

Projekt 2

Ekonometria finansowa i dynamiczna

Informatyka i Ekonometria rok V semestr I

Grzegorz Bylina

Kamila Kucharska

Andrzej Miczek

Spis treści

Wprowadzenie	3
Stopy zwrotu	3
Badanie stacjonarności	5
Budowa modelu i weryfikacja jego własności	5
Prognozowanie na podstawie modelu	6
Prognozowanie metodą Monte Carlo.....	9
Prognozowanie metodą bootstrap	10
Budowa modelu ARMA-GARCH i ARIMA-GARCH i powtórzenie badania	11
Podsumowanie	15

Wprowadzenie

W celu wykonania projektu zostały pobrane zbiory danych z notowań giełdowych dziennych ze strony *stoq.pl*. Spółka MOL, której dotyczy zbiór została wybrana z rankingu dotyczącego notowań spółek notowanych na Warszawskim Indeksie Giełdowym – WIG z zakresu miejsc od 60 do 70. MOL jest to węgierskie przedsiębiorstwo przetwórstwa ropy naftowej i gazu ziemnego, zajmujące się również dystrybucją tych produktów. Okres, dla którego zostały pobrane dane to wrzesień 2023 do 4 października 2024. W projekcie przeprowadzono prognozowanie 4 kolejnych notowań dla okresu od 1 października 2024 na podstawie danych z okresu od 1.09.2023 do 30.09.2024 roku. Otrzymane prognozy zostały porównane z rzeczywistymi wartościami notowań zarejestrowanymi w pierwszych czterech dniach października. Ważnym aspektem w kontekście analizy i predykcji jest wstępowanie dywidendy, które może wpływać negatywnie na jakość predykcji.

Stopy zwrotu

Na podstawie danych obliczono zwykłe i logarytmiczne stopy zwrotu:

$$R_t = \ln(C_t / C_{t-1})$$

R_t – logarytmiczna stopa zwrotu w dniu t

C_t – cena zamknięcia w dniu t

C_{t-1} – cena zamknięcia z poprzedniego dnia $t-1$

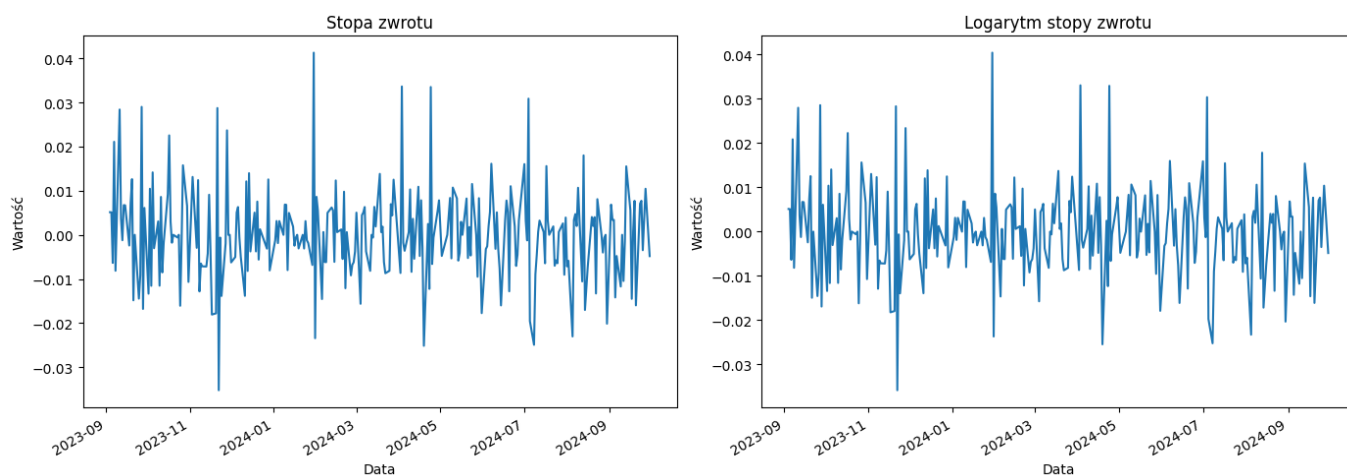
$$K_t = (C_t - C_{t-1}) / C_{t-1}$$

K_t – zwykła stopa zwrotu w dniu t

C_t – cena zamknięcia w dniu t

C_{t-1} – cena zamknięcia z poprzedniego dnia $t-1$

Wykres 1. Wartości zwykłych i logarytmicznych stóp zwrotu.



Na wykresach przedstawione zostały szeregi czasowe utworzone z przekształconych danych. Zarówno stopy zwykłe jak i logarytmiczne wyglądają podobnie.

Wykres 2. Świecowa reprezentacja cen akcji spółki MOL.



Cena zamknięć spółki MOL kształtowała się na podobnym poziomie w trakcie badanego okresu. W lipcu wystąpił nagły wzrost, a następnie gwałtowny spadek ceny.

