

# Projekt Egzaminacyjny

Karol Krawczykiewicz, Piotr Maszczak

Listopad 2022 – Styczeń 2023

## 1 Wstęp

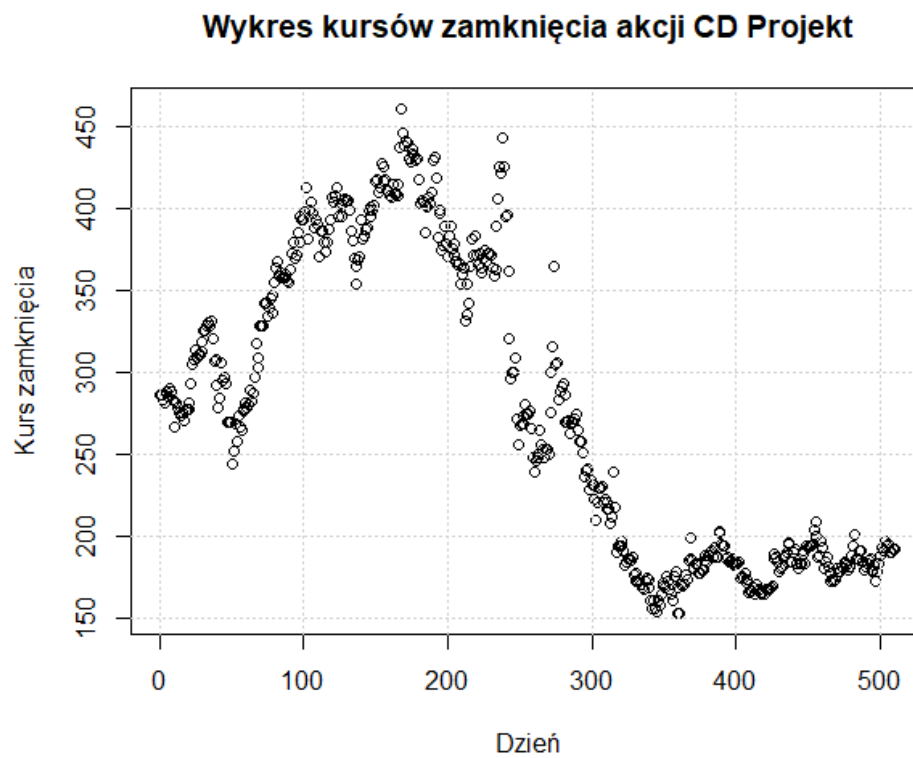
Następująca analiza dotyczy dwóch spółek działających na rynku gier wideo

CD Projekt SA (CDR) - Grupa działa w dynamicznie rozwijającej się branży elektronicznej rozrywki. Opiera się na dwóch silnych fundamentach – produkcji gier w ramach studia deweloperskiego CD Project Red oraz globalnej cyfrowej dystrybucji realizowanej przez serwis GOG.com.

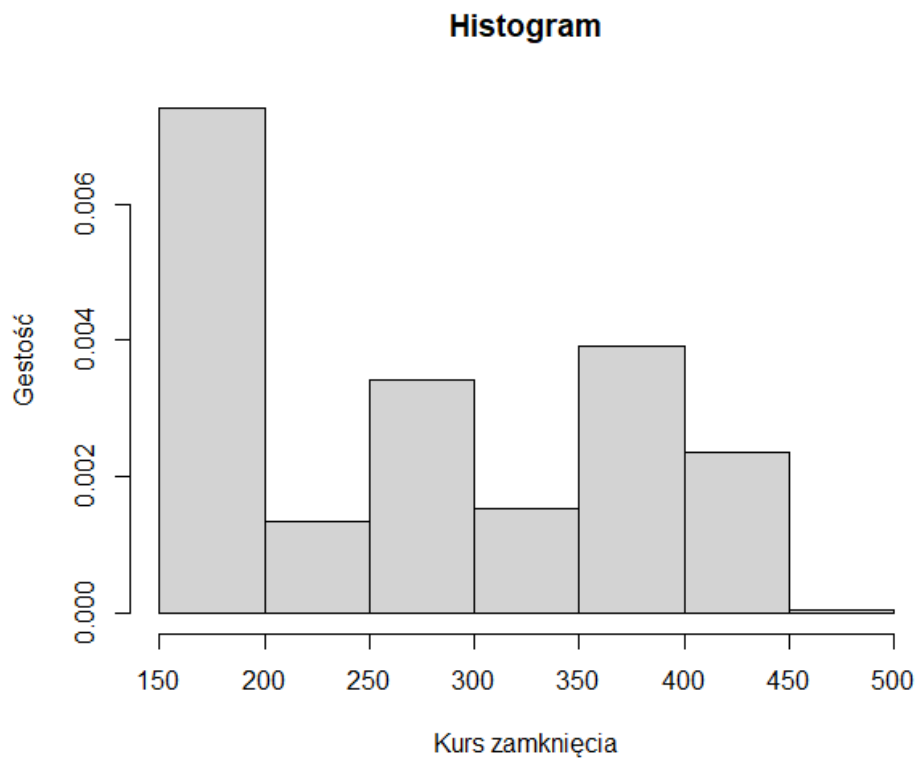
11 bit studios SA (11B) - Deweloper multiplatformowych gier sprzedawanych na całym świecie. Spółka zajmuje się każdym etapem tworzenia gier samodzielnie. Najbardziej znane gry to: Frostpunk i This War of Mine. Oprócz produkcji gier grupa zajmuje się również wydawaniem gier zewnętrznych oraz prowadzi internetową platformę sprzedaży.

