

Projekt Egzaminacyjny

Małgorzata Pyszyńska i Łukasz Kulpaczyński

Listopad 2024 - Styczeń 2025

Spis treści

1	Wstęp	2
1.1	Charakterystyka spółki Allegro.eu S.A. - Małgorzata Pyszyńska	2
1.2	Charakterystyka spółki LPP S.A. - Łukasz Kulpaczyński	2
2	Analiza log-zwrotów spółek Allegro i LPP	2
2.1	Wizualizacja danych	2
2.1.1	Allegro.eu S.A.	3
2.1.2	LPP S.A.	4
2.2	Podstawowa analiza statystyczna log-zwrotów	4
2.3	Analiza dobroci dopasowania rozkładów normalnego i t-Studenta	7
2.3.1	Estymowane parametry rozkładów	7
2.3.2	Wykresy diagnostyczne	8
2.3.3	Kryteria dopasowania i weryfikacja modelu	10
2.3.4	Interpretacja wyników	10
2.3.5	Test hipotezy o równości rozkładów dla Allegro i LPP	11
2.3.6	Podsumowanie	12
3	Analiza łącznego rozkładu log-zwrotów	12
3.1	Estymacja parametrów wektora losowego	12
3.2	Wykres rozrzutu z histogramami rozkładów brzegowych	13
3.3	Gęstość łączna i brzegowe rozkładu normalnego	14
3.4	Porównanie próbek rzeczywistych i generowanych	16
4	Regresja liniowa dla log-zwrotów	17
4.1	Przedziały ufności dla wartości oczekiwanej	17
4.2	Analiza modelu klasycznej regresji liniowej	18
4.2.1	Estymacja parametrów	18
4.2.2	Wizualizacja	19
4.2.3	Analiza reszt	20
4.2.4	Predykcja	22
4.3	Regresja uproszczona	22
4.4	Porównanie przedziałów ufności dla predykcji	23
4.4.1	Analityczny przedział ufności	23
4.4.2	Przedział ufności metodą bootstrapową	24
4.4.3	Porównanie przedziałów ufności	24
5	Podsumowanie	24

1 Wstęp

W niniejszej pracy podejmujemy się analizy log-zwrotów dwóch polskich spółek giełdowych: Allegro i LPP. Celem pracy jest szczegółowe zbadanie statystycznych właściwości danych finansowych oraz ocena ich dopasowania do modeli teoretycznych. Praca podzielona jest na trzy główne etapy.

Pierwszy etap, opisany w rozdziale *Analiza log-zwrotów spółek Allegro i LPP*, obejmuje wstępną wizualizację danych oraz podstawową analizę statystyczną. Celem tej części jest ogólne scharakteryzowanie zmienności log-zwrotów oraz ich rozkładu, co stanowi punkt wyjścia do dalszych analiz.

Drugi etap, omówiony w rozdziale *Analiza łącznego rozkładu log-zwrotów*, zakłada analizę wektora log-zwrotów dwóch spółek giełdowych, przyjmując, że log-zwroty są realizacjami wektora losowego (X, Y) . W tym rozdziale estymuje się wybrane parametry statystyczne wielowymiarowego rozkładu takie jak na przykład współczynnik korelacji i macierz kowariancji a następnie dokonuje analizy łącznego i brzegowych rozkładów.

Trzeci etap pracy to *Regresja liniowa*, w którym analizujemy zależności między log-zwrotami obu spółek za pomocą modelowania regresyjnego. Celem jest określenie wzajemnych powiązań między spółkami oraz zbadanie matematycznej struktury zależności między nimi.

Do przeprowadzenia obliczeń oraz wizualizacji wyników zastosowano środowisko RStudio, które umożliwiło efektywną realizację zaawansowanych analiz statystycznych i tworzenie wykresów ilustrujących wyniki.

1.1 Charakterystyka spółki Allegro.eu S.A. - Małgorzata Pyszńska

Allegro.eu S.A. to wiodąca polska platforma e-commerce, która łączy miliony klientów z tysiącami sprzedawców w Europie Środkowo-Wschodniej. Firma oferuje szeroki asortyment produktów w kluczowych kategoriach, takich jak motoryzacja, moda, dom i ogród, elektronika, książki, dzieci, zdrowie i uroda.

1.2 Charakterystyka spółki LPP S.A. - Łukasz Kulpaczyński

LPP S.A. to polska spółka odzieżowa, specjalizująca się w projektowaniu, produkcji oraz dystrybucji odzieży i akcesoriów pod markami Reserved, Cropp, House, Mohito i Sinsay. Firma prowadzi działalność zarówno poprzez sieć sklepów stacjonarnych, jak i kanały e-commerce, z sukcesem rozwijając swoją obecność na rynkach europejskich oraz pozaeuropejskich.

2 Analiza log-zwrotów spółek Allegro i LPP

2.1 Wizualizacja danych

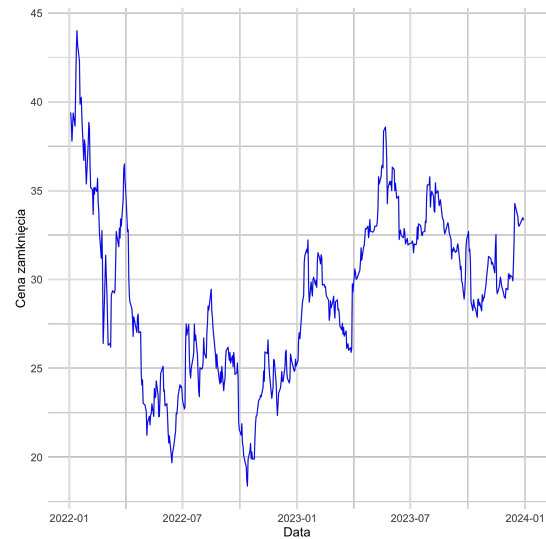
Dane analizowane w projekcie obejmują dzienne kursy zamknięcia akcji spółek Allegro.eu S.A. oraz LPP S.A. w okresie od 1 stycznia 2022 roku do 31 grudnia 2023 roku. Na ich podstawie obliczono dzienne log-zwroty według wzoru:

$$r_t = \ln \left(\frac{S_t}{S_{t-1}} \right),$$

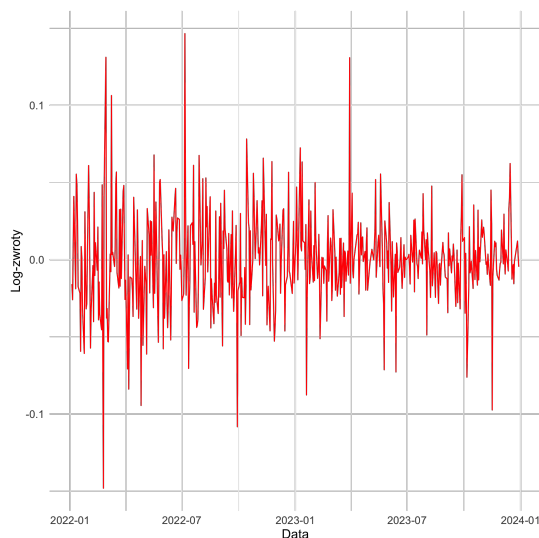
gdzie S_t oznacza kurs zamknięcia w dniu t , a S_{t-1} kurs zamknięcia w dniu poprzednim.

2.1.1 Allegro.eu S.A.

Komentarz: Kursy zamknięcia akcji Allegro.eu S.A. w latach 2022–2023. Wykres ilustruje zmienność cen akcji w czasie, z wyraźnymi wahaniami i potencjalnymi trendami.



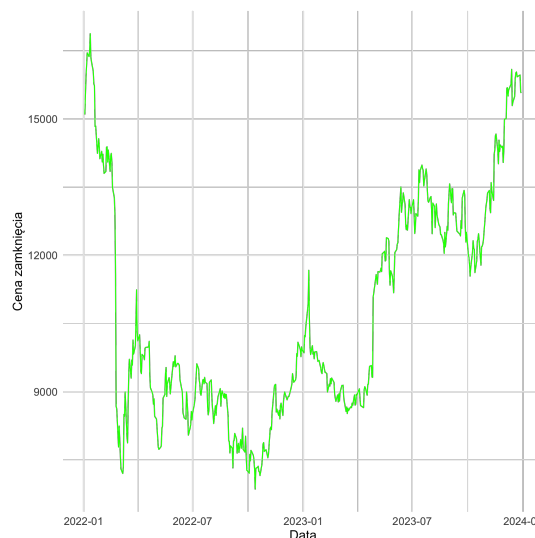
Rysunek 1: Kursy zamknięcia akcji Allegro.eu S.A. w latach 2022–2023.



