

Politechnika Gdańska  
Wydział Zarządzania i Ekonomii

Analityka Gospodarcza II, on-line  
Modelowanie Szeregów Czasowych, Semestr zimowy 2024/2025

# Prognozowanie szeregów czasowych

Pisemny Egzamin z Laboratorium  
ARIMA dla CCC Intelligent Solutions Holdings (CCCS)

Aleksandra Jaworska, s202280  
2025-02-06

## Spis treści

Wprowadzenie .....	3
Dane .....	4
Wstępna analiza danych .....	4
Testowanie stacjonarności.....	4
Analiza ACF i PACF .....	5
Budowa modelu ARIMA.....	6
Ocena dopasowania modelu .....	7
Diagnostyka reszt .....	8
Ręczny wybór parametrów ARIMA .....	9
Prognoza .....	10
Podsumowanie .....	11
Badanie dokładności prognozy - aktualizacja 8.02.2025.....	11

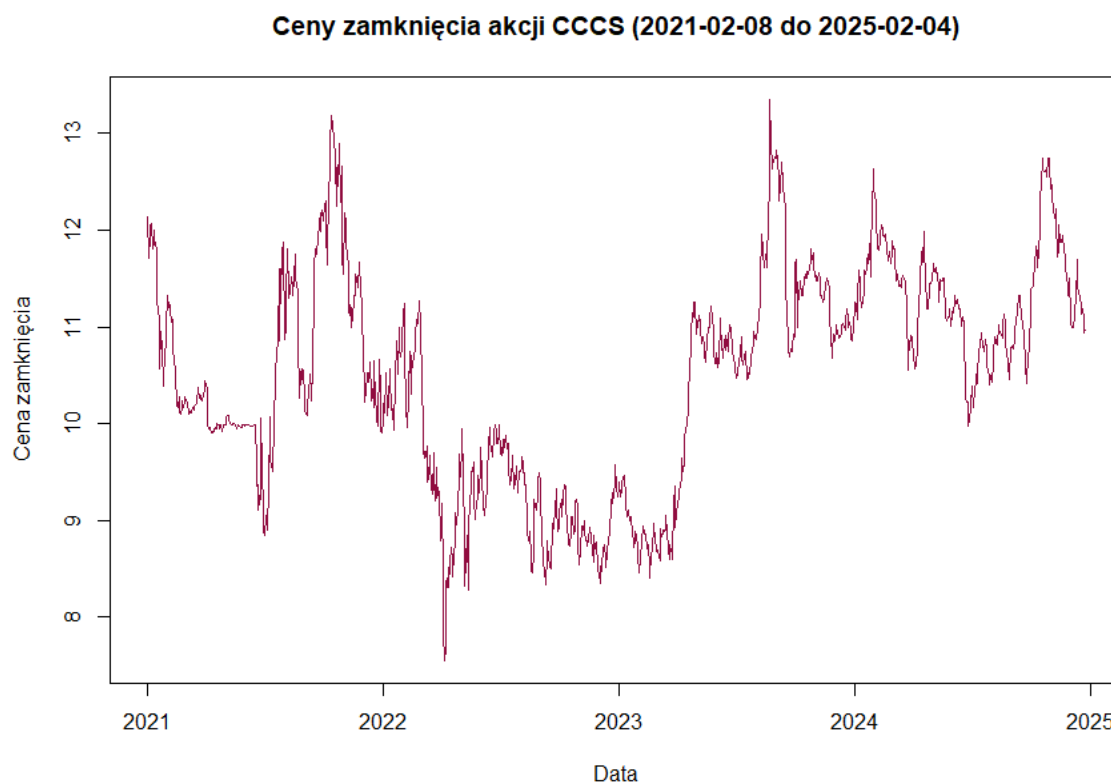
# Wprowadzenie

Głównym celem projektu jest opracowanie skutecznego modelu predykcyjnego, który pozwoli lepiej zrozumieć wzorce cenowe oraz dostarczyć wiarygodne prognozy dla krótkoterminowego horyzontu czasowego. W ramach tego projektu zajmę się prognozowaniem przyszłych cen zamknięcia akcji **CCC Intelligent Solutions Holdings** na dni **5, 6 i 7 lutego** przy zastosowaniu modelu ARIMA (Autoregressive Integrated Moving Average) – jednego z najpopularniejszych i najbardziej wszechstronnych narzędzi do analizy i prognozowania szeregów czasowych. Dzięki swojej zdolności do modelowania zarówno komponentów autoregresyjnych (AR), jak i średniej ruchomej (MA), a także uwzględnianiu różnicowania (I), ARIMA pozwala na dokładne odwzorowanie i przewidywanie skomplikowanych wzorców danych. Proces ten obejmuje kilka kluczowych etapów:

- **Zebranie i przygotowanie danych** – Pobranie historycznych cen zamknięcia akcji CCC Intelligent Solutions Holdings i przygotowanie ich do analizy.
- **Analiza wstępna i testowanie stacjonarności** – Sprawdzenie, czy dane są stacjonarne, co jest kluczowe dla wyboru odpowiedniego modelu.
- **Analiza autokorelacji (ACF) i częściowej autokorelacji (PACF)** – Identyfikacja zależności między wartościami szeregu czasowego w celu określenia potencjalnych parametrów modelu.
- **Budowa modelu ARIMA** – Dobór optymalnych parametrów modelu ARIMA za pomocą automatycznych narzędzi, które minimalizują błędy prognozy.
- **Ocena jakości modelu** – Weryfikacja poprawności modelu poprzez analizę reszt, stabilności oraz kryteriów dopasowania.
- **Prognozowanie** – Wykorzystanie modelu ARIMA do przewidywania przyszłych cen akcji oraz ocena dokładności prognozy.

## Dane

Analizowane dane dotyczą CCC Intelligent Solutions Holdings, spółki działającej w branży technologii ubezpieczeniowych i motoryzacyjnych. Dane obejmują dzienne ceny zamknięcia akcji w okresie od 8 lutego 2021 roku (poniedziałek) do 4 lutego 2025 roku, co daje 1003 okresy.



## Wstępna analiza danych

### Testowanie stacjonarności

Aby określić, czy szereg jest stacjonarny, wykonano test Augmented Dickey-Fuller (ADF) oraz Phillips-Perron Unit Root Test (PP)

*H0: Szereg czasowy ma pierwiastek jednostkowy (jest niestacjonarny).*

*H1: Szereg czasowy nie ma pierwiastka jednostkowego (jest stacjonarny).*

#### Augmented Dickey-Fuller Test

```
data: CCCS_ts  
Dickey-Fuller = -3.1599, Lag order = 10, p-value = 0.09484  
alternative hypothesis: stationary
```

### Phillips-Perron Unit Root Test

data: CCCS\_ts

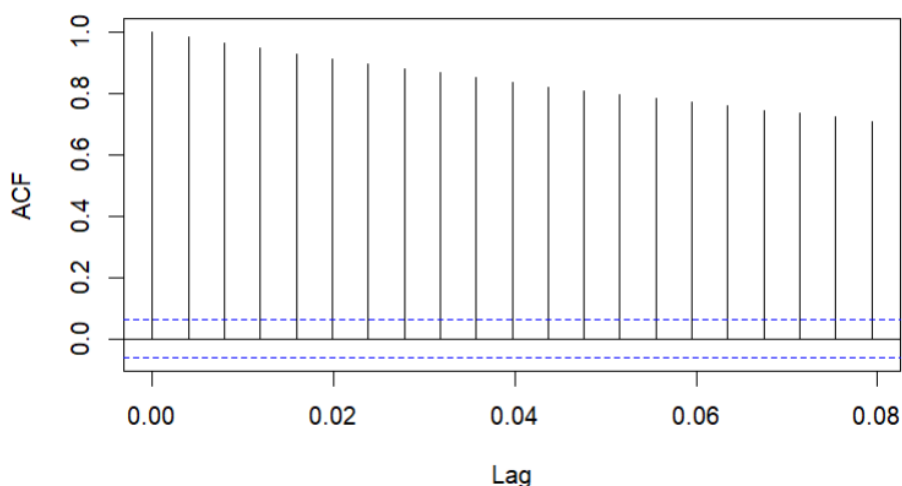
Dickey-Fuller  $Z(\alpha) = -21.259$ , Truncation lag parameter = 7, p-value = 0.05356  
alternative hypothesis: stationary

