Forecasting NVIDIA Stock Prices:

A Comparative Analysis of Financial Models

By Oday Najad and Antoni Valls

Why NVIDIA? History:

Five keypoints:

- Founded in 1993 by Jensen Huang, Chris Malachowsky, and Curtis Priem.
- Launched the first GPU in 1999, revolutionizing gaming graphics.
- Expanded into AI and data centers in the 2010s, driving deep learning innovation.
- Acquired ARM Holdings in 2020, aiming to dominate the semiconductor market.
- Pioneered GPU-based cryptocurrency mining, contributing to Bitcoin's rise.

- NVIDIA's GPUs are a cornerstone for AI development, powering deep learning frameworks and accelerating computations crucial for AI innovation. As AI adoption surges across industries, NVIDIA's stock has become a key indicator of technological progress.

- Simultaneously, the cryptocurrency market has seen a strong reliance on NVIDIA’s hardware for efficient mining operations.

- In June it became the most valuable company in the world after its market capitalization grew to $3.34 trillion.

- It is leading the S&P500 this year (2024).

Goal of the project

01

Which financial models most accurately predict NVIDIA's stock closing price?
Test different models learned in class to evaluate their performance in stock price modeling.

02

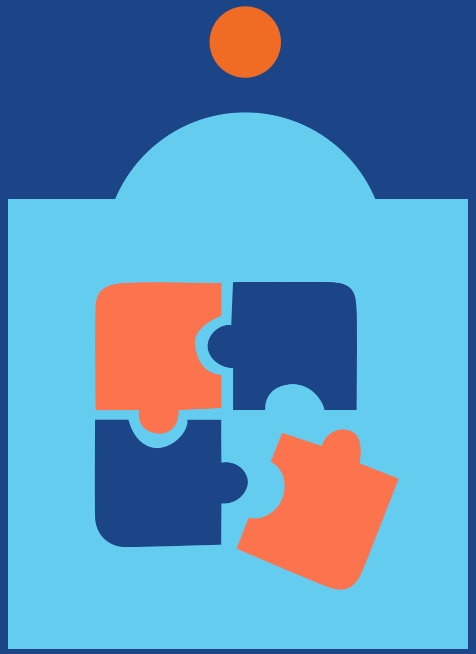
How does incorporating external data, like Bitcoin price and VIX, impact forecasting accuracy?
Analyze whether external indicators improve the predictive power of the models.

03

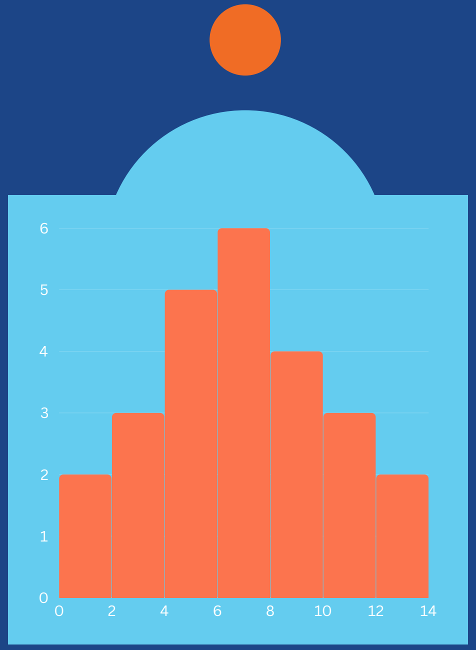
How do sector-specific factors influence NVIDIA's stock price movements?
Explore how incorporating technology and semiconductor industries features improves our models.

DATA

S&P500 data



The Standard and Poor's 500 or S&P 500 is the most famous financial benchmark in the world.
This stock market index tracks the performance of 500 large companies listed on stock exchanges in the United States.



Daily updated data can be downloaded from many, such as Nasdaq.com. We source it from a Kaggle repository, which provided access to three key datasets:

- sp500_companies.csv
- sp500_index.csv
- sp500_stocks.csv

s

S&P500 data

- sp500_companies.csv: Contains detailed information about S&P 500 companies, including their stock symbol, sector, and industry classification.

- sp500_index.csv: Tracks the S&P 500 index value from September 2, 2014, to August 30, 2024.

- sp500_stocks.csv: Provides daily stock prices for all S&P 500 companies from January 4, 2010, to August 30, 2024.

DATA

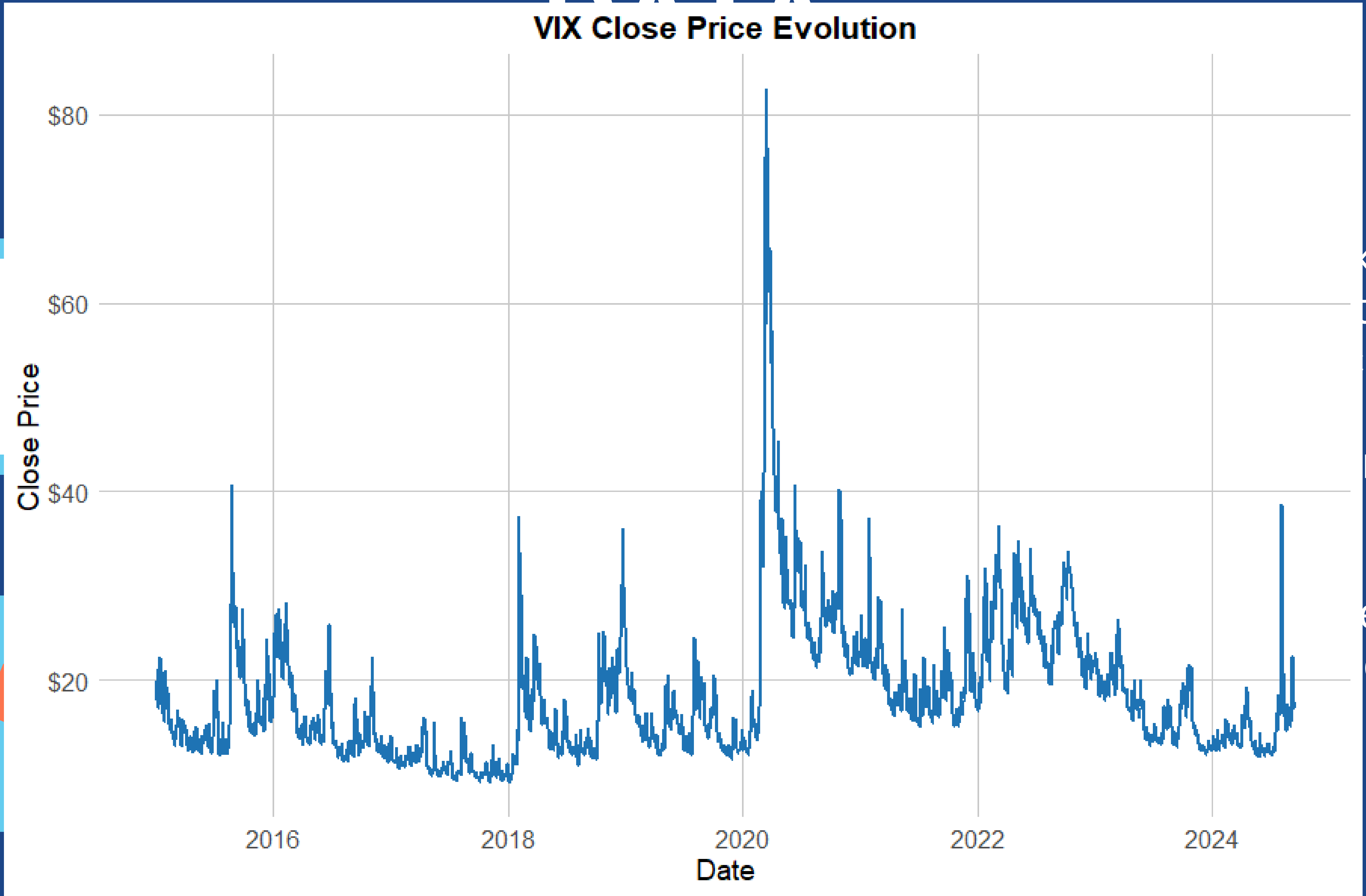
VIX Index

The VIX Index (Volatility Index) measures market expectations of future volatility and investor sentiment. It reflects market uncertainty that can impact stock prices, especially for tech firms, as they are normally sensitive to market sentiment and economic uncertainty.

Daily updated data since 1990 is available in Cboe’s webpage, the largest options exchange in the world.

s

DATA



ket
tor
an
as
nd

e’s
d.

DATA

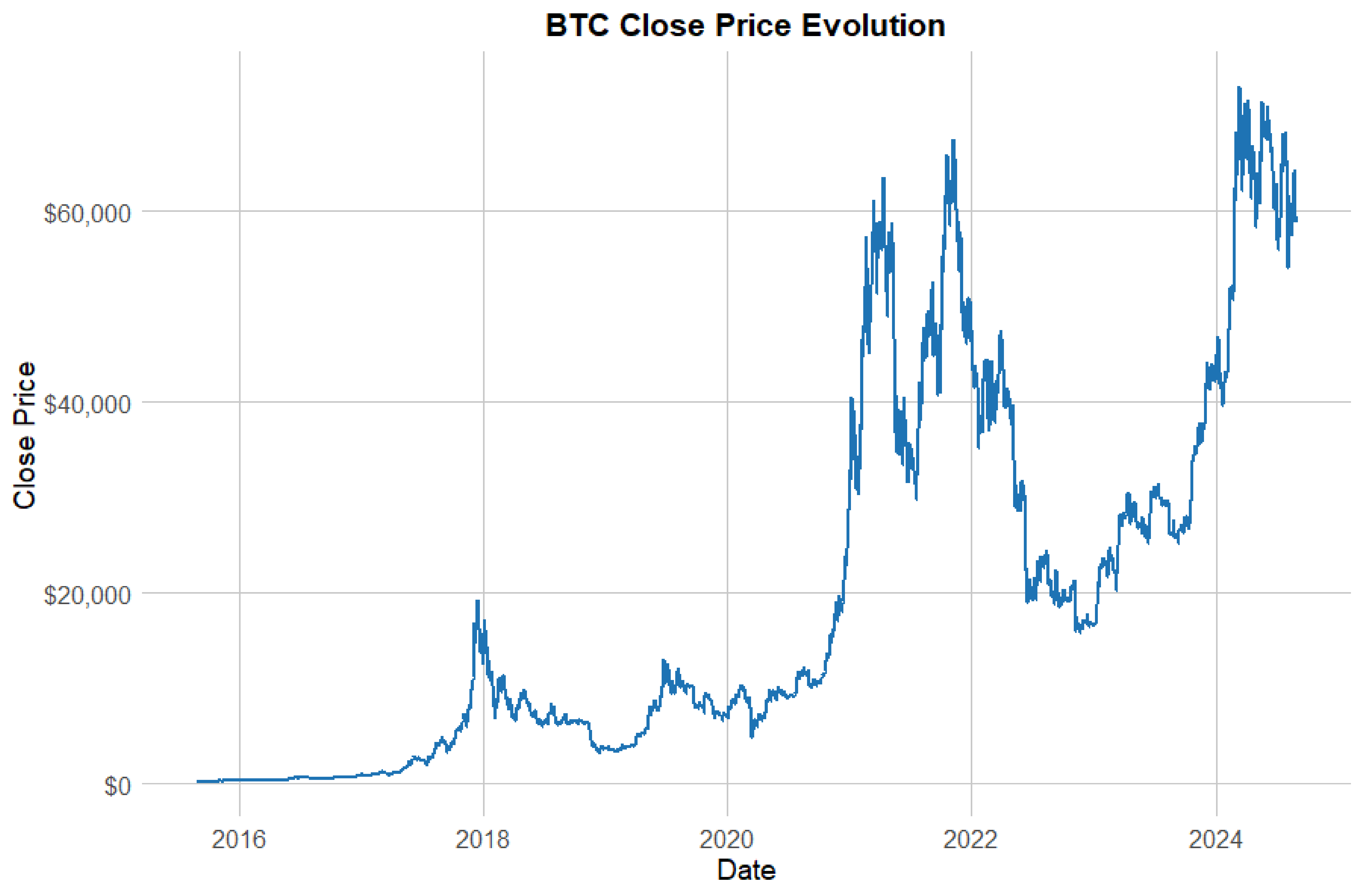
Bitcoin Value

BTC value is a key indicator of the cryptocurrency market’s evolution, reflecting trends and investor sentiment. As NVIDIA's GPUs are crucial for Bitcoin mining, fluctuations in Bitcoin’s price can directly impact NVIDIA's stock performance.

Daily updated BTC values can be downloaded from many sources. We got it from the financial platform Investing.com.

s

DATA



rrency
vestor
itcoin
irectly

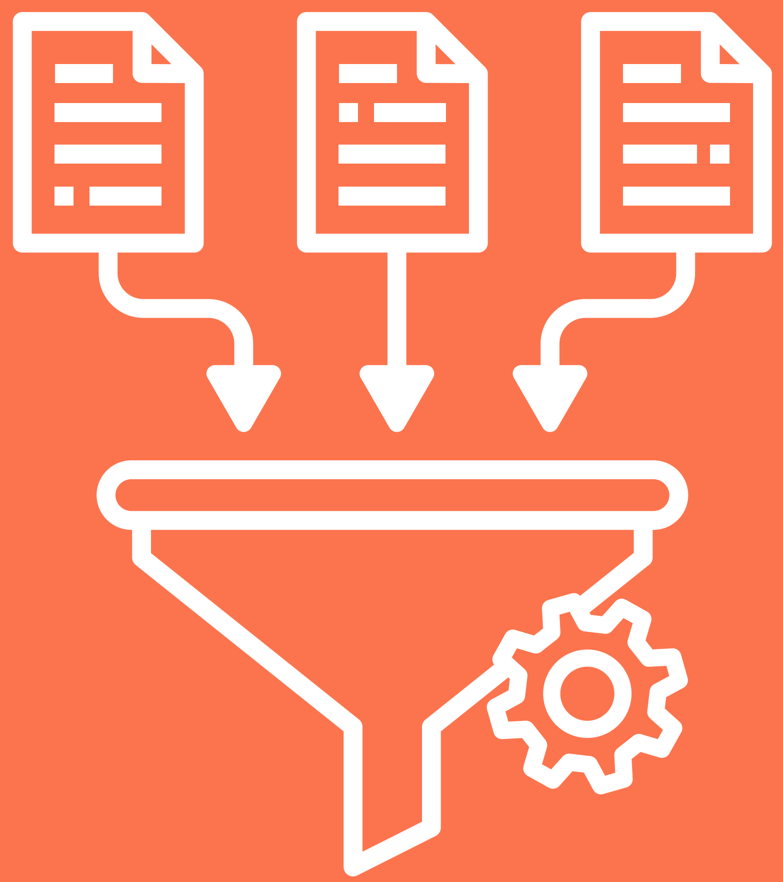
d from
atform

Data Preparation:

s

- Remove columns with missing values:
  - sp500_companies.csv contains 58 missing values.
  - No missing values in the other datasets.

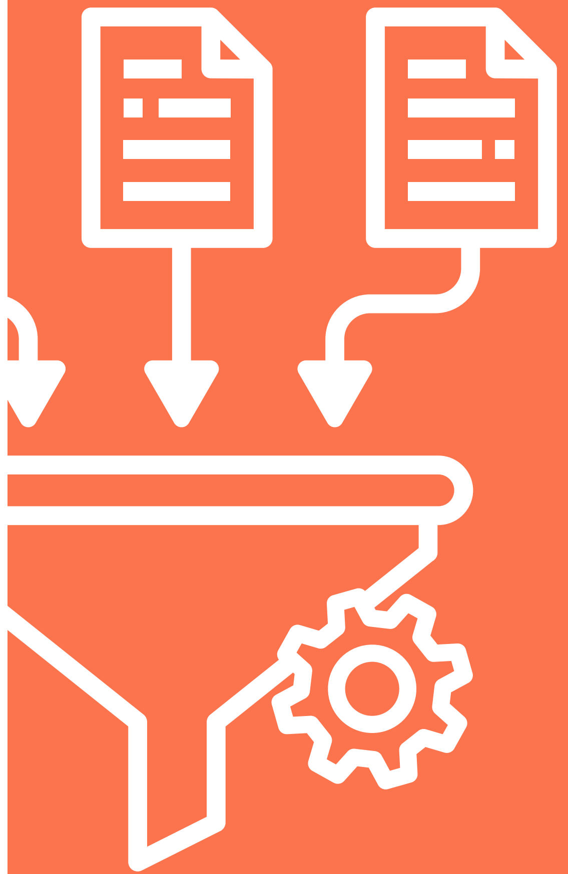
- Filter datasets to include data after 2015:
  - NVIDIA stock prices were significantly lower before 2015.



Data Preparation:

s

- Remove co
  - sp500_
  - No miss
- Filter datas
  - NVIDIA



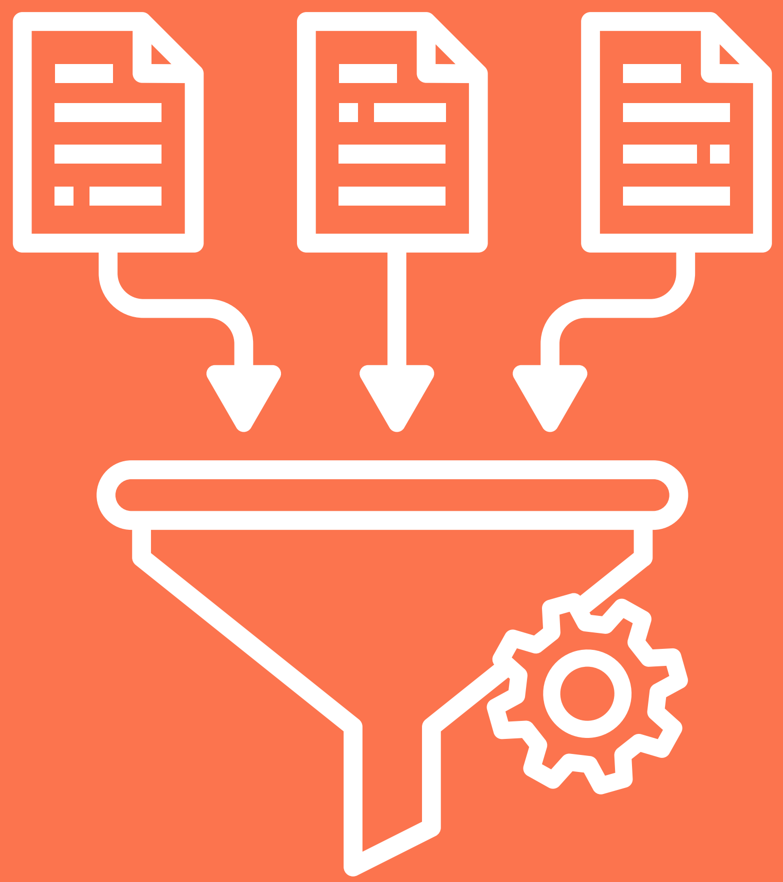
Data Preparation:

s

- Remove columns with missing values:
  - sp500_stocks.csv contains 183,396 missing values (about 9% of the total).
  - sp500_companies.csv contains 58 missing values.
  - No missing values in the other datasets.

- Filter datasets to include data after 2015:
  - NVIDIA stock prices were significantly lower before 2015.

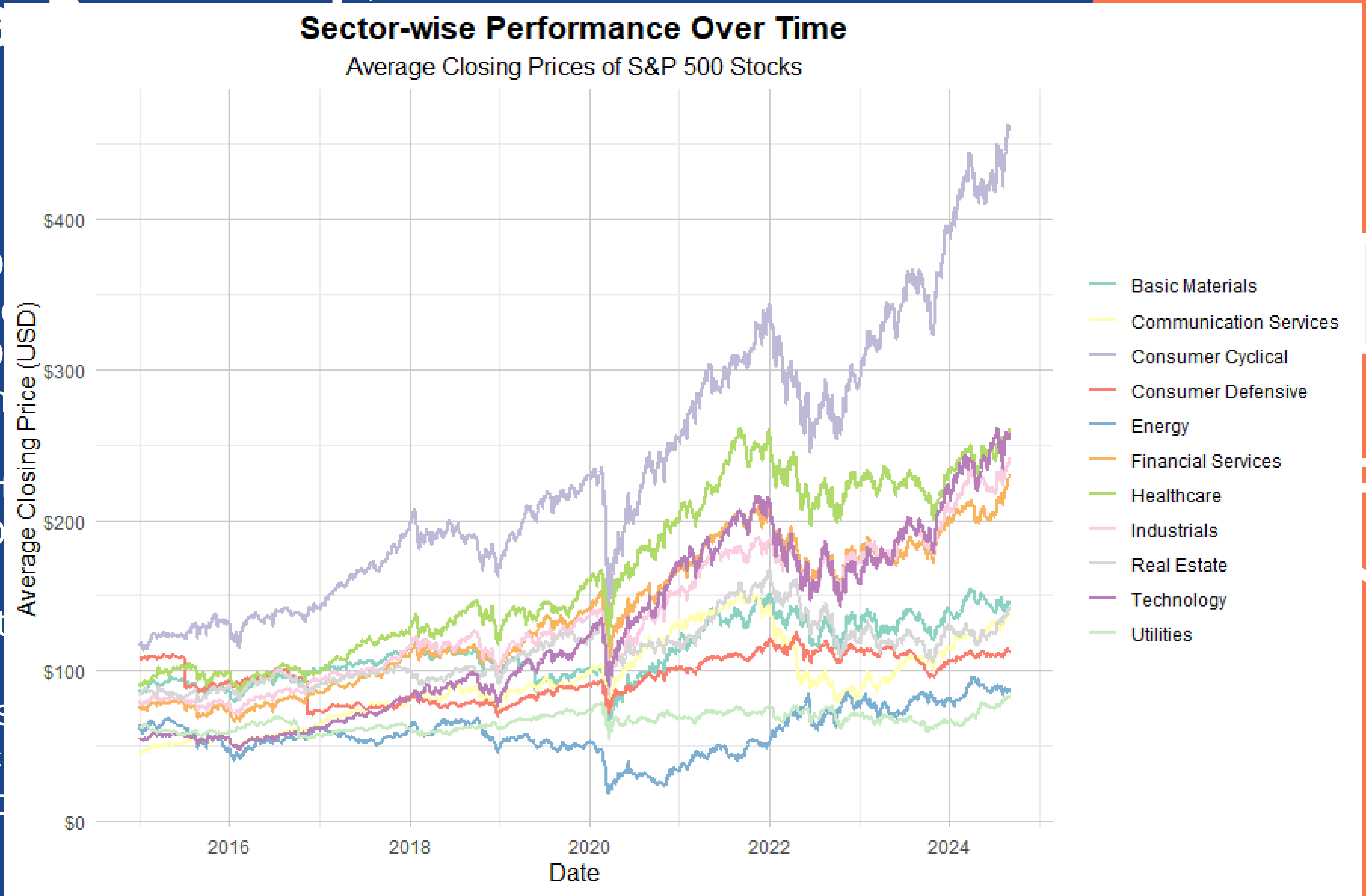
- Segment the S&P 500 data by sector and industry:
  - The data includes 11 distinct sectors and 115 different industries.
  - We collect the average stock values of the Technology sector and the Semiconductors industry.



Data

s

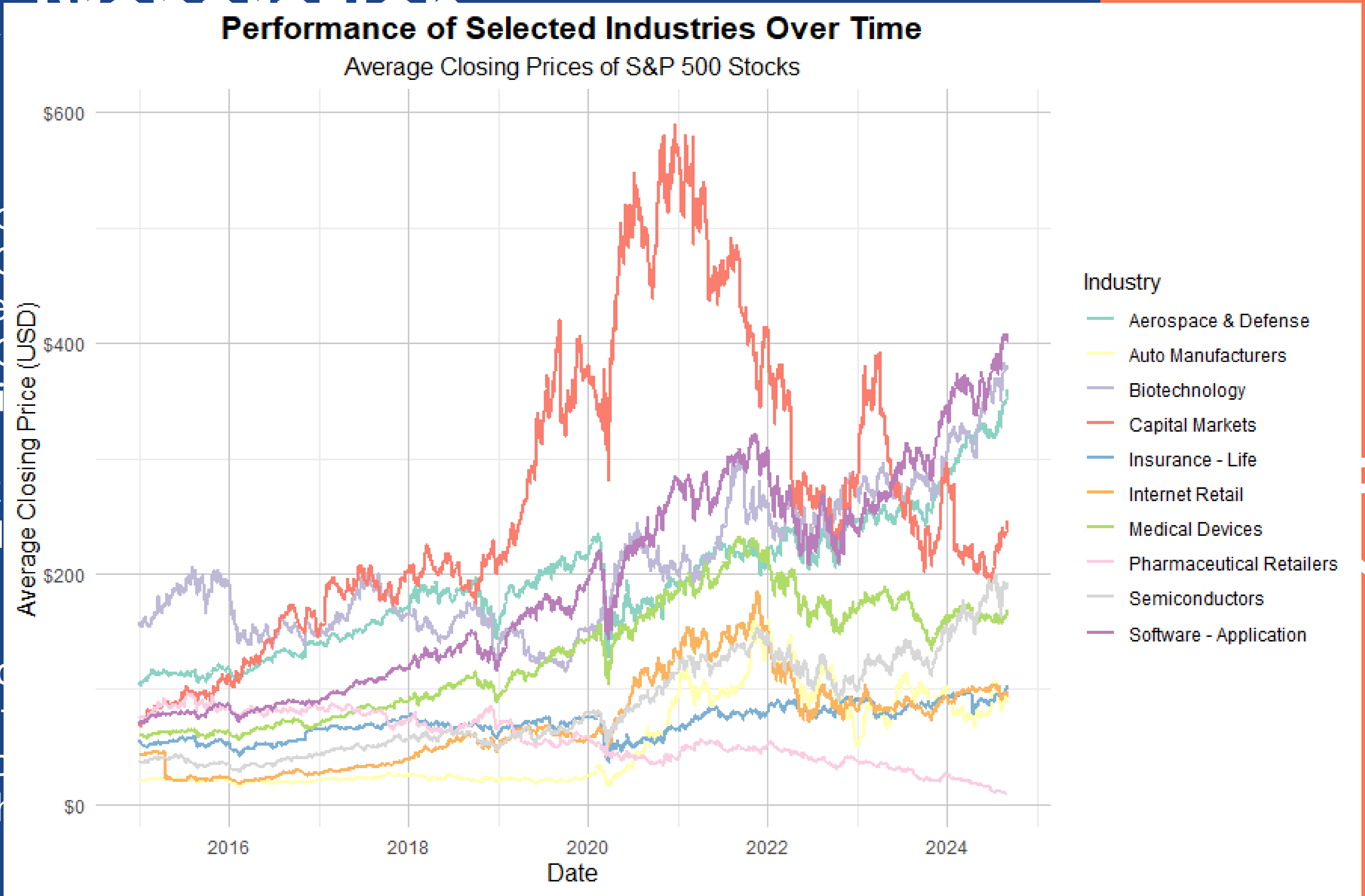
- Remove
  - sp50
    of th
  - sp50
  - No m
- Filter dat
  - NVID
- Segment
  - The
    indus
  - We c
    and t



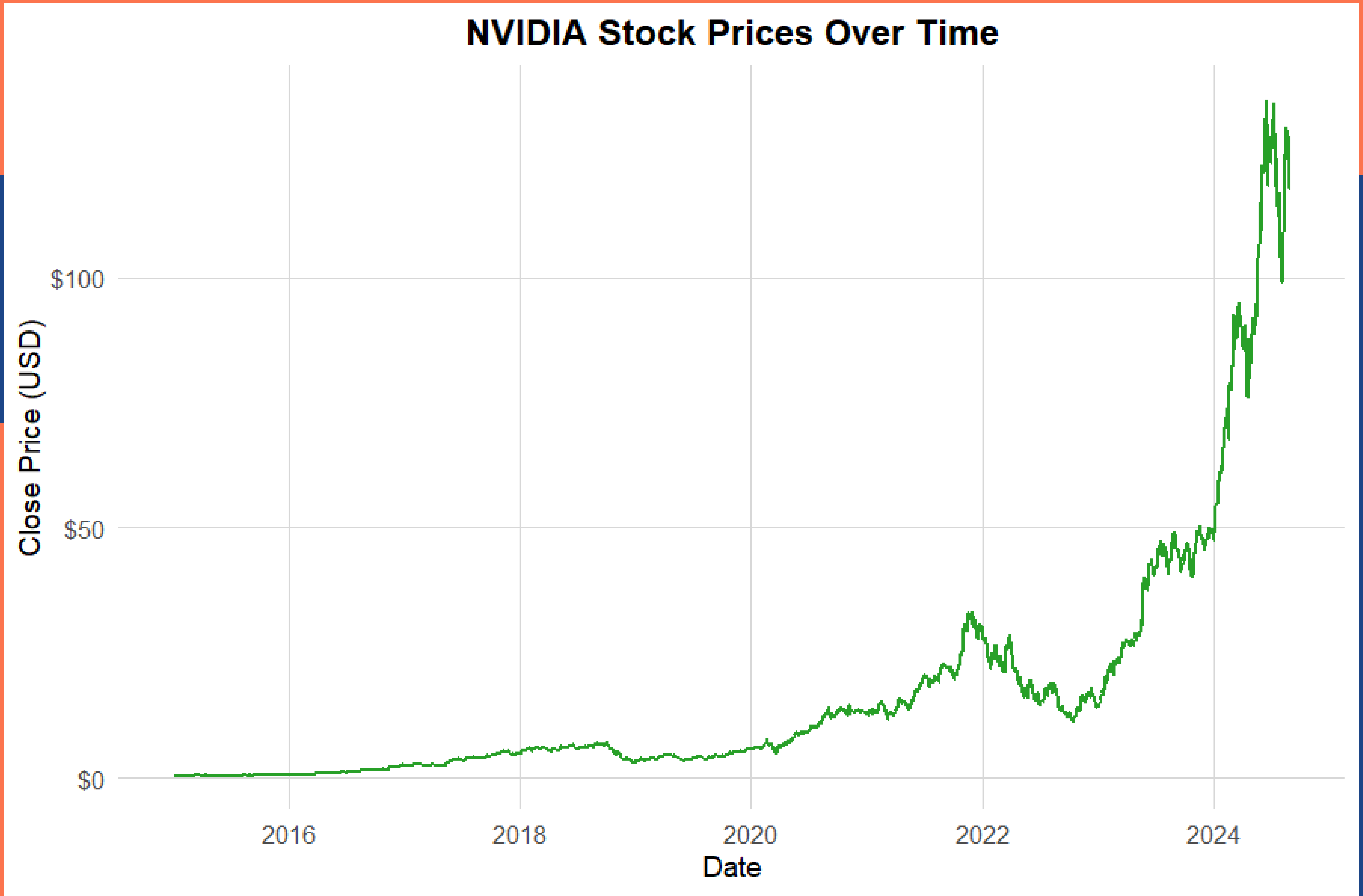
Data Preparation

s

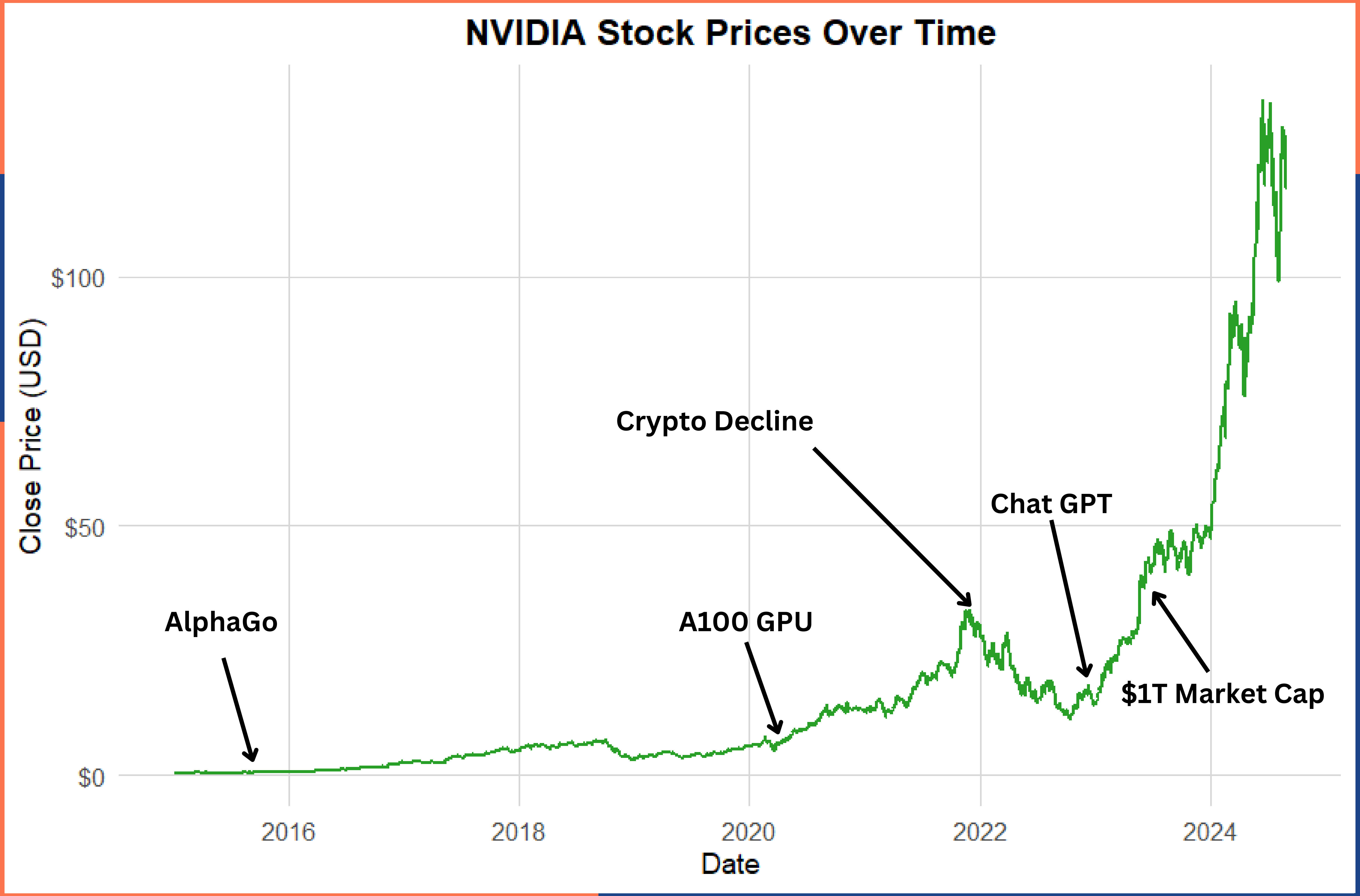
- Remove c
  - sp50(
    of the
  - sp50(
  - No mi
- Filter dat
  - NVIDI
- Segment
  - The c
    indus
  - We co
    and th