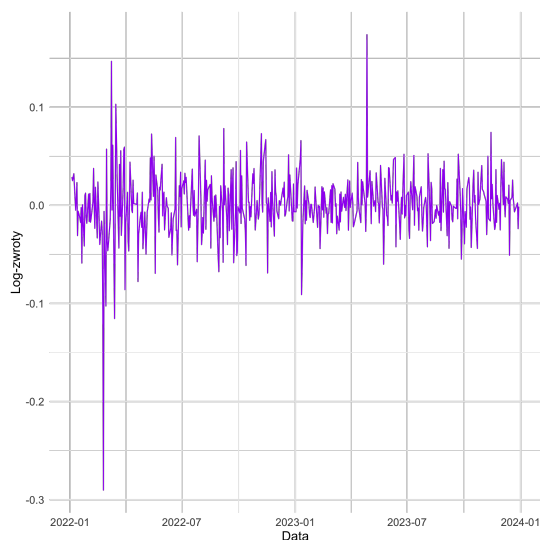
Rysunek 2: Dziennie log-zwroty akcji Allegro.eu S.A. w latach 2022–2023.

2.1.2 LPP S.A.

Komentarz: Kursy zamknięcia akcji LPP S.A. w latach 2022–2023. Wykres przedstawia zmienność cen akcji na przestrzeni dwóch lat, z wyraźnymi wahaniami sezonowymi.



Rysunek 3: Kursy zamknięcia akcji LPP S.A. w latach 2022–2023.



Rysunek 4: Dziennie log-zwroty akcji LPP S.A. w latach 2022–2023.

Komentarz: Dziennie log-zwroty akcji LPP S.A. w latach 2022–2023. Ukazano zmienność logarytmiczną cen, wskazując na dużą zmienność w wybranych okresach.

2.2 Podstawowa analiza statystyczna log-zwrotów

Zakładamy, że log-zwroty r_1, r_2, \dots, r_n są niezależnymi realizacjami zmiennej losowej X , o nieznanym dystrybucji F , wartości oczekiwanej μ i wariancji σ^2 . Estymatory klasyczne dla tych parametrów to:

- Estymator wartości oczekiwanej (średnia):

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- Estymator wariancji:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Estymator odchylenia standardowego:

$$S_n = \sqrt{S_n^2}$$

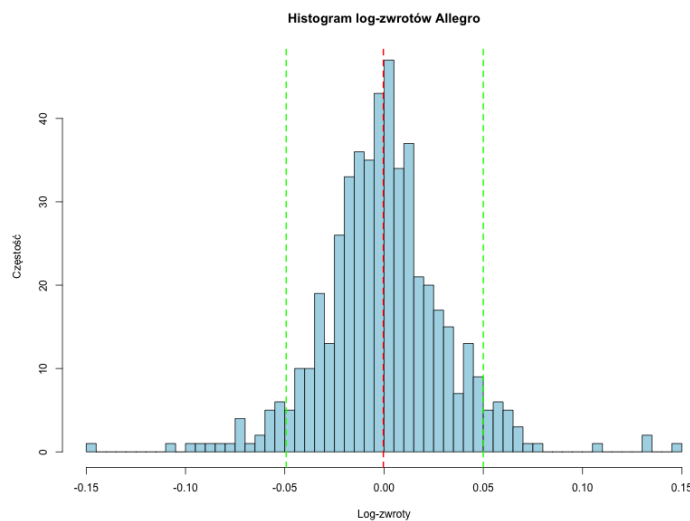
Dla kwantyli rzędu $\alpha = 5\%, 50\%, 95\%$ korzystamy z klasycznego estymatora kwantyli zaimplementowanego w R:

$$\hat{q}_\alpha = \text{quantile}(r, \alpha).$$

Spółka	Średnia	Wariancja	Odchylenie	Kwantyl 5%	Kwantyl 50%	Kwantyl 95%
Allegro	-0.00033	0.00099	0.03140	-0.04928	-0.00086	0.05000
LPP	0.00006	0.00109	0.03295	-0.04654	0.00000	0.05003

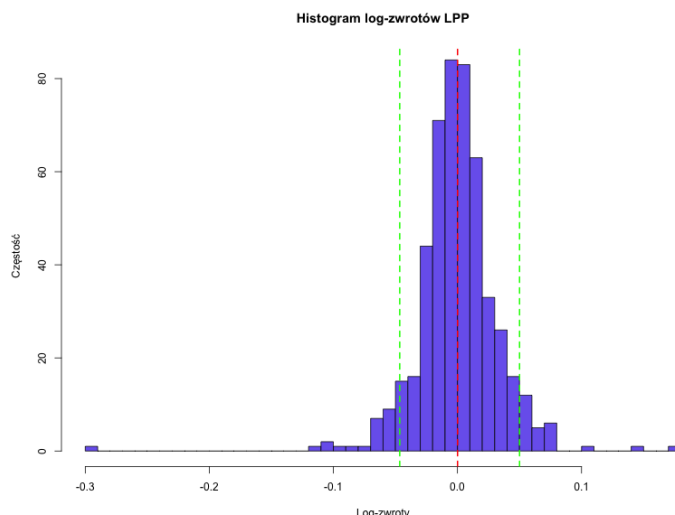
Tabela 1: Podstawowe statystyki log-zwrotów dla Allegro i LPP.

Histogram log-zwrotów dla spółki Allegro.eu S.A. pozwala na wizualną ocenę rozkładu dziennych zmian cen. Zauważalna jest koncentracja wartości wokół średniej, co sugeruje symetryczny charakter rozkładu. Zaznaczone kwantyle dodatkowo wskazują granice ekstremalnych zmian cen.



Rysunek 5: Histogram log-zwrotów dla Allegro.eu S.A. z zaznaczonymi średnią i kwantylami.

Histogram log-zwrotów dla spółki LPP S.A. również wskazuje na symetryczny charakter rozkładu, z log-zwrotami skupionymi wokół wartości średniej. Kwantyle uwidaczniają granice najczęstszych zmian cen oraz ich ekstremalne wartości.



Rysunek 6: Histogram log-zwrotów dla LPP S.A. z zaznaczonymi średnią i kwantylami.

Kwantyle pokazują, jak rozkładają się dzienne log-zwroty cen akcji. Wyznaczone wartości dla Allegro i LPP są następujące:

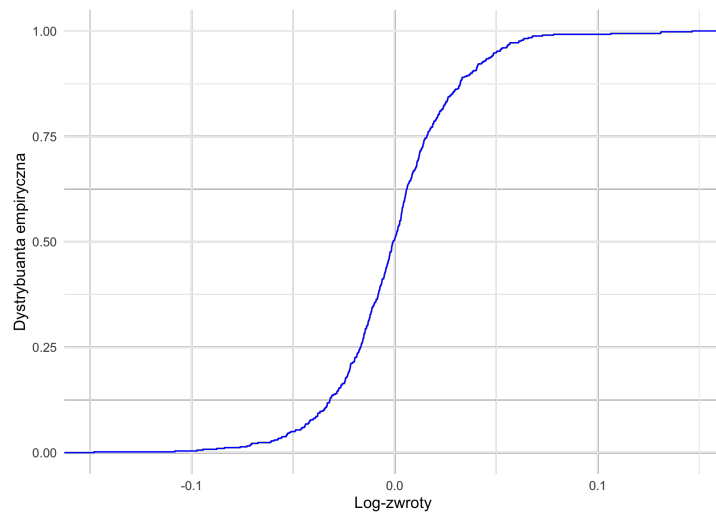
- **Kwantyl 5%:** W najgorszych 5% dni log-zwroty były niższe niż -4.93% dla Allegro i -4.65% dla LPP.
- **Kwantyl 50% (mediana):** Połowa log-zwrotów była niższa lub równa -0.086% dla Allegro i 0.00% dla LPP.
- **Kwantyl 95%:** W najlepszych 5% dni log-zwroty przekraczały 5.00% dla Allegro i 5.03% dla LPP.

