

Wprowadzenie

Celem niniejszego projektu jest analiza jednowymiarowych szeregów czasowych na podstawie danych finansowych spółki „DINO POLSKA” S.A. dotyczących zamknięcia, z naciskiem na prognozowanie przyszłych cen akcji. Dodatkowo, projekt obejmuje przeprowadzenie analizy porównawczej różnych modeli statystycznych, takich jak ARIMA i SARIMA, w celu wybrania najbardziej efektywnego podejścia do prognozowania.

Dino to ogólnopolska sieć supermarketów, która jest jednym z liderów w dynamicznie rozwijającym się segmencie średniej wielkości sklepów, usytuowanych w pobliżu miejsc zamieszkania klientów. Jest również jedną z najszybciej rozwijających się sieci w Polsce, zarówno pod względem liczby placówek, jak i wartości sprzedaży, na rynku detalicznym artykułów spożywczych (Dino Polska, 2025).

W ramach niniejszego projektu dokonano analizy danych finansowych, które zostały pobrane automatycznie z serwisu Yahoo Finance dzięki oprogramowaniu R. Proces ten pozwolił na bezpośrednie pobranie danych dotyczących kursów akcji wybranej spółki, które zostały wykorzystane do przeprowadzenia dalszej analizy. Wszelkie obliczenia, analizy statystyczne oraz tworzenie wykresów zostały przeprowadzone przy użyciu języka programowania R, który pozwolił na dokładne modelowanie i wizualizowanie wyników.

Podsumowując, projekt ma na celu opracowanie skutecznych modeli prognozujących ceny akcji spółki DINO, z wykorzystaniem zaawansowanych metod analizy szeregów czasowych. Ostateczne wyniki pozwolą na ocenę, który z testowanych modeli będzie najlepiej spełniał rolę narzędzia prognozującego zmiany cen w przyszłości.

Ujęcie teoretyczne

Analiza szeregów czasowych stanowi jeden z kluczowych obszarów badań w statystyce i ekonometrii, umożliwiając modelowanie i prognozowanie zmiennych, które są uporządkowane w czasie. W ramach tego typu analizy stosuje się szereg narzędzi i metod, które pozwalają na uchwycenie struktury czasowej danych oraz ich prognozowanie na podstawie dostępnych informacji historycznych. Wśród najczęściej wykorzystywanych modeli znajdują się modele ARIMA oraz SARIMA, które stanowią podstawę do prognozowania przyszłych wartości zmiennych na podstawie danych przeszłych.

Jednym z podstawowych założeń stosowania klasycznych modeli szeregów czasowych, takich jak ARIMA, jest stacjonarność szeregu czasowego. Oznacza to, że właściwości statystyczne szeregu, takie jak średnia, wariancja czy współczynniki autokorelacji, nie zmieniają się w czasie. W przypadku, gdy dane nie wykazują stacjonarności, stosuje się odpowiednie transformacje, takie jak różnicowanie szeregu, które prowadzą do osiągnięcia

stacjonarności. Aby ocenić stacjonarność szeregu czasowego, stosuje się rozszerzony test Dickey'a-Fuller'a (ADF) (Sirisha, Belavagi, Attigeri, 2022). Test ten sprawdza obecność pierwiastków jednostkowych w szeregu. Hipoteza zerowa testu zakłada obecność przynajmniej jednego pierwiastka jednostkowego, co wskazuje na niestacjonarność danych. Odrzucenie hipotezy zerowej oznacza, że dane są stacjonarne (Osińska, 2006). W przypadku niestacjonarności stosuje się różnicowanie danych w celu usunięcia trendów i sezonowości.

Dla identyfikacji rzędu modelu ARMA(p , q) kluczowe znaczenie mają funkcje autokorelacji (ACF) oraz częściowej autokorelacji (PACF). Analiza tych funkcji pozwala na określenie wartości parametrów modelu, w tym rzędu składnika autoregresyjnego (AR) oraz składnika średniej ruchomej (MA). Funkcje te bazują na badaniu zależności między kolejnymi obserwacjami szeregu czasowego, a także na minimalizacji wariancji składnika resztowego modelu (Osińska, 2006).

Funkcja autokorelacji (ACF) mierzy zależności między kolejnymi obserwacjami szeregu czasowego, wskazując, jak wartości są ze sobą powiązane na różnych opóźnieniach. Analiza ACF ujawnia charakterystyczne cechy (Sirisha, Belavagi, Attigeri, 2022):

- Szybki spadek wartości autokorelacji sugeruje, że niskie wartości parametru q mogą być wystarczające.
- Długi ogon autokorelacji może wskazywać na potrzebę dodatkowego różnicowania szeregu czasowego.

Funkcja autokorelacji dostarcza informacji na temat pamięci procesu stochastycznego, co oznacza, że wpływ impulsu w czasie t może utrzymywać się w kolejnych okresach, w zależności od liczby niezerowych współczynników autokorelacji (Osińska, 2006).

Funkcja częściowej autokorelacji (PACF) mierzy zależności pomiędzy obserwacjami na określonych opóźnieniach, uwzględniając jedynie bezpośredni wpływ tych opóźnień, po wyeliminowaniu wpływu wcześniejszych opóźnień. Analiza PACF wskazuje na kluczowe zależności (Sirisha, Belavagi, Attigeri, 2022):

- Gwałtowny spadek wartości PACF po określonym opóźnieniu sugeruje odpowiednią wartość parametru p .
- Brak istotnych wartości PACF może wskazywać, że składnik AR nie jest konieczny w modelu.

Współczynniki funkcji PACF określają korelację między błędami predykcji, mierząc zależność między wartościami Y_t a Y_{t-k} , po wyeliminowaniu wpływu pozostałych opóźnień.

Jednakże, analiza ACF i PACF nie zawsze daje jednoznaczne wyniki, szczególnie w przypadku modelu ARMA, który łączy składniki AR i MA. W takich przypadkach konieczne jest wykorzystanie dodatkowych metod diagnostycznych (Osińska, 2006).

Kryteria informacyjne stanowią istotne narzędzie przy identyfikacji modelu ARMA w empirycznych szeregach czasowych (Osińska, 2006). W celu wyznaczenia rzędu opóźnień modelu ARMA przyjmuje się relatywnie wysokie wartości p i q i estymuje modele dla wszystkich możliwych kombinacji tych parametrów, a następnie wyborze najlepszego modelu na podstawie wybranego kryterium informacyjnego:

1. Kryterium informacyjne Akaike'a (AIC), preferowany jest model, który minimalizuje utratę informacji, co oznacza najniższą wartość AIC.
2. Bayesowskie rozszerzenie minimum AIC (BIC), wybór pada na model charakteryzujący się najniższą wartością BIC

Dla małych próbek istnieje skorygowana wersja AIC, która jest bardziej odpowiednia niż standardowy AIC w przypadku mniejszych próbek oraz bardziej odpowiednia niż BIC dla bardzo małych próbek (Brewer, Butler, Cooksley, 2016). W przypadku rozbieżności między modelami wskazanymi przez kryteria AIC i BIC, preferowany jest model wybrany na podstawie BIC. Kryterium to stanowi modyfikację AIC, uwzględniającą korektę tendencji AIC do wyboru modeli z nadmierną liczbą parametrów (Osińska, 2006).

Do automatycznego dopasowania parametrów modeli ARIMA w oprogramowaniu statystycznym często wykorzystuje się funkcje, takie jak *auto.arima()*, która na podstawie wybranych kryteriów optymalizuje parametry modelu i zwraca najlepszy model ARIMA na podstawie wartości AIC, AICc lub BIC.

W ocenie modelu ARIMA często stosuje się standardowy błąd oszacowania modelu, który oblicza się według wzoru:

$$\hat{\sigma}_k = \sqrt{\frac{1}{n - p - q} \sum_{t=1}^n \hat{e}_t^2}$$

gdzie \hat{e}_t oznacza reszty modelu.

Błąd ten służy nie tylko do oceny dopasowania modelu, ale także do wyznaczania wariancji estymowanych parametrów oraz ich istotności statystycznej. Niższa wartość błędu wskazuje na lepsze dopasowanie modelu do danych.