EDA

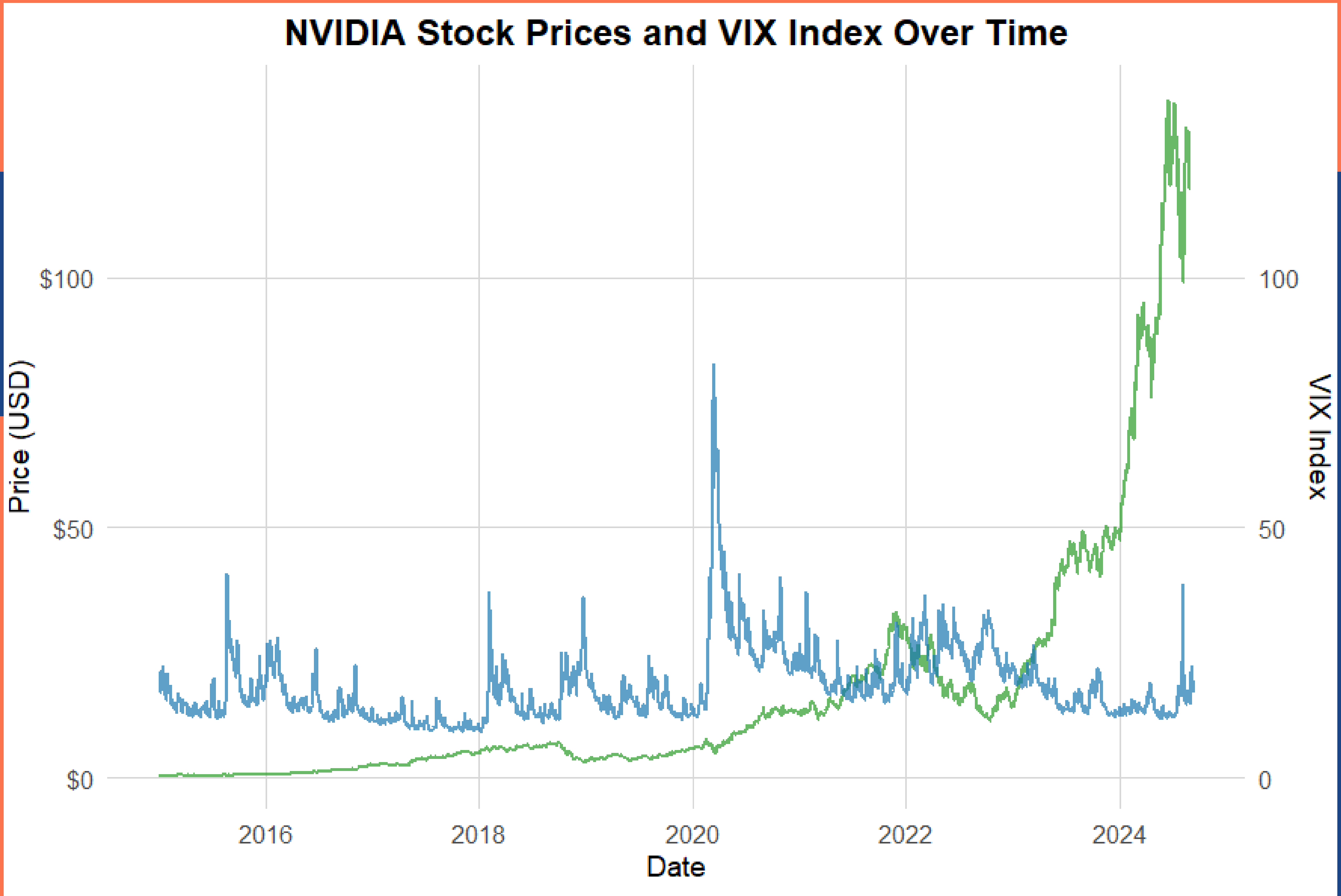

NVIDIA Stock Prices Over Time
Close Price (USD)
$100
$50
$0
2016
2018
2020
2022
2024
Date

EDA



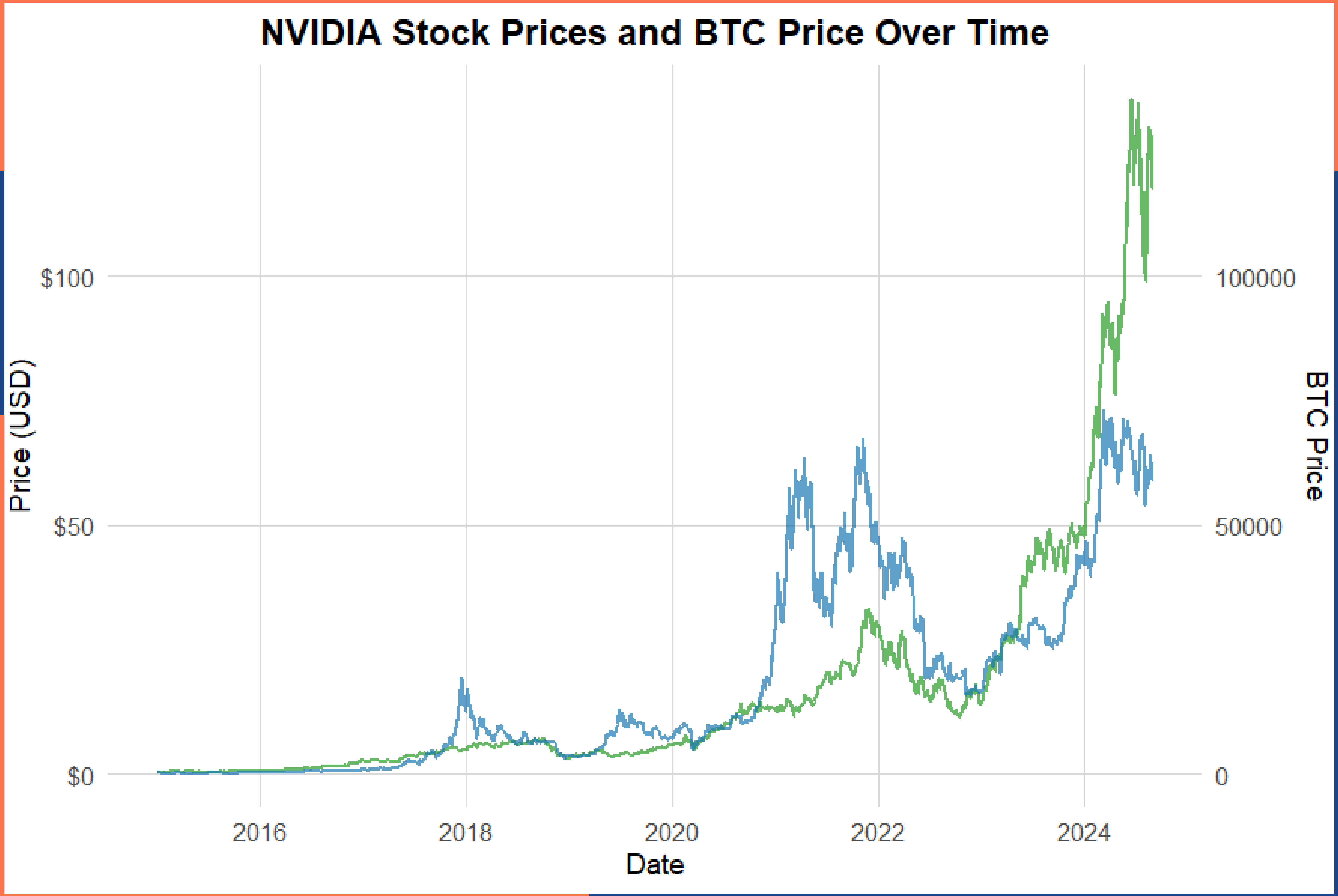
NVIDIA Stock Prices Over Time
Close Price (USD)
$100
$50
$0
2016
2018
2020
2022
2024
Date
AlphaGo
A100 GPU
Crypto Decline
Chat GPT
$1T Market Cap

EDA



NVIDIA Stock Prices and VIX Index Over Time
Price (USD)
$100
$50
$0
VIX Index
100
50
0
2016
2018
2020
2022
2024
Date

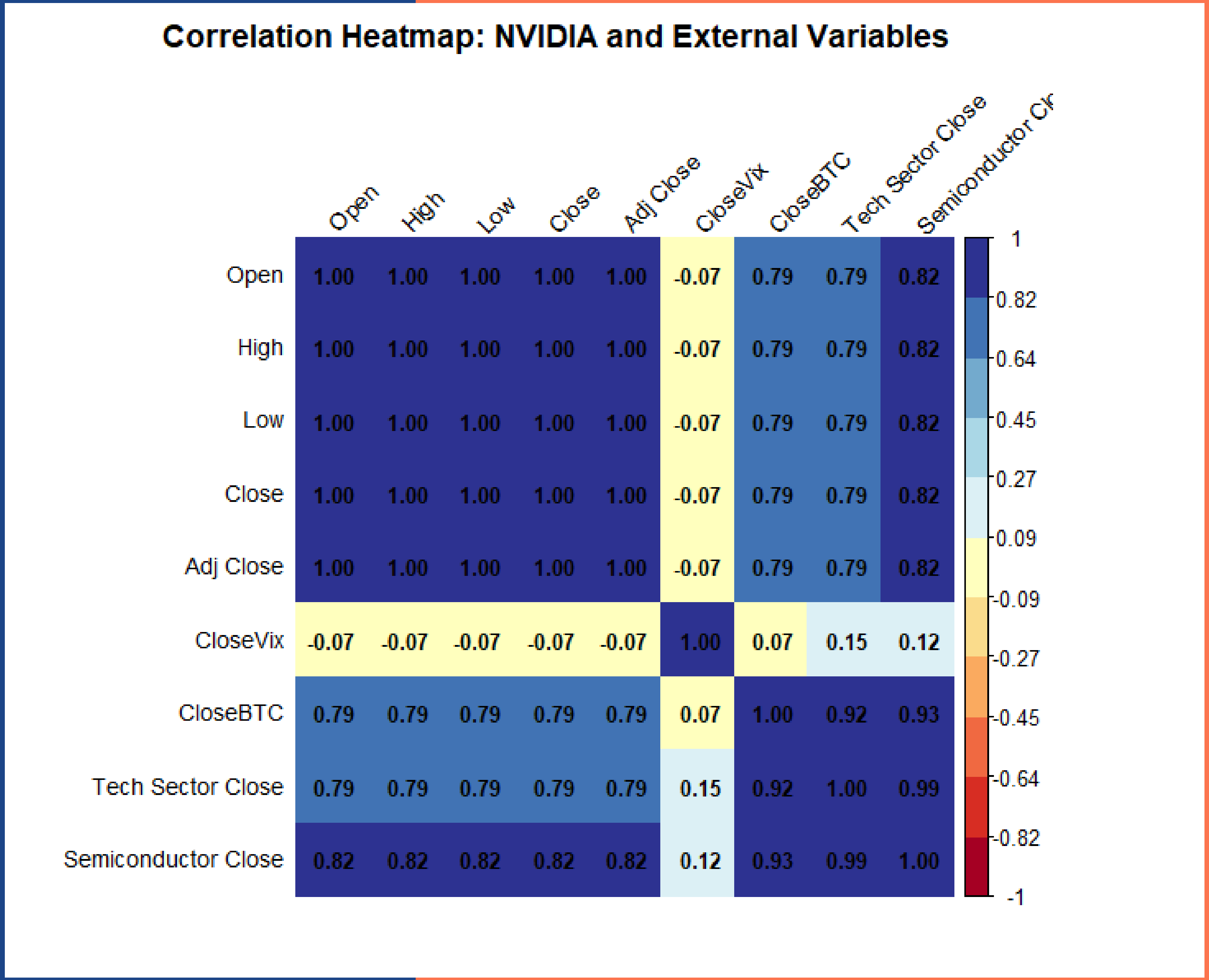
EDA


NVIDIA Stock Prices and BTC Price Over Time
Price (USD)
$100
$50
$0
BTC Price
100000
50000
0
2016
2018
2020
2022
2024
Date

Open, Close, High, Low

NVIDIA and BTC stock prices, and VIX index are represented by four key values: **Open**, **Close**, **High**, and **Low**. The Open marks the price at the start of the trading day, while the Close is the price at the end. High and Low indicate the day's highest and lowest prices. For this analysis, we will focus solely on the Close value, including for forecasting.

Correlation Matrix



Stationarity

ADF Test:

- Dickey-Fuller = 1.5607,
- Lag order = 15,
- p-value = 0.99

KPSS Test:

- KPSS Level = 18.903
- Truncation lag parameter = 9
- p-value = 0.01

Stationarity

ADF Test:

- Dickey-Fuller = 1.5607,
- Lag order = 15,
- p-value = 0.99

KPSS Test:

- KPSS Level = 18.903
- Truncation lag parameter = 9
- p-value = 0.01

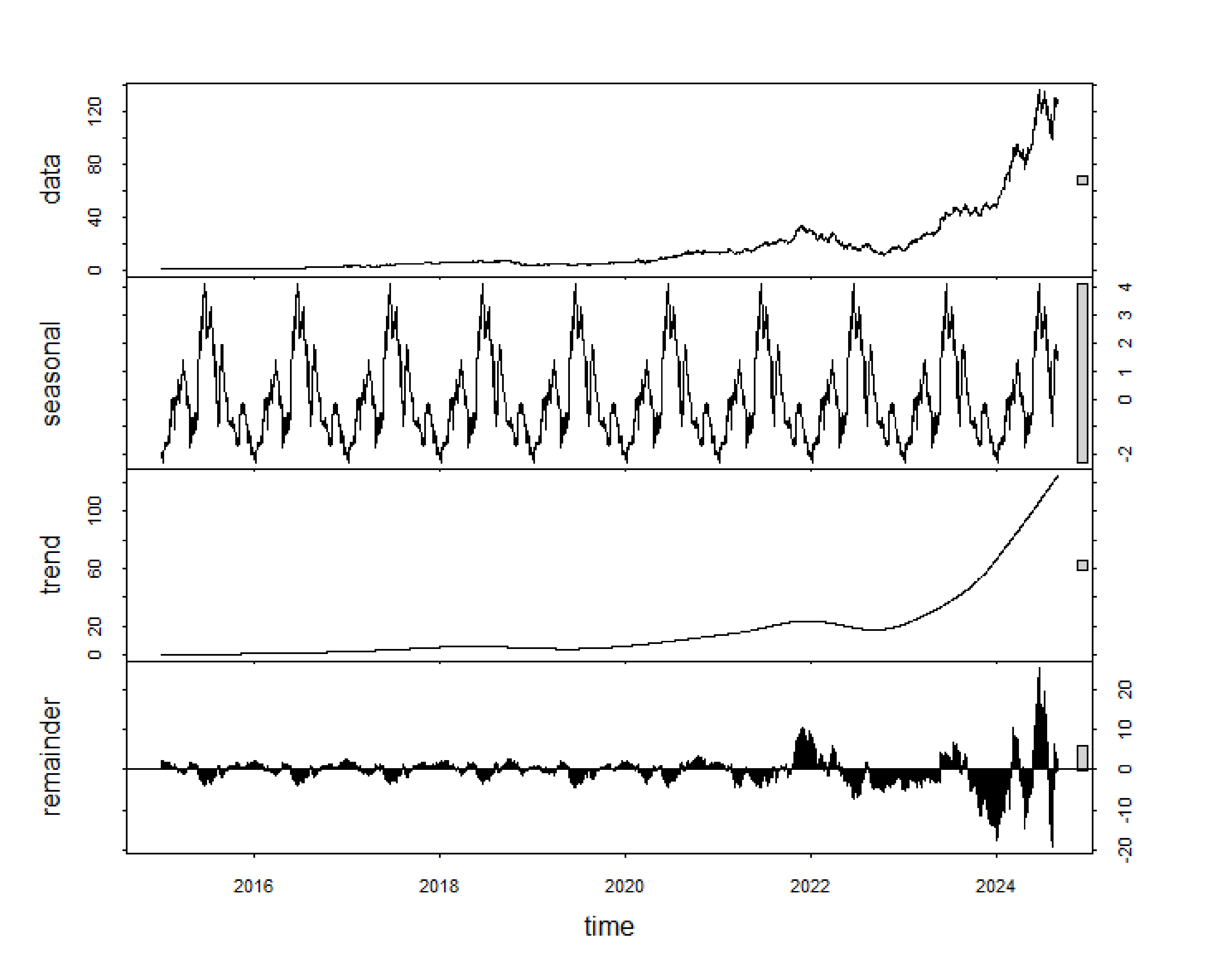
Both tests suggest that the data is **not** stationary

◆ STL Decomposition ◆

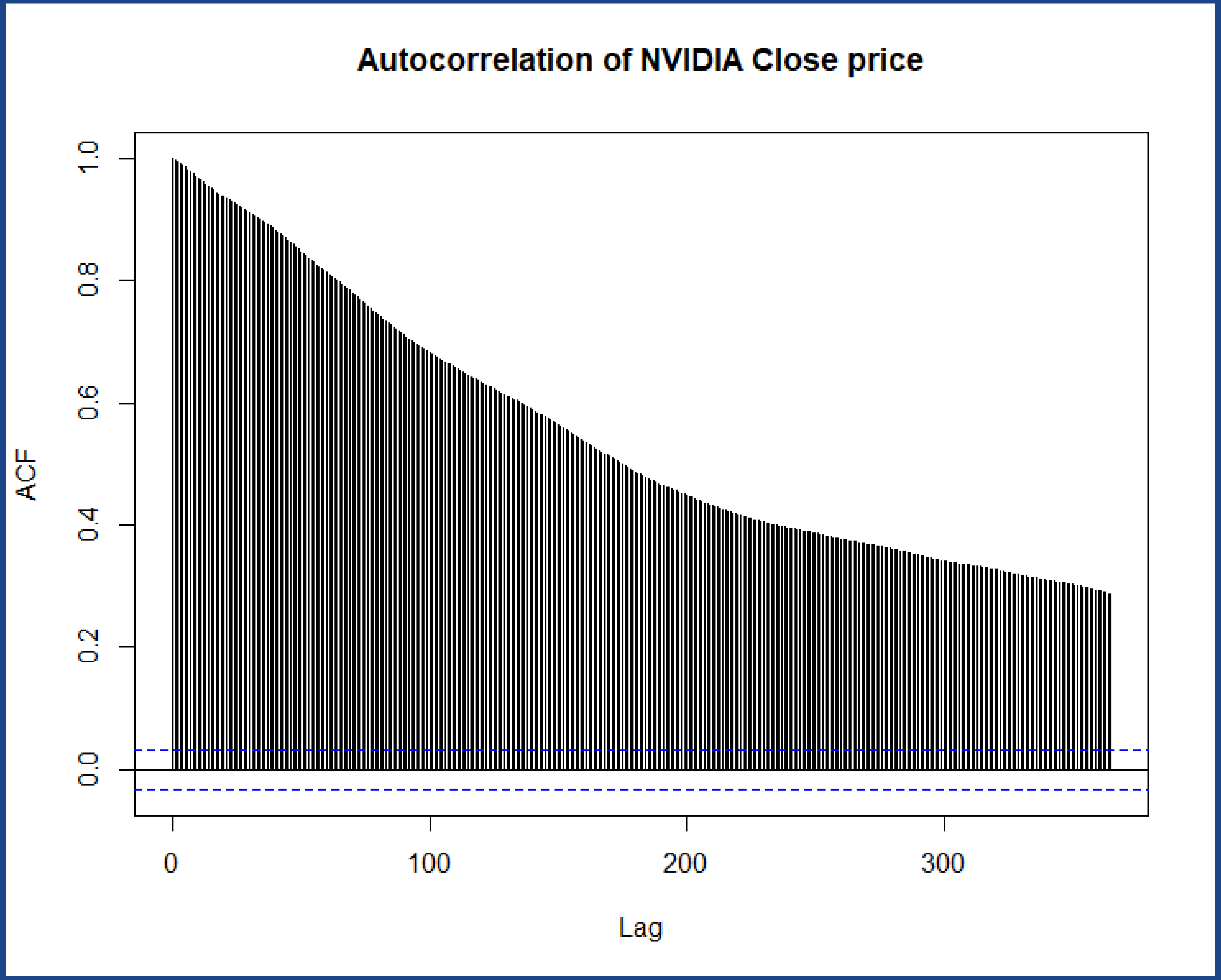
- **Clear Seasonality**: The data exhibits a strong and consistent seasonal pattern, suggesting that cyclical behaviors, likely linked to annual cycles

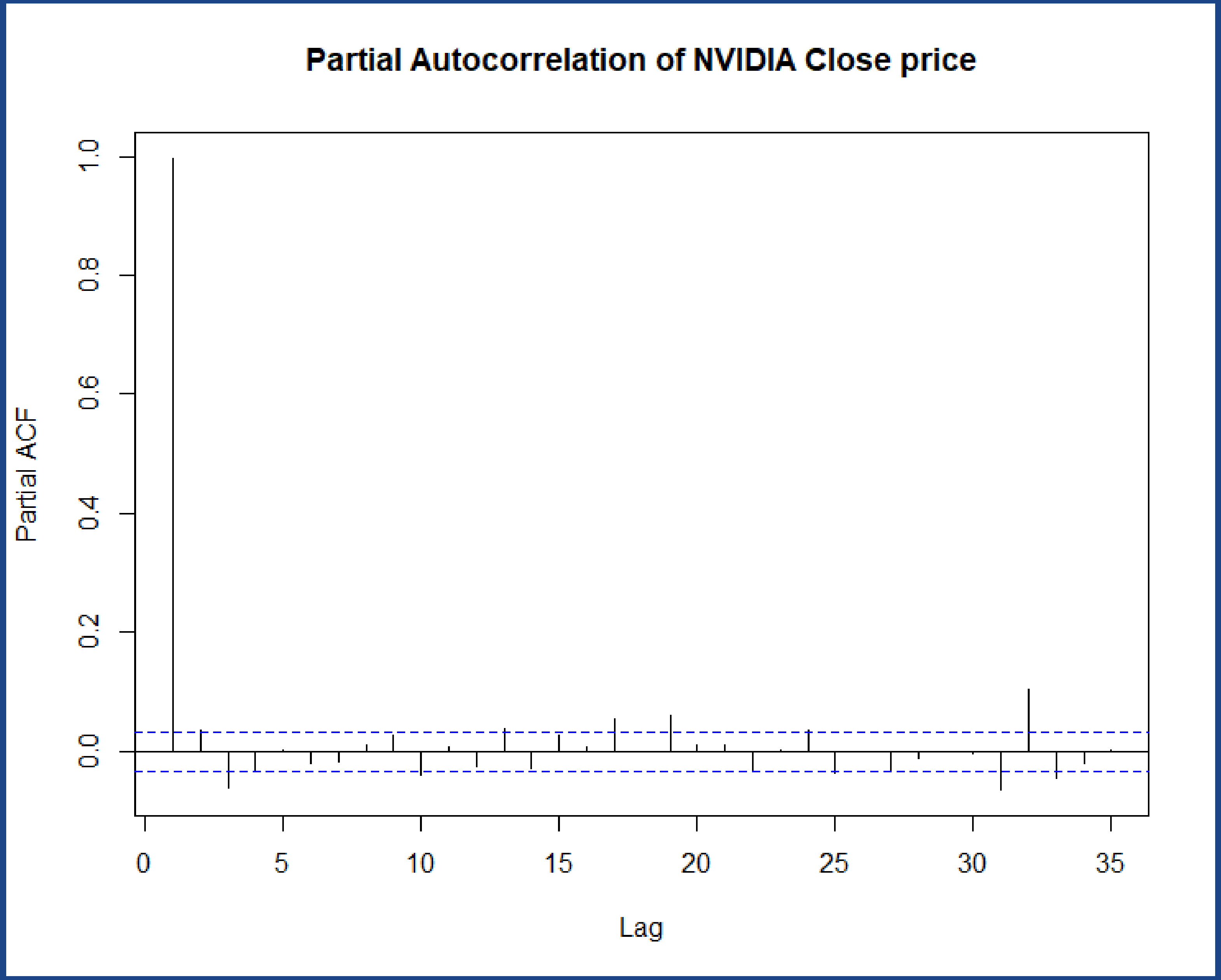
- **Huge trend**: Starting around 2020, there is a significant and accelerating upward trend.

- **Increased Volatility**: The remainder component shows growing volatility, particularly after 2020. This implies that the data is subject to irregular shocks or external influences, which become more pronounced as the trend increases.

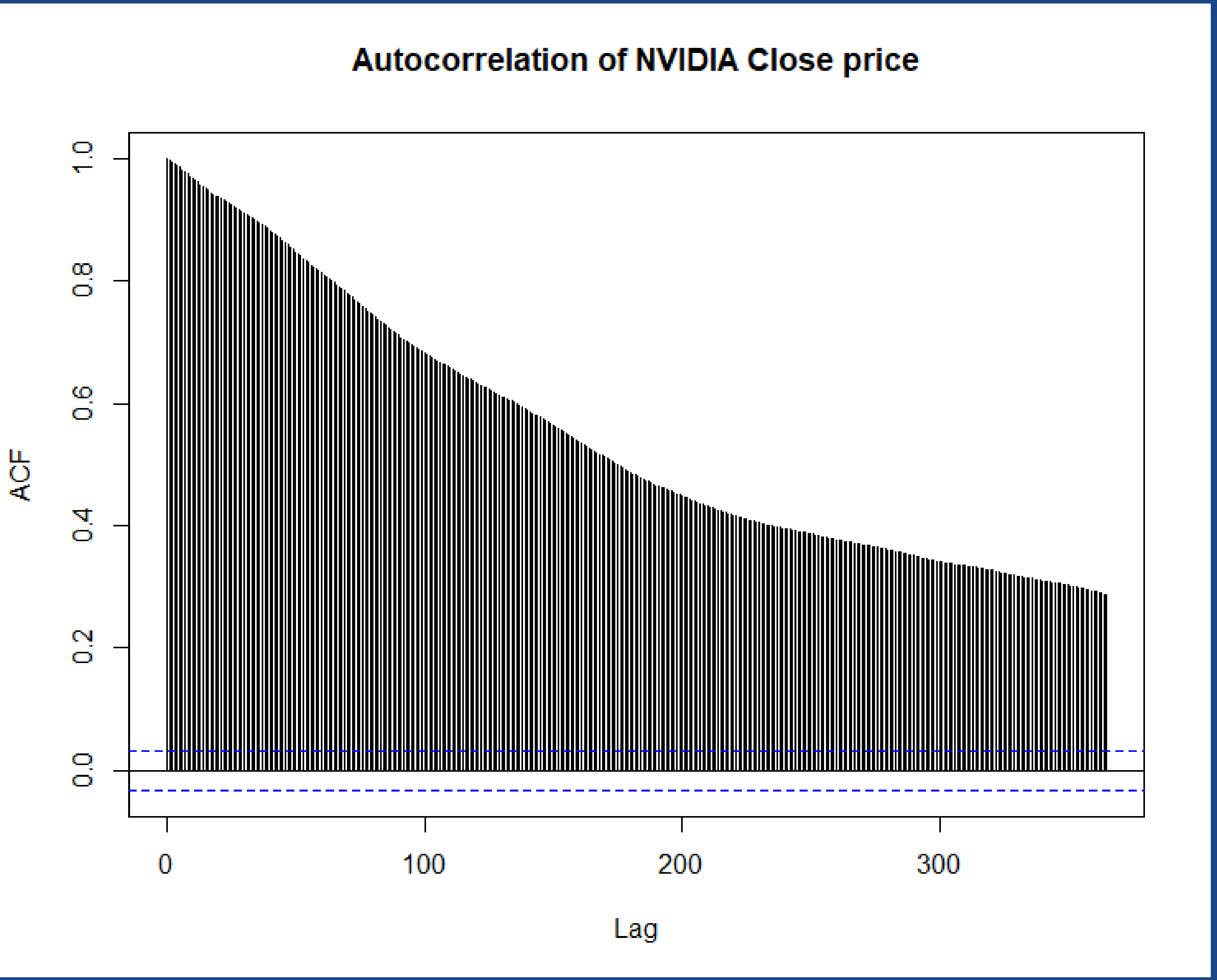


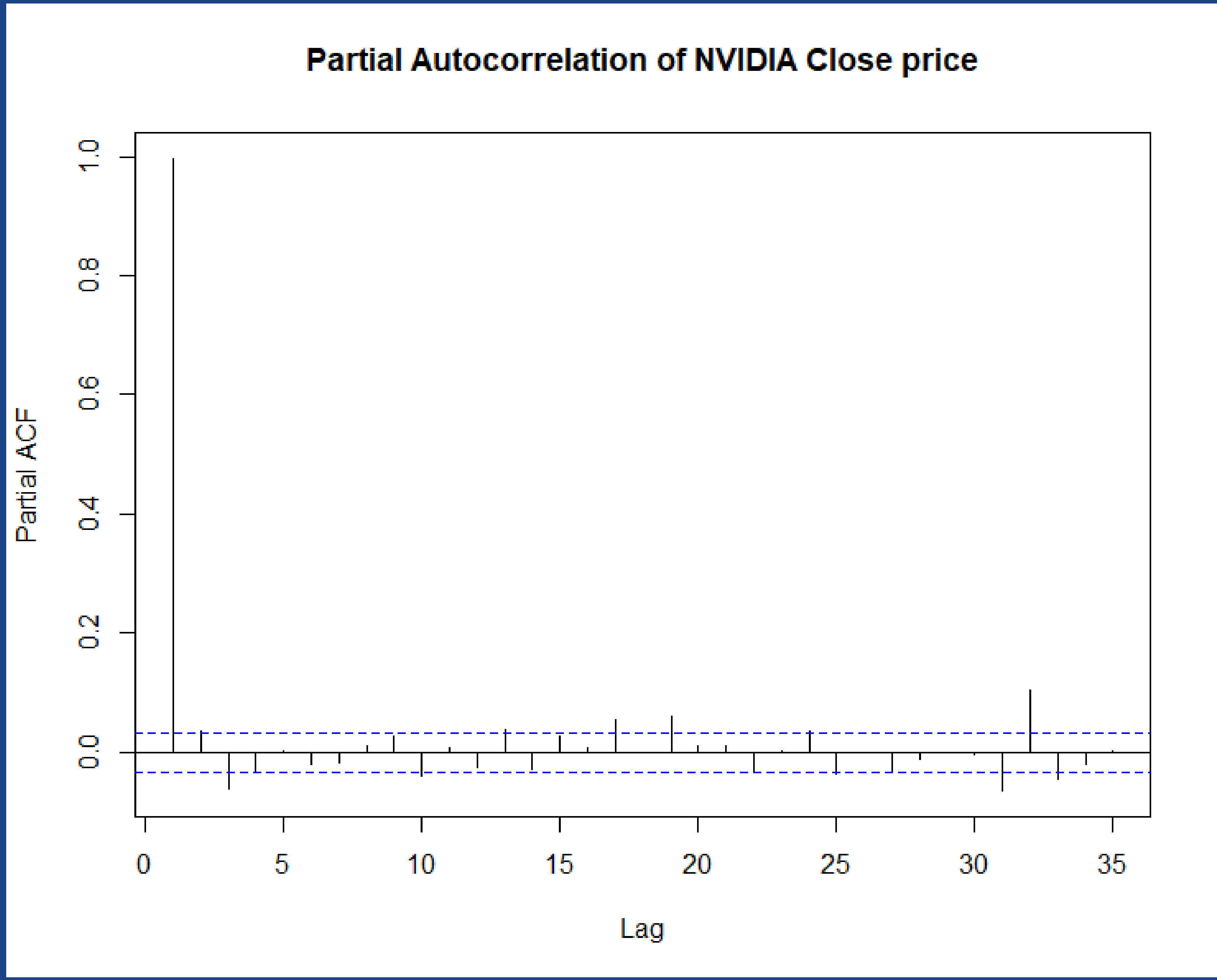
ACF and PACF Plots





ACF and PACF Plots





Exponentially decaying

It may be ARIMA (1,d,0)

Huge spike at Lag 1

First differentiation

ADF Test:

- Dickey-Fuller = -15.038
- Lag order = 15
- p-value = 0.01

KPSS Test:

- KPSS Level = 1.1733
- Truncation lag parameter = 9
- p-value = 0.01

First differentiation

ADF Test:

- Dickey-Fuller = -15.038
- Lag order = 15
- p-value = 0.01

KPSS Test:

- KPSS Level = 1.1733
- Truncation lag parameter = 9
- p-value = 0.01

ADF suggests that it is stationary but KPSS do not

Second differentiation

ADF Test:

- Dickey-Fuller = -23.062
- Lag order = 15
- p-value = 0.01

KPSS Test:

- KPSS Level = 0.0061815
- Truncation lag parameter = 9
- p-value = 0.1

Second differentiation

ADF Test:

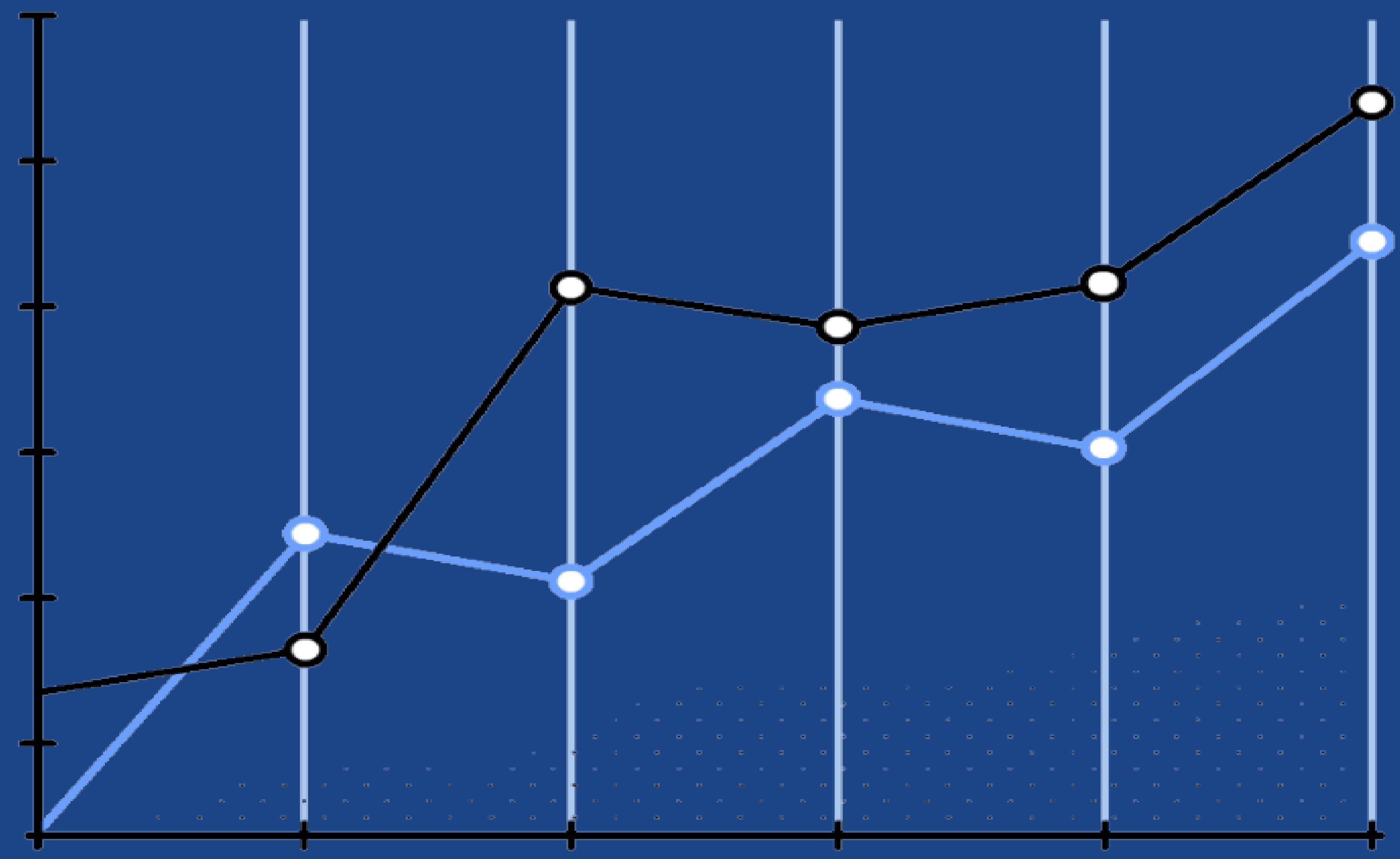
- Dickey-Fuller = -23.062
- Lag order = 15
- p-value = 0.01

KPSS Test:

- KPSS Level = 0.0061815
- Truncation lag parameter = 9
- p-value = 0.1

Both tests suggest that second differences are stationary

MODELING AND FORECASTING



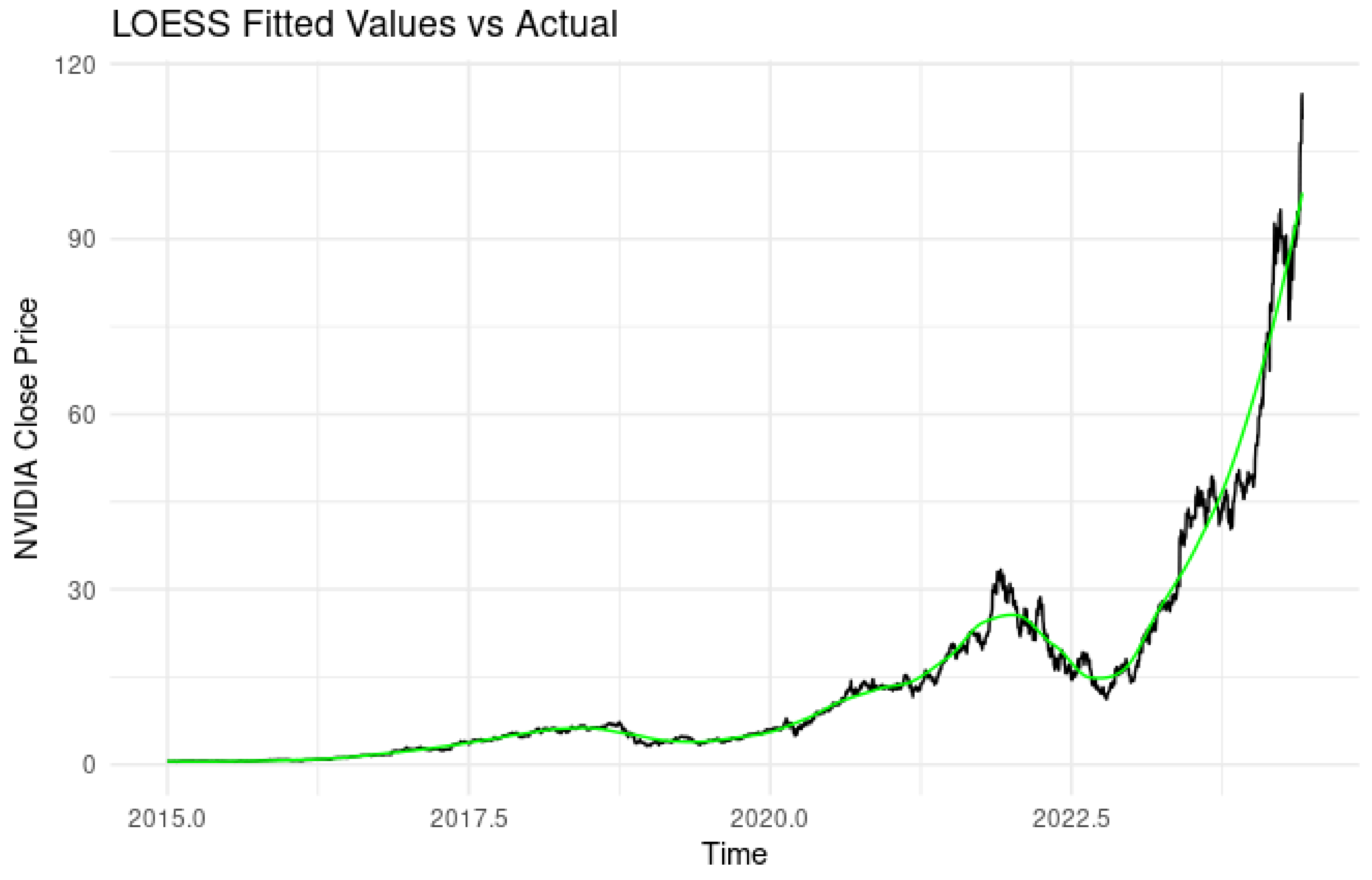
Modeling and Forecasting

For the modeling component, we have divided the dataset into training and test sets, with the test set comprising the last three months of data.

