AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Zarządzania

Projekt EFiD – część 1.

Autorzy:	*Patrycja Piła, Mateusz Strojek, Julia Szutka, Magdalena Wnuk*
Kierunek studiów:	*Informatyka i Ekonometria*
Prowadzący:	*dr hab. Tomasz Wójtowicz*

Kraków, 2025

Spis treści

Polecenie zadania

Wersja D Na podstawie danych dziennych obliczyć zwykłe i logarytmiczne tygodniowe stopy zwrotu (zamknięcie – zamknięcie) przyjmując za koniec tygodnia poszczególne dni tygodnia tzn. obliczyć stopy zwrotu poniedziałek-poniedziałek, wtorek-wtorek itd. (w przypadku braku notowań w danym dniu tygodnia wziąć pod uwagę ostatnią dostępną cenę przed tym dniem).

Obliczyć podstawowe charakterystyki wyznaczonych wcześniej zbiorów stóp zwrotu (średnia, minimum, maksimum, odchylenie standardowe, skośność, kurtoza). Podać krótką interpretację uzyskanych wartości.

Za pomocą co najmniej trzech różnych testów zbadań normalność rozkładu wyznaczonych stóp zwrotu. Zbadać występowanie i istotność autokorelacji stóp zwrotu.

Za pomocą odpowiednich testów porównać wartości oczekiwane, wariancje rozkładów uzyskanych stóp zwrotu, porównać wyniki uzyskane dla różnych rodzajów obliczonych stóp zwrotu.

Wstęp

Analiza tygodniowych stóp zwrotu (zwykłych i logarytmicznych) oraz ich charakterystyk statystycznych pozwala ocenić efektywność inwestycji, zmienność i ryzyko na rynku, a także wykryć ewentualne wzorce, takie jak efekt dnia tygodnia.

Celem zadania było obliczenie tygodniowych stóp zwrotu oraz przeanalizowanie ich charakterystyk. W tym celu został wybrany indeks S&P 500. Jest to indeks giełdowy ważony kapitalizacją, w skład którego wchodzi 500 przedsiębiorstw o największej kapitalizacji, notowanych między innymi na New York Stock Exchange.

Analiza obejmuje dni robocze (poniedziałek - piątek) ze względu na brak notowań w weekendy. Występujące braki danych zostały zastąpione ostatnio zanotowaną wartością.

Statystyki opisowe

Aby policzyć statystyki opisowe, najpierw wyznaczono tygodniowe stopy zwrotu.

Zwykła stopa zwrotu jest wyrażana wzorem:

$$R_t = \frac{(P_t - P_{t-7})}{P_{t-7}}$$

Gdzie R_t to stopa zwrotu, a P_t to wartość indeksu w czasie zamknięcia w danym dniu.

Stopę logarytmiczną oblicza się następująco:

$$R_t = ln(\frac{P_t}{P_{t-7}})$$

Gdzie R_t to stopa zwrotu, a P_t to wartość indeksu w czasie zamknięcia w danym dniu.

Finalne wyniki statystyk opisowych kształtują się następująco:

Tabela 1. Statystyki opisowe prostej stopy zwrotu

Dzień tygodnia	Średnia	Mediana	Min	Max	Odch. std	Wariancja	Skośność	Eksces
Poniedziałek	0,0023	0,0039	-0,1312	0,1740	0,0239	0,0006	-0,2389	8,3580
Wtorek	0,0023	0,0038	-0,1225	0,0702	0,0209	0,0004	-1,1520	5,4565
Środa	0,0023	0,0044	-0,1252	0,1131	0,0220	0,0005	-1,1007	5,8746
Czwartek	0,0022	0,0035	-0,1797	0,1040	0,0225	0,0005	-1,2602	10,7921
Piątek	0,0023	0,0037	-0,1498	0,1210	0,0232	0,0005	-0,6130	6,5406