Badanie stacjonarności

W celu badania stacjonarności stóp zwrotu i logarytmów cen wykorzystane Test Dickeya-Fullera (ADF) o hipotezach:

H_0 : Szereg nie jest stacjonarny

$H_1: \sim H_0$

Przy badaniu szeregu logarytmicznych stóp zwrotu otrzymaliśmy p -value w przybliżeniu mniejszy niż przyjęty poziom istotności ($\alpha = 0.05$), dlatego odrzucono hipotezę zerową co oznacza, że badany szereg logarytmicznych stóp zwrotu jest stacjonarny.

W przypadku logarytmu cen wartość p -value wyniosła w przybliżeniu 0.15 co stanowi wartość większą niż poziom istotności ($\alpha = 0.05$), dlatego nie ma podstaw do odrzucenia hipotezy zerowej. Szereg logarytmów cen jest niestacjonarny.

Budowa modelu i weryfikacja jego własności

Skuteczne modelowanie i prognozowanie szeregów czasowych opiera się na wykorzystaniu modeli ARMA dla szeregów stacjonarnych oraz modeli ARIMA dla szeregów niestacjonarnych. W przypadku modeli ARMA, istotne są dwa parametry:

p - odpowiada za rząd autoregresji,

q - określa rząd średniej ruchomej.

Dla modeli ARIMA(p,d,q) kluczowe jest ustalenie stopnia integracji (d), co wiąże się z oceną stacjonarności szeregu czasowego. Jeśli szereg jest stacjonarny, parametr d przyjmuje wartość 0.

W wybranym zbiorze danych szereg logarytmicznych stóp zwrotu jest stacjonarny, dlatego stworzono model ARMA. Za pomocą funkcji *auto_arima* najlepsze dopasowanie modelu na podstawie kryterium informacyjnego Akaika otrzymano dla ARIMA(1,0,0), czyli ARMA(1,0)

W obu przypadkach funkcja *summary* pozwoliła od razu stwierdzić **istotność wszystkich parametrów modelu** ($P > |z| = 0.00$). W ten sam sposób zostały przebadane składniki losowe w kontekście autokorelacji oraz reszty modeli jako potencjalny biały szum jak również normalność składników losowych. Wyniki zostały przedstawione w poniższej tabeli:

	Logarytmiczne stopy zwrotu		Logarytmy cen	
	p-value	interpretacja	p-value	interpretacja
Test Ljung-Box'a	0.96	Brak autokorelacji	0.96	Brak autokorelacji
Test na heteroskedastyczność	0.49	Brak heteroskedastyczności	0.50	Brak heteroskedastyczności
Test Jarque-Bera	0.00	Brak normalności rozkładu reszt	0.00	Brak normalności rozkładu reszt

Test Ljunga-Boxa w obu przypadkach wartości są znacznie większe niż poziom istotności (0.05), co oznacza, że nie mamy podstaw do odrzucenia hipotezy zerowej. W związku z tym możemy uznać, że reszty zarówno dla logarytmów stóp zwrotu, jak i logarytmów cen nie występuje autokorelacja, na tej podstawie zweryfikowano, że **reszty modelu są białym szumem**.

Test na heteroskedastyczność jest przeprowadzany, aby sprawdzić, czy wariancja reszt modelu pozostaje stała w czasie (homoskedastyczność) czy też się zmienia (heteroskedastyczność).

W analizie szeregów czasowych stała wariancja reszt jest pożądana, ponieważ wskazuje na stabilność modelu i brak nieregularności w jego błędach.

Test heteroskedastyczności został zastosowany zarówno dla logarytmicznych stóp zwrotu, jak i logarytmów cen. Wysokie wartości p-value w obu przypadkach (0,49 oraz 0,50) wskazują, że nie ma podstaw do odrzucenia hipotezy zerowej, co sugeruje, że wariancja reszt pozostaje stała w czasie. Oznacza to, że model nie wykazuje cech heteroskedastyczności, a zmienność reszt nie zmienia się w znaczący sposób w analizowanym okresie.

Test Jarque'a-Bery'ego sprawdza, czy rozkład danych różni się od rozkładu normalnego. Hipoteza zerowa zakłada normalność rozkładu. W obu przypadkach wartości są równe 0.00, co jest mniejsze od poziomu istotności. Oznacza to, że odrzucamy hipotezę zerową o normalności rozkładu. Wynika z tego, że reszty obu modeli **nie mają rozkładu normalnego**.

Prognozowanie na podstawie modelu

Za pomocą modelu ARMA dokonano prognozy dla 4 kolejnych notowań dla logarytmicznych stóp zwrotu, logarytmów cen oraz cen. Dodatkowo wyznaczono 95% przedział ufności dla prognozowanej wartości.

