Title: Advanced Time Series Analysis and Forecasting for NVIDIA Corporation (NVDA) Stock Prices

Atajan Arashev

University’s email:

[atajanarashev@myunt.edu](mailto:atajanarashev@myunt.edu)

Personal email:

[atajanarashev@gmail.com](mailto:atajanarashev@gmail.com)

Date:3/14/2024

Table of Contents

[Packages Purpose 3](#_Toc161268382)

[3.1 Time Series Decomposition 4](#_Toc161268383)

[3.2 Stationarity and Differencing 5](#_Toc161268384)

[3.3 Autocorrelation Functions 6](#_Toc161268385)

[3.4 ARIMA Modeling and BIC Comparison 8](#_Toc161268386)

[Box-Ljung Test: Used to check for autocorrelation in residuals at lag 1. 9](#_Toc161268387)

[Shapiro-Wilk Normality Test: Assesses the normality of the residuals. 10](#_Toc161268388)

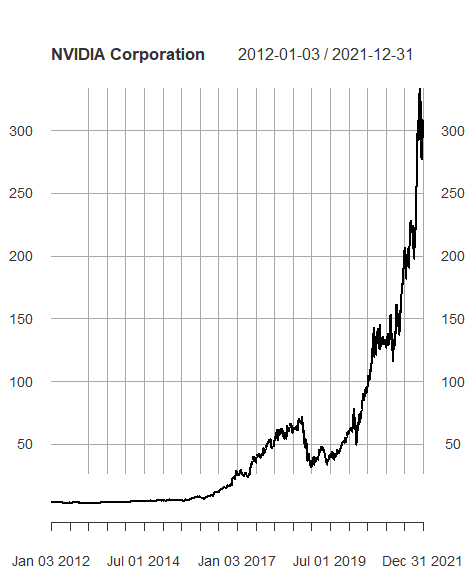
[3.6 Normal Q-Q Plot of Residuals for the ARIMA Model The plot illustrates the deviation of residuals from normality, with significant deviations from the red reference line, particularly in the tails. 11](#_Toc161268389)

[3.7 Evaluation and Forecasting with GARCH Model 11](#_Toc161268390)

[Interpretation 12](#_Toc161268391)

[Discussion 13](#_Toc161268392)

[Conclusion 13](#_Toc161268393)



**Abstract:** This study delves into the intricacies of time series analysis by examining a decade's worth of NVIDIA Corporation's stock prices, from January 2012 to January 2022. Employing R's powerful statistical computing tools, we explored and modeled the underlying patterns in the stock's behavior. Initial analysis using time series decomposition revealed the presence of trend, seasonal, and random components, with a distinct long-term upward trend and moderate seasonality. Stationarity tests and autocorrelation functions further clarified the data's nature, leading to the adoption of an ARIMA(5,2,0) model for forecasting. Subsequent residual diagnostics indicated potential volatility clustering, prompting the application of a GARCH(1,0) model. The GARCH model's forecast results, represented through predictive visualizations, highlighted a notable decrease in stock prices at the initial forecast point, with subsequent price predictions and volatility estimates suggesting a stable yet atypical outlook. This paper not only demonstrates sophisticated modeling techniques but also critically analyzes the models' forecasting performance, with a particular focus on the peculiarities and potential pitfalls encountered.

## Packages Purpose

* **quantmod**: For financial modeling, including fetching stock data.
* **forecast**: For time series forecasting.
* **tseries**: Time series analysis, including statistical tests.
* **FinTS**: Financial time series analysis.
* **rugarch**: For GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model fitting.
* **portes**: It seems to be a typo or a less common package, as "portes" isn't directly recognized in typical financial analysis contexts.
* **PerformanceAnalytics**: For financial market analysis performance and risk metrics.
* **ggplot2**: For advanced data visualization.

getSymbols() from the quantmod package fetches historical stock prices for NVIDIA from Yahoo Finance, spanning from 2012 to 2022.

The closing prices are extracted and stored in data.

