

# Projekt Egzaminacyjny

Karol Krawczykiewicz, Piotr Maszczak

Listopad 2022 – Styczeń 2023

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>3</b>
<b>2</b>	<b>Analiza cen spółek</b>	<b>3</b>
2.1	Spółka CD Projekt . . . . .	3
2.2	Spółka 11 BIT STUDIOS . . . . .	7
<b>3</b>	<b>Analiza łącznego rozkładu log-zwrotów</b>	<b>11</b>
<b>4</b>	<b>Regresja liniowa dla log-zwrotów</b>	<b>17</b>
<b>5</b>	<b>Podsumowanie</b>	<b>24</b>

# 1 Wstęp

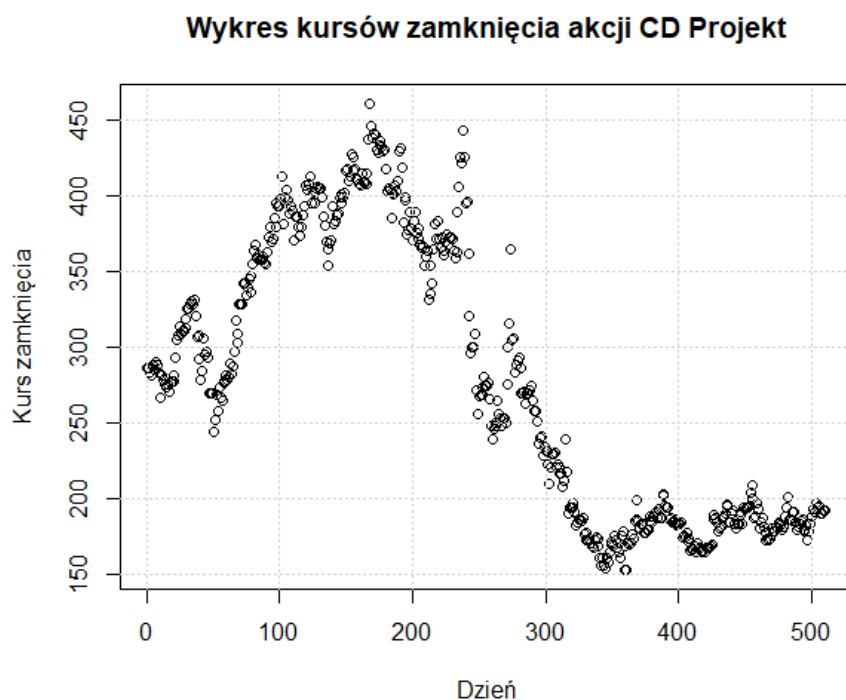
Następująca analiza dotyczy dwóch spółek działających na rynku gier wideo.

CD Projekt SA (CDR) to grupa działająca w dynamicznie rozwijającej się branży elektronicznej rozrywki. Z silnym naciskiem na produkcję gier przez swoje studio deweloperskie CD Project Red oraz globalną dystrybucję cyfrową za pośrednictwem platformy GOG.com, firma osiągnęła duże sukcesy i jest szanowana na całym świecie.

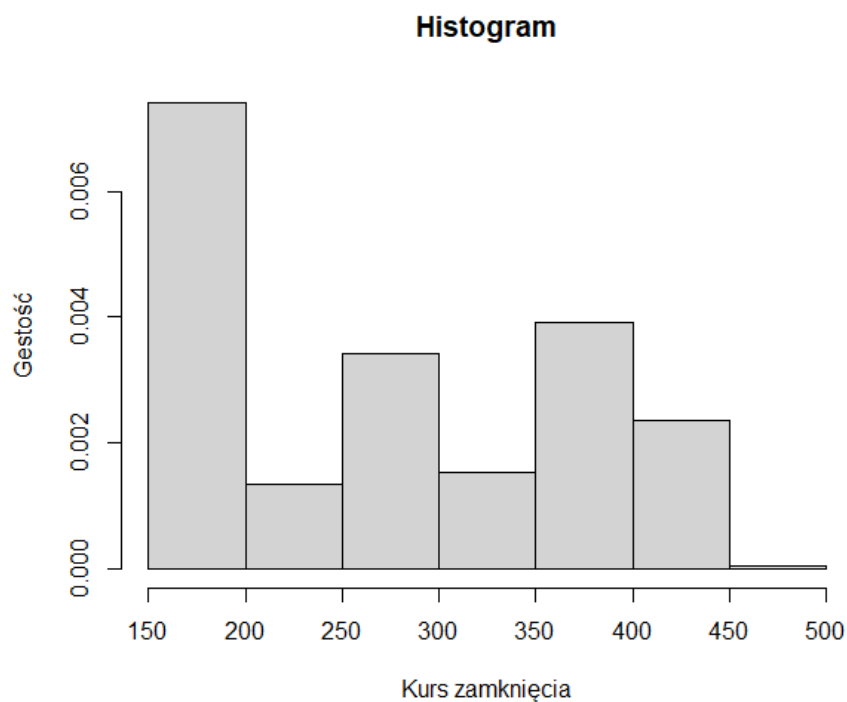
11 bit studios SA (11B) to wieloplatformowy deweloper gier. Firma jest aktywna na każdym etapie tworzenia gry, z znanymi tytułami takimi jak Frostpunk i This War of Mine. Oprócz produkcji gier, grupa również wydaje gry zewnętrzne i prowadzi platformę sprzedaży internetowej.