#### Interpretacja:

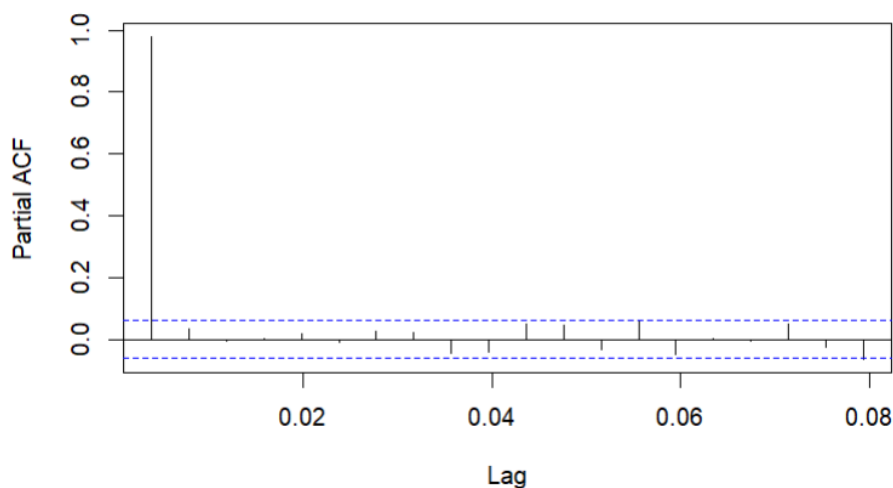
Oba z przeprowadzonych testów badają obecność jednostkowego pierwiastka (unit root) w szeregu czasowym. Dla analizowanych danych wartość p-value obu testów przekroczyła zadany poziom istotności 0.05 (ADF p-value = **0.09486**; PP p-value = **0.05356**). Oznacza to, iż nie ma podstaw do odrzucenia hipotezy zerowej, zatem możemy stwierdzić, iż szereg czasowy jest niestacjonarny. W tym momencie można zastanowić się nad zastosowaniem ręcznego różnicowania. Na ten moment jednak skupiono się na dalszej analizie danych.

#### Analiza ACF i PACF

##### ACF dla CCCS



##### PACF dla CCCS



### Interpretacja:

ACF (autokorelacja) wykazuje powolny spadek, co sugeruje obecność składnika MA (średniej ruchomej), natomiast PACF (częściowa autokorelacja) wykazuje znaczącą wartość tylko dla pierwszego opóźnienia. Na tej podstawie można wywnioskować, że szereg czasowy ma zarówno komponent AR, jak i MA, można byłoby zatem wykorzystać model ARMA(p, q). Należy jednak pamiętać, iż badany szereg czasowy **nie jest stacjonarny**. W związku z tym należy albo ręcznie dane zróżnicować za pomocą funkcji `diff()` w RStudio, albo przyjąć model ARIMA (AutoRegressive Integrated Moving Average), będący rozszerzeniem modelu ARMA, które uwzględnia różnicowanie. W tym projekcie zdecydowano się na implementację drugiego podejścia.

## Budowa modelu ARIMA

Do budowy modelu ARIMA postanowiono użyć funkcji `auto.arima()`, przeszukującą różne kombinacje parametrów p (AR), d (różnicowanie) i q (MA), w celu znalezienia modelu minimalizującego kryterium informacyjne (np. AIC lub BIC).

```
Series: stock_prices
ARIMA(1,1,0)

Coefficients:
      ar1
    -0.0326
s.e.    0.0316

sigma^2 = 0.04585:  log likelihood = 123.02
AIC=-242.05   AICc=-242.04   BIC=-232.23

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -0.001203145  0.2139069  0.1488948 -0.03247355  1.437378  0.9994143 -0.0003919161
```

