Modelowanie Matematyczne

Projekt Egzaminacyjny

Wersja oryginalna: Październik 2023 – Styczeń 2024
Wersja poprawiona: Październik 2025

Kacper Hołowaty

Spis treści

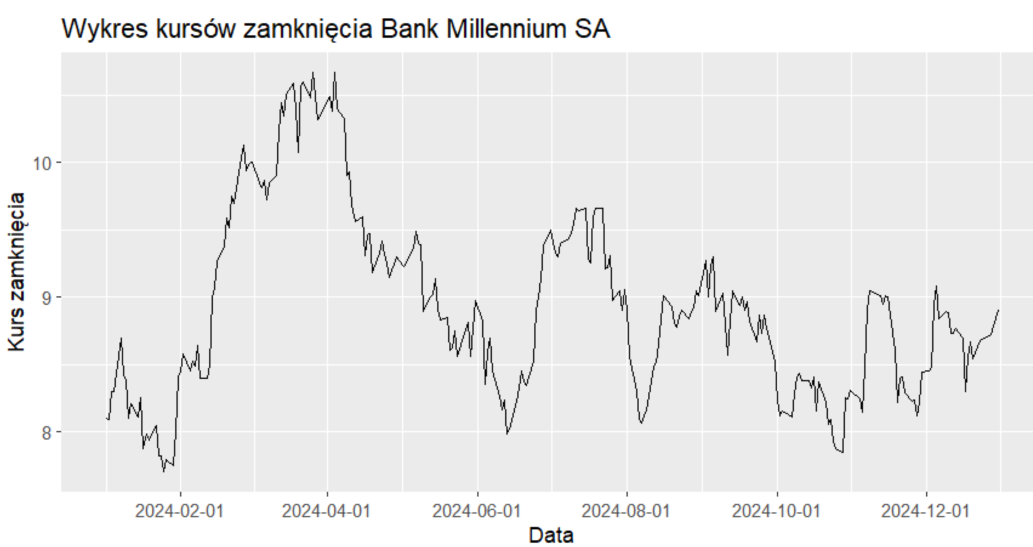
1 Wstęp

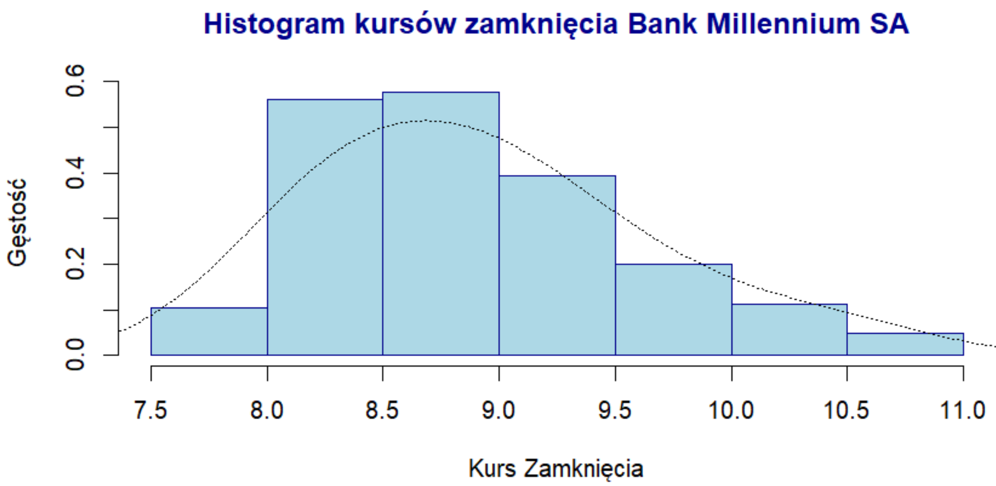
Motywem przewodnim projektu jest analiza dwóch spółek: Bank Millennium oraz mBank. Zawiera on trzy rozdziały, gdzie każdy z nich analizuje różne specyfikacje spółek, na przykład kursy zamknięcia czy log-zwroty za 2024 rok.

Bank Millennium oraz mBank to dwie wiodące instytucje finansowe, które odgrywają istotną rolę w polskim sektorze bankowym. Oba banki specjalizują się w kompleksowej obsłudze klientów detalicznych oraz korporacyjnych, oferując nowoczesne rozwiązania z zakresu bankowości elektronicznej, produktów kredytowych, oszczędnościowych oraz inwestycyjnych.

2 Analiza cen zamknięcia dla spółek

2.1 Bank Millennium S.A.





Statystyki opisowe Bank Millennium

	$\bar{x}$	odch. st.	skośność	kurtoza
Akcja	8.912269	0.6849215	0.6590286	2.85685

Kurtoza

Kurtoza dla kursu zamkniecia spółki Bank Millennium wynosi około 2.86 i ma ona rozkład lekko platykurtyczny, co oznacza, że intensywność wartości ekstremalnych jest mniejsza niż w przypadku rozkładu normalnego, i że dane są skoncentrowane wokół średniej. Kurtoza ta sugeruje również mniejsze ryzyko ekstremalnych wahań cen akcji niż w rozkładzie normalnym oraz rzadsze występowanie bardzo dużych wzrostów lub spadków.

Skośność

Skośność dla kursu zamknięcia spółki Bank Millennium wynosi około 0.66 i oznacza, że rozkład jest umiarkowanie skośny w prawo, czyli ma dłuższy ogon po prawej stronie, a większość wartości jest mniejszych od średniej. Średnia jest minimalnie większa od mediany (Me = 8.845).

Parametry wybranych rozkładów wyestymowane za pomocą estymatora największej wiarygodności dla spółki Bank Millennium

Rozkład normalny

	estimate	Std. Error
mean	8.9122691	0.04331789
sd	0.6835448	0.03063008

Rozkład log-normalny

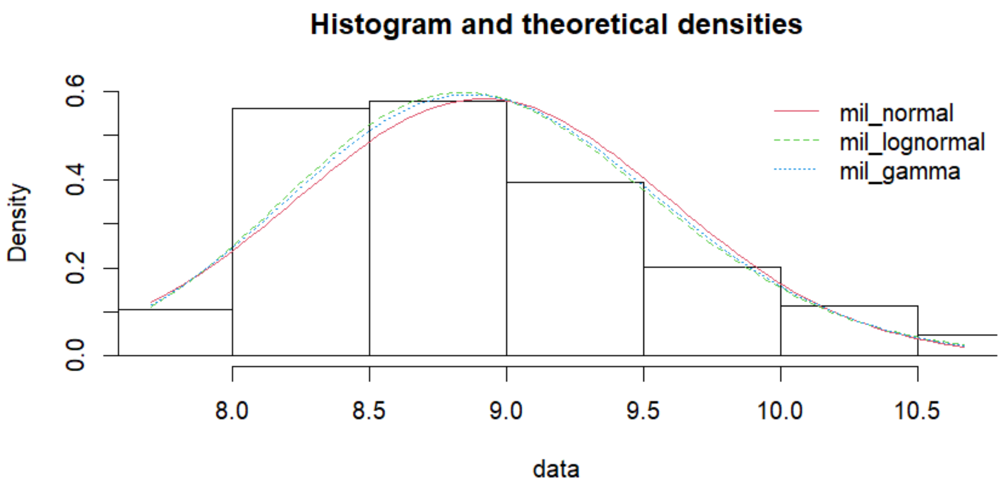
	estimate	Std. Error
meanlog	2.18456392	0.004767633
sdlog	0.07523199	0.003368546