Aby ocenić, czy reszty modelu są losowe, można zastosować test Ljunga-Boxa, który jest rozszerzeniem testu Boxa-Pierce'a. Test ten pozwala zweryfikować hipotezę zerową, zakładającą brak autokorelacji reszt (losowość), w porównaniu do hipotezy alternatywnej, sugerującej istnienie zależności w resztach. Jeśli hipoteza zerowa zostaje odrzucona, oznacza to, że reszty nie są losowe, co może wskazywać na niedostateczne dopasowanie modelu (Chodakowska, Nazarko, 2022).

Celem modelowania szeregów czasowych jest uzyskanie modelu, który będzie w stanie prognozować przyszłe wartości na podstawie danych historycznych. Prognozy te ocenia się na podstawie różnych miar błędów prognoz. W analizie błędów prognoz ex post stosowane są zarówno mierniki absolutne jak i względne. Do najczęściej stosowanych mierników absolutnych należą (symbole mierników składają się z pierwszych liter ich nazw w języku angielskim) :

- średni błąd predykcji

$$ME = \frac{1}{r} \sum_{t=1}^r (Y_t - Y_{tp})$$

- średni błąd absolutny

$$MAE = \frac{1}{r} \sum_{t=1}^r |Y_t - Y_{tp}|$$

- pierwiastek błędu średniokwadratowego

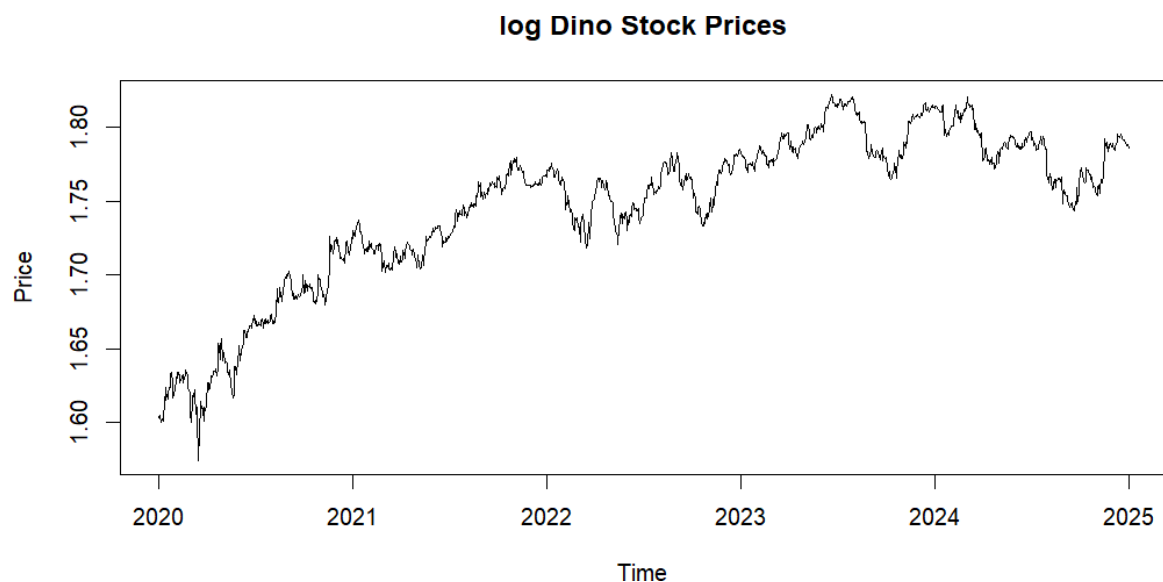
$$RMSE = \sqrt{\frac{1}{r} \sum_{t=1}^r (Y_t - Y_{tp})^2}$$

Między miernikami zachodzi relacja $ME \leq MAE \leq RMSE$. Relacja między miernikami (2) i (3) wynika bezpośrednio z relacji między średnią z sumy wartości bezwzględnych i pierwiastkiem ze średniej sumy kwadratów. (Szmuksta-Zawadzka, Zawadzki, 2012)

Porównując różne modele, można wybrać ten, który generuje najbardziej dokładne prognozy, a tym samym jest najbardziej odpowiedni do danej aplikacji. W praktyce porównuje się wyniki uzyskane za pomocą różnych metod, takich jak ARIMA i SARIMA, aby wybrać model najlepiej dopasowany do charakterystyki danych.

Analiza i dopasowanie modeli

W pierwszym etapie projektu dokonano zebrania danych finansowych spółki DINO z platformy Yahoo Finance dzięki bibliotece `quantmod` w języku R. Dane obejmują codzienne ceny zamknięcia akcji od 1 stycznia 2020 roku do 31 grudnia 2024 roku. Automatyzacja tego procesu za pomocą funkcji `getSymbols()` znacznie usprawniła etap pozyskiwania danych. Zebrane dane obejmują codzienne ceny zamknięcia akcji, które zostały przekształcone na logarytmiczny szereg czasowy w celu ułatwienia analizy i modelowania.



Logarytmiczne stopy zwrotów spółki Dino

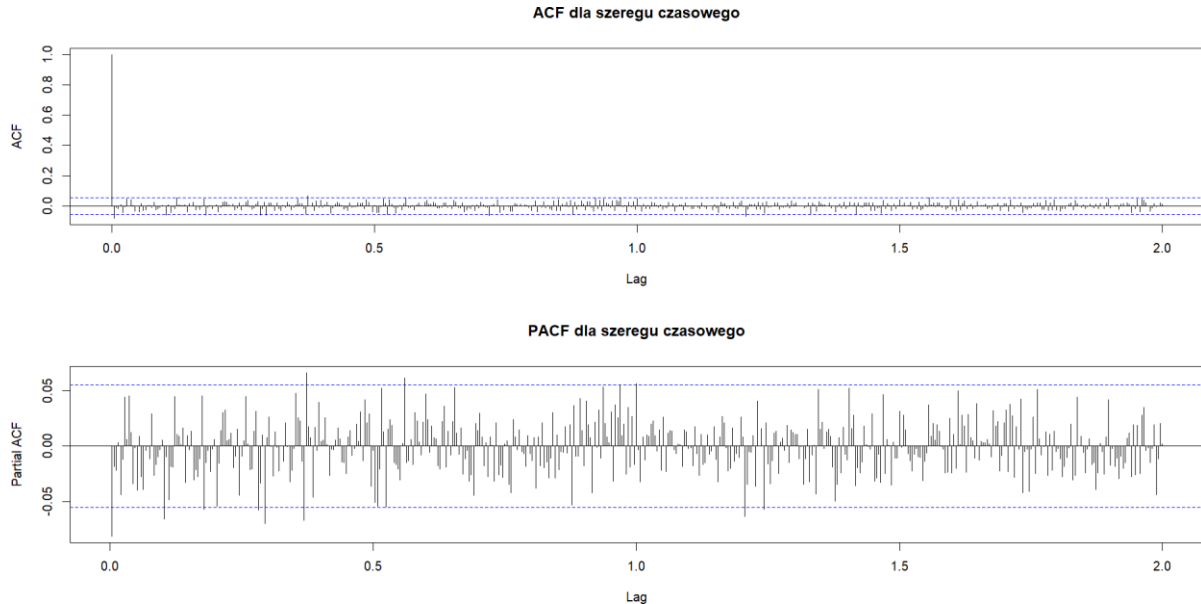
Po przygotowaniu danych przeprowadzono wstępną analizę stacjonarności szeregu czasowego, która ma kluczowe znaczenie w kontekście dalszego modelowania.

Tabela 1. Wyniki testu ADF

	Surowe dane	Dane po jednym zróżnicowaniu
Wartość statystyki	-2.7368	-10.594
Lag order	10	10
p-value	0.2664	0.01

Surowe dane nie wykazywały cech stacjonarności, dlatego zastosowano odpowiednie transformacje, takie jak różnicowanie szeregu, aby uzyskać stacjonarność, co jest niezbędne dla wielu metod modelowania. Szereg jest zintegrowany $I(1)$.

W projekcie zastosowano funkcje ACF (autokorelacji) i PACF (autokorelacji cząstkowej) jako narzędzia do analizy struktury szeregów czasowych.