Średnie tygodniowe stopy zwrotu dla wszystkich dni tygodnia są bardzo zbliżone (ok. 0,23%), co wskazuje na brak wyraźnego efektu dnia tygodnia w poziomie zwrotów. Największe zróżnicowanie wartości minimalnych i maksymalnych występuje w czwartki. Odchylenia standardowe są z zakresu 0,021 do 0,024. Ujemne wartości skośności świadczą o lekkiej asymetrii rozkładów w stronę strat, natomiast wysokie wartości kurtozy (szczególnie w czwartek - ponad 10) wskazują na występowanie tzw. „grubych ogonów". W obliczeniach użyto funkcji, która zwraca kurtozę nadwyżkową (eksces kurtozy), odejmując wartość 3 (kurtozę rozkładu normalnego) od kurtozy właściwej. W związku z tym, rozkład normalny charakteryzuje się kurtozą nadwyżkową równą 0. Rozkład zwrotów odbiega od normalności (skośność poniżej 0 i eksces powyżej 0) i wykazuje typowe cechy danych finansowych, czyli między innymi sporadyczne ekstremalne obserwacje.

Tabela 2. Statystyki opisowe logarytmicznej stopy zwrotu

Dzień tygodnia	Średnia	Mediana	Min	Max	Odch. std	Wariancja	Skośność	Eksces
Poniedziałek	0,0021	0,0039	-0,1407	0,1604	0,0240	0,0006	-0,5922	7,8865
Wtorek	0,0021	0,0038	-0,1307	0,0679	0,0211	0,0004	-1,3545	6,4280
Środa	0,0020	0,0044	-0,1338	0,1072	0,0222	0,0005	-1,3274	6,6615
Czwartek	0,0020	0,0035	-0,1980	0,0990	0,0228	0,0005	-1,6779	13,8649
Piątek	0,0020	0,0036	-0,1622	0,1142	0,0233	0,0005	-0,9094	7,5431

Analiza statystyk opisowych dla prostych i logarytmicznych stóp zwrotu ujawnia zbieżność miar centralnych, ale różnice w kształcie rozkładu. Potwierdza się, że średnia i mediana obu stóp są bardzo zbliżone, podobnie jak odchylenie standardowe, co sugeruje, że ogólny poziom ryzyka i tendencja centralna są porównywalne w obu szeregach. Kluczowe rozbieżności pojawiają się jednak w ekstremach: wartości minimalne są niższe dla stóp logarytmicznych, podczas gdy maksymalne są wyższe dla prostych, co wynika wprost z nieliniowej transformacji logarytmicznej. Najważniejsza różnica dotyczy skośności: logarytmiczne stopy zwrotu charakteryzują się wyższą ujemną skośnością (lewoskośnością), co oznacza, że rozkład ma grubszy lewy ogon, wskazując na większą częstotliwość występowania dużych, ujemnych zwrotów. Dodatkowo, stopy logarytmiczne wykazują wyższy eksces (kurtozę), co oznacza, że dane są bardziej skumulowane wokół średniej (leptokurtoza), co jest szczególnie widoczne w analizie dziennej (np. dla czwartków). Zarówno wysoka skośność, jak i eksces są silnym dowodem na to, że rozkłady obu stóp, zwłaszcza logarytmicznych, istotnie odbiegają od rozkładu normalnego.

Testy normalności

W celu sprawdzenia, czy rozkład stóp zwrotu można uznać za normalny, zastosowano trzy testy: Shapiro-Wilka, Jarque-Bera oraz Andersona-Darlinga. We wszystkich przypadkach hipotezy były następujące:

- H_0: rozkład badanej cechy jest rozkładem normalnym,
- H_1: rozkład badanej cechy nie jest rozkładem normalnym.

Test Shapiro-Wilka opiera się na dopasowaniu obserwacji do wartości oczekiwanych w rozkładzie normalnym i jest szczególnie czuły przy mniejszych próbach. Test Jarque-Bera bada zgodność skośności i kurtozy z wartościami odpowiadającymi rozkładowi normalnemu, dlatego dobrze wychwytuje asymetrię i grube ogony charakterystyczne dla danych finansowych. Test Andersona-Darlinga koncentruje się na dopasowaniu całego rozkładu, przy czym szczególną wagę przykłada do ogonów, czyli obszarów, gdzie w praktyce najczęściej pojawiają się odchylenia od normalności.

