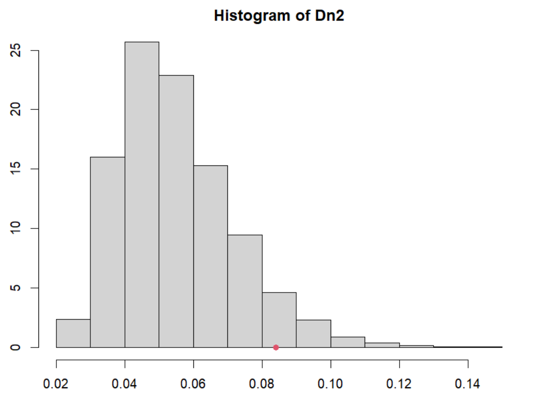
Rysunek 15: Wykres log-zwrotów wraz z dopasowanym rozkładem normalnym spółki JJB



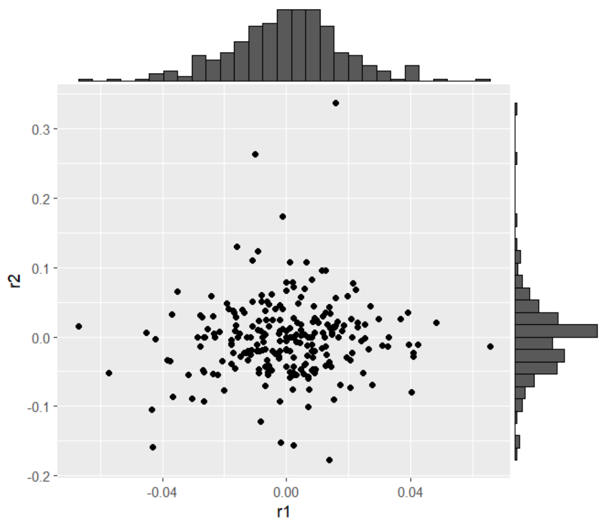
Rysunek 16: Wykres qqplot dla log-zwrotów spółki JJB



Rysunek 17: Wykres cdf dla log-zwrotów spółki JJB



Rysunek 18: Histogram danych w teście



Rysunek 19: Wykres rozrzutu z histogramami brzegowymi

spółek są słabo-dodatnio skorelowane

μ_1	μ_2
-0.0002403422	-0.0024970847

Tabela 6: Wektor średnich

	r_1	r_2
r_1	3.529462e-04	7.719315e-05
r_2	7.719315e-05	3.181408e-03

Tabela 7: Macierz kowariancji

	r_1	r_2
r_1	1.00000000	0.07284753
r_2	0.07284753	1.00000000

Tabela 8: Macierz korelacji

3.2.3 Wzór gęstości rozkładu dwuwymiarowego normalnego

Wyestymowane wcześniej parametry oznaczają że do log-zwrotów spółek został dopasowany rozkład

$$(X, Y) \sim\ N(-0.000240342, -0.0024970847, 0.01878686, 0.05640397, 0.0728)$$

Inaczej zapisując:

$$(X, Y) \sim\ N(\begin{bmatrix}-0.0002403 & -0.002497\end{bmatrix}, \begin{bmatrix}3.529462e-04 & 7.719315e-05 \\ 7.719315e-05 & 3.181408e-03\end{bmatrix})$$

Gęstość dopasowanego rozkładu

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi \cdot 0.01878686 \cdot 0.05640397 \cdot \sqrt{1-0.0728^2}} \\ & \exp\left(-\frac{1}{2(1-0.0728^2)}\left[\frac{(x+0.000240342)^2}{0.01878686^2}\right.\right. \\ & \left.\left. -2\cdot 0.0728 \frac{(x+0.000240342)(y+0.0024970847)}{0.01878686 \cdot 0.05640397} + \frac{(y+0.0024970847)^2}{0.05640397^2}\right]\right) \end{aligned} \tag{1}$$