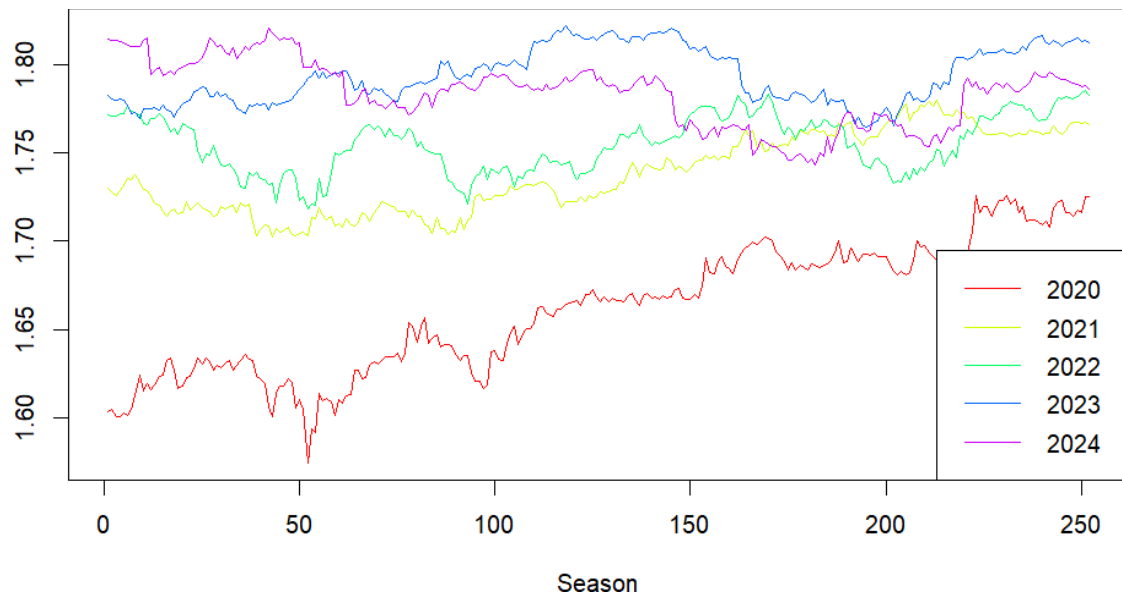
Wykresy autokorelacji i autokorelacji cząstkowej dla zróżnicowanego szeregu

Dla zróżnicowanego szeregu czasowego stopnia I, można wyciągnąć następujące wnioski:

1. Wykresy nie wskazują na obecność sezonowości, co sugeruje, że dane nie mają wyraźnych cyklicznych wzorców, które należałoby uwzględnić w modelu. Być może sezonowość jest subtelna i ujawnia się bardziej w wartościach modelu niż na podstawie wizualnej analizy ACF i PACF.
2. Na podstawie wykresów ACF i PACF, model ARIMA(0,1,1) (lub podobny) wydaje się być odpowiedni. Istotne pierwsze wartości na w wykresie ACF i brak wyraźnych wartości w wykresie PACF wskazują na potrzebę uwzględnienia jednego składnika średniej ruchomej (MA), co odpowiada strukturze modelu ARIMA(0,1,1).

Te wnioski sugerują, że przy dalszym modelowaniu warto rozważyć tę strukturę, jednak należy pamiętać o dalszym testowaniu i weryfikacji wyników na zbiorze testowym.

Kolejnym krokiem była analiza sezonowości i trendów w danych. Zastosowane wykresy sezonowe oraz dekompozycja szeregu czasowego pozwoliły na wyodrębnienie składników takich jak trend, sezonowość oraz reszta. Dzięki temu możliwe było także określenie, czy w modelu należy uwzględnić komponenty sezonowe, co ma duże znaczenie w kontekście prognozowania na przyszłość.

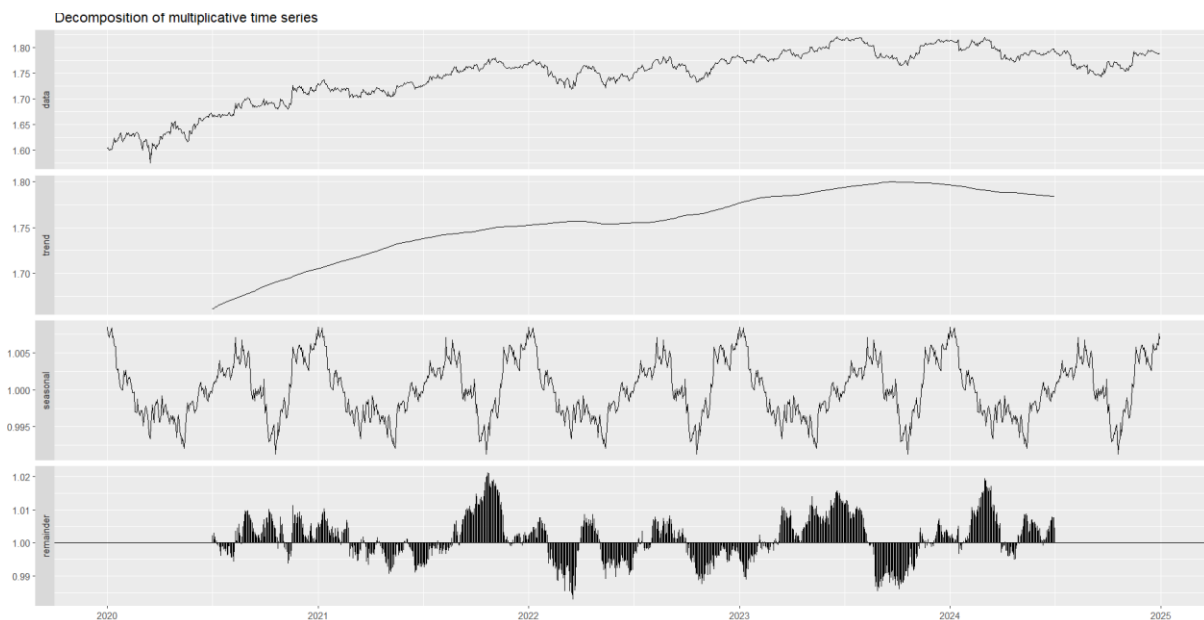


Wykres sezonowości

Wykres przedstawia sezonowość logarytmicznych cen zamknięcia akcji Dino Polska na przestrzeni lat 2020–2024. Każda linia reprezentuje dany rok i pokazuje zmiany cen akcji w ciągu roku (około 252 dni giełdowych).

Wyraźnie widoczne jest ogólne zwiększanie się poziomu cen akcji z roku na rok. Linia 2024 (fioletowa) znajduje się wyżej niż poprzednie lata, co sugeruje, że cena akcji systematycznie rośnie.

Chociaż ceny akcji zmieniają się w każdym roku, kształt wykresu dla różnych lat jest dość podobny. Istnieją pewne powtarzalne wzorce – np. dolki i szczyty występujące w podobnych okresach w różnych latach. Możliwe jest istnienie specyficznych miesięcznych lub kwartalnych wzorców w zachowaniu cen. Wartości w roku 2020 są niższe niż w kolejnych latach, co może wynikać z wpływu pandemii COVID-19 na rynek.



Dekompozycja szeregu czasowego

W wyniku dekompozycji szeregu czasowego zaobserwowano następujące cechy:

1. Wartości szeregu czasowego wykazują stopniowy wzrost do około 2024 roku, po czym trend zdaje się stabilizować. Taki wzrost może wskazywać na długoterminowy rozwój wartości akcji spółki DINO, z możliwymi okresami spowolnienia, które wpływają na dynamikę zmian.
2. Wzór w składniku sezonowym sugeruje istnienie cykliczności, która może być związana z rocznymi zjawiskami, takimi jak zmiany na rynku, wydarzenia gospodarcze, czy też efekty sezonowe związane z działalnością spółki. Cykl sezonowy może mieć wpływ na wartość akcji w zależności od pór roku.
3. Składnik resztowy po dekompozycji nie wykazuje wyraźnych wzorców, co sugeruje, że reszty są losowe i nie ma w nich istotnych systematycznych błędów. Taki wynik oznacza, że model dobrze uchwycił struktury w danych, a pozostałe wariacje są przypadkowe i nie wskazują na dodatkowe zależności, które należałoby uwzględnić w modelu.

