

IAD Szeregi Czasowe - Projekt Końcowy

Damian Spodar & Tomasz Hanusek gc1

Analiza Szeregów Czasowych dla cen akcji Chevron oraz Valero Energy

Cel Projektu

Celem projektu jest analiza szeregów czasowych cen akcji spółek energetycznych Chevron (CVX) oraz Valero Energy (VLO). Analiza pozwala zrozumieć zachowanie cen w okresie od początku 2022 roku, zweryfikować charakter trendu (deterministyczny vs. stochastyczny) oraz porównać dynamikę dwóch dużych spółek sektora energetycznego. Kluczowym elementem jest budowa modelu progностycznego klasy ARIMA poprzedzona weryfikacją stacjonarności.

Krótki opis spółek

Chevron (CVX): Amerykańska spółka z branży energetycznej, zajmująca się wydobyciem, rafinacją i sprzedażą ropy naftowej oraz gazu ziemnego. Chevron należy do największych producentów energii na świecie, a jego akcje są notowane na giełdzie NYSE. Firma charakteryzuje się stabilną obecnością w sektorze upstream i downstream oraz znaczącym udziałem w globalnym rynku ropy.

Valero Energy (VLO): Amerykańska spółka rafineryjna i dystrybutor paliw, koncentrująca się na przetwarzaniu ropy naftowej i produkcji paliw, w tym benzyny, oleju napędowego i biopaliw. Valero posiada rafinerie w Stanach Zjednoczonych i Kanadzie, a jej akcje są również notowane na giełdzie NYSE. Firma jest jednym z największych niezależnych producentów paliw na świecie i charakteryzuje się dużą zmiennością przychodów ze względu na wahania cen ropy i popytu na paliwa.

Użyte biblioteki

```
library(quantmod)
library(forecast)
library(tseries)
library(lmtest)
library(FinTS)
library(rugarch)
```

Chevron

Pobranie danych

Dane cen akcji Chevron (CVX) pobrano z serwisu Yahoo Finance w okresie od 01.01.2022 do 20.01.2026. Do analizy wykorzystano ceny zamknięcia (Close), które są standardem w badaniach finansowych i dobrze odzwierciedlają codzienne wahania wartości akcji. Dodatkowo obliczono logarytmiczne stopy zwrotu, które ułatwiają modelowanie zmienności i stosowanie modeli ARIMA.

```
[1] "CVX"
```

Analiza trendu - wybór stopnia wielomianu przez AIC

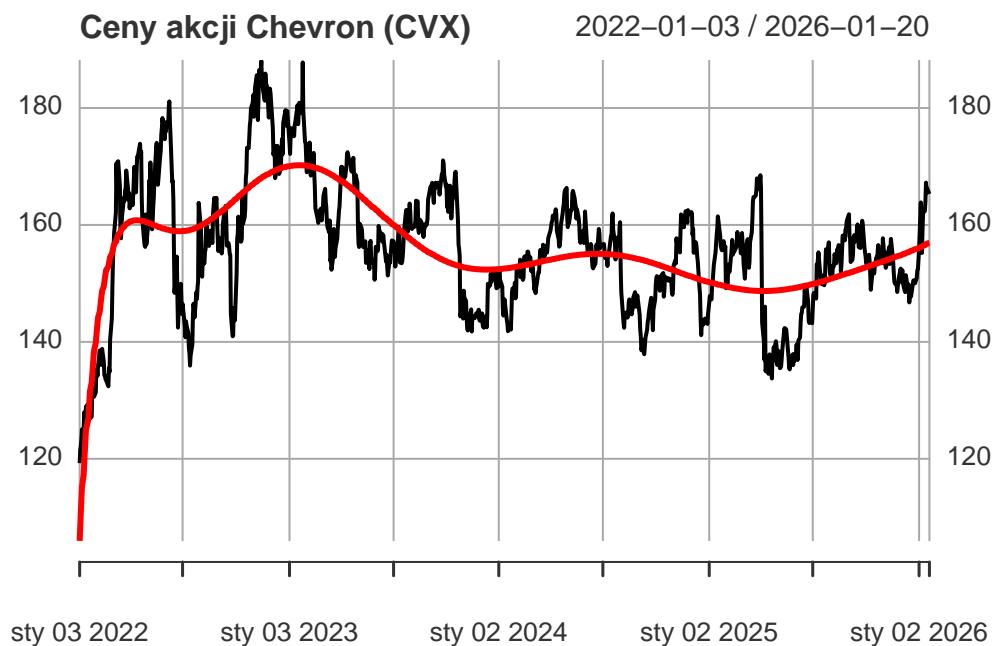
W pierwszym kroku podjęto próbę opisu szeregu za pomocą deterministycznego modelu trendu wielomianowego. Do wyboru najlepszego modelu wykorzystano kryterium informacyjne Akaike (AIC).