Rozkład brzegowy dla spółki KMR: $X \sim\ N(-0.000240342, 0.01878686)$. Gęstość rozkładu brzegowego:

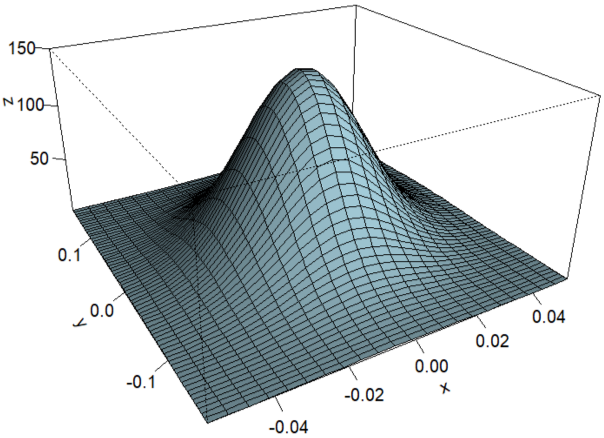
$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 0.01878686} \exp\left(-\frac{1}{2}\left(\frac{x+0.000240342}{0.01878686}\right)^2\right) \tag{2}$$

Rozkład brzegowy dla spółki JJB: $Y \sim\ N(-0.0024970847, 0.05640397)$. Gęstość rozkładu brzegowego:

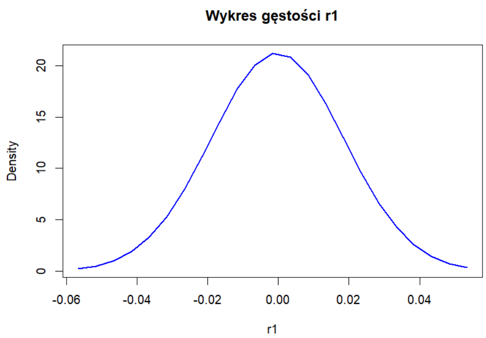
$$f(y) = \frac{1}{\sqrt{2\pi} \cdot 0.05640397} \exp\left(-\frac{1}{2}\left(\frac{y+0.0024970847}{0.05640397}\right)^2\right) \tag{3}$$

Wykres gęstości rozkładu łącznego (rysunek 20):

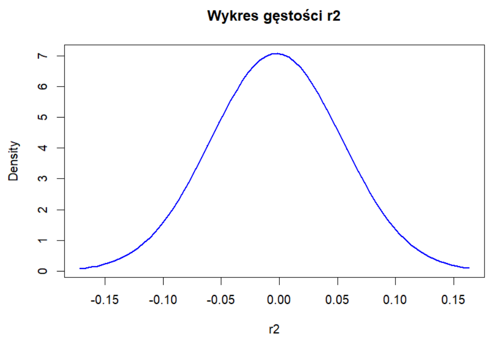
Wykres gęstości rozkładu brzegowego KMR (rysunek 21a). Wykres gęstości rozkładu brzegowego JJB (rysunek 21b)



Rysunek 20: Wykres gęstości rozkładu łącznego



(a) Wykres gęstości X



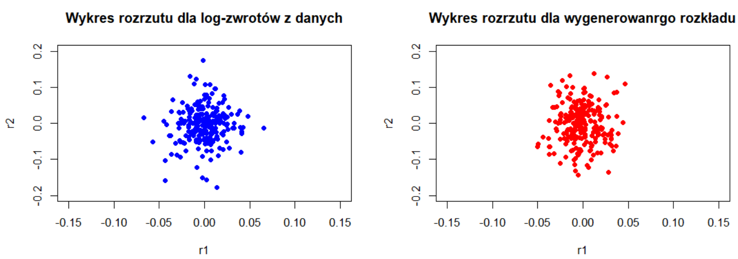
(b) Wykres gęstości Y

Rysunek 21: Wykresy gęstości rozkładów brzegowych

3.3 Analiza dopasowania rozkładu

3.3.1 Porównanie wykresów rozrzutu w oparciu o wygenerowaną próbę

Z rozkładu $(X,Y) \sim N(-0.000240342, -0.0024970847, 0.01878686, 0.05640397, 0.0728)$ została wygenerowana próba liczności 241 (ponieważ taką liczność mają dane log-zwrotów spółek), rysunek 22. Analizując powyższe wykresy zauważano podobieństwo kształtów