W procesie estymacji parametrów wykorzystano dwa podejścia do modelowania szeregu czasowego. Pierwszym modelem był SARIMA (0,1,1)(1,0,0)[252], uzyskany za pomocą funkcji `auto.arima()`. Model ten uwzględniał sezonowość i okazał się najlepszy pod względem dopasowania do danych, z wartością σ^2 równą 0.000559. Sezonowość, uwzględniona w modelu, była uzasadniona przez cykliczne wzorce widoczne w danych.

Wzór dla modelu SARIMA(0,1,1)(1,0,0)[252] przedstawia się następująco:

$$(1 - 0.0728B_{252})(1 - B)Y_t = (1 - 0.0756B)\varepsilon_t$$

gdzie:

- Y_t - wartość szeregu w chwili t
- B - operator opóźnienia ($BY_t = Y_{t-1}$)
- $\phi_1 = 0.0728$ - współczynnik sezonowy autoregresji
- $\theta_1 = -0.0756$ - współczynnik średniej ruchomej
- ε_t - składnik resztowy.

Drugim podejściem było manualne dopasowanie modelu ARIMA (0,1,1), który nie uwzględniał komponentu sezonowego, wykorzystano do tego funkcję Arima. Model ten miał nieznacznie gorszą wartość σ^2 (0.0005606), co sugeruje, że obecność sezonowości w poprzednim modelu rzeczywiście miała pozytywny wpływ na dopasowanie modelu. Choć różnica ta nie jest duża, analiza dalszych parametrów, takich jak reszty, może wskazać, czy sezonowość dodaje zbędną złożoność do modelu.

Wzór dla modelu ARIMA(0,1,1) przedstawia się następująco:

$$(1 - B)Y_t = (1 - 0.0759B)\varepsilon_t$$

gdzie:

- Y_t - wartość szeregu czasowego w chwili t
- B - operator opóźnienia ($BY_t = Y_{t-1}$)
- $\theta_1 = -0.0759$ - współczynnik średniej ruchomej
- ε_t - składnik losowy.

Trzecim podejściem było manualne dopasowanie modelu SARIMA (1,1,1)(0,0,1)[252]. Może on być alternatywą dla modelu SARIMA(0,1,1)(1,0,0)[252] ponieważ uwzględnia dodatkowy parametr sezonowy MA(1), co oznacza, że wpływ szumu błędu z poprzedniego sezonu jest uwzględniany - może to poprawić dopasowanie. Wykorzystując funkcję Arima otrzymano wzór, który przedstawia się następująco:

$$(1 - 0.1481B_{252})(1 - B)Y_t = (1 - 0.2229B)(1 + 0.0697B^{252})\varepsilon_t$$

gdzie:

- Y_t - wartość szeregu w chwili t
- B - operator opóźnienia ($BY_t = Y_{t-1}$)
- $\phi_1 = 0.1481$ - współczynnik sezonowy autoregresji
- $\theta_1 = -0.2229$ - współczynnik średniej ruchomej
- $\Theta_1 = 0.0697$ - parametr sezonowej średniej ruchomej (252 dni wstecz),
- ε_t - składnik resztowy.

Dalsze porównanie tych trzech modeli pod kątem ich wyników i zastosowań w prognozowaniu pozwoli na podjęcie ostatecznej decyzji o najlepszym modelu do wykorzystania w prognozowaniu wartości akcji spółki DINO.

Tabela 2. Analiza porównawcza parametrów poszczególnych modeli

Parametr	SARIMA(0,1,1)(1,0,0)[252]	ARIMA (0,1,1)	SARIMA(1,1,1)(0,0,1)[252]
σ^2	0.000559	0.0005606	0.0005595
Log-Likelihood	2928.47	2926.27	2928,47
AIC	-5850.95	-5848.54	-5848.94
AICc	-5850.93	-5848.53	-5848.91
BIC	-5835.54	-5838.26	-5828.39
Training Set Error Measures			
ME	0.0007859311	0.0008599722	0.0008038033
RMSE	0.02361472	0.02366874	0.02361589
MAE	0.01661762	0.01663805	0.01662511
MPE	0.01355695	0.01481563	0.01386529
MAPE	0.2928741	0.2932985	0.2930156
MASE	0.07295741	0.9937355	0.07299029
ACF1	-0.0005033273	-0.0006689677	-0.001325009

1. Porównanie kryteriów statystycznych (dopasowanie modelu):

- SARIMA (0,1,1)(1,0,0)[252] ma najniższą wartość AIC (-5850.95) i AICc (-5850.93), co oznacza, że spośród trzech modeli najlepiej dopasowuje się do danych.
- Wartość BIC (-5835.54) jest nieco wyższa niż dla ARIMA (0,1,1) (-5838.26), ale różnica jest niewielka. BIC jest bardziej konserwatywnym kryterium, więc nieznacznie lepsza wartość dla ARIMA (0,1,1) nie musi oznaczać, że jest to lepszy model.
- Wariancja reszt (σ^2) dla wszystkich modeli jest bardzo zbliżona, co oznacza, że różnice w precyzji dopasowania są niewielkie.

2. Analiza jakości prognoz (błędy modelu):

- SARIMA (0,1,1)(1,0,0)[252] osiąga najniższe wartości RMSE (0.02361472) i MAE (0.01661762), co oznacza, że przewiduje wartości najbardziej precyzyjnie.
- MAPE (0.2928741) również jest najniższe, co oznacza, że model lepiej przewiduje zmiany względne w cenach.
- ME (0.00078593), czyli średni błąd, jest najniższy w SARIMA (0,1,1)(1,0,0)[252], co oznacza, że model nie ma tendencji do systematycznego przeszacowywania lub niedoszacowywania wartości.

Porównując wyniki trzech modeli, model SARIMA(0,1,1)(1,0,0)[252] wydaje się najlepiej dopasowanym modelem, osiągającym najniższe wartości AIC i BIC, a także wykazującym bardzo zbliżone błędy prognozy do modelu ARIMA. Pomimo że różnice w błędach prognozy są niewielkie, model SARIMA może być bardziej korzystny w kontekście jego zdolności do uchwycenia sezonowości w danych.

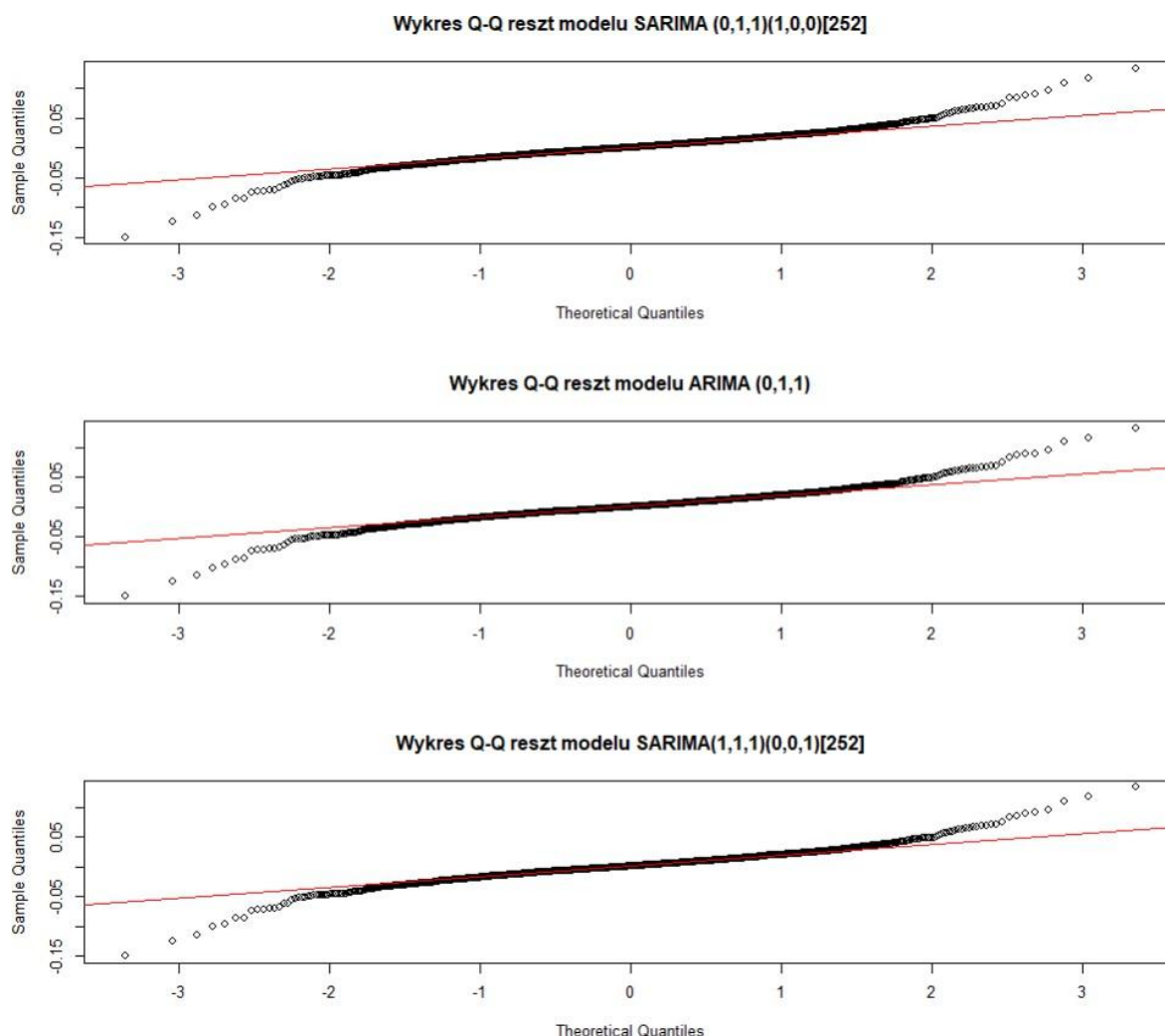
Następnie przeprowadzono ocenę jakości modelu poprzez analizę reszt oraz testy na obecność białego szumu, wykorzystując przy tym testy oparte na statystyce Ljunga Boxa.