## 2 Analiza cen spółek

### 2.1 Spółka CD Projekt



Rysunek 1: Wykres akcji w przedziale czasowym od Stycznia 2020 do Grudnia 2021



Rysunek 2: Histogram gęstości

	$\bar{x}$	odch. st.	skośność	kurtoza
Akcja	277.6719	90.62922	0.2497853	1.56275

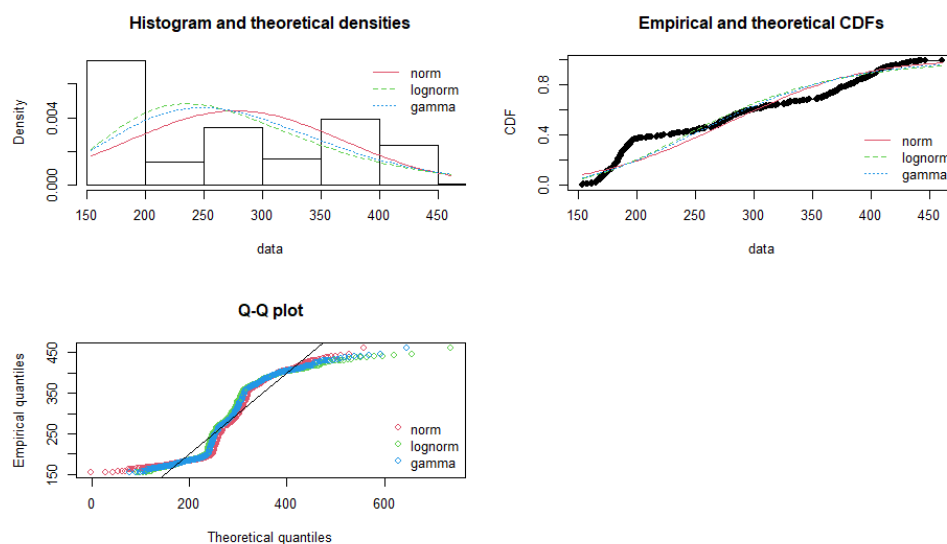
Współczynnik skośności powyżej 0 świadczy o prawostronnej asymetrii rozkładu (wydłużone jest prawe ramię rozkładu). Kurtoza większa od 0 oznacza iż rozkład jest bardziej wysmukły niż normalny, większe skupienie wartości wokół średniej.

Wystymowane parametry dla badanych rozkładów przy wykorzystaniu estymatora największej wiarygodności (MLE):

- Rozkład normalny - mean 277.67190, sd 90.54032
- Rozkład log-normalny - meanlog 5.5719177, sdlog 0.3322444
- Rozkład gamma - shape 9.33484470 , rate 0.03362056

