**Project Report**

**On**

# A STATISTICAL INVESTIGATION OF FORD MOTOR'S STOCK PRICE DYNAMICS (1972-2024)

***Submitted***

***in partial fulfilment of the requirements for the degree of***

**MASTER OF SCIENCE**

***in***

**APPLIED STATISTICS AND DATA ANALYTICS**

***by***


**AISWARYA C G**

**(Register No. SM23AS002)**

**(2023-2025)**

***Under the Supervision of***

**TANIA P R**



**DEPARTMENT OF MATHEMATICS AND STATISTICS**

**ST. TERESA'S COLLEGE (AUTONOMOUS)**

**ERNAKULAM, KOCHI – 682011**


**APRIL 2025**

## ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

# CERTIFICATE

This is to certify that the dissertation entitled, **A STATISTICAL INVESTIGATION OF FORD MOTOR'S STOCK PRICE DYNAMICS (1972-2024)** is a Bonafide record of the work done by **Ms. AISWARYA C G** under my guidance as partial fulfilment of the award of the degree of **Master of Science in Applied Statistics and Data Analytics** at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date:

Place: Ernakulam

**Tania P R**
Assistant professor,
Department of Mathematics & Statistics
St. Teresa's College (Autonomous),
Ernakulam



**Nisha Oommen**
Assistant professor,
Department of Mathematics & Statistics
St. Teresa's College (Autonomous),
Ernakulam

**External Examiners**

1. Sangeetha Chandran ........................ 30.4.25

2. Anju N B ........................ 30-04-25

ST. TERESA'S COLLEGE (AUTONOMOUS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
ERNAKULAM

# DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of **Mrs. TANIYA P R**, Assistant Professor, Department of Mathematics and Statistics, St. Teresa's College (Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Place: Ernakulam

Date: 30/04/2025

**AISWARYA C G**

**SM23AS002**

# ACKNOWLEDGEMENTS

I must mention several individuals who encouraged me to carry out this work. Their continuous invaluable knowledgeable guidance throughout this study helped me to complete the work up to this stage.

I am very grateful to my project guide Mrs. Tania P R for the immense help during the period of work.

In addition, the very energetic and competitive atmosphere of the Department had much to do with this work. I acknowledge with thanks to the faculty, teaching and non-teaching staff of the department and colleagues.

I am also very thankful to the HOD Mrs. Nisha Oommen for their valuable suggestions, critical examination of work during the progress.

Place: Ernakulam

Date: 30/04/2025

**AISWARYA C G**

**SM23AS002**

# ABSTRACT

Stock prices are an essential indicator of a company's financial health and market performance. They fluctuate based on various factors, including company earnings, economic conditions, and investor sentiment. Forecasting Stock price is a critical component in financial decision-making, enabling investors to anticipate future trends and optimize their investment strategies. This project concentrates on time series analysis in light of forecasting the future average stock price of Ford Motor Company over the period of 1972-2024 using statistical methods, ARIMA and Holt-Winter Exponential Smoothing, in particular. The objective of the project was to analyse the data for the time period of 1972-2024, forecast the production for 2025-2028 and compare findings made with the help of ARIMA and Holt Winter Exponential Smoothing. ARIMA and Holt-Winter Exponential Smoothing were necessary for the project to analyse the data, understand trends and patterns in the data. To begin with, ARIMA and Holt-Winter Exponential Smoothing will be described from the theoretical perspective and then each method will be applied to the data. In the end, the analysis will be compared to one another.

# INDEX

**Contents** **Page No.**

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

Stock price prediction plays a crucial role in the financial sector, providing investors, traders, and policymakers with valuable insights into market trends and potential investment opportunities (Faris, 2023). The ability to forecast stock price movements enables investors to make informed decisions, manage risks, and optimize their portfolios for better returns. Over the years, various methods have been explored to analyse stock price movements, including fundamental analysis, technical analysis, and machine learning-based forecasting techniques (Nti et al., 2020). These approaches allow researchers to identify patterns, assess market volatility, and develop predictive models that enhance investment strategies. Given the complexities of stock market dynamics, understanding historical price movements is essential for improving forecasting accuracy and minimizing financial risks (Edwards et al.,2018).

Ford Motor Company, one of the leading automobile manufacturers, has experienced fluctuations in its stock prices due to various macroeconomic and industry-specific factors (Vychytilová et al., 2019). Factors such as changes in consumer demand, technological advancements, economic conditions, and global trade policies significantly influence the company's stock performance. Additionally, market inefficiencies and investor sentiment play a crucial role in stock price movements (Agustin, 2019). Analysing historical stock price data of Ford provides an opportunity to identify trends, seasonality, and potential anomalies that may impact future stock price movements (Epps, 1979). A comprehensive understanding of these patterns helps investors and financial analysts to anticipate market shifts and adjust their trading strategies accordingly.

This study aims to analyse the stock price dynamics of Ford Motor Company, focusing on identifying patterns, trends, and seasonality in its historical stock prices. By employing statistical and machine learning techniques such as the Autoregressive Integrated Moving Average (ARIMA) model, Linear Regression, and Random Forest, this research seeks to develop and evaluate forecasting models for stock price prediction

(Nti et al., 2020). ARIMA, a widely used time series model, is effective in capturing linear dependencies in stock prices, while machine learning models such as Random Forest can capture complex non-linear relationships in data. Comparing these models will provide insights into their predictive performance and suitability for stock price forecasting (Dash & Sundarka, 2015).

The findings of this study will contribute to a better understanding of stock market behaviour by assessing the effectiveness of different forecasting methodologies. By leveraging historical data, investors and financial analysts can enhance their decision-making process and develop more accurate prediction models. Moreover, this research will provide valuable insights into the financial stability of Ford Motor Company, helping stakeholders make informed investment choices based on data-driven forecasting techniques.

## 1.1 OBJECTIVES

The main objectives of the study are as follows:

1. To perform Exploratory Data Analysis (EDA) on historical average stock price data to identify patterns, trends, and seasonality through visualization, time series decomposition, and stationarity tests with necessary transformation.
2. To develop and evaluate predictive models for stock price forecasting by fitting ARIMA model to the historical data.
3. To develop and evaluate predictive models for stock price forecasting by fitting Holt's Winter Exponential Smoothing model to the historical data.
4. To compare the performance of ARIMA and Holt's Winter Exponential Smoothing models using RMSE and MSE.

# CHAPTER 2
# REVIEW OF LITERATURE

This chapter presents the findings from related research that analysed various stock price datasets and applied different statistical methods, data mining techniques, and machine learning algorithms for stock price prediction.

Epps (1979) analysed short-term stock price co-movements, finding that correlations among price changes in common stocks decrease as the measurement interval lengthens. The study attributed this phenomenon to non-stationarity in security price changes and lagged adjustments in stock prices to relevant information. The findings contribute to understanding market microstructure and highlight inefficiencies in intraday stock price movements.

Wang et al. (2014) explored how integrating fundamental and technical analysis can improve risk-adjusted returns for investors in Taiwan's stock market. Using data from the Taiwan Economic Journal Database (1991–2009), they examined firm-level financial factors such as profitability, operating efficiency, and solvency. The study found that incorporating volume-price covariance as a trading signal led to portfolios that consistently outperformed the market. These findings support the effectiveness of combined investment strategies in achieving superior long-term returns.

Acharya et al. (2015) investigated the impact of the downgrade of General Motors (GM) and Ford to junk status in 2005 on corporate bond markets. Using a novel dataset, they documented a widespread sell-off in GM and Ford bonds, leading to significant liquidity risk for market-makers. Their findings indicated a substantial increase in co-movement between credit default swap (CDS) spreads across industries, particularly during the downgrade period. The study demonstrated that corporate bond market makers' imbalance towards sales in GM and Ford bonds explained a significant portion of this co-movement. These results supported models in which market prices are influenced by the limited risk-bearing capacity of financial intermediaries.

Dash and Sundarka (2015) examined the stationarity of beta in the Indian automotive and auto-ancillary sectors to assess the reliability of the Capital Asset Pricing Model (CAPM). Analysing eleven stocks over a decade, the study segmented the period into different market regimes—stagnant, growth, boom, depression, and steady phases—and applied univariate ANCOVA. The results indicated that beta values remained relatively stationary across different market regimes, suggesting that systematic risk factors in the Indian stock market exhibit stability within the automotive sector.