Tabela 3. Wyniki testów normalności rozkładu zwykłych tygodniowych stóp zwrotu

Dzień	Statystyka Shapiro-Wilka	Wartość p Shapiro-Wilka	Statystyka Jarque-Bera	Wartość p Jarque-Bera	Statystyka Anderson - Darlinga	Wartość krytyczna (5%) Andersona - Darlinga
Poniedziałek	0,9024	6,9878E-18	1544,7849	0,0000E+00	10,4456	0,781
Wtorek	0,9210	4,9793E-16	773,2416	1,2380E-168	9,2594	0,781
Środa	0,9138	8,8447E-17	867,4840	4,2487E-189	9,4214	0,781
Czwartek	0,9008	4,8865E-18	2707,1794	0,0000E+00	7,5604	0,781
Piątek	0,9262	1,7879E-15	977,9189	4,4424E-213	6,6815	0,781

Test Shapiro-Wilka wykazał bardzo niskie wartości p, co jednoznacznie prowadzi do odrzucenia hipotezy o normalności rozkładu. Test Jarque'a-Bery zwrócił wyjątkowo wysokie wartości statystyki, wskazujące na silne odchylenia od normalności, wynikające głównie z nadmiernej kurtozy. Z kolei test Andersona-Darlinga dał wartości statystyki znacząco przekraczające wartość krytyczną, co potwierdza, że rozkład stóp zwrotu różni się od normalnego, szczególnie w ogonach rozkładu.

Wszystkie testy prowadzą do wspólnego wniosku, że zwykłe tygodniowe stopy zwrotu nie mają rozkładu normalnego. Oznacza to, że w danych występują grube ogony i możliwe asymetrie, co sprawia, że skrajne zmiany cen pojawiają się częściej niż w modelu opartym na rozkładzie normalnym.

Tabela 4. Wyniki testów normalności rozkładu logarytmicznych tygodniowych stóp zwrotu

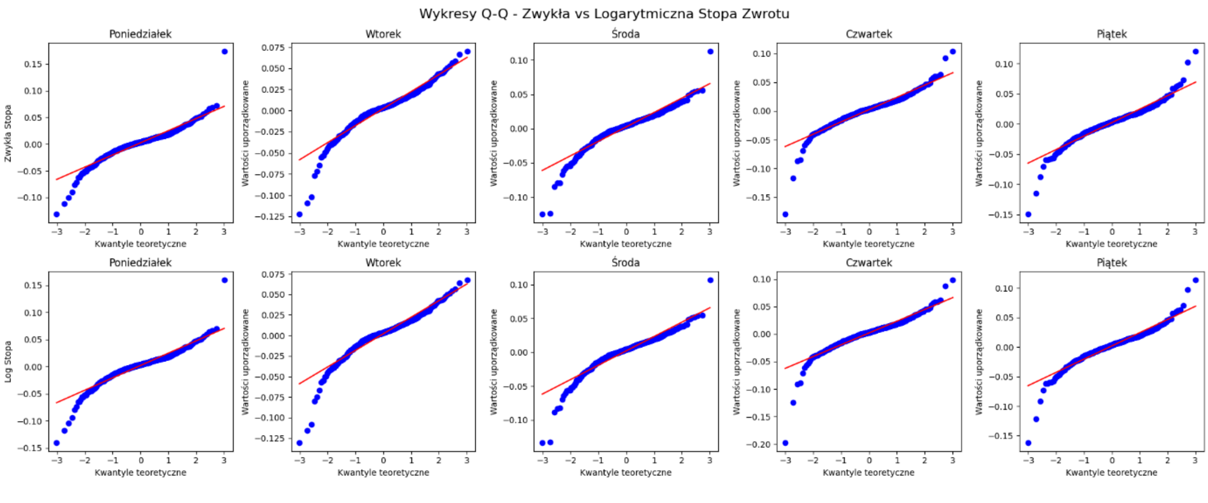
Dzień	Statystyka Shapiro-Wilka	Wartość p Shapiro-Wilka	Statystyka Jarque-Bera	Wartość p Jarque-Bera	Statystyka Andersona - Darlinga	Wartość krytyczna (5%) Andersona - Darlinga
Poniedziałek	0,8995	3,7107E-18	1401,8329	3,9432E-305	11,0678	0,781
Wtorek	0,9090	2,9476E-17	1072,5021	1,2856E-233	10,1993	0,781
Środa	0,9024	6,9627E-18	1133,4557	7,4683E-247	10,5265	0,781
Czwartek	0,8827	1,3690E-19	4485,4465	0,0000E+00	8,5851	0,781
Piątek	0,9174	2,0017E-16	1329,5708	1,9380E-289	7,3397	0,781