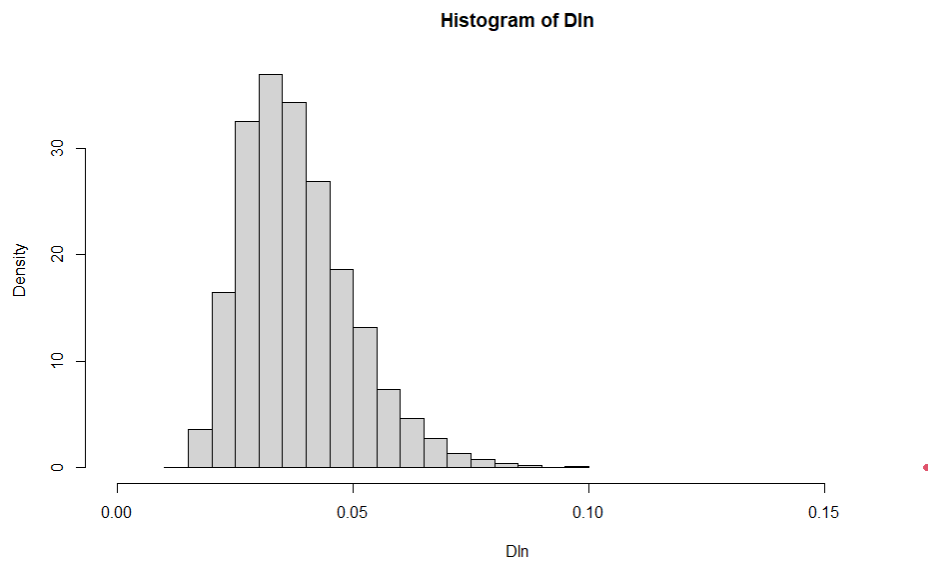
	Normalny	Log-normalny	Gamma
Kolmogorov-Smirnov statistic	0.1777046	0.1715694	0.1739537
Cramer-von Mises statistic	2.9612965	2.9246380	2.8490879
Anderson-Darling statistic	19.7989554	19.3686342	19.1126404
Akaike's Information Criterion	6047.228	6010.751	6013.699
Bayesian Information Criterion	6055.697	6019.220	6022.168

Z powyższych danych wynika, iż rozkład Log-Normalny najlepiej opisuje kursy zamknięcia spółki CDR. Trzy z pięciu przypadków optuje za rozkładem Log-Normalny, a dwa za rozkładem Gamma.



Rysunek 3: Wykresy diagnostyczne: Funkcje gęstości, Funkcja dystrybuanty, Kwantyle

Na podstawie wykresów diagnostycznych można wywnioskować, że prawdopodobnie żaden z rozkładów nie będzie w dobrym stopniu odwzorowywał danych.



Po przetestowaniu hipotezy zerowej o równości dystrybuant, przy wykorzystaniu metody Monte-Carlo i statystyk Kołmogorowa-Smirnowa, uzyskano wyniki:

$$Dn = 0.1715694$$

$$p = \text{length}(Dn[Dn > dn])/N = 0$$

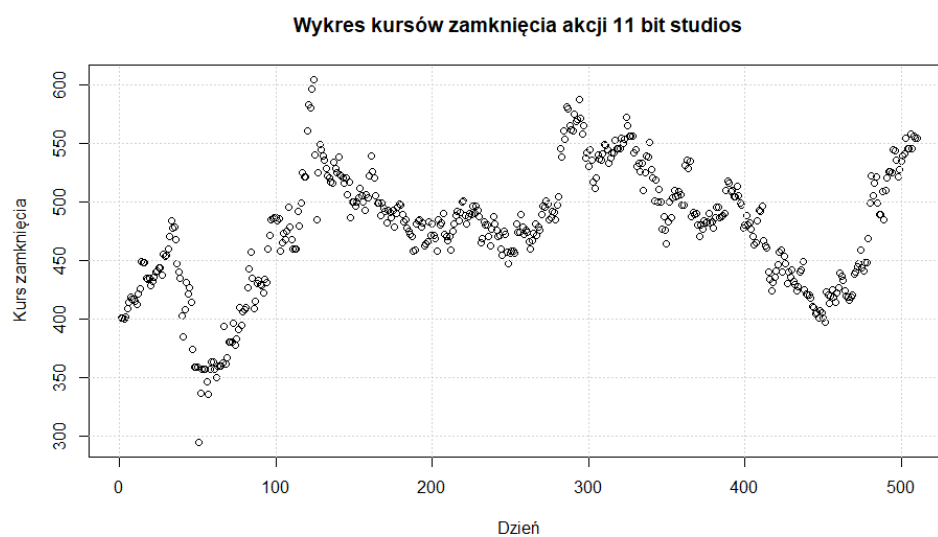
gdzie  $Dn$  to odległość dystrybuant empirycznych od rozkładu logarytmiczno-normalnego,  $dn$  to wartość statystyki  $Dn$ , dla danych akcji spółki,  $N = 10000$