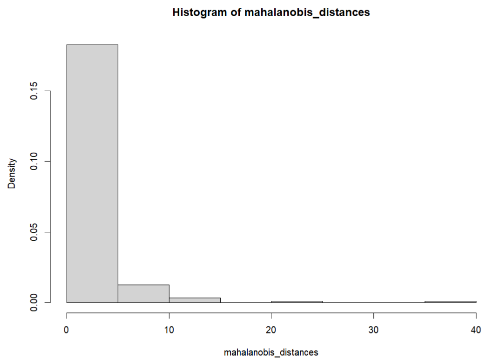
Rysunek 22: Wykresy rozrzutu wygenerowanej próby oraz danych

wykresu log-zwrotów rzeczywistych danych, oraz wykresu z wygenerowanego rozkładu. Na podstawie tego podobieństwa zrobiona hipoteza, że rozkład log-zwrotów może być $(X,Y) \sim N(-0.000240342, -0.0024970847, 0.01878686, 0.05640397, 0.0728)$ dobrze dopasowany. Ale żeby się upewnić zrobiono testowanie hipotezy, że kwadraty odległości Mahalanobisa wektora log-zwrotów od średniej, mają rozkład $\chi^2(2)$

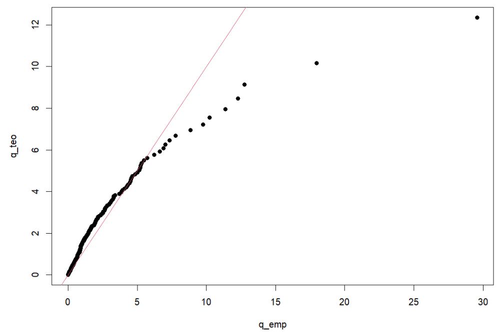
3.3.2 Analiza odległości Mahalanobisa

Obliczono kwadraty odległości Mahalanobisa dla każdej pary cen od średniej $\hat{\mu}$. Oraz zrobiono histogram, rysunek 23a

Zrobiono wykres diagnostyczny qqplot: porównywanie kwantyli empirycznych oraz kwantyli z rozkładu χ^2, rysunek 23b



(a) Histogram kwadratów odległości Mahalanobisa



(b) QQplot kwantyli empirycznych oraz z rozkładu χ^2

Rysunek 23: Analiza odległości Mahalanobisa

Punkty w miarę układają się blizko prostej $y = x$ co wskazuje, że rozkład kwadratów odległości Mahalanobisa może być z rozkładu χ^2

Żeby bardziej się upewnić przeprowadzono test zgodnosci oparty na statystyce Kolmogorova-Smirnova. Wynik testu $p - value = 0.0001707 < 5\%$ zatem odrzucono hipotezę, ze kwadraty odleglości Mahalanobisa mają rozkład χ^2, co skutkuje też odrzuceniem hipotezy o normalnosci rozkładu log-zwrotow

4 Regresja liniowa dla log-zwrotów

4.1 Wyznaczenie prostej regresji

Wyznaczono prostą regresji $R_2 = b_0 + b_1 \cdot R_1$, (gdzie R_2 to spółka JJB, a R_1 spółka KMR) zamieszczenie tej prostej na wykresie rozrzutu log-zwrotów można widzieć na rysunku 24

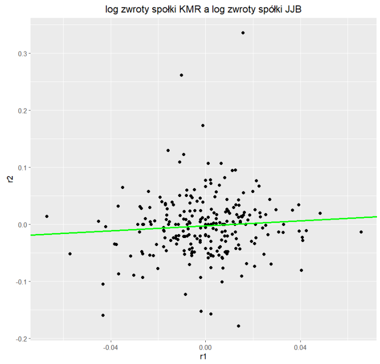
Wykorzystano model $R_2 = b_0 + b_1 \cdot R_1 + \epsilon$, zakłądając że $\epsilon \sim N(0, \sigma^2)$. Wyestymowano współczynniki b_0, b_1 metodą najmnijeszych kwadratów, otrzymano $\beta_1 = 0.2187108, \beta_0 = -0.002444519$. Skąd linia regresji to $y = -0.0024 + 0.22 \cdot x$.