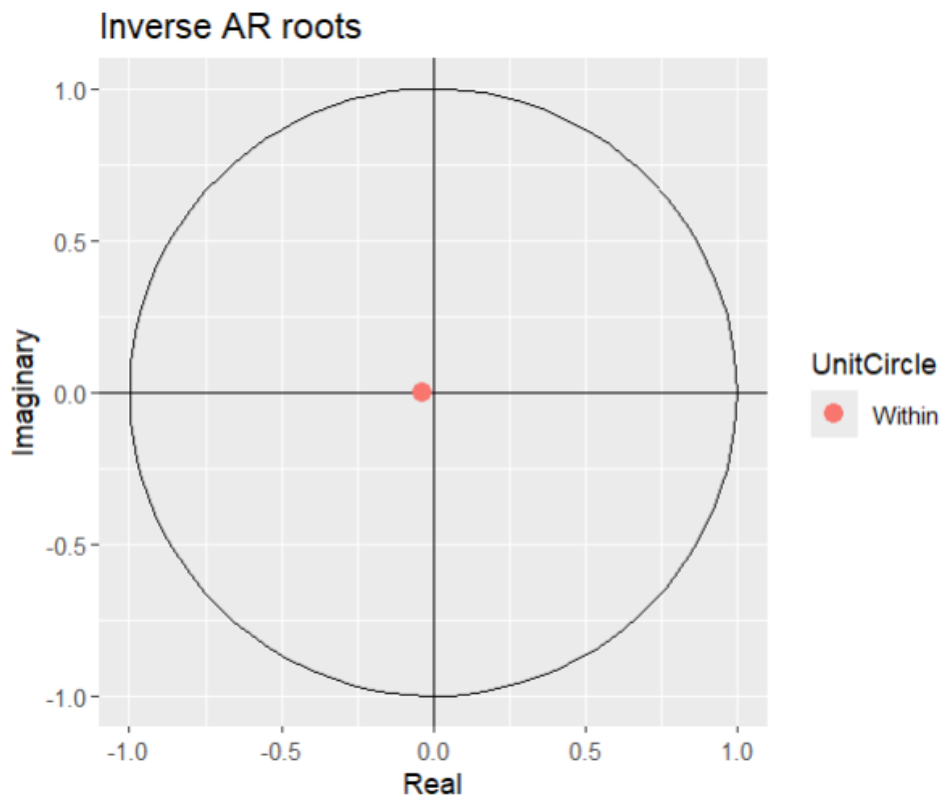
Model wybrany przez funkcję `auto.arima()` to **ARIMA(1,1,0)**, co oznacza:

- AR(1) – obecna wartość zależy od wartości z poprzedniego okresu, choć współczynnik AR (-0.0326) jest niewielki. Wzrost wartości w poprzednim okresie wpływa na **spadek** wartości w bieżącym okresie.
- I(1) – pierwsze różnicowanie zostało zastosowane do osiągnięcia stacjonarności;
- MA(0) – nie wykryto istotnych efektów średniej ruchomej

## Ocena dopasowania modelu

- $\sigma^2$  (wariancja reszt modelu) wyniosła 0.04585, jest to stosunkowo niska wartość, co sugeruje, że model dobrze opisuje dane;
- **AIC (Akaike Information Criterion)** = -242.05, niska wartość, co wskazuje na dobre dopasowanie;
- **BIC (Bayesian Information Criterion)** = -232.23, podobnie jak AIC również wskazuje na dobre dopasowanie;
- Niskie wartości **RMSE (Root Mean Squared Error)** = 0.2139 oraz **MAPE (Mean Absolute Percentage Error)** = 1.437% wskazują na niewielkie błędy prognozy.

W kolejnym kroku przeanalizowano poniższy wykres **Inverse AR roots** używany do sprawdzenia stabilności modelu.