W celu oszacowania dystrybuant empirycznych log-zwrotów $\hat{F}(x)$ dla spółek Allegro.eu S.A. oraz LPP S.A. zastosowano klasyczny estymator dystrybuanty:

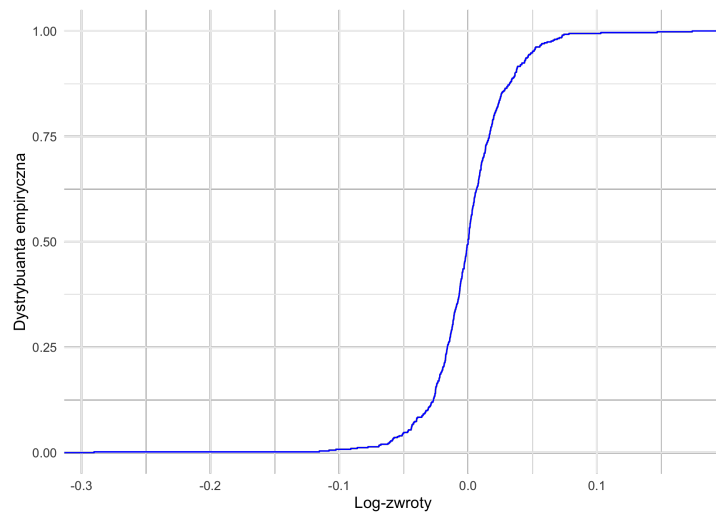
$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}},$$

gdzie $1_{\{X_i \leq x\}}$ to wskaźnik przyjmujący wartość 1, jeśli $X_i \leq x$, a 0 w przeciwnym przypadku. Funkcja \hat{F}_n , zwana **dystrybuantą empiryczną**, jest funkcją schodkową, której kształt odzwierciedla rozkład log-zwrotów.

Na rysunku 7 przedstawiono dystrybuantę empiryczną dla Allegro.eu S.A., natomiast na rysunku 8 zaprezentowano odpowiedni wykres dla LPP S.A. Oba wykresy ukazują charakterystyczny schodkowy przebieg funkcji, zgodny z istotą dystrybuant empirycznych.



Rysunek 7: Dystrybuanta empiryczna log-zwrotów dla Allegro.eu S.A.



Rysunek 8: Dystrybuanta empiryczna log-zwrotów dla LPP S.A.

2.3 Analiza dobroci dopasowania rozkładów normalnego i t-Studenta

W celu oceny dobroci dopasowania danych log-zwrotów spółek Allegro i LPP do rozkładów normalnego oraz t-Studenta zastosowano estymację parametrów metodą największej wiarygodności (MLE). Rozkłady te, ze względu na swoje właściwości, są szeroko stosowane w analizie danych finansowych.

2.3.1 Estymowane parametry rozkładów

Parametry dopasowanych rozkładów normalnego oraz t-Studenta oszacowano przy użyciu biblioteki `fitdistrplus`. Wyniki estymacji przedstawiono w tabeli 2.

Spółka	Rozkład	Parametry
Allegro	Normalny	$\mu = -0.0001, \sigma = 0.0201$
Allegro	t-Studenta	$\mu = -0.0001, \sigma = 0.0198, df = 9.7$
LPP	Normalny	$\mu = 0.0002, \sigma = 0.0189$
LPP	t-Studenta	$\mu = 0.0002, \sigma = 0.0186, df = 10.1$

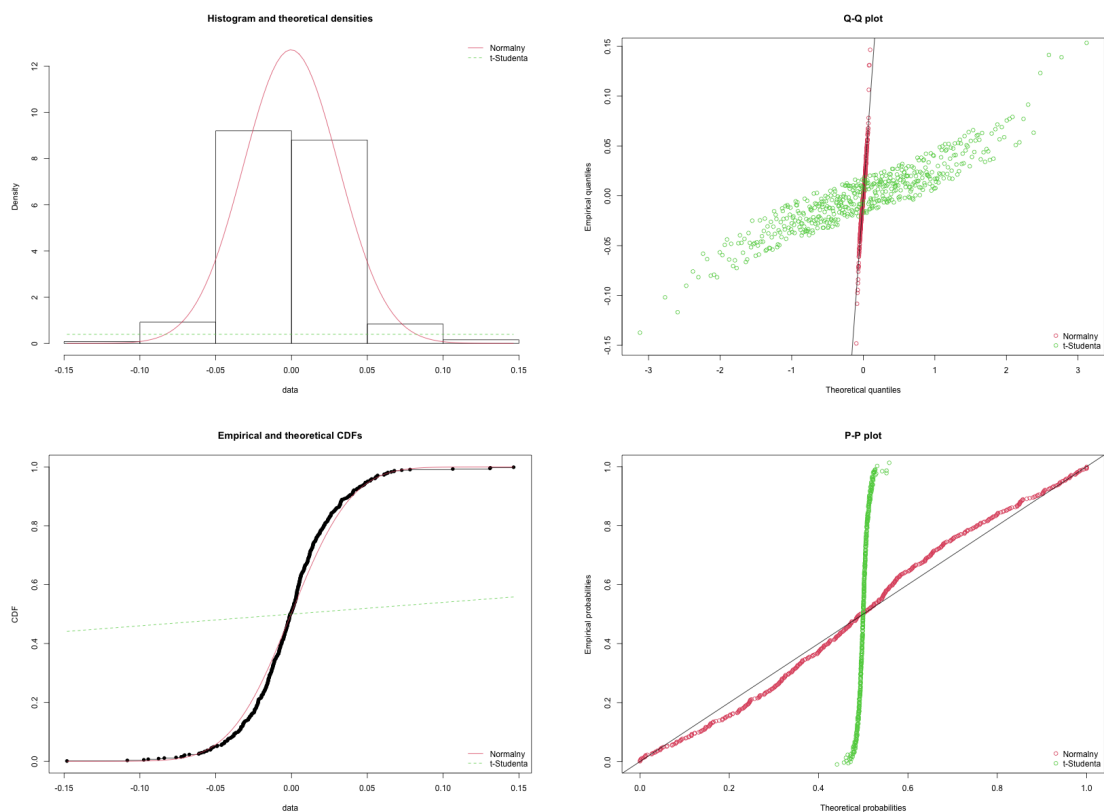
Tabela 2: Estymowane parametry dopasowanych rozkładów.

2.3.2 Wykresy diagnostyczne

Dla obu spółek wygenerowano cztery rodzaje wykresów diagnostycznych, które przedstawiono na rysunkach 9 i 10:

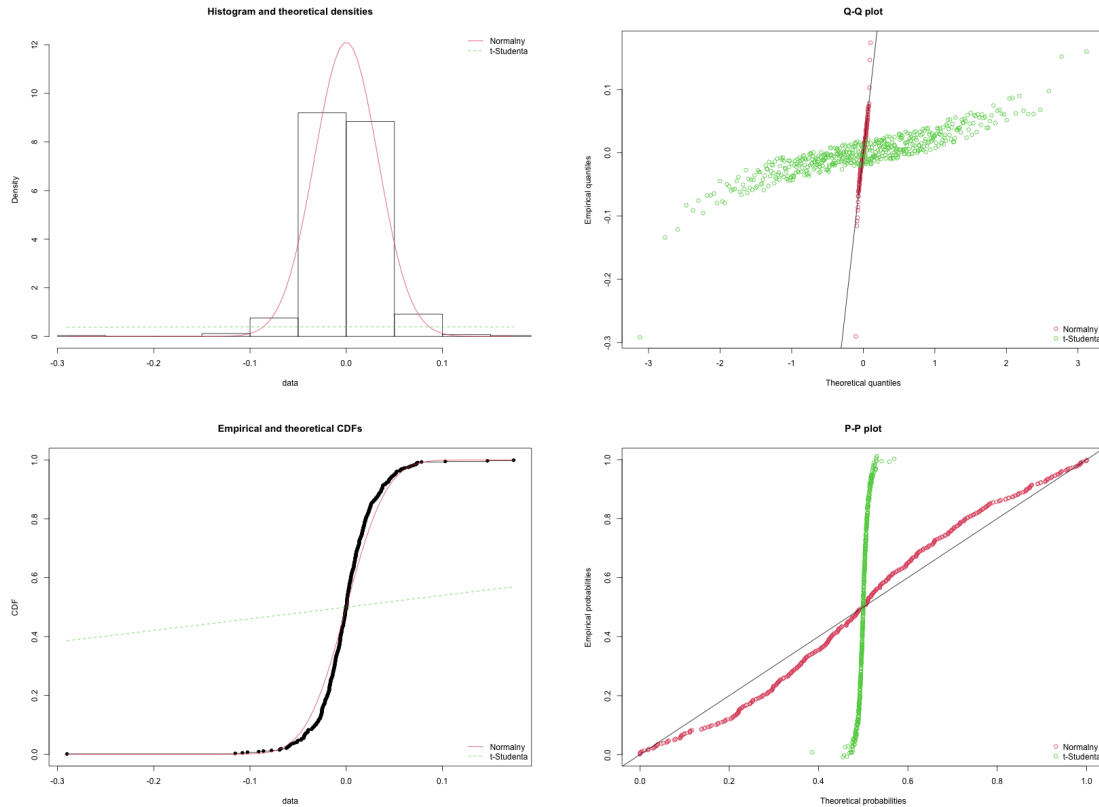
- **Porównanie gęstości empirycznych i teoretycznych:** umożliwia ocenę, jak dobrze rozkłady teoretyczne odwzorowują empiryczną gęstość danych.
- **QQ-plot:** przedstawia dopasowanie kwantyli danych do kwantyli teoretycznych, co pozwala ocenić zgodność rozkładów.
- **Dystrybuanta empiryczna vs teoretyczna:** porównanie funkcji rozkładów kumulacyjnych.
- **P-P plot:** ukazuje proporcję punktów danych w porównaniu z teoretyczną proporcją dla danego rozkładu.