Rozkład gamma

	estimate	Std. Error
shape	174.68275	15.640544
rate	19.59982	1.757421

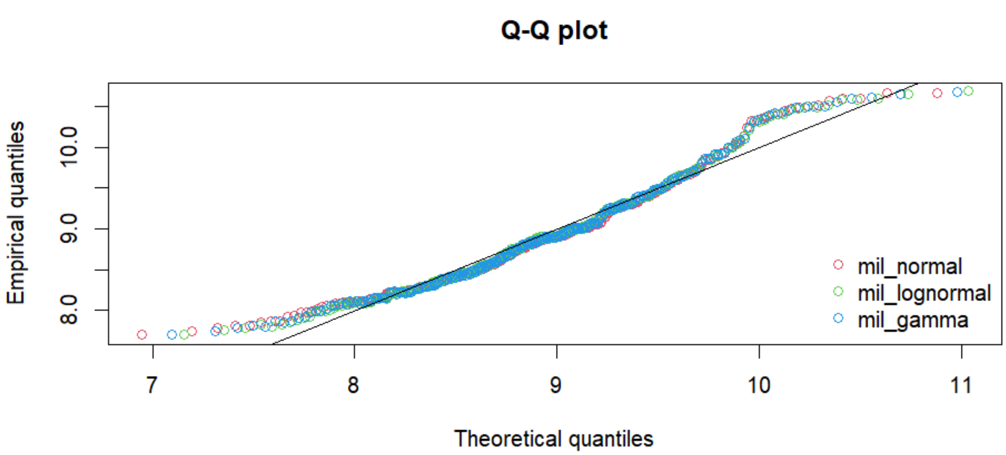
Wykresy diagnostyczne dla Bank Millennium

Wykres gęstości



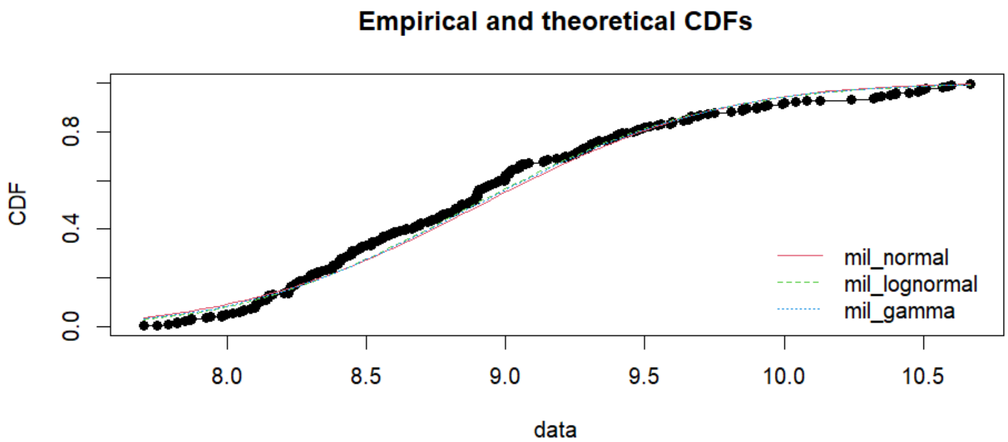
Ten wykres porównuje funkcje gęstości prawdopodobieństwa dla trzech rozkładów: log-normalnego, gamma i normalnego. Pozwala zobaczyć, jak te rozkłady różnią się pod względem kształtu i szerokości.

Wykres kwantyl-kwantyl



Ten wykres porównuje kwantyle (wartości, które dzielą rozkład na równe części) dla każdego z trzech rozkładów. Pomaga on zobaczyć, czy istnieją istotne różnice między kwantylami rozkładu teoretycznego (log-normalny, gamma, normalny), a rzeczywistymi danymi.

Wykres dystrybuanty



Ten wykres porównuje funkcje dystrybuanty kumulacyjnej dla trzech rozkładów. Pokazuje, jak szybko rośnie prawdopodobieństwo osiągnięcia danej wartości dla każdego z rozkładów.

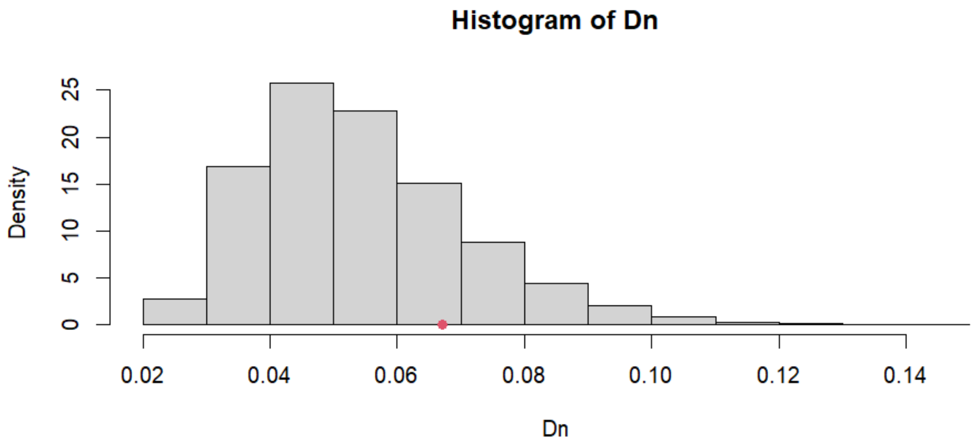
Statystyki dopasowania rozkładu

	norm	**lnorm**	**gamma**
KS	0.08280634	0.06706802	0.07243193
CM	0.40417807	0.26525777	0.30814087
AD	2.84176797	1.92126642	2.20444827
AIC	521.1608	510.1292	513.4298
BIC	528.1957	517.1641	520.4647

Najlepszy rozkład

Analizując wykresy, wartości statystyk KS, CM i AD oraz kryteria AIC i BIC, zauważam, że najmniejsze wartości, jak i najmniejsze odstępy na wykresie, występują dla rozkładu logarytmiczno-normalnego. Zakładam więc, że dla danych cen zamknięcia spółki Bank Millennium, rozkładem najlepiej opisującym dane jest właśnie rozkład logarytmiczno-normalny.

Hipoteza o równości rozkładów - Bank Millennium



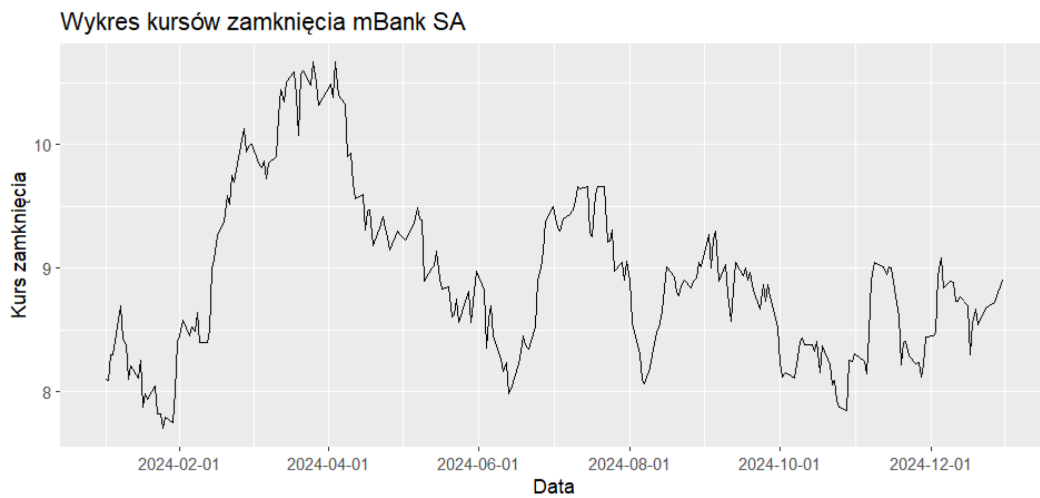
Histogram wizualizuje, gdzie w rozkładzie statystyk testu Kolmogorova-Smirnova znajduje się statystyka dla danych obserwowanych. Jeśli punkt ten znajduje się na skraju lub daleko od głównej masy rozkładu, może to sugerować, że dane obserwowane różnią się od założonego rozkładu log-normalnego.

Wartość D wynosi 0.067. Reprezentuje maksymalną różnicę między empiryczną dystrybuantą danych a dystrybuantą teoretyczną.