Tabela 3. Wyniki testu Ljunga-Boxa

	SARIMA(0,1,1)(1,0,0)[252]	ARIMA (0,1,1)	SARIMA(1,1,1)(0,0,1)[252]
p-value	0.8859466	0.8649924	0.8879232

Wyniki tych testów wykazały, że wszystkie modele dobrze uchwyciły struktury w danych, nie pozostawiając przy tym istotnych wzorców. Najlepszy wynik ma SARIMA(1,1,1)(0,0,1)[252], ale różnice między modelami są bardzo małe. SARIMA(0,1,1)(1,0,0)[252] nadal wydaje się najlepszym wyborem, ponieważ w poprzednich analizach miała najniższe wartości AIC/AICc, co oznacza, że lepiej dopasowuje się do danych.

Po przeprowadzeniu testów Ljunga-Boxa, które miały na celu ocenę autokorelacji reszt w modelach, wygenerowano wykresy Q-Q dla reszt. Wykresy te zostały zaprezentowane w celu wizualnej oceny rozkładu reszt oraz ich dopasowania do rozkładu normalnego.



Wykresy Q-Q reszt modeli

Na podstawie obserwacji wykresów można zauważyć, że reszty generalnie dobrze pasują do rozkładu normalnego, z drobnymi odstępstwami jedynie na końcach, co może sugerować obecność białego szumu. Tego rodzaju analiza potwierdza, że model dobrze odwzorowuje dane, chociaż końcowe reszty mogą wskazywać na pewne niuanse, które warto obserwować w dalszej analizie.

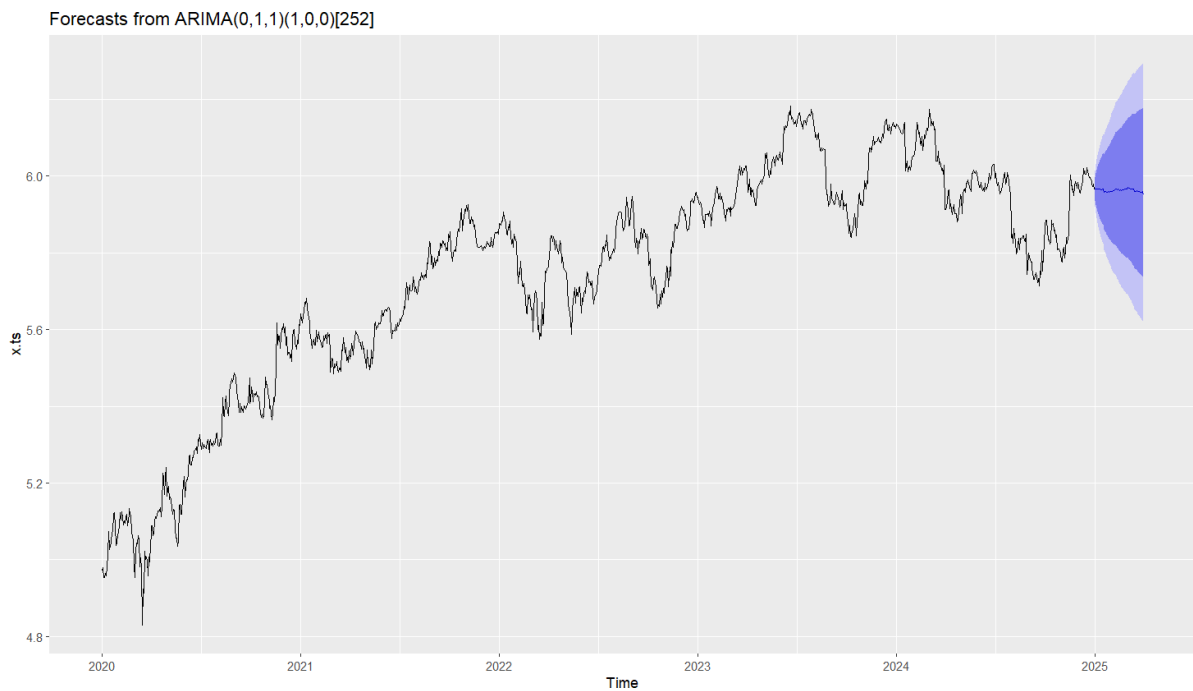
Najlepszy wykres Q-Q to ten dla modelu SARIMA(0,1,1)(1,0,0)[252], ponieważ punkty układają się najbliżej linii diagonalnej, co oznacza, że reszty są najbardziej zgodne z rozkładem normalnym. Dodatkowo, skrajne wartości mają najmniejsze odchylenia, co świadczy o mniejszej liczbie wartości odstających. Potwierdza to wcześniejsze wyniki testu Ljungu-Boxa, gdzie model ten również miał dobre wartości p-value.

Po przeprowadzeniu analiz diagnostycznych na resztach modeli ARIMA i SARIMA, przyszedł czas na etap prognozowania. Prognozowanie jest kluczowym krokiem w analizie szeregów czasowych, pozwalającym na przewidywanie przyszłych wartości na podstawie zebranych danych historycznych. W przypadku modeli ARIMA i SARIMA, prognozy będą

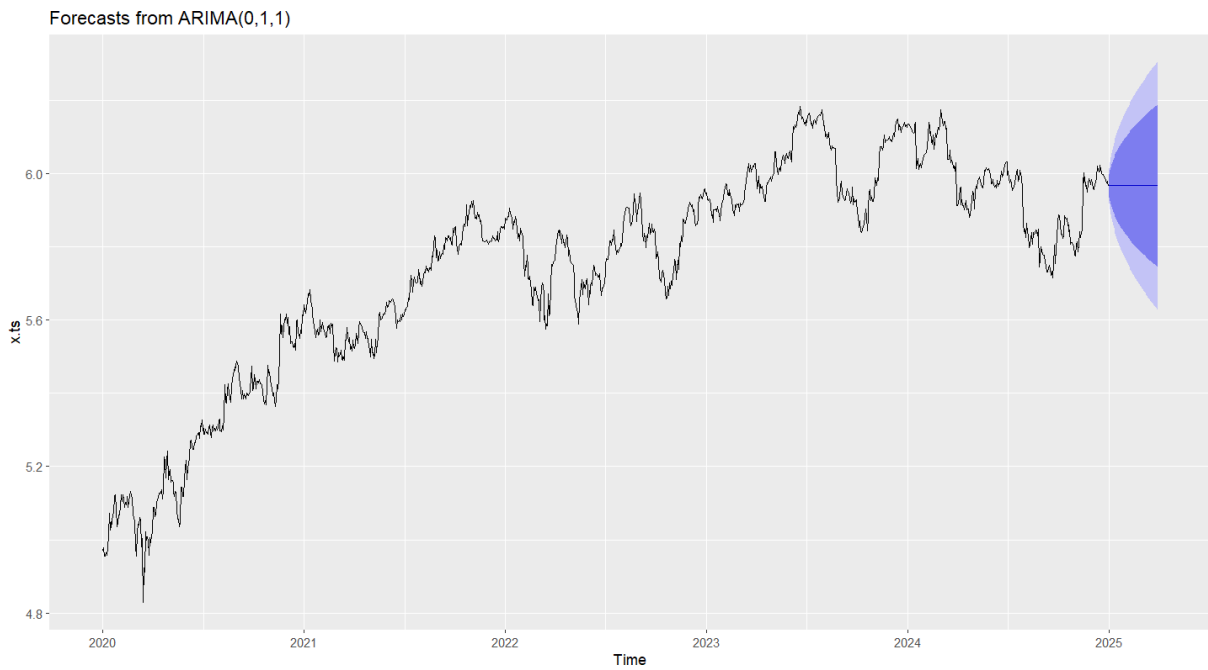
tworzone na podstawie optymalnych parametrów uzyskanych podczas procesu modelowania.