Drakopoulou (2016) highlighted the importance of fundamental analysis in daily equity trading. The study argued that while technical analysis helps traders identify patterns and trends, fundamental analysis provides insights into a stock's intrinsic value. The research advocated for a combined approach, suggesting that traders who integrate both methodologies can make more informed investment decisions and achieve better portfolio performance.

Edwards et al. (2018) provided an extensive guide to technical analysis in their book *Technical Analysis of Stock Trends*. The book covers key concepts such as the Dow Theory, reversal patterns, consolidation formations, trend channels, and technical indicators. It also introduces trading tactics, portfolio diversification strategies, and risk management techniques. This comprehensive work reinforces the importance of technical analysis in stock trading and provides investors with tools to enhance their decision-making processes.

Agustin (2019) explored the predictive ability of fundamental and technical analysis in stock price forecasting. The study focused on companies listed in the LQ45 index from 2007 to 2016, employing fundamental variables such as Earnings per Share (EPS), Dividend Payout Ratio (DPR), and Return on Equity (ROE), alongside technical variables including price momentum, positive extreme price increase (D-Up), and negative extreme price decline (D-Down). The findings revealed that technical analysis produced the highest predictive ability, while an integrated model combining fundamental and technical analysis outperformed fundamental analysis alone. This suggests that investors can achieve optimal stock returns by incorporating both methodologies.

Vychytilová et al. (2019) investigated macroeconomic factors influencing stock volatility in the automotive industry. Using quarterly panel data from 39 automakers across 11 countries between 2000 and 2017, they applied a mixed-effect model optimized with a genetic algorithm and AIC criterion. Their findings revealed positive correlations between stock return volatility and factors such as GDP growth, stock market development, and unemployment, while money supply and industrial production index (IPI) exhibited inverse relationships. The study demonstrated that macroeconomic factors play a crucial role in explaining stock price fluctuations in the automotive sector.

Nti et al. (2020) conducted a systematic review of stock market prediction methodologies using machine learning. Analysing 122 studies published between 2007 and 2018, they categorized methodologies into technical, fundamental, and hybrid approaches. The findings indicated that 66% of studies focused on technical analysis, 23% on fundamental analysis, and 11% on hybrid approaches. Support vector machines and artificial neural networks were the most commonly used machine learning algorithms. The study concluded that integrating fundamental and technical indicators enhances predictive accuracy in stock market forecasting.

Faris (2023) examined the financial performance of Ford Motor Company using both fundamental and technical analysis. The study covered the period from 2017 to 2021 and employed financial statement analysis, volatility modelling, and chart visualization. It found that in 2020, Ford had the lowest payout gap (0.15) and a weak EBIT margin (1.50%), leading investors to execute technical volatility strategies. The research emphasized the role of financial indicators in shaping stock valuation and investment risk assessment.

# CHAPTER 3
# MATERIALS AND METHODS

## 3.1 DATA COLLECTION

The data for this study was sourced from historical annual average stock price records of Ford Motor Company (F) from 1972 to 2024. The dataset used in this study is taken from a website called Yahoo Finance.

The dataset includes annual average stock prices from 1972 to 2024, enabling a comprehensive analysis of long-term trends, patterns, and forecasting potential. The prices are given in USD ($).

Below are all the features of the data:

1. Year: Year refers to the specific year in which the stock price data is recorded.
2. Annual Average Stock Price: Average Stock Price is the mean price of Ford Motor Company's stock over that entire year.

## 3.2 METHODOLOGY

The first step in this study was to carefully examine the dataset to understand the historical trends in Ford Motor Company's stock prices. The goal was to analyse past stock price movements and predict future prices using time series models.

To start, exploratory data analysis (EDA) was conducted to identify trends, patterns, and seasonality in the data. This helped in understanding how stock prices have changed over time.

Next, two forecasting models were applied:

1. ARIMA (Auto Regressive Integrated Moving Average) – a widely used statistical model for time series forecasting.
2. Holt-Winters Exponential Smoothing – a method that captures trends and seasonality in the data.

Finally, the accuracy of both models was compared using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to determine which model provided the best predictions.

## 3.3 TOOLS FOR ANALYSIS AND FORECASTING

1. EDA (Exploratory Data Analysis)

2. ARIMA (Auto regressive Integrated Moving Average)

3. Holt-Winter's Exponential Smoothening Technique

## 3.4 TOOLS FOR COMPARISON

The accuracy of the model is evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

**1. Mean Absolute Error (MAE)**

MSE is the average of the squared difference between predicted values and actual values. The formula for MAE is:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

where,

$y_i$ is the $i^{th}$ observed value,

$\hat{y}_i$ is the corresponding predicted value,

n is the number of observations.

**2. Root Mean Squared Error (RMSE)**

RMSE is calculated by taking the root of mean squared error. The formula for RMSE is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where,

$y_i$ is the actual value for the $i^{th}$ observation,

$\hat{y}_i$ is the predicted value for the $i^{th}$ observation,

n is the number of observations.

## 3.5 PYTHON PROGRAMMING LANGUAGE

Python is the most popular programming language which is known for its understandability, adaptability, readability and flexibility. Python is used in many fields such as data science, machine learning, web development.

## 3.6 MICROSOFT EXCEL

Microsoft Excel is a powerful spreadsheet application developed by Microsoft. It is widely used for data organization, analysis, and visualization. It is used to create spreadsheets with rows and columns for calculations, create charts and used to perform many other operations.

# CHAPTER 4

# EXPLORATORY DATA ANALYSIS (EDA)

## 4.1 EXPLORATORY DATA ANALYSIS (EDA)

It is the initial step of data analysis process, it enables us to know the underlying patterns and structure of data. It also enables us to know the relationship between dataset. EDA can be used to know the quality of data, also for checking the null values, missing values etc. After EDA with the check of null and missing values, further step is to summarize and visualize the data in order to know it more. This can involve the computation of summary statistics like mean, median, mode and standard deviation etc. Data visualization, breaking the data into trend, seasonal and residual also are included in this.

## 4.2 DESCRIPTIVE STATISTICS

Descriptive statistics provide description of variability, distribution and central tendency of the data set by summarizing mean, median, mode, standard deviation and quartiles. Standard deviation and quartiles are measures that help in identifying outliers and extreme values of the dataset. Descriptive statistics overall is a comprehensive summary of data's key characteristics, which will assist in understanding data more and conducting further analysis accurately.

## 4.3 TIME SERIES VISUALISATION

Time series visualization is the graphical display of data observed over consecutive time periods. It includes several methods like line plots, seasonal subseries plots, autocorrelation plots, histograms, and interactive visualizations. These techniques assist analysts in detecting trends, patterns, and outliers in time-dependent data for improved comprehension and decision-making.

## 4.4 SEASONAL DECOMPOSITION

When dealing with time series data, one of the primary techniques employed is seasonal decomposition. This technique allows the data to be decomposed into three broad components: trend, seasonality, and residuals. Through the study of these components, we can more effectively understand how stock prices change over time and choose the most appropriate forecasting model.

**Trend:** This is the long-term direction of movement in stock prices—whether they are generally moving up, down, or remaining the same over time.

**Seasonality:** Certain patterns in stock price data recur at fixed intervals, such as daily, monthly, or annual fluctuations. For instance, stock prices may exhibit seasonal trends, rising during specific periods due to market cycles, earnings reports, or economic factors and declining during others.

**Residuals:** These are the random fluctuations in the data that cannot be accounted for by trend or seasonality. They indicate how well our model fits the real data.

Knowledge of these elements is important in selecting an appropriate forecasting model. It enables us to perform the appropriate transformations, enhance the precision of the model, and have more accurate predictions.

Exploratory data analysis (EDA) also has an important role to play in this activity. It enables us to identify missing values, understand trends, and know which forecasting models are appropriate for the dataset.

# CHAPTER 5
# TIME SERIES ANALYSIS

Time Series Analysis is a statistical method for analysing and interpreting data points gathered or recorded over fixed time intervals. It detects patterns, trends, seasonality, and other forms of structure in time-series data to allow for predictions and forecasts of future values. A time series describes the association between two variables, one of which is time. Mathematically, a time series can be represented as the function relation $Y_t = f(t)$, where $Y_t$ denotes the value of the variable being considered at time t. Here time can be yearly, monthly, weekly, daily or even hourly typically in equal intervals of time. An example of hourly temperature reading, daily sales and monthly production constitute time series.