## 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

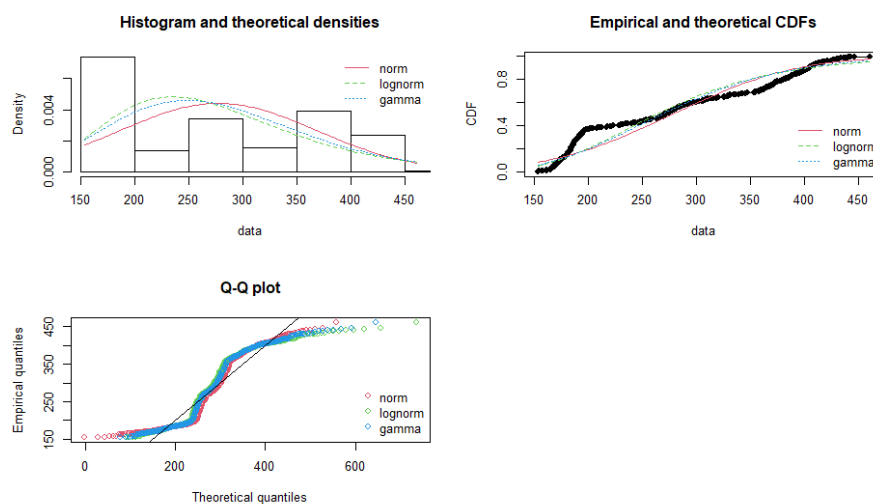
Współczynnik skośności powyżej 0 świadczy o prawostronnej asymetrii rozkładu (wydłużone jest prawe ramię rozkładu). Kurtoza większa od 0 oznacza iż rozkład jest bardziej wysmukły niż normalny, większe skupienie wartości wokół średniej.

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

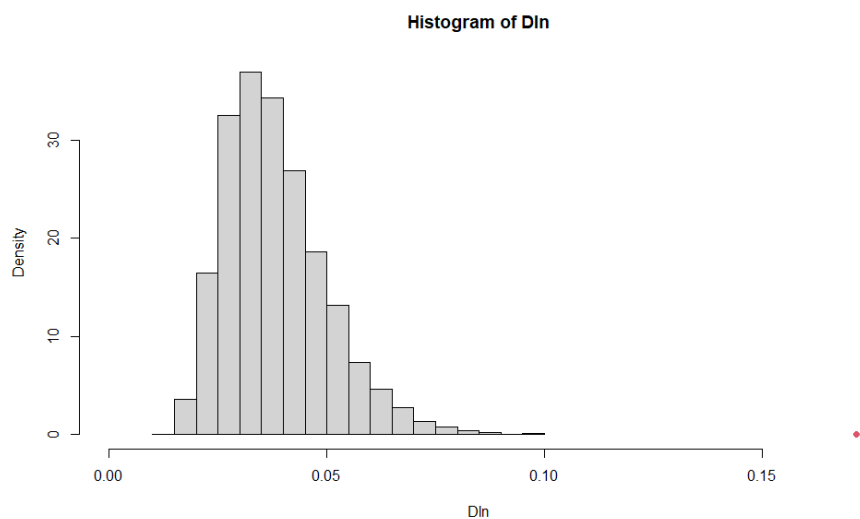
	Normalny	Log-normalny	Gamma
Kolmogorov-Smirnov statistic	0.1777046	0.1715694	0.1739537
Cramer-von Mises statistic	2.9612965	2.9246380	2.8490879
Anderson-Darling statistic	19.7989554	19.3686342	19.1126404
Akaike's Information Criterion	6047.228	6010.751	6013.699
Bayesian Information Criterion	6055.697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

$$Dn = 0.1715694$$

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

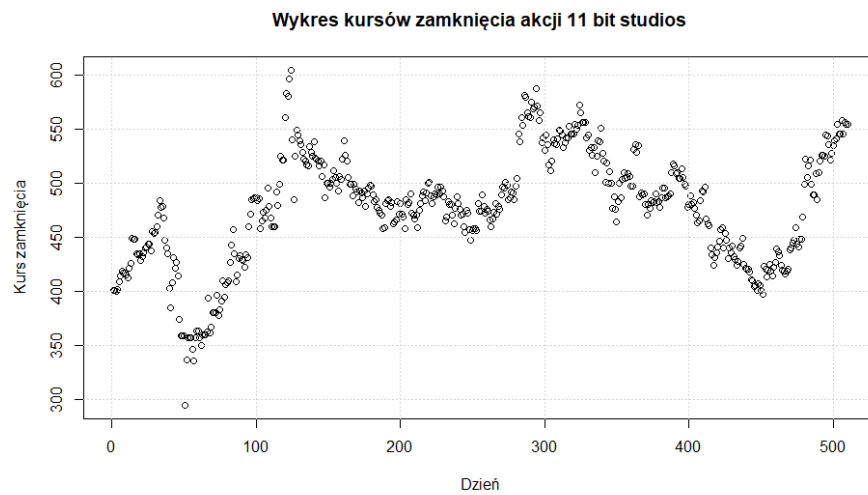
gdzie Dn to odległość dystrybuant empirycznych od rozkładu log-normalnego, dn to wartość statystyki Dn, dla danych akcji spółki,  $N = 10000$