Prognozowanie

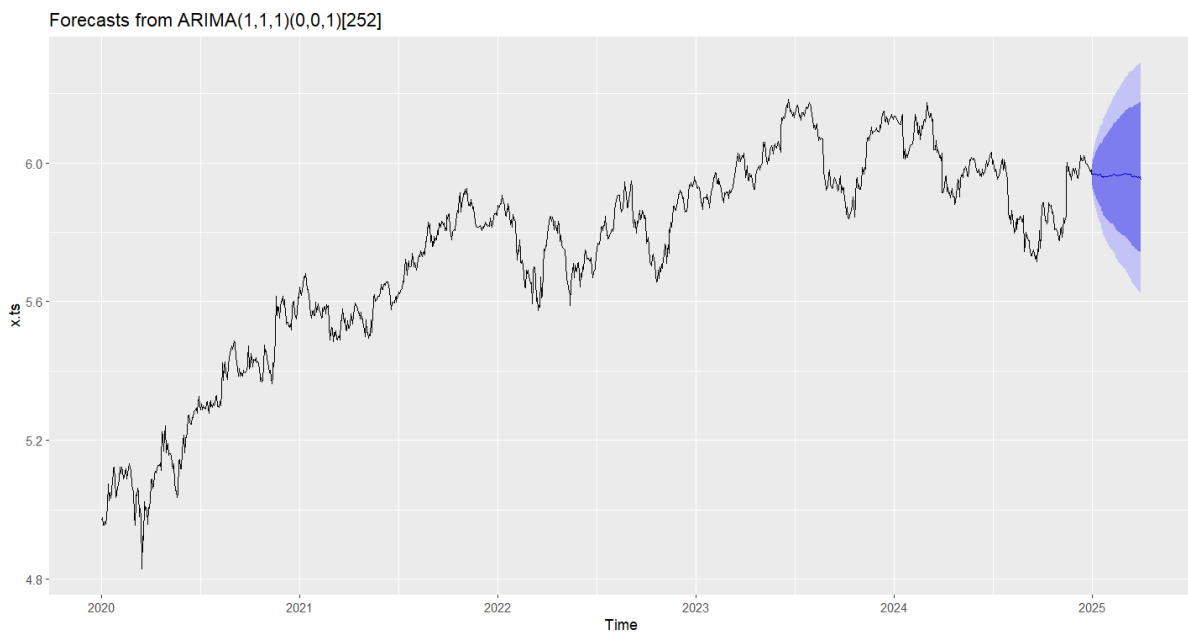
Prognozy modeli na poniższych wykresach obejmują przyszłe wartości wraz z pasmami niepewności (strefami ufności zaznaczonymi na niebiesko).



Wykres prognozy dla modelu SARIMA (0,1,1)(1,0,0)[252]



Wykres prognozy dla modelu ARIMA (0,1,1)



Wykres prognozy dla modelu SARIMA (1,1,1)(0,0,1)[252]

SARIMA(0,1,1)(1,0,0)[252]

- Prognoza uwzględnia sezonowość, co może być korzystne, jeśli dane wykazują powtarzające się wzorce roczne.

- Przedział ufności jest dość wąski, co sugeruje stabilność prognozy i wysoką pewność modelu.
- Prognoza wygląda bardziej dynamicznie, czyli podąża za trendem historycznym.

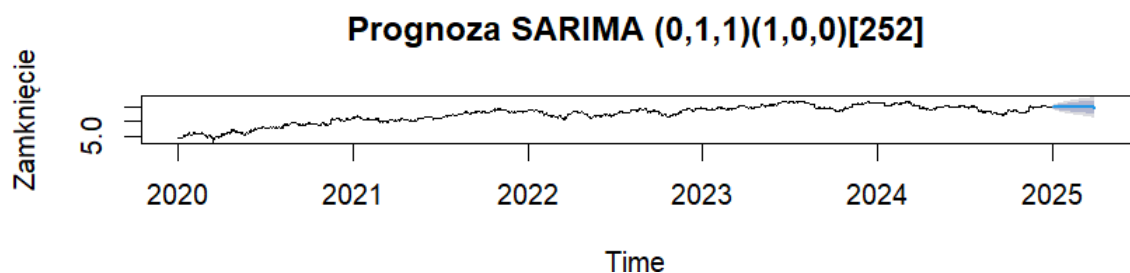
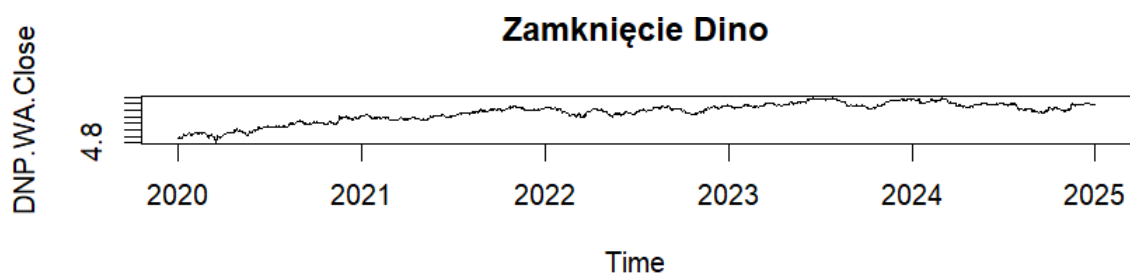
ARIMA(0,1,1)

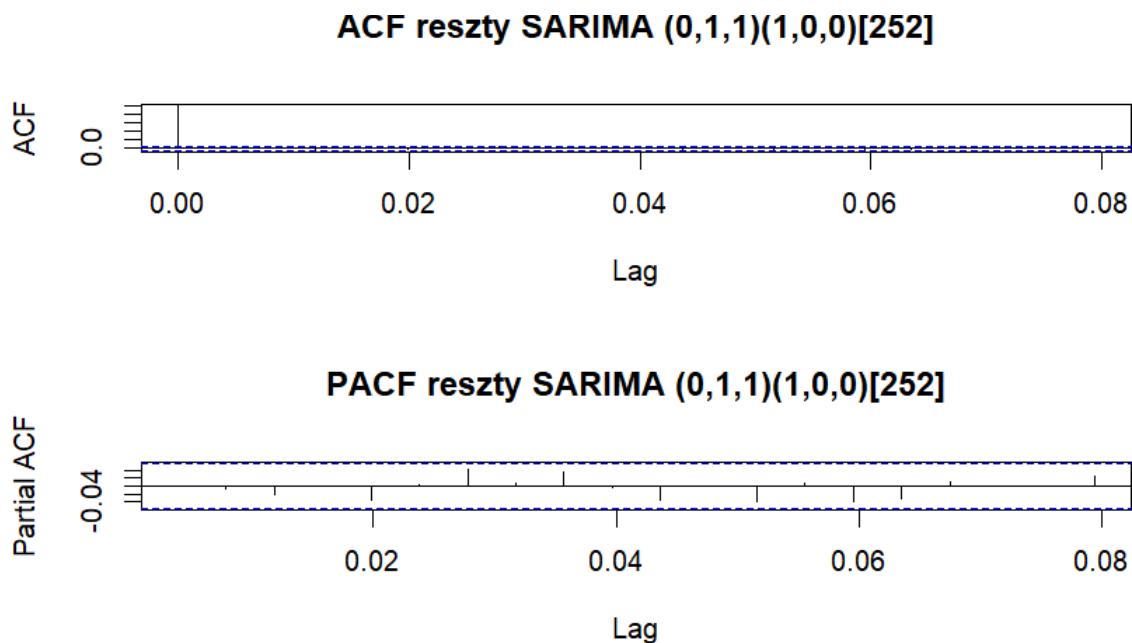
- Nie uwzględnia sezonowości, co może powodować zbyt uproszczoną prognozę.
- Przedział ufności jest nieco szerszy niż w modelu SARIMA, co oznacza większą niepewność przewidywań.
- Model prognozuje mniej dynamicznie, co może powodować gorsze odwzorowanie rzeczywistego ruchu cen.

