

# Projekt Egzaminacyjny

Maria Koren, Julian Jaroszyński

Listopad-Grudzień 2023

# Spis treści

<b>1</b>	<b>Wstęp</b>	<b>3</b>
<b>2</b>	<b>Analiza cen spółek</b>	<b>4</b>
2.1	Analiza cen spółki KMR.UK . . . . .	4
2.1.1	Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram	4
2.1.2	Statystyki opisowe . . . . .	5
2.1.3	Estymacja parametrów trzech rozkładów korzystając z estymatora największej wiarygodności (MLE) . . . . .	6
2.1.4	Wykresy diagnostyczne . . . . .	7
2.1.5	Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS . . . . .	8
2.2	Analiza cen spółki JJB . . . . .	10
2.2.1	Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram	10
2.2.2	Statystyki opisowe . . . . .	10
2.2.3	Estymacja parametrów trzech rozkładów korzystając z estymatora największej wiarygodności (MLE) . . . . .	11
2.2.4	Wykresy diagnostyczne . . . . .	12
2.2.5	Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS . . . . .	13
<b>3</b>	<b>Analiza łącznego rozkładu log-zwrotów</b>	<b>14</b>
3.1	Analiza rozkładów brzegowych . . . . .	14
3.1.1	Analiza rozkładów brzegowych spółki KMR . . . . .	14
3.1.2	Analiza rozkładów brzegowych spółki JJB . . . . .	15
3.2	Estymacja parametrów rozkładu dwuwymiarowego normalnego oraz analiza dobroci dopasowania . . . . .	17
3.2.1	Wykres rozrzutu z histogramami rozkładów brzegowych . . . . .	17
3.2.2	Wektor średnich, kowariancji, macierz korelacji, współczynnik korelacji	17
3.2.3	Wzór gęstości rozkładu dwuwymiarowego normalnego . . . . .	18
3.3	Analiza dopasowania rozkładu . . . . .	19
3.3.1	Porównanie wykresów rozrzutu w oparciu o wygenerowaną próbę . .	19
3.3.2	Mahalanobis . . . . .	20

# 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.

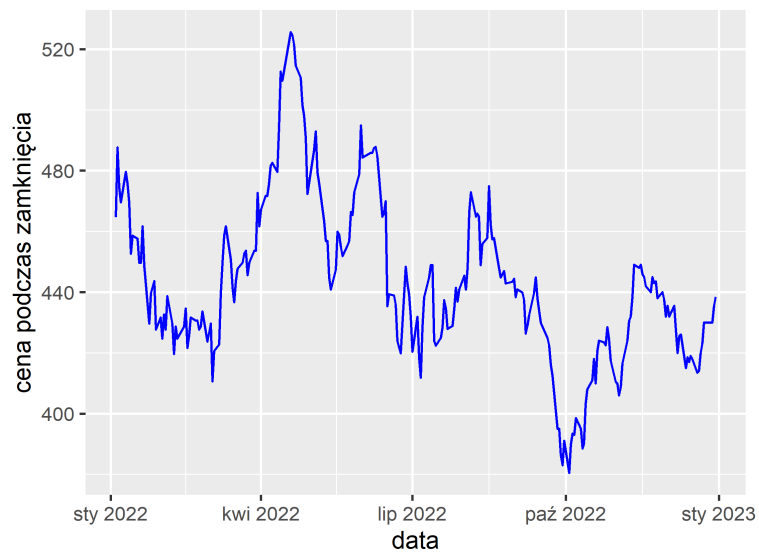
Jujubee S.A. to studio deweloperskie zajmujące się tworzeniem gier wideo, które ma swoim koncie takie tytuły jak: „FLASHOUT 3D”, „Suspect in Sight”, „Take Off – The Flight Simulator”, strategię czasu rzeczywistego „Realpolitiks”, grę przygodowo-dokumentalną „KURSK” oraz „Deep Diving Simulator”. Studio 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).