Uzyskane wartości AIC dla kolejnych stopni wielomianu były następujące:

Wartości AIC dla kolejnych stopni wielomianu:

```
[1] 7737.497 7714.970 7465.121 7381.144 7336.844 7338.844 7339.180 7254.783  
[9] 7182.055 7179.095
```

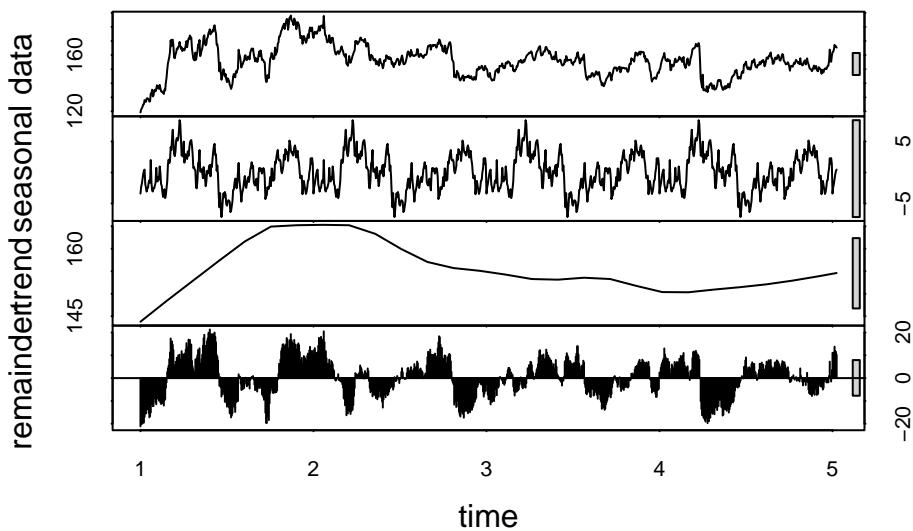
Najlepszy stopień wielomianu wg AIC: 10



Wnioski i krytyka podejścia: Najniższą wartość AIC uzyskano dla wielomianu 10. stopnia. Choć matematycznie model ten najlepiej dopasowuje się do danych historycznych, w kontekście finansowym wskazuje to na zjawisko “przeuczenia” (overfitting). Ceny akcji zazwyczaj charakteryzują się trendem stochastycznym (błądzenie losowe), a nie deterministycznym. Wysoki stopień wielomianu próbuje na siłę opisać losowe fluktuacje jako stałe prawo, co zweryfikujemy badając reszty tego modelu.

Analiza sezonowości

Sezonowość została zbadana za pomocą dekompozycji:

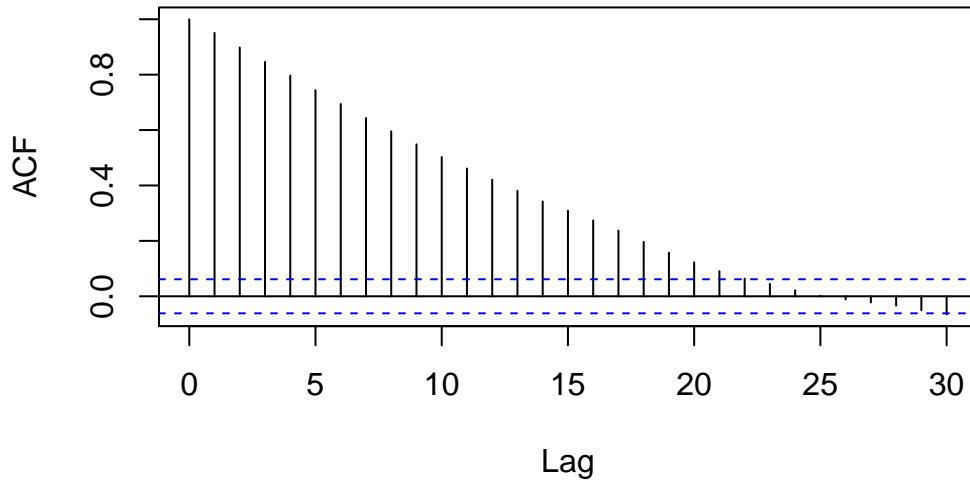


Analiza wykresów dekompozycji wskazuje, że dominującą strukturę szeregu stanowi trend, natomiast składnik sezonowy jest stosunkowo niewielki. Oznacza to, że w analizowanym okresie ceny akcji Chevron nie wykazują wyraźnej powtarzalnej sezonowości, co jest charakterystyczne dla dziennych danych giełdowych.

Analiza reszt modelu trendu

Reszty najlepszego modelu trendu zostały poddane testowi Ljunga–Boxa, który sprawdza, czy istnieje autokorelacja reszt:

ACF reszt modelu wielomianowego



Box-Ljung test

```
data: residuals_trend
X-squared = 6521.6, df = 20, p-value < 2.2e-16
```

Test Ljunga–Boxa dla reszt modelu trendu wykazuje $p\text{-value} < 2.2 \times 10^{-16}$. Odrzucamy hipotezę zerową o braku autokorelacji.

Wniosek: Reszty nie są białym szumem. Oznacza to, że model deterministyczny (wielomian) nie wyjaśnia dynamiki szeregu. Szereg prawdopodobnie jest niestacjonarny i wymaga podejścia stochastycznego (ARIMA).

Badanie stacjonarności (Test ADF)

Przed zastosowaniem modeli ARIMA, należy formalnie zbadać stacjonarność szeregu za pomocą Rozszerzonego Testu Dickeya-Fullera (ADF).

Hipotezy:

H_0 : Szereg jest niestacjonarny (posiada pierwiastek jednostkowy).

H_1 : Szereg jest stacjonarny.

```
Warning in adf.test(prices): p-value smaller than printed p-value
```

Augmented Dickey-Fuller Test

```
data: prices
Dickey-Fuller = -4.5617, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

```
Warning in adf.test(returns): p-value smaller than printed p-value
```

```
Augmented Dickey-Fuller Test
```

```
data: returns
Dickey-Fuller = -9.9862, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

Interpretacja:

1. Dla cen surowych $p\text{-value} \leq 0.01$, co pozwala przyjąć H_1 . Szereg jest stacjonarny.
2. Dla log-stóp zwrotu (pierwsze różniczki) $p\text{-value} \leq 0.01$, co pozwala przyjąć H_1 .

Konkluzja: Szereg jest zintegrowany w stopniu pierwszym ($I(1)$), co uzasadnia przyjęcie parametru $d=1$ w modelu ARIMA.