SARIMA(1,1,1)(0,0,1)[252]

- Model uwzględnia zarówno składnik autoregresyjny AR(1), jak i sezonową średnią ruchomą (SMA(1)).
- Prognoza jest bardziej stabilna niż w ARIMA(0,1,1), ale nie tak dokładna jak SARIMA(0,1,1)(1,0,0)[252].
- Przedziały ufności są podobne do pierwszego modelu, ale prognoza wydaje się mniej dynamiczna.

Biorąc pod uwagę powyższe analizy, modelem najlepiej prognozującym jest SARIMA(0,1,1)(1,0,0)[252], ponieważ uwzględnia sezonowość, co poprawia dopasowanie do danych giełdowych, ma najwęższe pasmo niepewności, co sugeruje, że model jest bardziej pewny swoich przewidywań, a sama prognoza wygląda bardziej dynamicznie i realistycznie, podążając za trendem





Zestawienie wykresów: rzeczywiste wartości cen zamknięcia, prognoza, wykres ACF i PACF dla modelu SARIMA (0,1,1)(1,0,0)[252]

Analizując wykres ACF reszty, można zauważyć brak istotnych wartości autokorelacji dla żadnego opóźnienia- oznacza to, że reszty modelu są białym szumem, czyli model nie pozostawił żadnej strukturalnej zależności w danych. W przypadku wykresu PACF obserwujemy brak istotnych wartości PACF - oznacza to, że model nie pozostawił żadnych dodatkowych zależności czasowych w resztach. Wszystkie wartości leżą w granicach nieistotności, co wskazuje, że model poprawnie przechwycił strukturę danych.

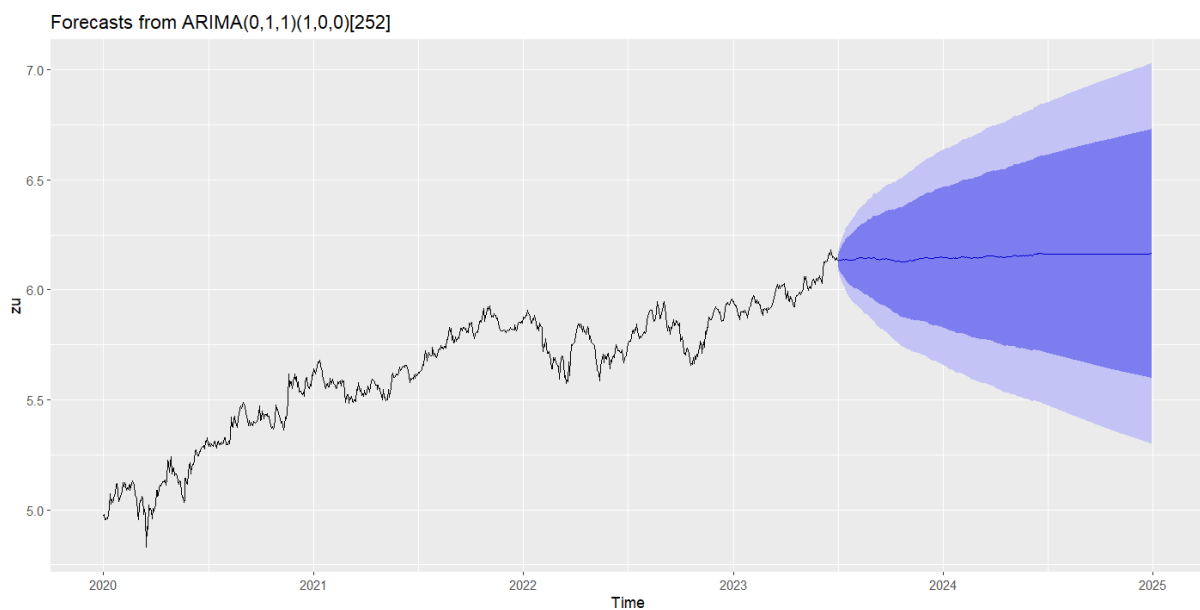
Ostatecznie, przeprowadzono walidację najlepszego modelu, tj. SARIMA (0,1,1)(1,0,0)[252], na zbiorze testowym, aby ocenić skuteczność w prognozowaniu przyszłych cen akcji. Wyniki te będą stanowiły podstawę do oceny, jak model radzi sobie z przewidywaniem zmian cen akcji spółki DINO w kontekście przyszłych okresów.

Tabela 4. Zestawienie parametrów modelu uczącego

Parametr	SARIMA(0,1,1)(1,0,0)[252] model uczący
σ^2	0.0005723
Log-Likelihood	2038.82
AIC	-4071.65

AICc	-4071.62
BIC	-4057.31
Training Set Error Measures	
ME	0.001355057
RMSE	0.02388244
MAE	0.01719666
MPE	0.02353379
MAPE	0.3085517
MASE	0.06358829
ACF1	-0.002447532

Model jest dobrze dopasowany – niskie wartości AIC (-4071.65) i BIC (-4057.31) sugerują wysoką jakość modelu. Błędy prognozy są niskie – RMSE (0.0239) i MAE (0.0172) wskazują na dobrą dokładność przewidywań. Reszty modelu są losowe – ACF1 (-0.0024) bliskie zero oznacza, że model nie pozostawił w danych ukrytych zależności.



Wykres prognozy dla zbioru testowego

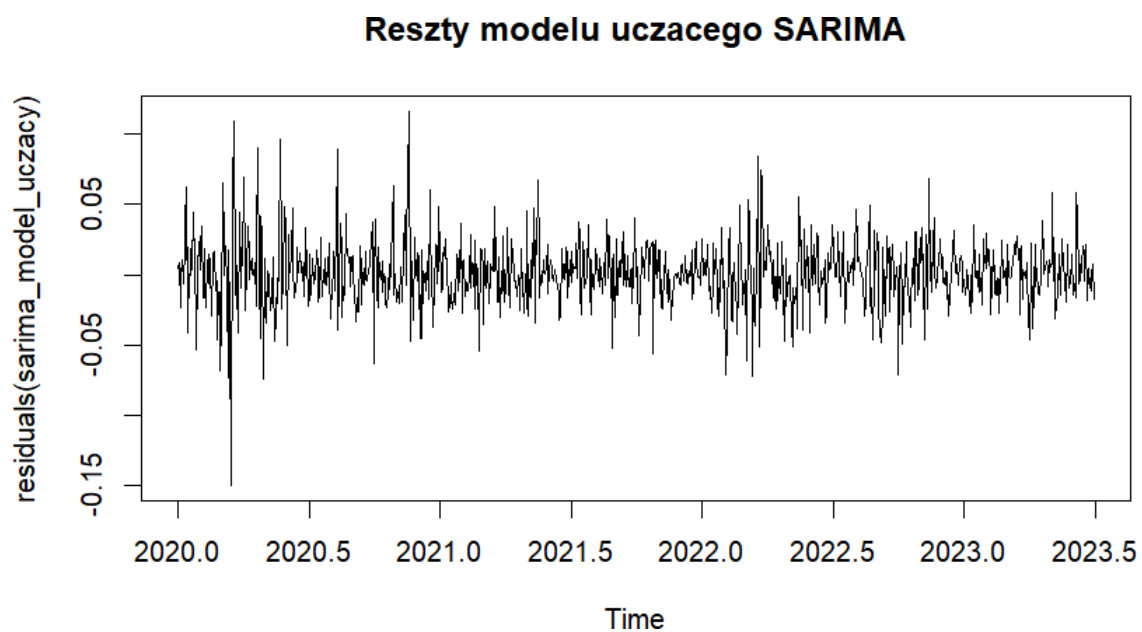
Prognoza kontynuuje trend wzrostowy, zgodny z danymi historycznymi, co sugeruje, że model prawidłowo przechwycił strukturę trendu. Przedziały ufności rosną z czasem, co oznacza, że niepewność prognozy zwiększa się w dalszej przyszłości – jest to typowe dla modeli szeregów czasowych. Środek prognozy wygląda realistycznie, ale szerokie przedziały ufności mogą oznaczać, że w długim okresie model nie daje bardzo precyzyjnych przewidywań.