Wartość  $p$  jest mniejsza od przyjętego poziomu istotności,

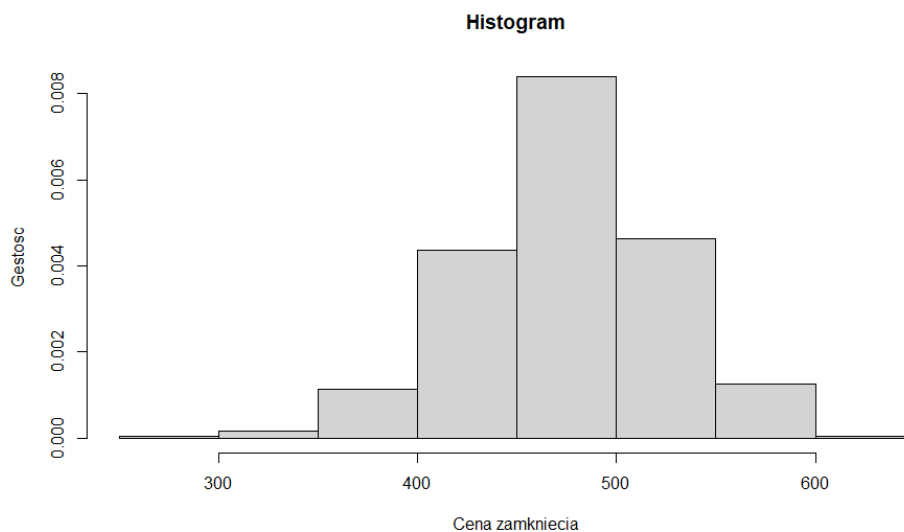
$$\alpha = 5\%$$

Zatem hipotezę o równości dystrybuant odrzucamy.

## 2.2 Spółka 11 BIT STUDIOS



Rysunek 4: Wykres akcji w przedziale czasowym: Styczeń 2020 do Grudnia 2021



Rysunek 5: Histogram gęstości

	$\bar{x}$	odch. st.	skośność	kurtoza
Akcja	476.8853	51.16337	-0.385939	3.071287

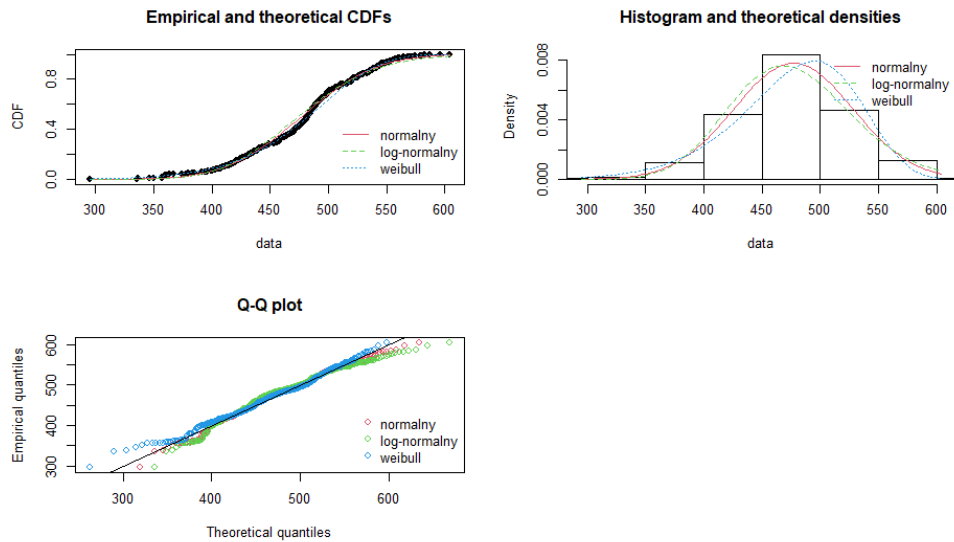
Współczynnik skośności poniżej 0 świadczy o lewostronnej asymetrii rozkładu (wydłużone jest lewe ramię rozkładu). Kurtoza większa od 0 oznacza iż rozkład jest bardziej wysmukły niż normalny, większe skupienie wartości wokół średniej.