Tabela 1.1: Wartości prognoz logarytmów cen

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	3.365094	3.345721	3.384467	3.365225
02.paź	3.364769	3.340659	3.388880	3.352007
03.paź	3.364854	3.336107	3.393600	3.347797
04.paź	3.364832	3.332260	3.397404	3.354804

Prognozy wartości dla poszczególnych dni wykazują wysoką zgodność z rzeczywistymi wartościami, a przedziały ufności wskazują na niewielką niepewność w prognozach, co sugeruje, że model jest skuteczny w przewidywaniu wartości.

Tabela 1.2: Wartości prognoz dla logarytmicznych stóp zwrotu

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	0.001283	-0.018071	0.020637	0.001383
02.paź	-0.000295	-0.020283	0.019693	-0.013218
03.paź	0.000112	-0.019917	0.020142	-0.004211
04.paź	0.000007	-0.020025	0.020039	0.007008

Wartości prognoz dla poszczególnych dni wskazują, że przewidywane stopy zwrotu mieszczą się w ustalonych przedziałach ufności, co sugeruje, że prognozy są otoczone niepewnością, ale pozostają w rozsądnych granicach. Rzeczywiste wartości stóp zwrotu pokazują, że w większości

przypadków prognozy są bliskie rzeczywistości, chociaż niektóre z nich, jak na przykład 02 października, znacząco odbiegają od prognozowanych wartości.

Prognozy cen zostały wyznaczone na podstawie modelu ARIMA przy użyciu prognoz logarytmicznych. Ceny mają warunkowy rozkład log-normalny. Jego wartość obliczamy za pomocą wzoru:

$$e^{\mu + \sigma^2 / 2}$$

Na początku zdefiniowano poziom istotności $\alpha = 0.05$. Na podstawie prognozowanych logarytmicznych cen oraz dolnych granic przedziałów ufności obliczono odchylenia standardowe. Następnie otrzymane wartości podstawiono do powyższego wzoru w celu przekształcenia prognozowanych wartości logarytmicznych cen na prognozy zwykłych cen. Dzięki temu otrzymaliśmy tabele z normalnymi wartościami prognozowanej ceny, dolnej i górnej granicy przedziału ufności.

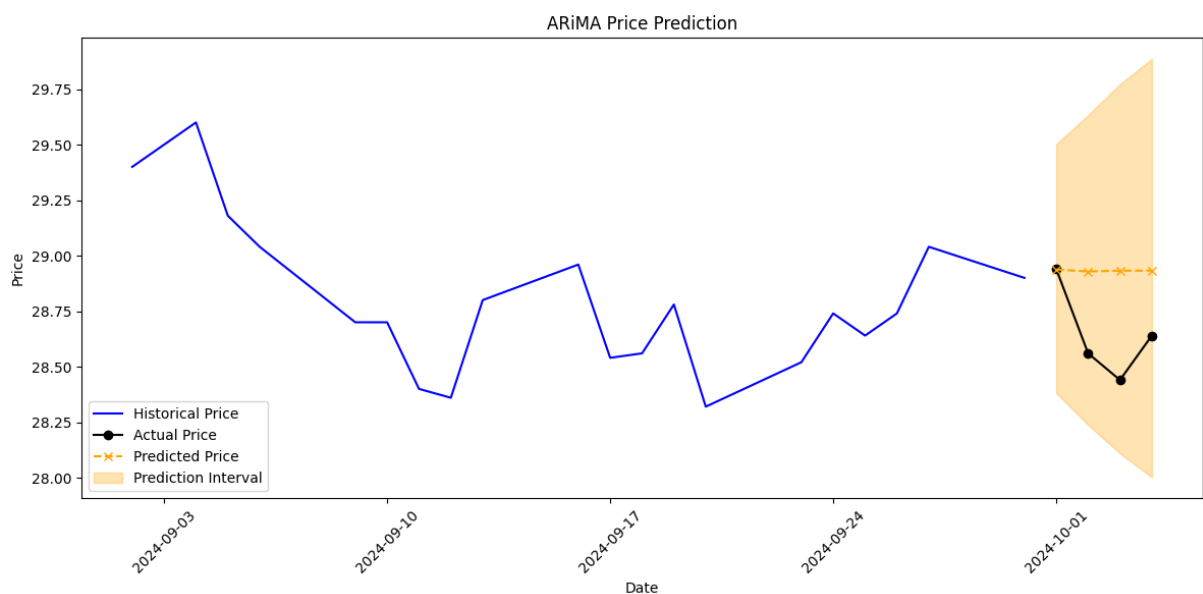
Tabela 1.3: Wartości prognoz dla cen

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	28.93762952502089	28.3810	29.5023	28.94
02.paź	28.929015040447545	28.2377	29.6327	28.56
03.paź	28.932371004645923	28.1095	29.7729	28.44
04.paź	28.932623834897342	28.0015	29.8864	28.64

Prognozy wartości dla podanych dni wskazują na stabilne przewidywania, z wartościami prognozowanymi bliskimi rzeczywistym oraz szerokimi przedziałami ufności, co sugeruje umiarkowany poziom niepewności. W każdym przypadku prognozy mieszczą się w ustalonych granicach, co potwierdza skuteczność modelu w odzwierciedleniu rzeczywistości.