For each model:

- Plot the fitted model against the actual data.
- Calculate evaluation metrics, including RMSE and MAPE.
- Compute the residuals and create a plot of these residuals.
- Conduct ACF and PACF tests on the residuals, accompanied by visualizations of the results.
- Perform the Box-Pierce test to assess autocorrelation in the residuals.
- Execute the Durbin-Watson test on the residuals.
- Assess the normality of the residuals by plotting a histogram and conducting the Shapiro-Wilk normality test.

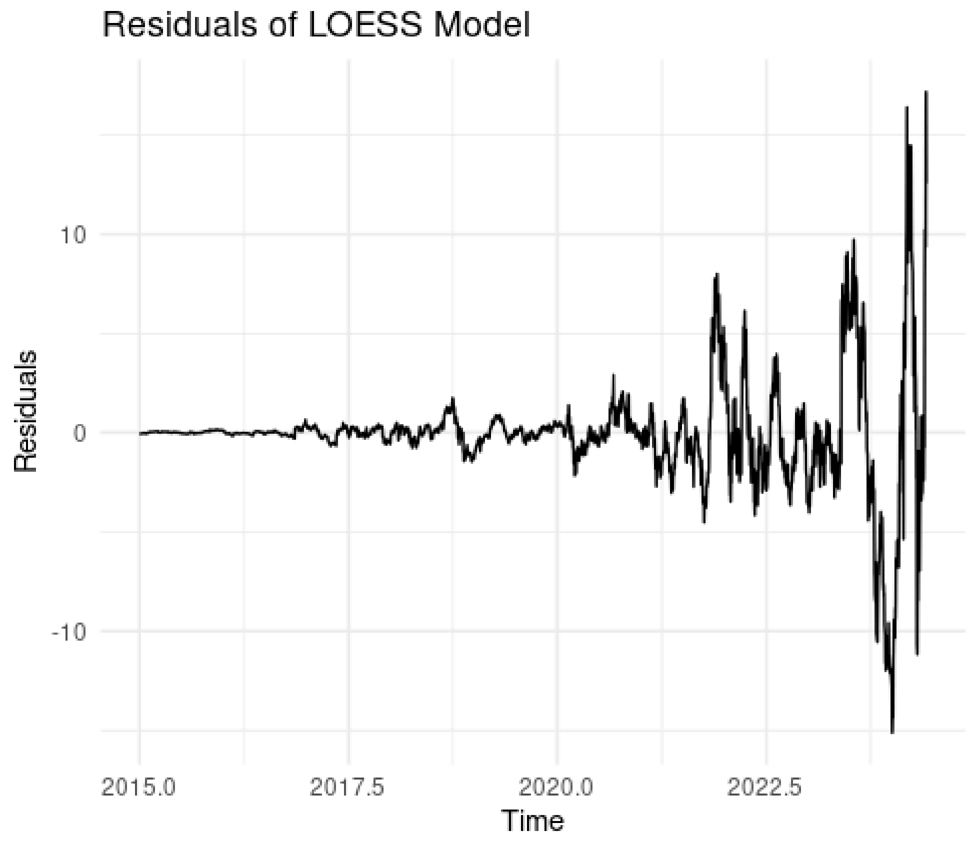
LOESS(modeling)

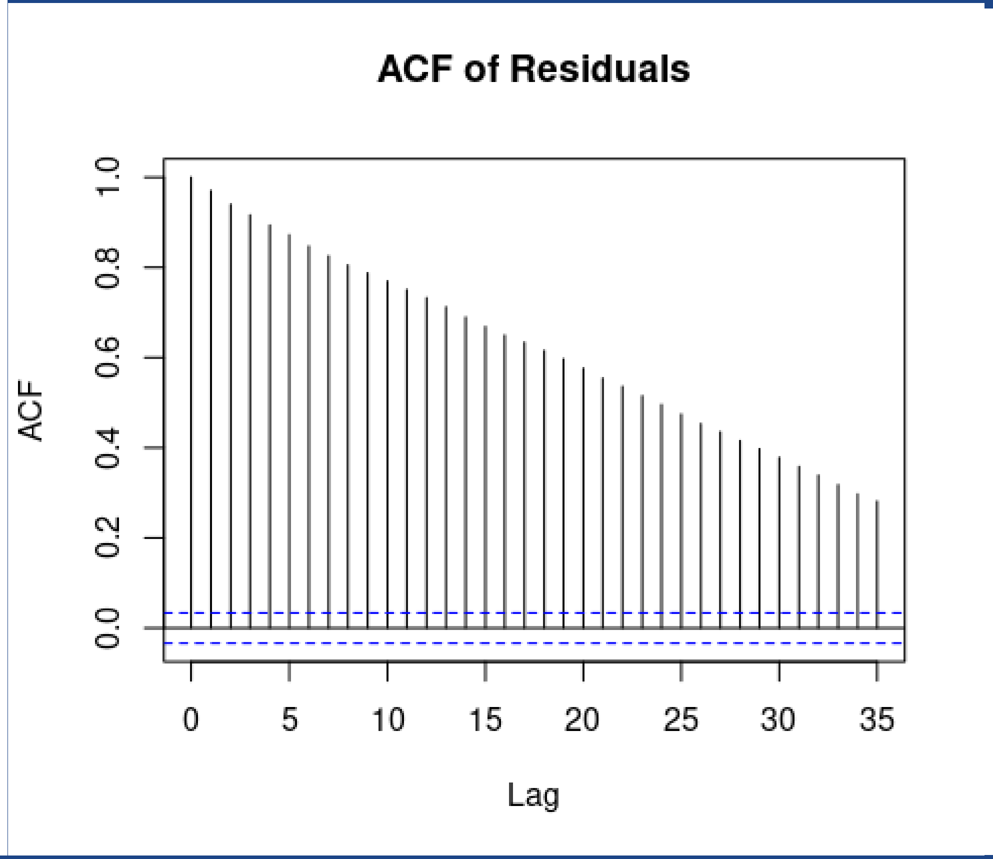


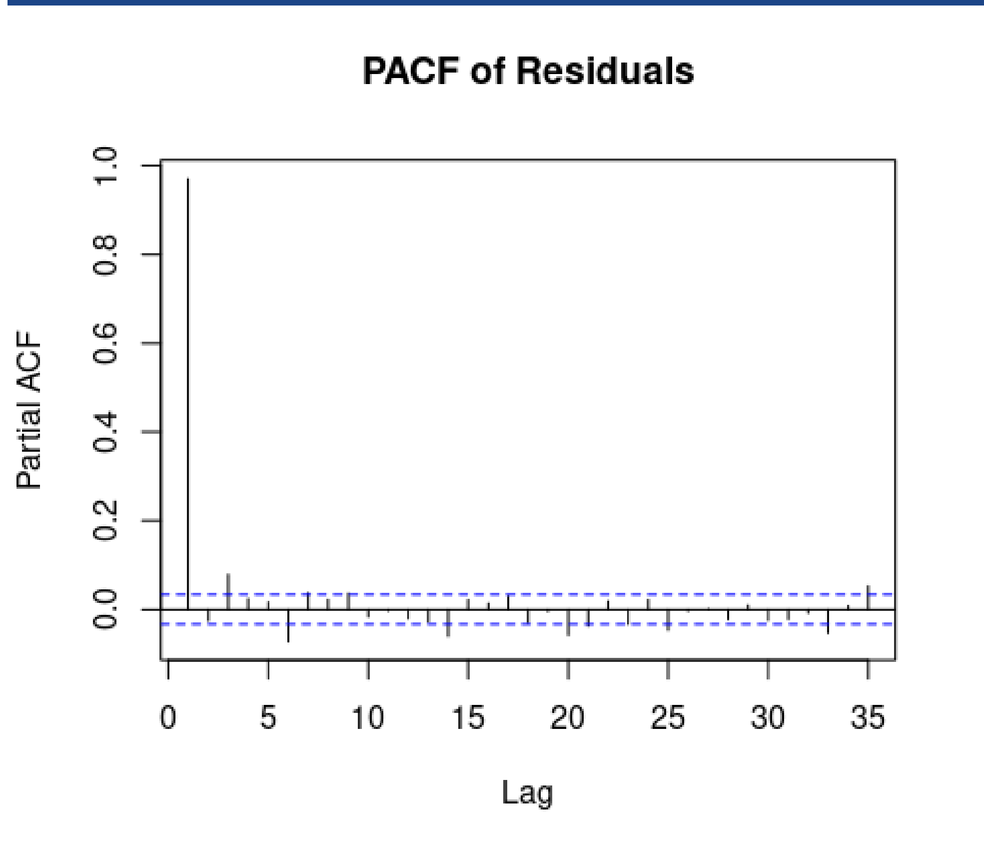
RMSE: 2.685305

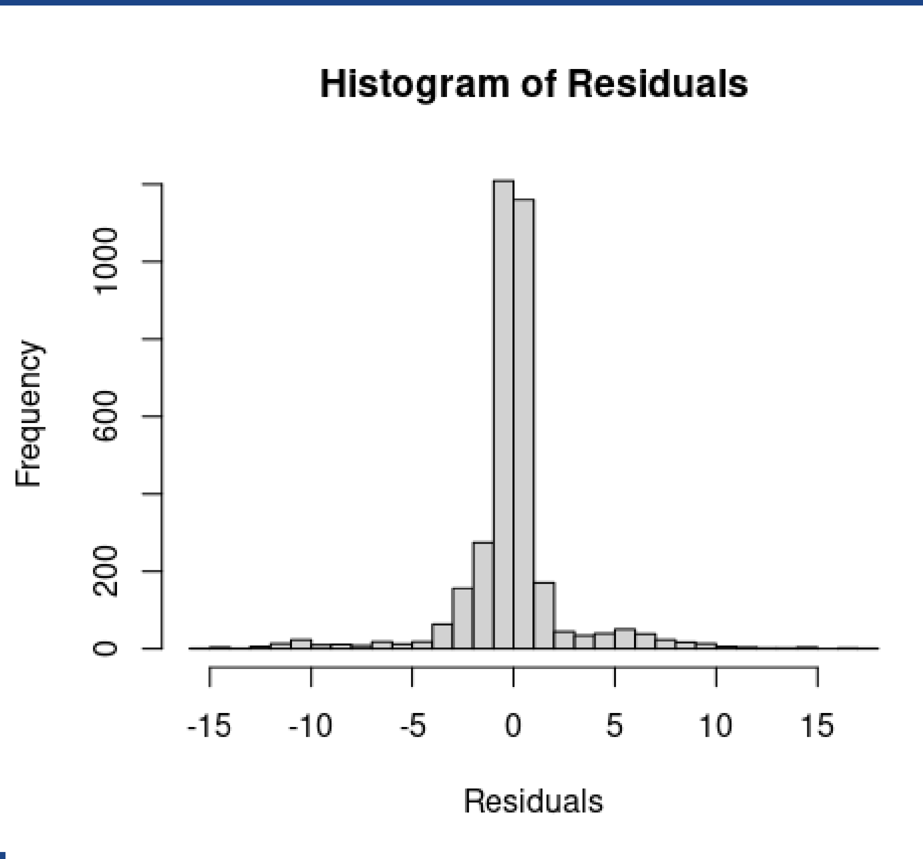
MAPE: 8.412563 %

LOESS(modeling) - Residuals









Box-Pierce test: p-value < 2.2e-16

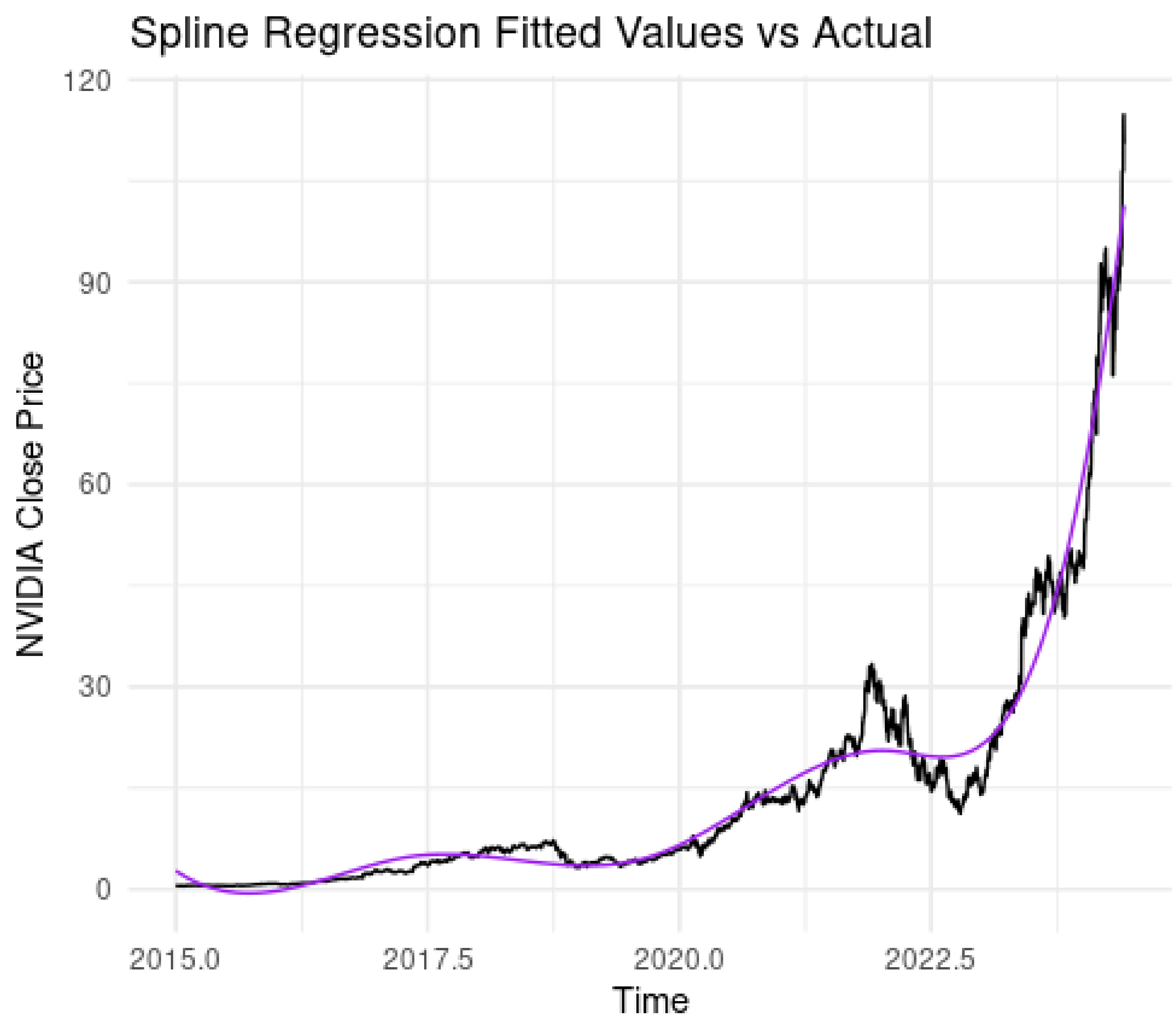
Durbin-Watson test: DW = 0.052321

Positive autocorrelation in the residuals

Shapiro-Wilk normality test:
W = 0.72926

the data deviates from a normal distribution

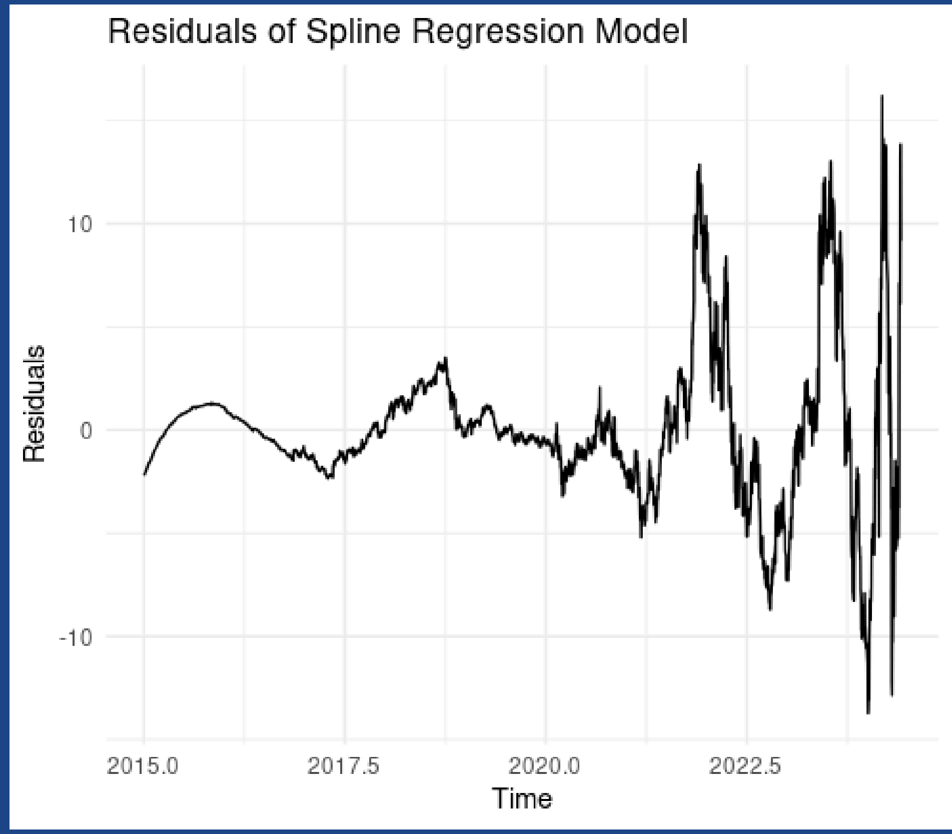
Spline Regression(modeling)

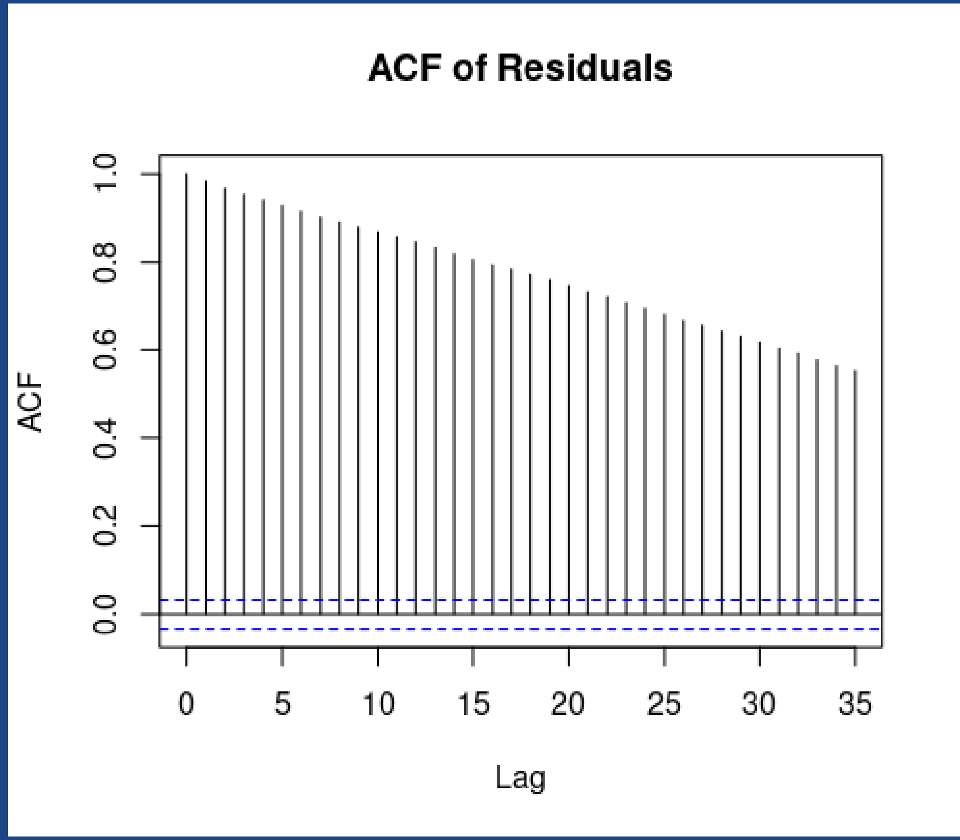


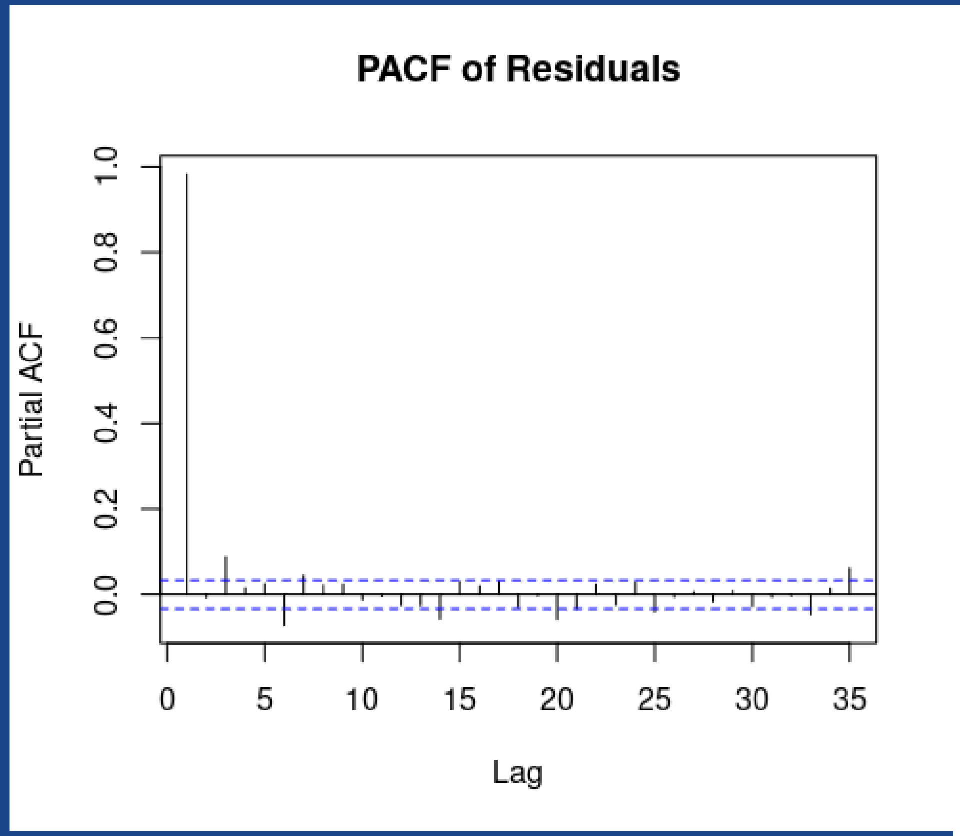
RMSE: 3.475954

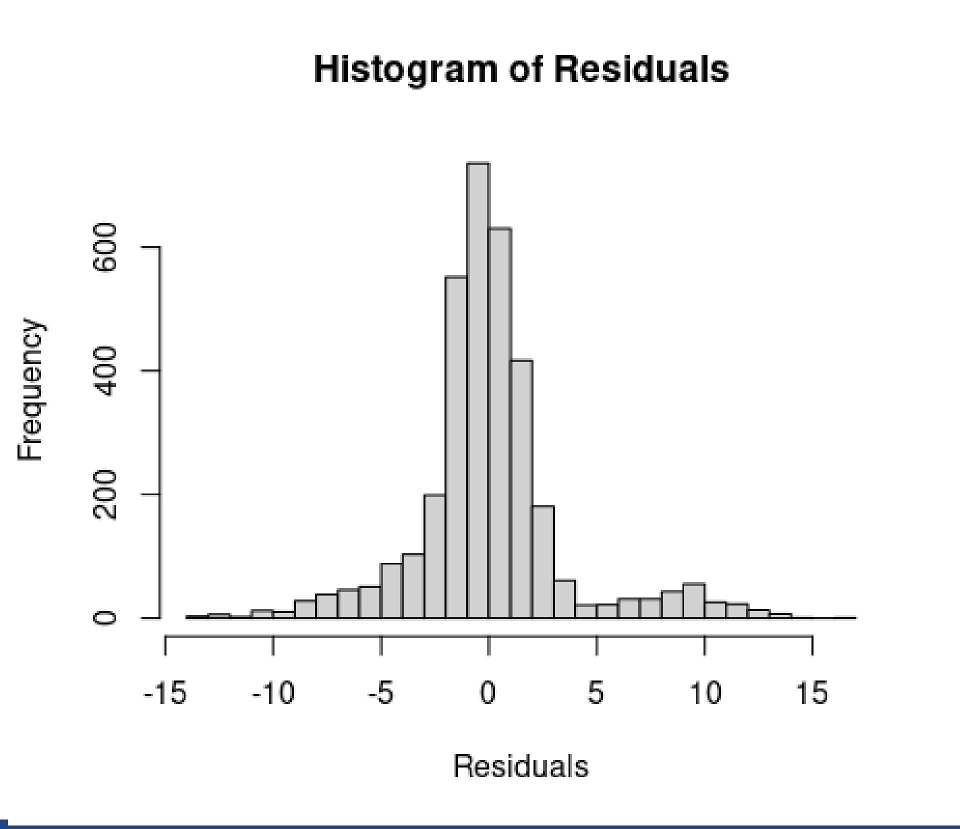
MAPE: 38.08316 %

Spline Regression(modeling) - Residuals









Box-Pierce test: p-value < 2.2e-16
Durbin-Watson test: DW = 0.031206
Positive autocorrelation in the residuals

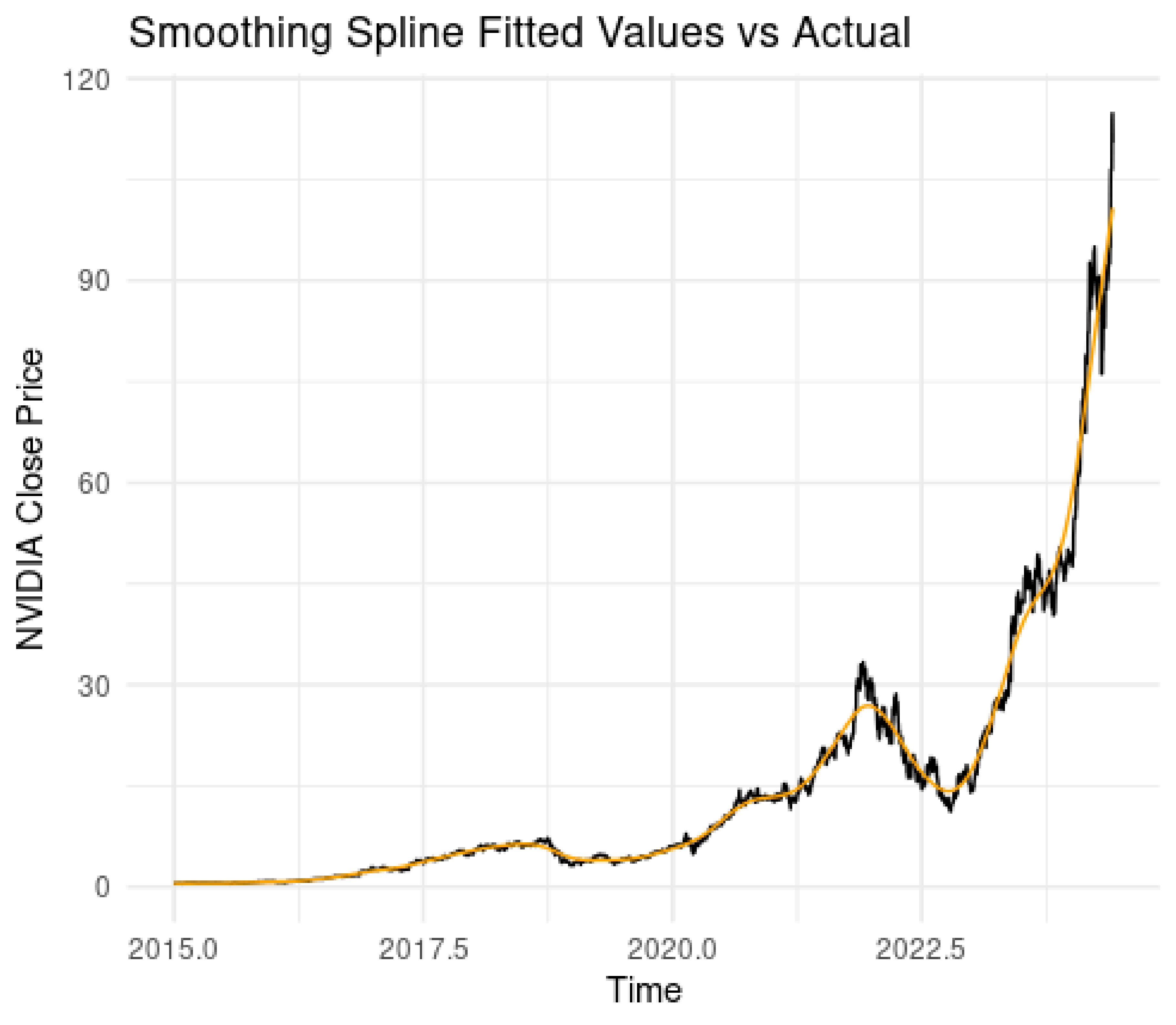
Shapiro-Wilk normality test:
W = 0.89017

the data deviates from a normal distribution

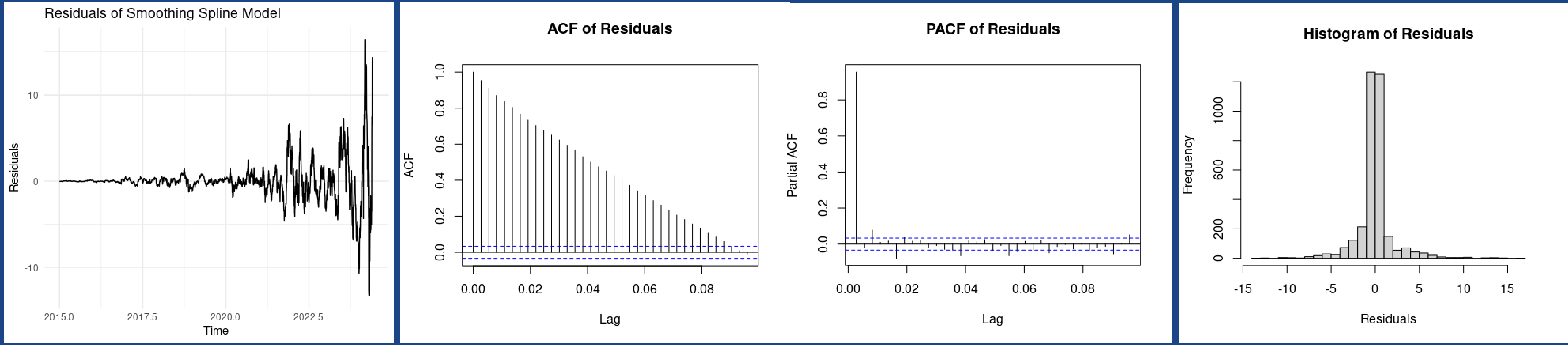
Smoothing Spline (Modeling)