Test Shapiro-Wilka dla logarytmicznych stóp zwrotu wykazał bardzo niskie wartości p, co jednoznacznie prowadzi do odrzucenia hipotezy o normalności rozkładu. Test Jarque'a-Bery zwrócił ekstremalnie wysokie wartości statystyki oraz zerowe wartości p, co dowodzi silnych odchyleń od normalności wynikających głównie z leptokurtozy. Z kolei test Andersona-Darlinga dał wartości statystyki w zakresie 7,34–11,07, znacznie przekraczające wartość krytyczną 0,781, co potwierdza brak dopasowania do rozkładu normalnego, zwłaszcza w ogonach.

Wszystkie testy zgodnie wskazują, że logarytmiczne tygodniowe stopy zwrotu nie spełniają założenia normalności.

Podsumowując, wyniki testów Shapiro-Wilka, Jarque'a-Bery oraz Andersona-Darlinga, zarówno dla zwykłych, jak i logarytmicznych tygodniowych stóp zwrotu, jednoznacznie prowadzą do odrzucenia hipotezy zerowej o normalności rozkładu.

Rysunek 1. Porównanie rozkładów zwykłych i logarytmicznych tygodniowych stóp zwrotu na wykresach Q–Q



Wykresy Q-Q pokazują, że rozkład tygodniowych stóp zwrotu tylko częściowo przypomina normalny – w centrum dane układają się blisko prostej, natomiast w ogonach widać istotne odchylenia. Potwierdza to wyniki testów, że rozkład charakteryzuje się grubymi ogonami i odstępstwami od normalności.

Autokorelacja

Tabela 5. Wyniki testu Ljunga–Boxa dla tygodniowych stóp zwrotu

Opóź-nienie	Poniedziałek		Wtorek		Środa		Czwartek		Piątek	
	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p
1	0,1600	0,6892	0,2299	0,6316	0,0177	0,8943	7,8501	0,0051	4,9560	0,0260
2	0,4096	0,8148	0,6874	0,7091	0,0437	0,9784	11,0539	0,0040	5,3771	0,0680
3	6,9591	0,0732	3,5673	0,3121	0,3146	0,9573	11,1055	0,0112	5,8088	0,1213
4	6,9626	0,1379	5,1257	0,2746	2,6716	0,6142	16,4955	0,0024	6,2113	0,1839
5	10,2634	0,0681	6,7550	0,2395	3,6201	0,6053	16,5561	0,0054	9,9618	0,0763
6	10,3054	0,1124	9,2369	0,1607	6,1832	0,4030	20,6767	0,0021	10,2072	0,1162
7	10,6043	0,1568	9,6434	0,2097	6,1980	0,5168	20,9095	0,0039	10,6835	0,1530
8	11,1998	0,1906	9,6756	0,2885	6,2599	0,6181	22,3162	0,0044	10,6972	0,2195
9	11,2410	0,2595	11,0983	0,2690	8,2698	0,5072	22,3681	0,0078	11,8082	0,2243
10	11,3014	0,3345	11,2330	0,3397	8,5866	0,5717	22,3826	0,0133	11,9464	0,2887

Test Ljunga-Boxa sprawdza hipotezę zerową o braku autokorelacji w szeregu czasowym do danego opóźnienia. W analizie tygodniowych stóp zwrotu otrzymano następujące wyniki:

- Poniedziałek, wtorek i środa - wartości p dla wszystkich opóźnień były wyraźnie większe niż 0,05, co oznacza brak podstaw do odrzucenia hipotezy zerowej. Stopy zwrotu w te dni nie wykazują istotnej autokorelacji.
- Czwartek - dla większości opóźnień wartości p były mniejsze od 0,01, co wskazuje na istotną autokorelację.
- Piątek - przy pierwszym opóźnieniu uzyskano istotny wynik ($p \approx 0,0260$), jednak na dalszych opóźnieniach autokorelacja zanika. Można więc mówić o słabej, krótkoterminowej zależności.

Podsumowując, wyraźna autokorelacja występuje jedynie w czwartkowych stopach zwrotu, natomiast w piątkowych jest ograniczona i krótkotrwała.