Wartość p jest mniejsza od przyjętego poziomu istotności,

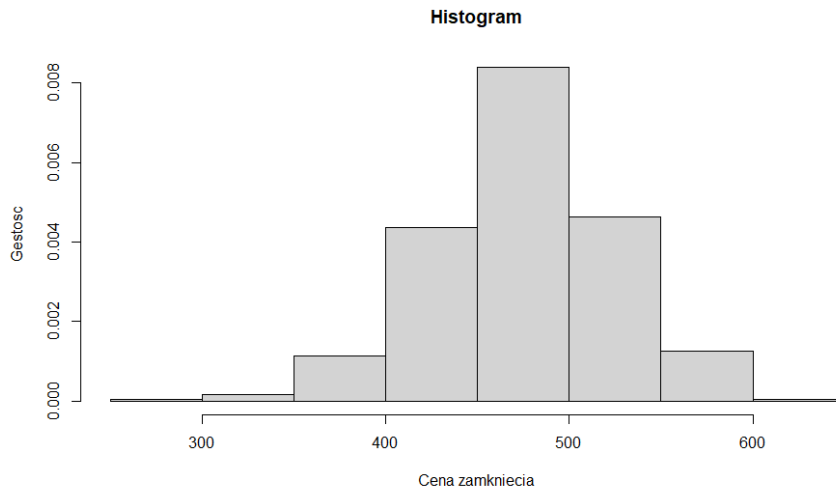
$$\alpha = 5\%$$

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.

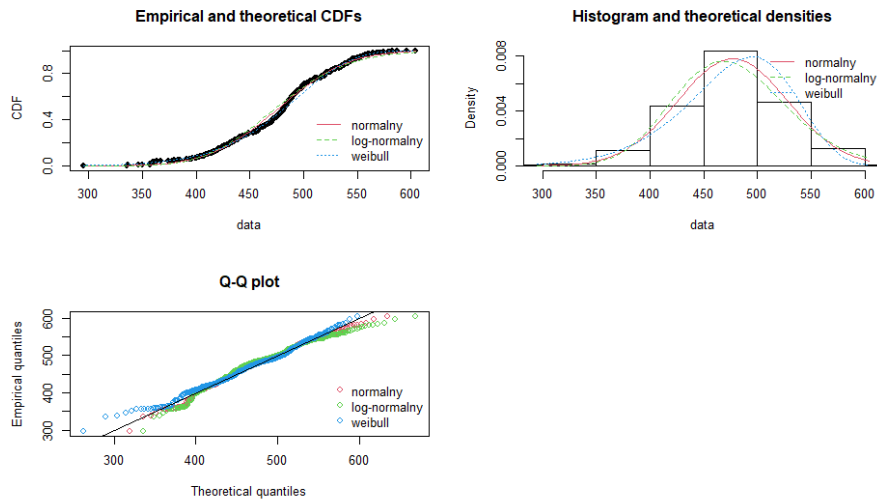
Wystymowane parametry dla badanych rozkładów przy wykorzystaniu estymatora największej wiarygodności (MLE):

- 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

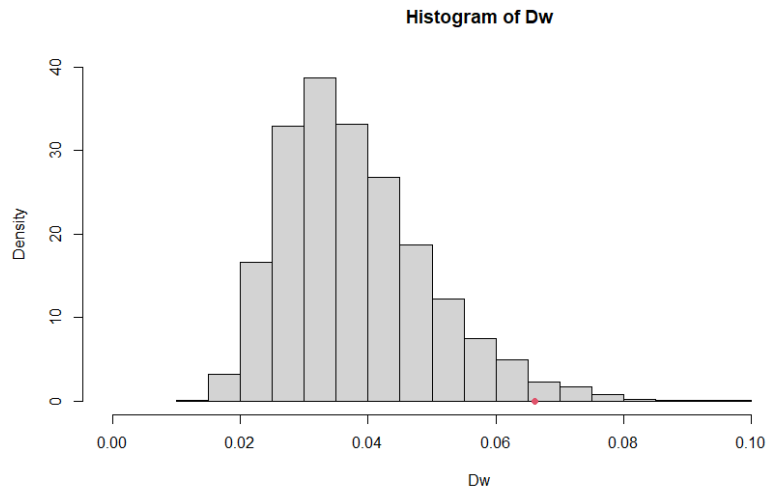


Z powyższych danych wynika, iż rozkład Weibulla najlepiej opisuje kursy zamknięcia spółki 11 bit studios. Wszystkie z pięciu przypadków wskazują za rozkładem Weibulla.