Wykresy dla Allegro. Na rysunku 9 przedstawiono wykresy diagnostyczne dla log-zwrotów spółki Allegro. Wykresy wskazują na pewne odchylenia od rozkładu normalnego i t-Studenta, szczególnie widoczne na QQ-plotach.



Rysunek 9: Wykresy diagnostyczne dla Allegro: gęstości (górną lewą), QQ-plot (górną prawą), dystrybuanta empiryczna (dół lewą) oraz P-P plot (dół prawą).

Wykresy dla LPP. Analogiczne wykresy diagnostyczne dla spółki LPP przedstawiono na rysunku 10. Rozkład normalny wydaje się lepiej odwzorowywać dane w porównaniu do t-Studenta, co widoczne jest szczególnie na QQ-plotach i P-P plotach.



Rysunek 10: Wykresy diagnostyczne dla LPP: gęstości (góra lewa), QQ-plot (góra prawa), dystrybuanta empiryczna (dół lewa) oraz P-P plot (dół prawa).

2.3.3 Kryteria dopasowania i weryfikacja modelu

Do oceny jakości dopasowania wykorzystano statystyki Kolmogorowa-Smirnowa (KS), Craméra-von Misesa (CM), Andersona-Darlinga (AD) oraz kryteria informacyjne Akaikego (AIC) i Bayesa (BIC). Wyniki przedstawiono w tabeli 3.

Spółka	Rozkład	KS	CM	AD	AIC	BIC
Allegro	Normalny	0.0603	0.5583	3.2033	-2038.016	-2029.587
Allegro	t-Studenta	0.4610	38.4045	180.0888	922.3521	926.5667
LPP	Normalny	0.0816	1.2052	6.9794	-1989.767	-1981.338
LPP	t-Studenta	0.4628	38.4521	180.2731	922.4021	926.6167

Tabela 3: Statystyki dopasowania i kryteria informacyjne dla rozkładów normalnego i t-Studenta.

2.3.4 Interpretacja wyników

Na podstawie wyników z tabeli 3 można zauważyć, że:

- Rozkład normalny lepiej opisuje dane niż rozkład t-Studenta, co widoczne jest w znacznie niższych wartościach AIC i BIC dla obu spółek.
- Statystyki dobroci dopasowania (KS, CM, AD) wskazują na mniejsze odchylenia od rozkładu normalnego w porównaniu do t-Studenta.

- Pomimo grubszych ogonów w rozkładzie t-Studenta, jego dopasowanie do analizowanych danych jest słabsze.

Na tej podstawie jako najlepszy model opisujący dane log-zwrotów dla obu spółek wybrano rozkład normalny.

2.3.5 Test hipotezy o równości rozkładów dla Allegro i LPP

W celu przetestowania hipotezy zerowej o zgodności rozkładów log-zwrotów spółek Allegro i LPP z rozkładem normalnym wykorzystano metodę Monte Carlo oraz statystykę Kolmogorowa-Smirnowa (KS). Algorytm testowania przedstawia się następująco:

1. Obliczono wartość statystyki D_n dla dystrybuant F_0 (teoretycznej) i \hat{F}_n (empirycznej).
2. Wyznaczono rozkład statystyki D_n przy użyciu $N = 10,000$ symulacji z rozkładu F_0 , każda o liczności $n = 500$.
3. Oszacowano wartość p jako proporcję próbek, dla których $D_n > d_n$.
4. Przyjęto poziom istotności $\alpha = 0.05$. Jeśli $p \leq \alpha$, odrzucano hipotezę zerową H_0 , w przeciwnym razie brak było podstaw do jej odrzucenia.

Test dla Allegro. Hipotezy:

- H_0 : Rozkład log-zwrotów spółki Allegro jest zgodny z rozkładem normalnym.
- H_1 : Rozkład log-zwrotów spółki Allegro nie jest zgodny z rozkładem normalnym.

Wyniki testu dla log-zwrotów spółki Allegro przedstawiono w tabeli 4.

Statystyka KS	p-wartość
0.0603	0.0452

Tabela 4: Wyniki testu Monte Carlo dla spółki Allegro.

Interpretacja wyników:

- Statystyka KS wynosi 0.0603, co wskazuje na niewielkie odchylenie między dystrybuantą empiryczną a teoretyczną.
- p-wartość 0.0452 jest mniejsza niż $\alpha = 0.05$, co prowadzi do odrzucenia hipotezy H_0 .
- Ostateczny wniosek: **Hipoteza o normalności rozkładu log-zwrotów dla Allegro została odrzucona.**

Test dla LPP. Hipotezy:

- H_0 : Rozkład log-zwrotów spółki LPP jest zgodny z rozkładem normalnym.
- H_1 : Rozkład log-zwrotów spółki LPP nie jest zgodny z rozkładem normalnym.

Wyniki testu dla log-zwrotów spółki LPP przedstawiono w tabeli 5.

Statystyka KS	p-wartość
0.0816	0.0026

Tabela 5: Wyniki testu Monte Carlo dla spółki LPP.

Interpretacja wyników:

- Statystyka KS wynosi 0.0816, co oznacza większe odchylenie od modelu normalnego w porównaniu do spółki Allegro.
- p-wartość 0.0026 jest znacznie mniejsza niż $\alpha = 0.05$, co prowadzi do odrzucenia hipotezy H_0 .
- Ostateczny wniosek: **Hipoteza o normalności rozkładu log-zwrotów dla LPP została odrzucona.**

2.3.6 Podsumowanie

Wyniki testów wskazują, że:

- Rozkład log-zwrotów dla spółki Allegro nie jest zgodny z rozkładem normalnym na poziomie istotności $\alpha = 0.05$.
- Dla spółki LPP hipoteza o zgodności rozkładu log-zwrotów z rozkładem normalnym została również odrzucona. Rzeczywisty rozkład wykazuje istotne odchylenia od modelu normalnego.

3 Analiza łącznego rozkładu log-zwrotów

3.1 Estymacja parametrów wektora losowego

Założmy, że log-zwroty dwóch akcji są niezależnymi realizacjami wektora losowego (X, Y) o nieznaną gęstość f , wektorze średnich μ , współczynniku korelacji ρ , macierzy kowariancji Σ i macierzy korelacji P .

Wzory estymatorów parametrów wektora losowego są następujące:

- Wektor średnich:

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_X \\ \hat{\mu}_Y \end{bmatrix}, \quad \hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\mu}_Y = \frac{1}{n} \sum_{i=1}^n Y_i$$

- Macierz kowariancji:

$$\hat{\Sigma} = \begin{bmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{bmatrix}$$

- Współczynnik korelacji:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

- Macierz korelacji:

$$\hat{P} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Na podstawie danych log-zwrotów dla spółek Allegro i LPP oszacowano kluczowe parametry statystyczne.

Wektor średnich

$$\hat{\mu} = \begin{bmatrix} \text{Allegro: } -0.0003349 \\ \text{LPP: } 0.0000620 \end{bmatrix}$$

Wektor średnich $\hat{\mu}$ wskazuje na niewielkie średnie log-zwroty dla obu spółek, zbliżone do zera, co sugeruje brak istotnych trendów w zmianach dziennych cen akcji.

Macierz kowariancji

$$\hat{\Sigma} = \begin{bmatrix} 0.0009879 & 0.0003974 \\ 0.0003974 & 0.0010880 \end{bmatrix}$$

Macierz kowariancji $\hat{\Sigma}$ obrazuje zróżnicowanie i współzmiennność log-zwrotów dla obu spółek. Element $\Sigma_{12} = 0.0003974$ wskazuje na umiarkowaną pozytywną współzmiennność pomiędzy Allegro a LPP. Wartości na diagonalu macierzy (Σ_{11} i Σ_{22}) odzwierciedlają zmienność log-zwrotów odpowiednio dla Allegro i LPP.

Współczynnik korelacji

$$\hat{\rho} = 0.3833$$

Współczynnik korelacji $\hat{\rho} = 0.3833$ wskazuje na umiarkowaną dodatnią zależność pomiędzy log-zwrotami Allegro i LPP. Oznacza to, że zmienność cen akcji obu spółek wykazuje pewne podobieństwo, ale nie jest to zależność silna.