## 2 Analiza cen spółek

### 2.1 Analiza cen spółki KMR.UK

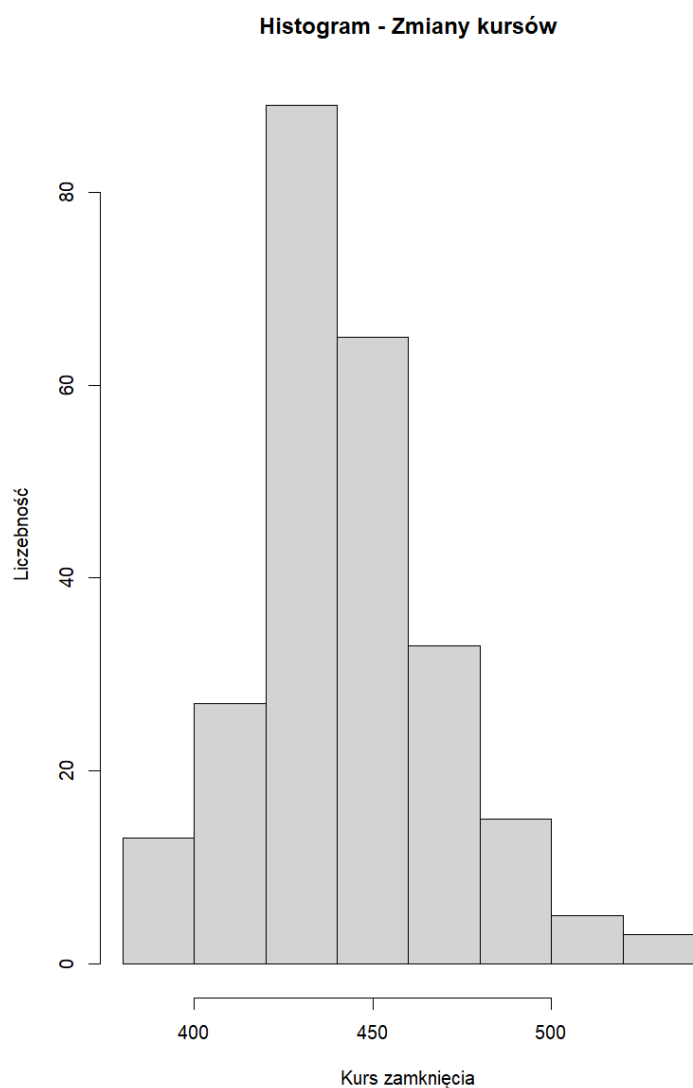
#### 2.1.1 Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram

Zrobiono wykres kursu zamknięcia pokazujący zmiany w czasie, rysunek 1



Rysunek 1: Cena podczas zamknięcia

Oraz histogram, pokazujący liczebność danych, rysunek 2



Rysunek 2: Histogram danych

### 2.1.2 Statystyki opisowe

Zostały obliczone następujące 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

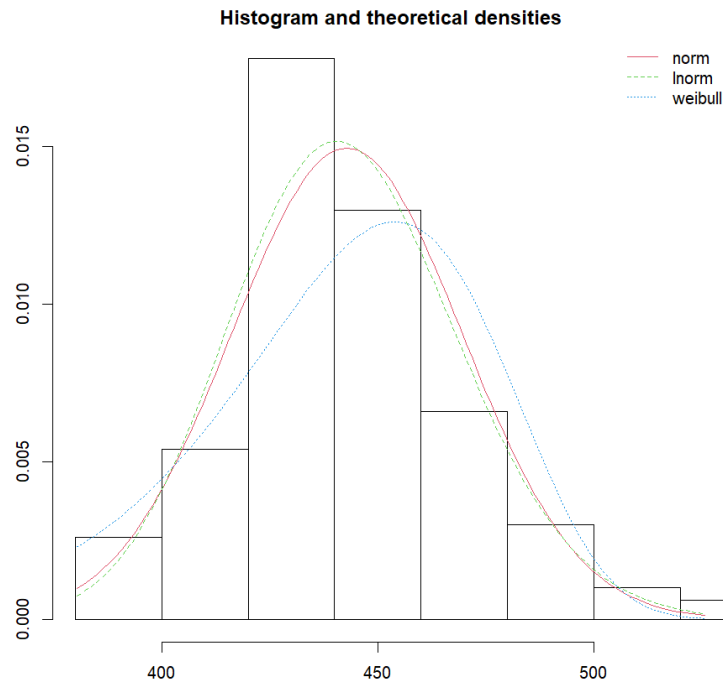
### Interpretacja wyników

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

- Otrzymana kurtoza mówi o 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)