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,  $N = 10000$

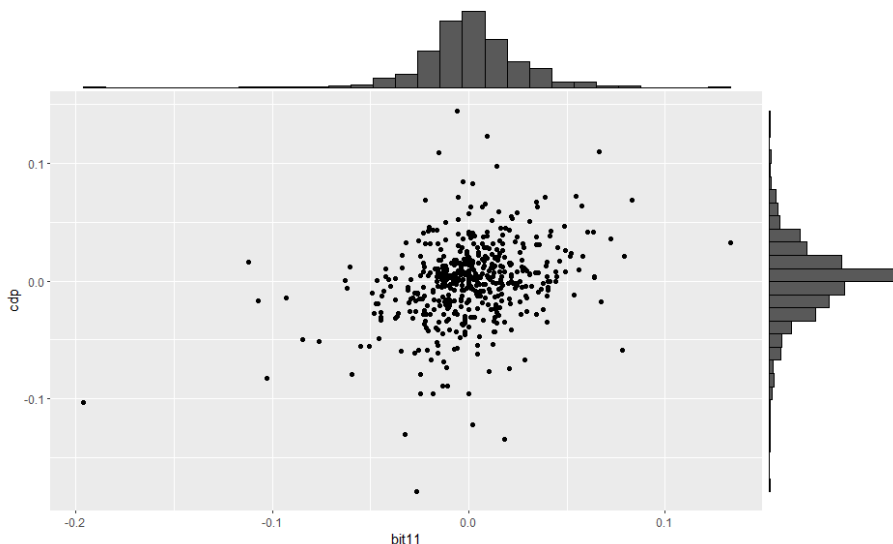
Wartość  $p$  jest mniejsza od przyjętego poziomu istotności,

$$\alpha = 5\%$$

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:

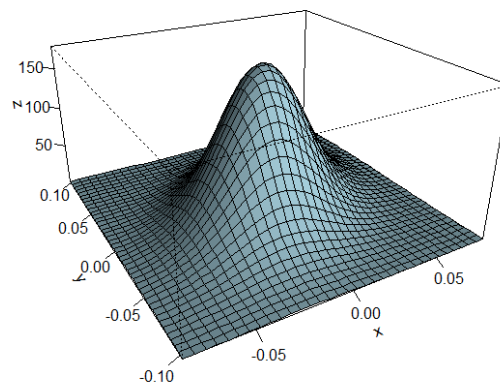
$$\hat{\Sigma} = \begin{bmatrix} 0.00077 & 0.00031 \\ 0.00031 & 0.00117 \end{bmatrix}$$

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:

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

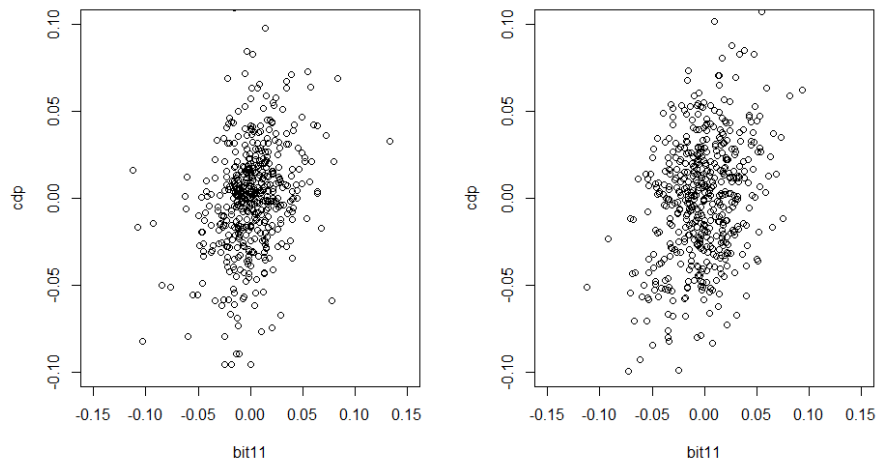
Gęstość rozkładów brzegowych jest zapisywana w następujący sposób:

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

Gęstość rozkładów brzegowych dla analizowanych danych:

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

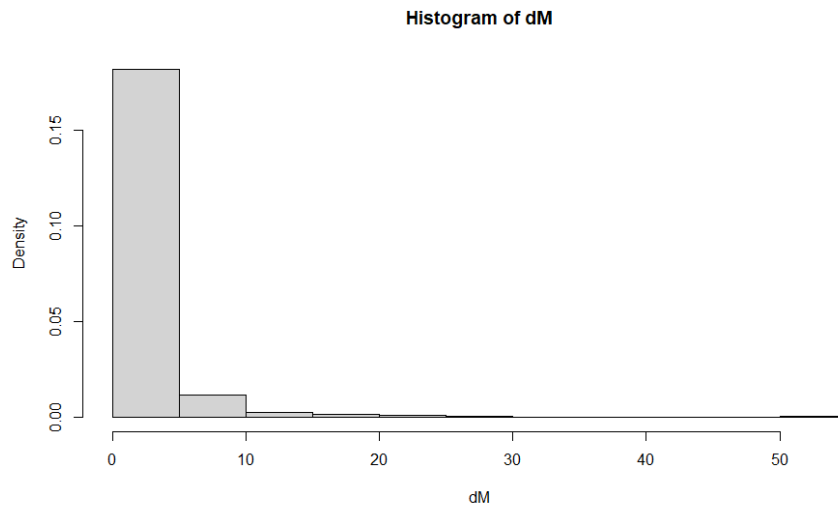
$$f_{cdr}(x) = 11.65201 \cdot \exp\left(-\frac{(x+0.00078)^2}{0.00234}\right)$$



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.