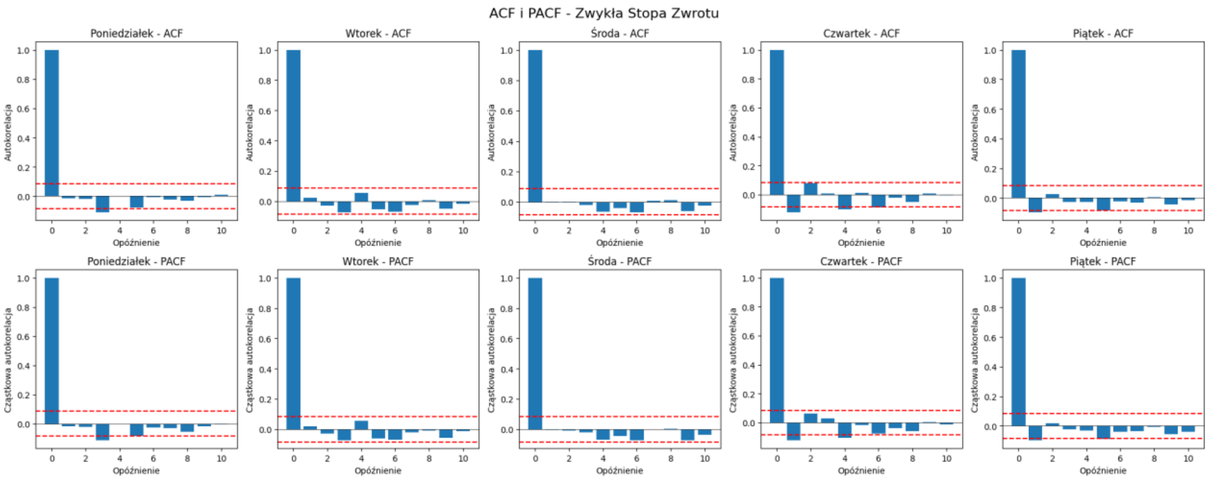
Tabela 7. Wyniki testu Ljunga–Boxa dla logarytmicznych tygodniowych stóp zwrotu

Opóź-nienie	Poniedziałek		Wtorek		Środa		Czwartek		Piątek	
	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p	Statystyka	Wartość p
1	0,0049	0,9441	0,3083	0,5787	0,0019	0,9655	6,8061	0,0091	4,1058	0,0427
2	0,1371	0,9337	0,7079	0,7019	0,0044	0,9978	10,4203	0,0055	4,5922	0,1007
3	6,2576	0,0997	3,0735	0,3804	0,2322	0,9722	10,4898	0,0148	4,8161	0,1858
4	6,2578	0,1807	4,5352	0,3384	2,5202	0,6410	15,8970	0,0032	5,2217	0,2653
5	9,3530	0,0958	5,9219	0,3139	3,3608	0,6446	15,9768	0,0069	8,7063	0,1214
6	9,4168	0,1515	8,4246	0,2086	6,0467	0,4180	19,7279	0,0031	8,9943	0,1739
7	9,7625	0,2024	8,7567	0,2706	6,0722	0,5313	20,0124	0,0055	9,4673	0,2208
8	10,3345	0,2423	8,7788	0,3613	6,1208	0,6337	21,3804	0,0062	9,4846	0,3031
9	10,3956	0,3194	10,2193	0,3330	8,2019	0,5139	21,4136	0,0109	10,7164	0,2956
10	10,4189	0,4045	10,3513	0,4102	8,5091	0,5792	21,4382	0,0182	10,9256	0,3633

- Poniedziałek, wtorek i środa - wartości p dla prawie wszystkich opóźnień były wyraźnie większe niż 0,05, co oznacza brak podstaw do odrzucenia hipotezy zerowej. Logarytmiczne stopy zwrotu w te dni nie wykazują istotnej autokorelacji.
- Czwartek - w tym przypadku dla większości opóźnień wartości p były mniejsze od 0,01, co jednoznacznie wskazuje na istotną autokorelację w szeregu logarytmicznych stóp zwrotu.
- Piątek - przy pierwszym opóźnieniu uzyskano istotny wynik ($p \approx 0{,}0427$), jednak na dalszych lagach autokorelacja zanika. Można więc mówić o słabej, krótkoterminowej zależności.

Wyniki te są zbliżone do uzyskanych dla zwykłych stóp zwrotu - jedynie w czwartkowych danych występuje silniejsza zależność czasowa.