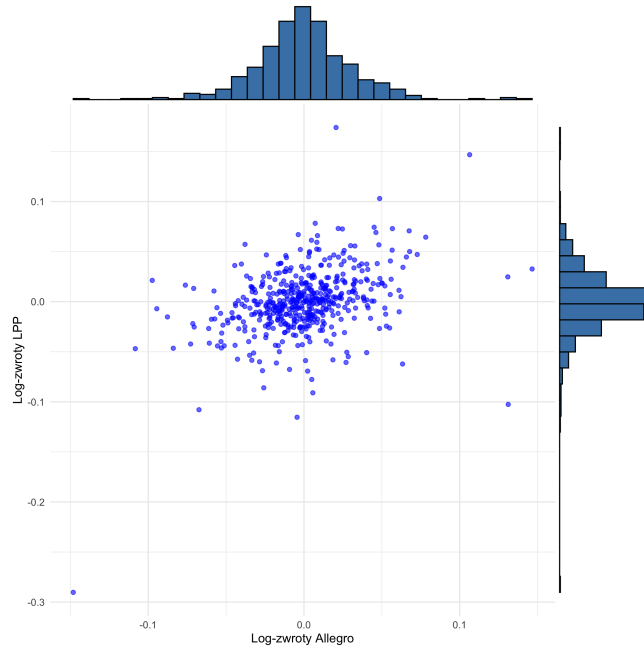
Macierz korelacji

$$\hat{P} = \begin{bmatrix} 1.0000 & 0.3833 \\ 0.3833 & 1.0000 \end{bmatrix}$$

Macierz korelacji \hat{P} przedstawia zależności między log-zwrotami obu spółek w sposób standaryzowany. Element $P_{12} = 0.3833$ potwierdza dodatnią korelację między Allegro a LPP, natomiast wartości $P_{11} = P_{22} = 1.0000$ wskazują na doskonałą korelację zmiennej z samą sobą.

3.2 Wykres rozrzutu z histogramami rozkładów brzegowych

Poniżej przedstawiono wykres rozrzutu log-zwrotów spółek Allegro i LPP, wraz z histogramami ich rozkładów brzegowych. Wykres ten umożliwia ocenę zależności między log-zwrotami obu spółek, a histogramy obrazują rozkłady zmiennych na osiach x i y .



Rysunek 11: Wykres rozrzutu log-zwrotów spółek Allegro i LPP z histogramami rozkładów brzegowych.

3.3 Gęstość łączna i brzegowe rozkładu normalnego

W analizie danych log-zwrotów spółek Allegro i LPP przyjęto założenie, że ich wektor log-zwrotów podąża za dwuwymiarowym rozkładem normalnym $N(\mu, \Sigma)$, gdzie:

- $\mu = (\mu_1, \mu_2)$ to wektor wartości oczekiwanych, reprezentujący średnie log-zwroty dla obu spółek,
- Σ to macierz kowariancji, definiowana jako:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

gdzie σ_1, σ_2 to odchylenia standardowe, a ρ to współczynnik korelacji między log-zwrotami spółek.

Wzór gęstości łącznej $f(x, y)$ pozwala opisać rozkład prawdopodobieństwa jednoczesnych log-zwrotów obu spółek:

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

Ten wzór pozwala zobrazować, w jaki sposób wartości log-zwrotów dla jednej spółki są powiązane z drugą.

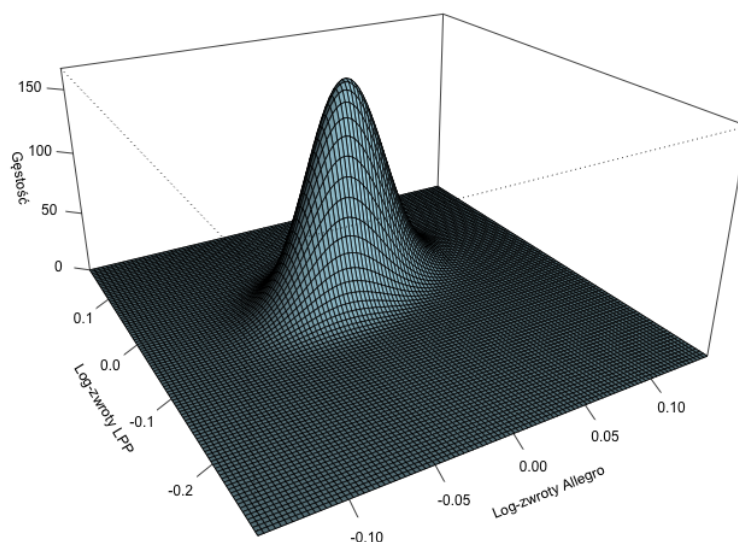
Gęstości brzegowe $f_X(x)$ i $f_Y(y)$ opisują rozkłady log-zwrotów dla poszczególnych spółek:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right),$$

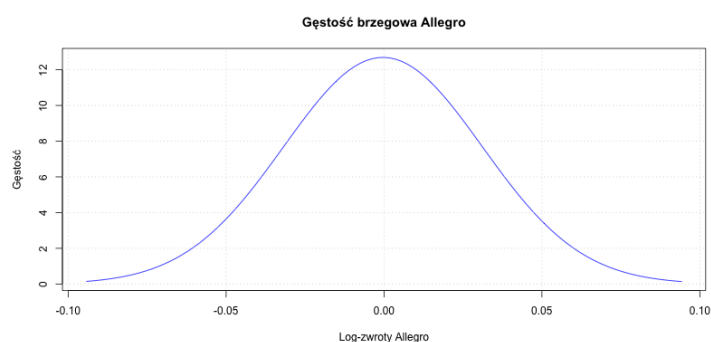
$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right).$$

Gęstości te odzwierciedlają prawdopodobieństwo uzyskania konkretnego log-zwrotu dla każdej ze spółek, niezależnie od wyników drugiej.

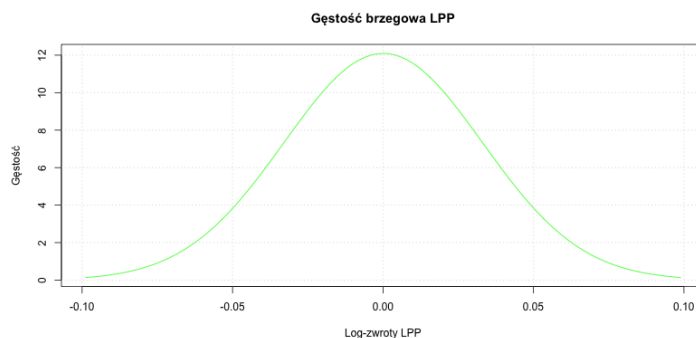
Na rysunku 12 zaprezentowano gęstość łączną, która ukazuje wzajemne zależności log-zwrotów spółek Allegro i LPP. Z kolei rysunki 13 i 14 przedstawiają gęstości brzegowe dla Allegro i LPP, umożliwiając analizę rozkładu log-zwrotów dla każdej spółki z osobna.



Rysunek 12: Wykres gęstości łącznej log-zwrotów spółek Allegro i LPP.



Rysunek 13: Gęstość brzegowa log-zwrotów spółki Allegro.

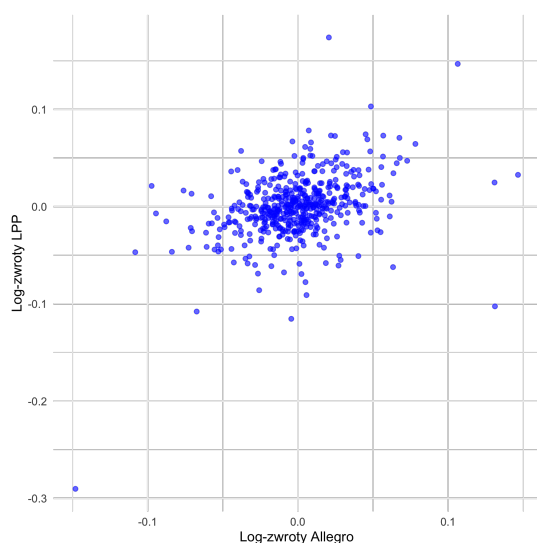


Rysunek 14: Gęstość brzegowa log-zwrotów spółki LPP.

