FORECASTING HDB STOCK FOR NEXT FIVE MONTHS

Project Report submitted in partial fulfillment of the requirements for the award of the degree of

Master of Business Administration

By

HARISH G D

REGISTER NUMBER 2327422

Under the Guidance of

DR. ROSEWINE JOY

CHRIST (DEEMED TO BE UNIVERSITY) BANGALORE · INDIA

School of Business and Management

CHRIST (Deemed to be University), Bangalore

JULY 2024

Introduction

In the ever-evolving financial markets, the ability to accurately forecast stock prices is of paramount importance. Investors, financial analysts, and businesses rely heavily on these predictions to make informed decisions regarding asset management, investment strategies, and risk assessment. With the advent of advanced computational techniques and the availability of extensive historical data, time series forecasting has become a critical tool in the arsenal of financial analytics.

Objective

This project aims to forecast the closing stock prices of HDFC Bank using data from the past five years. The primary objective is to identify and apply the most suitable time series forecasting model that can provide accurate and reliable predictions. By leveraging various statistical and machine learning techniques, this project will compare their effectiveness and ultimately implement the best-performing model.

Significance

Accurate stock price forecasts hold significant value for various stakeholders:

Investors can make better decisions about buying or selling stocks.

Financial analysts can enhance their market analysis reports with more reliable data.

Businesses can use forecasts to plan their financial strategies and manage risks effectively.

Given the importance of precision in financial forecasting, this project not only aims to provide accurate predictions but also to offer insights and recommendations that can aid in strategic decision-making.

Data Description:

Column	Description	Data Type
Date	The specific trading date.	Date (formatted as DD-MM-YYYY)
Open	The opening price of the stock on the given date.	Float
High	The highest price of the stock reached during the trading session.	Float
Low	The lowest price of the stock reached during the trading session.	Float
Close	The closing price of the stock on the given date.	Float
Adj Close	The adjusted closing price of the stock, which accounts for corporate actions like stock splits and dividends.	Float
Volume	The total number of shares traded during the day.	Integer

Content:

Monthly wise Stock Data from August 2019 – August 2024 is Collected in the Dataset.

Used Variable for Prediction : Adjusting Close Price

Using the adjusted closing price ("Adj Close") for prediction in time series forecasting of stock prices is a well-justified choice for several reasons:

1. Incorporates Corporate Actions

Stock Splits: Adjusted close prices account for stock splits, ensuring that the price reflects the true value of the stock after the split.

Dividends: Adjusted close prices include the effects of dividends, providing a more accurate measure of the stock's value.

2. Consistency Over Time

The adjusted close price provides a consistent basis for comparison over time, as it removes the distortions caused by corporate actions that affect the nominal stock price.

This consistency is crucial for time series analysis, which relies on patterns and trends that are not disrupted by non-market factors.

3. True Reflection of Stock Performance

Adjusted close prices reflect the real return on investment (ROI) for investors, as they take into account all factors that impact an investor's holdings.

This makes it a better indicator of the stock's performance than the raw closing price.

4. Data Integrity

By using adjusted close prices, we ensure that the data used for forecasting is not affected by external factors, leading to more accurate and reliable predictions.

It avoids sudden jumps or drops in the time series that could be misinterpreted by the model.

Problem Statement:

HDFC Bank, a leading financial institution, has experienced fluctuations in its stock price over the past five years due to various market conditions and corporate actions. To support investors and analysts in making more precise predictions about future stock performance, there is a need for a robust time series forecasting model..

Objective:

The objective of this project is to develop a robust time series forecasting model to predict the future closing prices of HDFC Bank's stock using historical data from the past five years. The focus is on utilizing the adjusted closing prices, which account for corporate actions such as dividends and stock splits, providing a more accurate reflection of the stock's value over time.