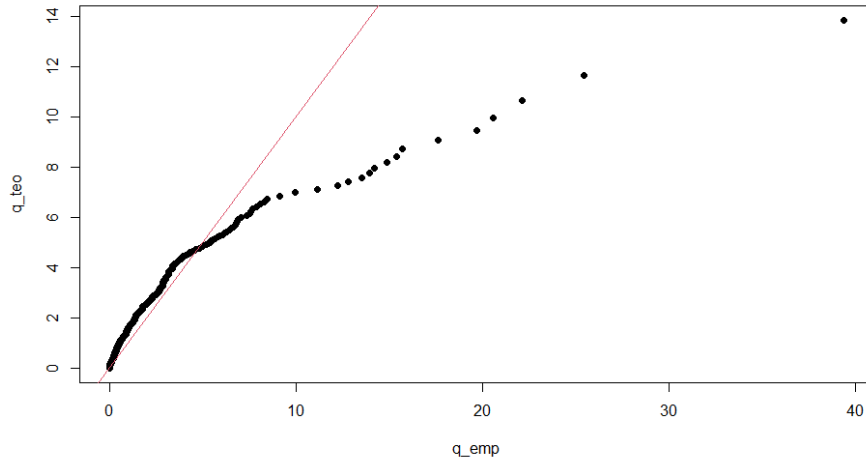
Badanie hipotezy mówiącej o tym, że kwadrat odległości Mahalanobisa wektora cen od średniej ma rozkład  $\chi^2$  (drugiego stopnia).



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

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  $\chi^2$  (stopnia drugiego). Celem testu jest zweryfikowanie hipotezy zerowej, oraz hipotezy alternatywnej, która głosi, że dane nie pochodzą z tego rozkładu. W teście uzyskamy wartość p, które pozwalają na ocenę poziomu istotności statystycznej odchyleń danych od rozkładu teoretycznego.



Po przetestowaniu hipotezy że kwadrat odległości Mahalanobisa wektora cen od średniej mają rozkład  $\chi^2$  (stopnia drugiego), 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  $\chi^2$  (stopnia drugiego). Wartość p jest mniejsza niż poziom istotności  $\alpha = 5\%$ , zatem odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną. Ponadto, przez to odrzucamy również hipotezę o normalności rozkładu log-zwrotów dla badanych spółek.

## 4 Regresja liniowa dla log-zwrotów

### Przedziały ufności dla wartości oczekiwanych log-zwrotów spółki CDR.

Wykorzystywany jest Model 2, z uwagi na dużą ilość próbek oraz na brak jej normalności.

Wzór dla przedziału ufności dla Modelu 2 będzie następujący:

$$\left[ \bar{X}_n - \mu(1 - \alpha/2) \cdot \frac{S_n}{\sqrt{n}}, \bar{X}_n + \mu(1 - \alpha/2) \cdot \frac{S_n}{\sqrt{n}} \right]$$

gdzie;  $\mu = -0.0007792008$ ,  $S_n = 0.03423808$ ,  $X_n = \text{log-zwroty spółki}$ , dla danych długości  $n = 509$  i  $\alpha = 5\%$

- przedział lewy wynosi: -0.003753595
- przedział prawy wynosi: 0.002195194

Ostateczny przedział ufności dla wartości oczekiwanych log-zwrotów spółki CDR wynosi: [-0.003753595, 0.002195194]

## Przedziały ufności dla wartości oczekiwanych log-zwrotów spółki 11B.

Wykorzystywany jest Model 2, z uwagi na dużą ilość próbek oraz na brak jej normalności.

Wzór dla przedziału ufności dla Modelu 2 będzie następujący:

$$[\bar{X}_n - \mu(1 - \alpha/2) \cdot \frac{S_n}{\sqrt{n}}, \bar{X}_n + \mu(1 - \alpha/2) \cdot \frac{S_n}{\sqrt{n}}]$$

gdzie;  $\mu = 0.0006374281$ ,  $S_n = 0.02771611$ ,  $X_n = \text{log-zwroty spółki}$ , dla danych długości  $n = 509$  i  $\alpha = 5\%$

- przedział lewy wynosi: -0.001770378
- przedział prawy wynosi: 0.003045234

Ostateczny przedział ufności dla wartości oczekiwanych log-zwrotów spółki 11B wynosi: [-0.001770378, 0.003045234]