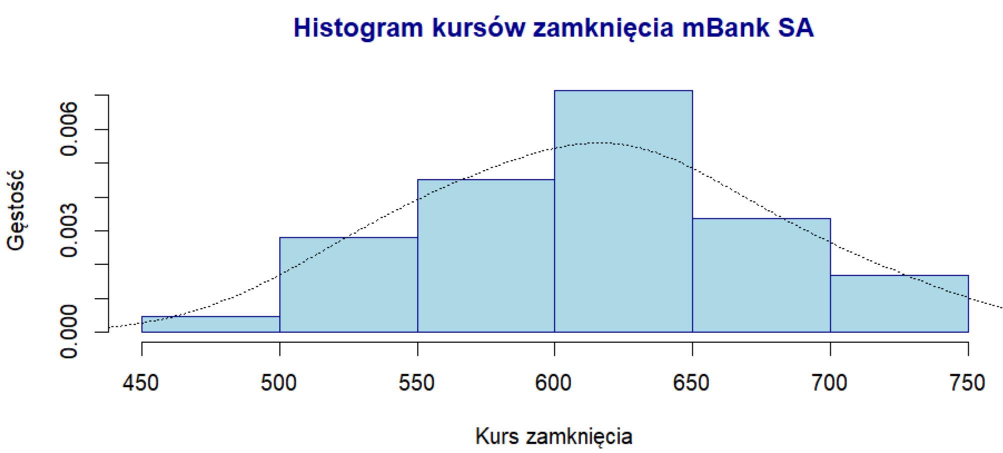
Natomiast wartość p (p-value) oznacza prawdopodobieństwo uzyskania statystyki testowej co najmniej tak ekstremalnej jak obserwowana, przy założeniu prawdziwości hipotezy zerowej, że rozkłady są takie same.

Wykorzystując metodę Monte Carlo, otrzymuję wartość p równą 0.2037. Jest ona większa od przyjętego poziomu istotności 5%, więc nie ma podstaw do odrzucenia hipotezy zerowej, dla rozkładu X $\sim$ LN(2.185, 0.075).

2.2 mBank SA



Wykres kursów zamknięcia mBank SA
Kurs zamknięcia
10
9
8
2024-02-01
2024-04-01
2024-06-01
2024-08-01
2024-10-01
2024-12-01
Data



Histogram kursów zamknięcia mBank SA
Gęstość
0.000
0.003
0.006
450
500
550
600
650
700
750
Kurs zamknięcia

Statystyki opisowe mBank

	$\bar{x}$	odch. st.	skośność	kurtoza
Akcja	613.894	59.95286	0.1337955	2.424685

Kurtoza

Kurtoza dla kursu zamkniecia spółki mBank wynosi około 2.42 i ma ona rozkład platykurtyczny, co oznacza, że intensywność wartości ekstremalnych jest mniejsza niż w przypadku rozkładu normalnego, i że dane są skoncentrowane wokół średniej. Kurtoza ta sugeruje również mniejsze ryzyko ekstremalnych wahań cen akcji niż w rozkładzie normalnym oraz rzadsze występowanie bardzo dużych wzrostów lub spadków.

Skośność

Skośność dla kursu zamknięcia spółki mBank wynosi około 0.13 i sugeruje, że rozkład jest niemal symetryczny z bardzo niewielką prawostronną asymetrią. Średnia jest niewiele mniejsza od mediany (Me = 617).

Parametry wybranych rozkładów wyestymowane za pomocą estymatora największej wiarygodności dla spółki mBank

Rozkład normalny

	estimate	Std. Error
mean	613.89398	3.791722
sd	59.83235	2.681151

Rozkład log-normalny

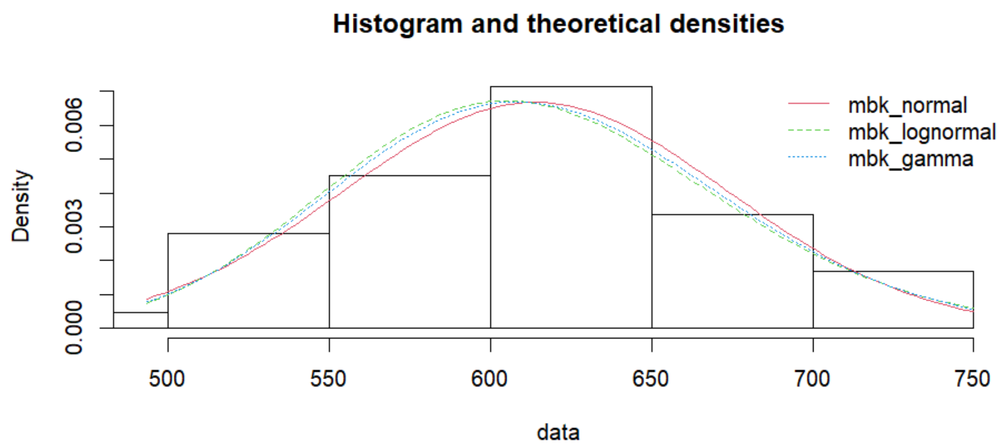
	estimate	Std. Error
meanlog	6.41505973	0.006193528
sdlog	0.09773223	0.004377423

Rozkład gamma

	estimate	Std. Error
shape	105.1513377	9.36446
rate	0.1712856	0.01529

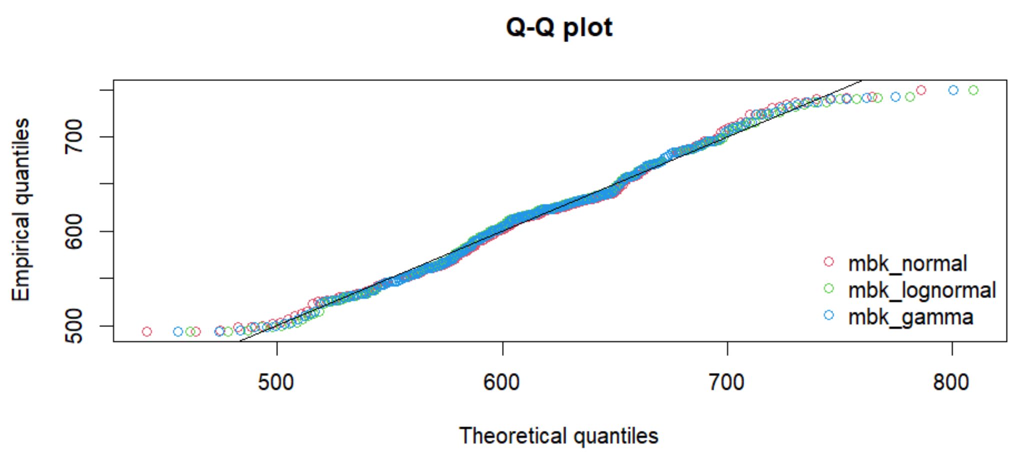
Wykresy diagnostyczne dla mBank

Wykres gęstości



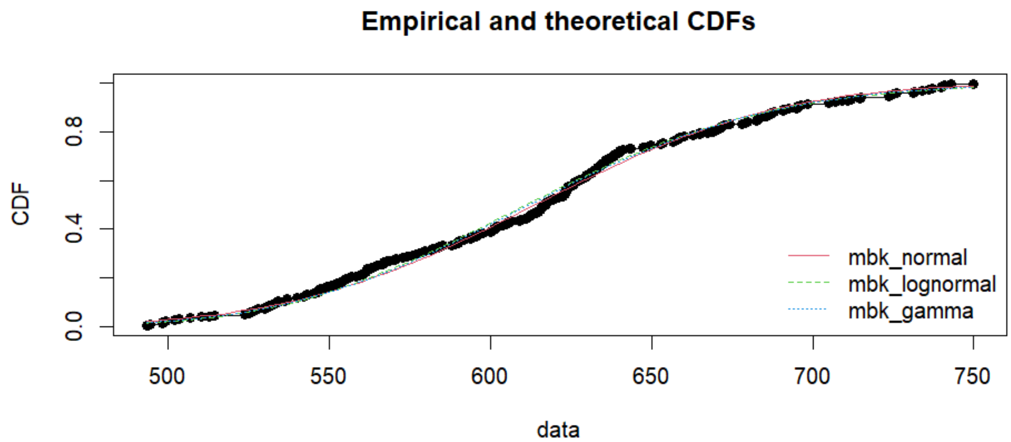
Ten wykres porównuje funkcje gęstości prawdopodobieństwa dla trzech rozkładów: log-normalnego, gamma i normalnego. Pozwala zobaczyć, jak te rozkłady różnią się pod względem kształtu i szerokości.