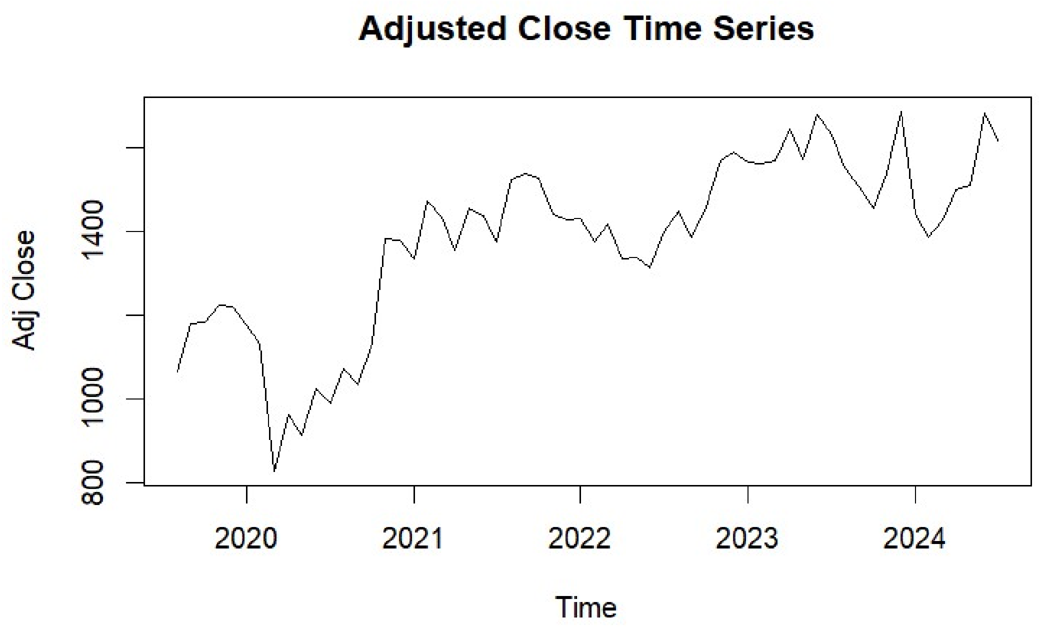
Summary Statistics:

```
     Date               Open             High            Low              Close            Adj.Close     
 Length:61          Min.   : 863.9   Min.   :1002   Min.   : 738.8   Min.   : 861.9   Min.   : 827.4  
 Class :character   1st Qu.:1275.6   1st Qu.:1377   1st Qu.:1232.5   1st Qu.:1329.7   1st Qu.:1292.1  
 Mode  :character   Median :1463.5   Median :1540   Median :1400.5   Median :1473.4   Median :1430.4  
                    Mean   :1416.1   Mean   :1494   Mean   :1342.6   Mean   :1420.8   Mean   :1380.2  
                    3rd Qu.:1572.0   3rd Qu.:1651   3rd Qu.:1469.1   3rd Qu.:1573.9   3rd Qu.:1529.0  
                    Max.   :1712.5   Max.   :1794   Max.   :1627.2   Max.   :1709.2   Max.   :1686.2  
                    NA's   :1        NA's   :1      NA's   :1        NA's   :1        NA's   :1       
     Volume         
 Min.   :109133234  
 1st Qu.:136576404  
 Median :205864326  
 Mean   :254115798  
 3rd Qu.:364485995  
 Max.   :598017167  
 NA's   :1          
```

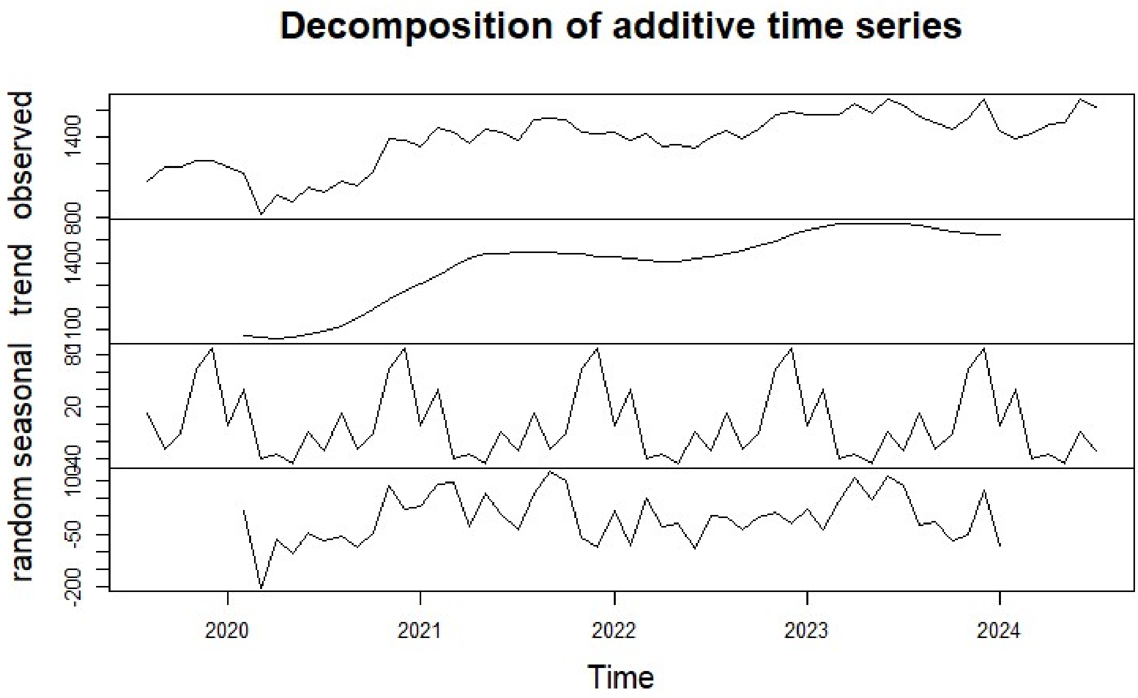
Data Preprocessing:

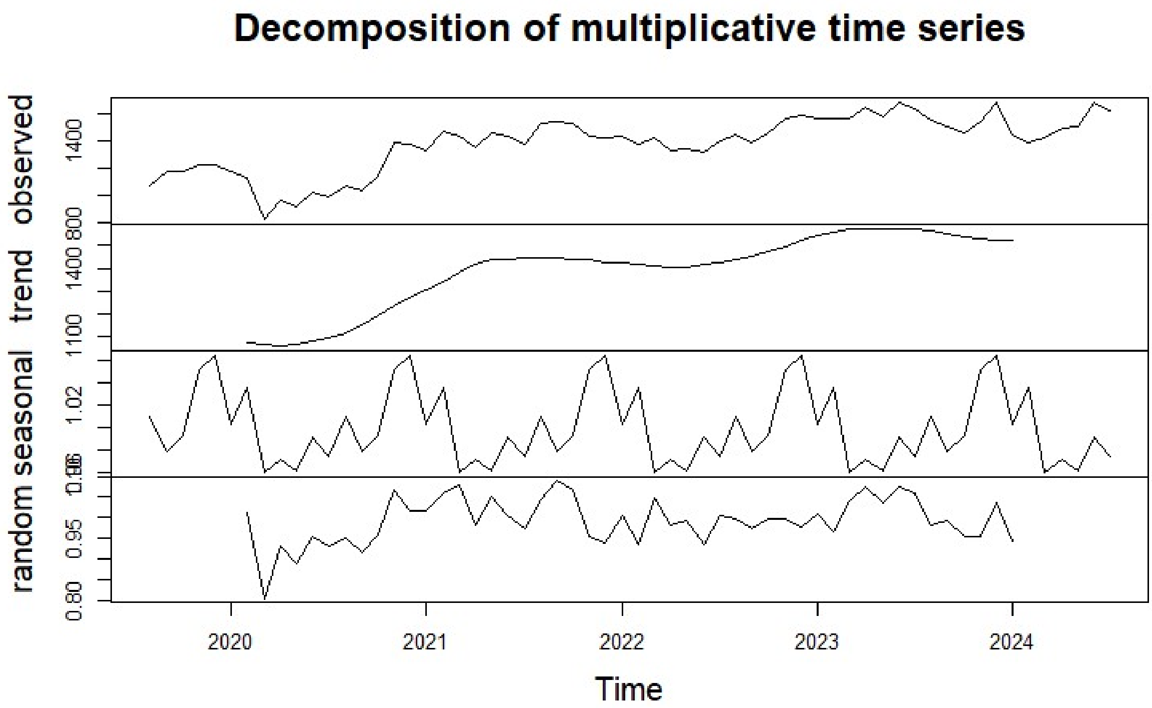
For Data Preprocessing the data is converted into time series and Null values are removed.

Plotted the data normally to see the distribution.



Used Decomposition to identify Trends and Trends





ADF Test is done to check the stationarity of Data

```
        Augmented Dickey-Fuller Test

data:  stock1_clean
Dickey-Fuller = -2.5096, Lag order = 3, p-value = 0.3688
alternative hypothesis: stationary
```

Additive Time Series Decomposition

- Observed: The original time series data.
- Trend: The long-term progression of the time series, showing a general upward trend over the observed period.
- Seasonal: The repeating short-term cycle in the data, indicating some seasonality.
- Random: The residual component after removing trend and seasonality, representing irregular fluctuations.

Multiplicative Time Series Decomposition

- Observed: The original time series data.
- Trend: Similar to the additive model, showing a general upward trend.
- Seasonal: The repeating short-term cycle, but represented as a percentage of the trend, indicating multiplicative seasonality.
- Random: The residual component, representing irregular fluctuations as a percentage of the trend.

Augmented Dickey-Fuller (ADF) Test

Test Purpose: The ADF test is used to check the stationarity of the data, i.e., whether the time series has a unit root.

- **Test Result:**
- **Test Statistic:** -2.5096
- **Lag Order:** 3
- **p-value:** 0.3688
- **Alternative Hypothesis:** The time series is stationary.