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



Rysunek 3: Histogram wraz z estymowanymi rozkładami

Wyniki estymacji parametrów są przedstawione w tabeli 2

	$\mu$	$\sigma$
normalny	442.8741	26.67269
	$\mu$	$\sigma$
log-normalny	6.091499	0.05957788
	$a$	$\sigma$
weibulla	15.61307	455.8914

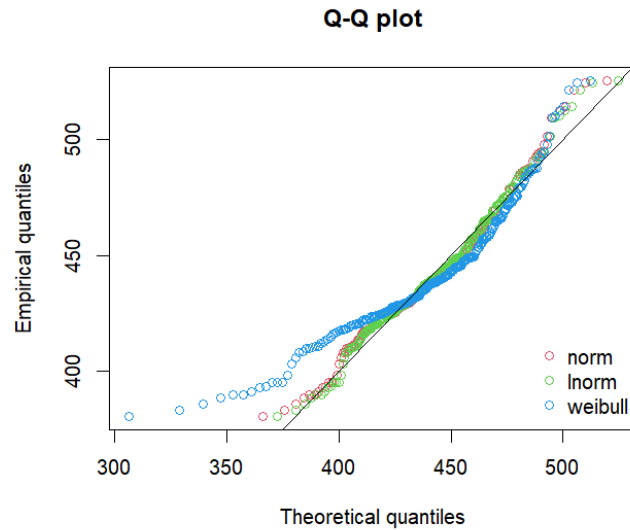
Tabela 2: Wyniki estymacji parametrów

Te wyniki oznaczają, że zostały dopasowane następujące rozkłady:

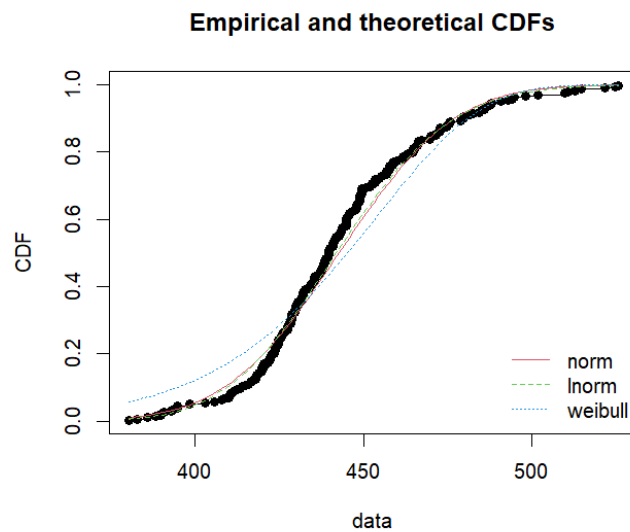
- $X \sim N(442.87, 26.67)$
- $X \sim LN(6.09, 0.059)$
- $X \sim W(15.61, 455.89)$

### 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 poziomej 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  $\alpha$ .

Najlepiej jest gdy te kwantyle są takie same bądź bardzo bliskie sobie. 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 teoretyczne 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 najbliższy do danych teoretycznych (ma najmniejszą odległość), czyli ma najmniejszą wartość statystyki. W kryteriach informacyjnych za najlepszy rozkład również jest uważany rozkład, mający najmniejszą wartość 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

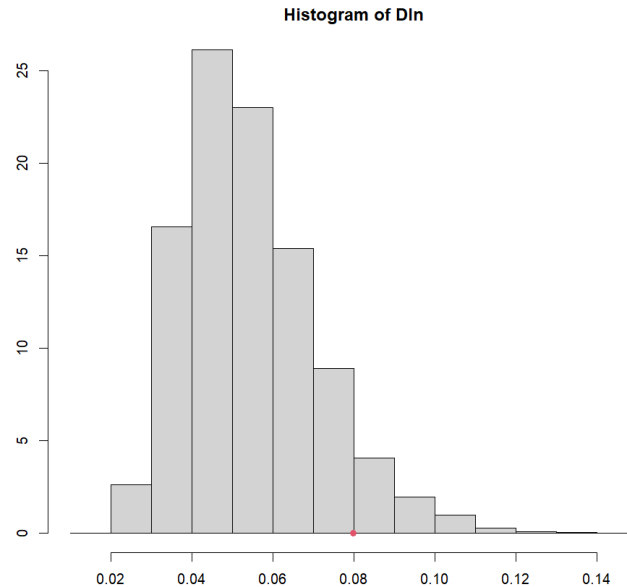
Zrobiona hipoteza  $H_0: F = LN(6.09, 0.0595)$  przeciwko hipotezie  $H_1: F$  nie jest równy  $LN(6.09, 0.059)$