RMSE: 2.084678

MAPE: 6.924252 %



Smoothing Spline (Modeling)- Residuals



Box-Pierce test: p-value < 2.2e-16

Durbin-Watson test: DW = 0.031206

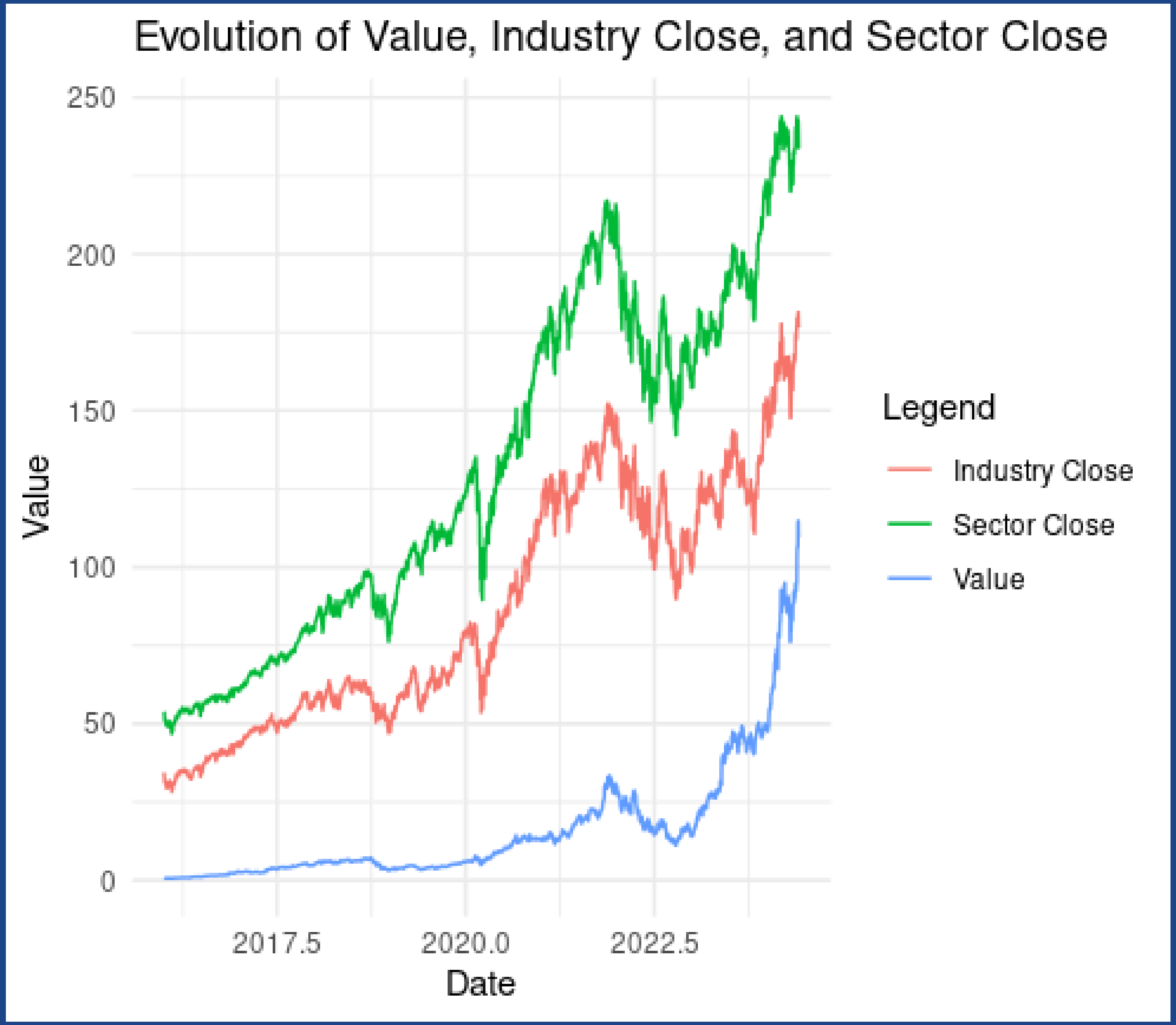
Positive autocorrelation in the residuals

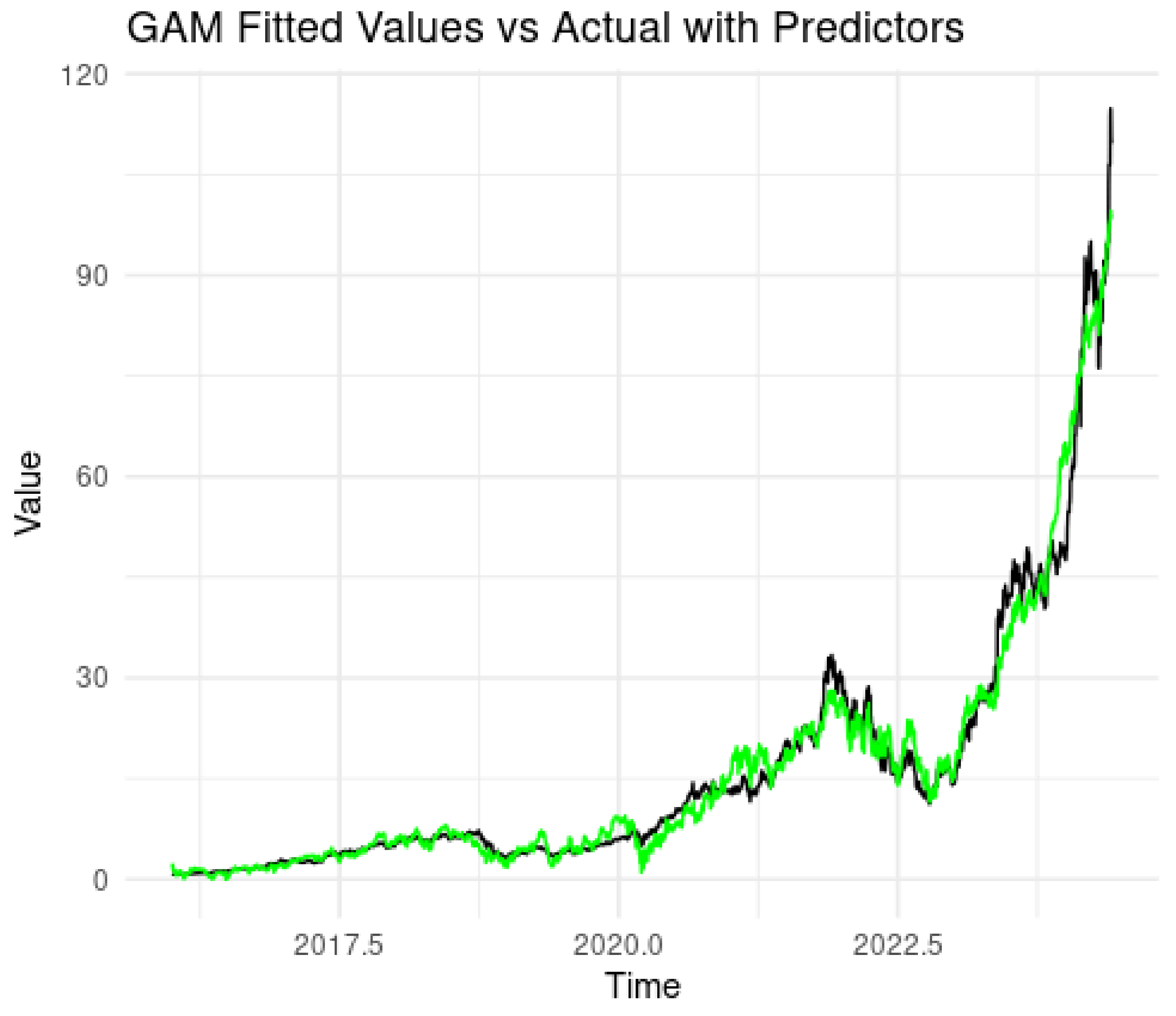
Shapiro-Wilk normality test:
W = 0.89017

the data deviates from a normal distribution

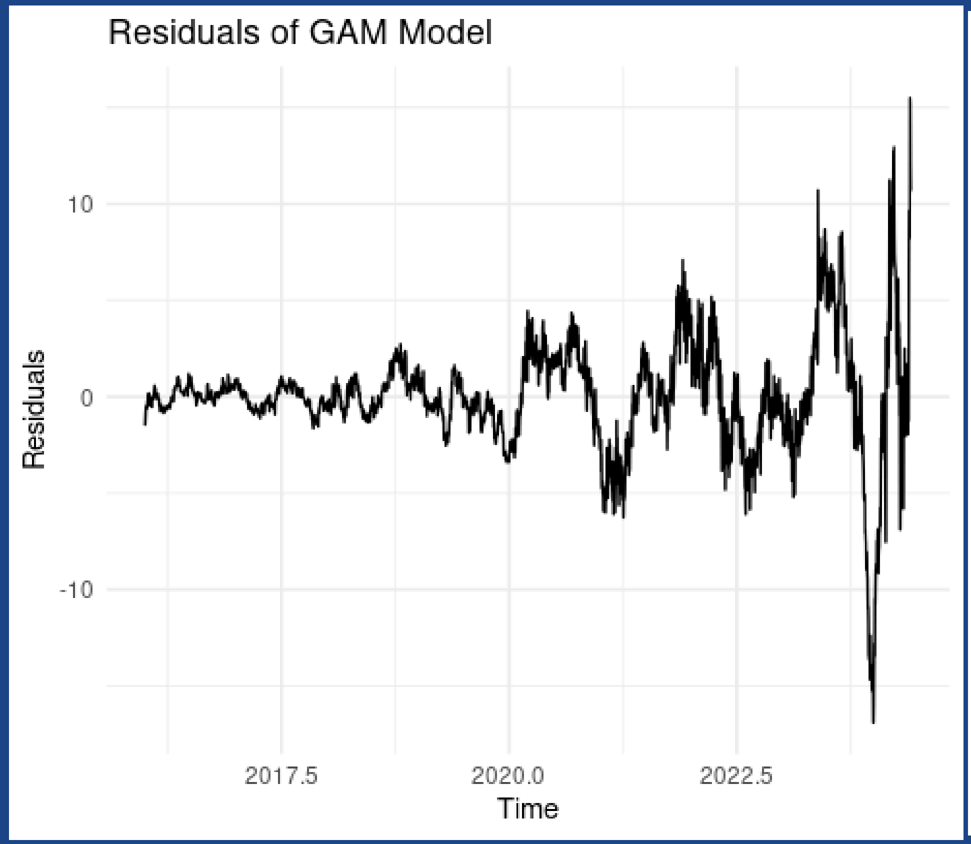
GAM with Tech and Semiconductors

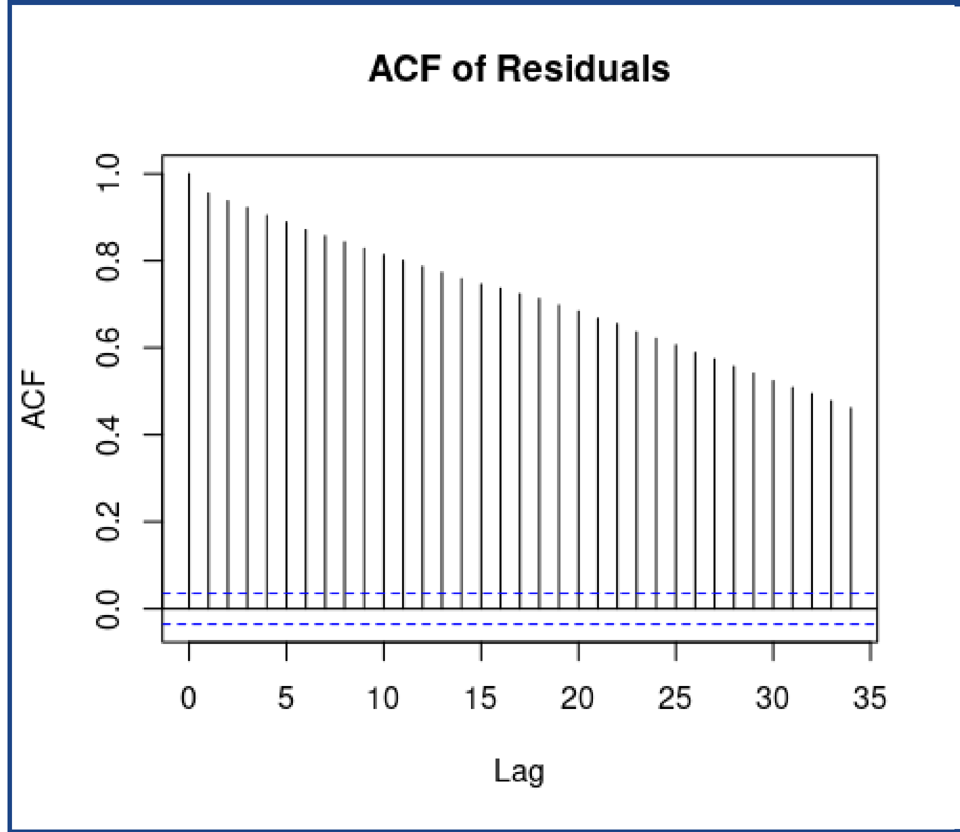
MAPE: 17.09476 % | RMSE: 2.875411 | AIC: 15228.23

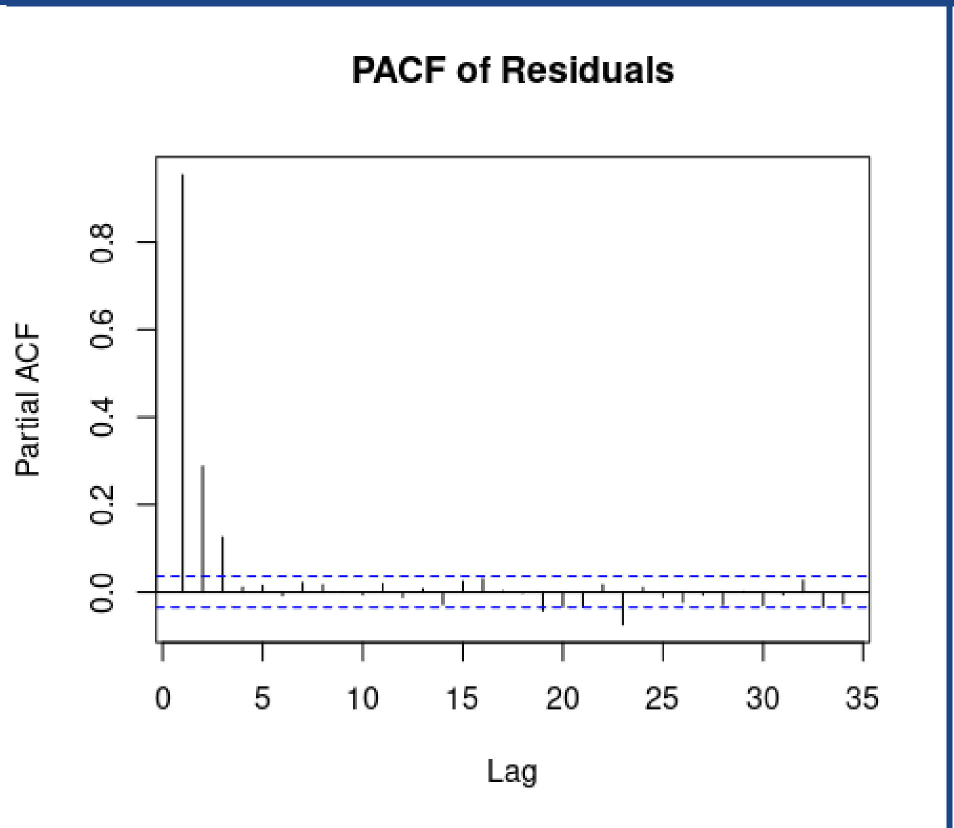


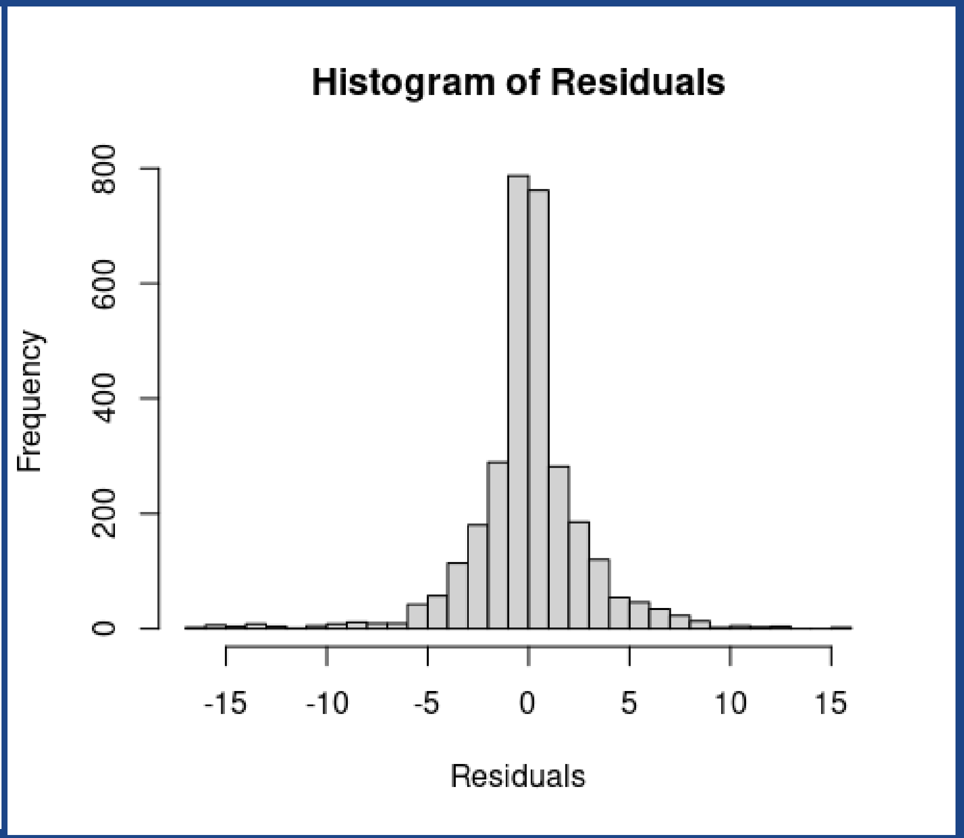


GAM with Tech and Semiconductors - Residuals









Box-Pierce test: p-value < 2.2e-16
Durbin-Watson test: DW = 0.085795
Positive autocorrelation in the residuals

Shapiro-Wilk normality test:
W = 0.88836

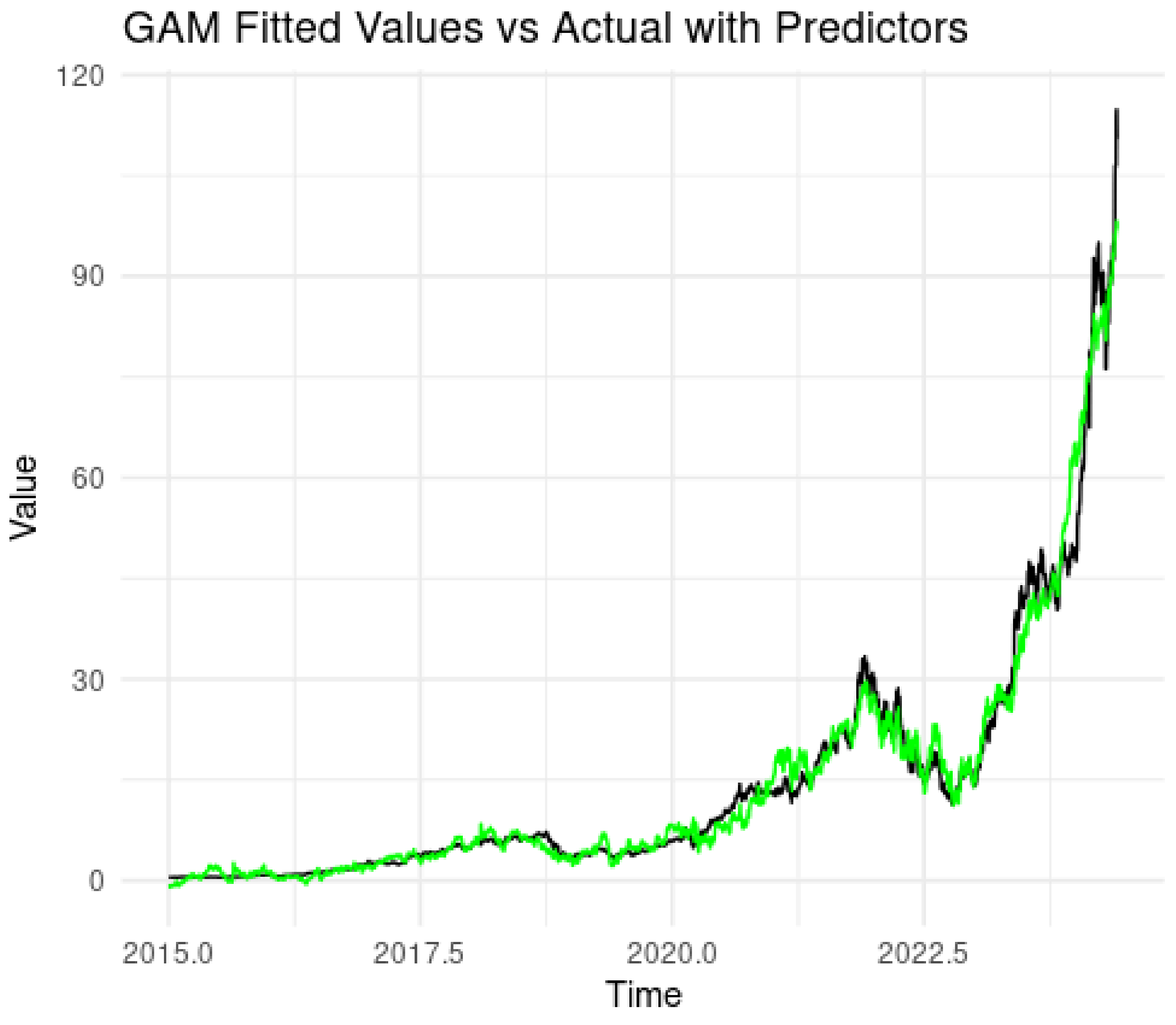
the data deviates from a normal distribution

GAM with Tech, Semiconductors and VIX

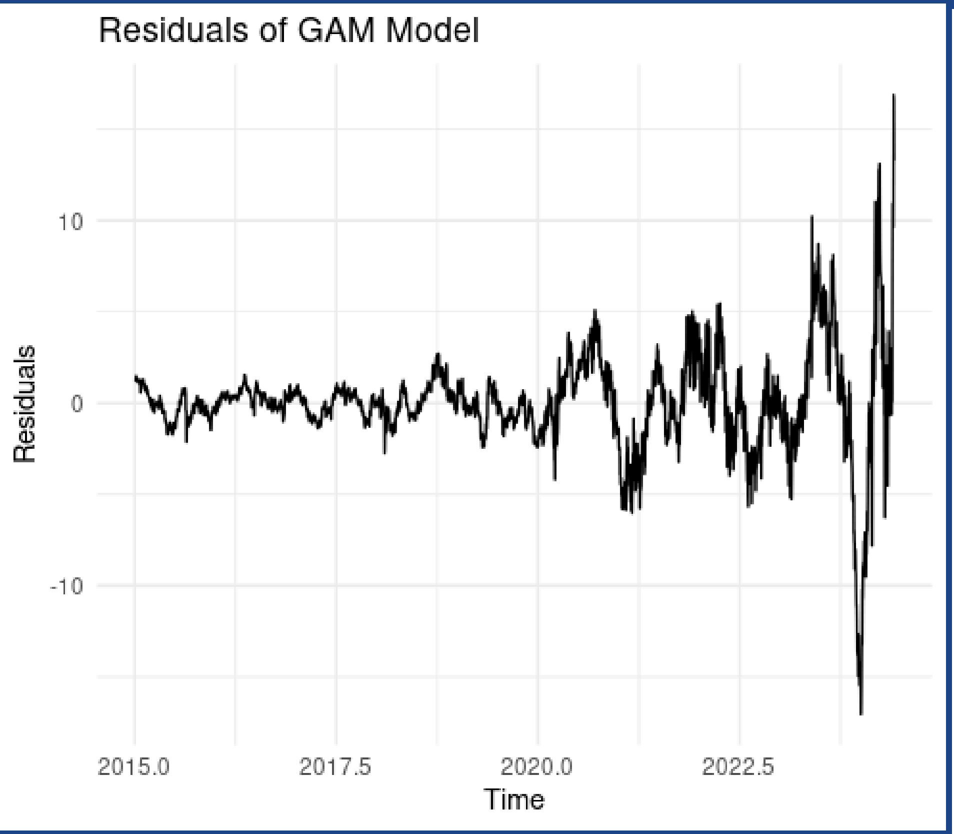
MAPE: 27.32582 %

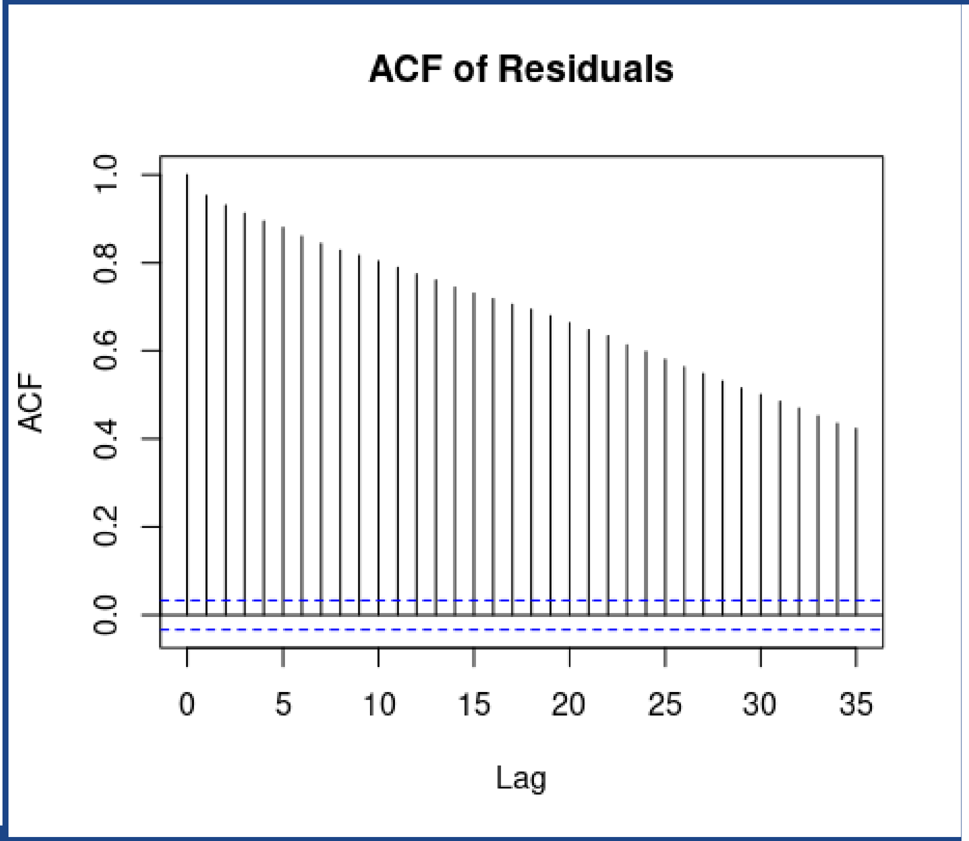
RMSE: 2.655296

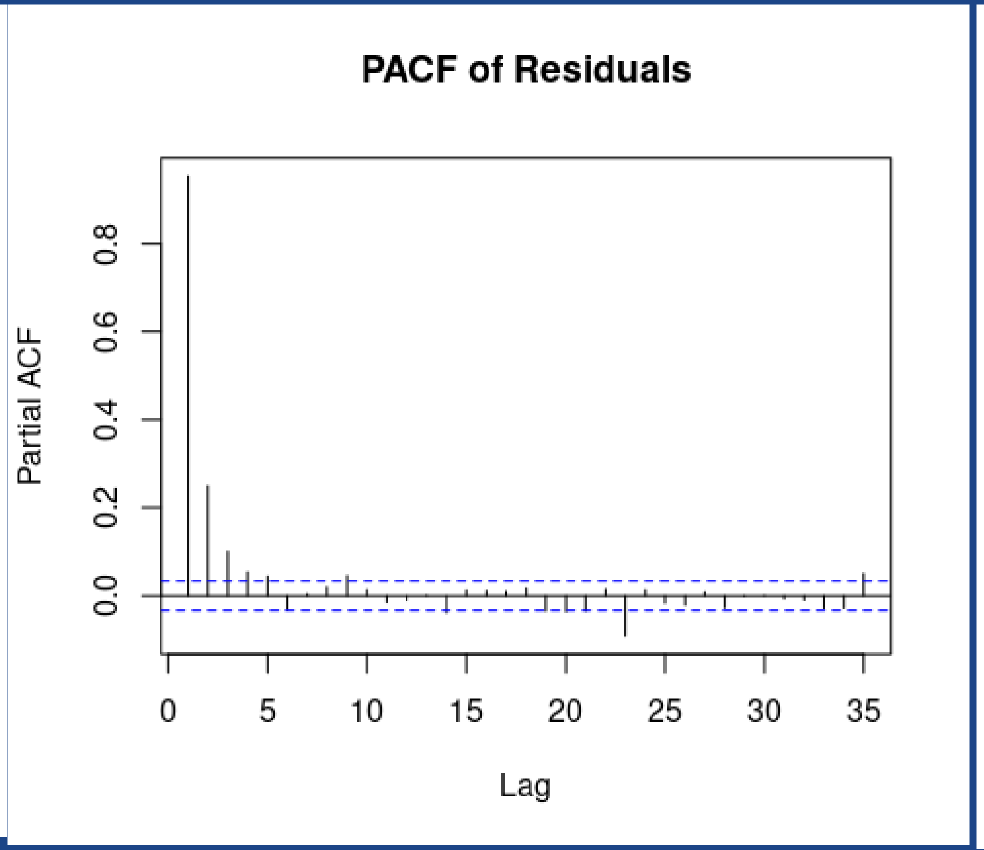
AIC: 16489.82

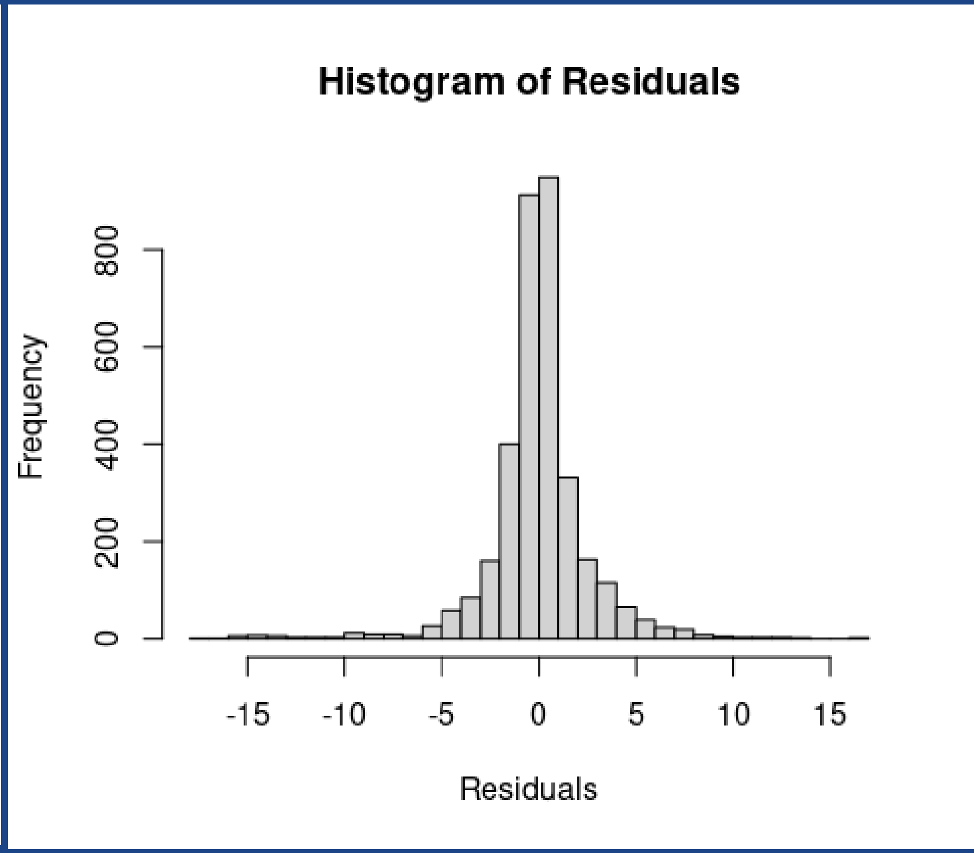


GAM with Tech, Semiconductors and VIX - Residuals









Box-Pierce test: p-value < 2.2e-16

Durbin-Watson test: DW = 0.087102

Positive autocorrelation in the residuals

Shapiro-Wilk normality test:
W = 0.85153

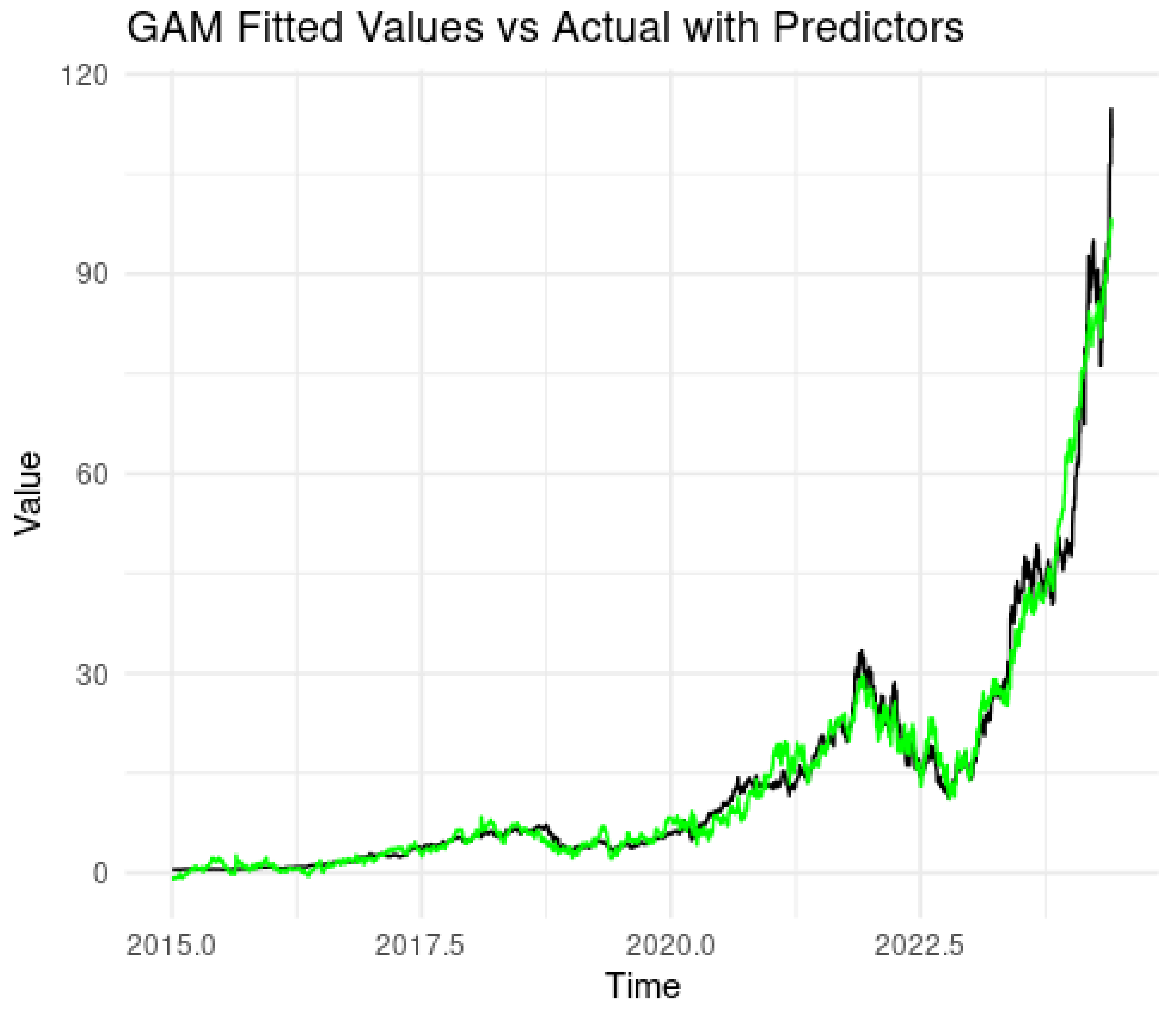
the data deviates from a normal distribution