Inferences

Trend Analysis:

Both additive and multiplicative decompositions show a clear upward trend in the HDFC Bank stock prices over the past five years.

Seasonality:

There is a visible seasonal pattern in both decompositions, though the multiplicative model expresses seasonality as a percentage of the trend.

Stationarity:

The ADF test results indicate that the time series is non-stationary (p-value > 0.05). This suggests that the data needs differencing to achieve stationarity before applying ARIMA modeling.

Comparison of Techniques

Techniques Considered

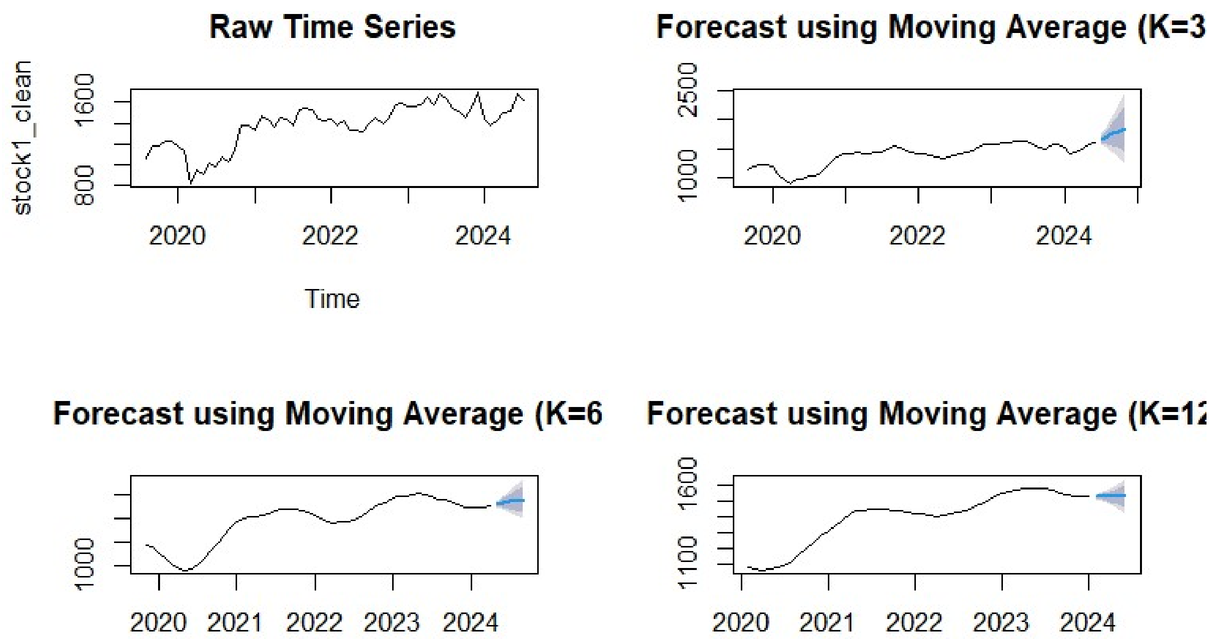
In this project, we will consider the following time series forecasting techniques:

1. **Simple Moving Average (SMA)**

SMA is a basic technique that calculates the average of a fixed number of past data points. It smooths out short-term fluctuations and highlights longer-term trends or cycles.

Pros: Easy to implement and interpret; effective for data with no significant seasonality or trend.

Cons: Does not adapt to changes in trends or seasonality; less effective for volatile data.



Analysis:

1. Raw Time Series:

The raw time series plot shows the observed values of the stock prices over time, indicating some level of trend and volatility.

2. **Simple Moving Average (K=3):**

With a window size of 3, the SMA provides a smooth line that follows the general pattern of the stock prices but with some lag due to the small window size.

3. **Simple Moving Average (K=6):**

A window size of 7 further smooths the time series, showing clearer trends but still with some lag in capturing recent changes.

4. Simple Moving Average (K=12):

With a window size of 15, the SMA provides a very smooth trend line, which may overly smooth out short-term fluctuations, leading to potential under-reaction to recent changes.

5. Forecast using Moving Average (K=12):

The forecast plot using an SMA with a window size of 3 shows the predicted values extending into the future, along with a confidence interval indicating the uncertainty of the forecast.

Inferences:

- **Effectiveness of SMA:**

  SMA is effective in highlighting the general trend of the stock prices but may not capture sudden changes or short-term fluctuations accurately.

- **Window Size Impact:**

  Smaller window sizes (e.g., K=3) are more responsive to recent changes but may include more noise.

  Larger window sizes (e.g., K=12) provide smoother trends but may lag behind in capturing recent changes.