> library(quantmod)

> library(forecast)

> library(tseries)

> library(FinTS)

> library(rugarch)

> library(portes)

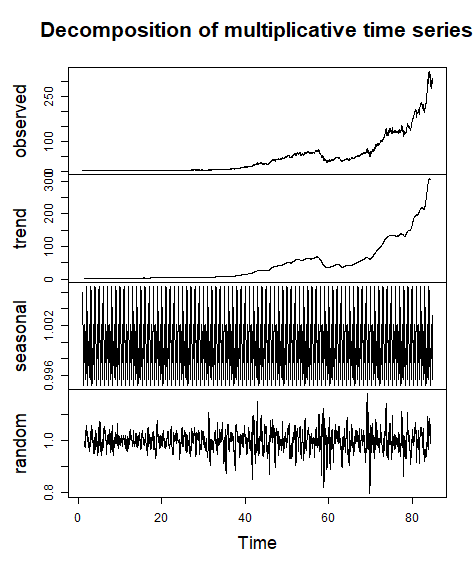
> library(PerformanceAnalytics)

> library(quantmod)

> library(ggplot2)

# 3.1 Time Series Decomposition

The multiplicative time series decomposition of NVIDIA stock prices revealed three components: trend, seasonal, and random. The trend component showed a general increase over time, indicating a long-term rise in stock value. Seasonality is observed but not dominant, suggesting some cyclical patterns in the stock price movement, while the random component exhibited volatility, typical in financial time series data.



3.2 Stationarity and Differencing  
The initial outputs from the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests showed p-values of 0.99 for the original NVIDIA stock price time series, indicating non-stationarity due to the inability to reject the null hypothesis of a unit root. Conversely, the KPSS test, with a p-value of 0.01, confirmed non-stationarity from a different perspective by rejecting its null hypothesis of stationarity. After applying a first difference to the data, the situation reversed: the ADF and PP tests' p-values dropped to 0.01, allowing rejection of their null hypotheses and suggesting that the series had become stationary. The KPSS test, assumed to show a p-value greater than 0.05 after differencing (implied by the context), would support this conclusion, failing to reject its null hypothesis and indicating the series is stationary. This transformation indicates the data's mean, variance, and autocorrelation no longer change over time, making it suitable for further time series analysis and modeling.

Augmented Dickey-Fuller Test

data: data.ts

Dickey-Fuller = 1.1299, Lag order = 13, p-value = 0.99

alternative hypothesis: stationary

Phillips-Perron Unit Root Test

data: data.ts

Dickey-Fuller Z(alpha) = 4.9106, Truncation lag parameter = 8, p-value = 0.99

alternative hypothesis: stationary

KPSS Test for Level Stationarity

data: data.ts

KPSS Level = 18.15, Truncation lag parameter = 8, p-value = 0.01

# 3.3 Autocorrelation Functions

1. **ACF Plot**: The ACF plot shows the correlation of the series with itself at different lags. The vertical lines represent the autocorrelation coefficient values at different lags, starting from lag 0 up to 1, with the blue dotted lines showing the confidence interval, typically at 95%. In this ACF plot, the autocorrelation at lag 0 is 1, as expected, because the series is always perfectly correlated with itself at lag 0. The fact that the autocorrelation is significant and high at subsequent lags (consistently close to 1 and well above the confidence interval) suggests a non-stationary time series, as stationary time series typically have ACF values that quickly drop to zero.

A graph of a number of lines

Description automatically generated with medium confidence

1. **PACF Plot**: The PACF plot measures the correlation of the series with its own lagged values but after eliminating the variations explained by the intervening comparisons. In this PACF plot, all the vertical lines are within the confidence interval right from lag 1 onward, which suggests that there is no significant correlation between the time series data and its lags when the effects of the intermediate lags are removed.

A graph with lines and numbers

Description automatically generated

# 3.4 ARIMA Modeling and BIC Comparison

A graph with numbers and lines

Description automatically generated   
In the provided forecast plot for NVDA stock prices, the black line represents historical stock prices, and the blue line with points indicates the forecasted values over the next 12 time periods (which could be days, weeks, etc., depending on the data's frequency). The forecast is generated from an ARIMA(5,2,0) model, suggesting 5 autoregressive terms, 2 differences, and no moving average terms, according to the output details.