## 5.1 COMPONENTS OF TIME SERIES

There are mainly 4 components in time series they are as follows:

**Secular Trend ($T_t$) :** Trend is basically long time change in time series data. It reflects clear and fundamental direction of statistical data through passage of time. It is smooth, consistent and long run movement. It can be portrayed as increase, reduction or constancy over time for a specific portion in a secular trend, likewise the aggregate trend may reveal increasing trend, declining trend or stable trend. If there is neither increasing nor decreasing then the time series is termed as stationary. For instance, in global average temperature data in a time series, the tendency to increase can be observed. This trend of increase in temperatures is consistently found over a period of time.

**Seasonality ($S_t$) :** Seasonal changes are fluctuations that occur in a consistent pattern during a particular period, usually within a year or less, and tend to recur in a similar fashion year after year. They are caused by factors like climatic conditions, social customs, and religious rituals. Seasonal fluctuations are to be measured to know their effect on the overall trend. For instance, ice cream sales are usually higher in summer and lower in winter. This cycle repeats every year owing to variation in temperature and consumer behaviour according to the season.

**Cyclic variations ($C_t$) :** Cyclic fluctuations are periodic movements that can be upward or downward movements in the time series data. The duration of this cyclic fluctuations is usually more than a year and also these variations are not periodic. Business cycles and cycles operate on Business and Economic series. These cyclic movements go through various stages such as prosperity, recession, depression and recovery. At these various stages, the time series undergo changes. These alterations are referred to as cyclic changes. Series on production of prices, demands etc. experience these cyclic changes.

**Irregular fluctuation ($I_t$) :** Irregular fluctuations are those fluctuations which are resulted due to sudden or accidental occurrences. These fluctuations can be resulted due to environmental disasters such as flood, drought etc. or can be due to sudden strikes, wars etc. these activities can result in sudden change in the time series data. These activities are based on random factors and these fluctuations cannot be forecasted as the other elements of time series.

## 5.2 MATHEMATICAL MODELS FOR TIME SERIES

In classical time series analysis, it is assumed that there exist two models commonly for the decompositions of a time series into its components.

**Additive Model:** According to the additive model the decomposition of time series is done on the assumption that the effects of various components are additive or in other words,

$$Y_t = T_t + S_t + C_t + I_t$$

Where $Y_t$ is the time series value

$T_t$ is the trend

$S_t$ is the seasonal variation

$C_t$ is the cyclical variation

$I_t$ is the irregular variation

In this model $S_t$, $C_t$ and $I_t$ are absolute quantities and can have positive or negative values. The model assumes that four components of the time series are independent of each other and none has any effect on the remaining three components. In actual practice, this hypothesis does not hold good as these factors affect each other.

**Multiplicative Model:** According to the multiplicative model the decomposition of a time series is done on the assumption that the effects of the four components of the time series are not independent of each other. According to the multiplicative model,

$$Y_t = T_t * S_t * C_t * I_t$$

In this model $T_t$, $S_t$, $C_t$ and $I_t$ are not absolute amounts as in the case of the additive model. There are relative variations and are expressed as rates or indices fluctuating above or below unity. The multiplicative model can be expressed in terms of the logarithm.

$$log Y_t = log T_t + log S_t + log C_t + log I_t$$

## 5.3 TIME SERIES MODELLING

### 5.3.1 BASIC DEFINITIONS

### 1. Stationary Time series

Stationary time series refers to series of observations where the mean, variance and the Autocorrelation remains constant over time. It is a time series data which exhibit a stable behaviour without trend and seasonality.

### 2. Non-stationary Time series

Non-stationary time series refers to series of observation where the mean variance and the Autocorrelation varies over time. It is a time series data exhibit unstable behaviour with trend, seasonality and other patterns. Non-stationary data cannot be used for analysis.

### 3. Auto Correlation Function (ACF)

Autocorrelation function in time series is a tool used to measure the correlation between a time series and its lagged value at different time intervals. ACF value of 1 or -1 indicate strong positive or negative autocorrelation. Patterns of ACF give idea about seasonality and other random behaviours. Stationarity can be assessed by ACF, ACF plots with lags dying to zero represents stationarity. The autocorrelation function of a stationary time series $\{Z_t\}$, $\rho(k)$ at lag k is defined as the correlation at lag k between $Z_t$ and $Z_{t+k}$. Thus, the autocorrelation function at lag k is given by,

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Where $\gamma(k) = Cov(Z_{t\,,}\, Z_{t+k})$

## 4. Partial Autocorrelation Function (PACF)

The Partial Autocorrelation Function (PACF) is used to assess the direct link between two observations in a time series while taking additional data effect into consideration. PACF is used to find the MA parameter of ARIMA model. The partial autocorrelation function, like the autocorrelation function, conveys vital information regarding the dependence structure of a stationary process. In the context of time series, a large portion of the correlation between $Z_t$ and $Z_{t+k}$ can be due to the correlation these variables have with $(Z_{t-1,}\, Z_{t-2,}\, \ldots, Z_{t-k+1})$. The partial autocorrelation of lag k can be thought of as the partial regression coefficients kk in the representation.

$$Z_t = \emptyset_{k1} Z_{t-1} + \emptyset_{k2} Z_{t-2} + \cdots + \emptyset_{kk} Z_{t-k+1}$$

Thus, the partial autocorrelation at lag k, $\emptyset_{kk}$, measures the correlation between $Z_t$and $Z_{t-k}$ after adjusting for the effects of $(Z_{t-1}$, $Z_{t-2}$, ..., $Z_{t-k+1})$.

## 5. Augmented Dickey Fuller Test

The ADF test belongs to a category of tests called 'Unit Root Test', which is the proper method for testing the stationarity of a time series. Augmented Dickey-Fuller (ADF) test is a common statistical test used to test whether a given time series is stationary or not. It is one of the most commonly used statistical tests when it comes to analysing the stationary of a series. It tests the following two null and alternate hypotheses:

$H_0$: The time series is regarded as non-stationary.

$H_1$: The time series is regarded as stationary.

Now, if the p-value from this test comes out to be less than a particular level (e.g. $\alpha$ = 0.05) then in such cases the null hypothesis is rejected and concludes that the time series is stationary.

## 6. Stationarity

The ARIMA method is appropriate only for a stationary data series. Stationarity implies that the AR coefficients must satisfy certain conditions for an ARIMA model to be stationary. There is a reason for requiring stationarity: we could not get useful estimates of the parameters of a process otherwise. If p=0, we have either a pure MA model or a white noise series. All pure MA models and white noise are stationary, so there are no stationarity conditions to check. For an AR (1) or ARMA (1, q) process, the stationary requirement is that the absolute value of $\emptyset_1$ must be less than one: $|\emptyset_1| < 1$.

For an AR (2) or ARMA (2, q) process, the stationary requirement is a set of three conditions: $|\emptyset_2| < 1$, and $\emptyset_1 - \emptyset_2 < 1$.

## 7. Invertibility

There is another condition that ARIMA models must satisfy called invertibility. This requirement implies that the MA coefficients must satisfy certain conditions. There is a commonsense reason for invertibility: a non-invertible ARIMA model implies that the weights placed on past Z observations do not decline as we move further into the past, but commonsense says that larger weights should be attached to more recent observations. Invertibility ensures that these results hold.

If q=0, we have either a pure AR process or a white noise series. All pure AR processes and white noise are invertible and no further checks are required. For an MA (1) or ARMA (p, 1) process, invertibility requires that the absolute value of $\theta_1$ must be less than one, $|\theta_1| < 1$, $\theta_1 + \theta_2 < 1$, and $\theta_1 - \theta_2 < 1$.

## 8. Auto Regressive (AR) Process

A time series $\{Zt\}$ is said to be an autoregressive process of order p, abbreviated as AR(p) if it is a weighted linear sum of the past p values plus a random shock so that

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t$$

where $\{\varepsilon t\}$ denotes a purely random process with 0 mean and constant variance $\sigma^2$ . Using the backward shift operator B, such that $BZ_t = Z_{t-1}$ , the AR (p) model may be written more succinctly in the form,

$$\emptyset(B)Z_t = \varepsilon_t$$

where $\emptyset(B)$ = 1 - $\emptyset_1 B_1$ - $\emptyset_2 B_2$ -...- $\emptyset_p B_p$ is a polynomial in B of order p.