Analizując otrzymane wyniki można stwierdzić, że wartości prognozowane są zbliżone do wartości rzeczywistych, ponadto wartości rzeczywiste mieszczą się w przedziale ufności co świadczy o poprawności modelu.

Wykres 3. Zestawienie predykcji cen, a faktycznych wartości dla modelu ARiMA.



Rzeczywiste ceny znalazły się w prognozowanym przedziale ufności. Natomiast prognozowane ceny różnią się od rzeczywistych.

Tabela 1.4: Błędy prognozy ex-post dla cen

ME	MSE	RMSE
0.289095	0.116059	0.340674

Interpretacja wyników:

ME: Wartość średniego błędu wynosi 0,289. Ten wskaźnik pokazuje średnią różnicę między prognozowanymi a rzeczywistymi wartościami cen. Pozytywna wartość ME sugeruje, że prognozy mają tendencję do zaniżania rzeczywistych wartości. To sugeruje, że model często przewiduje ceny za niskie.

MSE: Wartość średniego błędu kwadratowego to 0,116. MSE mierzy przeciętną wielkość błędów prognozy podniesionych do kwadratu, co powoduje, że większą wagę zyskują większe odchylenia. Niższa wartość MSE oznacza, że prognozy są stosunkowo dokładne, ale w tym przypadku błędy są na umiarkowanym poziomie.

RMSE: RMSE to 0,341, co oznacza, że średnio prognozy różnią się od rzeczywistych cen o 0,341 jednostki. Dzięki temu możemy lepiej zrozumieć, jak duże są błędy prognozy, ponieważ RMSE jest w tej samej skali co ceny.

Model prognozuje ceny z umiarkowaną dokładnością, ale tendencja do przewidywania może wymagać pewnych poprawek, aby uzyskać lepsze wyniki.

Prognozowanie metodą Monte Carlo

Na podstawie rozkładu reszt modelu ARIMA, za pomocą symulacji Monte Carlo dla 1000 losowych wartości wyznaczono prognozy logarytmów cen i samych cen 4 notowań.

Tabela 2.1: Wartości prognoz logarytmów cen

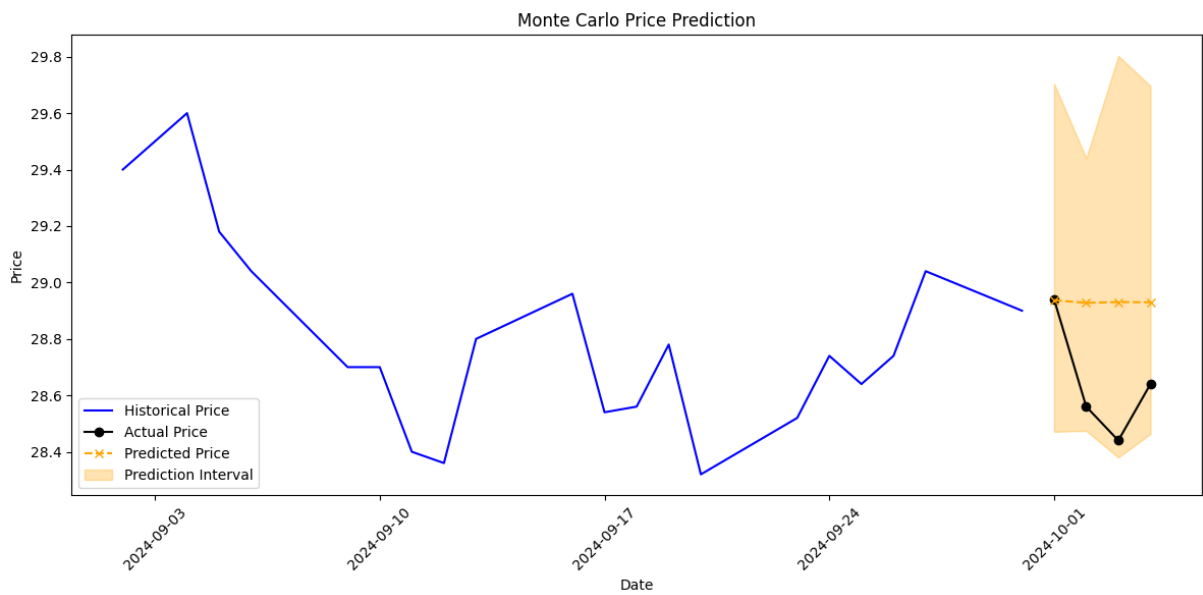
	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	3.365094	3.349323	3.391290	3.365225
02.paź	3.364769	3.344371	3.390965	3.352007
03.paź	3.364854	3.349083	3.390731	3.347797
04.paź	3.364832	3.346726	3.388506	3.354804

Tabela 2.2: Wartości prognoz dla cen

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	28.937153	28.381026	29.502267	28.94
02.paź	28.928393	28.237727	29.632742	28.56
03.paź	28.930196	28.109487	29.772939	28.44
04.paź	28.929863	28.001540	29.886412	28.64

Z analizy uzyskanych wyników wynika, że prognozowane wartości znajdują się w ustalonym przedziale ufności.