Zrobiona klasyczna regresja w R za pomocą funkcji $lm()$. Otrzymane wyniki można obejrzeć na rysunku 25

Opis otrzymanych wyników modelu regresji

1. (Estimate) są to współczyniki $\hat{b_0} = \beta_0, \hat{b_1} = \beta_1$ opisujące linię regresji ($\hat{b_0} = -0.002445, \hat{b_1} = 0.218711$)

2. (Residual standard error) Błąd standardowy reszt modelu $\hat{\sigma}$ ($\hat{\sigma} = 0.05637$)



Rysunek 24: Linia regresji

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.002445   0.003632  -0.673    0.502
r1           0.218711   0.193687   1.129    0.260

Residual standard error: 0.05637 on 239 degrees of freedom
Multiple R-squared:  0.005307,  Adjusted R-squared:  0.001145
F-statistic: 1.275 on 1 and 239 DF,  p-value: 0.2599
```

Rysunek 25: Wyniki regresji

3. (Multiple R-squared) Współczynnik determinacji R ($R = 0.005307$)
4. (t value, Pr $(> |\text{t}|)$) Wynik testu na istotność współczynników b_0, b_1 – wartość statystyki testowej oraz p-value
5. (F-statistic) Wynik dodatkowego testu na istotność współczynnika b_1

Analiza reszt modelu

Przeprowadzono również analizę reszt modelu, zakładając rozkład normalny $\epsilon \sim N(0, \sigma^2)$. W rozważanym przykłądzie błędem standardowym reszt jest $\hat{\sigma} = 0.05637$. Dlatego analizowano rozkłąd $\epsilon \sim N(0, 0.05637^2)$

Histogram reszt zamieszczono niżej, rysunek 26a

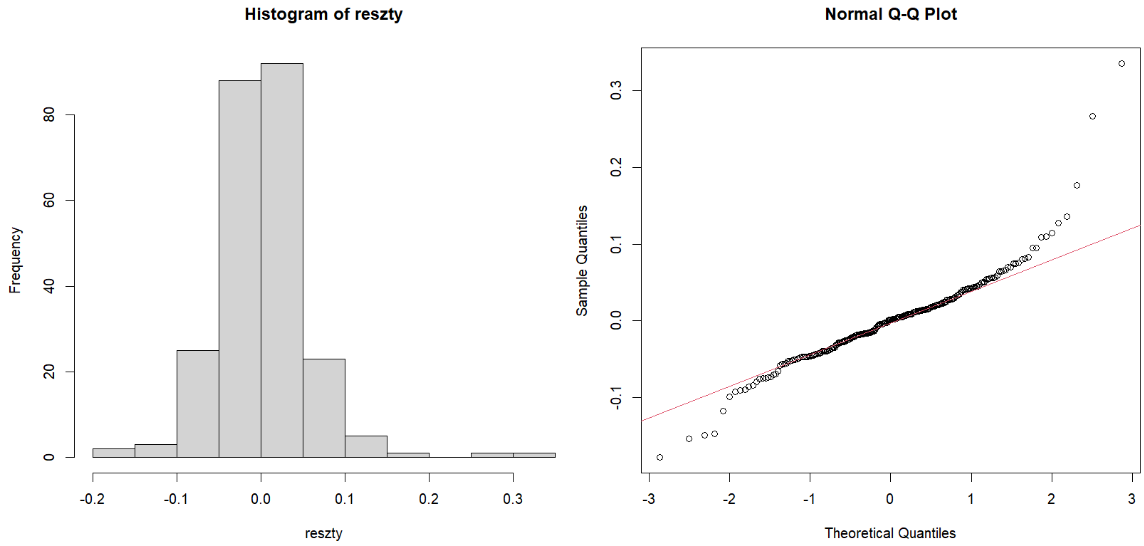
Zrobiono wykres diagnostyczny qq-plot dla reszt (rysunek 26b), czyli wykres porównujący kwantyle empiryczne oraz teoretyczne. Przy dobrze dopasowanym rozkładzie kwantyle muszą układać się obok prostej $y = x$. Analizując wykres qq-plot, widać że kwantyle układają się w miare obok prostej $y = x$, co wskazuje że rozkład reszt może być rozkładem normalnym $\epsilon \sim N(0, 0.05637^2)$