Wykres kwantyl-kwantyl



Ten wykres porównuje kwantyle (wartości, które dzielą rozkład na równe części) dla każdego z trzech rozkładów. Pomaga on zobaczyć, czy istnieją istotne różnice między kwantylami rozkładu teoretycznego (log-normalny, gamma, normalny), a rzeczywistymi danymi.

Wykres dystrybuanty



Ten wykres porównuje funkcje dystrybuanty kumulacyjnej dla trzech rozkładów. Pokazuje, jak szybko rośnie prawdopodobieństwo osiągnięcia danej wartości dla każdego z rozkładów.

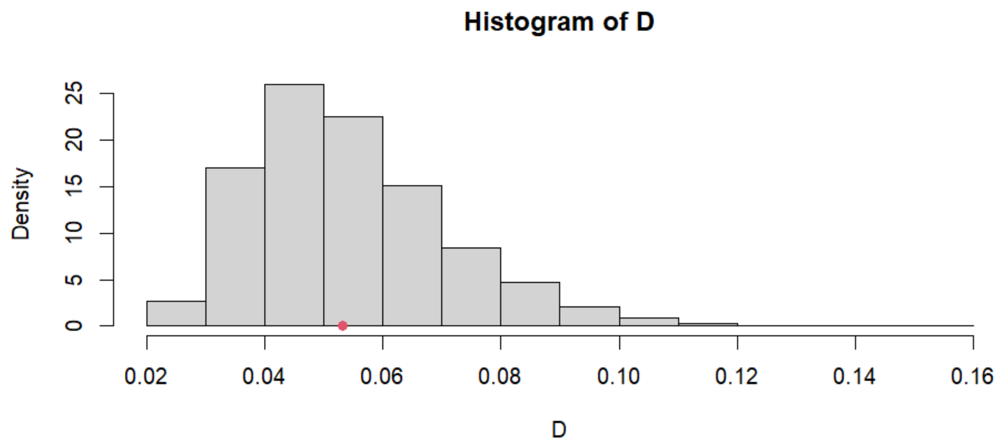
Statystyki dopasowania rozkładu

	norm	**lnorm**	**gamma**
KS	0.05419402	0.05972085	0.05322526
CM	0.13111529	0.15131857	0.13677518
AD	0.86581508	0.86625996	0.81774844
AIC	2748.222	2747.220	2746.921
BIC	2755.256	2754.255	2753.956

Najlepszy rozkład

Analizując wykresy, wartości statystyk KS, CM i AD oraz kryteria AIC i BIC, zauważam, że najmniejsze wartości, jak i najmniejsze odstępy na wykresie, występują dla rozkładu gamma. Zakładam więc, że dla danych cen zamknięcia spółki mBank, rozkładem najlepiej opisującym dane jest właśnie rozkład gamma.

Hipoteza o równości rozkładów - mBank



Histogram wizualizuje, gdzie w rozkładzie statystyk testu Kolmogorova-Smirnova znajduje się statystyka dla danych obserwowanych. Jeśli punkt ten znajduje się na skraju lub daleko od głównej masy rozkładu, może to sugerować, że dane obserwowane różnią się od założonego rozkładu log-normalnego.

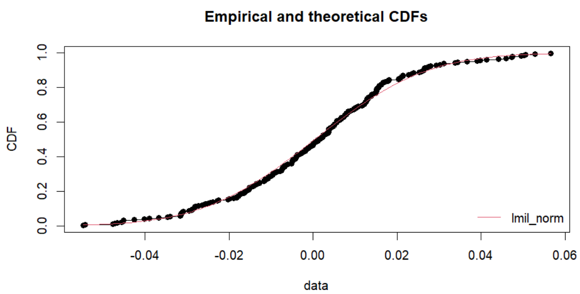
Wartość D wynosi 0.05323. Reprezentuje maksymalną różnicę między empiryczną dystrybuantą danych a dystrybuantą teoretyczną.

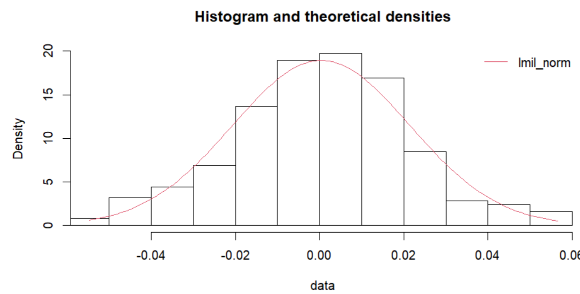
Natomiast wartość p (p-value) jest to prawdopodobieństwo uzyskania statystyki testowej równej lub bardziej skrajnej niż obserwowana, pod warunkiem prawdziwości hipotezy zerowej (rozkłady są takie same).

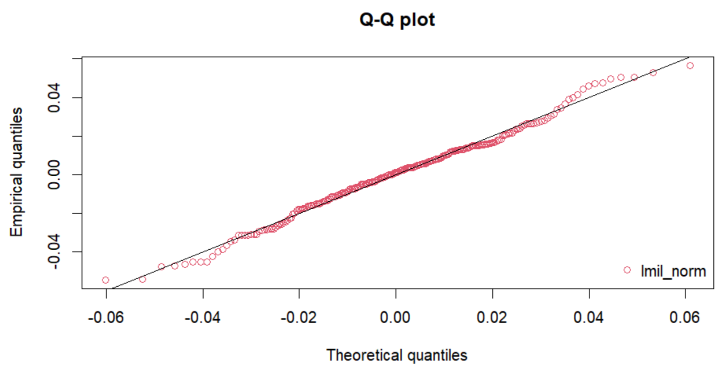
Wykorzystując metodę Monte Carlo, otrzymuję wartość p równą 0.4645. Jest ona większa od poziomu istotności 5%, więc nie ma podstaw do odrzucenia hipotezy zerowej, dla rozkładu X $\sim \Gamma(105.1513, 0.1713)$. Oznacza to również bardzo dobre dopasowanie rozkładu do danych.

3 Analiza łącznego rozkładu log-zwrotów

Wykresy diagnostyczne dla log-zwrotów Bank Millennium

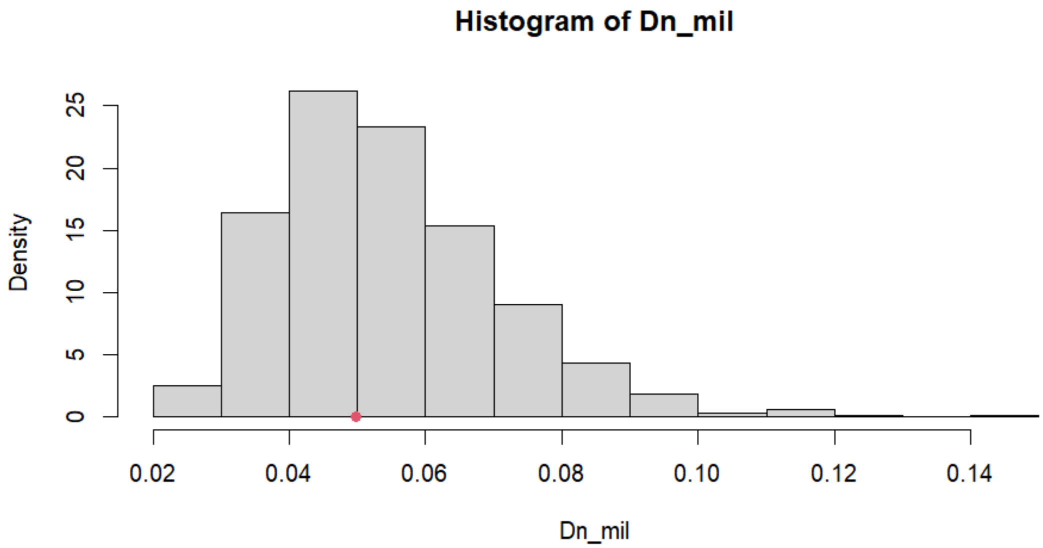






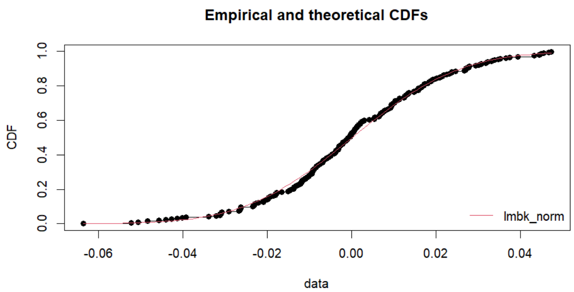
Dla Banku Millennium można zaobserwować, że odległości między rzeczywistymi danymi a kwantylami rozkładu teoretycznego, na większej części wykresu są niewielkie. Praktycznie niezauważalna jest także różnica między empiryczną funkcją dystrybuanty a rzeczywistymi danymi. Rozkład normalny wydaje się być tym, który dobrze opisuje dane dla spółki, choć niekoniecznie jest on najlepszym.