Rysunek 2. Wykresy funkcji autokorelacji (ACF) i cząstkowej autokorelacji (PACF) dla zwykłych tygodniowych stóp zwrotu

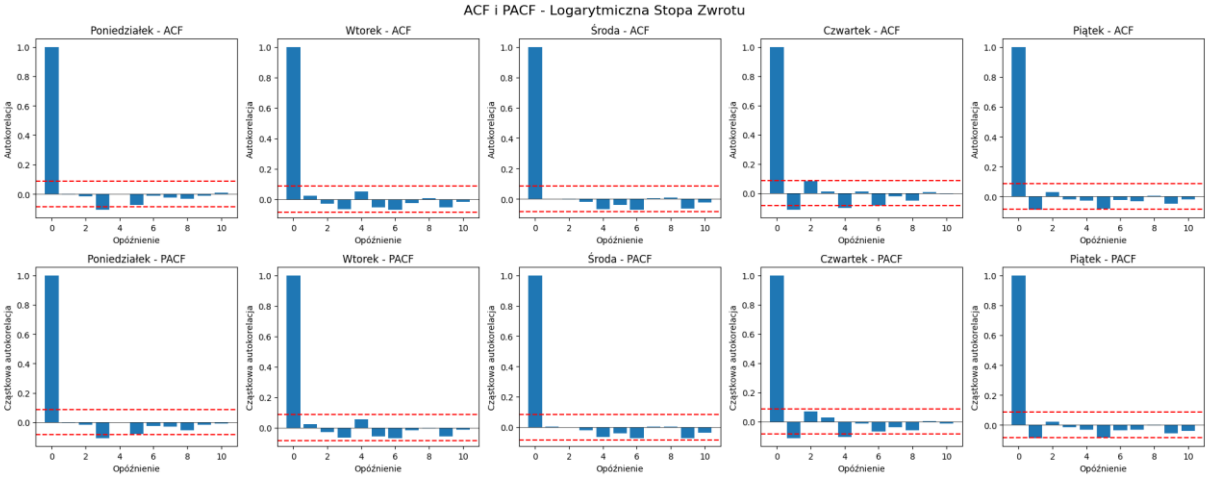


Wykresy ACF i PACF pokazują, że tygodniowe stopy zwrotu dla poniedziałku, wtorku i środy nie wykazują istotnej autokorelacji, co oznacza, że kolejne obserwacje są od siebie niezależne.

W przypadku czwartku widoczne są wyraźniejsze zależności między obserwacjami przy pierwszych kilku opóźnieniach, co wskazuje na obecność autokorelacji. Dla piątku zależność ta jest jedynie delikatna i krótkookresowa, ograniczona do najniższych opóźnień.

Wykresy potwierdzają, że istotna autokorelacja występuje jedynie w czwartkowych stopach zwrotu.

Rysunek 3. Wykresy funkcji autokorelacji (ACF) i cząstkowej autokorelacji (PACF) dla logarytmicznych tygodniowych stóp zwrotu



Wykresy ACF i PACF wskazują, że dla poniedziałku, wtorku i środy logarytmiczne stopy zwrotu nie wykazują istotnej autokorelacji, ponieważ wartości autokorelacji mieszczą się w granicach istotności. Oznacza to brak zależności pomiędzy kolejnymi obserwacjami w tych dniach.

Dla czwartku można zauważyć wyraźniejsze zależności przy pierwszych opóźnieniach, co potwierdza obecność autokorelacji w tym dniu. W przypadku piątku zależność jest słaba i krótkookresowa, ograniczona do najniższych opóźnień.

Wnioski z wykresów są spójne z wynikami testu Ljunga-Boxa - istotna autokorelacja występuje jedynie w czwartkowych logarytmicznych stopach zwrotu, natomiast w pozostałe dni szeregi można uznać za procesy losowe bez istotnej pamięci.

Porównanie wartości oczekiwanych

W następnym kroku przeprowadzono testy w celu porównania wartości oczekiwanych zwykłych, jak i logarytmicznych stóp zwrotu pomiędzy badanymi grupami. Zastosowano test jednoczynnikowej analizy wariancji ANOVA oraz test Kruskala-Wallisa, który stanowi jego odpowiednik w sytuacji, gdy założenie o normalności rozkładu może być naruszone.