W celu badania normalności rozkładu reszt wykorzystano metodę Monte Carlo wraz ze statystyką Kołmogorowa-Smirnowa, statystykę Andersona-Darlinga oraz test Shapiro-Wilka. Wartości p-value dla każdego z tych testów umieszczono w tabeli 9. Ponieważ każda ze statystyk wynosi mniej niż przyjęty poziom istotności 5%, hipotezę o normalności rozkładu reszt odrzucono

Współczynnik determinacji R

KS	AD	Shapiro-Wilk
0.04645	4.617e-08	9.376e-11

Tabela 9: Wartości p-value statystyk



(a) Histogram reszt regresji

(b) Wykres diagnostyczny qq-plot reszt

Rysunek 26: Analiza reszt modelu regresji

Współczynnik determinacji jest jedną z miar jakości dopasowania modelu. Informuje jaka część zmienności zmiennej r_2 pokrywa się z korelacjami ze zmiennych zawartych w rozważanym modelu. W rozważanym przykładzie $R = 0.005307$ oznacza słabe wyjaśnienie zmienności dannych. Mając R=0.5% wynika że 95% danych log-zwrotów spółki JJB nie są wyjaśnione przez log-zwroty spółki KMR, można więc rozważyć dodanie do modelu innych zmiennych oprócz log-zwrotów spółki KMR

Istotnośc współczynników regresji

Testowana istotność współczynników regresji b_0, b_1, czyli dwie hipotezy $(i = 0, 1)$ $H_0: \ b_i = 0$ przeciwko $H_1 : b_i \neq 0$

Statystyką testową jest

$$t_i = \frac{b_i - \beta_i}{\mathrm{se}(\beta_i)}, i = 0, 1$$

gdzie wielkości $se(\beta_i)$ to błędy standardowe estymatorów Jeśli hipoteza zerowa jest prawdziwa, statystyka T ma rozkład t-Studenta o $n - 2 = 241 - 2 = 239$ stopniach swobody. Wykres tego rozkładu t-Studenta zamieszczono niżej (rysunek 27)