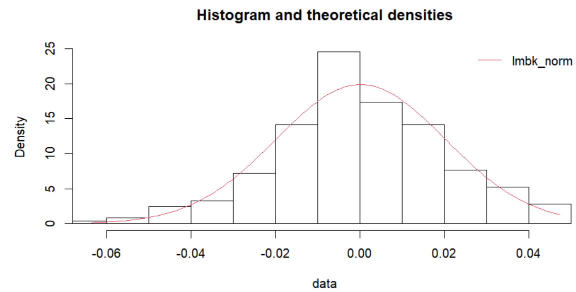
Hipoteza o równości rozkładów dla log-zwrotów Bank Millennium

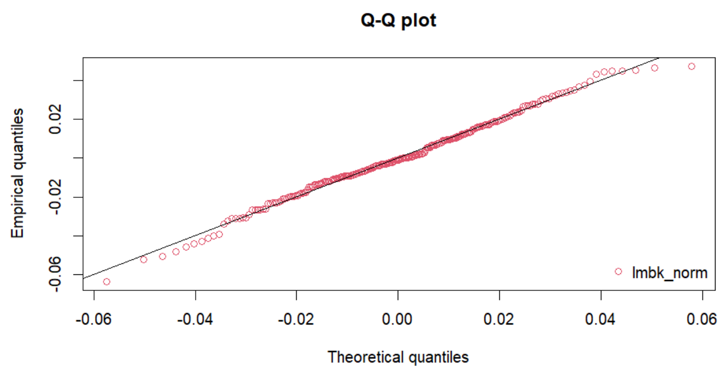


Dla rozkładu X $\sim$ N(0.00038, 0.02106) sprawdzam metodą KS, czy hipoteza jest spełniona. Wyliczone p-value wynosi 0.557 i jest znacznie większe od przyjętego poziomu istotności 5%. Oznacza to, że nie ma podstaw do odrzucenia hipotezy.

Wykresy diagnostyczne dla log-zwrotów mBank

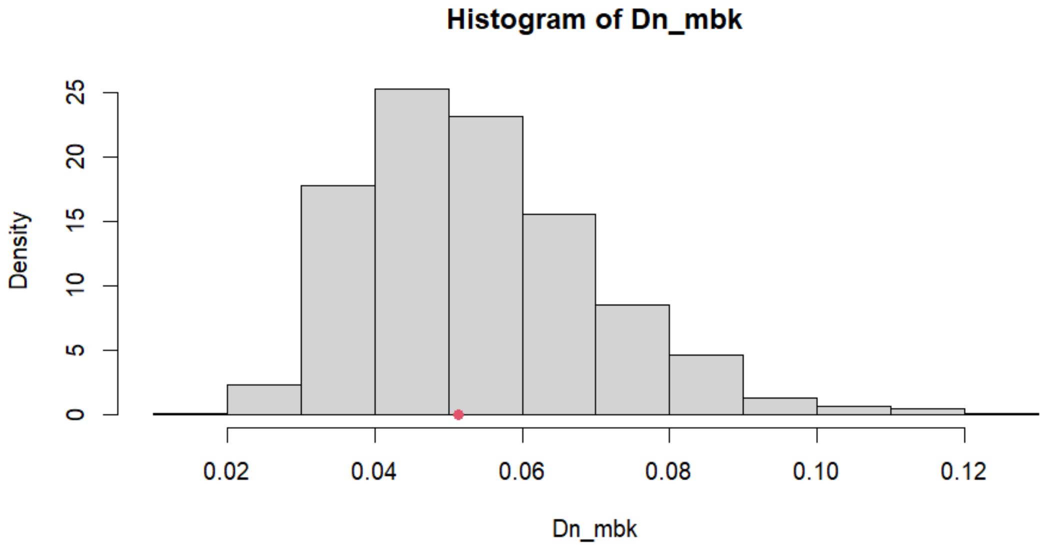






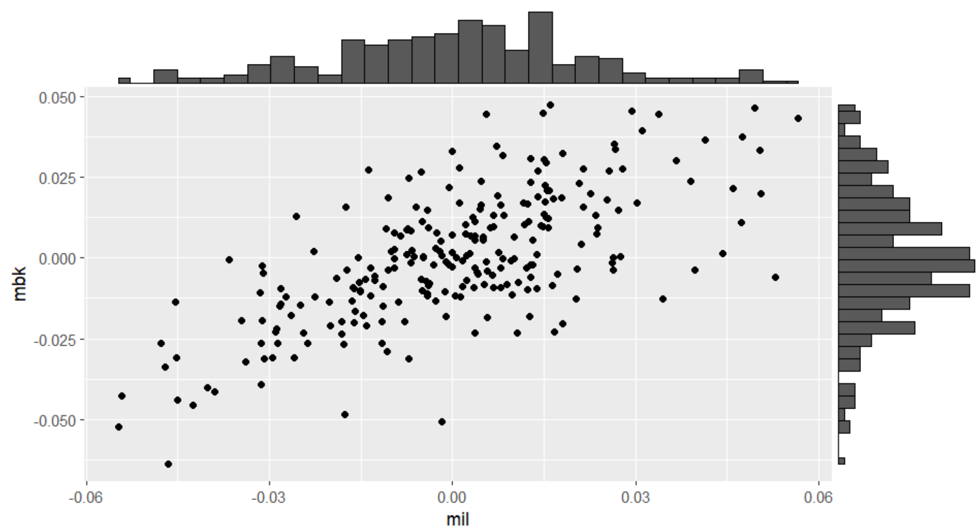
Dla spółki mBank można zaobserwować, że odległości między rzeczywistymi danymi, a kwantylami rozkładu teoretycznego są niewielkie. Zauważalna również jest niewielka różnica między empiryczną funkcją dystrybuanty a rzeczywistymi danymi. Pozwala to stwierdzić, że rozkład normalny dobrze opisuje dane dla tej spółki.

Hipoteza o równości rozkładów dla log-zwrotów mBank



Dla rozkładu X $\sim$ N(0.00017, 0.02006) sprawdzam metodą KS, czy hipoteza jest spełniona. Wyliczone p-value wynosi 0.515 i jest większe od przyjętego poziomu istotności 5%, co oznacza, że nie ma powodu do odrzucenia hipotezy.

Wykres rozrzutu z histogramami rozkładów brzegowych



Wykres rozrzutu wskazuje na **silną dodatnią korelację** między log-zwrotami obu banków. Punkty układają się wzdłuż rosnącej linii prostej, co sugeruje, że akcje mBanku i Banku Millennium reagują podobnie na te same czynniki rynkowe.