## 9. Moving Average (MA) Process

A time series $\{Zt\}$ is said to be a moving average process of order q, abbreviated as MA(q) if it is a weighted linear sum of the last q random shocks so that

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon t\}$ denotes a purely random process with 0 mean and constant variance $\sigma^2$ . Using the backward shift operator B the MA (q) model may be written in the form,

$$Z_t = \theta(B)\varepsilon_t$$

where $\theta(B)$= $1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is a polynomial in B of order q.

## 10. Auto Regressive Moving Average Process (ARMA)

Mixed autoregressive moving average model with p auto regressive terms and q moving average terms is abbreviated as ARMA (p, q) is given by

$$Z_t - \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Using the backward shift operator B, the ARMA (p, q) can be written in the form,

$$Z_t - \emptyset_1 B Z_t - \emptyset_2 B^2 Z_t - \cdots - \emptyset_p B^p Z_t = \varepsilon_t - \theta_1 B \varepsilon_t - \theta_2 B^2 \varepsilon_t - \cdots - \theta_q B^q \varepsilon_t$$

$$\emptyset(B) Z_t = \theta(B)\varepsilon_t$$

where $\emptyset(B)$ and $\theta(B)$ are polynomial in B of degree of order p and q respectively

## 11. Akaike Information Criteria (AIC)

The Akaike Information Criterion (AIC) is a statistical metric used to evaluate and compare models in time series forecasting. It helps identify the best-fitting model by balancing goodness of fit and model complexity.

The AIC is calculated using the formula:

$$\text{AIC}=2k-2\ln(L)$$

where k represents the number of parameters in the model, and L is the maximum likelihood of the model. A lower AIC value indicates a better model, as it suggests a good trade-off between accuracy and simplicity.

## 12. Diagnostic Checking

In diagnostic checking, we have to assess model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that $\{\varepsilon_t\}$ is white noise. Hence, model diagnostic checking is accomplished through careful analysis of residual series $\{\varepsilon_t\}$. To check whether the errors are normally distributed, one can construct a histogram of standardized residuals and compare it with the standard normal distribution. The assumption random shocks have a mean of zero and constant variance can be checked using a residuals plot. To check whether the residuals are approximately white noise, we compute the sample ACF of the residuals to see whether they are all statistically insignificant. The iterative nature of the three-stage UBJ (Univariate Box-Jenkins models) modelling procedure is important. The estimation and diagnostic-checking stages provide warning signals telling us when, and how, a model should be reformulated. It is continued to re-identify, re-estimate, and re-check until a model is got that is satisfactory according to several criteria.

## 13. Q-Q Plot

The quantile-quantile (q-q plot) plot is a graphical method for determining if a dataset follows a certain probability distribution or whether two samples of data came from the same population or not. Q-Q plots are particularly useful for assessing whether a dataset is normally distributed or if it follows some other known distribution

## 14. Correlogram

A correlogram is a visual representation of the autocorrelation function (ACF) for a time series, plotted against lag values. It shows the degree of correlation between observations at different time lags. It helps identify patterns, autocorrelation, seasonality, and whether residuals behave like white noise (no significant autocorrelations).

## 15. Histogram

A histogram represents the frequency distribution of residuals, KDE (Kernel Density Estimate) provides a smoothed estimate of the density, and the standard normal curve overlays to compare the residual distribution against a normal distribution.

## 16. Ljung-Box Test

The Ljung-Box test is a statistical test used to determine whether a time series exhibits significant autocorrelation (i.e., whether past values influence future values). It helps assess whether the residuals (errors) of a model are independently distributed (white noise) or if there is remaining structure that the model has not captured. The test outputs a p-value, which is used to draw conclusions: If p-value > 0.05, then the residuals are independent (no significant autocorrelation), the model adequately captures the time series structure and no significant patterns are left in the residuals.

## 5.4 HOLT WINTERS EXPONENTIAL SMOOTHING TECHNIQUE

Holt's method can be extended to deal with time series which contain both trend and seasonal variations. The Holt-Winters method has two versions, additive and multiplicative, the use of which depends on the characteristics of the particular time series. Let $L_t$ , $T_t$ , $I_t$ denote the local level, trend and seasonal index, respectively at time t.

The interpretation of it depends on whether seasonality is thought to be additive or multiplicative. In the additive case I, $y_t - I_t$ is the deseasonalized value, while in the multiplicative class it is $y_t/I_t$ . The values of the 3 quantities $L_t$ , $T_t$ , $I_t$ all need to be estimated and so we need 3 updating equations with three smoothing parameters, say ∝, γ and δ. As before the smoothing parameters are usually chosen in the range (0, 1). The form of the updating equations is again intuitively plausible.

Suppose the seasonal variation is multiplicative. Then the (recurrence form) equations for updating $L_t$ , $T_t$ , $I_t$ when a new observation $yt$ becomes available are

$$L_t = \alpha\,(y_t / I_{t-s}) + (1-\alpha)\,(L_{t-1} + T_{t-1})$$

$$T_t = \gamma\,(L_t - L_{t-1}) + (1-\gamma)\,T_{t-1}$$

$$I_t = \delta\,(y_t / L_t) + (1-\delta)\,I_{t-s}$$

and the forecasts from time (t) are then,

$$\hat{y}_{t+h} = (L_t + hT_t)I_{t-s+h}$$

for h = 1, 2, …, s

where s denotes the seasonal period (For example: s = 4 for quarterly data and 12 for monthly data). If the seasonal variation is additive, the equations for updating $L_t$ , $T_t$ , $I_t$ when a new observation $yt$ becomes available are

$$L_t = \alpha\ (y_t / I_{t-s}) + (1- \alpha)\ (L_{t-1} + T_{t-1})$$

$$T_t = \gamma\ (L_t - L_{t-1}) + (1- \gamma)\ T_{t-1}$$

$$I_t = \delta\ (y_t / L_t) + (1- \delta)\ I_{t-s}$$

and the forecasts from time (t) are then

$$\hat{y}_{t+h} = (L_t + hT_t)I_{t-s+h}$$

for h = 1, 2, ……., s

For starting values, it seems sensible to set the level component $L_O$ , equal to the average observation in the first year that is,

$$L_O = \sum_{t=1}^{s} y_t$$

where s is the number of seasons. The starting values for the slope component can be taken from the average difference per period between the first and second-year averages. That is,

$$T_0 = \frac{\frac{\sum_{t=s+1}^{2s} y_t}{s} - \frac{\sum_{t=1}^{s} y_t}{s}}{s}$$

Finally, the seasonal index starting value can be calculated after allowing for a trend adjustment, as follows:

$$I_0 = (y_k - (k-1)T_0)/2 \quad (multiplicative)$$

$$I_o = (y_k - (L_O + (k-1)T_0)/2) \quad (additive)$$

Where k = 1, 2 ……., s. This will lead to (s) separate values for $I_o$, which is what is required to gain the initial seasonal pattern.

# CHAPTER 6

# RESULT AND ANALYSIS

This chapter provide a detailed examination and interpretation of the results obtained from the study of Ford Motor's stock price forecasting using time series analysis.

## 6.1 ARIMA MODEL

### 6.1.1 TIME SERIES PLOT

The initial step in time series is to draw a time series plot. The fig 6.1 shows the time series plot of average annual stock price of Ford company from 1972 to 2024.