Zgenerowano  $N = 10000$  próbek licznosci  $n$  (równej ilości danych) z rozkładu  $F_0 = LN(6.09, 0.059)$  wybranego wcześniej jako najlepszego rozkładu i obliczono odległość dystrybuant empirycznych od rozkładu  $F_0$  (wartosc statystyki  $D_n$ )



Obliczona również wartość statystyki dla rzeczywistych 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 teoretycznych

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

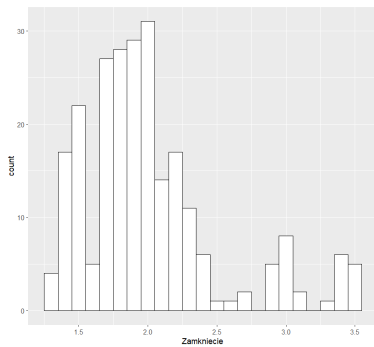
## 2.2 Analiza cen spółki JJB

### 2.2.1 Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram

Wykres przedstawiający kursy zamknięcia firmy Jujubee S.A. (JJB) wraz z histogramem



Rysunek 7: Cena podczas zamknięcia



Rysunek 8: Histogram danych

### 2.2.2 Statystyki opisowe

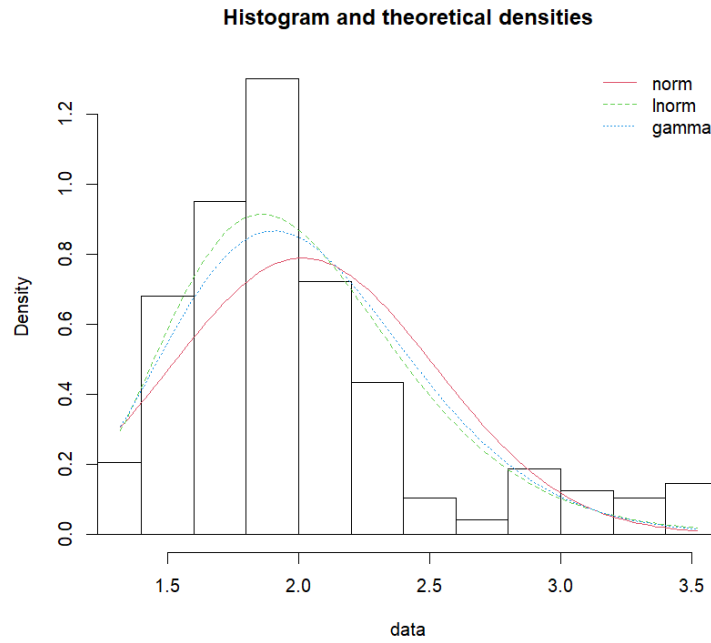
Jest to skośność prawostronna (liczba jest dodatnia), oznacz to że jest dużo przypadków mniejszych od średniej, firma miała parę dużych kursów podbijających średnią; Jest to rozkład leptokurtyczny, oznacz to że wartości są skoncentrowane wokół średniej oraz to że istnieje duża szansa na pojawienie się odstających obserwacji

	$\bar{x}$	odch. st.	skośność	kurtoza
Akcja	2.017201	0.5117912	1.319479	1.417186

Tabela 6: statystyki opisowe

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

Wystymowano wyniki trzech rozkładów: normalnego, log-normalnego oraz rozkładu Weibulla za pomocą estymatora MLE. Wyżej wymienione wykresy dodano do wcześniejszego histogramu, co widać za rysunku 9



Rysunek 9: Histogram wraz z estymowanymi rozkładami

Parametry trzech rozkładów

	estymator
średnia	2.0172008
sd	0.5107625

Tabela 7: parametry rozkładu normalnego

	estymator
średnia	0.6736043
sdlog	0.2306277

Tabela 8: parametry rozkładów log-normalnego

Te wyniki oznaczają, że zostały dopasowane następujące rozkłady:

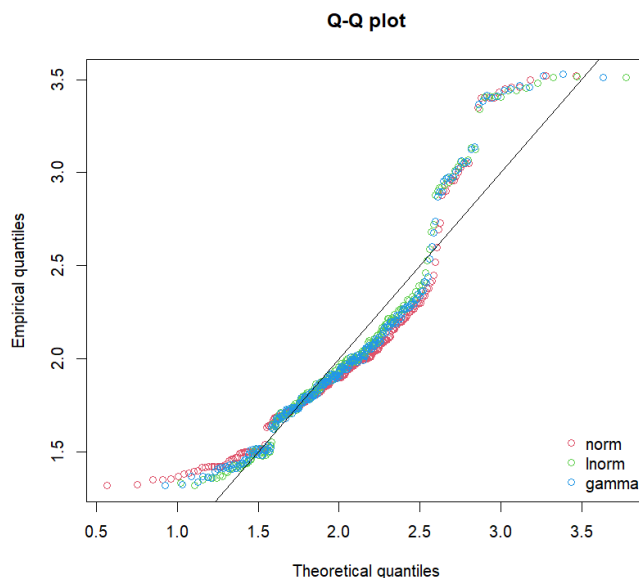
- $X \sim N(2.02, 0.51)$
- $X \sim LN(0.67, 0.23)$
- $X \sim \Gamma(17.95, 8.9)$

	estymator
kształt	17.954308
rate	8.900479

Tabela 9: parametry rozkładów gamma

## 2.2.4 Wykresy diagnostyczne

Zrobiono wykresy diagnostyczne qqplot oraz cdf



Rysunek 10: Wykres qq-plot

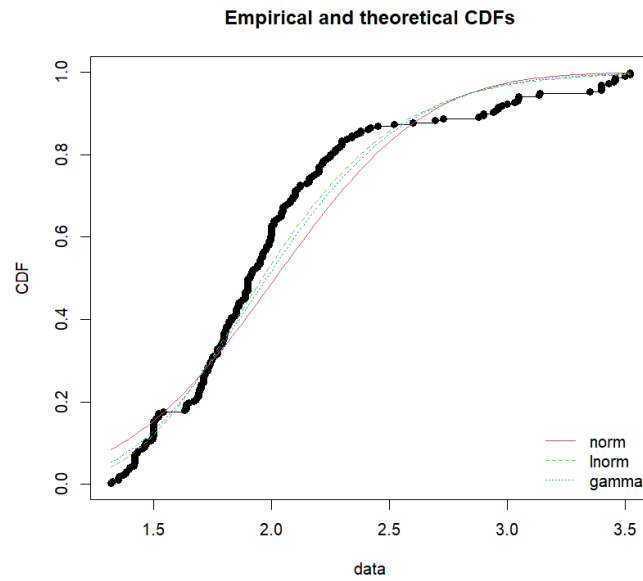
Na podstawie tych wykresów można wywnioskować że wykres lnorm jest najlepiej opisujący kursy zamknięcia, ale aby to dokładnie stwierdzić trzeba wykonać analizę statystyk KS, CM i AD oraz kryteriów informacyjnych AIC i BIC

### KS, CM, AD

Analiza dopasowania rozkładów za pomocą testów Kolmogorova-Smirnova, Cramera-von Misesa i Andersona-Darlinga pokazuje, że dla rozkładu normalnego (norm), log-normalnego (lnorm) i rozkładu gamma (gamma) wartości statystyk są odpowiednio 0.1465, 0.0946 i 0.1122 dla testu Kolmogorova-Smirnova, 1.5942, 0.5890 i 0.8612 dla testu Cramera-von Misesa, oraz 10.0248, 4.1675 i 5.7877 dla testu Andersona-Darlinga.

	norm	lnorm	gamma
Kolmogorov-Smirnov statistic	0.1464534	0.09463034	0.1121969
Cramer-von Mises statistic	1.5941731	0.58899369	0.8611976
Anderson-Darling statistic	10.0247605	4.16749359	5.7876581

Tabela 10: Goodness-of-fit statistics



Rysunek 11: Wykres cdf

### AIC, BIC

Porównanie kryteriów doboru modelu, takich jak Akaike's Information Criterion (AIC) i Bayesian Information Criterion (BIC), wskazuje, że dla rozkładu normalnego (norm), log-normalnego (lnorm) i rozkładu gamma (gamma) wartości wynoszą odpowiednio 376.0498, 315.5450 i 331.6349 dla AIC, oraz 383.0847, 322.5800 i 338.6698 dla BIC.