◆◆ GAM with Tech, Semiconductors VIX and BTC

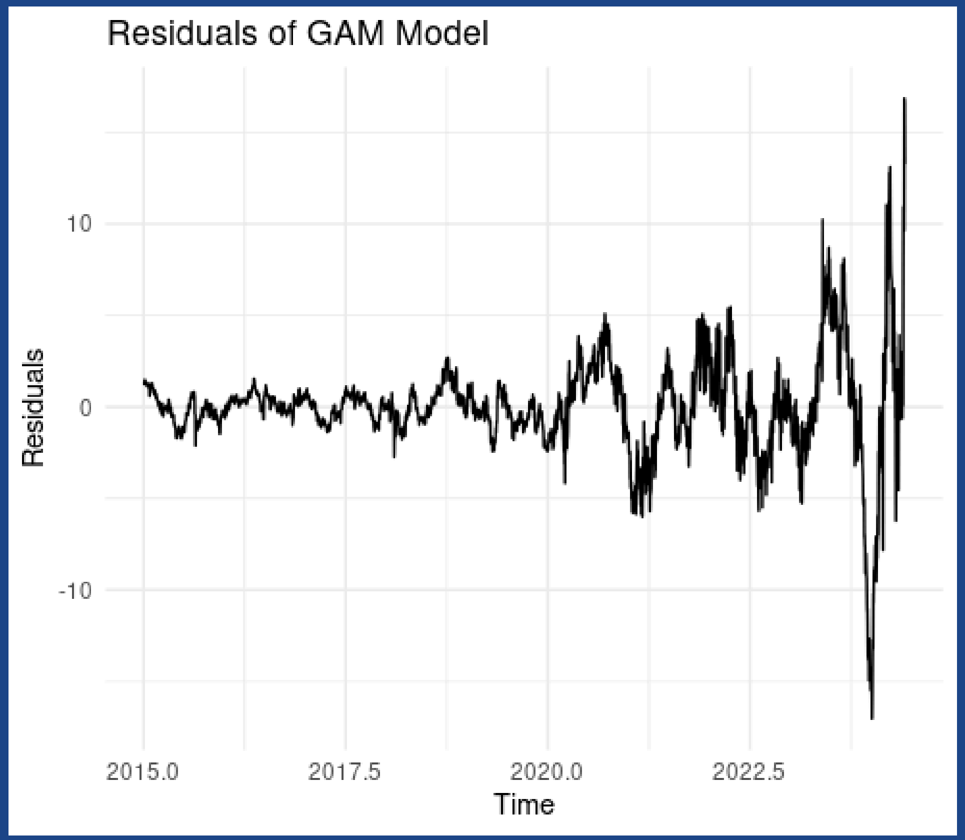
MAPE: 27.34299 %

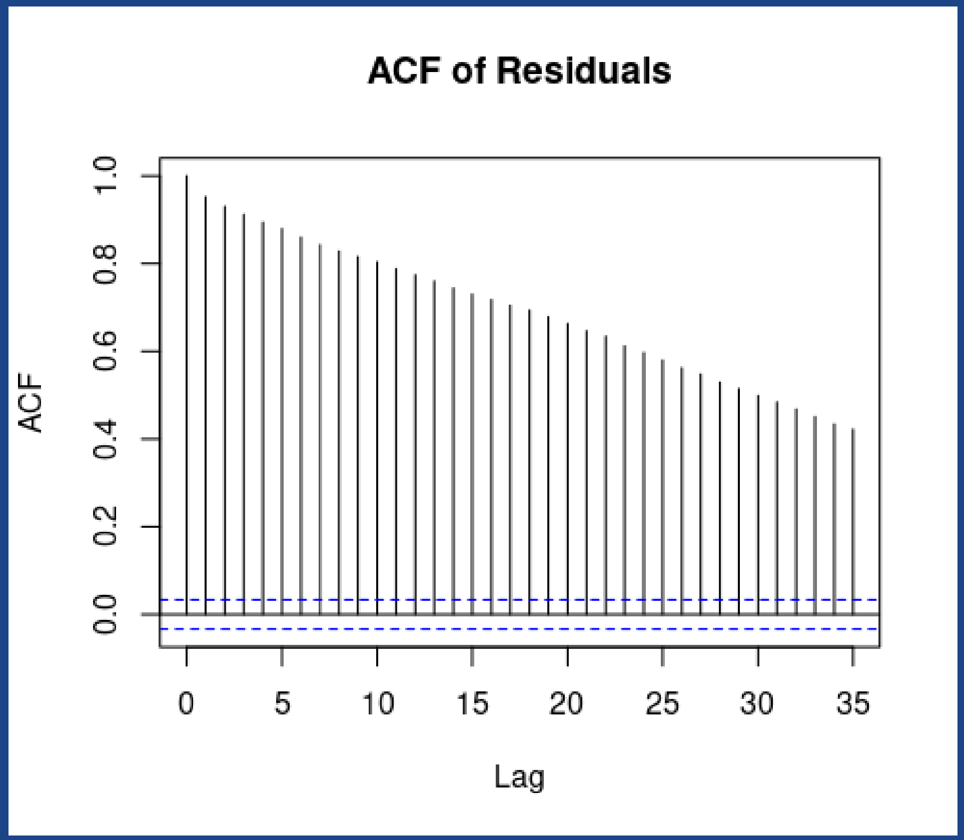
RMSE: 2.65525

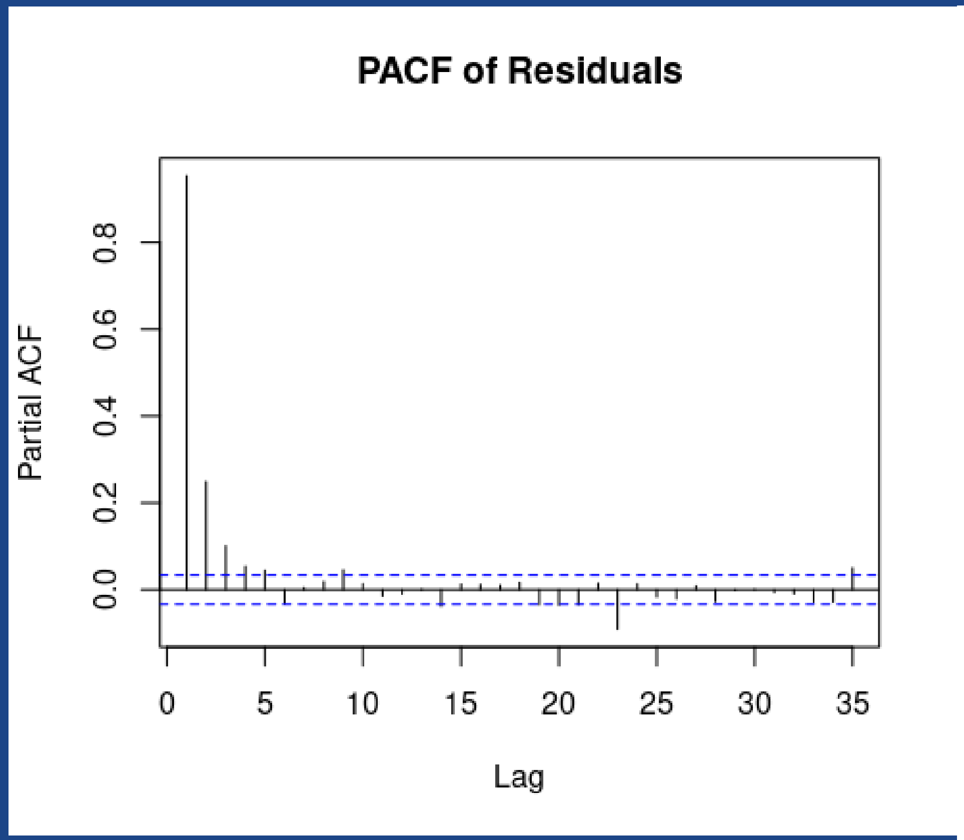
AIC: 16491.7

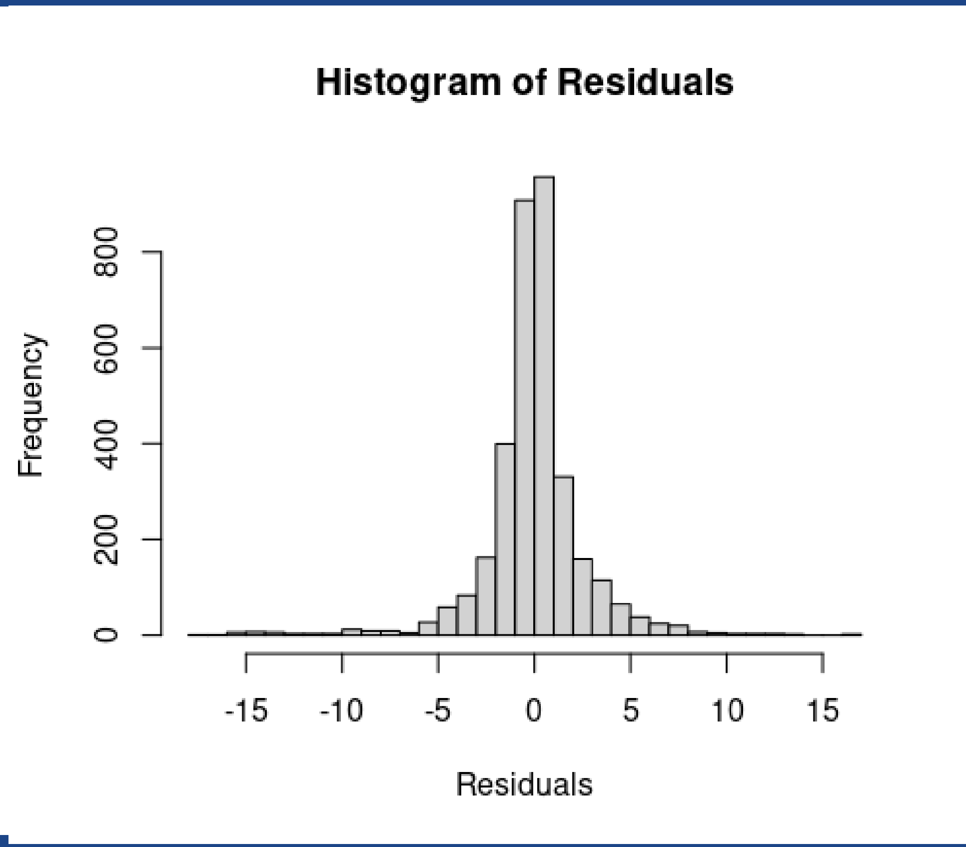


GAM with Tech, Semiconductors, VIX and BTC- Residuals









Box-Pierce test: p-value < 2.2e-16

Durbin-Watson test: DW = 0.087243

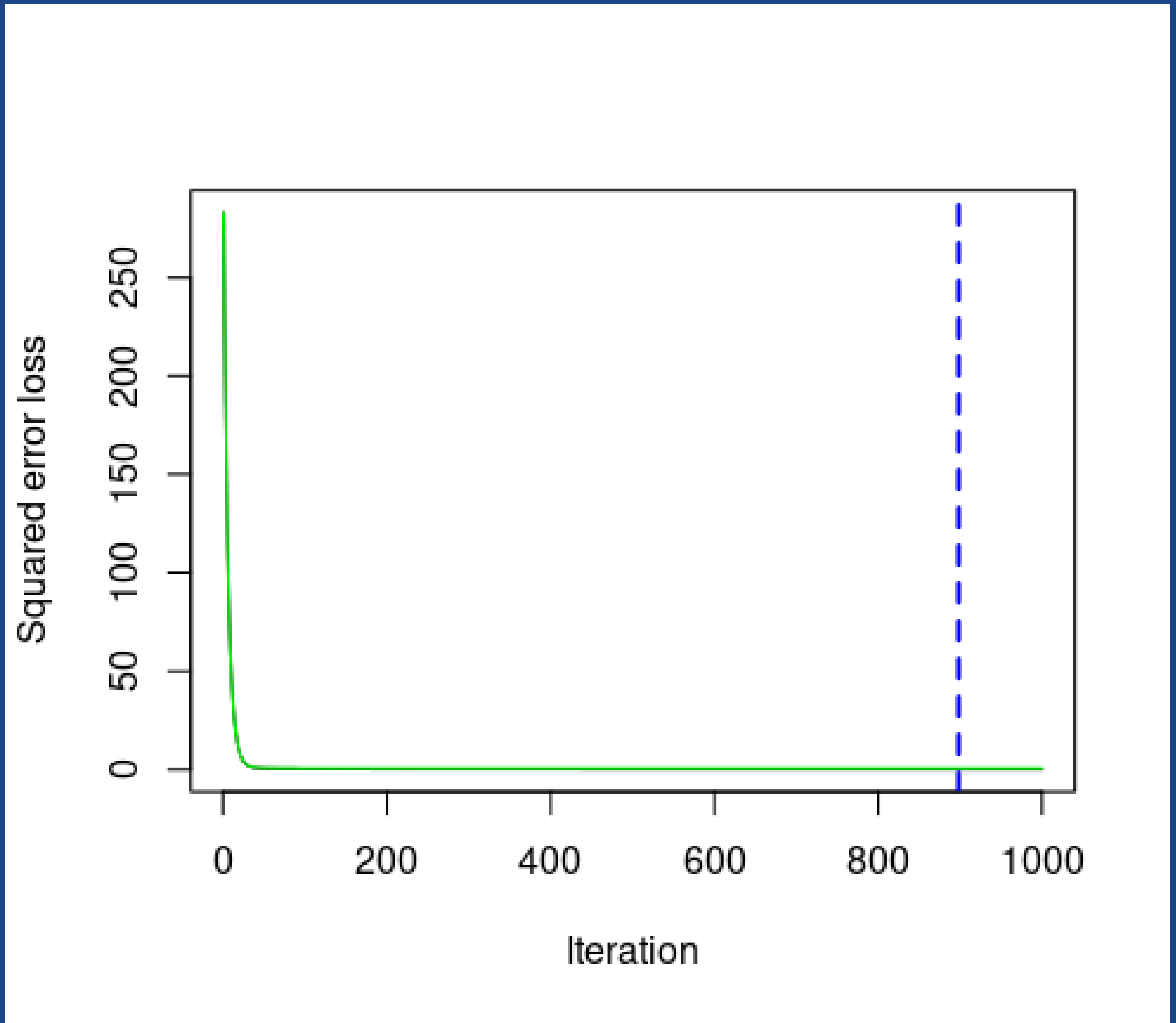
Positive autocorrelation in the residuals

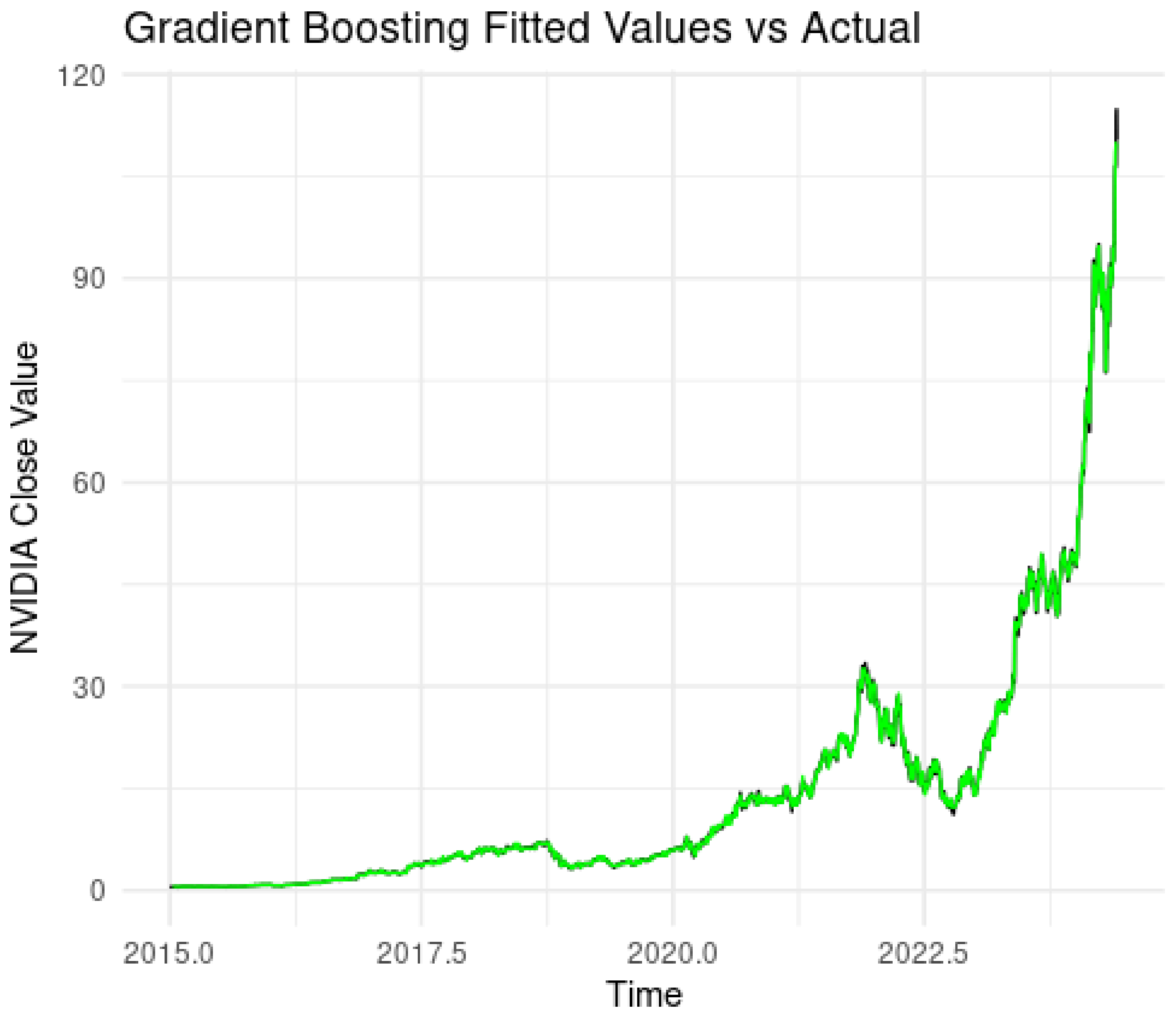
Shapiro-Wilk normality test:
W = 0.85096

the data deviates from a normal distribution

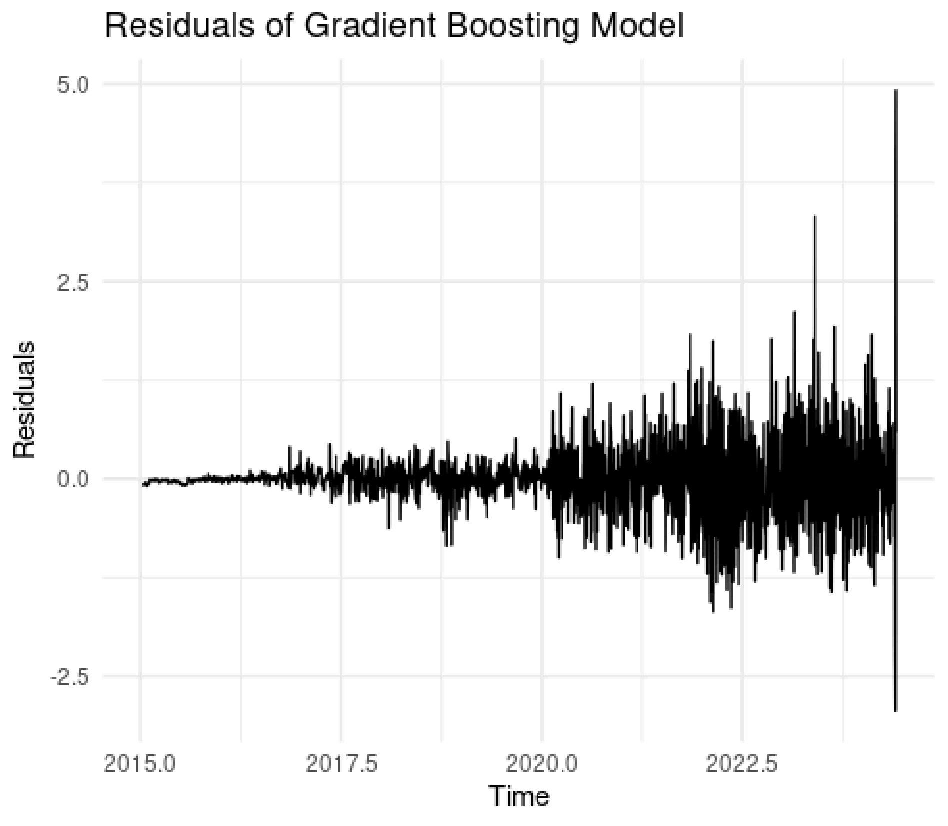
Gradient Boosting

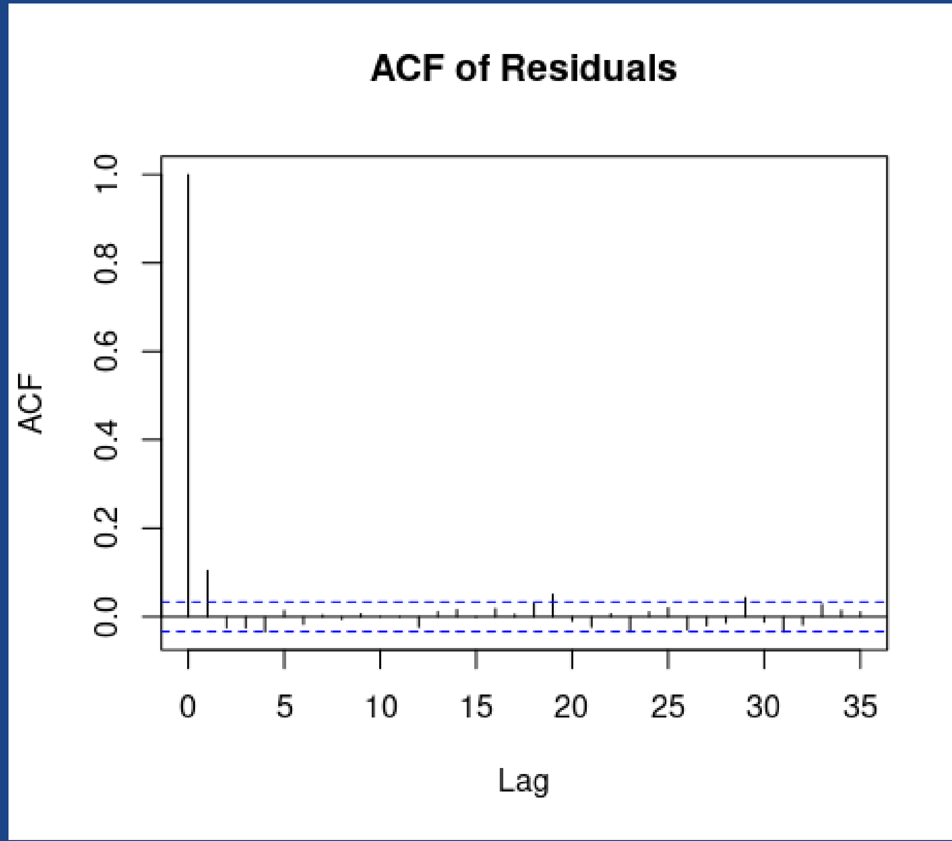
MAPE: 2.469149 % | RMSE: 0.3627375

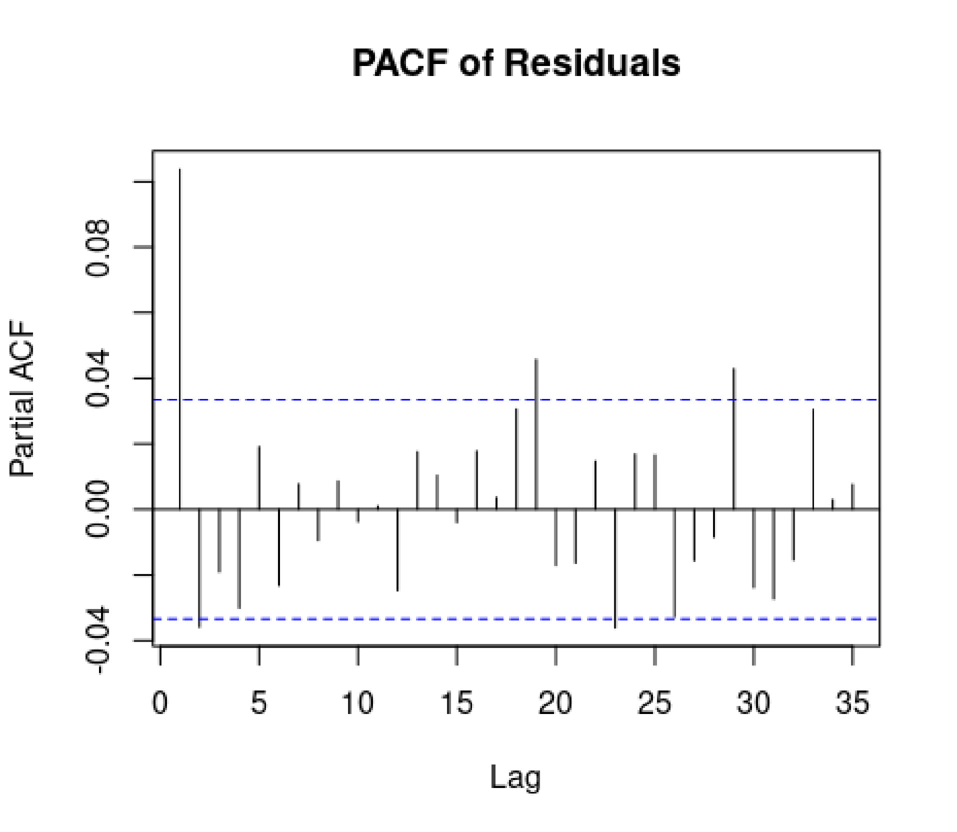


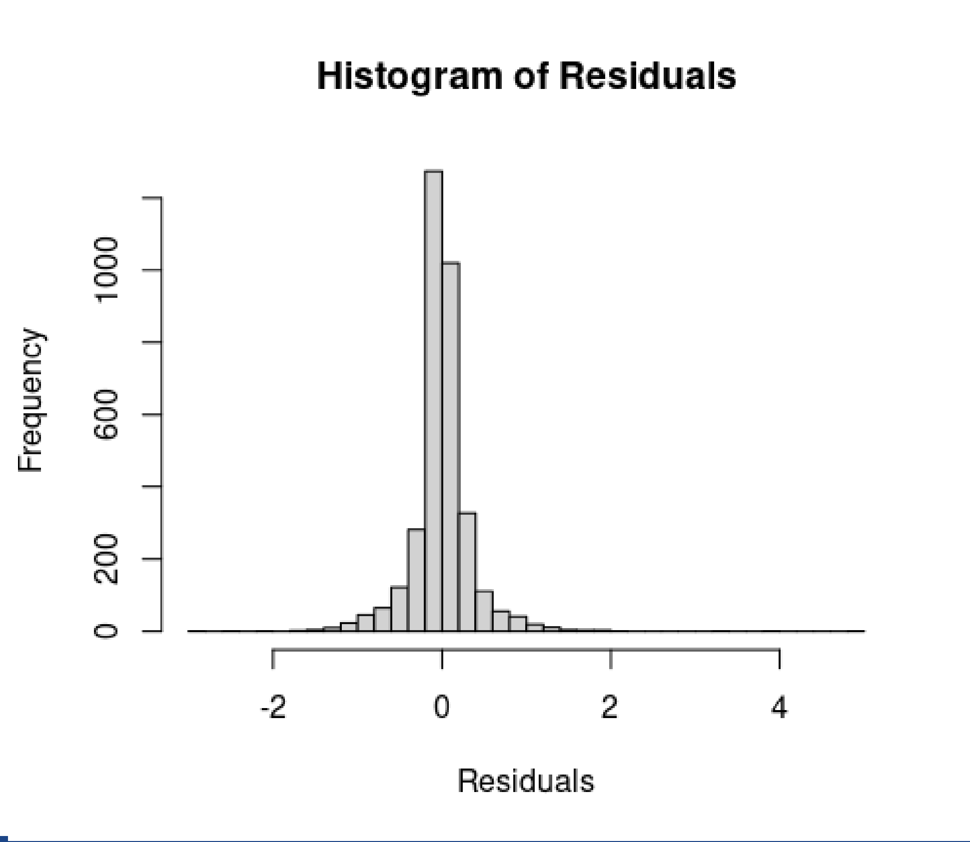


Gradient Boosting - Residuals









Box-Pierce test: p-value < 2.067e-06
Durbin-Watson test: DW = 1.7919
Positive autocorrelation in the residuals

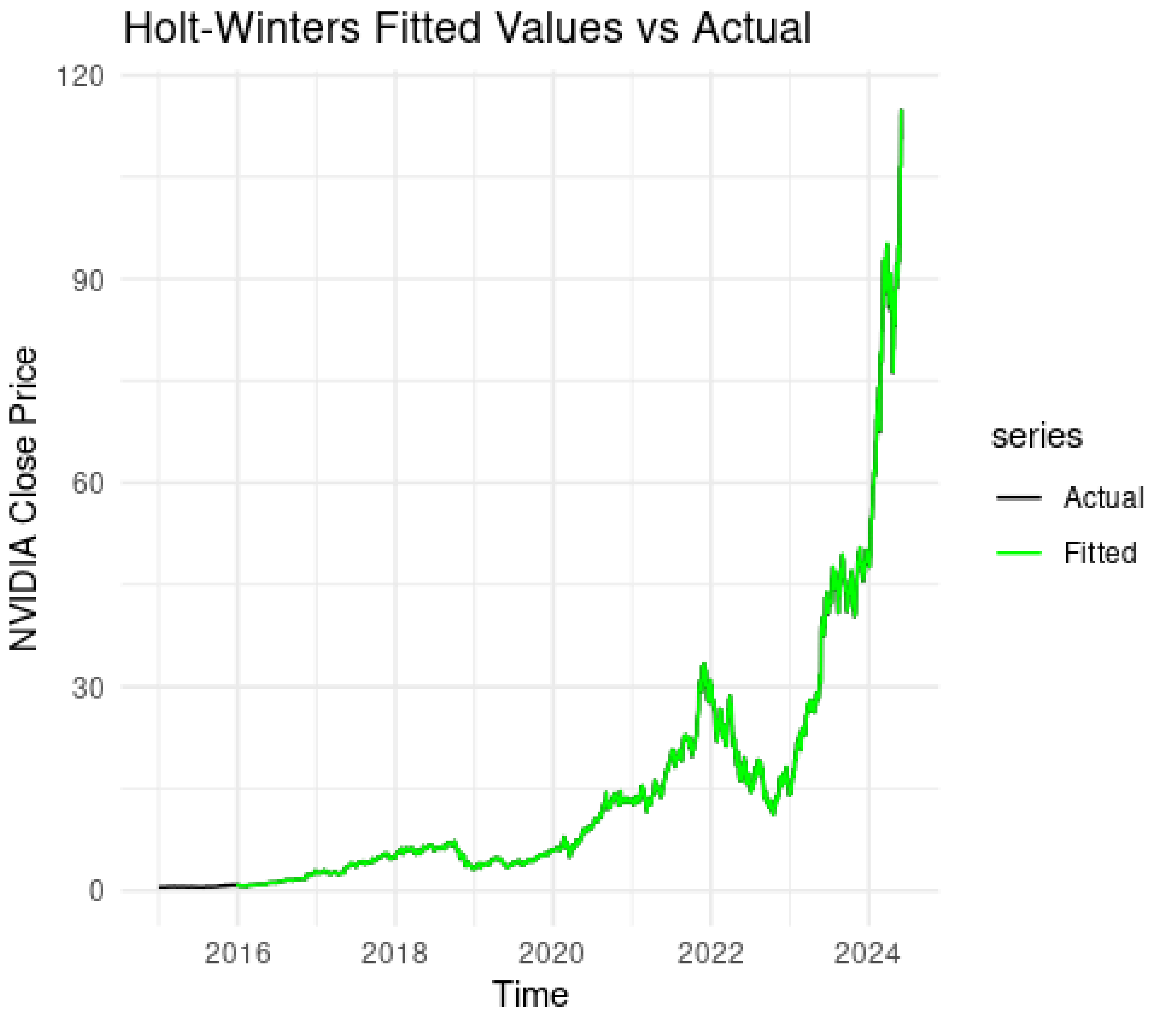
Shapiro-Wilk normality test:
W = 0.82237