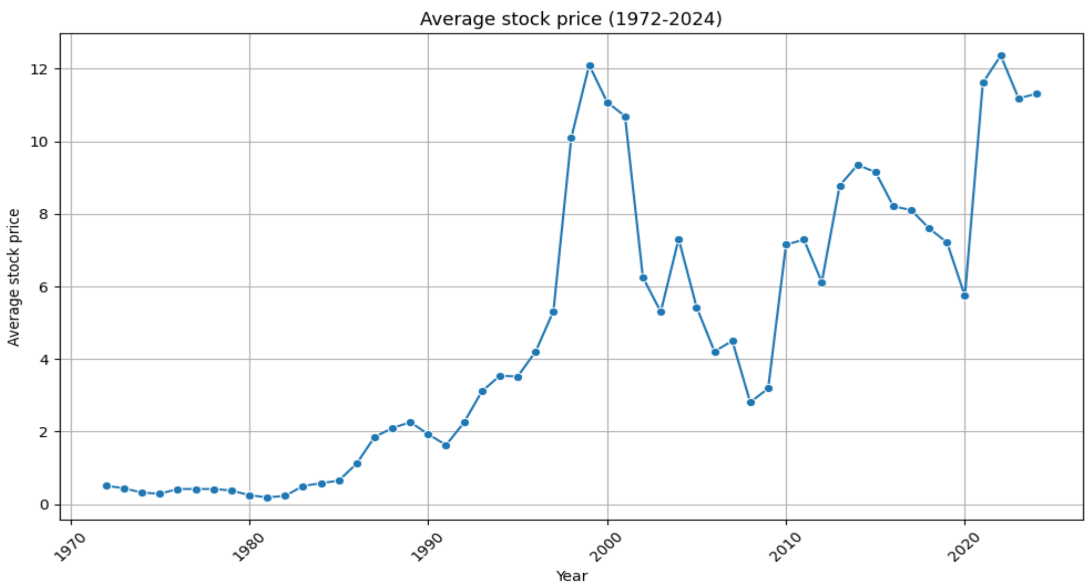


Fig 6.1

### 6.1.2 SEASONAL DECOMPOSITION OF DATA

The next step involves performing seasonal decomposition to extract the trend, seasonal, and random components from the time series data. The fig 6.2 shown below illustrates this decomposition.

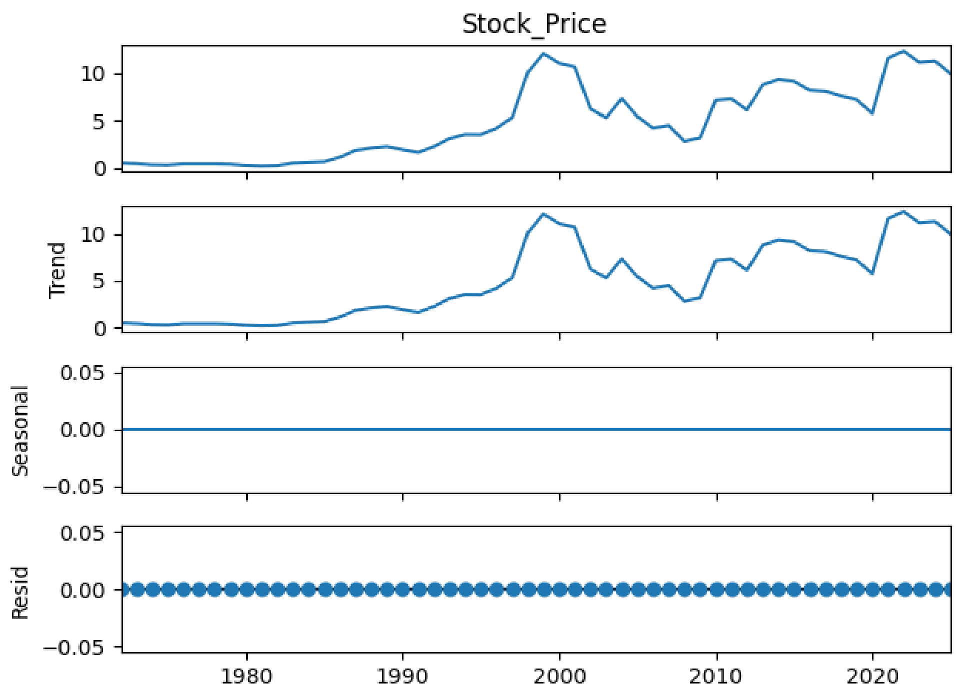


Fig 6.2

### 6.1.3 STATIONARITY CHECK USING AUGMENTED DICKEY-FULLER TEST

To test the stationarity of the time series data using the ADF test, we follow a hypothesis testing approach. The null and alternative hypotheses are stated as follows:

$H_o$: The data is non-stationary.

$H_1$: The data is stationary.

**RESULT**

ADF Statistic: -1.176211

p-value: 0.683744

Critical Values:

1%: -3.562878534649522

5%: -2.918973284023669

10%: -2.597393446745562

The p-value is greater than 0.05, so we fail to reject the null hypothesis. The time series is non-stationary.

Therefore, we apply differencing of order n until the time series becomes stationary in both cases. For this analysis, we perform differencing with n=1. Afterward, we assess the stationarity of the data again using the ADF test. The hypothesis tested is:

- $H_o$: The data is non-stationary.
- $H_1$: The data is stationary.

**RESULT**

ADF Statistic: -5.825795

p-value: 0.000000

Critical Values:

1%: -3.568485864

5%: -2.92135992

10%: -2.5986616

The p-value is less than 0.05, so we reject the null hypothesis. The differenced series is stationary. Fig 6.3 shows the time series plot after differencing.


Fig 6.3

### 6.1.4 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTION

Next step in Time Series Analysis is to plot and examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Fig 4.4 shows the ACF &PACF Plot.

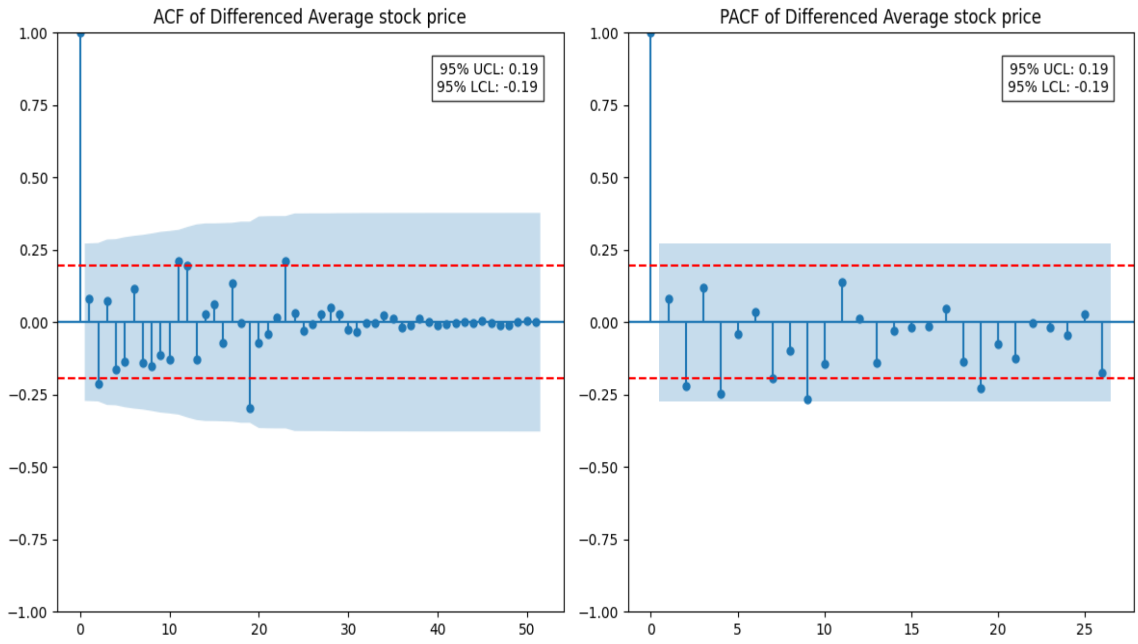


Fig 6.4

### 6.1.5 ARIMA MODELS AND CORRESPONDING AIC VALUES

In this step we choose the best model for forecasting the values. It is done by choosing one model from all possible models according to Akaike Information Criterion (AIC). The model with lowest AIC value is chosen as the best model. Table 4.1 below show the possible values of ARIMA model with their AIC value.

| SL NO | Model ARIMA (p, d, q) x (P, D, Q) | AIC value |
|---|---|---|
| 1 | ARIMA (0, 1, 0) x (0,0,0) | 194.96015275911608 |
| 2 | ARIMA (0, 1, 1) x (0,0,0) | 193.40346060533975 |
| 3 | ARIMA (0, 1, 2) x (0,0,0) | 189.27547904837502 |
| 4 | ARIMA (0, 1, 3) x (0,0,0) | 188.4813278395898 |
| 5 | ARIMA (1, 1, 0) x (0,0,0) | 196.5006232248018 |
| 6 | ARIMA (1, 1, 1) x (0,0,0) | 193.02251141360537 |
| 7 | ARIMA (1, 1, 2) x (0,0,0) | 191.22119719866836 |
| 8 | ARIMA (1, 1, 3) x (0,0,0) | 190.46903561225844 |
| 9 | ARIMA (2, 1, 2) x (0,0,0) | 193.93719955058677 |
| 10 | ARIMA (2, 1, 1) x (0,0,0) | 194.10190744046756 |
| 11 | ARIMA (2, 1, 2) x (0,0,0) | 189.0284981342809 |
| **12** | **ARIMA (2, 1, 3) x (0,0,0)** | **186.9793300581165** |
| 13 | ARIMA (3, 1, 0) x (0,0,0) | 192.35360320076123 |
| 14 | ARIMA (3, 1, 1) x (0,0,0) | 192.89025133633007 |
| 15 | ARIMA (3, 1, 2) x (0,0,0) | 190.72296826137227 |
| 16 | ARIMA (3, 1, 3) x (0,0,0) | 189.9283827654314 |