Wektor średnich μ

Bank Millennium	mBank
0.0003797872	0.0001670004

Kowariancja

cov = 0.0002854486

$$\hat{\sum} = \begin{bmatrix} 0.0004455209 & 0.0002854486 \\ 0.0002854486 & 0.0004038974 \end{bmatrix}$$

Korelacja

p = 0.672912

$$P = \begin{bmatrix} 1.000000 & 0.672912 \\ 0.672912 & 1.000000 \end{bmatrix}$$

Wzór gęstości rozkładu normalnego

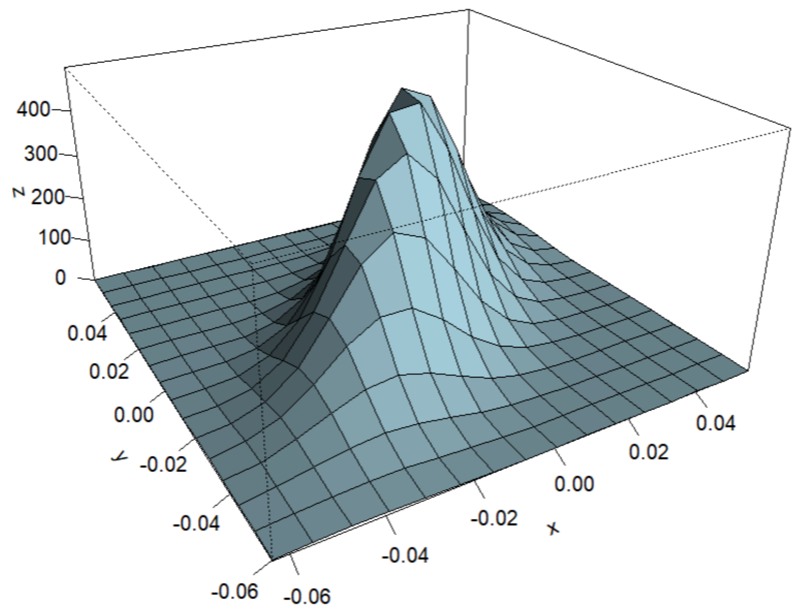
$f(x,y) = \frac{1}{2 \cdot \pi \cdot 0.021 \cdot 0.02 \cdot \sqrt{1-0.673^2}} exp(-\frac{1}{2 \cdot (1-0.673^2)} \cdot [\frac{(x-0.00038)^2}{0.021^2} - 2 \cdot 0.673 \cdot \frac{(x-0.00038) \cdot (y-0.00017)}{0.021 \cdot 0.02} + \frac{(y-0.00017)^2}{0.02^2}])$

Wzory gęstości rozkładów brzegowych

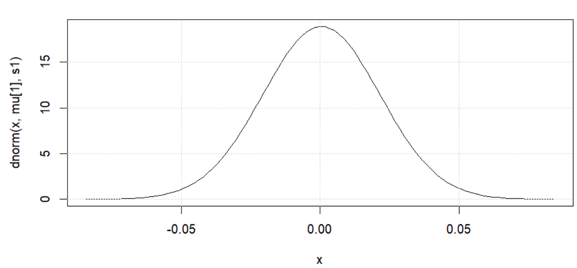
$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2 \cdot \pi} 0.021} \cdot e^{\frac{(x-0.00038)^2}{2}}$

$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2 \cdot \pi} \cdot 0.02} \cdot e^{\frac{(x-0.00017)^2}{2}}$

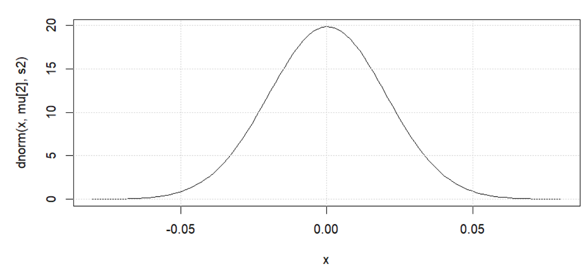
Wykres gęstości



Rysunek 1: Wykres gęstości rozkładu łącznego



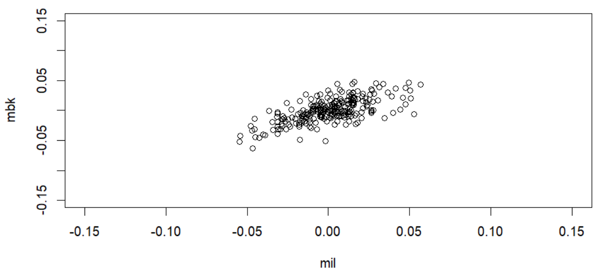
Rysunek 2: Gęstość rozkładu normalnego - Bank Millennium



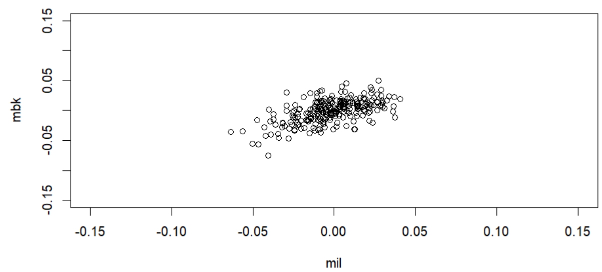
Rysunek 3: Gęstość rozkładu normalnego - mBank

Wykresy rozkładów brzegowych oraz trójwymiarowy wykres gęstości dwuwymiarowego rozkładu normalnego pokazują, że dane mają cechy typowe dla rozkładu normalnego. Widoczny jest eliptyczny kształt podstawy, wskazujący na dodatnią korelację między zmiennymi ($r = 0.67$), oraz jeden wyraźny szczyt w pobliżu punktu (0, 0), co oznacza, że najczęściej występują dni z niewielkimi zmianami cen obu akcji. Rozkład jest symetryczny i unimodalny, a gęstość szybko maleje wraz z oddalaniem się od centrum, co potwierdza, że ekstremalne zwroty są rzadkie.