Wyestymowane parametry dla badanych rozkładów przy wykorzystaniu estymatora największej wiarygodności (MLE):

- Rozkład normalny - mean 476.88529, sd 51.11319
- Rozkład log-normalny - meanlog 6.1612582, sdlog 0.1111746
- Rozkład Weibulla - shape 10.73364, rate 499.28283

	Normalny	Log-normalny	Weibull
Kolmogorov-Smirnov statistic	0.06995126	0.0929663	0.06613342
Cramer-von Mises statistic	0.38880200	0.7849020	0.30039685
Anderson-Darling statistic	2.12780717	4.4814222	1.55241020
Akaike's Information Criterion	5464.041	5495.215	5455.947
Bayesian Information Criterion	5472.510	5503.683	5464.416

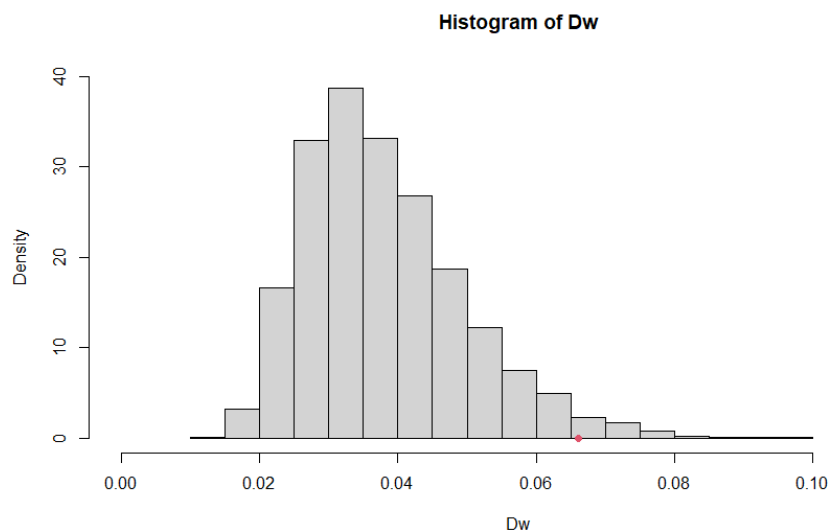
Z powyższych danych wynika, iż rozkład Weibulla najlepiej opisuje kursy zamknięcia spółki 11 bit studios. Wszystkie z pięciu przypadków wskazują za rozkładem Weibulla.



Rysunek 6: Wykresy diagnostyczne: Funkcja dystrybuanty, Funkcje gęstości, Kwantyle

Na podstawie wykresów diagnostycznych można wywnioskować, że rozkład Weibulla najbardziej odwzorowuje dane.





Po przetestowaniu hipotezy zerowej o równości dystrybuant, przy wykorzystaniu metody Monte-Carlo i statystyk Kołmogorowa-Smirnowa, uzyskano wyniki:

$$Dw = 0.06613342$$

$$p = \text{length}(Dw[Dw > dw])/N = 0.0238$$

gdzie  $Dw$  to odległość dystrybuant empirycznych od rozkładu Weibulla,  $dw$  to wartość statystyki  $Dw$ , dla danych akcji spółki,  $N = 10000$

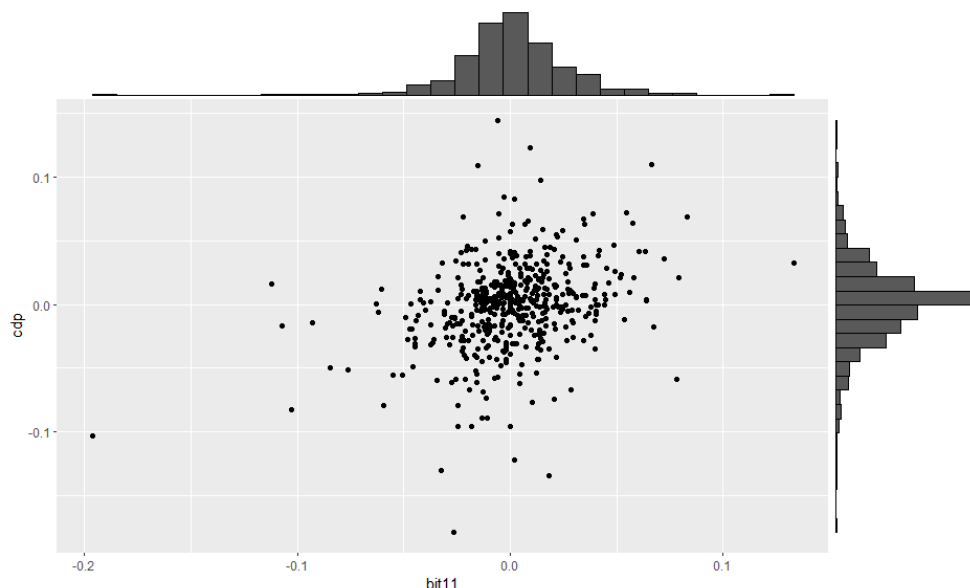
Wartość  $p$  jest mniejsza od przyjętego poziomu istotności,