Table 6.1

**Best ARIMA order: (2, 1, 3) x (0,0,0)**

**Best AIC: 186.9793300581165**

| | ar. L1 | ar. L2 | ma. L1 | ma. L2 | ma. L3 |
|---|---|---|---|---|---|
| Coefficients | -1.1515 | -0.9169 | 1.3588 | 1.3588 | 1.9796 |
| Std error | 0.160 | 0.109 | 11.771 | 17.244 | 33.237 |

Table 6.2

## 6.1.6 DIAGNOSTIC CHECKING

Diagnostic checking is a crucial step in confirming the validity, effectiveness, and reliability of statistical models. The main objective of this process is to identify and select the most appropriate and optimal model for the data.

### 6.1.6.1 Ljung-Box Test

The Ljung-Box test is employed to check for the presence of autocorrelation in the residuals of the model. The hypotheses for the test are as follows:

- Null Hypothesis ($H_0$): The residuals of the model are independently distributed (i.e., there is no significant autocorrelation remaining in the residuals).
- Alternative Hypothesis ($H_1$): The residuals of the model are not independently distributed (i.e., significant autocorrelation exists in the residuals).

The results of the Ljung-Box test are used to evaluate whether the residuals exhibit any remaining autocorrelation, helping to assess the adequacy of the model.

**Ljung test result**

Test statistic: 6.207767

P value: 0.797516

Since p=0.797516 is significantly greater than 0.05, the residuals appear to be uncorrelated. This suggests that the fitted ARIMA model effectively captures the autocorrelation structure of the data, and the residuals resemble white noise. Therefore, the model is considered a god fit. The diagnostic plot is shown below.

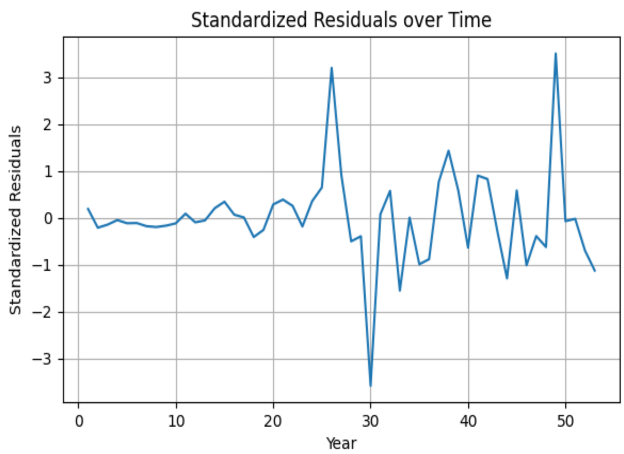


Fig 6.5 (i)

From the fig 6.5 (i) the standardized residuals plot is not showing any pattern, just a random fluctuation around zero, indicating that the homoscedasticity assumption is fair and outliers were well-managed.

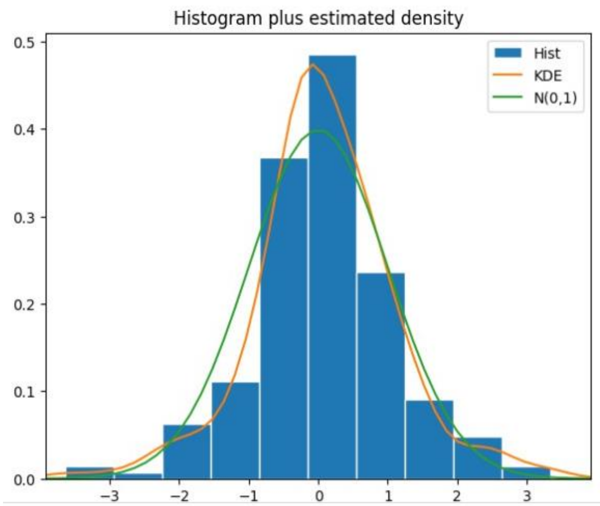


Fig 6.5 (ii)

Figure 6.5 (ii) is the residuals' bell-shaped distribution histogram, which was centred around zero, indicated that the residuals' normality condition was satisfied.

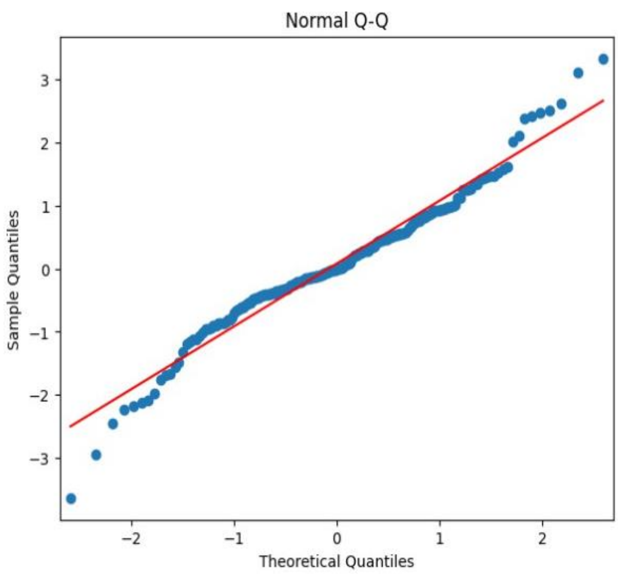


Fig 6.5 (iii)

Figure 6.5 (iii) shows that the residuals' strong adherence to a straight line, as seen by the Q-Q plot, suggests that they are roughly regularly distributed.

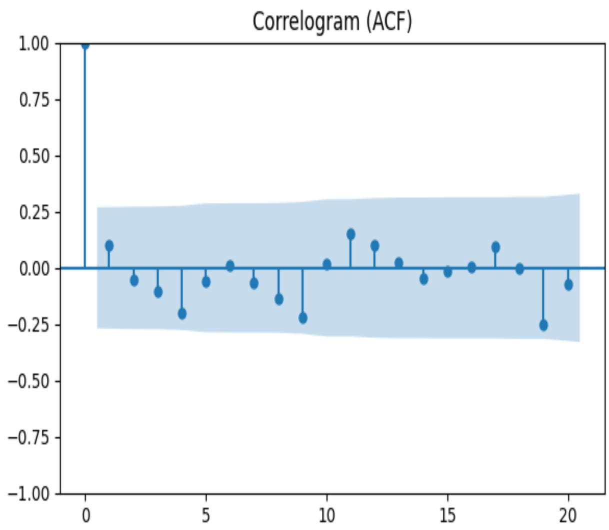


Fig 6.5 (iv)

The figure 6.5 (iv) is the correlogram which shows random fluctuations around zero in the autocorrelation function, which supports the assumption of independence of residuals.

## 6.1.7 FORECASTING THE SAMPLE

Forecasting the sample refers to predicting the actual data points or the training data points used in building the model. This allows us to evaluate the model's performance on the training dataset, providing insights into how well the model fits the historical data.