Porównanie wykresów rozrzutu



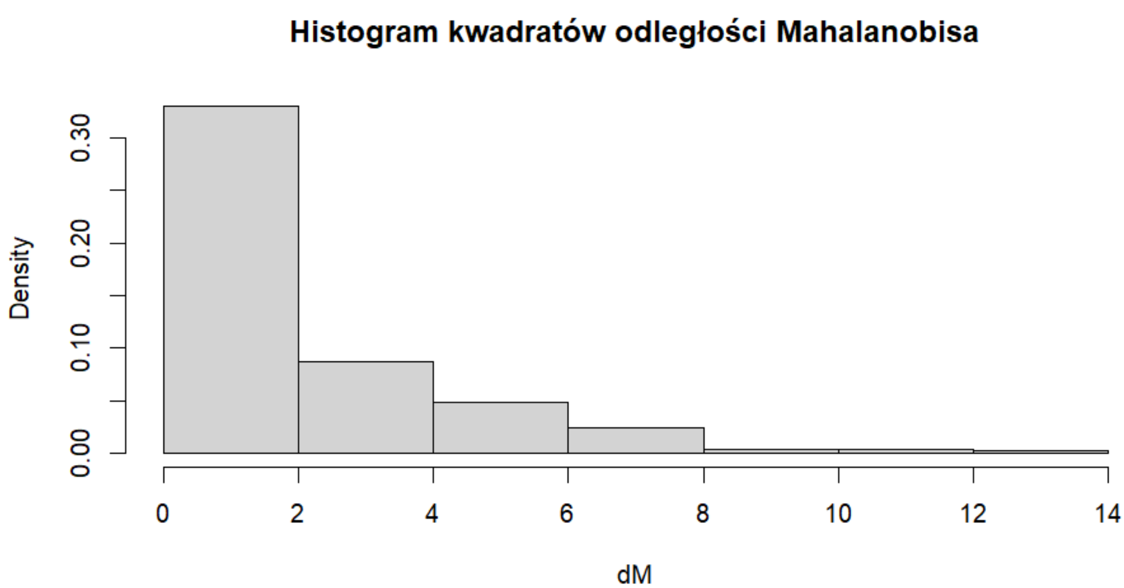
Rysunek 4: Wykres rozrzutu z danych



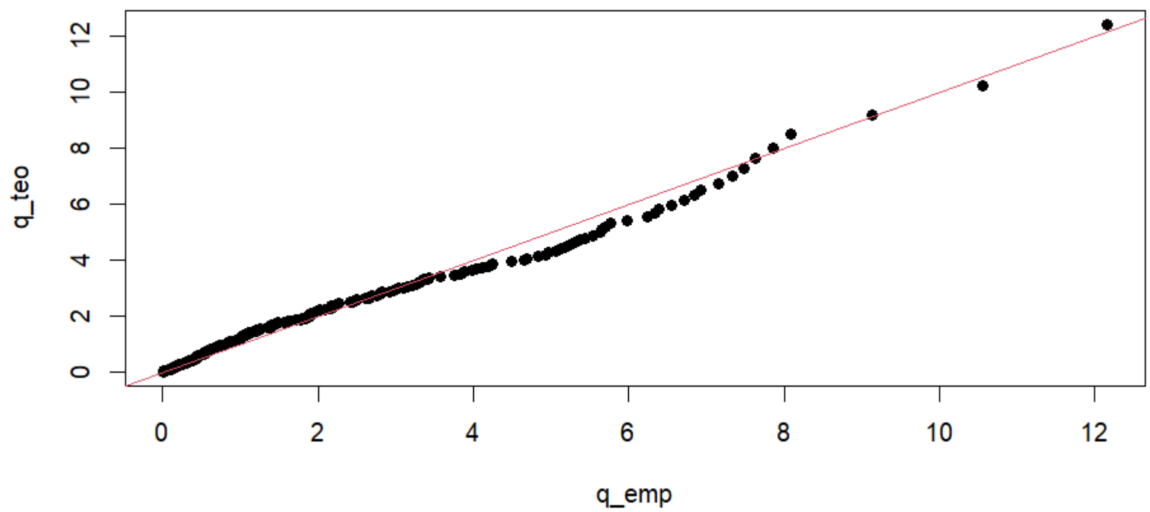
Rysunek 5: Wykres rozrzutu z próby

Na wykresach widać, że większość punktów skupia się blisko zera, co oznacza, że dzienne log-zwroty obu spółek są zazwyczaj niewielkie. Punkty układają się wzdłuż linii o dodatnim nachyleniu, co wskazuje na dodatnią korelację między mBankiem a bankiem Millennium - gdy jedna spółka zyskuje, druga zwykle też. Nie widać dużych odchyleń, więc ekstremalne zmiany są rzadkie. Oba wykresy wyglądają podobnie, co pokazuje, że próbka dobrze odzwierciedla dane rzeczywiste.

Kwadraty odległości Mahalanobisa



Analizując kształt histogramu dla log-zwrotów, stwierdzam, że jest on podobny do histogramu kwadratów odległości Mahalanobisa dla rozkładu normalnego.

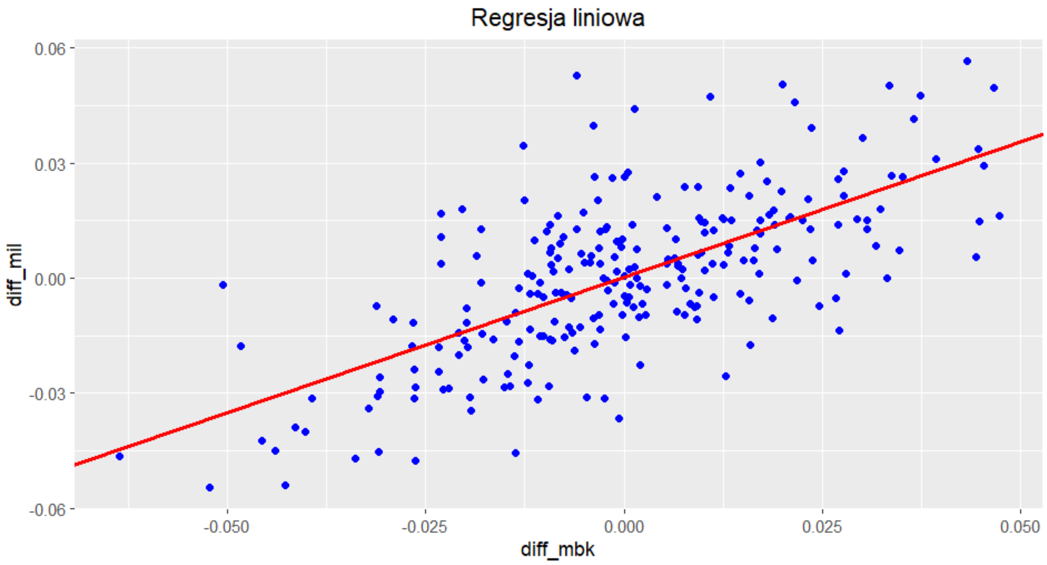


Punkty na wykresie leżą stosunkowo blisko linii prostej, więc można założyć, że rozkład empiryczny kwadratów odległości Mahalanobisa dobrze odpowiada rozkładowi teoretycznemu chi-kwadrat.

Przeprowadzam więc test zgodności oparty na statystyce Kolmogorova-Smirnova, z którego dowiaduję się, że p-value ma wartość 0.04978. Jako że p-value jest minimalnie mniejsze niż 5%, muszę odrzucić hipotezę, że kwadrat odległości Mahalanobisa ma rozkład chi(2). W konsekwencji można stwierdzić, że rozkład log-zwrotów nie jest idealnie normalny. Jednak warto zauważyć, że p-value jest bardzo bliskie granicy istotności, co sugeruje, że dane tylko nieznacznie odbiegają od normalności.

4 Regresja liniowa dla log-zwrotów

Analiza regresji



Rysunek 6: Prosta regresji na wykresie

$\beta_1 = 0.7067$
$\beta_0 = 0.00026$
Wynika z tego, że linia regresji to: $y = 0.7067x + 0.00026$

```
Call:
lm(formula = diff_mil ~ diff_mbk, data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.036469 -0.010158 -0.000886  0.008521  0.056764

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0002618  0.0009935   0.263    0.792
diff_mbk    0.7067353  0.0495336  14.268   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01565 on 246 degrees of freedom
Multiple R-squared:  0.4528,    Adjusted R-squared:  0.4506
F-statistic: 203.6 on 1 and 246 DF,  p-value: < 2.2e-16
```

Powyższy wynik pochodzi z analizy regresji liniowej, gdzie zmienna zależna `diff_mil` (log-zwroty spółki Bank Millennium) jest przewidywana przez zmienną niezależną `diff_mbk` (log-zwroty spółki mBank).

Można z niego wyczytać między innymi informację o resztach modelu (*Residuals*), takie że reszty są to różnice między rzeczywistymi wartościami a wartościami przewidywanymi przez model. W tym przypadku mają one wartości: minimalne: -0.036469, kwartyl 1Q: -0.010158, mediana: -0.000886, kwartyl 3Q: 0.008521 oraz maksymalne: 0.056764.

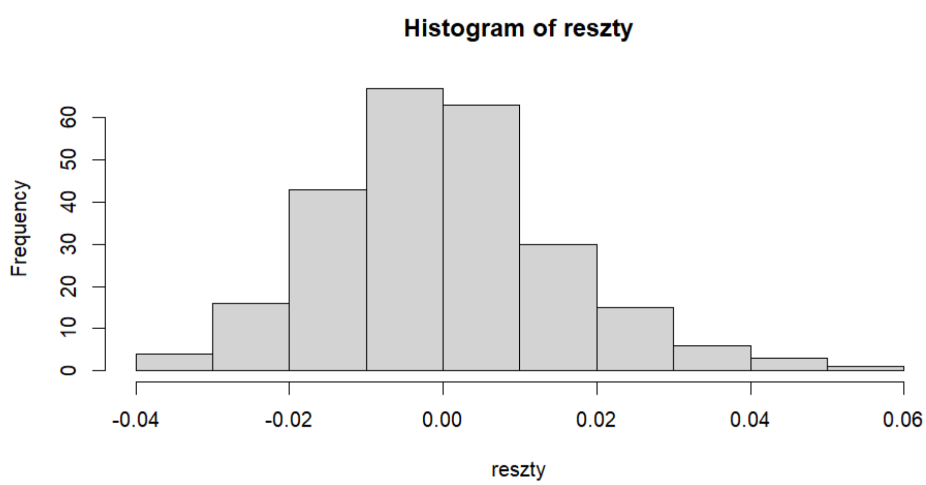
W przypadku modelu regresji dla log-zwrotów spółek mBank S.A. i Bank Millennium:

$$y = 0.00026 + 0.7067x + \epsilon, \quad \epsilon \sim N(0, 0.0.01565^2) \tag{1}$$