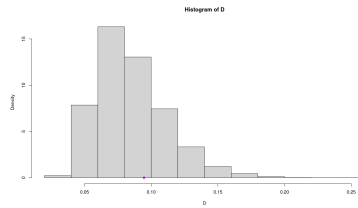
	norm	lnorm	gamma
Akaike's Information Criterion	376.0498	315.545	331.6349
Bayesian Information Criterion	383.0847	322.580	338.6698

Tabela 11: Goodness-of-fit criteria

Na podstawie wyżej wymienionych statystyk najlepszym jest rozkład log-normalny  $X \sim LN(0.67, 0.23)$

### 2.2.5 Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS

Zrobiona hipoteza  $H_0: F = LN(0.67, 0.23)$  przeciwko hipotezie  $H_1: F$  nie jest równy  $LN(0.67, 0.23)$ . Do przetestowania hipotezy został użyty sposób Kolmogorova-Smirnova (KS), którego wynikiem jest: 0.3202. Ta wartość jest większa niż 0.05, przez co hipoteza zerowa jest potwierdzona. Histogram znajduj się poniżej, rysunek 12



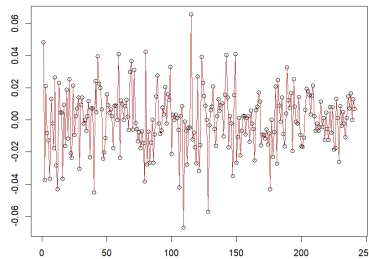
Rysunek 12: histogram metody ks

### 3 Analiza łącznego rozkładu log-zwrotów

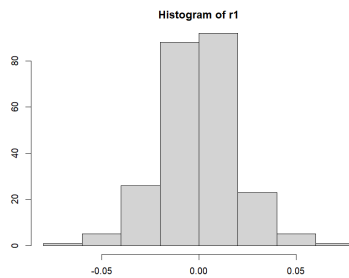
#### 3.1 Analiza rozkładów brzegowych

##### 3.1.1 Analiza rozkładów brzegowych spółki KMR

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



Rysunek 13: Wykres log-zwrotów spółki KMR

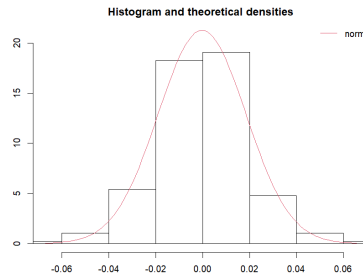


Rysunek 14: Wykres log-zwrotów spółki KMR

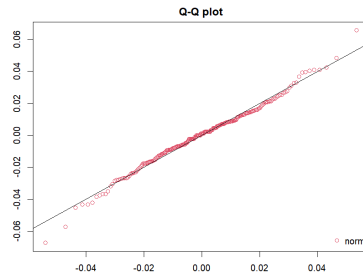
Został dopasowany do tych danych rozkład normalny  $X \sim N(-0.0023, 0.018)$ , rysunek 15

Zrobiono wykresy diagnostyczne cdf oraz qqplot, rysunki 16, 17

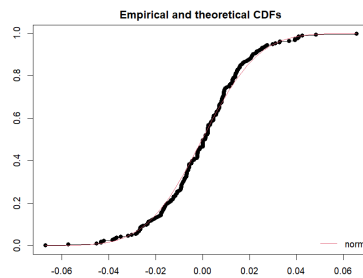
Przeprowadzono również test równości metodą Monte-Carlo. Skorzystano ze statystyki Kolmogorova-Smirnova. Wynik p-value otrzymany w tym teście jest równy 0.5656, co oznacza przyjęcie hipotezy że rozkład log-zwrotów jest  $X \sim N(-0.0023, 0.018)$ , rysunek 18



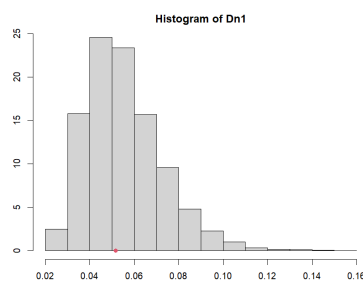
Rysunek 15: Wykres log-zwrotów wraz z dopasowanym rozkładem normalnym spółki KMR