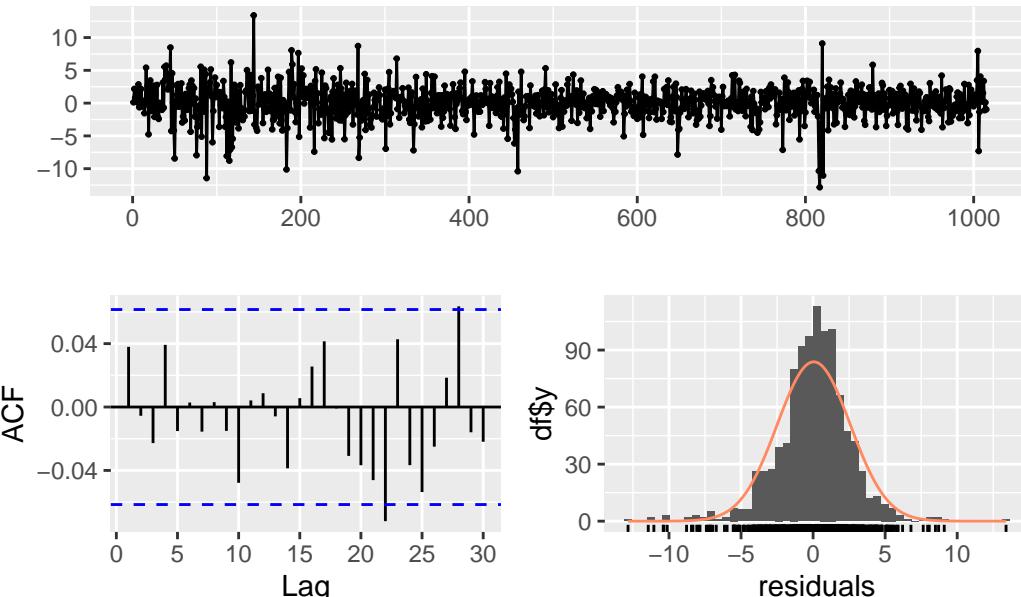
Modelowanie stochastyczne - ARIMA

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

sigma^2 = 6.428: log likelihood = -2382.16
AIC=4766.33   AICc=4766.33   BIC=4771.25

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.0454771 2.534113 1.836728 0.01912789 1.177037 0.9990787
          ACF1
Training set 0.03795515
```

Residuals from ARIMA(0,1,0)



Ljung-Box test

```
data: Residuals from ARIMA(0,1,0)
Q* = 6.6771, df = 10, p-value = 0.7555

Model df: 0.  Total lags used: 10
```

Dopasowano model **ARIMA(0,1,0)**, który najlepiej opisuje strukturę szeregu po uwzględnieniu trendu.

Model ARIMA(0,1,0) to tzw. Błądzenie Losowe (Random Walk). Można go zapisać równaniem: $Y_t = Y_{t-1} + \epsilon_t$, gdzie ϵ_t to biały szum. Oznacza to, że najlepszą prognozą ceny na jutro jest cena dzisiejsza, a zmiany cen są nieprzewidywalne. Jest to zgodne z Hipotezą Rynku Efektywnego (EMH) w jej słabej formie.

Jakość dopasowania oceniono za pomocą błędów prognozy (RMSE = 2.535, MAE = 1.838) oraz testu Ljunga–Boxa na resztach modelu:

```
Q* = 6.6771, df = 10, p-value = 0.7555
```

Wysoka wartość p-value w teście Ljunga–Boxa oznacza, że reszty modelu ARIMA nie wykazują istotnej autokorelacji, co świadczy o poprawnym dopasowaniu modelu. Model ten skutecznie uchwycił zależności czasowe w średniej szeregu, umożliwiając dalszą analizę zmienności i potencjalne prognozy cen.

Podsumowanie:

Ceny akcji Chevron w okresie od początku 2022 roku wykazały silny, nieliniowy trend, najlepiej opisany przez wielomian 10. stopnia. Reszty modelu trendu wykazywały istnienie autokorelacji, co uzasadniało zastosowanie modelu ARIMA(0,1,0). Dopasowany model dobrze uchwycił zależności czasowe, a reszty nie wykazywały istotnej autokorelacji, co potwierdził test Ljunga–Boxa.

Valero Energy

Pobranie danych

Dane cen akcji Valero Energy (VLO) pobrano z serwisu Yahoo Finance w okresie od 01.01.2022 do 20.01.2026. Do analizy wykorzystano ceny zamknięcia (Close), które są standardem w badaniach finansowych i dobrze odzwierciedlają codzienne wahania wartości akcji. Dodatkowo obliczono logarytmiczne stopy zwrotu, które ułatwiają modelowanie zmienności i stosowanie modeli ARIMA.