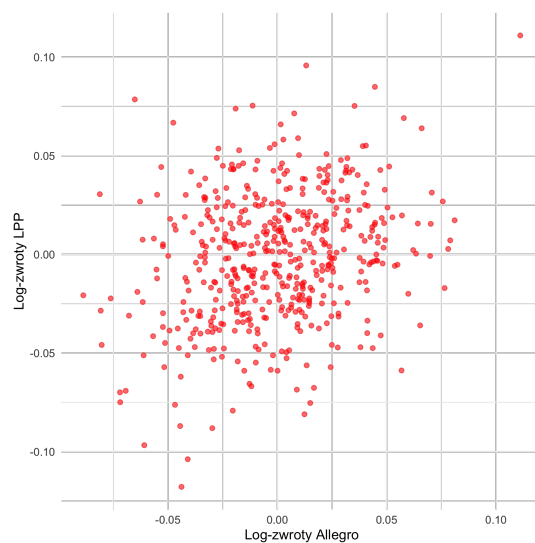
3.4 Porównanie próbek rzeczywistych i generowanych

W celu oceny, czy rozkład normalny dobrze opisuje dane log-zwrotów dla spółek Allegro i LPP, wygenerowano próbkę o liczności odpowiadającej liczbie obserwacji w danych rzeczywistych. Do generowania próby wykorzystano wyestymowane parametry $\hat{\mu}$ (wektor średnich) i $\hat{\Sigma}$ (macierz kowariancji). Następnie porównano wykresy rozrzutu dla danych rzeczywistych oraz próbki generowanej z rozkładu $N(\hat{\mu}, \hat{\Sigma})$.

Na rysunku 15 przedstawiono wykresy rozrzutu obok siebie: po lewej stronie dane rzeczywiste, a po prawej próbkę generowaną z rozkładu normalnego. Taki układ umożliwia łatwe porównanie obu rozkładów i ocenę zgodności.



(a) Dane rzeczywiste.



(b) Próbką z rozkładu $N(\hat{\mu}, \hat{\Sigma})$.

Rysunek 15: Porównanie wykresów rozrzutu dla danych rzeczywistych (a) i próbki generowanej z rozkładu normalnego (b).

Porównując wykresy, można zauważyć, że próbka generowana z rozkładu normalnego zachowuje ogólne właściwości struktury danych rzeczywistych. Niemniej różnice w koncentracji punktów w niektórych obszarach wskazują, że rozkład normalny nie w pełni opisuje rzeczywisty charakter log-zwrotów. Wyniki sugerują, że rozkład normalny w pewnym stopniu modeluje dane, ale nie uwzględnia bardziej złożonych zależności.

4 Regresja liniowa dla log-zwrotów

W celu analizy zależności między log-zwrotami spółek Allegro i LPP zastosowano klasyczny model regresji liniowej. Analizowane dane obejmują dzienne log-zwroty, których definicję podano w sekcji *Analiza log-zwrotów spółek Allegro i LPP*. Model regresji pozwala na ocenę wzajemnych powiązań między spółkami oraz ich predykcję na podstawie wartości historycznych.

4.1 Przedziały ufności dla wartości oczekiwanej

Przedziały ufności (CI – Confidence Interval) umożliwiają oszacowanie przedziału, w którym z określonym poziomem ufności $1 - \alpha$ znajduje się nieznany parametr (np. średnia μ). Ze względu na odrzucenie hipotezy o normalności rozkładu log-zwrotów w teście Monte Carlo, klasyczne podejście do wyznaczania przedziałów ufności, oparte na rozkładzie t -Studenta, nie jest odpowiednie. W związku z tym zastosowano metodę bootstrapową, umożliwiającą nieparametryczne oszacowanie przedziału ufności dla wartości oczekiwanej.

Przy wykorzystaniu $B = 1000$ próbek bootstrapowych, przedział ufności został wyznaczony jako:

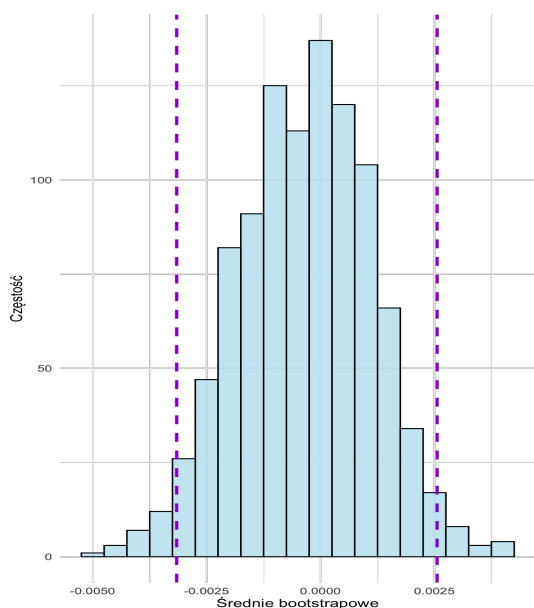
$$[q_{\alpha/2}^*, q_{1-\alpha/2}^*],$$

gdzie $q_{\alpha/2}^*$ i $q_{1-\alpha/2}^*$ oznaczają odpowiednie kwantyle empiryczne uzyskane z rozkładu statystyki bootstrapowej.

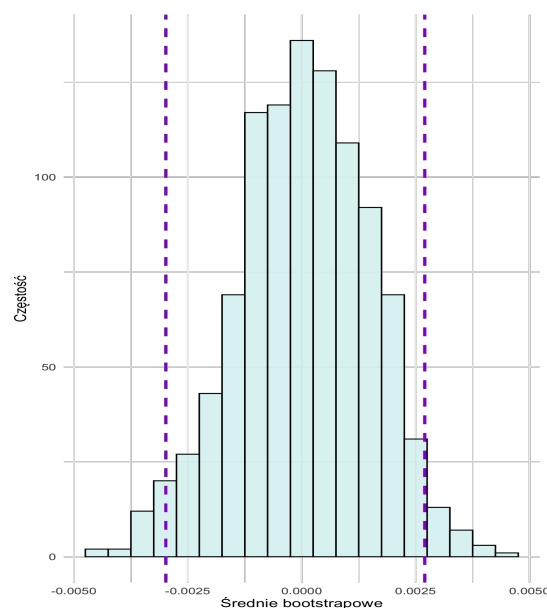
Wyniki dla obu spółek:

- Allegro: $[-0.003216, 0.002267]$,
- LPP: $[-0.003243, 0.002879]$.

Na Rysunkach 16a i 16b przedstawiono histogramy bootstrapowych średnich.



(a) Histogram bootstrapowych średnich dla log-zwrotów spółki Allegro.



(b) Histogram bootstrapowych średnich dla log-zwrotów spółki LPP.

Rysunek 16: Histogramy bootstrapowych średnich dla spółek Allegro i LPP.

Oba histogramy wykazują kształt zbliżony do rozkładu normalnego, co sugeruje, że estymacja średniej przy użyciu metody bootstrapowej jest stabilna. Granice przedziałów ufności oznaczone liniami przerywanymi wskazują zakres zmienności oszacowań średnich w kolejnych próbach bootstrapowych.

W przypadku spółki Allegro, przedział ufności dla wartości oczekiwanej jest stosunkowo wąski, co wskazuje na niską wariancję estymatora. Natomiast dla spółki LPP rozstęp przedziału ufności jest nieco większy, co może wynikać z większej zmienności log-zwrotów tej spółki.

4.2 Analiza modelu klasycznej regresji liniowej

Model regresji liniowej opisuje zależność między log-zwrotami R_1 i R_2 za pomocą równania:

$$R_2 = b_0 + b_1 R_1 + \varepsilon,$$

gdzie b_0 i b_1 to współczynniki regresji, a składnik losowy ε spełnia założenie $\varepsilon \sim N(0, \sigma^2)$.

4.2.1 Estymacja parametrów

Parametry modelu, b_0 oraz b_1 , estymowane są metodą najmniejszych kwadratów, minimalizującą sumę kwadratów reszt. Współczynnik b_1 interpretuje się jako miarę siły zależności między zmiennymi, natomiast b_0 reprezentuje wyraz wolny modelu.

Sformułowanie hipotez statystycznych Aby ocenić istotność współczynników regresji, przeprowadzamy testy statystyczne:

- **Hipoteza zerowa (H_0):** Współczynnik regresji jest równy zero ($b_i = 0$), co oznacza brak wpływu R_1 na R_2 .
- **Hipoteza alternatywna (H_1):** Współczynnik regresji jest różny od zera ($b_i \neq 0$), co oznacza, że istnieje istotny wpływ R_1 na R_2 .

Test przeprowadzamy na poziomie istotności $\alpha = 0.05$. Wartość p (p-value) określa prawdopodobieństwo uzyskania wyników równie ekstremalnych, co obserwowane, zakładając prawdziwość hipotezy zerowej. Jeśli $p \leq 0.05$, odrzucamy H_0 , co oznacza istotność statystyczną danego współczynnika.

Wyznaczenie współczynników regresji Parametry b_0 i b_1 estymujemy na podstawie wzorów:

$$b_1 = \frac{\sum_{i=1}^n (R_{1,i} - \bar{R}_1)(R_{2,i} - \bar{R}_2)}{\sum_{i=1}^n (R_{1,i} - \bar{R}_1)^2}, \quad b_0 = \bar{R}_2 - b_1 \cdot \bar{R}_1.$$

Wyniki regresji Wyniki estymacji parametrów modelu regresji przedstawiono w tabeli 6.