The table 6.3 presents the actual values alongside the in-sample forecasted values, enabling us to compare the model's predictions against the observed data.

| YEAR | ACTUAL PRICE | PREDICTED PRICE |
|---|---|---|
| 2004 | 7.31 | 6.564296 |
| 2005 | 5.43 | 7.263527 |
| 2006 | 4.21 | 5.707786 |
| 2007 | 4.5 | 5.137264 |
| 2008 | 2.81 | 3.946199 |
| 2009 | 3.19 | 2.771276 |

| | | |
|---|---|---|
| 2010 | 7.15 | 4.133911 |
| 2011 | 7.29 | 7.383057 |
| 2012 | 6.12 | 6.819375 |
| 2013 | 8.78 | 6.456656 |
| 2014 | 9.35 | 8.417856 |
| 2015 | 9.15 | 8.485278 |
| 2016 | 8.21 | 9.782129 |
| 2017 | 8.1 | 7.395219 |
| 2018 | 7.6 | 8.127566 |
| 2019 | 7.21 | 7.431914 |
| 2020 | 5.73 | 7.536698 |
| 2021 | 11.61 | 5.521650 |
| 2022 | 12.36 | 13.159350 |
| 2023 | 11.18 | 10.766022 |
| 2024 | 11.31 | 12.243780 |

Table 6.3

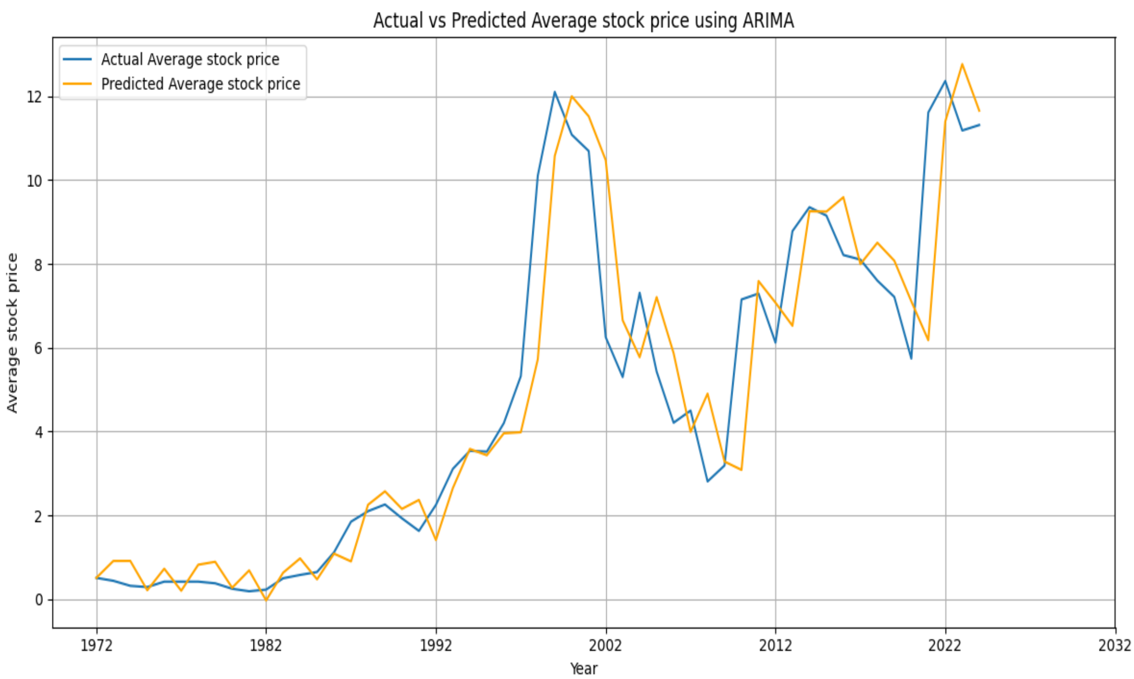


Fig 6.6

### 6.1.8 FORECASTED FUTURE VALUES

Table 6.4 is the forecasted value of the average annual stock price from 2025 to 2028 which also contains the LCL and UCL that is the upper control limit and lower control limit.

| YEAR | FORECASTED PRICE | LOWER AVERAGE STOCK PRICE | UPPER AVERAGE STOCK PRICE |
|---|---|---|---|
| 2025 | 10.023830 | 6.991113 | 13.056546 |
| 2026 | 9.319722 | 4.623258 | 14.016187 |
| 2027 | 9.916283 | 4.413936 | 15.418629 |
| 2028 | 9.948518 | 3.621282 | 16.275678 |

Table 6.4

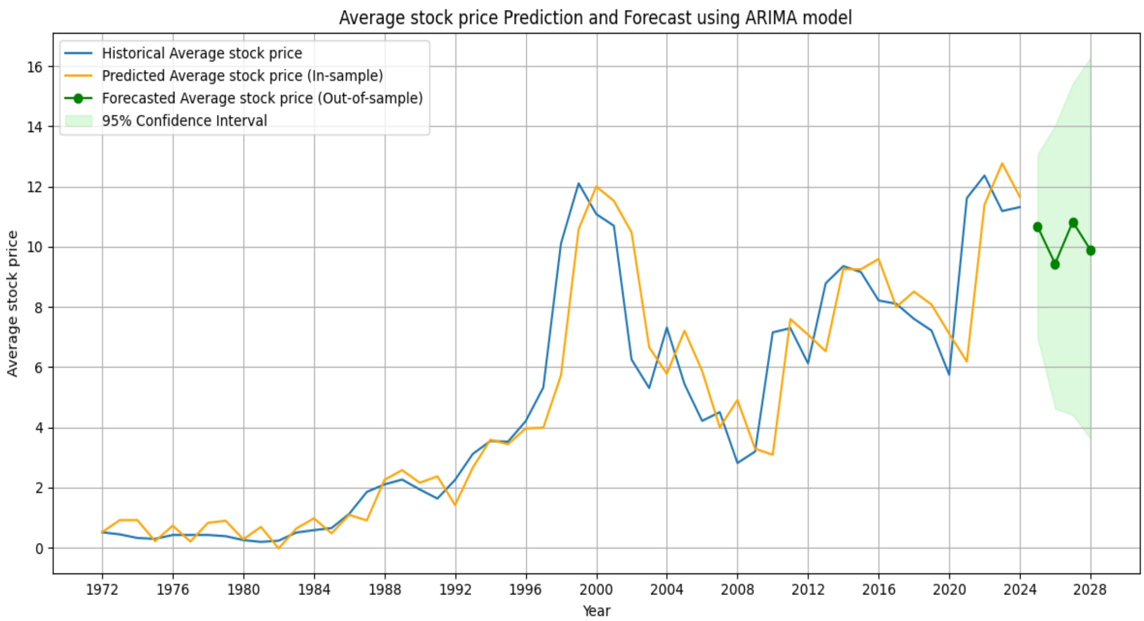


Fig 6.7

## 6.2 HOLT-WINTER EXPONENTIAL SMOOTHING TECHNIQUE

### 6.2.1 FORECASTING THE SAMPLE

The table 6.5 presents the actual values alongside the in-sample forecasted values, enabling us to compare the model's predictions against the observed data.

| YEAR | ACTUAL PRICE | PREDICTED PRICE |
|---|---|---|
| 2004 | 7.31 | 5.773626 |
| 2005 | 5.43 | 7.205005 |
| 2006 | 4.21 | 5.867990 |
| 2007 | 4.5 | 3.992998 |
| 2008 | 2.81 | 4.902717 |
| 2009 | 3.19 | 3.283626 |
| 2010 | 7.15 | 3.085005 |
| 2011 | 7.29 | 7.587990 |
| 2012 | 6.12 | 7.072998 |
| 2013 | 8.78 | 6.522717 |
| 2014 | 9.35 | 9.253626 |
| 2015 | 9.15 | 9.245005 |
| 2016 | 8.21 | 9.587990 |
| 2017 | 8.1 | 7.992998 |
| 2018 | 7.6 | 8.502717 |
| 2019 | 7.21 | 8.073626 |
| 2020 | 5.73 | 7.105005 |
| 2021 | 11.61 | 6.177990 |
| 2022 | 12.36 | 11.392998 |
| 2023 | 11.18 | 12.762717 |
| 2024 | 11.31 | 11.653626 |