### Interpretacja:

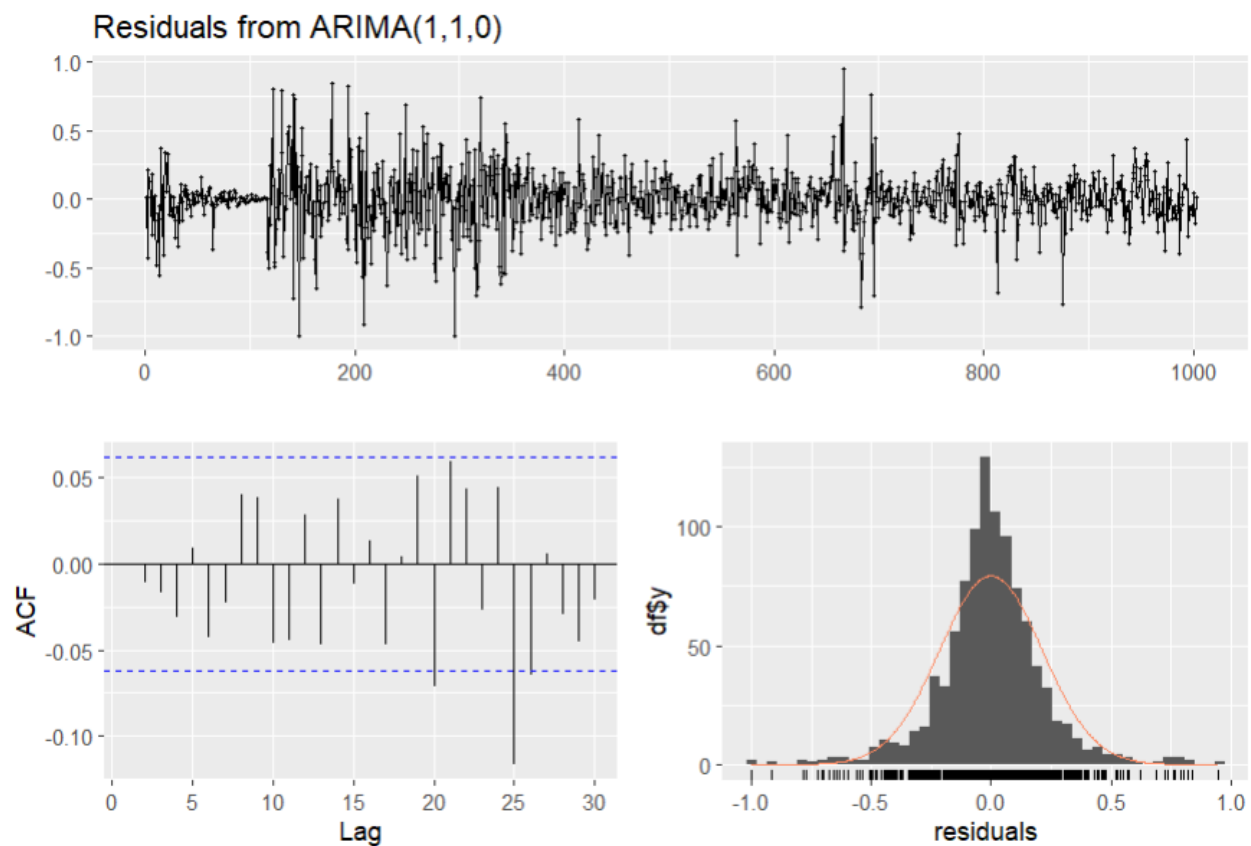
Na powyższym wykresie można zauważyć, że wszystkie odwrotne pierwiastki AR leżą wewnątrz okręgu. Oznacza to, że model ARIMA jest stabilny i można go użyć w prognozowaniu. Z praktycznego punktu widzenia stabilność modelu gwarantuje, że prognozy będą zbieżne i nie będą wykazywać nierealistycznych oscylacji.

Dopełnieniem powyższego wykresu byłoby sprawdzenie odwracalności komponentu MA oraz sezonowego komponentu MA. W związku z tym, że ARIMA wykazała brak istotnych efektów średniej ruchomej, przeprowadzenie tego testu jest niepotrzebne. Możemy jednak sprawdzić stabilność komponentu AR:

AR Inverse Roots:  
0.03260268  
The AR component is stationary.

Odwrotne pierwiastki AR wynoszą **0.03260268**, co jest mniejsze niż 1. Oznacza to, że komponent AR jest stacjonarny.

## Diagnostyka reszt



## Wykres reszt

Reszty na powyższym wykresie oscylują wokół zera. Występują pewne odchylenia jednak nie zauważono wyraźnego trendu/sezonowości - sugeruje to, że model dobrze pasuje do danych. Reszty również nie maleją/rosną w czasie, a zatem możemy stwierdzić, że ich wariancja jest stała (odrzucaamy hipotezę o heteroskedastyczności).



### Wykres ACF reszt

Wszystkie wartości ACF mieszczą się w granicach ufności, co sugeruje, że reszty są białym szumem, a zatem nie zawierają one żadnej systematycznej struktury ani wzorców, które mogłyby zostać wykorzystane do poprawy modelu.