### Estymatory:

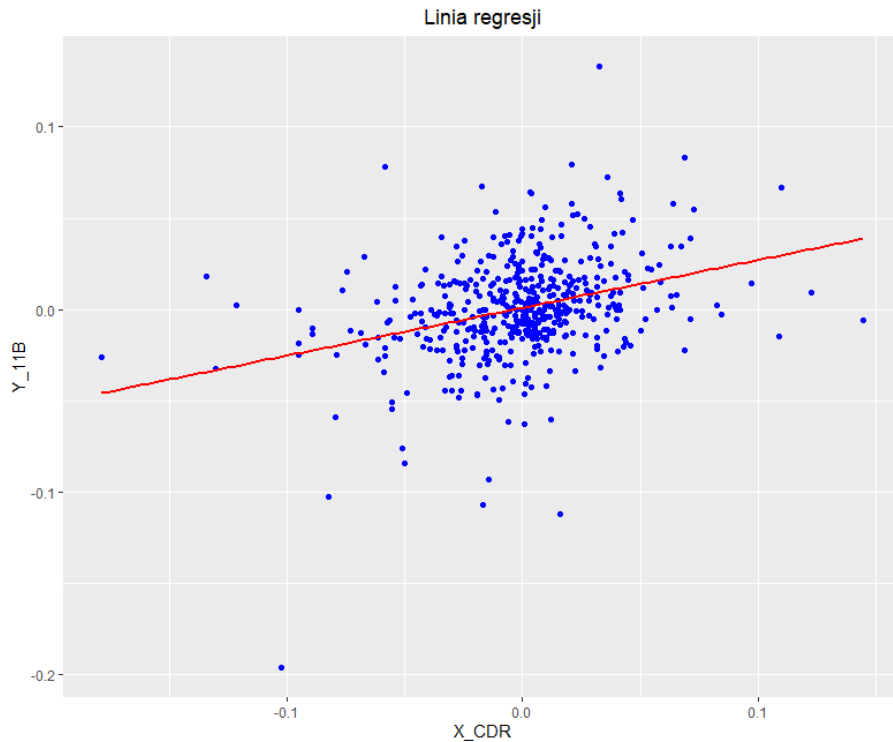
$$\beta_1 = \frac{\text{cov}(X, Y)}{\sigma_x^2} = 0.2617344$$

$$\beta_0 = \bar{X} - \bar{Y} * \beta_1 = -0.0009460377$$

gdzie X to log-zwroty dla spółki CDP, a Y dla spółki 11B  
Nasz model to

$$y = \beta_0 + \beta_1 \cdot x = -0.0009460377 + 0.2617344 \cdot x$$

## Analiza Regresji



Rysunek 12: Linia regresji na wykresie

```
Call:
lm(formula = Y_11B ~ X_CDR, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.170289 -0.012718 -0.000841  0.014242  0.123715

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0008414   0.0011640    0.723   0.47
X_CDR        0.2617344   0.0340206   7.693 7.52e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02625 on 507 degrees of freedom
Multiple R-squared:  0.1045,    Adjusted R-squared:  0.1028
F-statistic: 59.19 on 1 and 507 DF,  p-value: 7.524e-14
```

Powyższy wynik pochodzi z analizy regresji liniowej, gdzie zmienna zależna ( $Y_{11B}$ ) jest przewidywana przez zmienną niezależną ( $X_{CDR}$ ) i jest to przedstawiane za pomocą prostej na wykresie.

Tabela "Coefficients" pokazuje oszacowane współczynniki dla przecięcia i zmiennej niezależnej. Wartość  $t$  dla zmiennej niezależnej wynosi 7.693, a wartość  $p$  i  $F$  – *statistic* wynoszą kolejno:  $7.52e^{-14}$  i 59.19, co sugeruje, że zmienna niezależna jest istotnym czynnikiem wpływającym na zmienną zależną.

W przedstawionych danych widzimy, że reszty mają wartości minimalne - 0.170289, maksymalne 0.123715 i kwartyle: 1Q -0.012718, Mediana -0.000841, 3Q 0.014242. Z tego możemy wywnioskować, że reszty są dość małe i rozłożone wokół zera, co jest zgodne z założeniem, że reszty powinny być rozłożone normalnie o średniej 0 i stałym rozkładzie.

Średni błąd ( $\sigma^2$ ) jest równy 0.02625, ma on małą wartość co świadczy o dobrym dopasowaniu modelu. Stopnie swobody oznaczają liczbę niezależnych punktów danych, które służą do estymacji parametrów modelu. W tym przypadku jest to 507, co sugeruje, że model jest dobrze dopasowany do danych i ma wysoką precyzję.

R-kwadrat (współczynnik determinacji) to stosunek wariancji składników modelu do wariancji danych. Wynosi on 0.1028, co oznacza, że zmienna niezależna wyjaśnia około 10% wariancji zmiennej zależnej. Zatem 90% zmienności nie jest wyjaśniona przez obecne zmienne. Można więc rozważyć dołączenie do naszego modelu jeszcze innych zmiennych.

### Test istotności współczynników $\beta_0, \beta_1$ .

Sprawdzamy hipotezę zerową  $\beta_0 = 0$ , przeciwko alternatywnej hipotezie  $\beta_0 \neq 0$  oraz sprawdzamy hipotezę zerową  $\beta_1 = 0$ , przeciwko alternatywnej hipotezie  $\beta_1 \neq 0$ .