Wykres 4. Zestawienie predykcji cen, a faktycznych wartości dla symulacji Monte Carlo.



Rzeczywiste ceny znalazły się w prognozowanym przedziale ufności. Natomiast prognozowane ceny różnią się od rzeczywistych.

Tabela 2.3: Błędy prognozy ex-post dla cen

ME	MSE	RMSE
0.287704	0.114967	0.339067

Model prognozuje ceny z umiarkowaną dokładnością, ale tendencja do przewidywania może wymagać pewnych poprawek, aby uzyskać lepsze wyniki.

Prognozowanie metodą bootstrap

w tym etapie wyznaczono także 95% przedziały ufności dla wartości logarytmów cen i samych cen. Uzyskane wyniki zestawiono z poprzednimi prognozami.

Na podstawie oszacowanych reszt, za pomocą metod bootstrapowych wyznaczono prognozy logarytmów cen i samych cen 4 sesji. Następnie wyznaczono 95% przedziały ufności dla wartości logarytmów cen i samych cen. Uzyskane wyniki zestawiono z poprzednimi prognozami:

Tabela 3.1: Wartości prognoz logarytmów cen

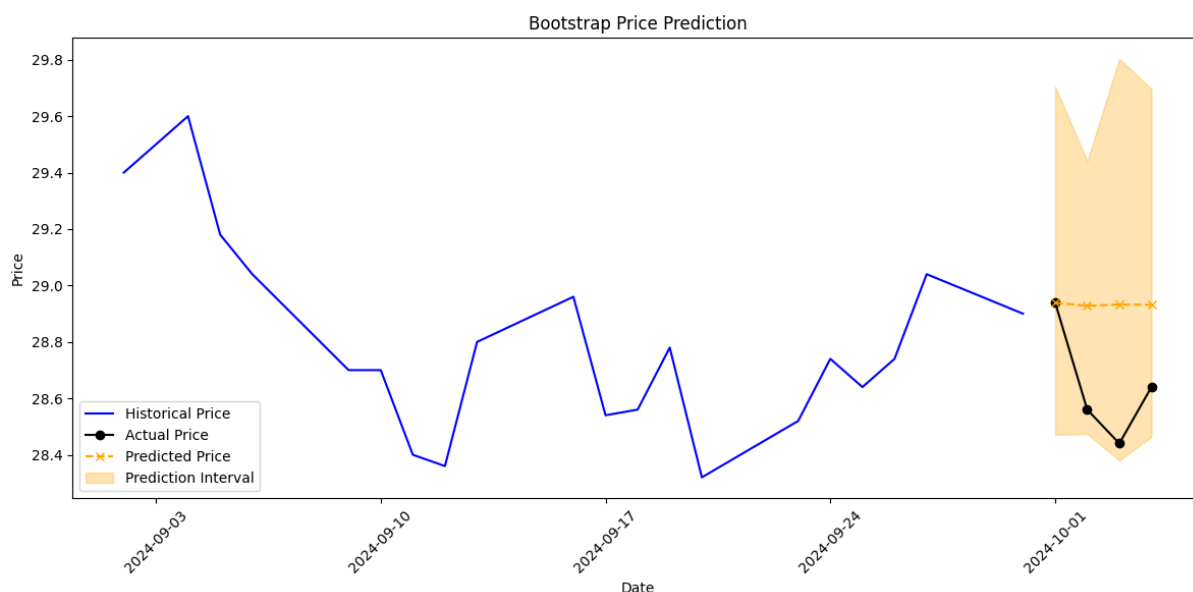
	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	3.365094	3.348891	3.391290	3.365225
02.paź	3.364769	3.348999	3.382376	3.352007
03.paź	3.364854	3.345680	3.394607	3.347797
04.paź	3.364832	3.348629	3.391027	3.354804

Tabela 3.2: Wartości prognoz dla cen

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
01.paź	28.938801	28.471137	29.704233	28.94
02.paź	28.927993	28.474211	29.440634	28.56
03.paź	28.932593	28.379878	29.802945	28.44
04.paź	28.931213	28.463672	29.696444	28.64

Wartości prognozy mieszczą się w wyznaczonych przedziałach ufności.

Wykres 5. Zestawienie predykcji cen, a faktycznych wartości dla metody Bootstrapowej.



Rzeczywiste ceny znalazły się w prognozowanym przedziale ufności. Natomiast prognozowane ceny różnią się od rzeczywistych.