Na potwierdzenie, że reszty są białym szumem przeprowadzono test **Ljung-Boxa**.

*H0: Reszty są białym szumem.*

*H1: Reszty nie są białym szumem.*

Ljung-Box test

data: Residuals from ARIMA(1,1,0)  
Q\* = 9.1119, df = 9, p-value = 0.427

Model df: 1. Total lags used: 10

P-value przyjęło wartość 0.427 (> niż 0.05), więc nie ma podstaw do odrzucenia hipotezy zerowej. Test udowodnił, że reszty są białym szumem.

### Histogram reszt

Rozkład reszt jest zbliżony do normalnego, jest to widoczne dzięki czerwonej linii gęstości przypominającej krzywą Gaussa - oznacza to, że model dobrze pasuje do danych i żadne problemy z modelem nie zostały zidentyfikowane. Na wykresie zaobserwowano również lekko wydłużone ogony sugerujące istnienie kilku ekstremalnych wartości.

### Ręczny wybór parametrów ARIMA

W ramach projektu zbadano również czy funkcja `auto.arima()` rzeczywiście wybiera najlepsze z możliwych parametrów modelu. musimy porównać wartości AIC (Akaike Information Criterion) i BIC (Bayesian Information Criterion). Niższe wartości obu kryteriów wskazują na lepsze dopasowanie modelu do danych, przy czym BIC dodatkowo uwzględnia złożoność modelu, preferując prostsze modele.

#### AIC (Akaike Information Criterion)

- **ARIMA (1, 1, 0) = -242.0479** ← parametry ustalone przez funkcję `auto.arima()`
- ARIMA(0, 1, 2) = -240.1906
- ARIMA(2, 1, 0) = -240.1594
- ARIMA(1, 1, 1) = -241.5413
- ARIMA(1, 1, 2) = -238.1983

Najniższą wartość dla kryterium AIC ma **ARIMA (1, 1, 0)**

### BIC (Bayesian Information Criterion)

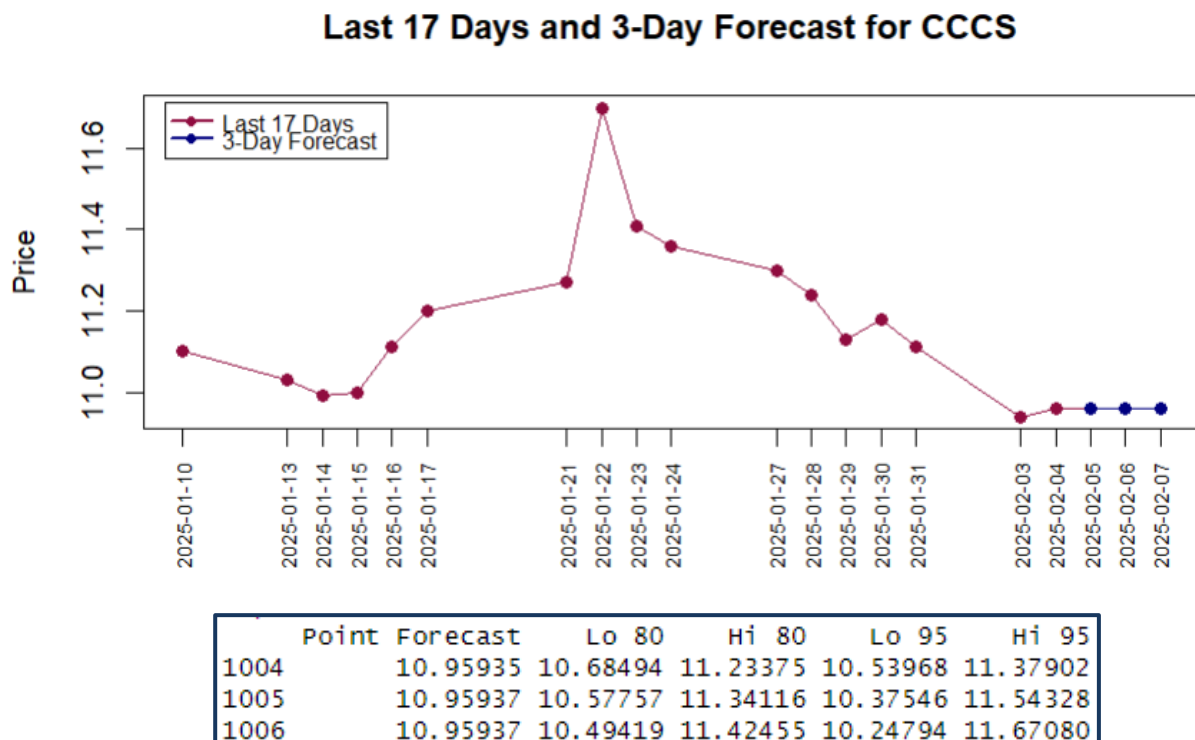
- **ARIMA (1, 1, 0) = -232.2284** ← parametry ustalone przez funkcję *auto.arima()*
- ARIMA(0, 1, 2) = -225.4614
- ARIMA(2, 1, 0) = -225.4301
- ARIMA(1, 1, 1) = -226.812
- ARIMA(1, 1, 2) = -218.5593