```
[1] "VLO"
```

Analiza trendu - wybór stopnia wielomianu przez AIC

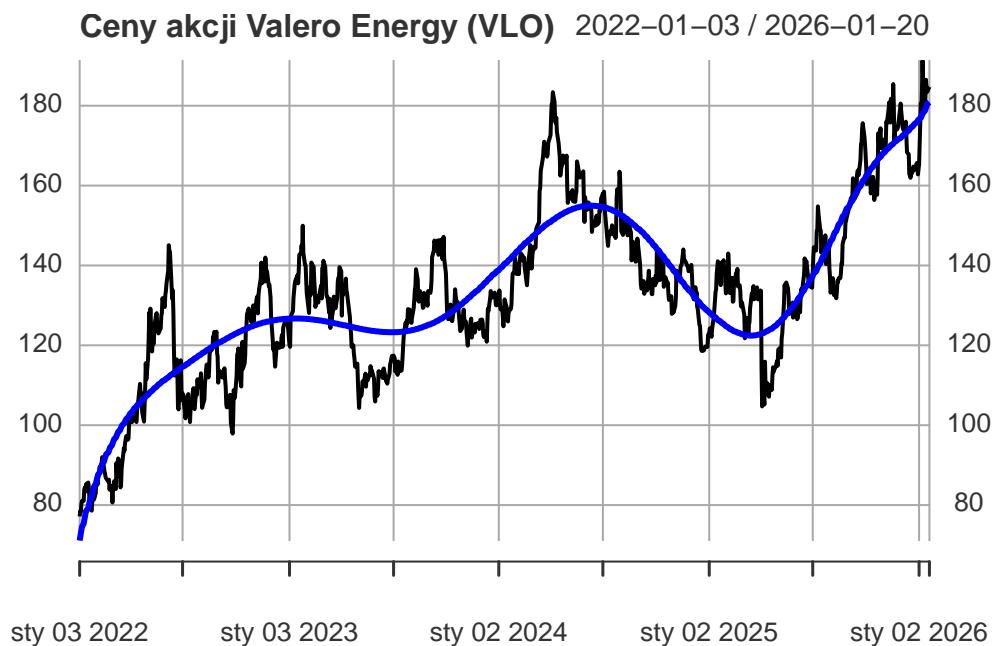
W pierwszym kroku podjęto próbę opisu szeregu za pomocą deterministycznego modelu trendu wielomianowego. Do wyboru najlepszego modelu wykorzystano kryterium informacyjne Akaike (AIC).

Uzyskane wartości AIC dla kolejnych stopni wielomianu były następujące:

Wartości AIC dla kolejnych stopni wielomianu:

```
[1] 8467.150 8433.556 8225.631 8147.319 7885.332 7837.302 7599.763 7575.726  
[9] 7516.727 7509.159
```

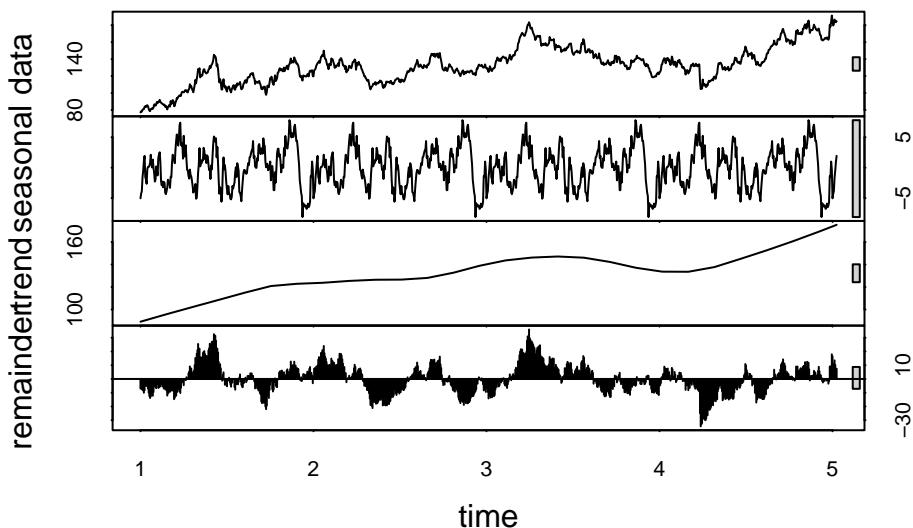
Najlepszy stopień wielomianu wg AIC: 10



Wnioski i krytyka podejścia: Najniższą wartość AIC uzyskano dla wielomianu 10. stopnia. Choć matematycznie model ten najlepiej dopasowuje się do danych historycznych, w kontekście finansowym wskazuje to na zjawisko “przeuczenia” (overfitting). Ceny akcji zazwyczaj charakteryzują się trendem stochastycznym (błądzenie losowe), a nie deterministycznym. Wysoki stopień wielomianu próbuje na siłę opisać losowe fluktuacje jako stałe prawo, co zweryfikujemy badając reszty tego modelu.

Analiza sezonowości

Sezonowość została zbadana za pomocą dekompozycji:

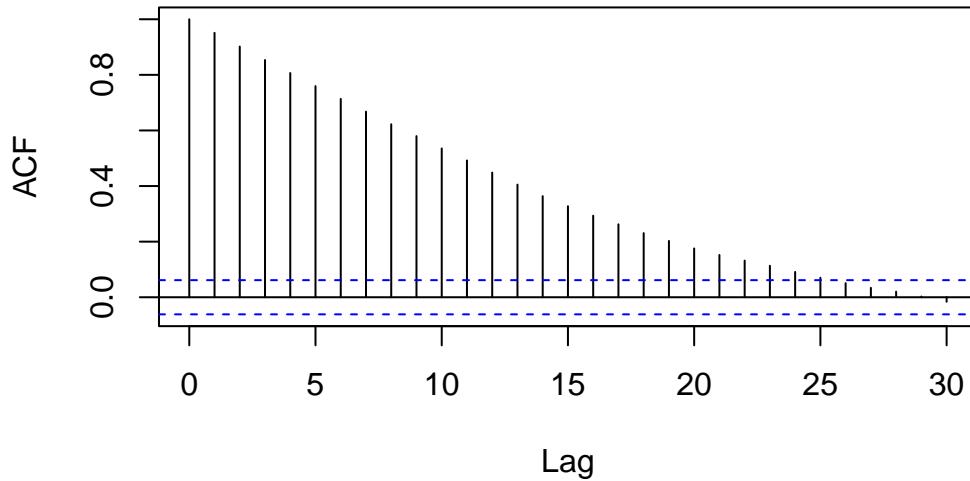


Analiza wykresów dekompozycji wskazuje, że dominującą strukturę szeregu stanowi trend, natomiast składnik sezonowy jest stosunkowo niewielki. Oznacza to, że w analizowanym okresie ceny akcji Valero Energy nie wykazują wyraźnej powtarzalnej sezonowości, co jest charakterystyczne dla dziennych danych giełdowych.

Analiza reszt modelu trendu

Reszty najlepszego modelu trendu zostały poddane testowi Ljunga–Boxa, który sprawdza, czy istnieje autokorelacja reszt:

ACF reszt modelu wielomianowego



Box-Ljung test

```
data: residuals_trend
X-squared = 6921.8, df = 20, p-value < 2.2e-16
```

Test Ljunga–Boxa dla reszt modelu trendu wykazuje $p\text{-value} < 2.2 \times 10^{-16}$. Odrzucamy hipotezę zerową o braku autokorelacji.

Wniosek: Reszty nie są białym szumem. Oznacza to, że model deterministyczny (wielomian) nie wyjaśnia dynamiki szeregu. Szereg prawdopodobnie jest niestacjonarny i wymaga podejścia stochastycznego (ARIMA).

Badanie stacjonarności (Test ADF)

Przed zastosowaniem modeli ARIMA, należy formalnie zbadać stacjonarność szeregu za pomocą Rozszerzonego Testu Dickeya-Fullera (ADF).

Hipotezy:

H_0 : Szereg jest niestacjonarny (posiada pierwiastek jednostkowy).
 H_1 : Szereg jest stacjonarny.

Augmented Dickey-Fuller Test

```
data: prices
Dickey-Fuller = -3.0962, Lag order = 10, p-value = 0.1143
alternative hypothesis: stationary
```

```
Warning in adf.test(returns): p-value smaller than printed p-value
```

Augmented Dickey-Fuller Test

```
data: returns
Dickey-Fuller = -9.4616, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

Interpretacja:

1. Dla cen surowych $p\text{-value} > 0.05$, co oznacza brak podstaw do odrzucenia H_0 . Szereg jest niestacjonarny.
2. Dla log-stóp zwrotu (pierwsze różnice) $p\text{-value} \leq 0.01$, co pozwala przyjąć H_1 .

Konkluzja: Szereg jest zintegrowany w stopniu pierwszym ($I(1)$), co uzasadnia przyjęcie parametru $d=1$ w modelu ARIMA.