Dla współczynników  $\beta_0$  i  $\beta_1$  odchylenia estymatorów wynoszą:

- $\beta_0 = 0.001163958$
- $\beta_1 = 0.03402064$

Wartości  $t_0$  i  $t_1$  to wartości statystyki T, które służą do testowania hipotezy zerowej o braku istotności współczynników regresji. wynoszą odpowiednio:

- $t_0 = -0.8127763$
- $t_1 = 7.693402$

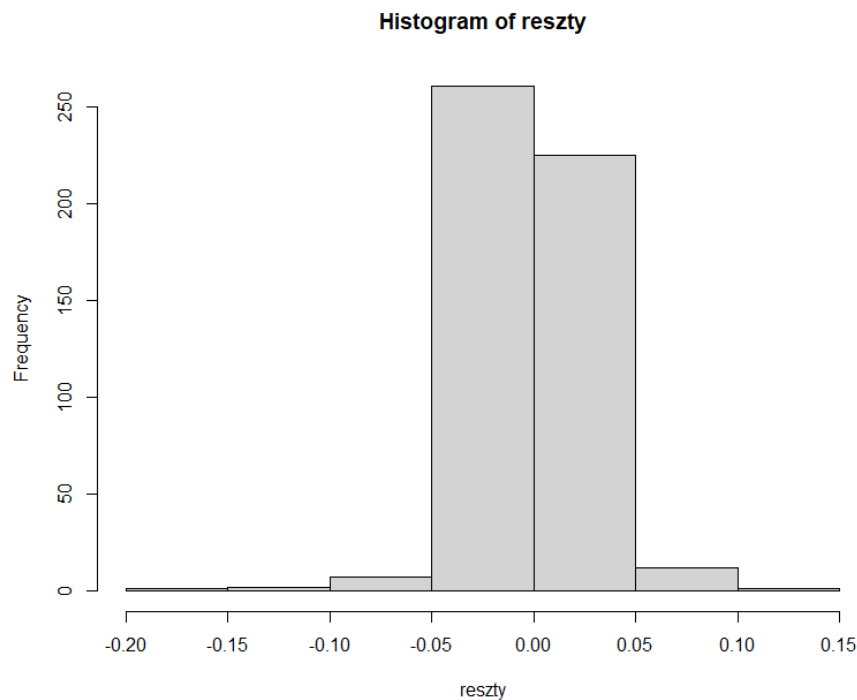
Wartość  $p$ -value ( $p = P(|T| > t_0)$ ,  $p = P(|T| > t_1)$ ) to prawdopodobieństwo, że wartość statystyki  $T$  jest większa niż wartość obliczona dla danych.

- $p$ -value = ( $p = P(|T| > t_0)$ ,  $p = P(|T| > t_1)$ )
- $p$ -value dla  $t_0 = 0.4183785$
- $p$ -value dla  $t_1 = 1.325384^{-11}$

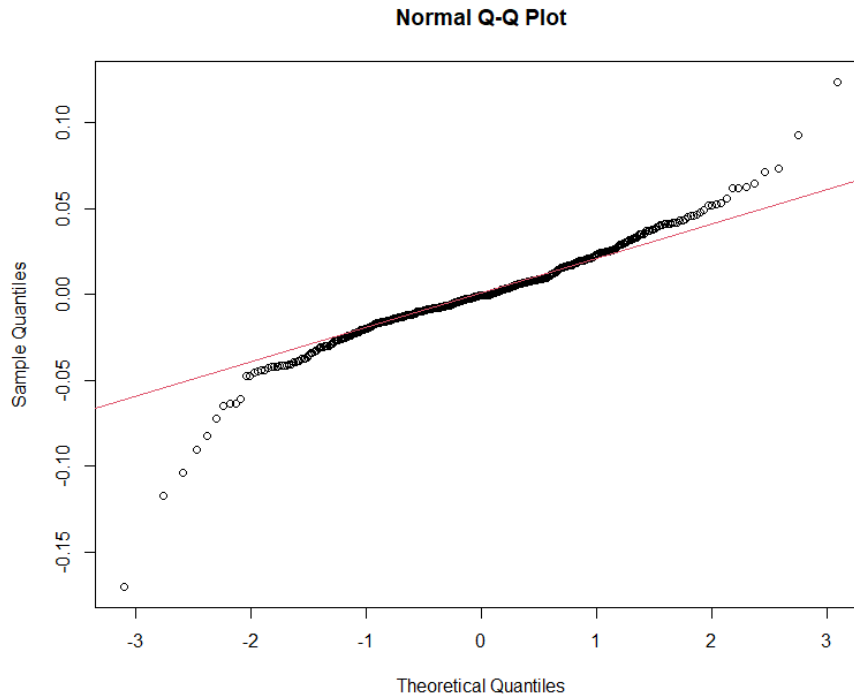
Wartość  $p$ -value dla współczynnika  $\beta_0$  jest równa 0.4183785, co oznacza, że można przyjąć hipotezę zerową  $\beta_0 = 0$  i pominąć  $\beta_0$  z modelu. Wartość  $p$ -value dla współczynnika  $\beta_1$  jest równa  $1.325384^{-11}$ , co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej, że współczynnik  $\beta_1$  jest równy zero.