1. Zrobiona hipoteza zerowa H_0 $b_0 = 0$ przeciwko hipotezie alternatywnej H_1: $b_0 \neq 0$ Podstawiając dane w powyższy wzór otrzymano:

   $$t_0 = \frac{-0.002444519}{0.003632} \approx -0.673$$

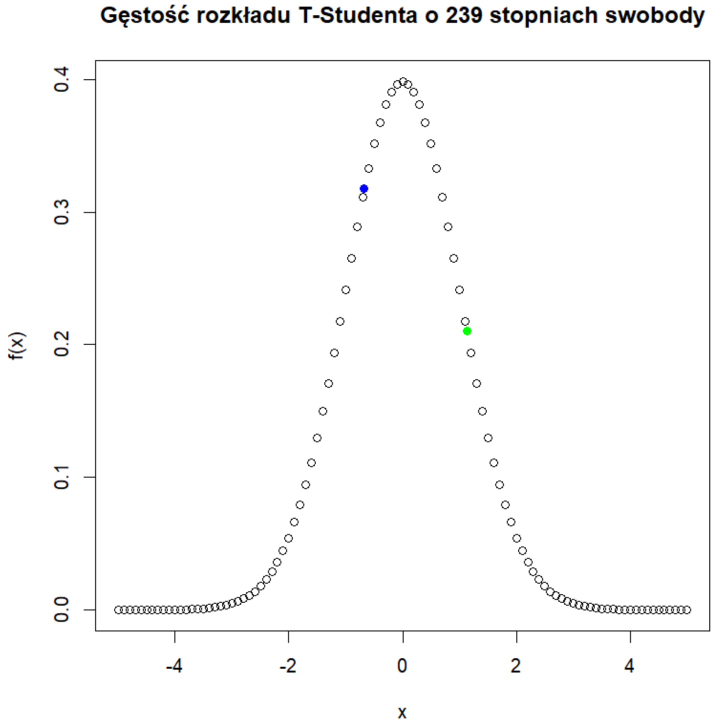
   Otrzymano wyniki $t - value = -0.673$, oraz $Pr(> |t|) = 0.502$. Wartość testu t_0 zamiszczono na wykresie kolorem niebieskim. Ponieważ $Pr(> |t|) = 0.502 > 5\%$ hipotezę, że współczynnik b_0 może być nieistotny przyjęto

2. Zrobiona hipoteza zerowa H_0': $b_1 = 0$ przeciwko hipotezie alternatywnej H_1': $b_1 \neq 0$

   Analogiczno dla wspołczynnika b_1 podstawiając dane w powyższy wzór otrzymano:

   $$t_1 = \frac{0.2187108}{0.193687} \approx 1.129$$

Otrzymano wyniki $t - value = 1.129$, oraz $Pr(>|t|) = 0.260$. Wartość testu t_1 zamiszczono na wykresie kolorem zielonym. Ponieważ $Pr(>|t|) = 0.260 > 5\%$ hipotezę, że współczynnik b_1 może być nieistotny przyjęto



Rysunek 27: Wykres rozkładu t-Studenta

4.2 Regresja dla uproszczonego modelu

Ponieważ okazało się, że każdy ze współczynników b_0, b_1 może być nieistotny, wykonano 2 regresji dla uproszczonych modeli

4.2.1 Uproszczony model przy założeniu $b_0 = 0$

Przy założeniu $b_0 = 0$, otrzymano model regresji $y = b_1 \cdot x = 0.22 \cdot x$. Wyniki regresji można zobaczyć na rysunku 28a

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.18062 -0.03365 -0.00157  0.02197  0.33300 

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
r1   0.2204     0.1935   1.139    0.256

Residual standard error: 0.05631 on 240 degrees of freedom
Multiple R-squared:  0.005378,	Adjusted R-squared:  0.001234 
F-statistic: 1.298 on 1 and 240 DF,  p-value: 0.2557
```