W teście ANOVA uzyskano wartość statystyki F = 0,0015 oraz wartość p = 0,9999. Otrzymany wynik jest znacznie powyżej przyjętego poziomu istotności, co oznacza brak podstaw do odrzucenia hipotezy zerowej mówiącej o równości średnich stóp zwrotu we wszystkich grupach.

Analogicznie, w teście Kruskala-Wallisa wartość statystyki wyniosła H = 0,6659, a odpowiadająca jej wartość p przyjęła 0,9555. Wynik ten również wskazuje na brak istotnych różnic pomiędzy porównywanymi grupami.

W przypadku logarytmicznych stóp zwrotu po przeprowadzeniu testu ANOVA uzyskano wartość statystyki F = 0,0013 oraz wartość p = 0,9999. Wynik ten podobny jest do wartości otrzymanych dla zwykłych stóp zwrotu. Ponownie nie odrzucono hipotezy zerowej.

W teście Kruskala-Wallisa wartość statystyki wyniosła H = 0,6659, a odpowiadająca jej wartość p = 0,9555. Otrzymane wyniki są bardzo zbliżone do tych uzyskanych dla zwykłych stóp zwrotu i prowadzą do identycznych wniosków.

Ponieważ testy normalności jednoznacznie odrzuciły normalność rozkładów, wyniki testu Kruskala-Wallisa są bardziej wiarygodne niż wynik testu ANOVA.

Zarówno dla zwykłych, jak i logarytmicznych stóp zwrotu, wyniki testów parametrycznych i nieparametrycznych jednoznacznie wskazują na brak statystycznie istotnych różnic w wartościach oczekiwanych pomiędzy analizowanymi grupami.

Porównanie wariancji rozkładów

Przeprowadzono analizę porównawczą wariancji w dwóch zbiorach danych z wykorzystaniem dwóch testów statystycznych.

Test Levene'a jest nieparametrycznym testem służącym do weryfikacji jednorodności wariancji w dwóch lub więcej grupach. Jest on odporny na odstępstwa od rozkładu normalnego. Test Bartletta również służy do weryfikacji jednorodności wariancji, jednak jest czuły na odstępstwa od rozkładu normalnego. Hipotezy są takie same w obu testach:

- H_0: Wariancje w badanych grupach są równe
- H_1: Wariancje w badanych grupach różnią się

Tabela 9. Porównanie wariancji zwykłych stóp zwrotu

Test	Statystyka testowa	Wartość p
Levene	1,1392	0,3361
Bartlett	11,1338	0,0251

Tabela 10. Porównanie wariancji logarytmicznych stóp zwrotu

Test	Statystyka testowa	Wartość p
Levene	1,0747	0,336113
Bartlett	9,8356	0,043291

Przeprowadzona wykazała rozbieżność wyników w obu zbiorach danych zarówno dla zwykłej stopy, jak i stopy logarytmicznej. Ponieważ test Bartletta jest skrajnie wrażliwy na naruszenie założenia o normalności, a testy normalności wcześniej wykazały jednoznacznie, że stopy zwrotu nie mają rozkładu normalnego, wyniki testu Bartletta (odrzucenie) są niewiarygodne i należy je zignorować. Decydujący jest wynik testu Levene'a, który jest odporny na brak normalności. Jego wynik nie daje podstaw do odrzucenia hipotezy o jednorodności wariancji (homoskedastyczności) dla poszczególnych dni tygodnia.

Wnioski

Transformacja logarytmiczna stóp zwrotu nie wprowadza istotnych zmian w kluczowych wnioskach ekonomicznych i ekonometrycznych. Zarówno stopy zwykłe, jak i logarytmiczne, charakteryzują się zbliżonymi wartościami średniej i odchylenia standardowego. Kluczowe jest, że obie formy szeregów jednoznacznie odrzuciły hipotezę o normalności rozkładu, wykazując typową dla rynków finansowych leptokurtyczność (grube ogony) oraz ujemną skośność, przy czym logarytmiczne stopy zwrotu charakteryzowały się nieco wyższą kurtozą nadwyżkową. Ponadto, w obu przypadkach istotna autokorelacja została wykryta głównie w szeregu czwartkowym, co dowodzi, że zaobserwowana słaba zależność czasowa wynika z samych danych, a nie z przeprowadzonych logarytmicznych przekształceń.