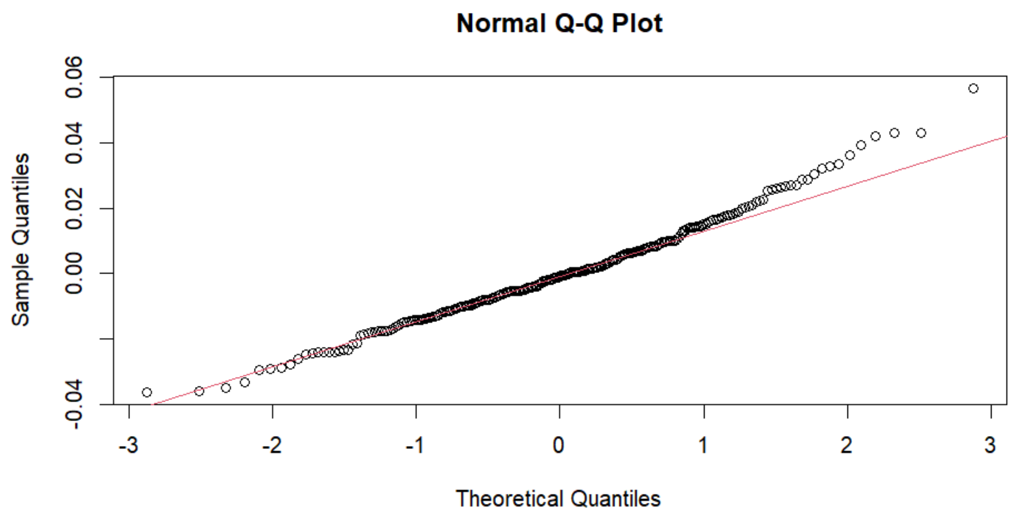
Współczynnik determinacji (R^2) wynosi 0.4528. Oznacza to, że około 45.28% zmienności log-zwrotów spółki Bank Millennium jest wyjaśnione przez log-zwroty spółki mBank. Zatem około 54.72% zmienności nie jest wyjaśnione przez log-zwroty spółki mBank.

Istotność statystyczna współczynnika β_1 jest bardzo wysoka (p-value $< 2.2 \times 10^{-16}$), co wskazuje na silną i statystycznie istotną zależność między log-zwrotami obu banków. Dodatnia wartość współczynnika $\beta_1 = 0.7067$ sugeruje, że wzrost log-zwrotów mBanku o 1 jednostkę wiąże się ze wzrostem log-zwrotów Banku Millennium o około 0.71 jednostki.

Analiza reszt



Z histogramu można wyczytać, że większość reszt ma wartości zbliżone do zera. Można zauważyć, że przypomina on rozkład normalny.



Na wykresie kwantyl-kwantyl można zauważyć, że duża część danych empirycznych pokrywa się z linią, która reprezentuje dane teoretyczne, co może sugerować, że reszta może być opisywana za pomocą rozkładu normalnego.

Testy normalności rozkładu

W tym podpunkcie sprawdzam hipotezę zerową, że rozkład reszt jest rozkładem normalnym.

Test Kolmogorova-Smirnova

$$D = 0.06209, p - value = 0.2946$$

Test Andrsona-Darlinga

$$A = 0.90617, p - value = 0.02065$$

Test Shapiro-Wilka

$$W = 0.98549, p - value = 0.01272$$

Analizując wyniki powyższych testów, test Kolmogorova-Smirnova ($p = 0.2946 > 0.05$) nie daje podstaw do odrzucenia hipotezy o normalności reszt. jednak bardziej czułe testy Andersona-Darlinga ($p = 0.02065$) i Shapiro-Wilka ($p = 0.01272$) wykrywają istotne statystycznie odchylenia od rozkładu normalnego. Wyniki testów są niejednoznaczne, jednak biorąc pod uwagę większą moc testów AD i SW, należy uznać, że reszty wykazują niewielkie, ale statystycznie istotne odchylenia od rozkładu normalnego, co skutkuje odrzuceniem hipotezy zerowej, że rozkład reszt jest rozkładem normalnym.

Test istotności współczynników b_0 i b_1

Sprawdzamy hipotezę zerową $b_0 = 0$ przeciwko hipotezie alternatywnej $b_0 \neq 0$.

Wartość statystyki testowej t_0 wynosi około 0.263. Stąd wynika, że $P(|T| > t_0) = 0.792$ (p-value) jest większe od poziomu istotności $\alpha = 5\%$, nie ma więc podstaw do odrzucenia hipotezy zerowej, że współczynnik b_0 jest równy zero.

Sprawdzamy hipotezę zerową $b_1 = 0$ przeciwko hipotezie alternatywnej $b_1 \neq 0$.

W tym przypadku wartość statystyki testowej t_1 wynosi około 14.268. Stąd też mamy wartość $P(|T| > t_1) < 2.2 \times 10^{-16}$ (p-value). Możemy więc odrzucić hipotezę zerową, że współczynnik $b_1 = 0$, na dowolnym poziomie istotności.

Wyniki testu istotności współczynników wskazują, że być może powinno się rozważyć prostszy model, w którym współczynnik b_0 miałby wartość zero.

Wynik regresji liniowej dla modelu: $y = b_1 x$

```
Call:
lm(formula = diff_mil ~ diff_mbk - 1, data = df)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.036207 -0.009895 -0.000626  0.008783  0.057026 

Coefficients:
         Estimate Std. Error t value Pr(>|t|)    
diff_mbk  0.70684    0.04944    14.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01562 on 247 degrees of freedom
Multiple R-squared:  0.4528,	Adjusted R-squared:  0.4506 
F-statistic: 204.4 on 1 and 247 DF,  p-value: < 2.2e-16
```

Można zauważyć, że oba modele są dość podobne, jednakże wydaje się, że model ze współczynnikiem b_0 może delikatnie lepiej przedstawiać dane, lecz niewiele to zmienia w kontekście analizy regresji liniowej.

Predykcja wielkości log-zwrotów spółki Bank Millennium, gdy log-zwroty spółki mBank będą na poziomie średniej z posiadanej próby

Wartość średniej log-zwrotów dla spółki mBank wynosi: m = 0.000167.

Predykcja dla modelu 1

$$\beta_0 + \beta_1 * m = 0.0003797872$$

Predykcja dla modelu 2

$$\beta_{1model2} * m = 0.0001180432$$

Na podstawie otrzymanych predykcji log-zwrotów, można stwierdzić, że oba modele przewidują bardzo zbliżone wartości, bliskie zeru. Model 1 (z wyrazem wolnym) przewiduje log-zwroty Banku Millennium na poziomie 0.0003797872, podczas gdy Model 2 (bez wyrazu wolnego) przewiduje wartość 0.0001180432.

Różnica między modelami jest bardzo mała, co potwierdza, że wyraz wolny w modelu 1 nie ma znaczącego wpływu na predykcję. Oznacza to, że gdy log-zwroty mBanku znajdują się na średniej wartości z próby, oczekiwane log-zwroty Banku Millennium będą również bliskie zeru, czyli bank nie powinien przynosić ani zysków, ani strat.

5 Podsumowanie

Celem projektu była analiza spółek Bank Millennium oraz mBank. Podzielono go na trzy etapy.

Pierwszy etap skupiał się na cenach zamknięcia akcji spółek za rok 2024. Przedstawiono w nim wykresy kursów zamknięcia, histogramy oraz szukano rozkładu, który najlepiej opisywałby te dane za pomocą wartości statystyk (KS, CM, AD), kryteriów informacyjnych (AIC, BIC) i metody Monte-Carlo. Doszedłem do wniosku, ze dla Banku Millennium najlepszym rozkładem jest rozkład log-normalny, natomiast dla mBanku, rozkład gamma.

Drugi etap skupiał się na anlizie dziennych log-zwrotów spółek za rok 2024. Analizowano tutaj rozkłady brzegowe, dobroć dopasowania danych do rozkładu normalnego oraz kwadraty odległości Mahalanobisa. Ostatecznie odrzucono hipotezę o normalności rozkładu log-zwrotów.

W trzecim etapie podjęto się analizy regresji liniowej dla log-zwrotów. Wykonano model regresji liniowej, który udało się uprościć, oraz zrobiono predykcję dla log-zwrotów Banku Millennium, gdy log-zwroty mBanku będą na poziomie średniej z posiadanej próby.