the data deviates from a normal distribution

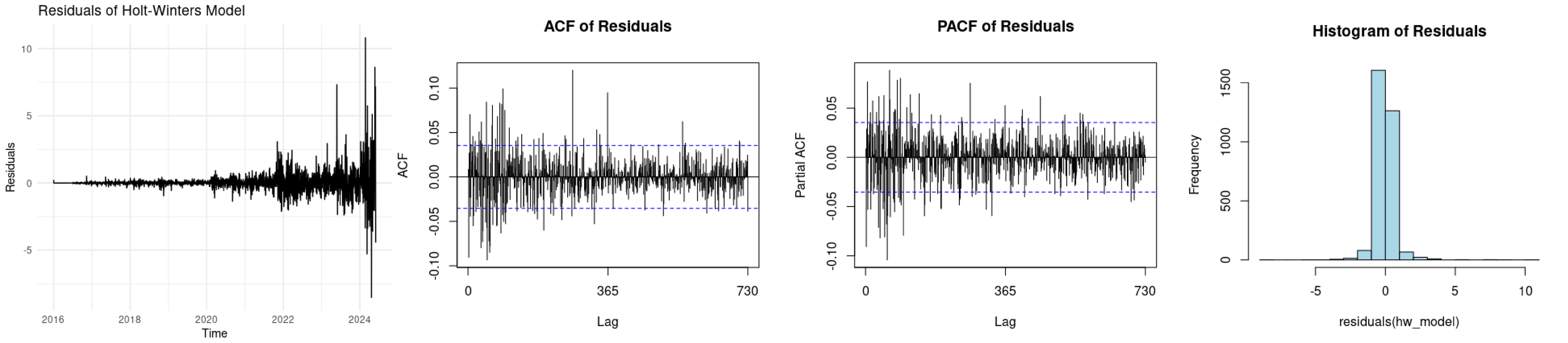
Holt - Winter's Exponential Smoothing

MAPE: 1.556645 %

RMSE: 0.6512679



Holt - Winter's Exponential Smoothing - Residuals



Box-Pierce test: p-value = 0.6354

Durbin-Watson test: DW = 2

No autocorrelation in the residuals ✌️

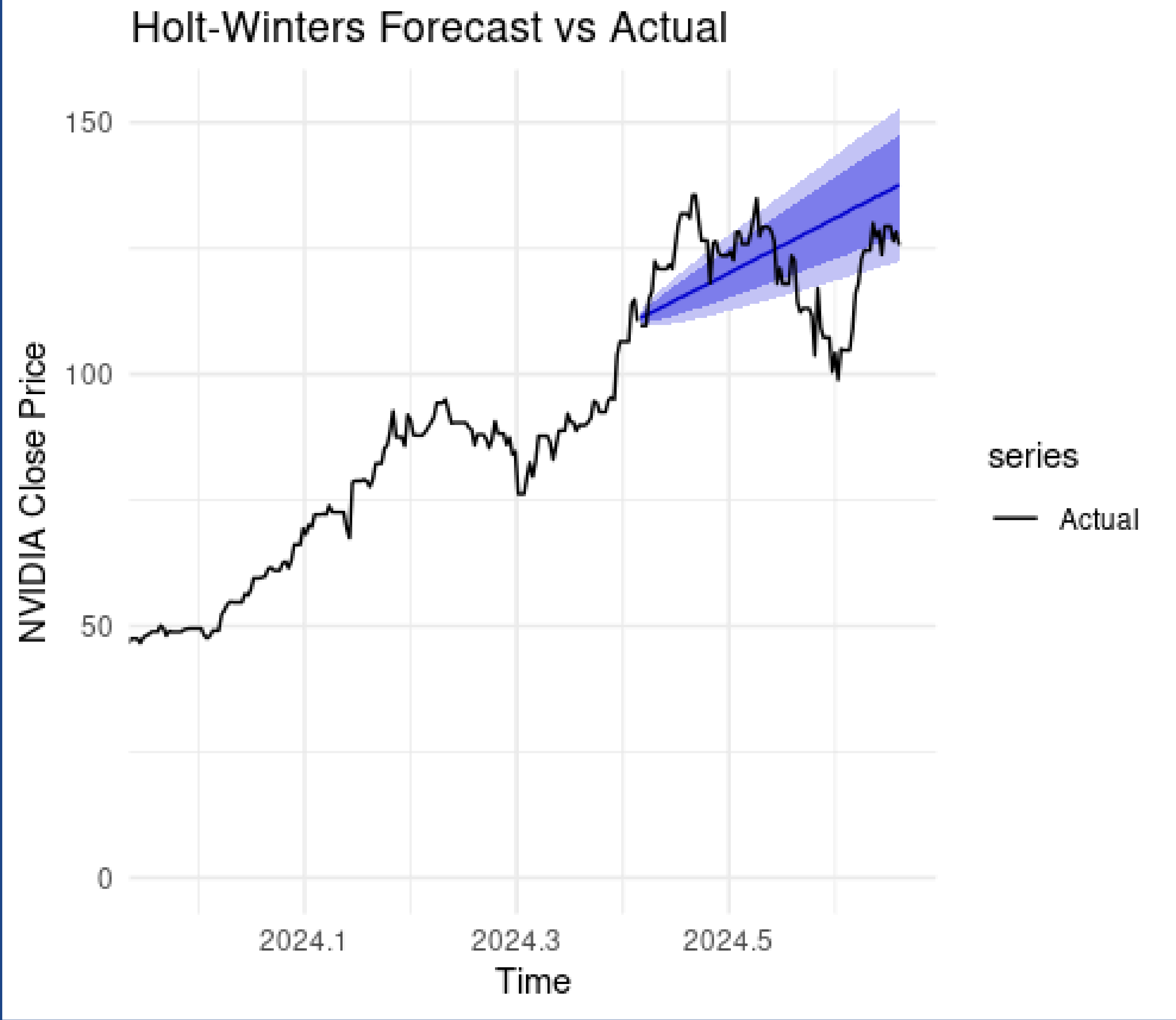
Shapiro-Wilk normality test:
W = 0.56552

the data deviates from a normal distribution

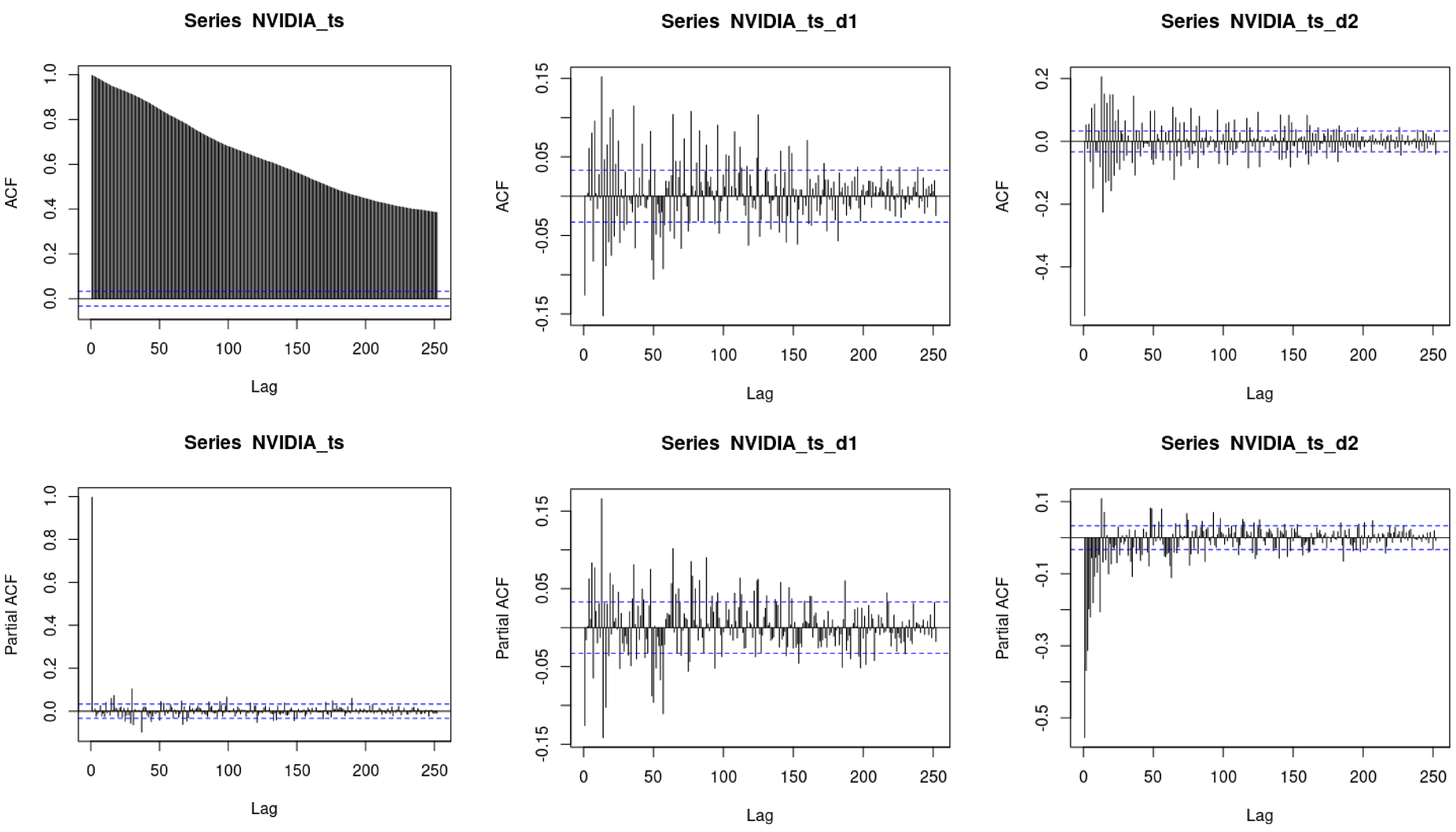
◆◆ Holt - Winter's Exponential Smoothing - Forecasting

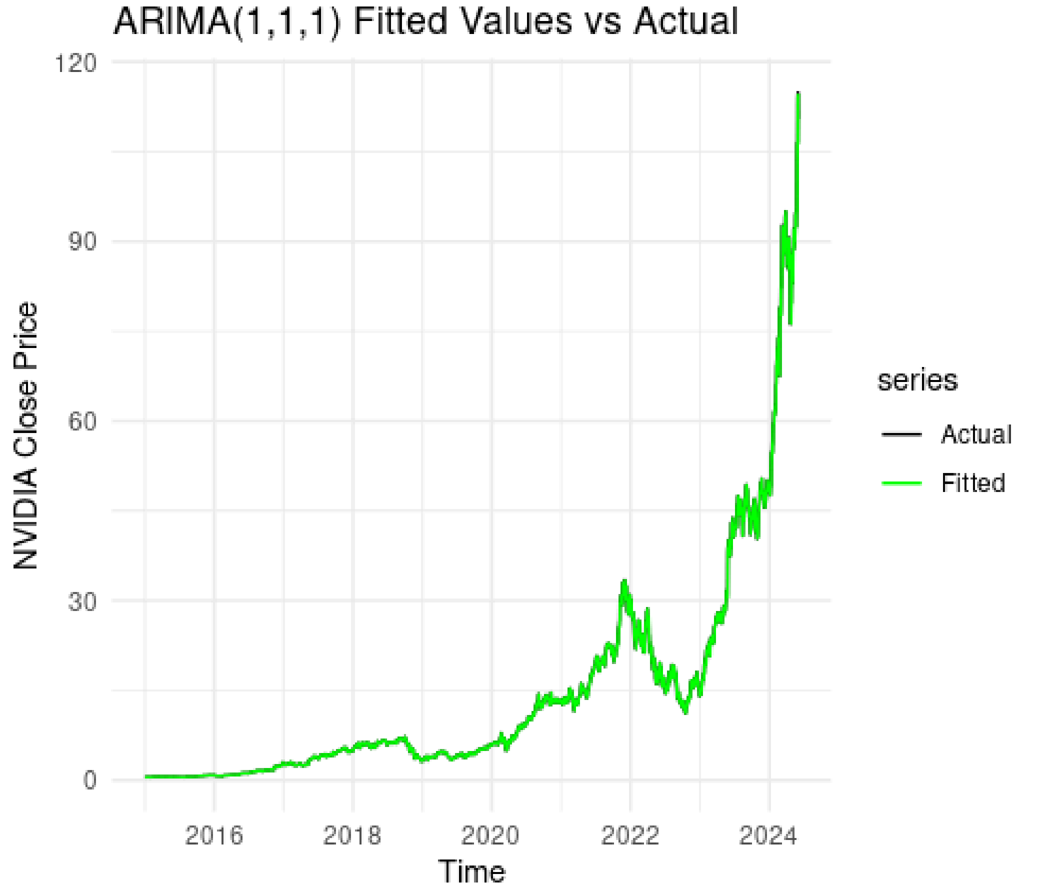
MAPE (forecast): 9.428002 %

RMSE (forecast): 13.33353



ARIMA(1,1,1)



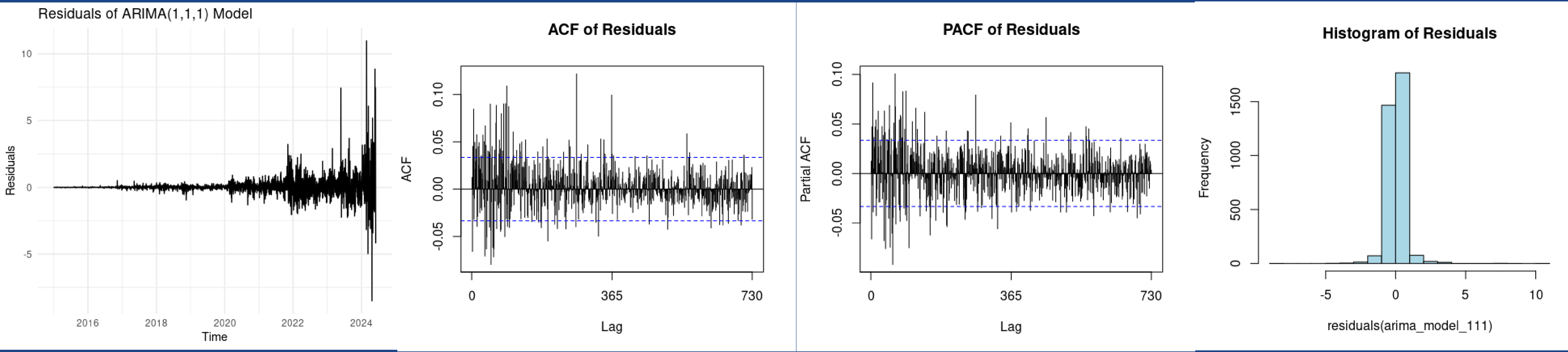


AIC: 6462.886

RMSE: 0.6190986

MAPE: 1.469759 %

ARIMA (1,1,1) - Residuals



Box-Pierce test: p-value = 3.442e-15
Durbin-Watson test: DW = 1.9614, p-value = 0.1291
Maybe ARIMA (1,1,1) is still not a good fit but let's forecast and verify

Shapiro-Wilk normality test:
W = 0.51687

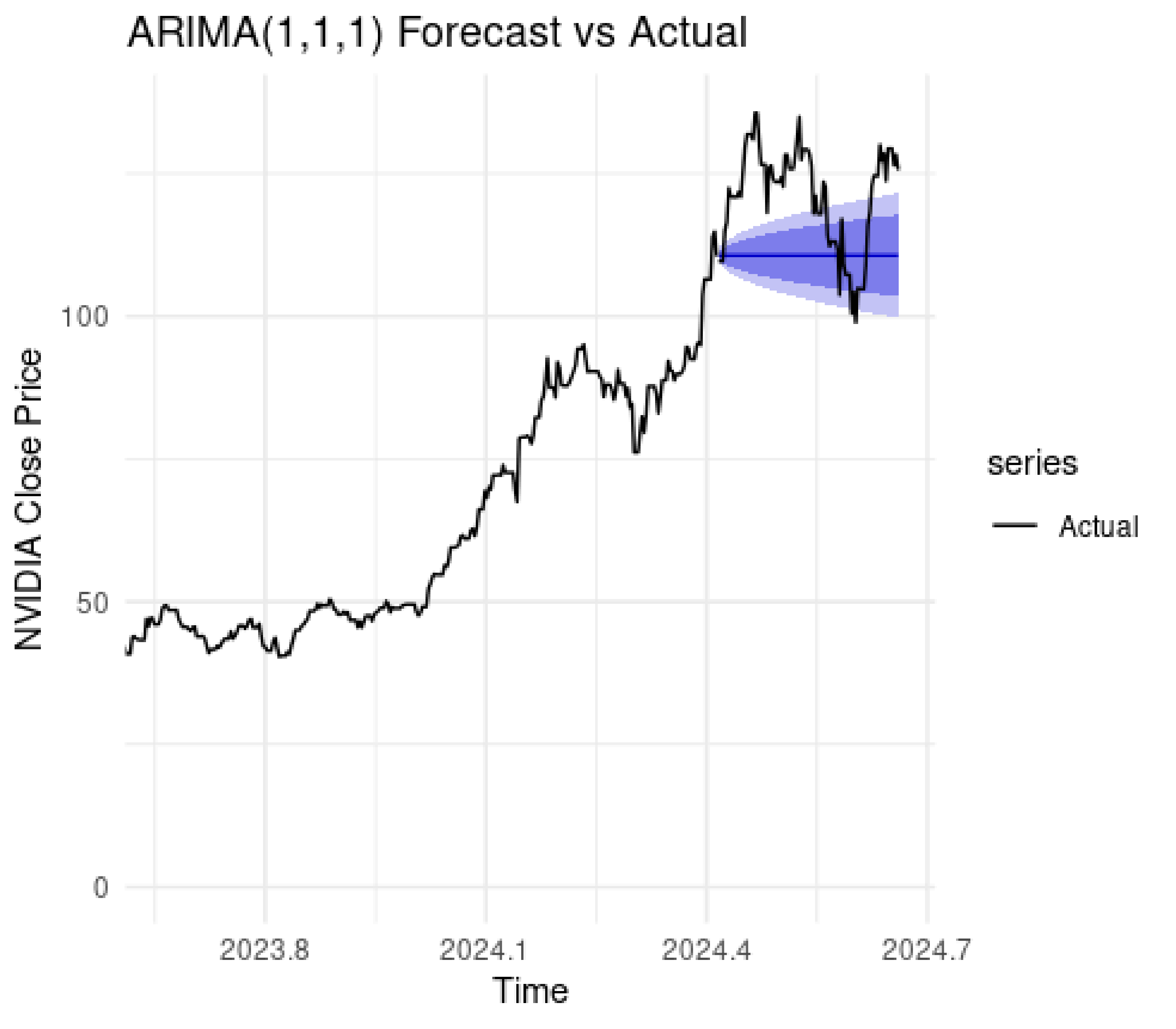
the data deviates from a normal distribution

ARIMA (1,1,1) - Forecasting

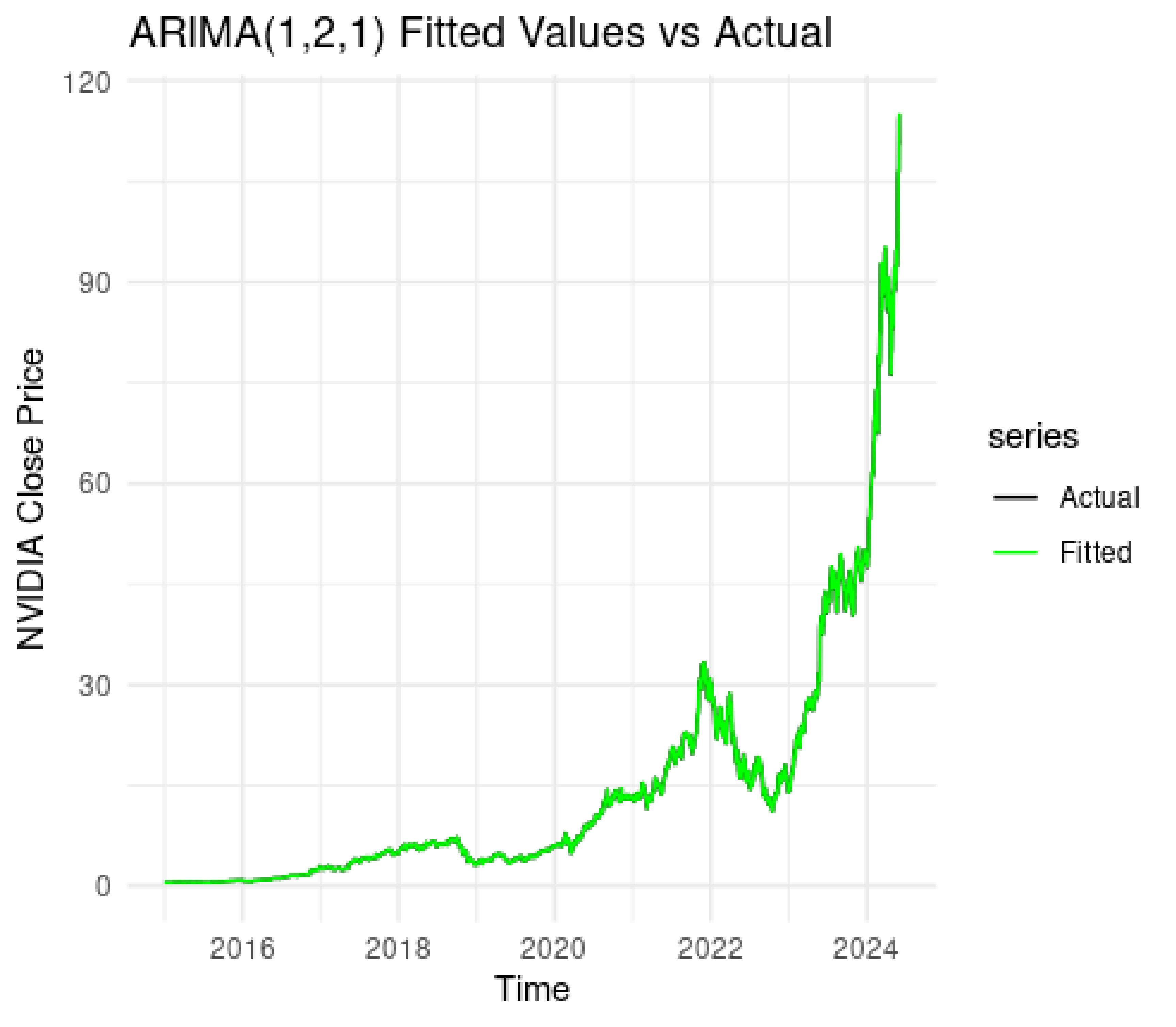
MAPE (forecast): 9.665037 %

RMSE (forecast): 13.62299

AIC (forecast): 6462.886



ARIMA (1,2,1)

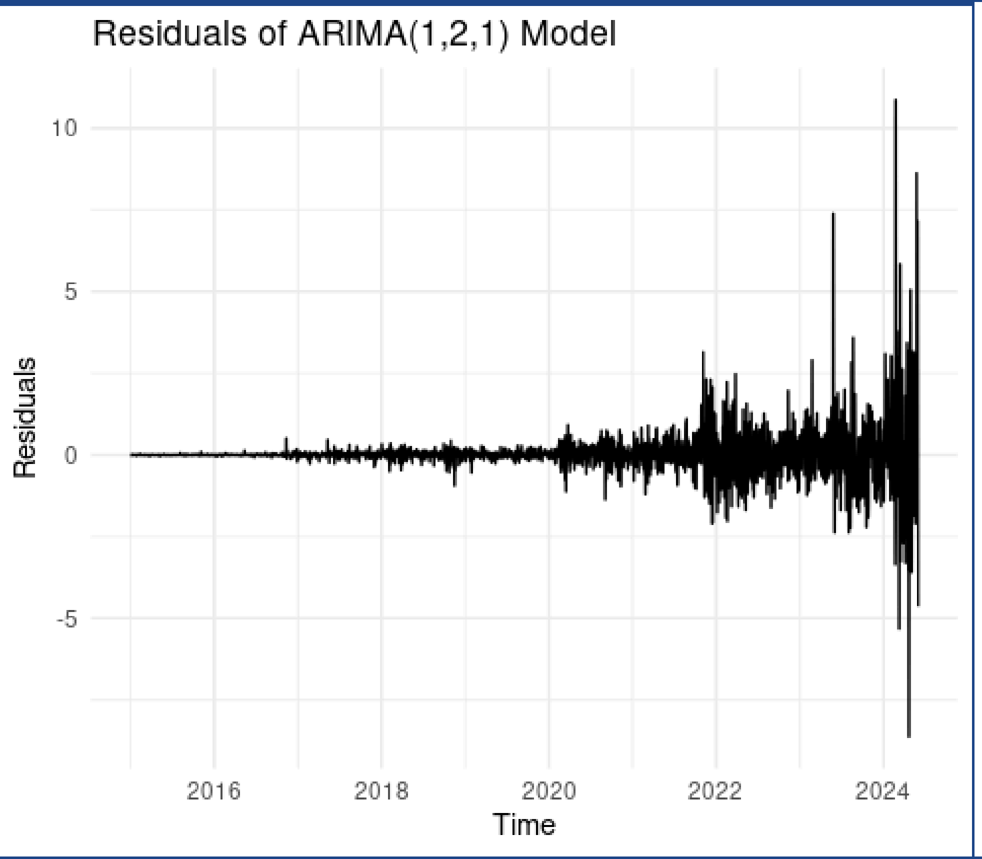


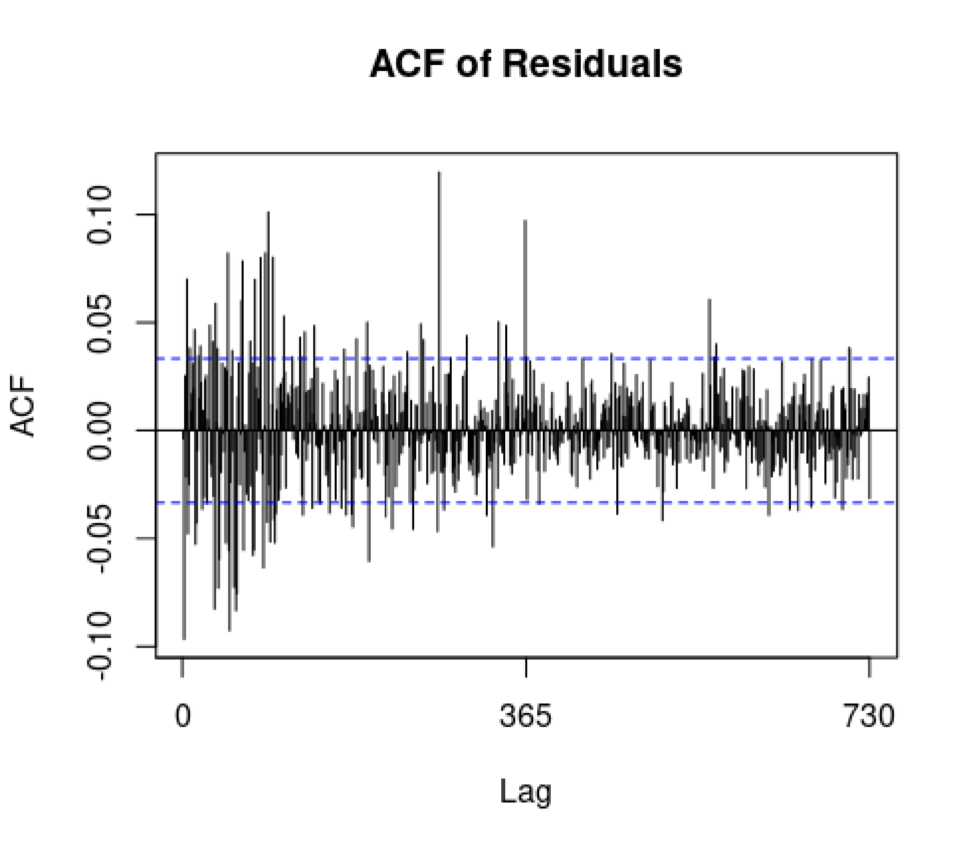
AIC: 6435.882

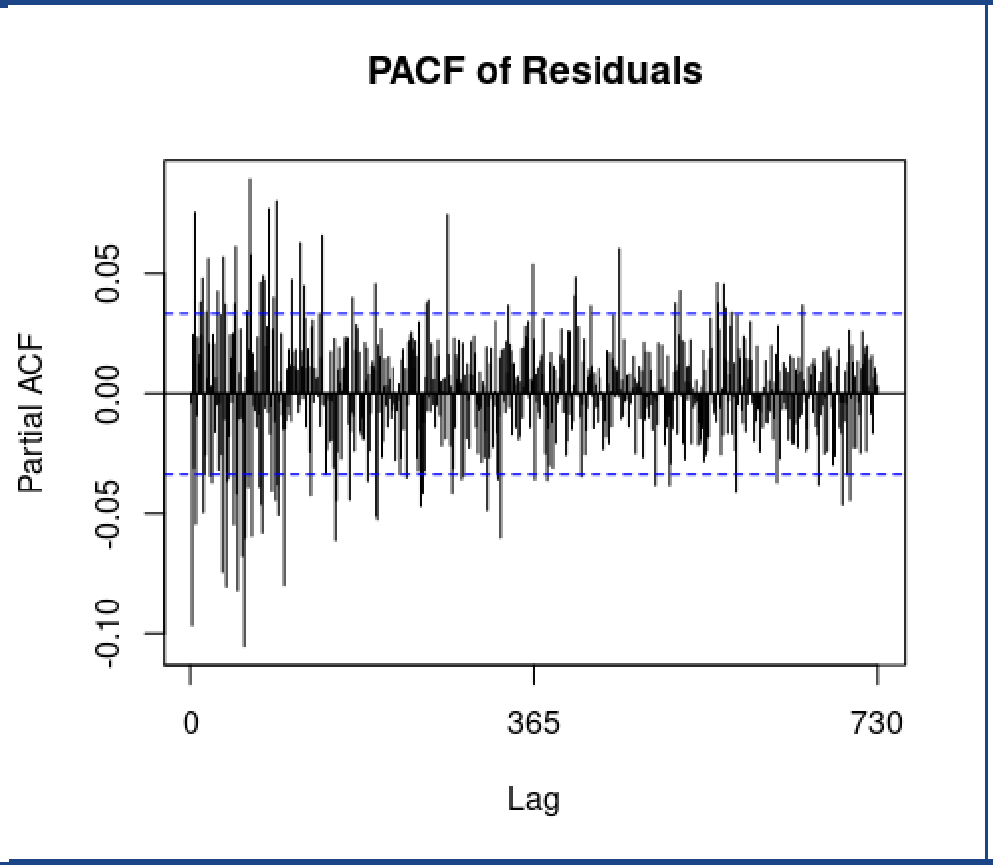
RMSE: 0.6163594

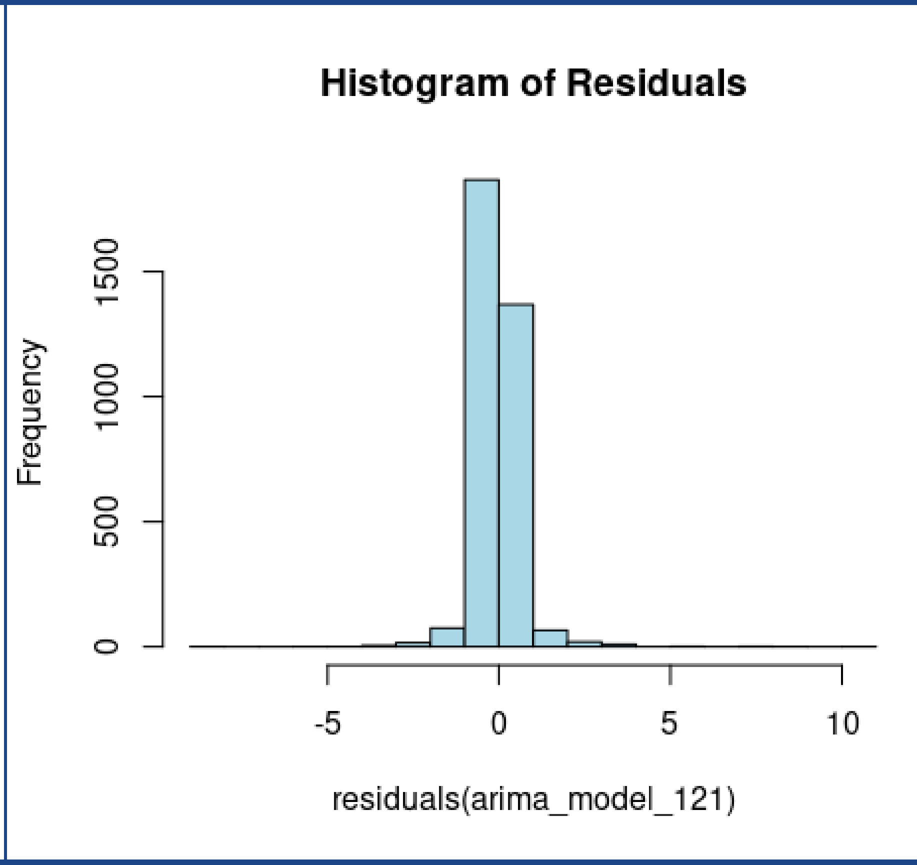
MAPE: 1.499127 %

ARIMA (1,2,1) - Residuals









Box-Pierce test: p-value = 2.809e-14

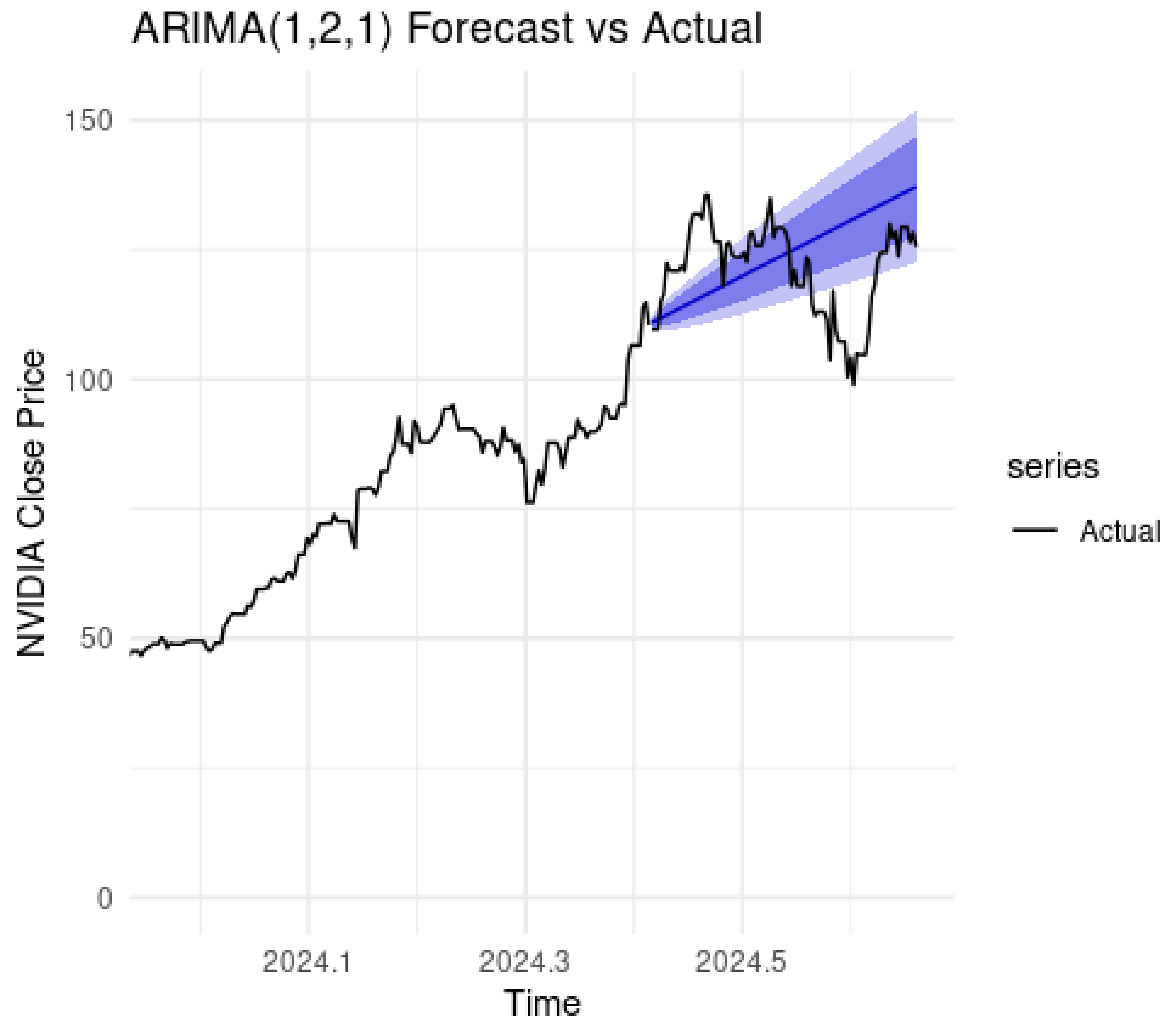
Durbin-Watson test: DW = 1.9911, p-value = 0.3975