Najniższą wartość dla kryterium BIC ma **ARIMA (1, 1, 0)**

**Najlepszy model** według obu kryteriów (AIC i BIC) to **ARIMA(1, 1, 0)**. Jest to również model wybrany przez funkcję *auto.arima()*, co potwierdza jego trafność.

## Prognoza

Zgodnie z wytycznymi projektu wykonano prognozę dla cen akcji wybranej spółki na dni 4, 5 oraz 6 lutego. Prognoza została oparta na modelu ARIMA (1, 1, 0), którego kryterium wyboru była powyżej przeprowadzona analiza.



Analizując przedziały ufności możemy stwierdzić, że:

- Z 80% prawdopodobieństwem cena akcji będzie się poruszać w zakresie 10.68494 - 10.49419 (dolna granica); 11.23375 - 11.42455 (górna granica)
- Z 90% prawdopodobieństwem cena akcji będzie się poruszać w zakresie 10.53968 - 10.24794 (dolna granica); 11.37902 - 11.67080 (górna granica)

Na analizowany okres model prognozuje stabilizację ceny CCCS Holdings w przedziale **10.95935 - 10.95937**. Oznacza to, że w najbliższych trzech dniach nie przewiduje się istotnych zmian wartości akcji, a kurs powinien oscylować wokół tego poziomu.