$$\alpha = 5\%$$

Zatem hipotezę o równości dystrybuant odrzucamy.

### 3 Analiza łącznego rozkładu log-zwrotów

W tym rozdziale wykonujemy analizę dziennych log-zwrotów wcześniej wymienionych spółek: 11 bit studios oraz CD Projekt.



Rysunek 7: Histogram rozkładów brzegowych

Z powyższego wykresu widać, że dane są znacznie rozproszone. Zauważalna część punktów skupia się wokół wektora średnich. Występuje tu dodatnia zależność. Duża ilość odstających wartości.

- Wektor średnich  $\hat{\mu} = (\text{bit11}: 0.000641, \text{cdp}: -0.00078)$
- Kowariancja - cov 0.000311
- Współczynnik korelacji - cor 0.32332

Wektor średnich  $\hat{\mu}$  świadczy o średniej wartości poszczególnych zmiennych losowych.

Kowariancja jest bliska 0, oznacza to, że zmienne losowe nie są ze sobą związane w żaden szczególny sposób, istnieje słaby związek pomiędzy zmiennymi losowymi. W takim przypadku zmiany w jednej zmiennej nie są skorelowane z zmianami w drugiej zmiennej, nie występuje żadna zależność liniowa.

Jeśli współczynnik korelacji wynosi 0.32, oznacza to, że istnieje słaba, dodatnia korelacja pomiędzy dwoma zmiennymi losowymi. To znaczy, że wzrost jednej zmiennej wiąże się z wzrostem drugiej zmiennej, ale związek ten nie jest silny.

Wzór na macierz kowariancji:

$$\Sigma = \begin{bmatrix} Var(X) & cov(X, Y) \\ cov(Y, X) & Var(Y) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

gdzie  $\sigma_1, \sigma_2$  to odchylenia standardowe i  $\rho$  współczynnik korelacji

Macierz kowariancji dla analizowanych danych:

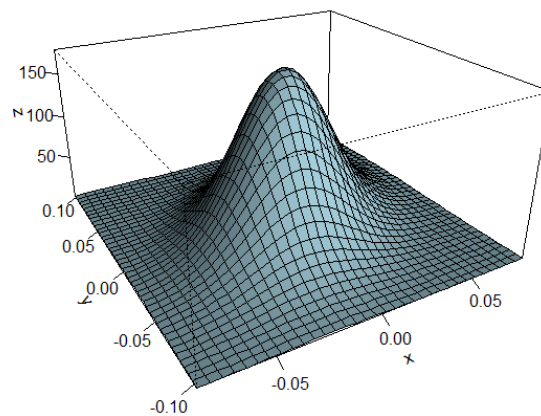
$$\hat{\Sigma} = \begin{bmatrix} 0.00077 & 0.00031 \\ 0.00031 & 0.00117 \end{bmatrix}$$

Wzór na macierz korelacji:

$$\rho = \begin{bmatrix} 1 & \frac{cov(X,Y)}{\sigma_X\sigma_Y} \\ \frac{cov(X,Y)}{\sigma_X\sigma_Y} & 1 \end{bmatrix}$$

Macierz korelacji dla analizowanych danych:

$$\begin{bmatrix} 1 & 0.32332 \\ 0.32332 & 1 \end{bmatrix}$$



Rysunek 8: Wykres gęstości

Gęstość rozkładu normalnego o wyestymowanych parametrach jest zapisywana w następujący sposób:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]\right)$$

gdzie  $x$  i  $y$  są wektorami zmiennych losowych,  $\hat{\mu}$  jest wektorem średnich,  $\sigma_1, \sigma_2$  to odchylenia standardowe i  $\rho$  współczynnik korelacji.