Tabela 3.3: Błędy prognozy ex-post dla cen

ME	MSE	RMSE
0.288250	0.115718	0.340174

Model prognozuje ceny z umiarkowaną dokładnością, ale tendencja do przewidywania może wymagać pewnych poprawek, aby uzyskać lepsze wyniki.

Budowa modelu ARMA-GARCH i ARIMA-GARCH i powtórzenie badania

W następnym etapie projektu zostały utworzone nowe modele, przy uwzględnieniu występowania efektu ARCH, czyli wymagające zastosowania modelu GARCH: dla logarytmicznych stóp zwrotu model ARMA-GARCH, a dla logarytmów cen model ARIMA-GARCH. W tym celu zostały wykorzystane reszty z poprzednich modeli oraz został dopasowany model GARCH(1,1). Następnie wykonano ponownie wcześniejsze etapy projektu z wykorzystaniem tych modeli.

	Wartość prognozy	Przedział ufności		Rzeczywista wartość
		Dolna granica	Górna granica	
Prognoza na podstawie modelu - logarytmiczne stopy zwrotu				
01.paź	0.001637	-0.01622	0.019896	0.001383
02.paź	0.000059	0.018730	0.018848	-0.013218
03.paź	0.000466	-0.018598	-0.019530	-0.004211
04.paź	0.000361	-0.018847	0.019568	0.007008
Prognoza na podstawie modelu - logarytmy cen				
01.paź	3.365932	3.349001	3.382863	3.365225
02.paź	3.365938	3.344560	3.387316	3.352007
03.paź	3.365936	3.341960	3.389912	3.347797
04.paź	3.365937	3.340317	3.391556	3.354804
Prognoza na podstawie modelu - ceny				
01.paź	28.961562	28.474283	29.454981	28.94
02.paź	28.962368	28.348103	29.586423	28.56
03.paź	28.962770	28.274503	29.663351	28.44
04.paź	28.963088	28.228083	29.712154	28.64
Prognoza metodą Monte Carlo - logarytmy cen				
01.paź	3.365094	3.348891	3.391290	3.365225
02.paź	3.364769	3.342092	3.390965	3.352007
03.paź	3.364854	3.349094	3.390731	3.347797
04.paź	3.364832	3.346726	3.388506	3.354804
Prognoza metodą Monte Carlo - ceny				
01.paź	28.937205	28.471137	29.704233	28.94
02.paź	28.928763	28.278225	29.694594	28.56
03.paź	28.930194	28.476922	29.687642	28.44
04.paź	28.929863	28.409554	29.621670	28.64
Prognoza metodą bootstrap - logarytmy cen				
01.paź	3.365094	3.364954	3.365320	3.365225
02.paź	3.364769	3.364064	3.365557	3.352007
03.paź	3.364854	3.363701	3.366642	3.347797
04.paź	3.364832	3.364170	3.365888	3.354804
Prognoza metodą bootstrap - ceny				
01.paź	28.936216	28.932166	28.942764	28.94
02.paź	28.926828	28.906432	28.949611	28.56
03.paź	28.929264	28.895935	28.981050	28.44
04.paź	28.928630	28.909501	28.959202	28.64

Najlepsze wyniki w zakresie trafności i precyzji uzyskano dla metody bootstrap zarówno w prognozach cen, jak i logarytmów cen, gdzie rzeczywiste wartości mieszczą się w wąskich przedziałach ufności, co wskazuje na dużą dokładność tej metody. Symulacja Monte Carlo również pokazuje dobrą trafność, ale jej przedziały ufności są szersze, zwłaszcza dla prognoz cen. Model oparty na logarytmicznych stopach zwrotu oraz bezpośredni model cen generują bardziej rozbieżne wyniki, z szerszymi przedziałami ufności, co sugeruje, że są mniej precyzyjne w prognozowaniu danych w tym przypadku.