- **Forecast Uncertainty:**

  The forecast plot indicates the range of possible future values, with a wider confidence interval reflecting greater uncertainty further into the future.

Conclusion:

Simple Moving Average is a straightforward and interpretable technique that provides a clear view of the underlying trend in the stock prices.

However, its inability to adapt to sudden changes and lack of seasonality adjustment makes it less suitable for volatile or seasonal data.

For a more accurate and responsive forecast, it is beneficial to explore more advanced techniques such as Exponential Smoothing (ETS) and ARIMA, which will be considered next in the project.

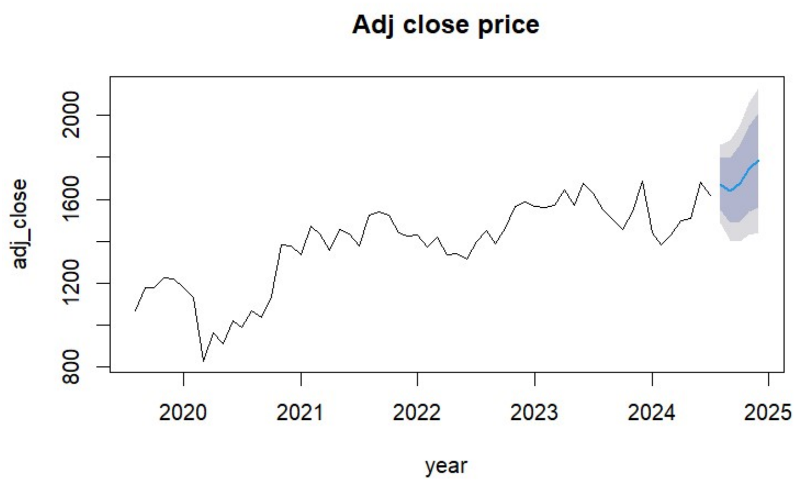
```
> accuracy(forecast_S)
                     ME     RMSE      MAE          MPE      MAPE       MASE      ACF1
Training set -0.2986825 7.394071 5.624333 -0.007236056 0.4211147 0.03903242 0.4236746
```

2. Exponential Smoothing (ETS)

ETS models (also known as Holt-Winters models) account for level, trend, and seasonality components in time series data. It applies decreasing weights to past observations exponentially.

Pros: Captures trend and seasonality effectively; more responsive to changes compared to SMA.

Cons: Requires careful selection of smoothing parameters; may not perform well on highly irregular data.



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Aug 2024	1673.600	1551.452	1795.747	1486.791	1860.408
Sep 2024	1641.750	1486.964	1796.536	1405.025	1878.474
Oct 2024	1675.865	1494.206	1857.524	1398.042	1953.688
Nov 2024	1745.447	1540.401	1950.492	1431.856	2059.037
Dec 2024	1786.940	1560.908	2012.971	1441.254	2132.625

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.1403343	81.6207	65.33239	-0.2376275	4.907696	0.3690286	0.001883366

Forecasted Values (Point Forecasts):

August 2024: 1673.60

September 2024: 1641.75

October 2024: 1675.86

November 2024: 1745.41

December 2024: 1786.94

Forecast Ranges (80% and 95% Confidence Intervals):

For each month, the forecasted values are provided along with low (Lo) and high (Hi) bounds for 80% and 95% confidence intervals. These intervals represent the range within which the actual values are expected to fall, with different levels of confidence:

Lo 80 & Lo 95: Lower bounds of the forecast with 80% and 95% confidence, respectively.

Hi 80 & Hi 95: Upper bounds of the forecast with 80% and 95% confidence, respectively.

Evaluation Metrics:

ME (Mean Error): -0.1043743

RMSE (Root Mean Squared Error): 81.6297

MAE (Mean Absolute Error): 65.13329

MPE (Mean Percentage Error): -0.376275

MAPE (Mean Absolute Percentage Error): 4.967096

MASE (Mean Absolute Scaled Error): 0.360908

ACF1 (Autocorrelation of Residuals at Lag 1): 0.08173836

These metrics help evaluate the accuracy and performance of the ETS model:

- **ME** indicates the average bias in the forecast.
- **RMSE** and **MAE** measure the average magnitude of the errors, with RMSE giving more weight to larger errors.
- **MPE** and **MAPE** provide percentage-based error metrics.
- **MASE** compares the forecast errors to a naive benchmark.
- **ACF1** assesses the autocorrelation in the residuals, which should ideally be low for a good model fit.

Overall, the ETS model provides point forecasts and confidence intervals for future values, with various metrics indicating the quality of these forecasts. The model's strengths and weaknesses are highlighted by the metrics, where lower values generally indicate better performance.