Let’s forecast and verify

Shapiro-Wilk normality test:
W = 0.52849

the data deviates from a normal distribution

ARIMA (1,2,1) - Forecasting



AIC: 6435.882

RMSE: 0.6163594

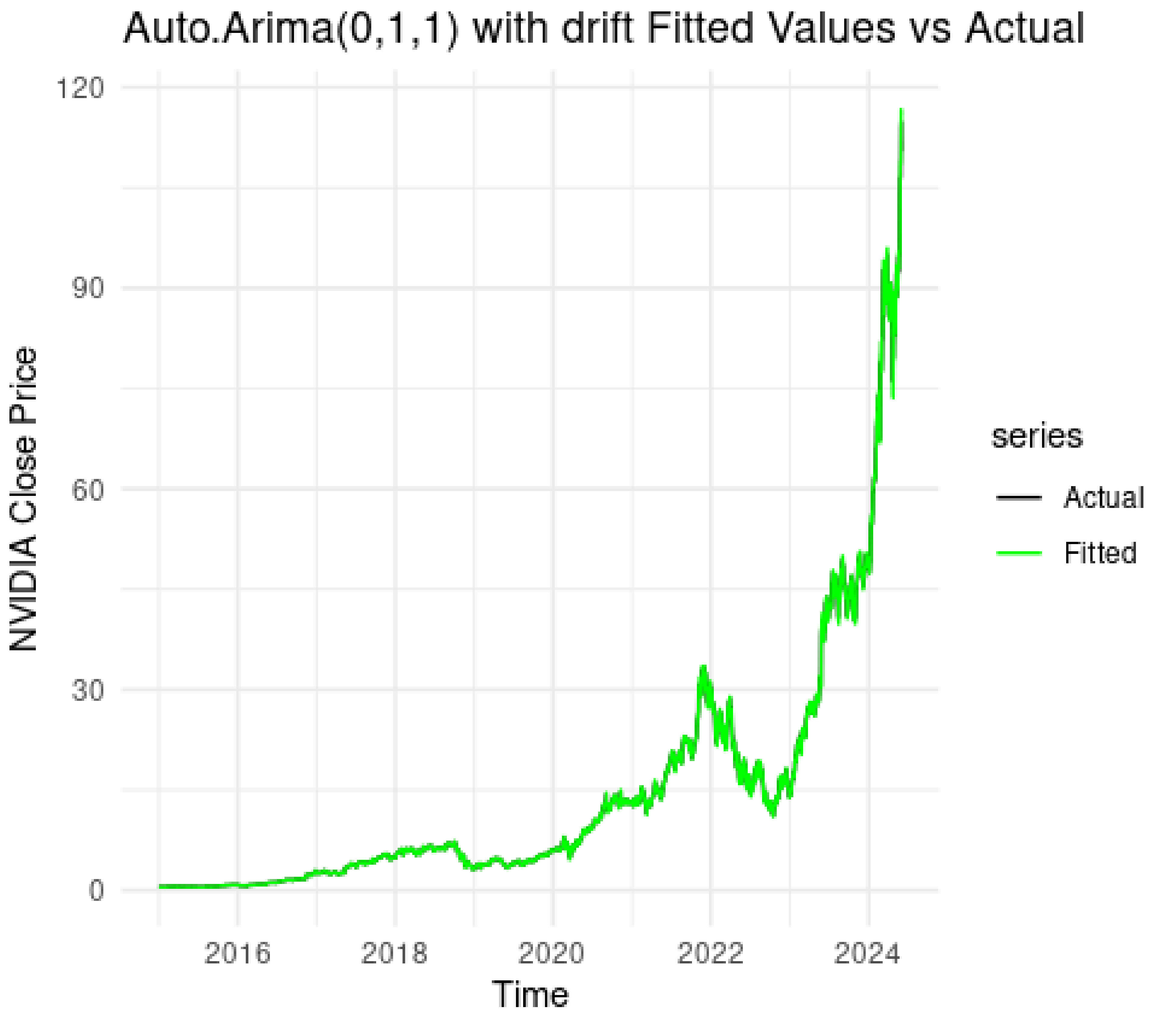
MAPE: 1.499127 %

AutoARIMA

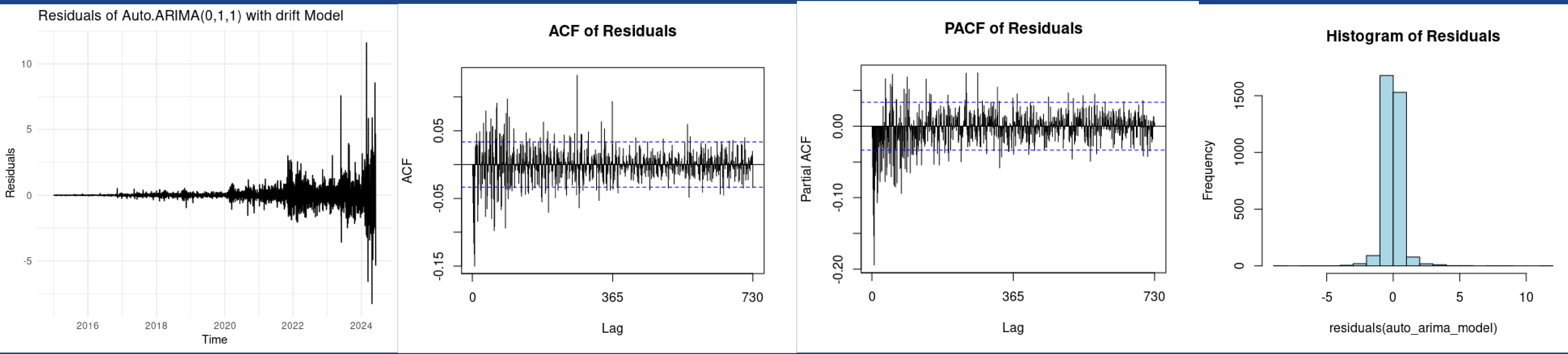
AIC: 6901.23

RMSE: 0.6592804

MAPE: 1.805527 %



AutoARIMA- Residuals



Box-Pierce test: p-value = 0.3392
Durbin-Watson test: DW = 2
No Autocorrelation

Shapiro-Wilk normality test:
W = 0.59315

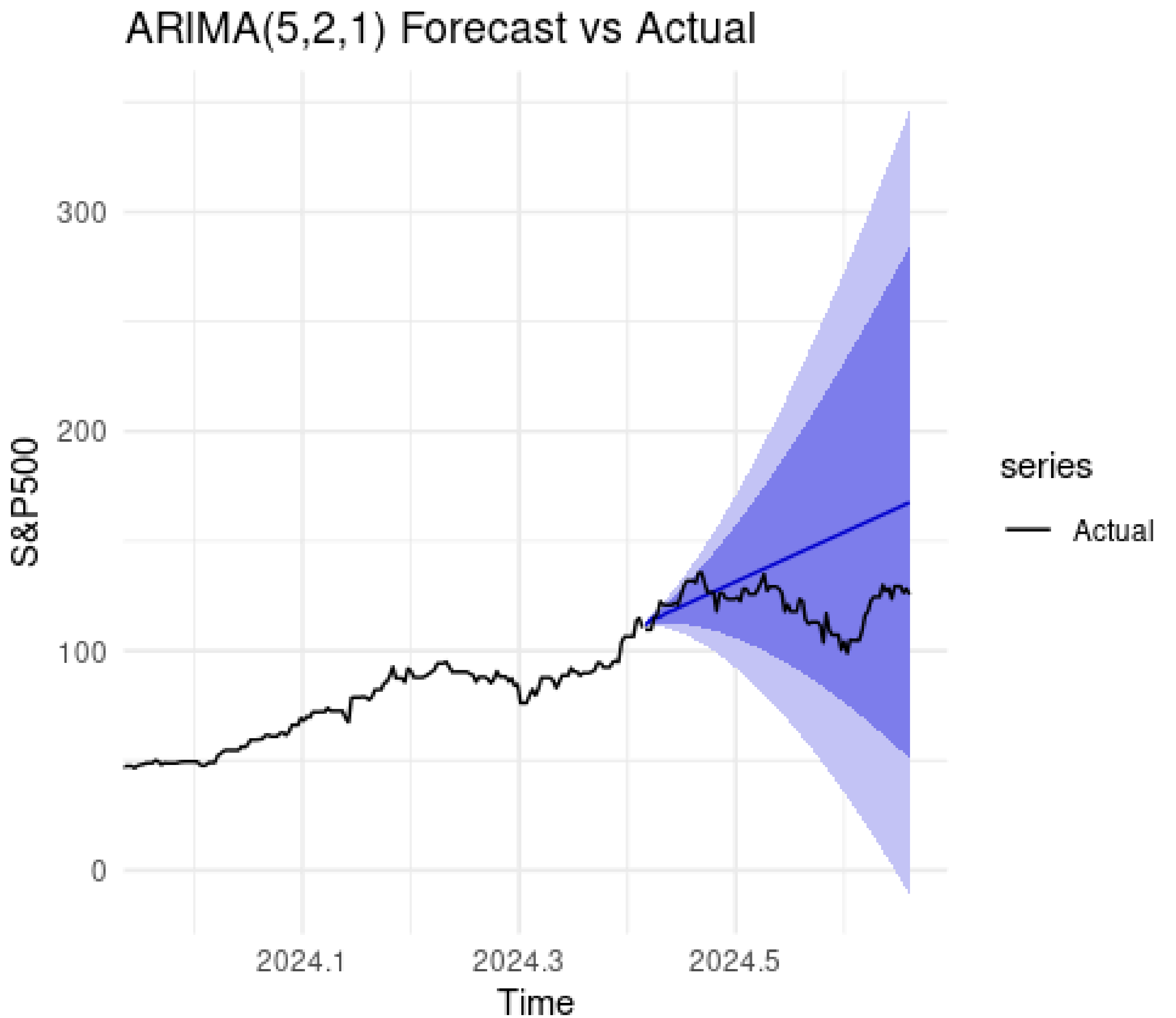
the data deviates from a normal distribution

AutoARIMA - Forecasting

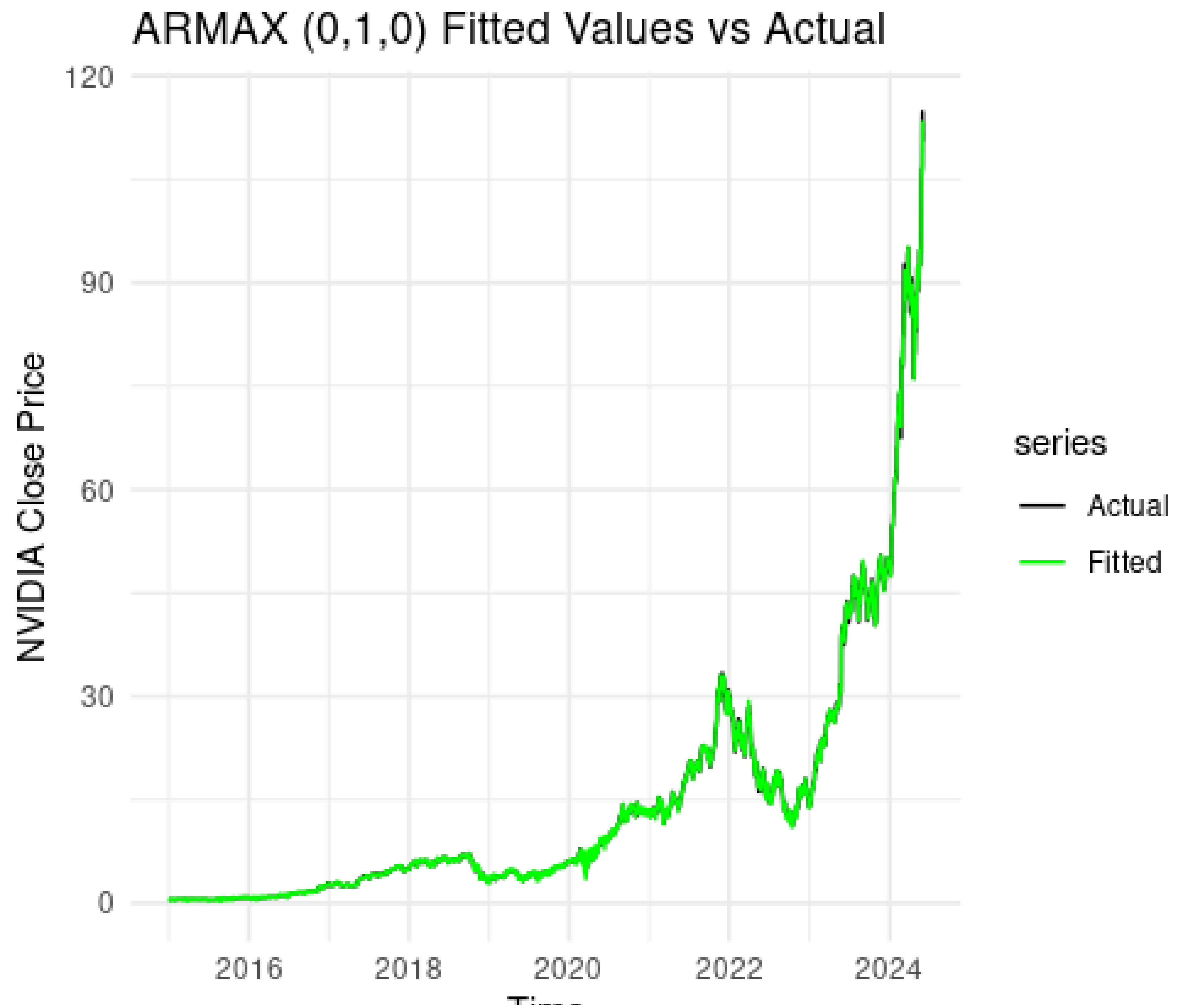
AIC: 6901.23

RMSE: 27.60483

MAPE: 18.55904 %

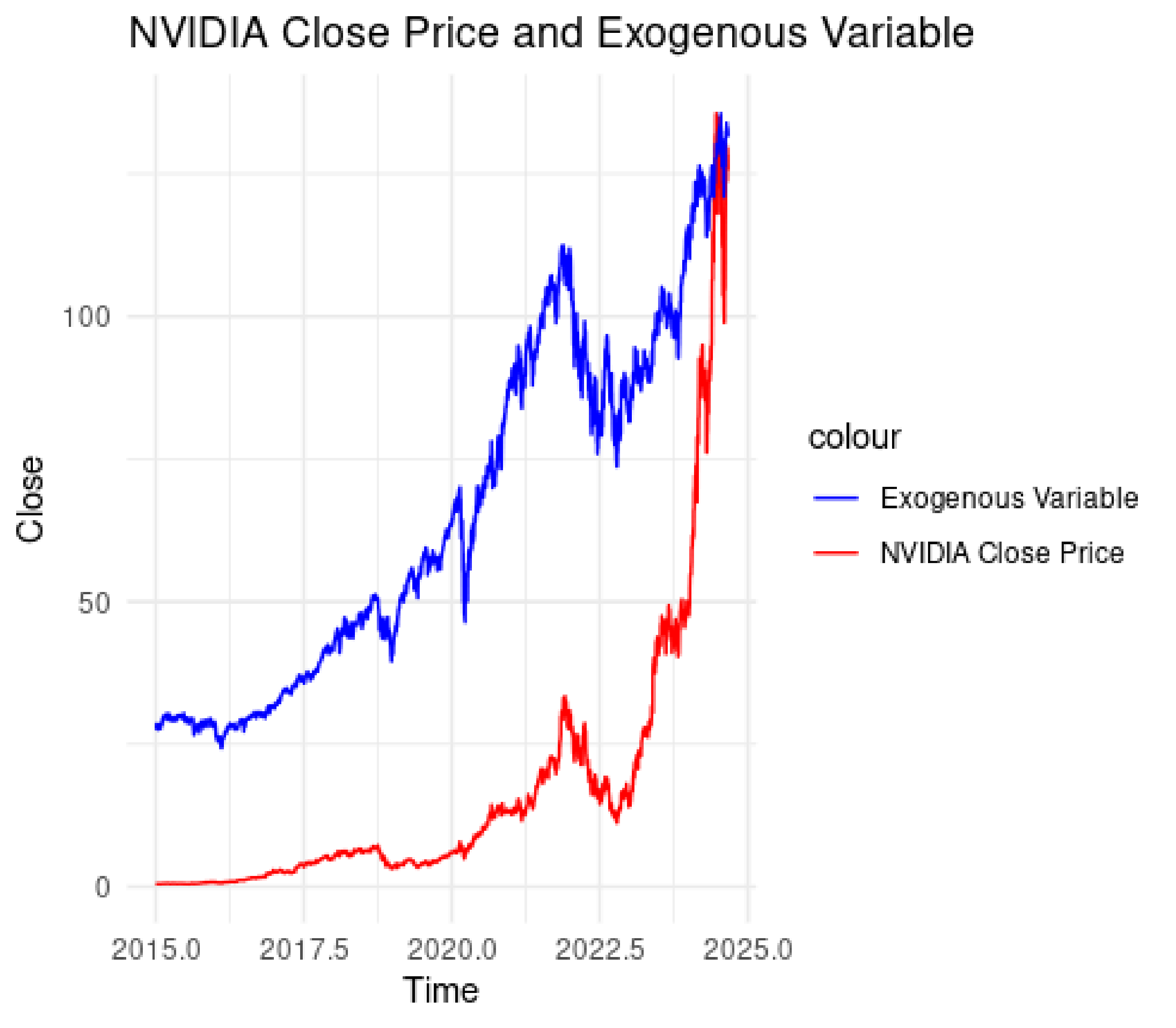


AutoARMAX with Tech Data

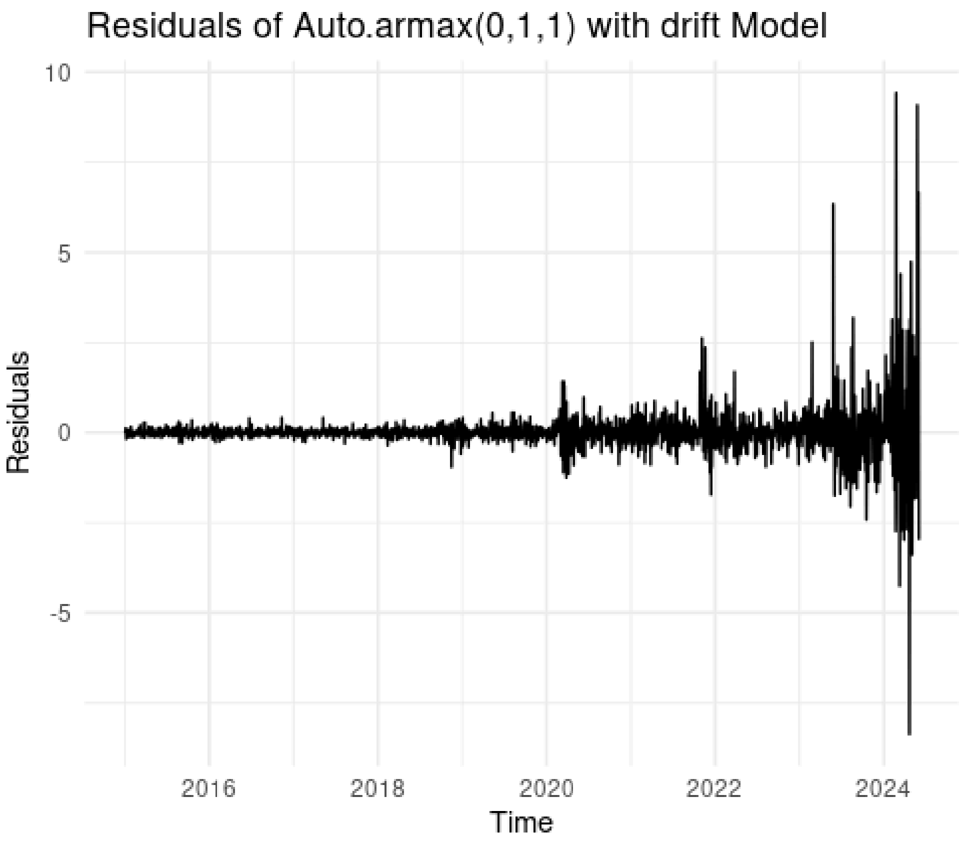


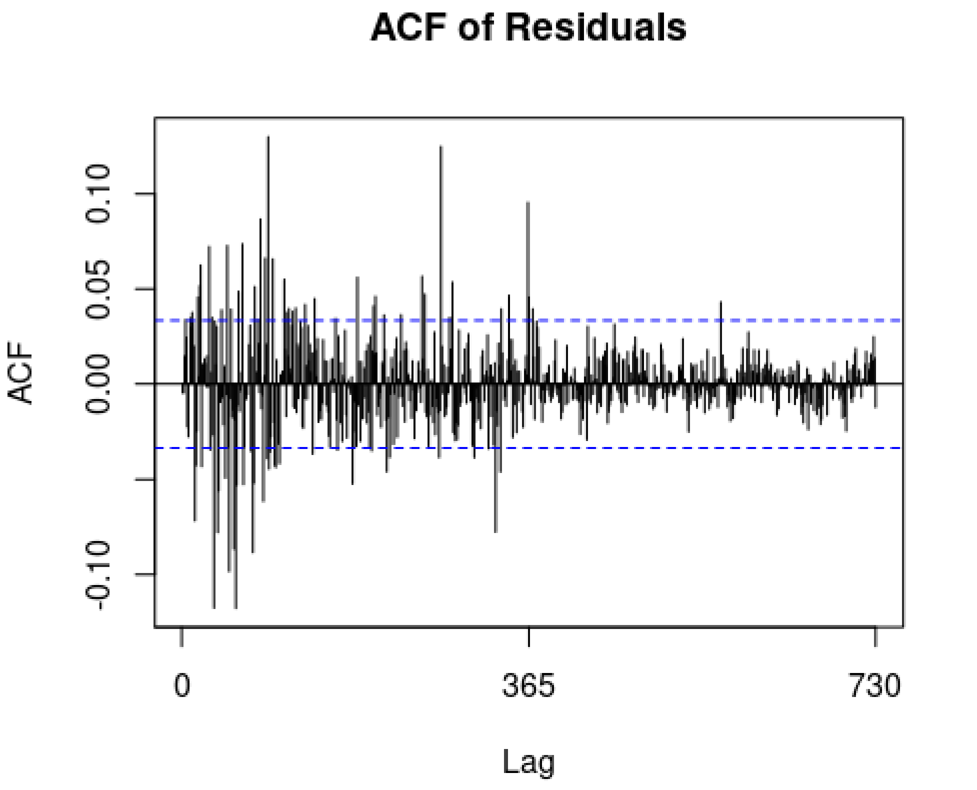
AIC: 5241.503

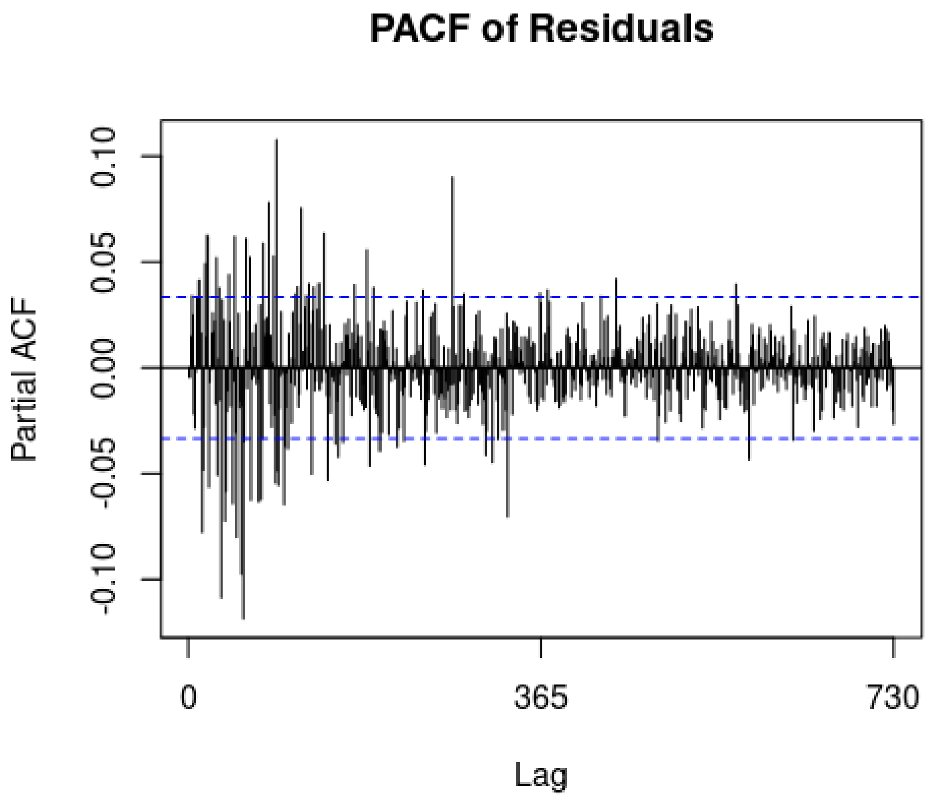
RMSE: 0.5171188 | MAPE: 2.791269 %

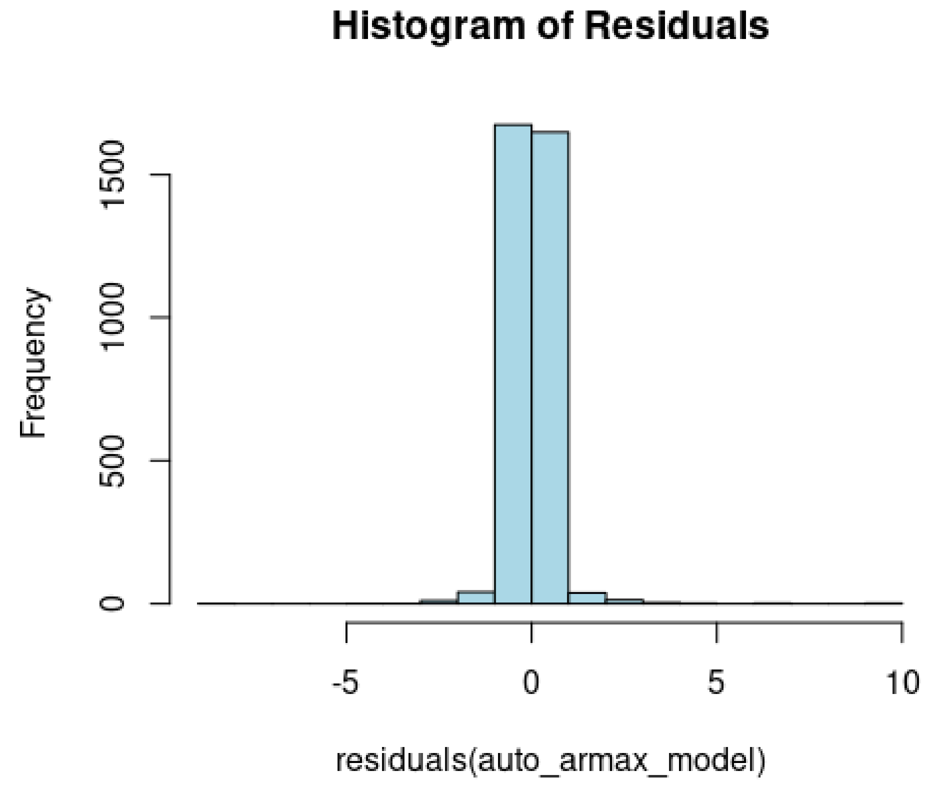


AutoARMAX with Tech Data - Residuals









Box-Pierce test: p-value = 0.8002
Durbin-Watson test: DW = 2
No Autocorrelation

Shapiro-Wilk normality test:
W = 0.50173

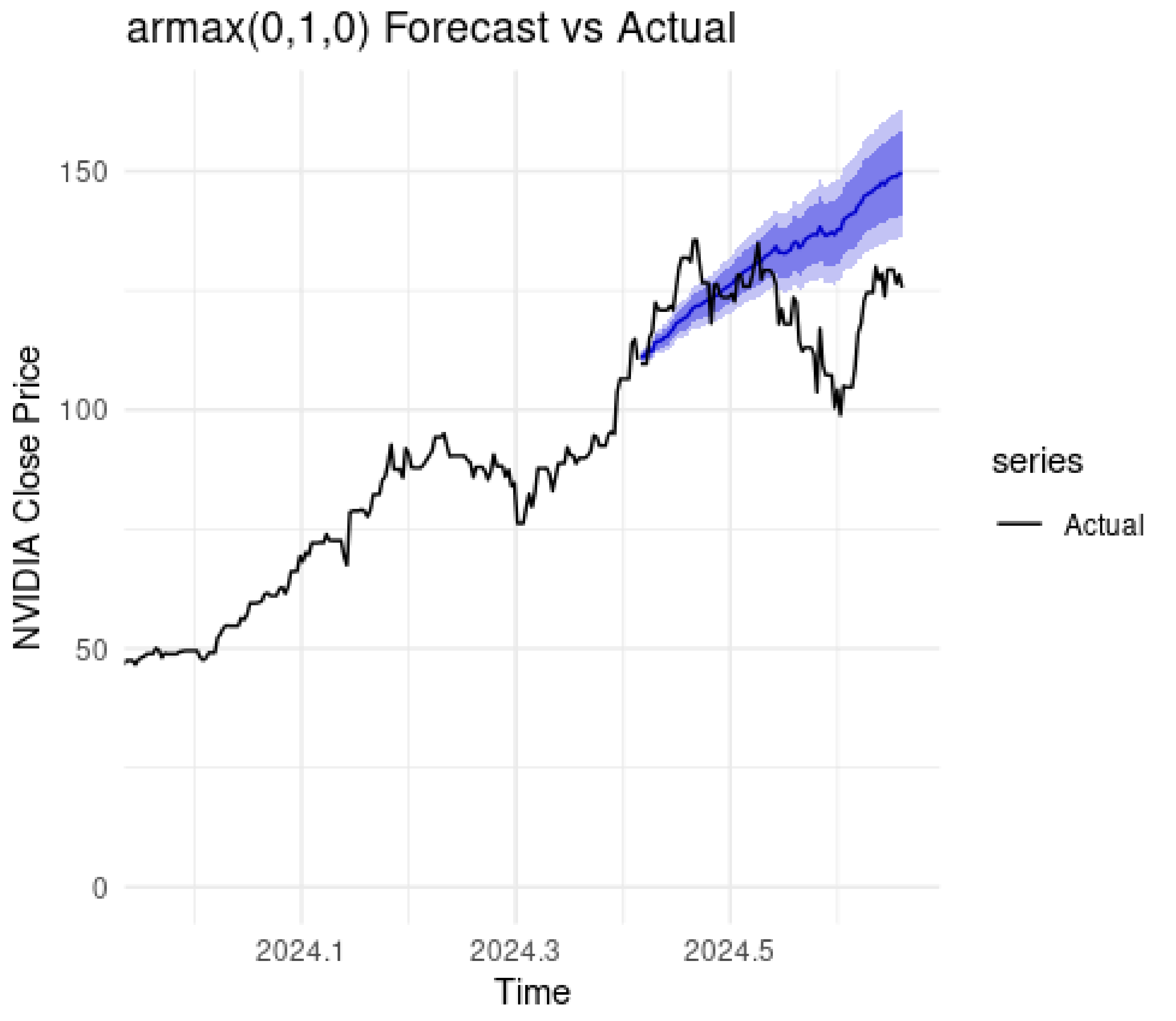
the data deviates from a normal distribution