### Analiza reszt

Testujemy hipotezę czy dystrybucja reszt ma rozkład normalny o średniej i wariancji z próby, oraz kontr hipotezę, że nie ma ona rozkładu normalnego.



Rysunek 13: Na histogramie widać, że większość reszt wpada w przedziały od -0.05 do 0.05 i wpisuje on się w kształt rozkładu normalnego



Rysunek 14: Na wykresie kwantyl-kwantyl widać iż znaczna część empirycznych pokrywa się z linią która reprezentuje dane teoretyczne, co sugeruje że reszta będzie z rozkładu normalnego

- Średnia reszt =  $6.525989^{-19}$
- odchylenie = 0.02622743

Średnia reszt jest bliska zeru, co oznacza, że w średnim model przewiduje dobrze wartości zmienności  $R^2$ . Natomiast odchylenie standardowe reszt wynosi 0.02622743, co oznacza, że większość reszt znajduje się w przedziale od -0.02622743 do 0.02622743 od wartości przewidywanej.

Z pomocą testu Kołmogorowa-Smirnova sprawdzamy czy dane należą do rozkładu normalnego. Uzyskane wyniki:

- $D = 0.081221$
- $p - value = 0.002424$

$p - value$  jest mniejsze niż 0.05, co sugeruje, że nie ma podstaw, by odrzucić hipotezę, że reszty są normalnie rozłożone.

$$RSE = 0.02624$$

Błąd standardowy reszty RSE jest miarą, jak dobrze model matematyczny opisuje dane. Oznacza on średni odchylenie reszt (różnica między wartością rzeczywistą a przewidywaną przez model) od wartości średniej. Im mniejszy błąd standardowy reszty, tym lepiej model opisuje dane. Wartość 0.02624 wskazuje, że nasz model jest precyzyjny.

Na podstawie powyższych analiz można uprościć model o pominięciu  $\beta_0$ , Po dokonaniu ponownej regresji model prezentuje się tak:

$$model.lm2 = lm(Y_{11B} \sim X_{CDR} - 1, data = df)$$

$$y = \beta_1 * x$$

a jego podsumowanie tak:

```
Call:
lm(formula = Y_11B ~ X_CDR - 1, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.16950 -0.01187  0.00000  0.01507  0.12458

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
X_CDR      0.2612      0.0340   7.683 8.09e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02624 on 508 degrees of freedom
Multiple R-squared:  0.1041,    Adjusted R-squared:  0.1023
F-statistic: 59.02 on 1 and 508 DF,  p-value: 8.094e-14
```

Rysunek 15: Nowy model po pominięciu b0

Nowy model jest minimalnie lepszy, ponieważ ma więcej stopni swobody (508 w porównaniu z 507 w pierwszym modelu).

**Wielkości log-zwrotów spółki 11B, gdy log-zwroty spółki CDR będą na poziomie średniej z posiadanej próby.**

predykcja będzie następująca:

$$\beta_{1model2} * m = -0.0002035073$$

$$m_{CDR} = -0.0007792008$$

Przedziały ufności:

fit	lwr	upr
-0.0002035073	-0.0002555499	-0.0001514647

Przedział ufności jest to zakres, w którym znajduje się prawdziwa wartość zmiennej, z pewnym poziomem pewności. Wartość "fit" jest oszacowaną wartością, a "lwr" i "upr" są dolnym i górnym ograniczeniem przedziału ufności, odpowiednio. W tym przypadku mówimy, że z pewnym poziomem ufności, prawdziwa wartość zmiennej znajduje się pomiędzy "lwr" a "upr".

## 5 Podsumowanie

Celem projektu była analiza spółek CDR i 11B, którą podzieliśmy na 3 etapy.

Pierwszy etap skupiał się na cenach zamknięcia akcji spółek, przedstawialiśmy wykresy kursów, szukaliśmy rozkładu który by jak najlepiej opisywał powyższe dane, za pomocą wartości statystyk, kryteriów informacyjnych i metody Monte-Carlo. Niestety, dla żadnej ze spółek dobrane rozkłady nie były wystarczająco pasujące.

Drugi etap obejmował analizę dziennych log-zwrotów spółek. Dokonaliśmy analizy dopasowania rozkładu normalnego do danych, przy użyciu m.in kwadratów odległości Mahalanobisa. Badanie zakończyło się odrzuceniem hipotezy o normalności rozkładu.

W trzeciej części zajęliśmy się tworzeniem modelu regresji liniowej. Po przeanalizowaniu parametrów model udało się uprościć. Następnie wyliczyliśmy predykcję dla wielkości log-zwrotów spółki 11B, gdy log-zwroty spółki CDR będą na poziomie średniej z posiadanej próby.

Ostatecznie, nasze wyniki sugerują konieczność bardziej zaawansowanego podejścia do analizy cen, aby lepiej zrozumieć rynkowe trendy i przewidzieć możliwe scenariusze.