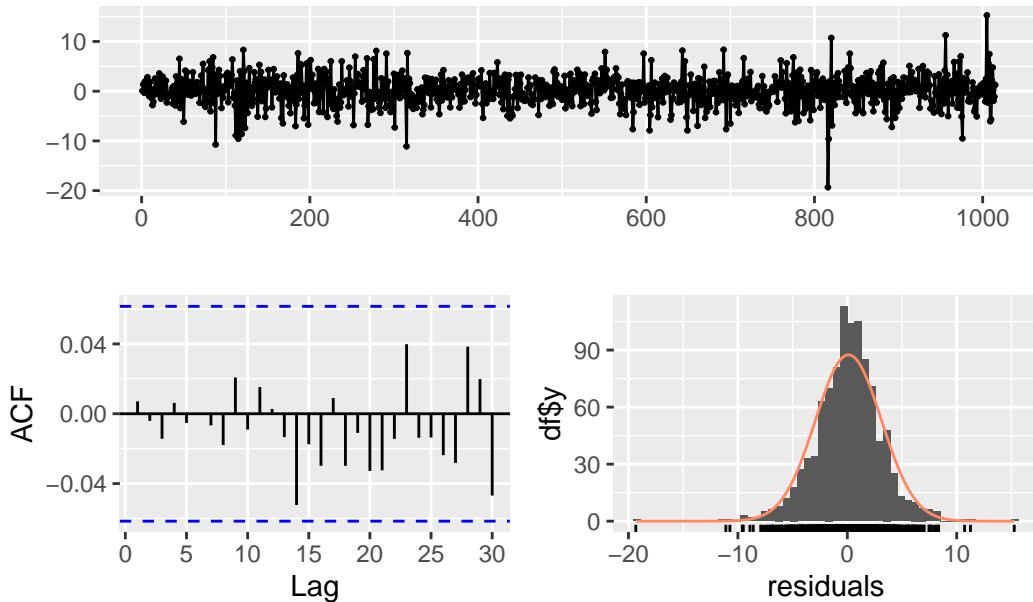
Modelowanie stochastyczne - ARIMA

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

sigma^2 = 9.136: log likelihood = -2560.42
AIC=5122.85   AICc=5122.85   BIC=5127.77

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.106076 3.021165 2.246017 0.05895511 1.725712 0.9990486
          ACF1
Training set 0.007087307
```

Residuals from ARIMA(0,1,0)



Ljung-Box test

```
data: Residuals from ARIMA(0,1,0)
Q* = 1.2448, df = 10, p-value = 0.9995

Model df: 0. Total lags used: 10
```

Dopasowano model **ARIMA(0,1,0)**, który najlepiej opisuje strukturę szeregu po uwzględnieniu trendu.

Model ARIMA(0,1,0) to tzw. Błądzenie Losowe (Random Walk). Można go zapisać równaniem: $Y_t = Y_{t-1} + \epsilon_t$, gdzie ϵ_t to biały szum. Oznacza to, że najlepszą prognozą ceny na jutro jest cena dzisiejsza, a zmiany cen są nieprzewidywalne. Jest to zgodne z Hipotezą Rynku Efektywnego (EMH) w jej słabej formie.

Jakość dopasowania oceniono za pomocą błędów prognozy (RMSE = 3.021, MAE = 2.246) oraz testu Ljunga–Boxa na resztach modelu:

```
Q* = 1.2448, df = 10, p-value = 0.9995
```

Wysoka wartość p-value w teście Ljunga–Boxa oznacza, że reszty modelu ARIMA nie wykazują istotnej autokorelacji, co świadczy o poprawnym dopasowaniu modelu. Model ten skutecznie uchwycił zależności czasowe w średniej szeregu, umożliwiając dalszą analizę zmienności i potencjalne prognozy cen.

Podsumowanie

Ceny akcji Valero Energy również wykazały nieliniowy trend, najlepiej opisany przez wielomian 10. stopnia. Sezonowość była słaba, a reszty trendu nie były losowe, co uzasadniało zastosowanie ARIMA(0,1,0). Dopasowany model dobrze uchwycił dynamikę średniego poziomu cen, a reszty nie wykazały autokorelacji (p-value w teście Ljunga–Boxa równe 0.9995), co potwierdza poprawność modelu.