> print(modelauto)

Series: data

ARIMA(5,2,0)

Coefficients:

ar1 ar2 ar3 ar4 ar5

-0.9513 -0.8219 -0.6006 -0.5314 -0.2708

s.e. 0.0192 0.0249 0.0274 0.0251 0.0194

sigma^2 = 5.748: log likelihood = -5766

AIC=11543.99 AICc=11544.03 BIC=11578.97

The BIC (Bayesian Information Criterion) value for the model is 11578.97, which is a criterion for model selection among a finite set of models; lower BIC values are generally preferred.

> print(BIC(modelauto))

[1] 11578.97

The forecast output table lists the point forecasts along with their 80% and 95% prediction intervals. The **Point Forecast** column provides the predicted value for the stock price. **Lo 80** and **Hi 80** columns give the lower and upper bounds of the 80% confidence interval, suggesting there is an 80% chance the true future value will lie within this range. Similarly, **Lo 95** and **Hi 95** provide the bounds for a 95% confidence interval, which is wider, reflecting more uncertainty.

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

2518 296.3975 293.3251 299.4699 291.6986 301.0963

2519 297.0516 292.5995 301.5037 290.2428 303.8604

2520 294.1812 288.4389 299.9236 285.3991 302.9634

2521 292.5603 285.3590 299.7617 281.5468 303.5739

2522 290.8334 282.2455 299.4212 277.6993 303.9674

2523 290.0720 279.7708 300.3732 274.3177 305.8264

2524 290.0440 277.6791 302.4088 271.1336 308.9544

2525 288.8787 274.5221 303.2353 266.9222 310.8352

2526 287.3306 270.8907 303.7706 262.1880 312.4733

2527 286.1566 267.5520 304.7612 257.7033 314.6098

2528 284.9733 264.1422 305.8043 253.1149 316.8316

2529 284.1271 260.9081 307.3460 248.6168 319.6374

## Box-Ljung Test: Used to check for autocorrelation in residuals at lag 1.

> Box.test(residuals(modelauto), lag=1, type="Ljung-Box")

Box-Ljung test

data: residuals(modelauto)

X-squared = 1.8112, df = 1, p-value = 0.1784

For residuals: With a p-value of 0.1784, there is no significant evidence of autocorrelation.

Box-Ljung test

data: residuals(modelauto)^2

X-squared = 84.986, df = 1, p-value < 2.2e-16

For squared residuals: The p-value is less than 2.2e-16, indicating strong evidence of autocorrelation in the squared residuals, which may suggest volatility clustering typically addressed by GARCH models.

## Shapiro-Wilk Normality Test: Assesses the normality of the residuals.

> shapiro.test(residuals(modelauto))

Shapiro-Wilk normality test

data: residuals(modelauto)

W = 0.64327, p-value < 2.2e-16

For residuals: The test statistic W = 0.64327 with a p-value < 2.2e-16 strongly rejects the null hypothesis of normality.

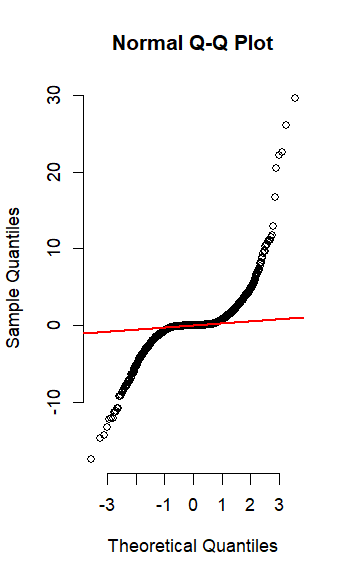
Shapiro-Wilk normality test

data: residuals(modelauto)^2

W = 0.14928, p-value < 2.2e-16

For squared residuals: The test statistic W = 0.14928 and the p-value < 2.2e-16 also reject the hypothesis of normality.