Gęstość rozkładu normalnego dla analizowanych danych:

$$f(x, y) = 177.23 \cdot \exp(-725.15(x-0.00064)^2 + 380.07(x-0.00064)(y+0.00078) - 477.24(y+0.00078)^2)$$

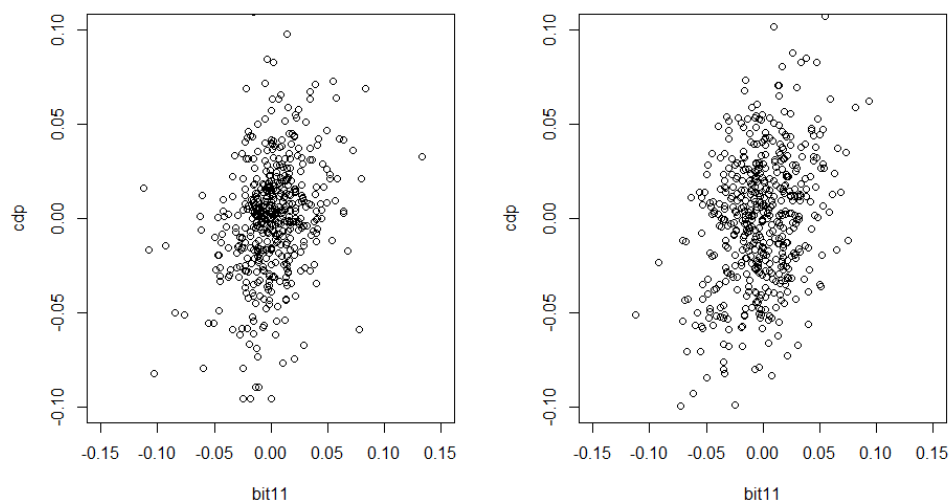
Gęstość rozkładów brzegowych jest zapisywana w następujący sposób:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Gęstość rozkładów brzegowych dla analizowanych danych:

$$f_{11b}(x) = 14.39388 \cdot \exp\left(-\frac{(x-0.00064)^2}{0.00154}\right)$$

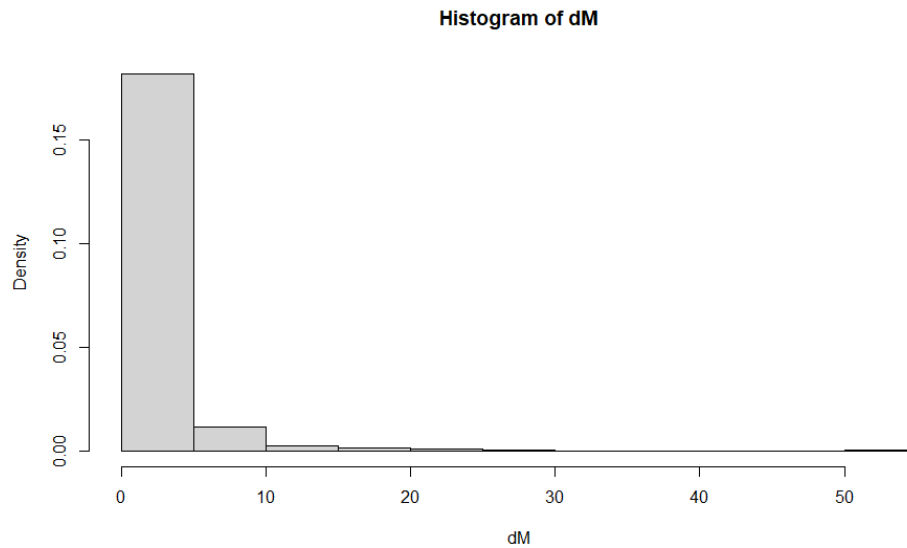
$$f_{cdr}(x) = 11.65201 \cdot \exp\left(-\frac{(x+0.00078)^2}{0.00234}\right)$$



Rysunek 9: Wykresy rozrzutów: pierwszy -otrzymany na podstawie danych, drugi - w oparciu o wygenerowaną próbę

Po wygenerowaniu próbie liczności danych = 509, na powyższym wykresie widać, że nasze dane nie przypominają rozkładu normalnego. Skupienie danych na obu rozkładach jest różne. Jest też wiele wartości odstających.