(a) Wyniki regresji dla uproszczoego modelu przy $b_0 = 0$

```
Residuals:
       Min         1Q     Median         3Q        Max 
-4.830e-17 -4.830e-17 -4.830e-17 -4.830e-17  1.159e-14 

Coefficients:
             Estimate Std. Error   t value Pr(>|t|)    
(Intercept) 2.187e-01  4.827e-17 4.531e+15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.494e-16 on 240 degrees of freedom
```

(b) Wyniki regresji dla uproszczoego modelu przy $b_1 = 0$

Rysunek 28: Wyniki regresji dla uproszczonych modeli

Duża wartość $Pr(>|t|) = 0.256 > 5\%$ wskazuje, że zostawiony współczynnik może być nieistotny. Współczynnik determinacji $R = 0.05631$ wskazuje, że zmienność jest niewyjaśniona tylko zmienną r_1

4.2.2 Uproszczony model przy założeniu $b_1 = 0$

Przy założeniu $b_1 = 0$, otrzymano model regresji $y = b_0 = -0.0024$. Wyniki regresji można zobaczyć na rysunku 28b

Mała wartość $Pr(>|t|) =< 2e^{-}16 < 5\%$ wskazuje, że zostawiony współczynnik jest istotnym. Współczynnik determinacji $R = 7.494e^{-}16$ wskazuje, że zmienność jest niewyjaśniona tylko stałą b_0

4.3 Predykcja log-zwrotów

Ponieważ rozważano 3 modeli regresji, zrobiono zostało 3 predykcji wielkości log-zwrotów R_2 gdy R_1 bedą na poziomie średniej z rozważanej próbki. W tym celu wyliczono średnią wartość log-zwrotów R_1. W tym przykłądzie wynosi ona $x^* = -0.0002403422$

1. Predykcja dla modelu $y = -0.0024 + 0.22 \cdot x$

   Predykcja $\hat{y}^*$ gdy wartość $x = \hat{x}^*$ to wartość: $\hat{y}^* = \beta_0 + \beta_1 x^*$. Podstawiając dane otrzymano $\hat{y}^* = -0.002497085$. Czyli jeżeli wartość log-zwrotów spółki KMR będzie wynosiła -0.0002403 wartość log-zwrotów spółki JJB będzie wynosiła -0.002497085

2. Predykcja dla modelu $y = 0.22 \cdot x$

   Predykcja $\hat{y}^*$ gdy wartość $x = \hat{x}^*$ to wartość: $\hat{y}^* = \beta_1 x^*$. Podstawiając dane otrzymano $\hat{y}^* = -5.256543e - 05$

3. Predykcja dla modelu $y = -0.0024$. Czyli jeżeli wartość log-zwrotów spółki KMR będzie wynosiła -0.0002403 wartość log-zwrotów spółki JJB będzie wynosiła $-5.256543e-05$

   Predykcja $\hat{y}^*$ gdy wartość $x = \hat{x}^*$ to wartość: $\hat{y}^* = \beta_0$. Podstawiając dane otrzymano $\hat{y}^* = -0.002444519$. Czyli jeżeli wartość log-zwrotów spółki KMR będzie wynosiła -0.0002403 wartość log-zwrotów spółki JJB będzie wynosiła -0.002444519

5 Podsumowanie

W projekcie została przeprowadzona analiza cen spółek KMR oraz JJB. Wyliczono statystyki opisowe, wyestymowano parametry rozkładów za pomocą funkcji w języku R z biblioteki *fitdistrplus*. Na podstawie analizy wykresów diagnostycznych CDF oraz qqplot, statystyk KS, CM, AD oraz kryteriów informacyjnych AIC, BIC dla każdej ze spółek wybrano najlepsze wykresy. Dla spółki KMR to jest rozkład $LN(6.09, 0.059)$. Dla spółki JJB rozkład $LN(0.67, 0.23)$. Dalej przeprowadzono testowanie hipotez tych rozkładów metodą Monte-Carlo za pomocą statystyki Kolmogorova-Smirnova. W wyniku tego testowania hipotezy o rozkładach logarytmiczno-normalnych spółek nie odrzucone.

W drugiej części projektu badano łączny rozkład log-zwrotów. W celu tej analizy w języku R wykorzystano biblioteki *mvtnorm* oraz *MASS*. Na podstawie wykresu rozrzutu zrobiona została hipoteza, że rozkład jest dwuwymiarowy normalny
$(X, Y) \sim N(-0.00024, -0.002497, 0.0188, 0.0564, 0.0728)$. Żeby się upewnić w prawdziwości tej hipotezy analizowano również kwadraty odległości Mahalanobisa. Po przeprowadzeniu tej analizy hipotezę o normalności rozkładu log-zwrotów

W trzeciej części projektu wykonano regresję dla rozkładu log-zwrotów. Na podstawie regresji zrobiona predykcja wartości log-zwrotów spółki KMR jeżeli wartość log-zwrotów spółki będą na poziomie średniej. Otrzymano wynik: jeżeli wartość log-zwrotów spółki KMR będzie wynosiła -0.0002403 wartość log-zwrotów spółki JJB będzie wynosiła -0.002