Parametr	Wartość	p -value
Wyraz wolny (b_0)	0.0002	0.885 (nieistotny)
Współczynnik nachylenia (b_1)	0.4023	$< 2 \cdot 10^{-16}$ (istotny)
Współczynnik determinacji (R^2)	0.147	–
Błąd standardowy reszt (RSE)	0.0305	–

Tabela 6: Wyniki estymacji parametrów modelu regresji.

Analiza wyników regresji

- **Współczynnik nachylenia (b_1):**

- Wartość $b_1 = 0.4023$ oznacza, że wzrost R_1 o jednostkę skutkuje średnim wzrostem R_2 o około 0.4023 jednostki.
- Bardzo niska wartość $p < 2 \cdot 10^{-16}$ sugeruje, że współczynnik jest **statystycznie istotny**, co pozwala odrzucić hipotezę zerową (H_0).
- Oznacza to, że istnieje istotna zależność między R_1 i R_2 .

- **Wyraz wolny (b_0):**

- Wartość $b_0 = 0.0002$ jest bardzo bliska zeru.
- Wysoka wartość $p = 0.885$ sugeruje, że nie można odrzucić hipotezy zerowej (H_0), co oznacza, że wyraz wolny **nie jest istotny**.
- Brak istotności b_0 wskazuje, że model dobrze odwzorowuje relację między zmiennymi bez konieczności uwzględniania stałej.

- **Współczynnik determinacji (R^2):**

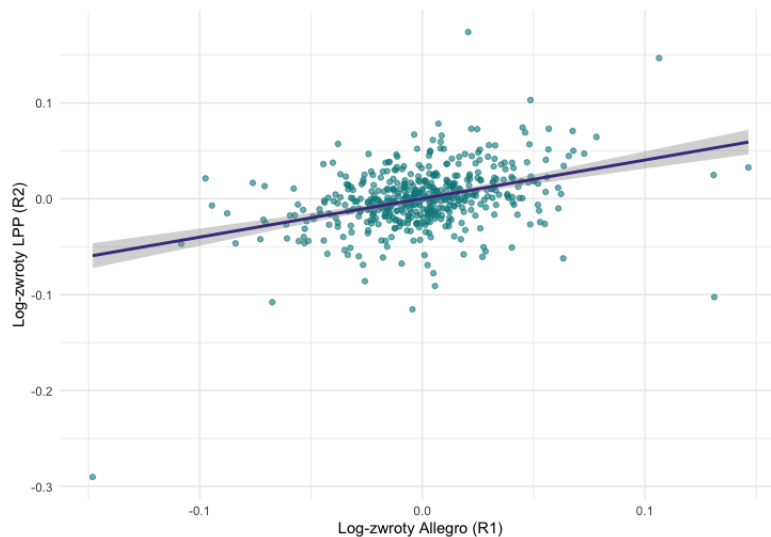
- Wartość $R^2 = 0.147$ oznacza, że model wyjaśnia **tylko 14.7%** zmienności R_2 .
- Niska wartość R^2 sugeruje, że istnieją inne istotne czynniki wpływające na R_2 , które nie zostały uwzględnione w modelu.

- **Błąd standardowy reszt (RSE):**

- Wartość $RSE = 0.0305$ wskazuje na przeciętne odchylenie wartości przewidywanych od rzeczywistych.
- Relatywnie wysoka wartość RSE sugeruje, że model może nie być bardzo dokładny.

4.2.2 Wizualizacja

Rysunek 17 przedstawia wykres rozrzutu R_1 i R_2 z naniesioną linią regresji.



Rysunek 17: Wykres rozrzutu log-zwrotów spółek Allegro (R_1) i LPP (R_2) z linią regresji.

Na przedstawionym wykresie możemy zauważyć, że dane wykazują umiarkowaną dodatnią zależność pomiędzy log-zwrotami spółek Allegro (R_1) i LPP (R_2). Punkty danych są rozproszone wokół linii regresji, co wskazuje, że chociaż istnieje pewien związek między zmiennymi, to nie jest on bardzo silny.

Obecność punktów znacznie oddalonych od linii regresji sugeruje istnienie wartości odstających, które mogą wpływać na dokładność modelu. Ponadto szerokie rozproszenie punktów wokół linii regresji potwierdza niską wartość współczynnika determinacji $R^2 = 0.147$, co oznacza, że model wyjaśnia tylko niewielką część zmienności log-zwrotów LPP.

4.2.3 Analiza reszt

Reszty modelu (ε) zostały poddane analizie normalności, aby ocenić, czy spełniają jedno z kluczowych założeń klasycznej regresji liniowej. Normalność reszt jest istotna, ponieważ wpływa na poprawność wnioskowania statystycznego, konstrukcję przedziałów ufności oraz testów istotności współczynników.

Sformułowanie hipotez

- **Hipoteza zerowa (H_0):** Reszty są rozłożone zgodnie z rozkładem normalnym.
- **Hipoteza alternatywna (H_1):** Reszty nie mają rozkładu normalnego.

Jeśli wartość p -value jest mniejsza niż przyjęty poziom istotności ($\alpha = 0.05$), odrzucamy hipotezę zerową, co oznacza, że reszty nie są normalnie rozłożone.

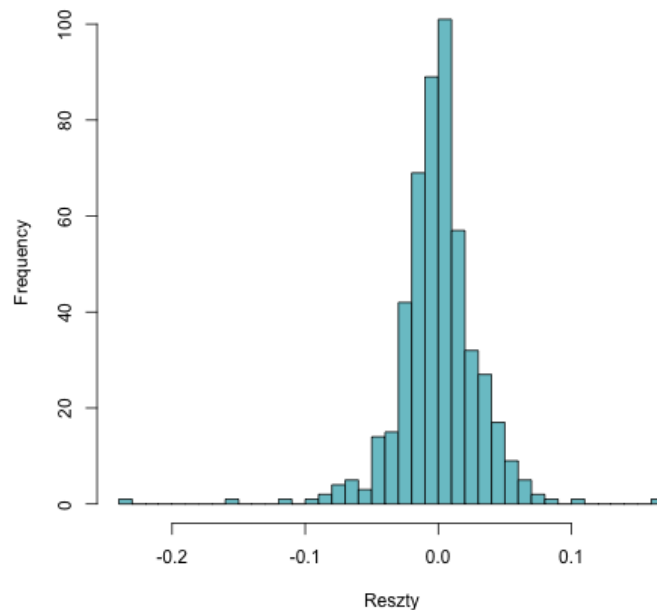
Testy normalności W celu sprawdzenia normalności reszt przeprowadzono następujące testy:

Test	p-value
Test Shapiro-Wilka	$p < 2.2 \cdot 10^{-16}$
Test Andersona-Darlinga	$p < 2.2 \cdot 10^{-16}$
Test Kołmogorowa-Smirnowa	$p = 0.0022$

Tabela 7: Wyniki testów normalności reszt.

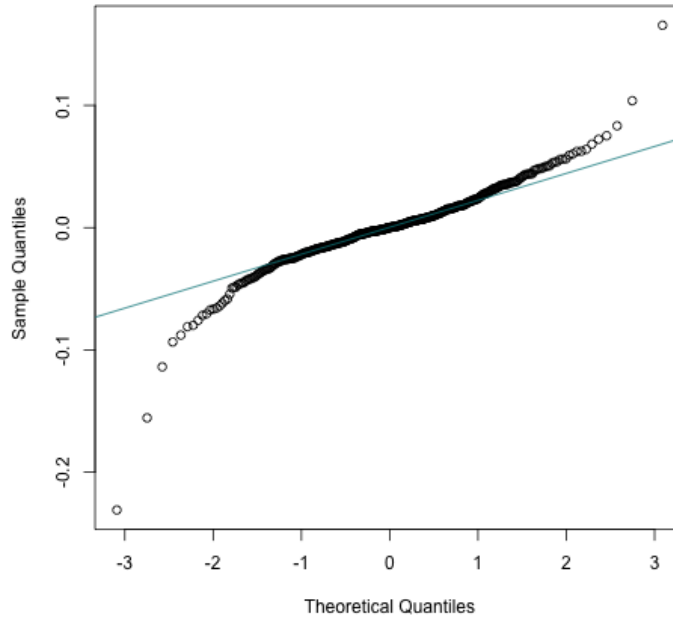
Bardzo niskie wartości p wskazują na silne odrzucenie hipotezy zerowej o normalności reszt. Oznacza to, że rozkład reszt **istotnie odbiega od normalnego**, co może wpłynąć na interpretację wyników regresji.