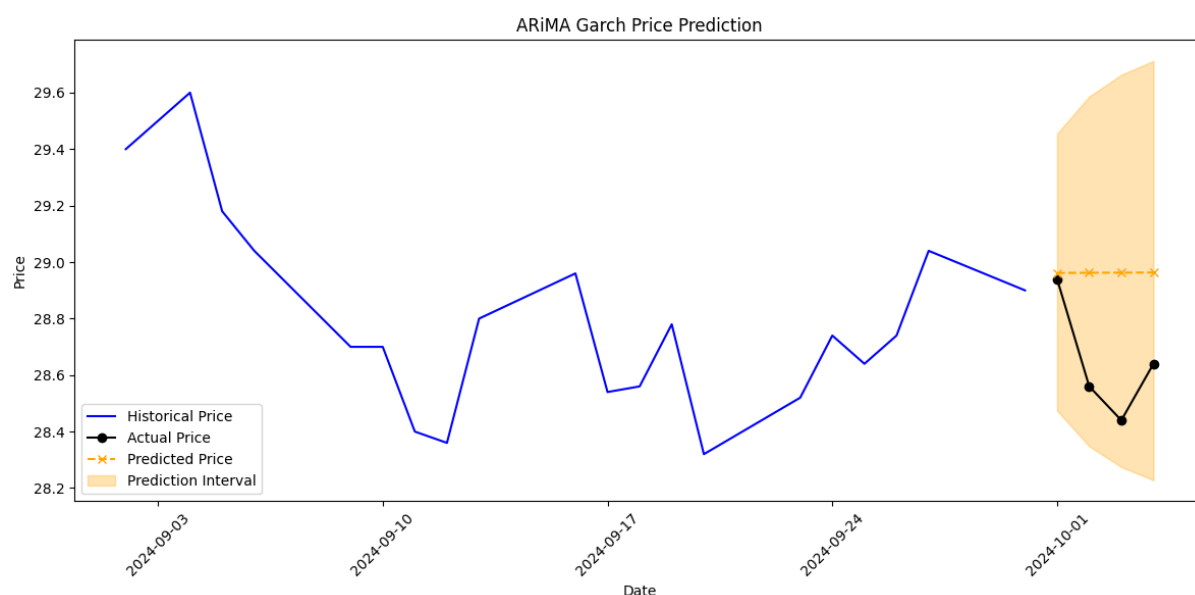
Błędy prognozy ex-post		
Prognoza na podstawie modelu - ceny		
ME	MSE	RMSE
0,317447	0,135010	0,367437
Prognoza metodą Monte Carlo - ceny		
ME	MSE	RMSE
0,287904	0,115076	0,339229
Prognoza metodą bootstrap - ceny		
ME	MSE	RMSE
0,287127	0,114316	0,338107

Najwyższe błędy występują w przypadku modelu ARIMA-GARCH, z wartościami ME wynoszącym 0,317447, MSE równym 0,135010 oraz RMSE równym 0,367437. Sugeruje to, że prognozy z tego modelu są najbardziej oddalone od rzeczywistych wartości, co oznacza niższą dokładność w porównaniu z pozostałymi metodami.

Symulacja Monte Carlo osiąga lepsze wyniki, z ME równym 0,287904, MSE 0,115076 i RMSE 0,339229, co wskazuje na zmniejszenie przeciętnego błędu i lepsze odwzorowanie rzeczywistych wartości w stosunku do modelu ARIMA-GARCH. Jednak najniższe wartości błędów osiąga metoda bootstrap, z ME na poziomie 0,287127, MSE równym 0,114316 i RMSE równym 0,338107. Te wartości świadczą o tym, że metoda bootstrap najdokładniej odwzorowuje rzeczywiste dane, minimalizując średnie i kwadratowe błędy prognoz.

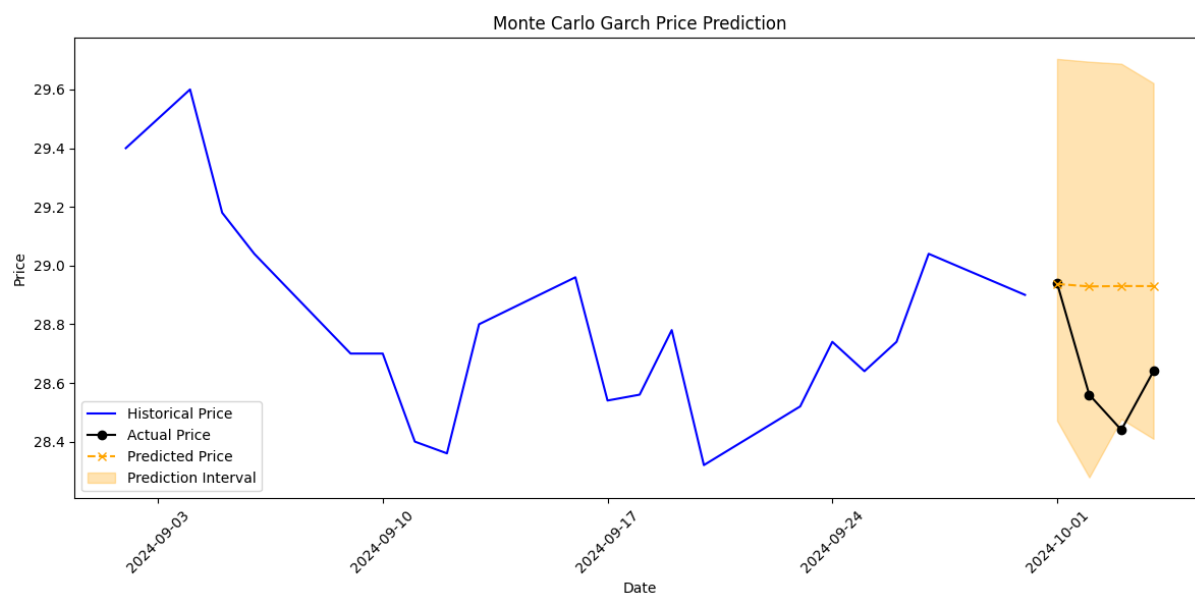
Podsumowując, analiza wskazuje, że metoda bootstrap jest najbardziej efektywna w prognozowaniu, generując najniższy poziom błędów, co czyni ją najlepszą z zaprezentowanych technik.