AutoARMAX with Tech Data - Forecasting

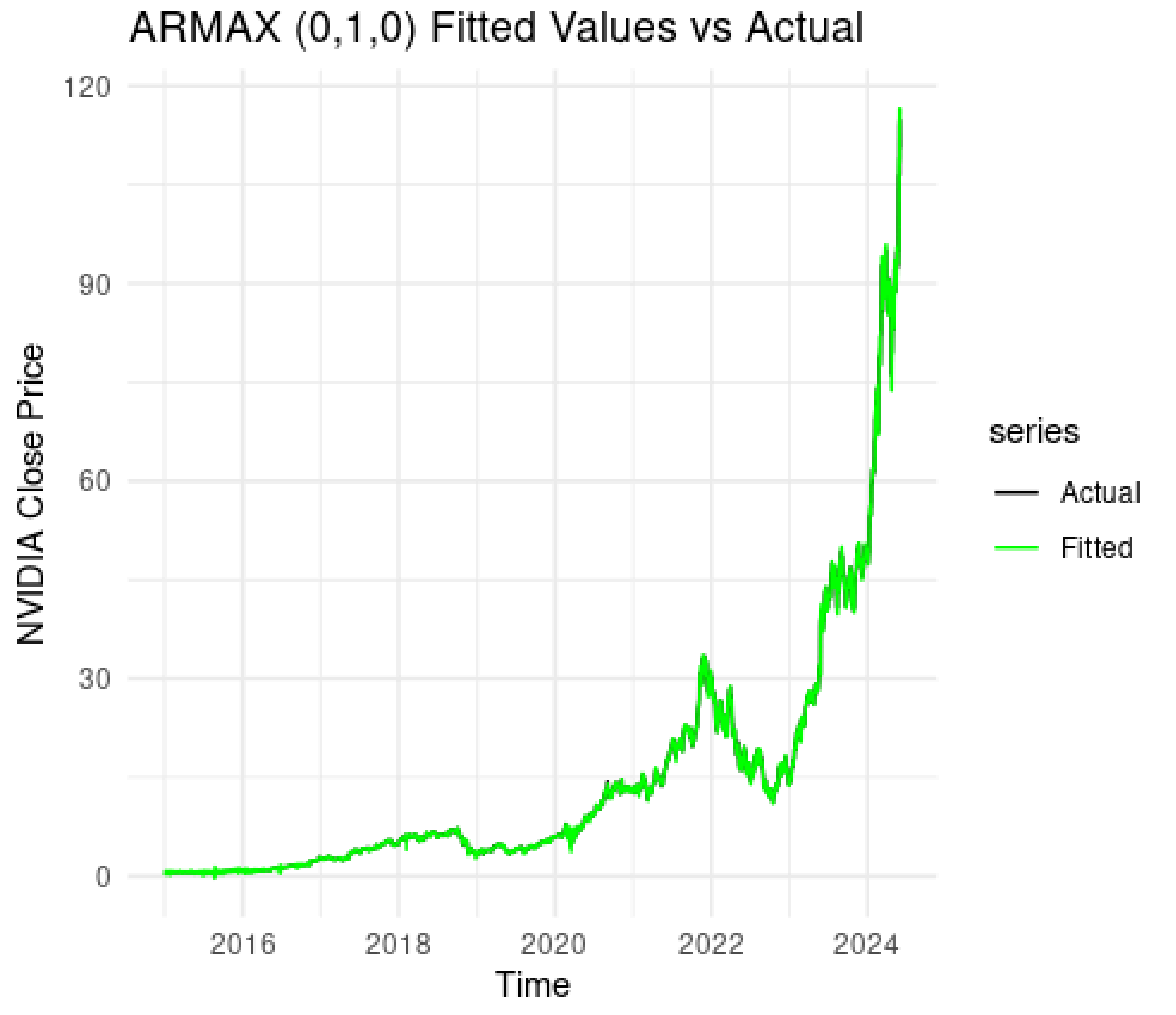
AIC: 5241.503

RMSE: 17.84104

MAPE: 12.11639 %

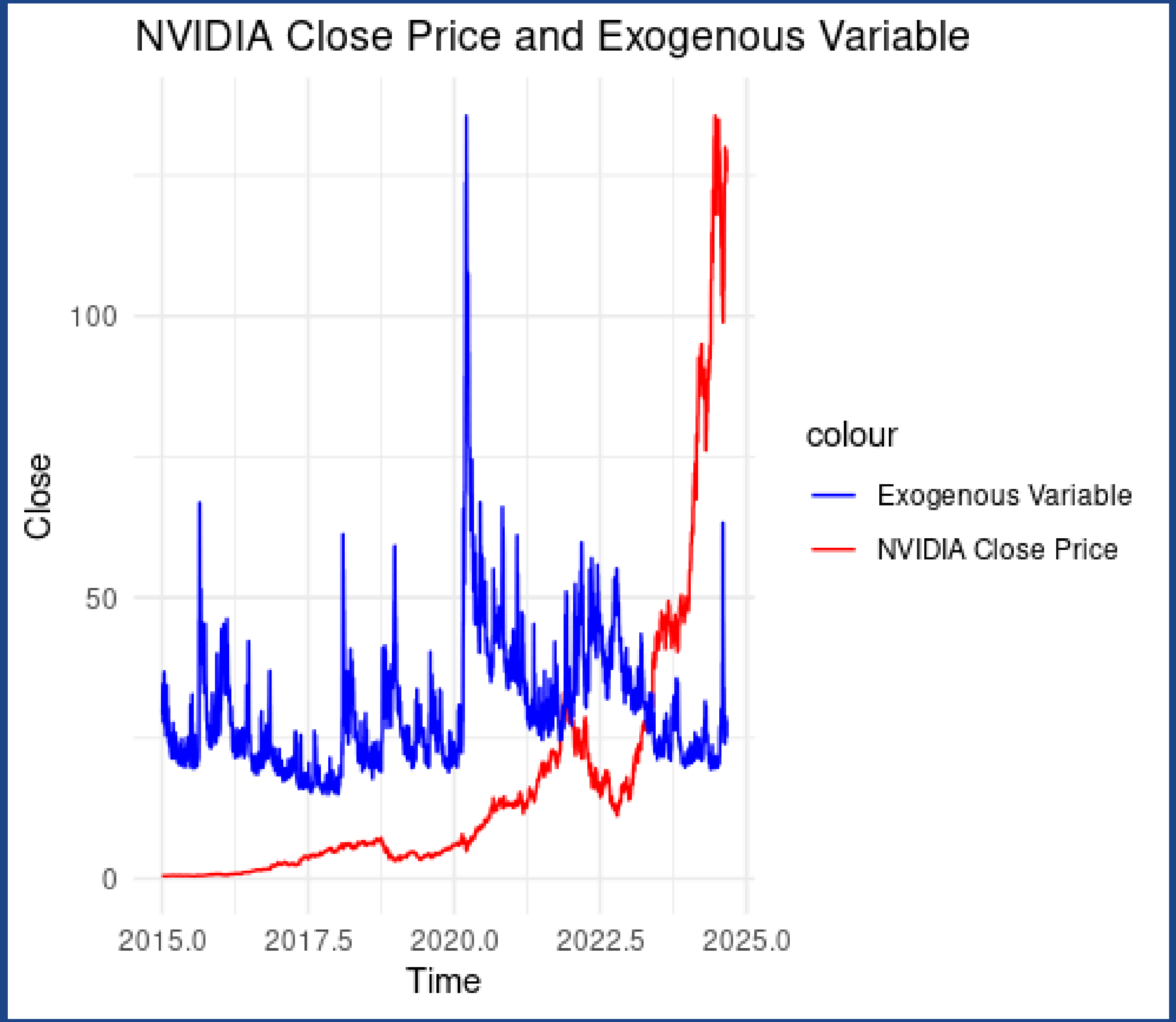


AutoARMAX with VIX Close

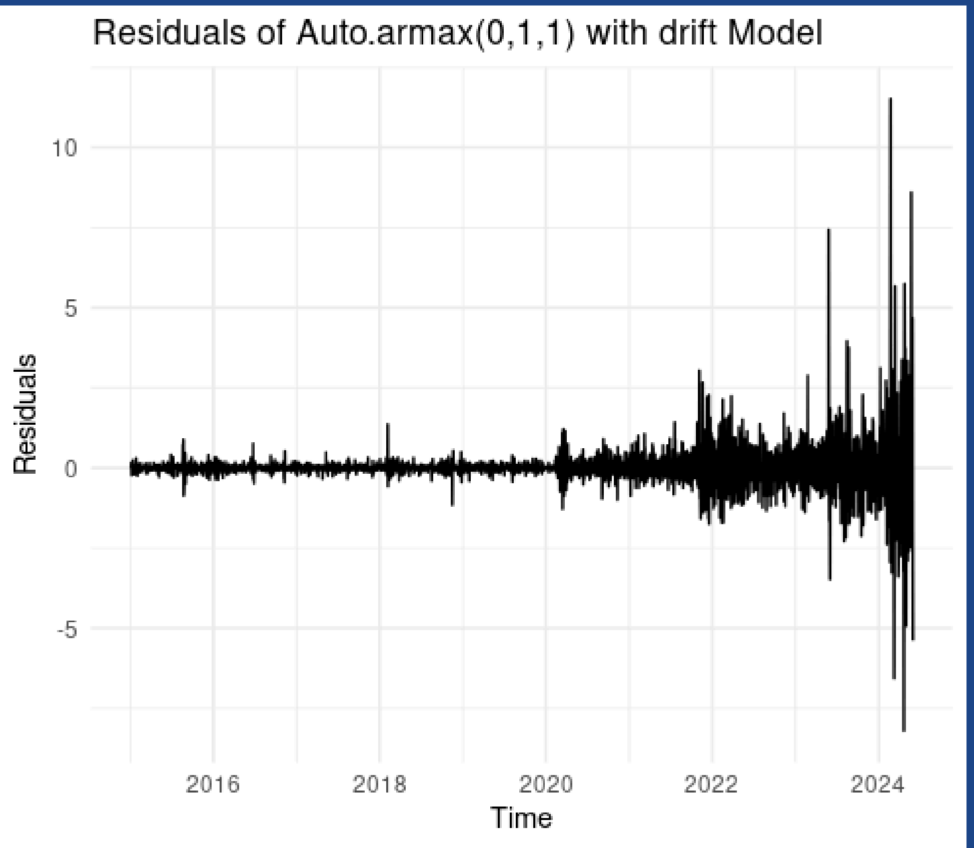


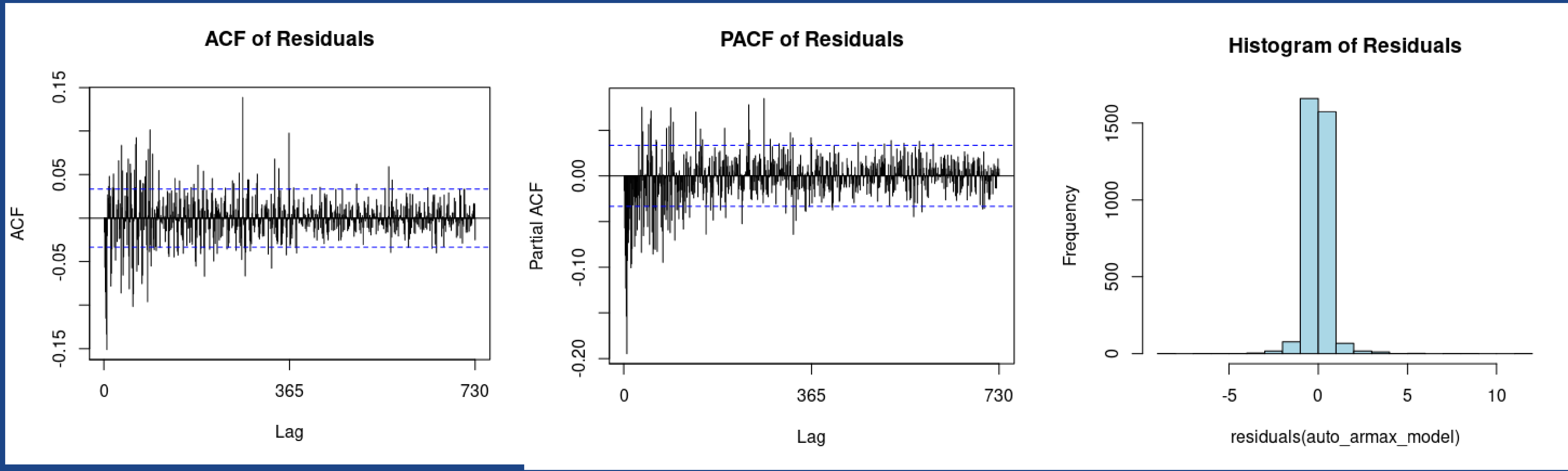
AIC: 6693.911

RMSE: 0.6394961 | MAPE: 3.852677 %



AutoARMAX with VIX Close - Residuals





Box-Pierce test: p-value = 0.3476

Durbin-Watson test: DW = 2

No Autocorrelation

Shapiro-Wilk normality test:

W = 0.58892

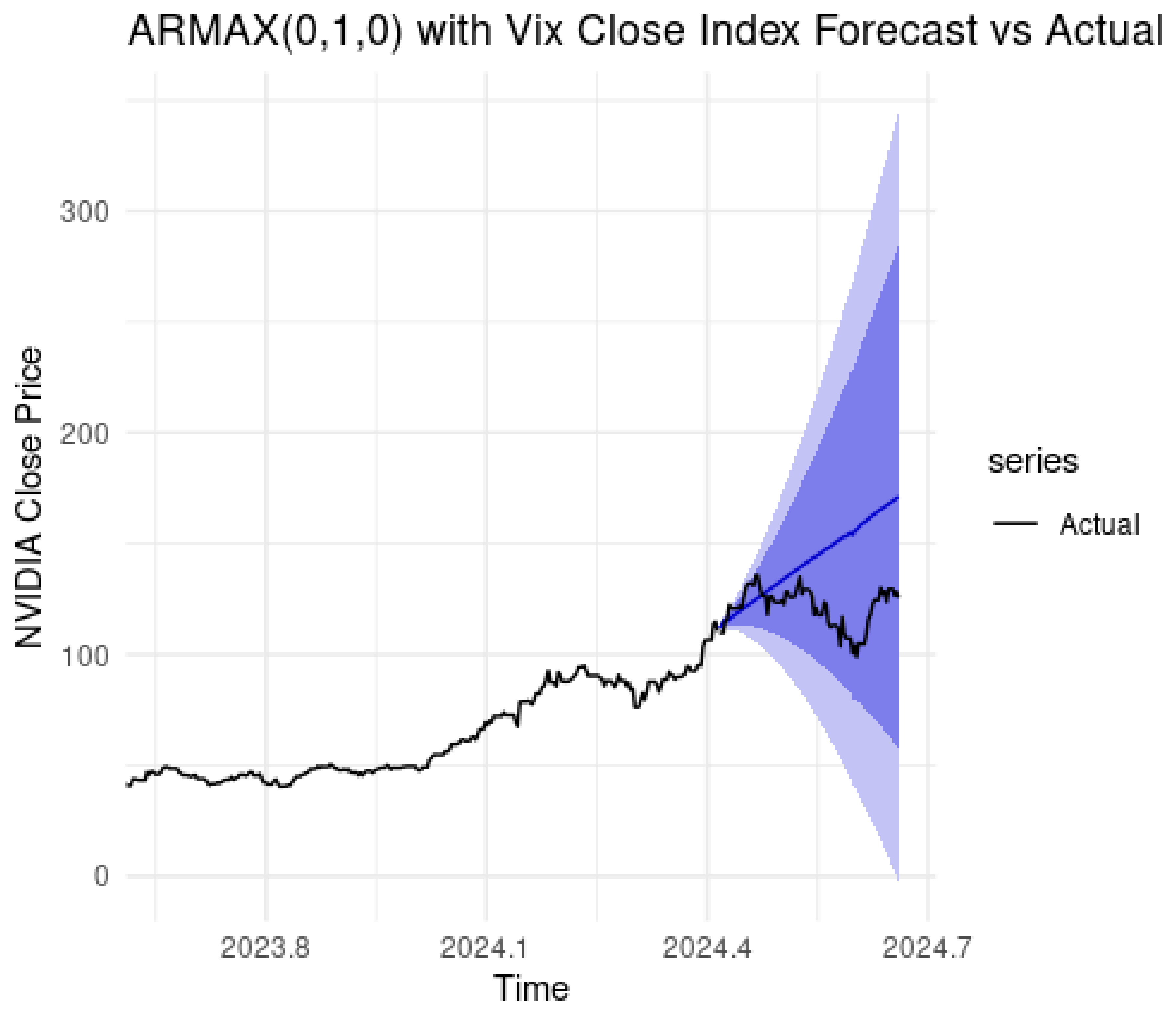
the data deviates from a normal distribution

AutoARMAX with VIX Close - Forecasting

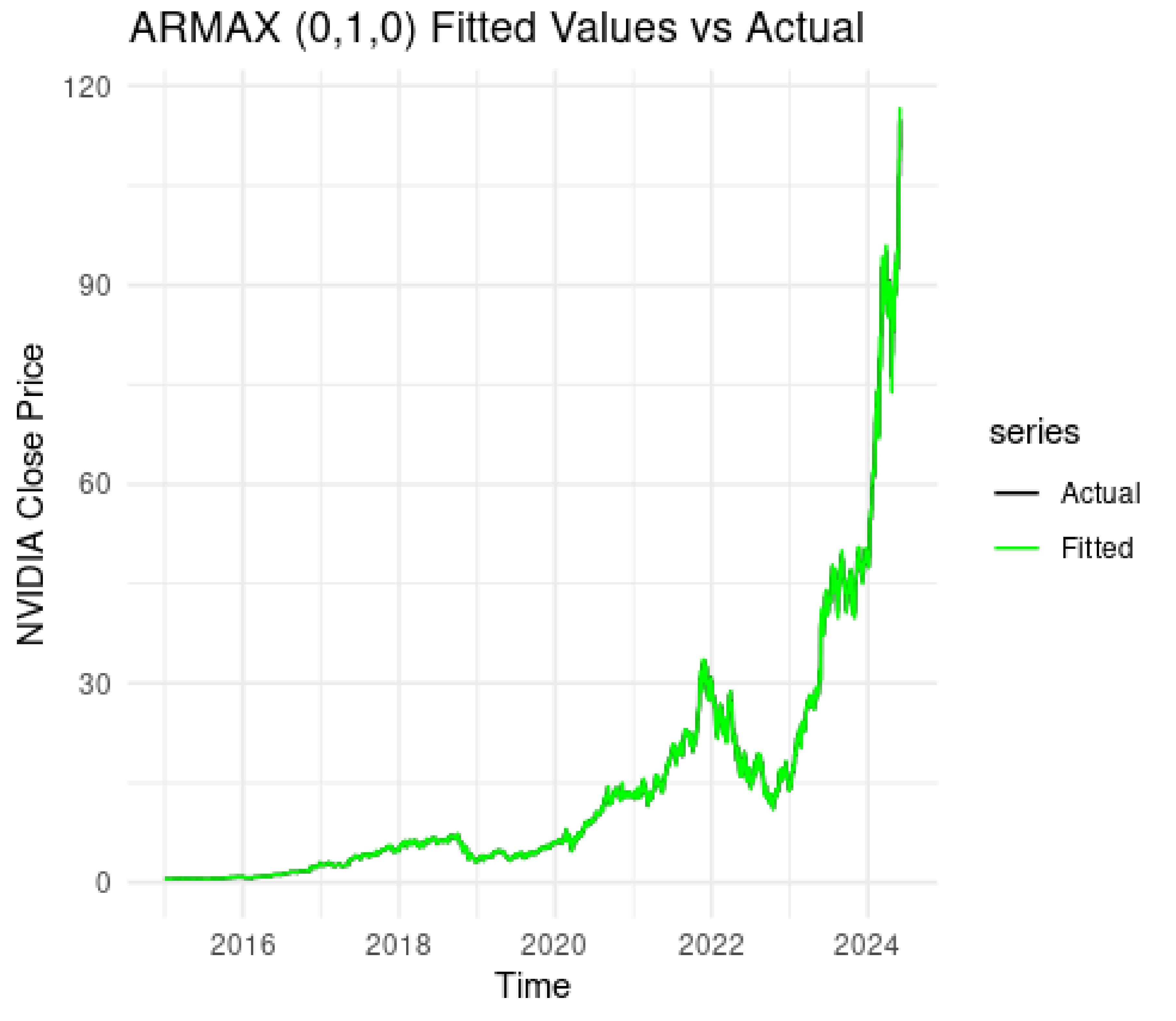
AIC: 6693.911

RMSE: 29.3777

MAPE: 19.83791 %

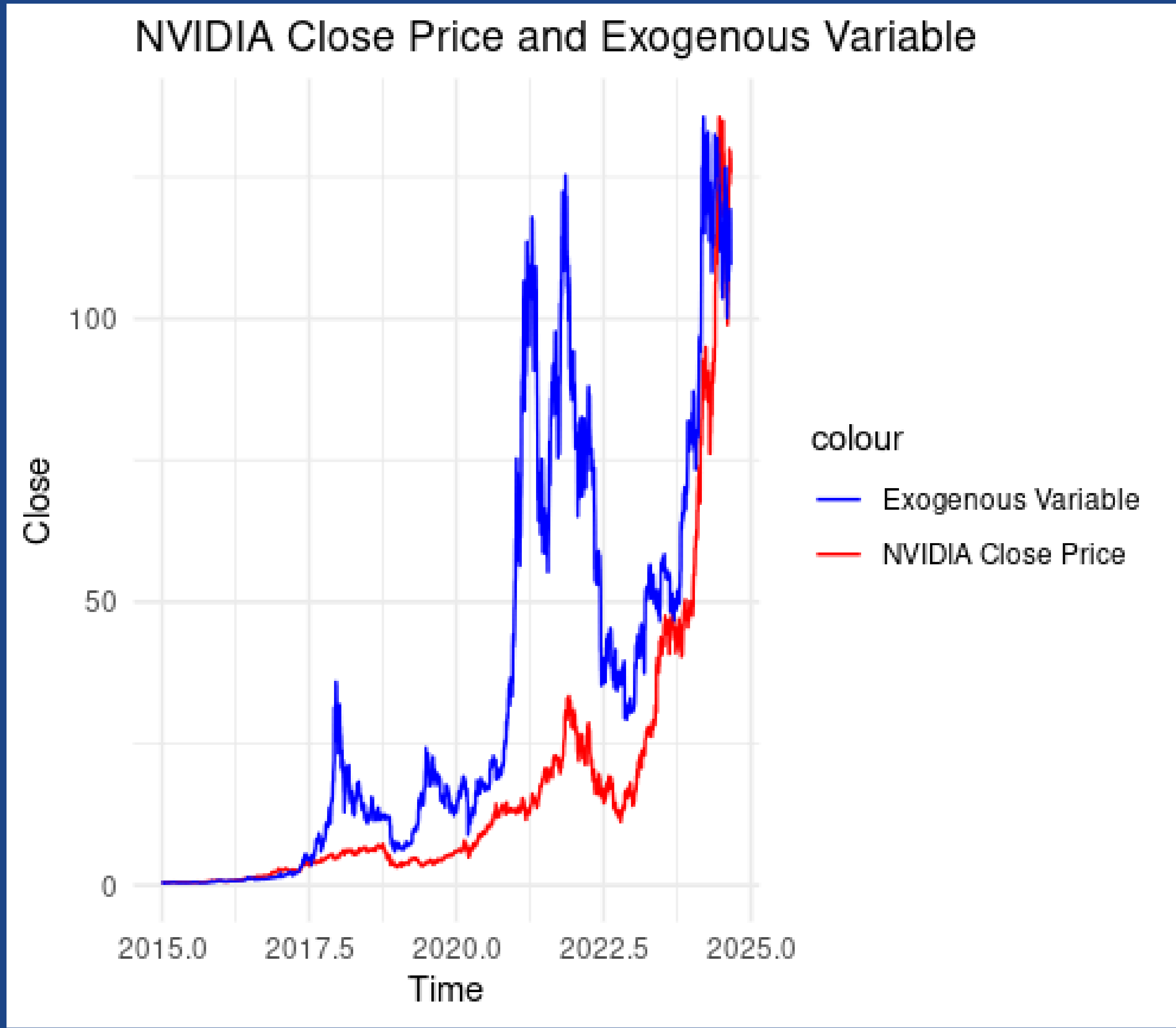


AutoARMAX with BTC Close

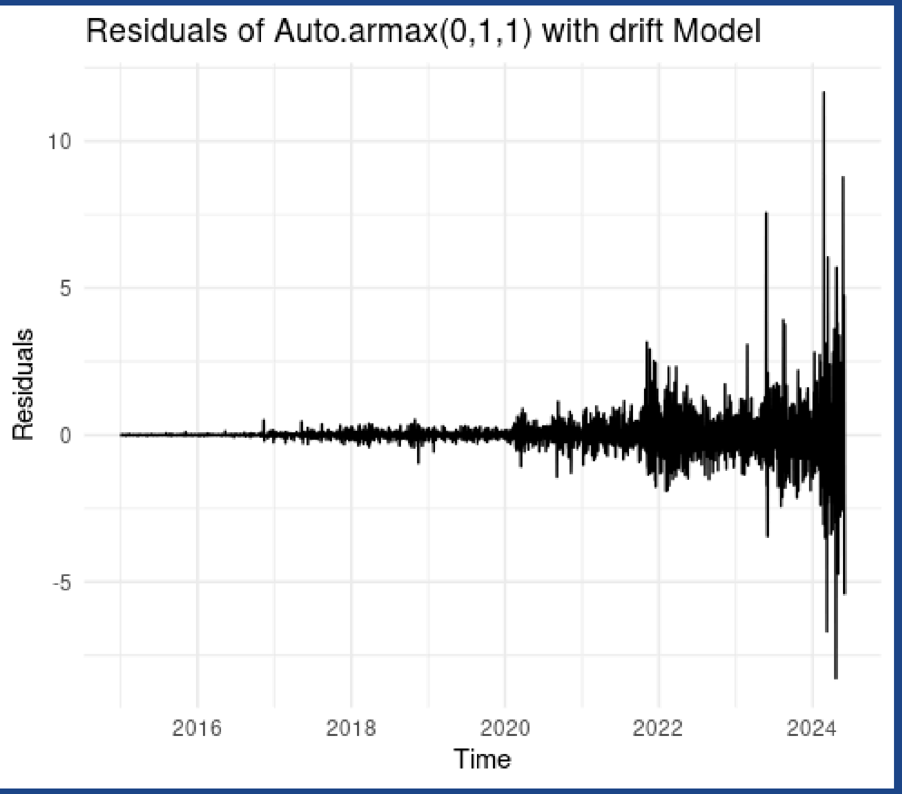


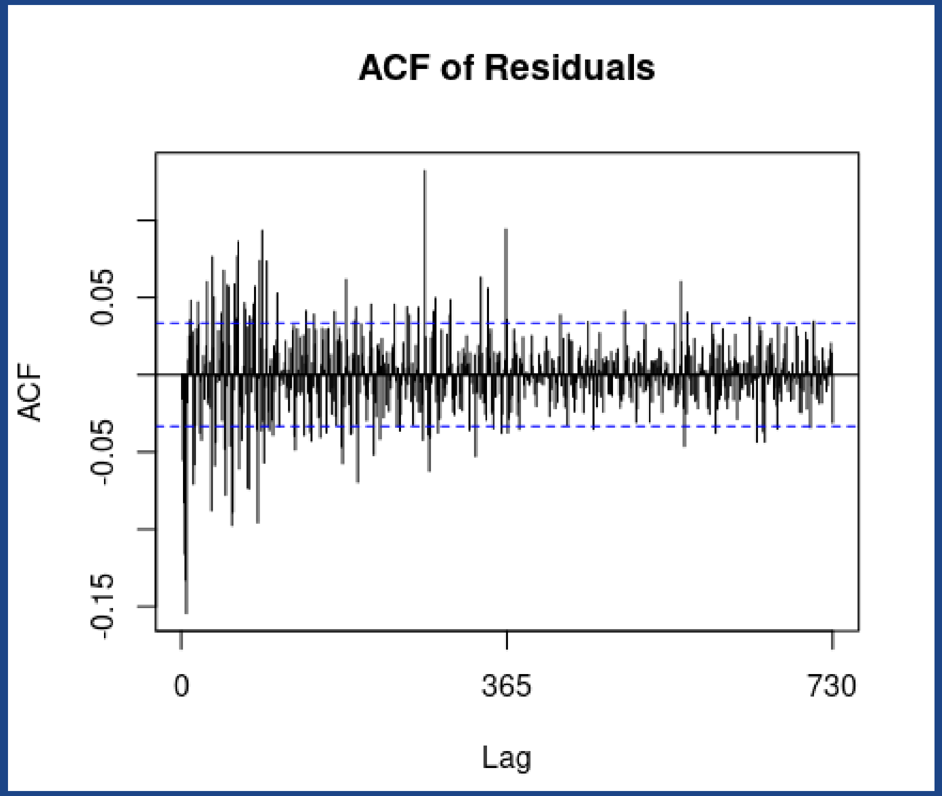
AIC: 6830.114

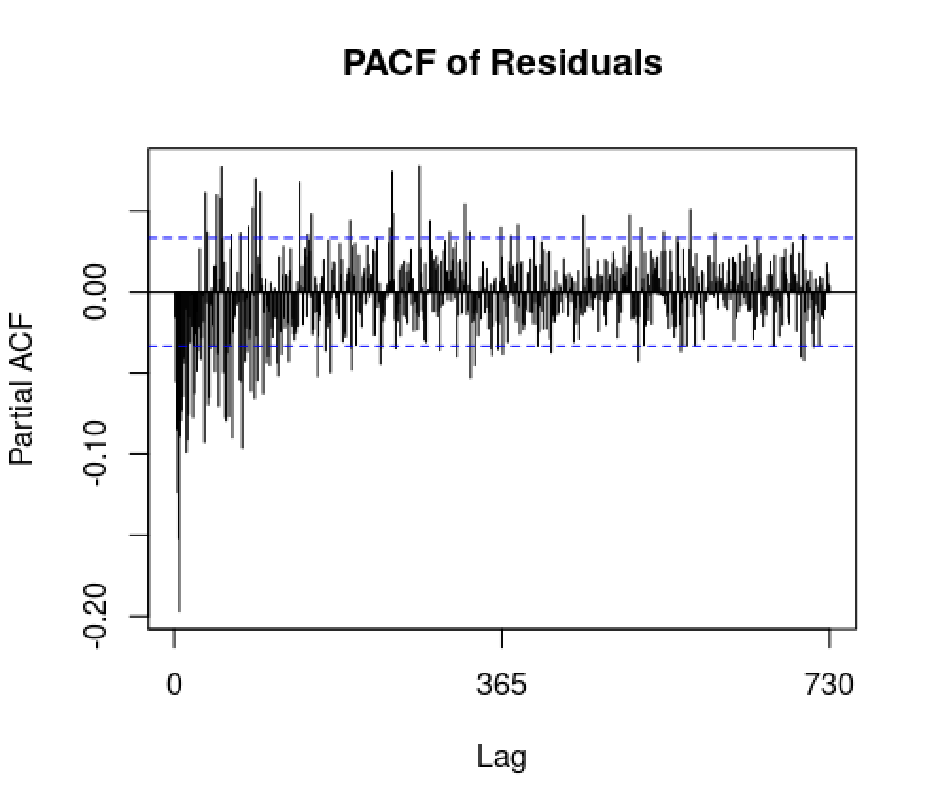
RMSE: 0.6523022 | MAPE: 1.819586 %

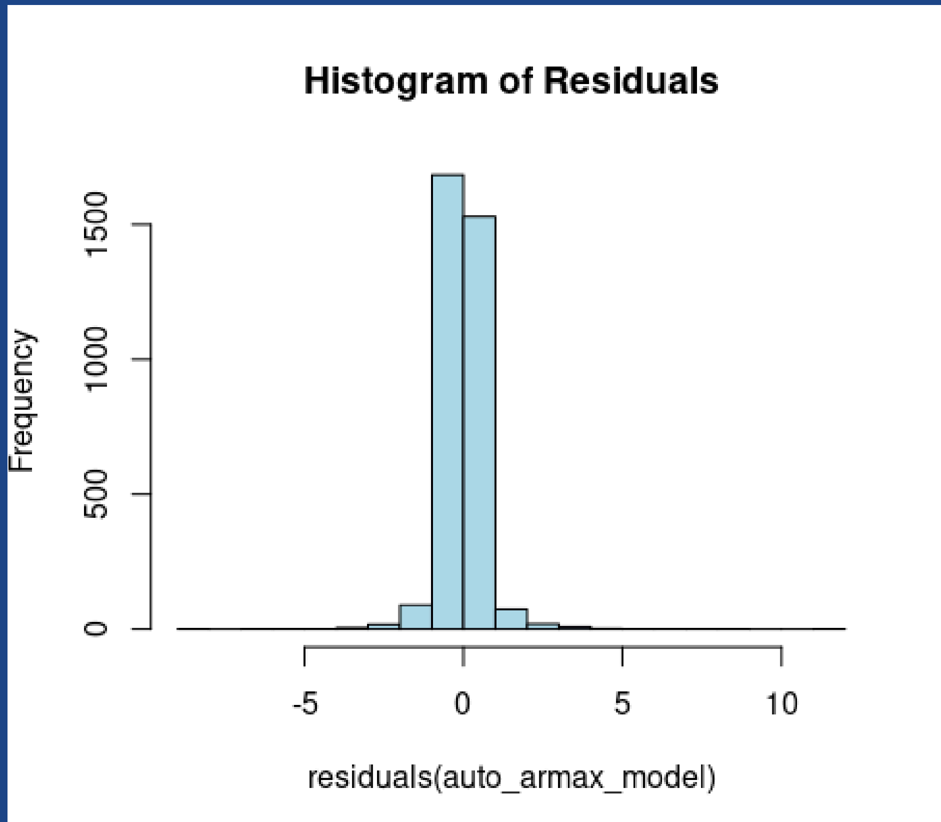


AutoARMAX with BTC Close - Residuals









Box-Pierce test: p-value = 0.3619
Durbin-Watson test: DW = 2
No Autocorrelation

Shapiro-Wilk normality test:
W = 0.58751

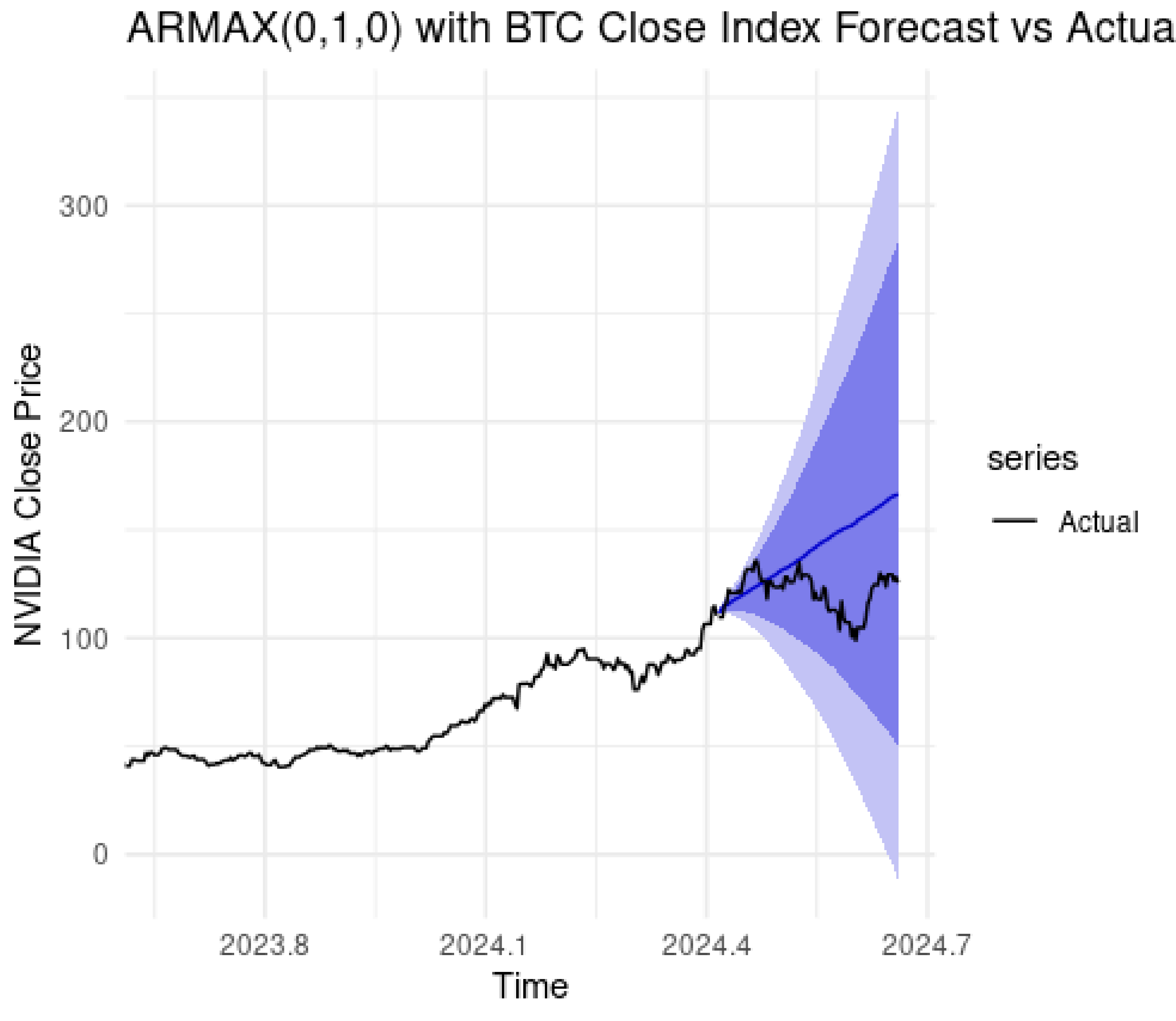
the data deviates from a normal distribution

AutoARMAX with BTC Close - Forecasting

AIC: 6830.114

RMSE: 27.0654

MAPE: 18.11172 %



Comparative Analysis

	Model	AIC	RMSE	MAPE
1	Loess	NA	2.6853047	8.412563
2	Spline Regression	NA	3.4759537	38.083161
3	Smoothing Spline	NA	2.0846777	6.924252
4	GAM w/TechSemi	15228.227	2.8754107	17.094755
5	GAM w/TechSemiVix	16489.822	2.6552962	27.325817
6	GAM w/TechSemiVixBTC	16491.703	2.6552502	27.342989
7	Gradient Boosting	NA	0.3627375	2.469149
8	Holt-Winters	NA	0.6512679	1.556645
9	ARIMA (1,1,1)	6462.886	0.6190986	1.469759
10	ARIMA (1,2,1)	6435.882	0.6163594	1.499127
11	AutoARIMA	6901.230	0.6592804	1.805527
12	ARMAX w/Tech	5241.503	0.5171188	2.791269
13	ARMAX w/Vix	6693.911	0.6394961	3.852677
14	ARMAX w/BTC	6830.114	0.6523022	1.819586

	Model	RMSE	MAPE
1	Holt-Winters	13.33353	9.428002
2	ARIMA (1,1,1)	13.62299	9.665037
3	ARIMA (1,2,1)	13.21806	9.362236
4	AutoARIMA	27.60483	18.559039
5	ARMAX w/Tech log	17.84104	12.116394
6	ARMAX w/Vix	29.37770	19.837911
7	ARMAX w/BTC	27.06540	19.837911

Conclusion

NVIDIA's stock is highly responsive to sector trends, external market forces, and cryptocurrency volatility. As it continues to grow and adapt to evolving technological demands, sophisticated forecasting models are crucial for accurately predicting its stock movements. This suggests that NVIDIA’s future stock trajectory will likely remain tied to innovations in AI, shifts in cryptocurrency, and overall market volatility, making it a stock that demands careful, data-driven analysis for accurate forecasting.