Analiza histogramu reszt Histogram reszt (Rysunek 18) sugeruje rozkład zbliżony do normalnego, lecz z wyraźnymi skośnościami i wartościami odstającymi. Można zauważyć lekką asymetrię oraz większą koncentrację wartości w pobliżu zera, a także kilka ekstremalnych wartości, które mogą wpływać na model.



Rysunek 18: Histogram reszt modelu.

Analiza QQ Plota QQ Plot (Rysunek 19) umożliwia graficzną ocenę normalności reszt. Wykres ten porównuje kwantyle teoretyczne rozkładu normalnego z empirycznymi kwantylami reszt.



Rysunek 19: QQ Plot reszt modelu.

Na wykresie widać, że większość punktów znajduje się wzdłuż linii teoretycznej normalności, jednak na końcach wykresu występują odchylenia, co wskazuje na wartości odstające i lekką nienormalność rozkładu.

4.2.4 Predykcja

Dla średniego log-zwrotu $R_1 = -0.0003$, przewidywany log-zwrot R_2 wynosi:

$$\hat{R}_2 = 0.000062.$$

Interpretacja wyniku Przewidywany log-zwrot R_2 dla średniego log-zwrotu R_1 jest bliski zeru ($\hat{R}_2 = 0.000062$), co sugeruje, że gdy log-zwroty spółki Allegro (R_1) oscylują wokół swojej średniej, to log-zwroty spółki LPP (R_2) pozostają niemal niezmiennione. Oznacza to, że przeciętnie ruchy cen akcji Allegro nie mają istotnego wpływu na ceny akcji LPP w ujęciu liniowym.

4.3 Regresja uproszczona

Analiza istotności współczynników modelu pełnego $R_2 = b_0 + b_1 \cdot R_1 + \varepsilon$ wykazała, że wyraz wolny (b_0) okazał się nieistotny ($p > 0.05$). Oznacza to, że na poziomie istotności $\alpha = 0.05$ nie mamy podstaw do odrzucenia hipotezy zerowej $H_0 : b_0 = 0$. W związku z tym uproszczono model, eliminując wyraz wolny:

$$R_2 = b_1 \cdot R_1 + \varepsilon.$$

Dla uproszczonego modelu współczynnik nachylenia został oszacowany jako:

$$b_1 = 0.40221.$$

Test istotności współczynnika nachylenia zweryfikował hipotezę:

$$H_0 : b_1 = 0 \quad \text{przeciwko} \quad H_1 : b_1 \neq 0.$$

Wartość p-value wyniosła $p < 2 \cdot 10^{-16}$, co oznacza, że możemy odrzucić hipotezę zerową na poziomie istotności $\alpha = 0.05$. Współczynnik nachylenia jest istotny statystycznie, co sugeruje, że istnieje zależność między log-zwrotami analizowanych spółek.

Dopasowanie modelu można ocenić na podstawie podstawowych miar jakości:

- Błąd standardowy reszt: 0.03047,
- Współczynnik determinacji R^2 : 0.147.

Wartość $R^2 = 0.147$ oznacza, że model wyjaśnia jedynie około 14.7% zmienności log-zwrotów LPP. Relatywnie niska wartość współczynnika determinacji sugeruje, że zależność liniowa między zmiennymi nie jest silna i inne czynniki mogą znacząco wpływać na zmienność log-zwrotów LPP.

Na podstawie oszacowanego modelu można również przeprowadzić predykcję. Dla średniego log-zwrotu Allegro $R_1 = -0.00033$ przewidywana wartość log-zwrotu LPP wynosi:

$$\hat{R}_2 = b_1 \cdot R_1 = 0.40221 \cdot (-0.00033) = -0.00013.$$

Oznacza to, że jeśli średni log-zwrot Allegro wynosi -0.00033 , to zgodnie z modelem, oczekiwany log-zwrot LPP będzie wynosił -0.00013 .

4.4 Porównanie przedziałów ufności dla predykcji

Dla wartości $\hat{R}_2 = -0.000062$ wyznaczono przedziały ufności zarówno metodą analityczną, jak i bootstrapową.

4.4.1 Analityczny przedział ufności

Zgodnie z teorią regresji liniowej, przedział ufności dla predykcji \hat{R}_2 na poziomie ufności $1 - \alpha$ można obliczyć jako:

$$CI_{\text{analityczny}} = \left[\hat{R}_2 - t_{1-\alpha/2, n-2} \cdot SE, \hat{R}_2 + t_{1-\alpha/2, n-2} \cdot SE \right],$$

gdzie:

- $t_{1-\alpha/2, n-2}$ – kwantyl rozkładu t-Studenta o $n - 2$ stopniach swobody,
- $SE = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(R_1 - \bar{R}_1)^2}{\sum_{i=1}^n (R_{1,i} - \bar{R}_1)^2} \right)}$ – błąd standardowy predykcji,
- $\hat{\sigma}^2$ – oszacowanie wariancji reszt.

Wyznaczony przedział ufności:

$$CI_{\text{analityczny}} = [-0.00262, 0.00274].$$

4.4.2 Przedział ufności metodą bootstrapową

Technika bootstrapowa pozwala na wyznaczenie przedziałów ufności bez konieczności spełniania założeń o normalności rozkładu reszt. Zastosowano następujący algorytm:

1. Z $n = 500$ obserwacji wygenerowano $B = 1000$ próbek bootstrapowych poprzez losowanie z powtórzeniami.
2. Dla każdej próbki oszacowano model regresji oraz wykonano predykcję dla średniego R_1 .
3. Wyznaczono kwantyle empiryczne rzędu $\alpha/2$ i $1 - \alpha/2$ z predykcji bootstrapowych, co pozwoliło wyznaczyć przedział ufności:

$$CI_{\text{bootstrap}} = [q_{\alpha/2}, q_{1-\alpha/2}].$$

Dla $\alpha = 0.05$ uzyskano przedział ufności:

$$CI_{\text{bootstrap}} = [-0.00249, 0.00277].$$

4.4.3 Porównanie przedziałów ufności

Wyniki analitycznego oraz bootstrapowego przedziału ufności są następujące:

- **Analityczny przedział ufności:** $[-0.00262, 0.00274]$
- **Bootstrapowy przedział ufności:** $[-0.00249, 0.00277]$

Obserwujemy, że oba przedziały są zbliżone, jednak przedział bootstrapowy jest nieco szerszy. Może to wynikać z większej elastyczności tej metody, która nie wymaga założenia o normalności rozkładu reszt, w przeciwieństwie do podejścia analitycznego, bazującego na rozkładzie t-Studenta.

5 Podsumowanie

W niniejszej pracy przeprowadziliśmy szczegółową analizę log-zwrotów spółek Allegro.eu S.A. oraz LPP S.A. Głównym celem było zbadanie ich statystycznych właściwości, dopasowanie modeli probabilistycznych oraz analiza zależności między log-zwrotami za pomocą regresji liniowej.

W pierwszej części dokonaliśmy wizualizacji danych oraz przeprowadziliśmy podstawową analizę statystyczną. Obliczone estymatory wartości oczekiwanej, wariancji i odchylenia standardowego pozwoliły na scharakteryzowanie zmienności log-zwrotów. Wyniki wskazały, że choć log-zwroty obu spółek wykazują podobne trendy, różnią się pod względem wariancji. Przeprowadziliśmy również analizę dobroci dopasowania rozkładów normalnego i t-Studenta, co pozwoliło stwierdzić, że model normalny lepiej opisuje dane.

Kolejnym krokiem była analiza łącznego rozkładu log-zwrotów obu spółek. Obliczone macierze kowariancji i korelacji pozwoliły na ocenę wzajemnych zależności. Współczynnik korelacji wykazał umiarkowaną dodatnią zależność między spółkami, co sugeruje, że ceny akcji obu podmiotów są w pewnym stopniu skorelowane.

Ostatnim etapem analizy była regresja liniowa, której celem było określenie relacji między log-zwrotami Allegro i LPP. Wyniki estymacji parametrów regresji wykazały

statystycznie istotną zależność między zmiennymi, choć współczynnik determinacji R^2 sugerował, że model liniowy wyjaśnia jedynie część zmienności log-zwrotów. Dodatkowo porównaliśmy przedziały ufności wyznaczone metodą analityczną i bootstrapową, co pozwoliło na ocenę stabilności oszacowań.

Podsumowując, przeprowadzona analiza dostarczyła kompleksowego obrazu zmienności log-zwrotów oraz ich wzajemnych powiązań. Otrzymane wyniki wskazują, że log-zwroty Allegro i LPP wykazują istotne podobieństwa, choć występują różnice w ich dynamice. Zastosowane metody statystyczne pozwoliły na precyzyjną ocenę właściwości rozkładu i zależności między spółkami, co może stanowić podstawę do dalszych badań oraz podejmowania decyzji inwestycyjnych.