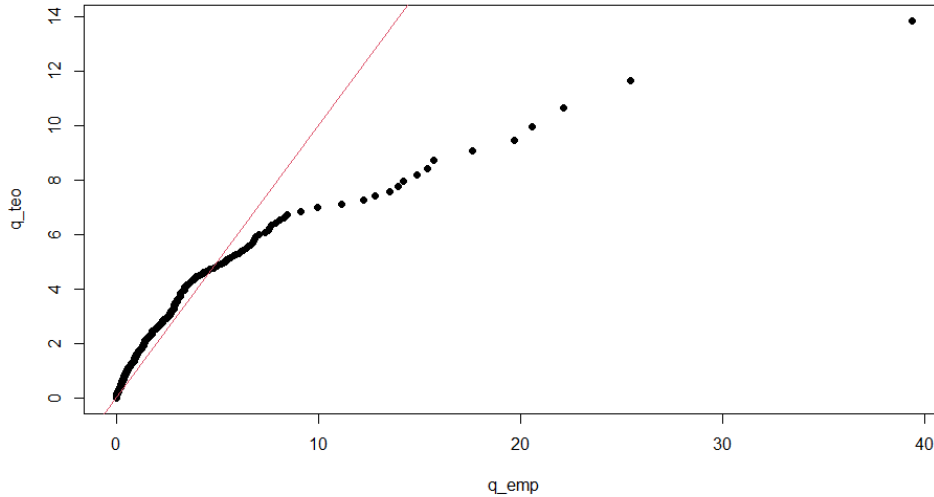
Badanie hipotezy mówiącej o tym, że kwadrat odległości Mahalanobisa wektora cen od średniej ma rozkład  $\chi^2$  (drugiego stopnia).



Rysunek 10: Histogram kwadratów odległości Mahalanobisa

Zasadnicza większość wartości znajduje się w pierwszym przedziale. Histogram nie ma zbliżonego kształtu do rozkładu  $\chi^2$  (drugiego stopnia) - ma dużą wartość na początku, a następnie bardzo małe, wręcz bliskie zeru (nie jest tak "płynny" jak powinien być).

Po przeprowadzeniu analizy statystycznej rozkładów kursów ustaliliśmy, iż najbardziej odpowiedni model dla opisanie tych danych jest rozkład normalny dwuwymiarowy. W konsekwencji, kwadrat odległości Mahalanobisa od średniej powinien być rozłożony według rozkładu  $\chi^2(2)$ .



Rysunek 11: Wykres diagnostyczny typu QQ-plot

Wykres kwantyl-kwantyl (Q-Q plot) to wykres, który porównuje rozkład kwantyli danych empirycznych z rozkładem kwantyli danych teoretycznych. Na osi x umieszcza się kwantyle teoretyczne, a na osi y kwantyle empiryczne. Analiza Q-Q plot wykazała, że rozkład  $\chi^2$  (stopnia drugiego) nie jest odpowiedni do opisu naszych danych empirycznych. Porównanie danych teoretycznych (przedstawionych przez czerwoną linię) z danymi empirycznymi (przedstawionymi przez czarne punkty) wykazało istotne odchylenia, co sugeruje potrzebę poszukiwania innego rozkładu statystycznego dla opisu tych danych.

Za pomocą testu zgodności badamy czy dane empiryczne pochodzą z rozkładu  $\chi^2$ . Celem testu jest zweryfikowanie hipotezy zerowej, oraz hipotezy alternatywnej, która głosi, że dane nie pochodzą z tego rozkładu. W teście uzyskamy p-value, które pozwali na ocenę poziomu istotności statystycznej odchyień danych od rozkładu teoretycznego.

Po przetestowaniu hipotezy że kwadrat odległości Mahalanobisa wektora cen od średniej mają rozkład  $\chi^2(2)$ , przy wykorzystaniu metody Monte-Carlo i statystyk Kołmogorowa-Smirnowa, uzyskano wyniki:

$$D = 0.16836$$

$$p = 5.882e^{-13}$$

Wartość  $p$  jest mniejsza od przyjętego poziomu istotności:

$$\alpha = 5\%$$

Oznacza to, że jest bardzo mało prawdopodobne, że dane empiryczne pochodzą z rozkładu  $\chi^2(2)$ . Wartość  $p$ -value jest mniejsza niż poziom istotności  $\alpha$  (0.05), zatem odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną. Ponadto, przez to odrzucamy również hipotezę o normalności rozkładu log-zwrotów dla badanych spółek.

## **4   Rozdział 3**

## **5   Podsumowanie**