Table 6.5

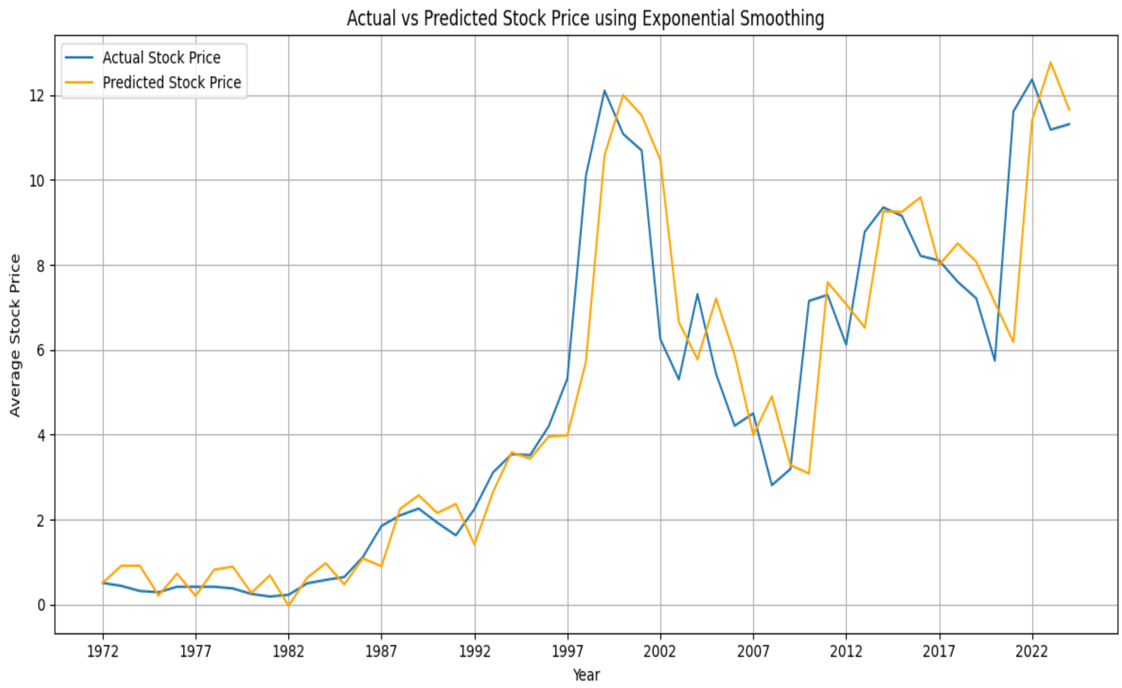


Fig 6.8

## 6.2.2 FORECASTED FUTURE VALUES

Table 6.6 is the forecasted value of the average annual stock price from 2025 to 2028 which also contains the LCL and UCL that is the upper control limit and lower control limit.

| YEAR | FORECASTED PRICE | LOWER AVERAGE STOCK PRICE | UPPER AVERAGE STOCK PRICE |
|---|---|---|---|
| 2025 | 10.672342 | 7.690565 | 13.654119 |
| 2026 | 9.432178 | 6.450401 | 12.413955 |
| 2027 | 10.823614 | 7.841837 | 13.805391 |
| 2028 | 9.892735 | 6.910958 | 12.874512 |

Table 6.6

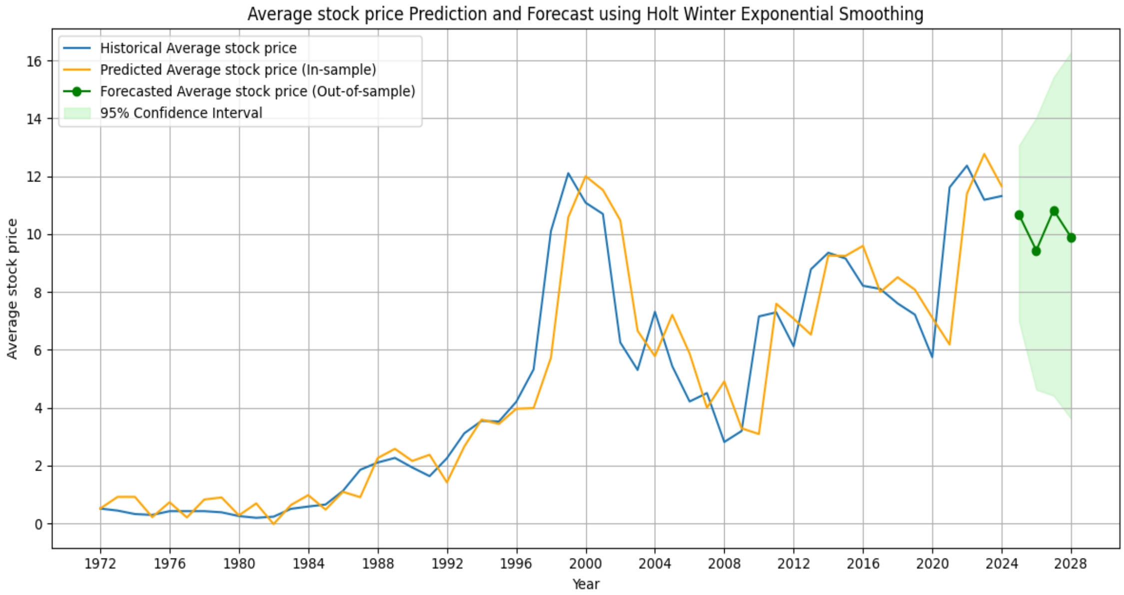


Fig 6.9

## 6.3 COMPARISON OF ARIMA AND HOLT-WINTER EXPONENTIAL SMOOTHING MODELS

Table 6.7 is the comparison of the MSE and RMSE values of two models that is the ARIMA model and Holt-Winter Exponential Smoothing technique.

| MODEL | MSE | RMSE |
|---|---|---|
| ARIMA MODEL | 0.354274 | 0.595209 |
| HOLT-WINTER EXPONENTIAL SMOOTHING | 2.314398 | 1.521315 |

Table 6.7

# CHAPTER 7
# CONCLUSION

This study analysed the historical stock price data of Ford Motor Company using time series forecasting models, specifically ARIMA and Exponential Smoothing. The primary objective was to identify trends, patterns, and predict future stock prices based on historical data. By evaluating multiple ARIMA models, the best-performing model was selected based on the Akaike Information Criterion (AIC). The ARIMA (2,1,3) model was identified as the most suitable, having the lowest AIC value of 186.9793. To ensure the reliability of the model, diagnostic tests such as the Ljung-Box test and residual analysis were conducted. The results confirmed that the residuals were independently distributed, resembling white noise, indicating a good model fit.

Exponential Smoothing was also employed as an alternative forecasting technique. To compare the effectiveness of both models, performance was evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The results showed that the ARIMA model achieved an MSE of 0.354274 and RMSE of 0.595209, whereas the Exponential Smoothing model had an MSE of 2.314398 and RMSE of 1.521315. These findings suggest that ARIMA provided slightly better predictive accuracy than Exponential Smoothing.

The forecasted stock prices for 2025 to 2028 were generated along with confidence intervals, ensuring an understanding of possible fluctuations in stock price trends. While both models demonstrated reasonable accuracy, the ARIMA model was preferred due to its better performance. However, Exponential Smoothing remains a viable alternative, especially for capturing short-term trends.

In conclusion, the ARIMA model is a more suitable choice for forecasting Ford's stock prices due to its superior performance in capturing patterns and minimizing prediction errors. Future research could explore more advanced machine learning models or hybrid approaches to further enhance forecasting accuracy.

# REFERENCES


1. Acharya, V. V., Schaefer, S., & Zhang, Y. (2015). Liquidity risk and correlation risk: A clinical study of the General Motors and Ford Downgrade of May 2005. *The Quarterly Journal of Finance*, *5*(02), 1550006.
2. Agustin, I. N. (2019). The integration of fundamental and technical analysis in predicting the stock price. *Jurnal Manajemen Maranatha*, *18*(2), 93-102.
3. Dash, M., & Sundarka, S. S. (2015). Testing the stationarity of beta for automotive and auto-ancillary sector stocks in Indian stock market. *Journal of Applied Management and Investments*, *4*(2), 76-81.
4. Drakopoulou, V. (2016). A review of fundamental and technical stock analysis techniques. *Journal of Stock & Forex Trading*, *5*.
5. Edwards, R. D., Magee, J., & Bassetti, W. C. (2018). *Technical analysis of stock trends*. CRC press.
6. Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, *74*(366a), 291-298.
7. Faris, M. (2023). Analysis of Fundamental and Technical Financial Ford Motor Company with The Arrangements of Implication Black Volatility. *Available at SSRN 4482880*.
8. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007-3057.
9. Vychytilová, J., Pavelková, D., Pham, H., & Urbánek, T. (2019). Macroeconomic factors explaining stock volatility: multi-country empirical evidence from the auto industry. *Economic research-Ekonomska istraživanja*, *32*(1), 3327-3341.
10. Wang, Y. C., Yu, J., & Wen, S. Y. (2014). Does fundamental and technical analysis reduce investment risk for growth stock? An analysis of Taiwan stock market. *International Business Research*, *7*(11), 24.