Rysunek 16: Wykres qqplot dla log-zwrotów spółki KMR



Rysunek 17: Wykres cdf dla log-zwrotów spółki KMR

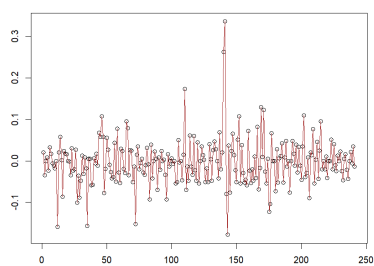


Rysunek 18: Histogram danych w teście

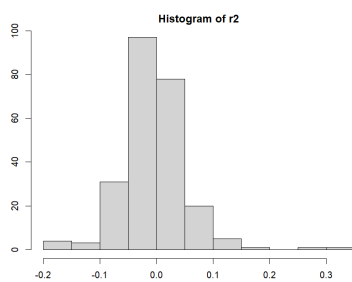
### 3.1.2 Analiza rozkładów brzegowych spółki JJB

Na rysunkach 19 oraz 20 są przedstawione odpowiednio wykresy log-zwrotów oraz histogram

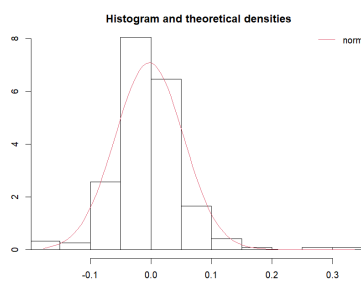
Został dopasowany do tych danych rozkład normalny  $Y \sim N(-0.0025, 0.057)$



Rysunek 19: Wykres log-zwrotów spółki JJB

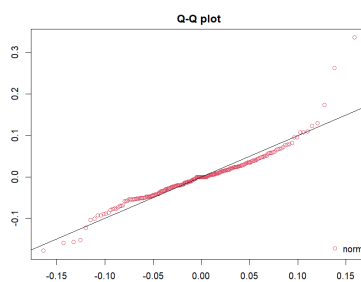


Rysunek 20: Wykres log-zwrotów spółki JJB



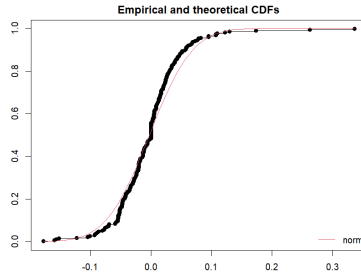
Rysunek 21: Wykres log-zwrotów wraz z dopasowanym rozkładem normalnym spółki JJB

Zrobiono wykresy diagnostyczne cdf oraz qqplot, rysunki 22, 23



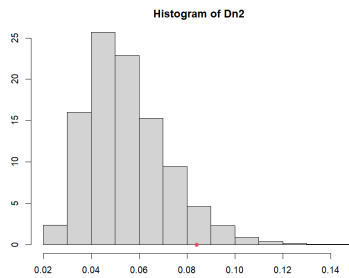
Rysunek 22: Wykres qqplot dla log-zwrotów spółki JJB





Rysunek 23: Wykres cdf dla log-zwrotów spółki JJB

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

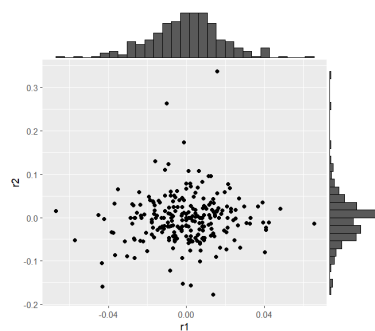


Rysunek 24: Histogram danych w teście

### 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 25



Rysunek 25: Wykres rozrzutu z histogramami brzegowymi

#### 3.2.2 Wektor średnich, kowariancji, macierz korelacji, współczynnik korelacji

Zostały obliczone następujące estymatory:  
Wektor średnich:

$\mu_1$	$\mu_2$
-0.0002403422	-0.0024970847

Tabela 12: Wektor średnich

Macierz kowariancji (rysunek 13), współczynnik kowariancji

$$\text{cov}(X, Y) = 7.719315e - 05$$

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

Tabela 13: Macierz kowariancji

Macierz korelacji (rysunek 14), współczynnik korelacji

$$\rho = 0.0728$$

	$r_1$	$r_2$
$r_1$	1.00000000	0.07284753
$r_2$	0.07284753	1.00000000

Tabela 14: Macierz korelacji

### 3.2.3 Wzór gęstości rozkładu dwuwymiarowego normalnego

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

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

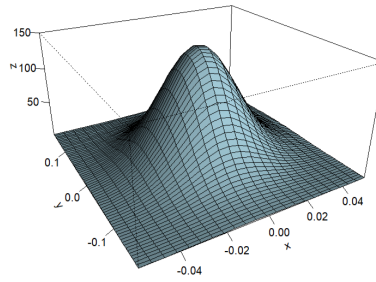
Inaczej zapisując:

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

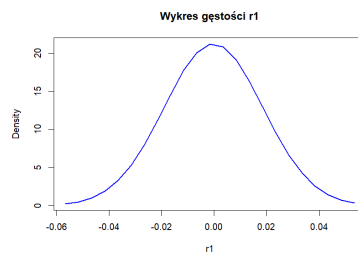
Rozkład brzegowy dla spółki KMR:  $X \sim N(-0.000240342, 0.01878686)$  Rozkład brzegowy dla spółki JJB:  $Y \sim N(-0.0024970847, 0.05640397)$  Wykres gęstości rozkładu łącznego (rysunek 26):

Wykres gęstości rozkładu brzegowego KMR (rysunek 27):

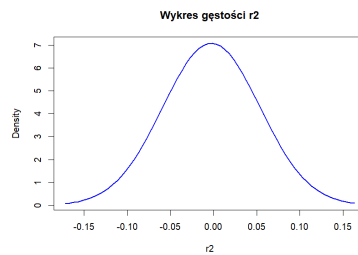
Wykres gęstości rozkładu brzegowego JJB (rysunek 28):



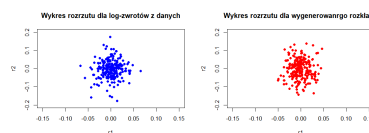
Rysunek 26: Wykres gęstości rozkładu łącznego



Rysunek 27: Wykres gęstości X



Rysunek 28: Wykres gęstości Y



Rysunek 29: Wykresy rozrzutu wygenerowanej próby oraz danych

### 3.3 Analiza dopasowania rozkładu

#### 3.3.1 Porównanie wykresów rozrzutu w oparciu o wygenerowaną próbę

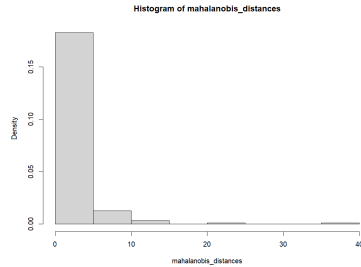
Z rozkładu  $(X, Y) \sim N(-0.000240342, -0.0024970847, 0.01878686, 0.05640397, 0.0728)$  została wygenerowana próba licząca 242 (poniższą liczbę mają dane log-zwrotów spółek), rysunek 29. Z wykresu robiona hipoteza, że rozkład log-zwrotów może być

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

. Ale żeby upewnić zrobiono testowanie hipotezy, że kwadrat odległości Mahalanobisa wektora log-zwrotów od średniej, mają rozkład  $\chi^2(2)$

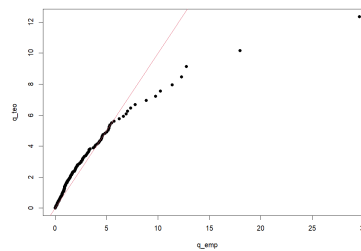
### 3.3.2 Mahalanobis

Obliczono kwadraty odległości Mahalanobisa dla każdej pary cen od średniej  $\hat{\mu}$ . Oraz zrobiono histogram, rysunek 30



Rysunek 30: Histogram kwadratów odległości Mahalanobisa

Zrobiono wykres diagnostyczny qqplot: porównywanie kwantyli empirycznych oraz kwantyli z rozkładu  $\chi^2$



Rysunek 31: QQplot kwantyli empirycznych oraz z rozkładu  $\chi^2$

Punkty w miarę układają się blisko prostej  $y = x$  co wskazuje, że rozkład kwadratów odległości Mahalanobisa mogą być z rozkładu  $\chi^2$

Żeby bardziej się upewnić przeprowadzono test zgodności oparty na statystyce Kolmogorowa-Smirnowa. Wynik testu  $p\text{-value} = 0.0001707 < 5\%$  zatem odrzucono hipotezę, że kwadraty odległości Mahalanobisa mają rozkład  $\chi^2$ , co skutkuje też odrzuceniem hipotezy o normalności rozkładu log-zwrotów