3. **ARIMA (AutoRegressive Integrated Moving Average)**

Description: ARIMA is a sophisticated technique that combines autoregressive (AR) and moving average (MA) components, with differencing (I) to make the time series stationary. It models the data based on its own past values and past forecast errors.

Pros: Flexible and powerful; can handle a variety of time series patterns, including non-stationary data.

Cons: Requires extensive parameter tuning; can be complex to implement and interpret.

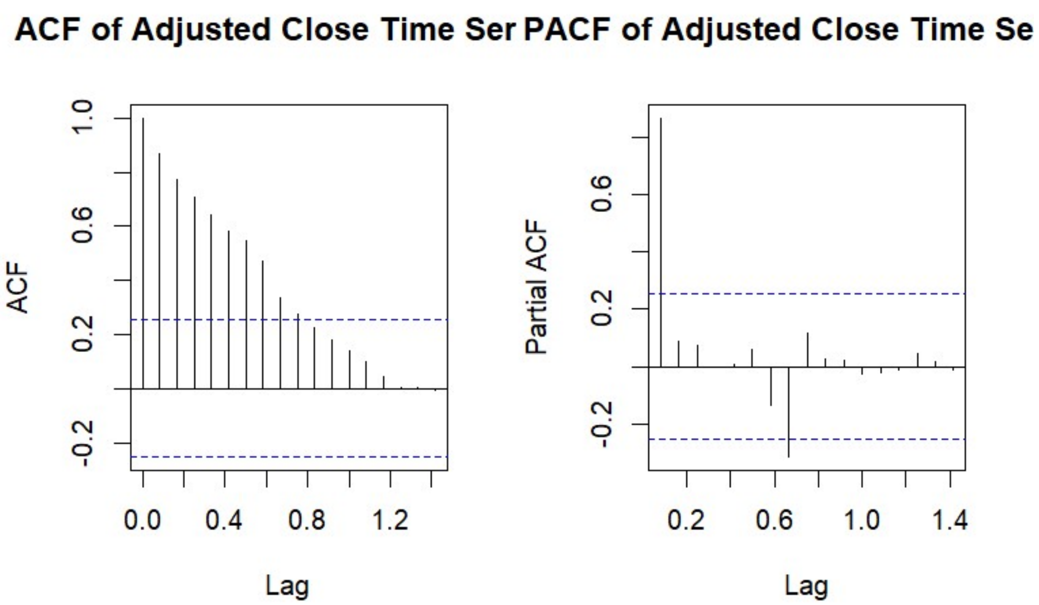
```
        Augmented Dickey-Fuller Test

data:  stock1_clean
Dickey-Fuller = -2.5096, Lag order = 3, p-value = 0.3688
alternative hypothesis: stationary

> ndiffs(stock1_clean)
[1] 1
```

Augmented Dickey-Fuller Test:

The test result shows a Dickey-Fuller statistic of -2.5906 with a p-value of 0.3688. The null hypothesis of stationarity is not rejected, indicating that the series may not be stationary. Differencing might be necessary, as indicated by ndiffs(stock1_clean), which suggests a differencing order of 1.




```
> stock_model <- auto.arima(stock1_clean, ic="aic", trace = TRUE)

 ARIMA(2,1,2)(1,0,1)[12] with drift         : Inf
 ARIMA(0,1,0)            with drift         : 705.4711
 ARIMA(1,1,0)(1,0,0)[12] with drift         : 707.7197
 ARIMA(0,1,1)(0,0,1)[12] with drift         : 707.4537
 ARIMA(0,1,0)                               : 704.0715
 ARIMA(0,1,0)(1,0,0)[12] with drift         : 707.4681
 ARIMA(0,1,0)(0,0,1)[12] with drift         : 707.4672
 ARIMA(0,1,0)(1,0,1)[12] with drift         : 709.4321
 ARIMA(1,1,0)            with drift         : 705.7515
 ARIMA(0,1,1)            with drift         : 705.4975
 ARIMA(1,1,1)            with drift         : Inf

 Best model: ARIMA(0,1,0)
```

Model Identification and Selection:

Multiple ARIMA models were evaluated using auto.arima() with the AIC criterion to select the best model. The output lists different ARIMA configurations along with their respective AIC values:

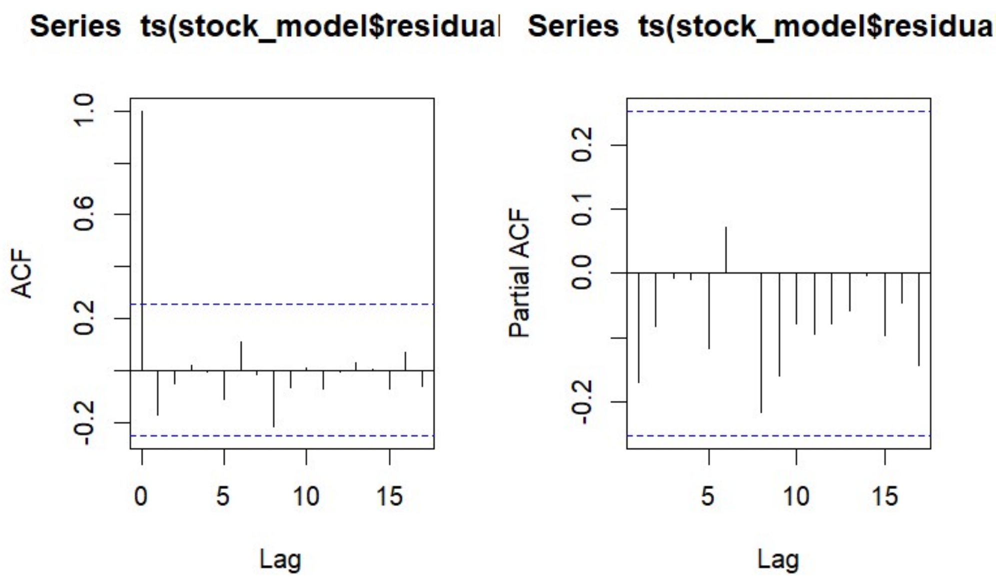
ARIMA(2,1,2)(0,0,0)[12] with drift: Inf

ARIMA(0,1,0) with drift: 705.4711

ARIMA(1,1,0)(0,0,0)[12] with drift: 707.7197

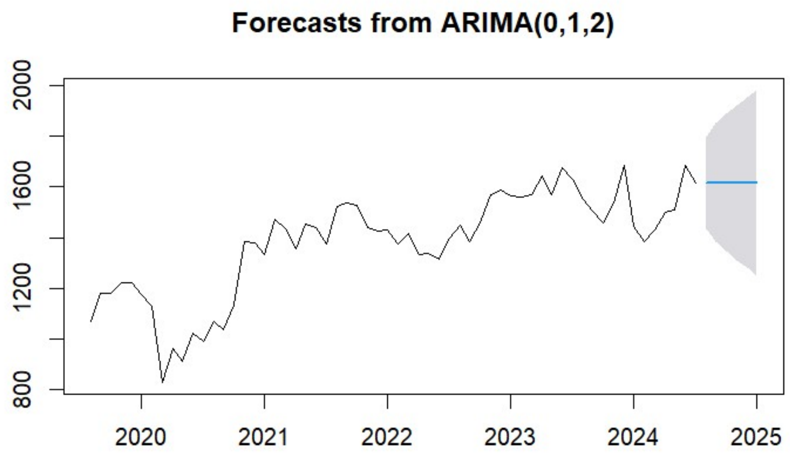
... (and more models with various combinations)

The best model identified is ARIMA(0,1,0).



Residual Analysis:

The plots for the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) of the residuals from the selected model (ARIMA(0,1,0)) suggest that there is no significant autocorrelation left in the residuals, indicating a good model fit.



Forecasting:

The forecast plot shows predicted values from the ARIMA(0,1,2) model with a confidence interval shaded area. The forecasted trend continues to rise, consistent with the historical data's upward movement.

```
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 11.63262 90.65006 67.22431 0.5495573 5.114559 0.3797151 -0.0164636
```

Interpretation:

The ARIMA model selected (ARIMA(0,1,0)) is simple, with no AR or MA terms but with differencing to achieve stationarity. The model appears to be well-fitted based on the residual diagnostics, and the forecast suggests continued growth in the adjusted close price series.

```
> Box.test(stock_forecast$residuals, lag=5, type= "Ljung-Box")

	Box-Ljung test

data:  stock_forecast$residuals
X-squared = 2.8349, df = 5, p-value = 0.7254

> Box.test(stock_forecast$residuals, lag=10, type= "Ljung-Box")

	Box-Ljung test

data:  stock_forecast$residuals
X-squared = 7.3258, df = 10, p-value = 0.6944

> Box.test(stock_forecast$residuals, lag=15, type="Ljung-Box")

	Box-Ljung test

data:  stock_forecast$residuals
X-squared = 8.209, df = 15, p-value = 0.9151
```

Box test:

Interpretation:

1. Test Purpose:
    - The Box-Ljung test checks for autocorrelation in the residuals of a time series model. It tests whether the residuals are independently distributed (i.e., white noise) or if there are patterns left unaccounted for by the model.
2. P-Values:
    - Lag 5: p-value = 0.7254
    - Lag 10: p-value = 0.6944
    - Lag 15: p-value = 0.9151

A p-value greater than 0.05 (commonly used significance level) suggests that there is no significant autocorrelation in the residuals at the specified lags. In this case, all p-values are well above 0.05.