## Podsumowanie

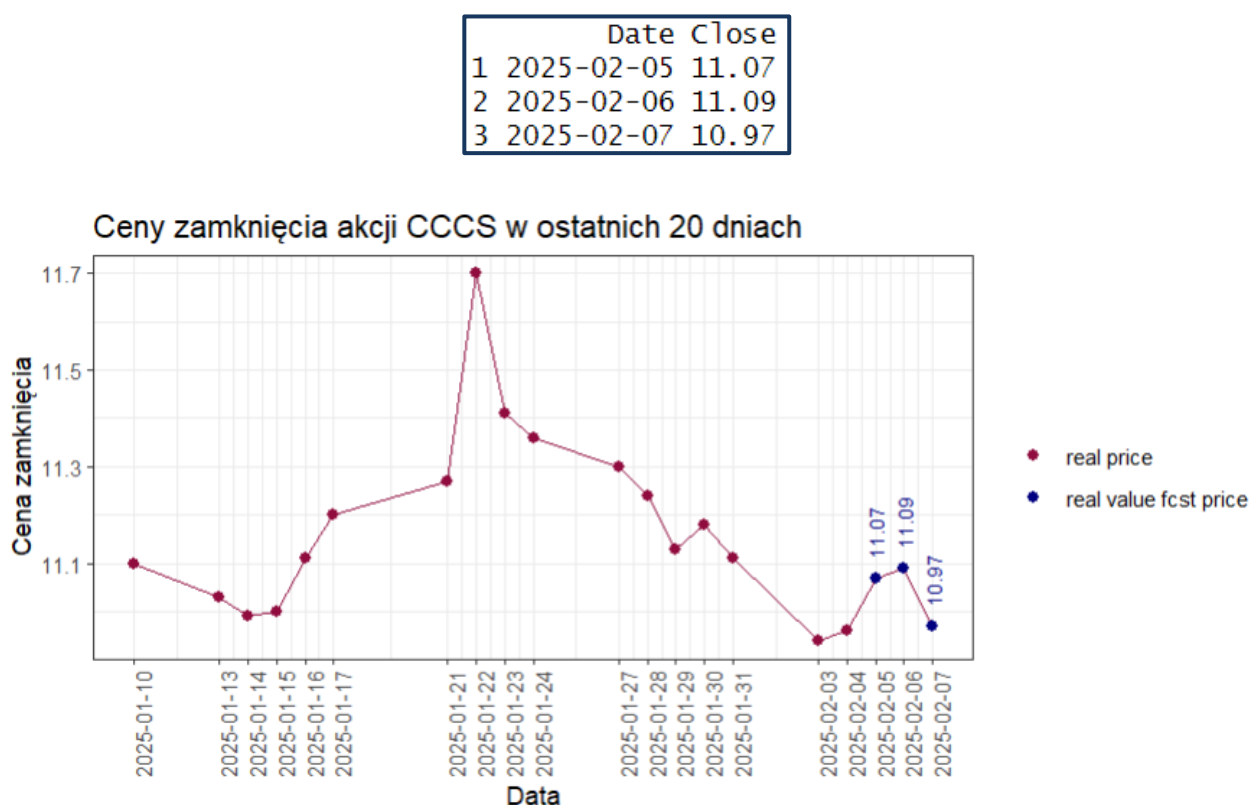
W ramach projektu dokonano predykcji cen akcji CCC Intelligent Solutions Holdings na dni 5, 6 i 7 lutego, wykorzystując model ARIMA(1,1,0). Analiza danych wykazała, że szereg czasowy był niestacjonarny, co uzasadniało zastosowanie różnicowania. Model został dobrany na podstawie funkcji `auto.arima()`, a jego skuteczność potwierdzono niskimi wartościami kryteriów AIC i BIC oraz dobrym dopasowaniem reszt do rozkładu normalnego.

Ocena jakości modelu wykazała niskie wartości błędów predykcji (RMSE = 0.2139, MAPE = 1.437%), stabilność modelu oraz brak istotnych efektów średniej ruchomej. Testy statystyczne (ADF, PP, Ljung-Boxa) potwierdziły poprawność i trafność modelu.

Wyniki prognozy wskazują na stabilizację ceny akcji w przedziale 10.95935 - 10.95937 w analizowanym okresie. Przedziały ufności sugerują, że przy 80% i 90% prawdopodobieństwie kurs nie powinien znacząco odbiegać od tych wartości, co oznacza brak oczekiwanych gwałtownych zmian cenowych. Model ARIMA(1,1,0) okazał się skutecznym narzędziem do krótkoterminowej prognozy cen akcji, dostarczając wiarygodnych wyników.

## Badanie dokładności prognozy – aktualizacja 8.02.2025

W ostatnim kroku postanowiono porównać otrzymane prognozy z rzeczywistymi wartościami zamknięcia akcji CCCS. W tym celu dnia 8.02.2025 r. do wcześniej wygenerowanego kodu podstawiono wartości rzeczywiste dni prognozowanych (5, 6, 7 luty) oraz zwizualizowano je na wykresie.



Wartości rzeczywiste postanowiono zestawzić z wygenerowanymi prognozami oraz przedziałami ufności:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1004	10.95935	10.68494	11.23375	10.53968	11.37902
1005	10.95937	10.57757	11.34116	10.37546	11.54328
1006	10.95937	10.49419	11.42455	10.24794	11.67080

Wartości rzeczywiste prognozy nie były dokładne w 100% (sytuacja pożądana, aczkolwiek mało prawdopodobna), jednak dla każdego z prognozowanych dni mieściły się w przedziałach ufności, zarówno na poziomie 80% jak i 95%. Dodatkowo obliczono MAPE oraz MAE, których niskie wartości potwierdziły dokładność prognozy

```
> cat("Mean Absolute Error (MAE):", round(mae, 4), "\n")
Mean Absolute Error (MAE): 0.1865
> cat("Mean Absolute Percentage Error (MAPE):", round(mape, 2), "%\n")
Mean Absolute Percentage Error (MAPE): 1.72 %
```