Następnie przeprowadzono test białego szumu dla reszt.

Tabela 5. Wyniki testu Ljunga-Boxa

	SARIMA(0,1,1)(1,0,0)[252] model uczący
p-value	0.8165525

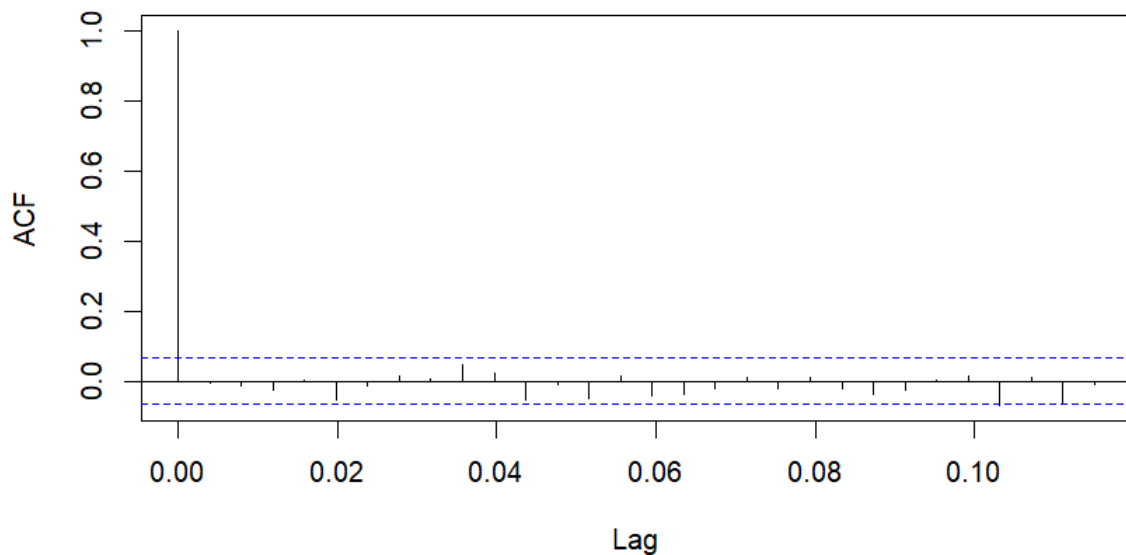
Wynik tego testu wykazał, że model dobrze uchwycił struktury w danych, nie pozostawiając przy tym istotnych wzorców.



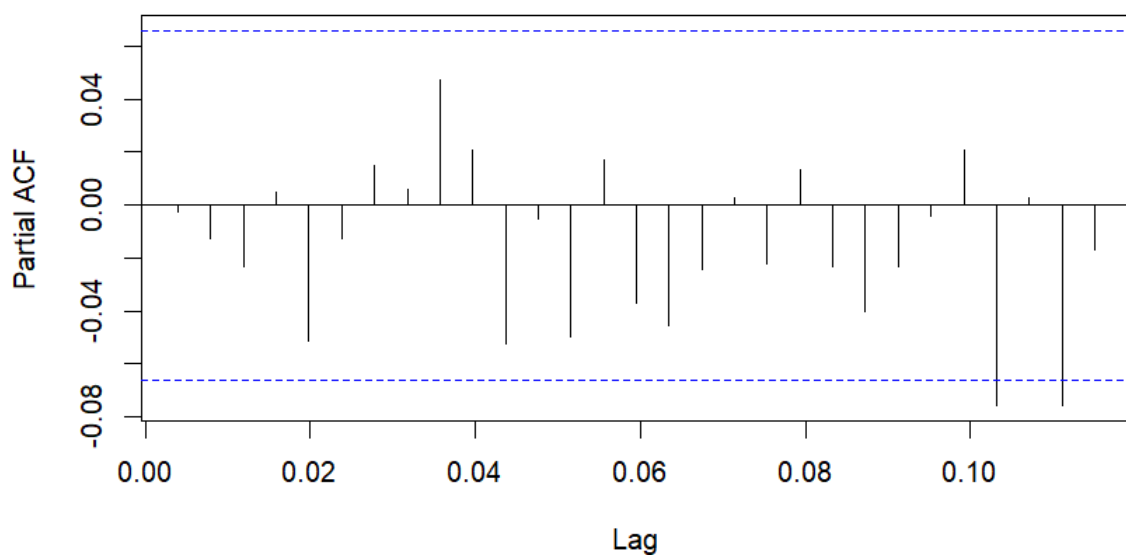
Reszty oscylują wokół zera, co oznacza, że model nie ma tendencji do systematycznego przeszacowywania ani niedoszacowywania wartości. Brak jest wyraźnego trendu w resztach, co sugeruje, że model skutecznie usunął zależności czasowe z danych. Wartości reszt nie są nadmiernie zmienne, co oznacza, że model dobrze dopasował się do danych,

bez skrajnych odchyień. Brak widocznych cyklicznych wzorców – oznacza, że model dobrze uchwycił sezonowość i trend. Jednorodna wariancja reszt (brak wzrostu lub spadku amplitudy w czasie) sugeruje, że model nie ma problemu ze zmienną wariancją (heteroskedastycznością).

ACF reszt



PACF reszt



Wykres ACF pokazuje, że brak jest istotnych wartości lagów, co oznacza, że reszty nie są skorelowane. Model nie pozostawił w danych struktury zależności czasowych, a więc dobrze

przechwycił sezonowość i trend. Brak autokorelacji w resztach potwierdza, że model jest dobrze dopasowany.

Na podstawie wykresu PACF można stwierdzić, że większość wartości mieści się w granicach istotności, co sugeruje losowość reszt. Pojedyncze wartości mogą lekko przekraczać granice, ale nie obserwujemy silnych skoków, model nie pozostawił więc struktury w danych.

Bibliografia

1. Brewer, M. J., Butler, A., Cooksley, S. L. (2016). The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution*, 5th Anniversary Special Feature. Pobrano z <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12541>, dn. 05.02.2025.
2. Chodakowska, E., Nazarko, J. (2022). *Prognozowanie na podstawie modeli ARIMA*. Oficyna Wydawnicza Politechniki Białostockiej. Pobrano z: https://pb.edu.pl/oficyna-wydawnicza/wp-content/uploads/sites/4/2022/07/Chodakowska-Nazarko-Prognozowanie-na-podstawie-modeli-ARIMA.pdf?utm_source=chatgpt.com, dn. 06.02.2025
3. Dino Polska (2025). Profil działalności. Pobrano z <https://grupadino.pl/profil-dzialalnosci/>, dn. 05.02.2025.
4. Osińska, M. (2006). *Ekonometria finansowa*. Polskie Wydawnictwo Ekonomiczne.
5. Sirisha, U. M., Belavagi, M. C., Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA, and LSTM models in time series forecasting: A comparison. Pobrano z <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9964190>, dn. 05.02.2025.
6. Szmuksta-Zawadzka, M., Zawadzki, J. (2012). O miernikach dokładności prognoz ex post w prognozowaniu zmiennych o silnym natężeniu sezonowości. *Metody Ilościowe w Badaniach Ekonomicznych*, 13(1), 212–223.
7. Yahoo Finance. Financial data for DNP.WA. Pobrano z <https://finance.yahoo.com>, dn. 05.02.2025