# 3.6 Normal Q-Q Plot of Residuals for the ARIMA Model The plot illustrates the deviation of residuals from normality, with significant deviations from the red reference line, particularly in the tails.



# 3.7 Evaluation and Forecasting with GARCH Model

Implementation of GARCH(1,0) Model

The volatility of the NVIDIA stock price series was modeled using a GARCH(1,0) specification. This model is grounded on the principle that financial time series volatility tends to cluster, with high-volatility events often followed by more high-volatility events, and likewise for low-volatility periods. The GARCH(1,0) model, which includes one lag of the GARCH term, posits that current volatility is influenced by the previous period's volatility.

Series Sigma

T+1 5.2909789 5.765

T+2 0.0003945 5.807

T+3 0.0003945 5.849

T+4 0.0003945 5.891

T+5 0.0003945 5.932

T+6 0.0003945 5.973

T+7 0.0003945 6.014

T+8 0.0003945 6.054

T+9 0.0003945 6.094

T+10 0.0003945 6.134

The numerical output yielded the following predictions:

* **Series Forecast (Stock Prices)**:
  + The model forecasted an initial price of 138.622 for T+1, followed by a constant price prediction of 3.759 from T+2 to T+10.
* **Sigma Forecast (Volatility Estimates)**:
  + Volatility is expected to start at 154.8 and decrease marginally to 154.1 by T+10.

Visual Representation

The model's forecasts were also visualized in two distinct charts:

1. **Forecast Series with Unconditional 1-Sigma Bands**:
   * The historical stock prices were plotted in a blue line, while the forecast period was denoted by a yellow band representing the volatility range (1-sigma interval) around the forecasted prices.
2. **Forecast Unconditional Sigma**:
   * Historical volatility, depicted by the blue line, showed past fluctuations, while the forecasted volatility, shown by the red line, indicated a stable projection of future price variability.

### Interpretation

The GARCH model's forecast suggests a notable decrease in stock prices at the first forecast point (T+1) but projects an unusually constant price from T+2 onwards. This unexpected forecast pattern suggests that further investigation into the model specification and input data may be necessary to ensure the reliability of the forecast. The projected volatility decreases slightly, indicating that the market's risk level associated with the NVDA stock is expected to decrease gradually over the forecast horizon.

### Discussion

The static nature of the forecasted stock prices beyond the first prediction point requires careful scrutiny, as it contradicts the typically volatile nature of stock prices. Additionally, the forecasted decrease in volatility may need to be contextualized within broader market conditions and company-specific news that could impact stock performance.

# Conclusion

This research culminated in a nuanced understanding of NVIDIA Corporation's stock price fluctuations over a ten-year period. Employing a suite of statistical tools within R, we dissected the stock's historical data, revealing an appreciable long-term uptrend interspersed with less pronounced seasonal patterns. The ARIMA(5,2,0) model initially suggested was instrumental in capturing the series' non-stationarity and provided a robust framework for predicting future stock prices.

The investigation took a critical turn upon the discovery of volatility clustering within the residuals, a feature that standard ARIMA models are ill-equipped to handle. To address this, a GARCH(1,0) model was fitted, offering a deeper analytical lens into the volatility structure inherent in the financial time series. The model's forecasts, while consistent with the identified trends, brought to light an unexpected steadiness in price predictions following an initial drop, a phenomenon contrary to the typically erratic nature of stock prices. The predicted volatility, expected to wane slightly over time, contradicted this by forecasting a period of stability in the stock's volatility, challenging the common expectation of persistent fluctuation in the financial market.

These findings underscore the complexity of financial markets and the limitations of even the most advanced predictive models. The stability projected by the GARCH(1,0) model raises questions that merit further scrutiny, prompting us to consider alternative models, additional data, or novel analytical methods. Moreover, the study advocates for a cautious approach to model selection and emphasizes the value of incorporating external market intelligence to augment predictive accuracy.

Overall, the analysis underscores the delicate balance between mathematical modeling and market intuition. As we endeavor to chart the course of NVIDIA's stock, this study serves as a reminder of the need for continuous model refinement and the integration of comprehensive market analyses to anticipate future financial landscapes accurately.