Wykres 6. Zestawienie predykcji cen, a faktycznych wartości dla modelu ARiMA-Garch.



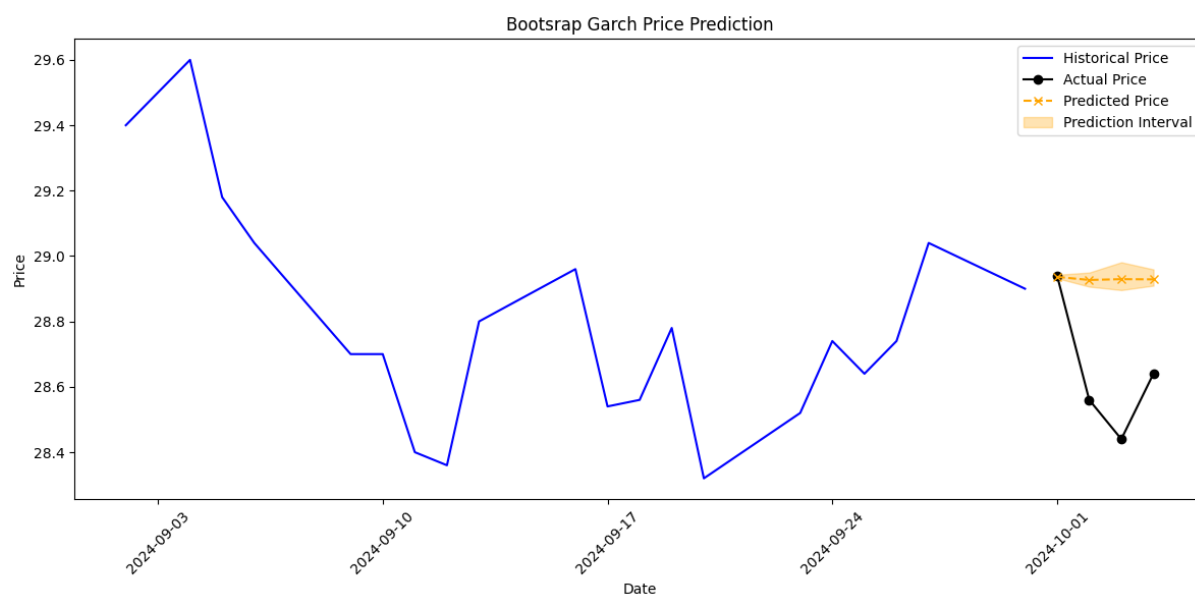
Rzeczywiste ceny znalazły się w prognozowanym przedziale ufności. Natomiast prognozowane ceny różnią się od rzeczywistych.

Wykres 7. Zestawienie predykcji cen, a faktycznych wartości dla symulacji Monte Carlo Garch.



Rzeczywiste ceny znalazły się w prognozowanym przedziale ufności. Natomiast prognozowane ceny różnią się od rzeczywistych.

Wykres 8. Zestawienie predykcji cen, a faktycznych wartości dla metody Bootstrap Garch.



Ta metoda jako jedyna utworzyła predykcje, w której 3 ostatnie rzeczywiste ceny znajdują się poza wyznaczonym przedziałem ufności.

Podsumowanie

W niniejszym badaniu analizowano dzienne logarytmiczne stopy zwrotu kursu spółki MOL. W oparciu o wyestymowany model, symulacje Monte Carlo, prognozowanie metodą bootstrap przeprowadzono prognozy logarytmicznych stóp zwrotu oraz logarytmów cen dla czterech kolejnych sesji. Do każdej prognozy wyznaczono 95% przedział ufności. Następnie z prognoz logarytmów cen wyliczono prognozy cen, przyjmując ich warunkowy rozkład log-normalny. Prognozowane ceny porównano z rzeczywistymi wartościami dla pierwszych czterech sesji października 2024.

W kontekście praktycznego podejścia do prognozowania notowań giełdowych można stwierdzić, że na podstawie przeprowadzonego badania, mimo uzyskania zadowalających wyników, nie zaobserwowano rzeczywistego odzwierciedlenia prognoz w kolejnych dniach po 1 października. Badanie wykazało, że wszystkie metody dobrze prognozowały wartości notowań dla pierwszego dnia, natomiast w przypadku kolejnych dni stwierdzono znacząco gorsze wyniki w porównaniu do prognoz dla dnia pierwszego.

Ogólnie, analizując wykresy notowań spółki MOL w danym roku, zauważono duże wzrosty i spadki notowań, a także wypłacenie dywidendy akcjonariuszom, co z pewnością negatywnie wpływa na możliwości predykcji.