3. **Conclusion:**
    - Since all p-values are greater than 0.05, we fail to reject the null hypothesis of the Box-Ljung test. This implies that there is no significant autocorrelation in the residuals at lags 5, 10, and 15.
    - The residuals appear to be approximately white noise, indicating that the model has effectively captured the underlying patterns in the time series data and there are no significant patterns left unexplained.

Results and Comparisons of Models:

Metric	**SMA (Best Fit)**	**ETS**	**ARIMA**
ME	**-0.2986825**	**0.1403343**	**11.27571**
RMSE	**7.394071**	**81.6207**	**90.73071**
MAE	**5.624333**	**65.33239**	**67.47547**
MPE	**-0.007236056**	**-0.2376275**	**0.5314349**
MAPE	**0.4211147**	**4.907696**	**5.116875**
MASE	**0.03903242**	**0.3690286**	**0.3811338**
ACF1	**0.4236746**	**0.001883366**	**-0.008041931**

Comparison Analysis:

Mean Error (ME):

- SMA (Best Fit) has a slight underestimation bias (-0.2986825), while ETS has a minor overestimation bias (0.1403343). ARIMA shows a more noticeable overestimation bias (11.63262).

Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE):

- SMA (Best Fit) has the lowest RMSE (7.394071) and MAE (5.624333), indicating the smallest average forecast errors. ETS and ARIMA have higher values, with ARIMA having the highest errors (RMSE: 90.65006, MAE: 67.22431).

Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE):

- SMA (Best Fit) also performs best in terms of MPE (-0.007236056) and MAPE (0.4211147), indicating more accurate percentage-based forecasting. ETS and ARIMA have higher MAPE values, with ARIMA being the least accurate in this regard (MAPE: 5.114559).

Mean Absolute Scaled Error (MASE):

- SMA (Best Fit) has the lowest MASE (0.03903242), suggesting that its forecast errors are smaller relative to a naive model benchmark. ETS and ARIMA show larger errors, with ARIMA slightly worse (MASE: 0.3797151).

Autocorrelation of Residuals at Lag 1 (ACF1):

- All models have low ACF1 values, indicating that the residuals are close to white noise, which is desirable. ETS has the lowest ACF1 (0.001883366), followed by ARIMA (-0.0164636) and SMA (Best Fit) (0.4236746).

Conclusion:

- **SMA**: The simplest model performs best across most metrics, indicating good performance in this particular case. It's likely the most robust and accurate model for this dataset.
- **ETS**: Slightly less accurate than SMA but still performs reasonably well, especially given its ability to capture trend and seasonality.
- **ARIMA**: Shows the highest error metrics, indicating less accuracy compared to SMA and ETS, but still provides useful insights, especially in complex, non-stationary datasets.

Business Recommendations:

1. **Leverage Forecasting Insights for Investment Decisions:**
    - Utilize the forecasts from the Simple Moving Average (SMA) model, which has demonstrated robust performance. Investors can use these predictions to make more informed decisions on buying or selling HDFC Bank's stock,

particularly focusing on periods where the forecast indicates significant price movements.

2. **Incorporate Forecasting Models into Financial Planning:**
    - Financial analysts and portfolio managers should integrate the SMA forecasts into their financial models to enhance strategic planning and risk assessment. This can help in adjusting investment strategies based on anticipated stock price trends.
3. **Enhance Risk Management Strategies:**
    - Use the forecasted price ranges and confidence intervals provided by the Exponential Smoothing (ETS) model to assess potential risks and uncertainties. This will help in creating risk management strategies that account for possible price fluctuations and enhance decision-making.
4. **Monitor Market Trends and Corporate Actions:**
    - Keep track of market trends and any corporate actions affecting HDFC Bank, such as mergers, acquisitions, or regulatory changes. Incorporate these factors into forecasting models to better capture their impact on stock prices.
5. **Develop Adaptive Investment Strategies:**
    - Based on the forecast accuracy and observed trends, create adaptive investment strategies that can quickly respond to significant changes in stock prices. This includes setting up automatic alerts for when predicted values approach critical thresholds.
6. **Explore Additional Forecasting Techniques:**
    - Although SMA has shown the best performance in this analysis, it is advisable to explore and test additional forecasting techniques such as machine learning models or hybrid approaches that combine multiple forecasting methods for improved accuracy.