BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

THE ROLE OF INFLATION IN PREDICTING STOCK MARKET RETURNS IN ZIMBABWE: A MACHINE LEARNING APPROACH (2018-2022)

BY

TAFADZWA CHIRUME

B211952B

TO BE SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR BSC.HONOURS IN STATISTICS AND FINANCIAL MATHEMATICS

PROJECT

2025

AUTHORSHOP DECLARATION STATEMENT

Title of the Thesis: The role of inflation in predicting stock market returns in Zimbabwe: a machine learning approach.

Author: Tafadzwa Chirume

Program: Bachelor of Science Honors Degree in Statistics and Financial Mathematics.

I, the undersigned author of the above-mentioned thesis, hereby declare that:

1. This thesis is my original work and has been prepared by me in accordance with the institution's requirements.
2. All sources, data, and references used in this thesis have been acknowledged and cited appropriately.
3. This thesis has not been submitted elsewhere for any degree or diploma.
4. I have obtained all necessary permissions for the inclusion of third-party content where applicable.

I affirm that this declaration is made with full integrity and in compliance with the institution's policies and academic practice.

Author's Signature: ___________________

Date: _____16/06/2025______

APPROVAL FORM

This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.

Tafadzwa Chirume ____________________ ____16/06/2025_____

B211952B Signature Date

Certified by

Ms. P. HLUPO .... ____16/06/2025_____

Supervisor Signature Date

Dr Magodora ______________________ ______16/06/2025_______

Chairperson Signature Date

DEDICATION

This research is dedicated to my mothers, Memory Dube and Eris Rwanga. Their shared qualities of unwavering support, boundless love, and constant encouragement have been my guiding lights throughout this journey. Thank you for always believing in me, even when I doubted myself.

ACKNOWLEDGEMENTS

I am deeply grateful for the support and guidance that made this research possible. I extend my sincere appreciation to my dedicated supervisor, Ms. P. Hlupo, whose expertise, insightful feedback, and unwavering patience shaped the direction and rigor of this study. I am also indebted to my lecturers, esteemed professionals, and experts, who imparted invaluable knowledge and skills, inspiring me to pursue this research with diligence. I acknowledge the Department of Statistics and Mathematics at Bindura University of Science Education for providing a conducive learning environment and necessary resources. My sincere thanks also go to my friend, W. Dekeshe, whose unwavering support, encouragement, and camaraderie were invaluable throughout this process, serving as a constant source of strength and motivation. Heartfelt gratitude is also extended to my family, whose love, understanding, and sacrifices have been the bedrock of my success. Finally, I offer my deepest gratitude to God for the gift of life, for the opportunities afforded to me, for the strength and perseverance to overcome challenges along the way.

ABSTRACT

This study explores the impact of inflation on stock market returns in Zimbabwe during the economically challenging period from 2018 to 2022. It aims to investigate the relationship between inflation and stock returns and to compare the forecasting accuracy of Random Forest and XGBoost machine learning models. Analysing monthly macroeconomic data, both models were trained and evaluated for their ability to capture complex, non-linear relationships between inflation and market performance. The findings reveal a strong correlation between inflation and stock returns, indicating that inflation significantly influences market dynamics. The XGBoost slightly outperformed the Random Forest model, achieving a lower Mean Squared Error (0.00093), a lower Root Mean Squared Error (0.03042), and a higher R-squared value (0.97), which explains 97% of the variance in stock returns. The study suggests that Zimbabwean policymakers should monitor inflation trends to stabilize financial markets, and investors should consider inflation when making decisions, given its significant impact on market returns. It also recommends using advanced machine learning models, particularly XGBoost and Random Forest models, for future economic forecasting and investment strategies in volatile environments to enhance predictive accuracy and risk management. Overall, the research demonstrates the effectiveness of sophisticated machine learning techniques in modelling complex economic relationships, providing policymakers and investors in Zimbabwe's evolving market with valuable insights.

TABLE OF CONTENTS

Contents

AUTHORSHIP DECLARATION STATEMENT ii

APPROVAL FORM iii

DEDICATION iv

ACKNOWLEDGEMENTS v

ABSTRACT vi

Table of content vii

List of Tables xii

List of Figures xiii

ACRONYMS xiv

CHAPTER 1: INTRODUCTION 1

1.0 Introduction 1

1.1 Background of the Study 1

1.2 Statement of the Problem 3

1.3 Research Objectives 3

1.4 Research Questions and or Hypothesis 4

1.5 Significance of the Study 4

1.6 Delimitations of the Study 4

1.7 Assumptions of the study 5

1.8 Limitations of the Study 5

1.9 Definition of Terms 6

1.10 Chapter Summary 7

CHAPTER 2: LITERATURE REVIEW 8

2.0 Introduction 8

2.1 Theoretical Review....8
2.1.1 The Fisher Hypothesis....8
2.1.2 The Monetorist Theories....9
2.1.3 The Keynesian Theory....9
2.1.4 The Inflation Illusion Hypothesis....10
2.1.5 The Proxy Hypothesis....10
2.1.6 Thee Efficient Market Hypothesis....10
2.1.7 Dividend Discount Model....11
2.1.8 Random Forest Models....11
2.1.9 XGBoost Models....12
2.2 Emperical Literature Review....12
2.3 Research Gaps....13
2.4 Proposed Conceptual Framework....14
2.5 Chapter Summary....15
CHAPTER 3: RESEACH METHODOLOY....16
3.0 Introduction....16
3.1 Research Paradigm and Research design....16
3.2 Population and Sampling....16
3.3 Data sources....17
3.3.1 Data Processing....17
3.3.2 Data Splitting....18
3.4 Description of Variables and Priori expectations....18
3.5 Analytical Model Specification and Justification....19
3.5.1 Random Forest....19
3.5.2 Model Specification....19
3.5.3 XGBoost Model....20

3.6 Model Diagnostic Tests 21
3.6.1 Residual Analysis test 21
3.6.2 Multicollinearity Tests 22
3.6.3 Model Fit Statistics 22
3.6.4 Model Validation and Reliability Test 22
3.6.5 Cross-Validation 22
3.6.6 Holdout Method 23
3.6.7 Performance Metrics 23
3.6.8 Reliability Testing 23
3.6.9 Residual Analysis 23
3.7 Ethical Consideration 24
3.8 Chapter Summary 24
CHAPTER 4: DATA ANALYSIS AND PRESENTATION OF FINDINGS 25
4.0 Introduction 25
4.1.0 Descriptive Statistics 25
4.2.0 Time Series of Economic Indicators 26
4.3.0 Diagnostic test/Pre-test 27
4.3.1 Checking Missing Values 27
4.3.2 Normality Test(Shapiro-Wilk Test) 27
4.3.3 Data transformation 27
4.3.4 Correlation 28
4.3.5 Multicollinearity check 29
4.3.6 Addressing Multicollinearity in Machine Learning Models 29
4.3.7 Feature Engineering 29
4.3.8 Data Splitting 30
4.3.9 Normality Check 30

4.4 Analytical Models....31
4.4.1 Random Forest development....32
4.6.3 Partial Dependences of ZSE on Inflation....32
4.4.3 Decision Tree....32
4.4.4 Predicting using Random Forest Model....33
4.4.5 Feature Importance....34
4.4.6 Random Forest Performance Metrics....35
4.4.7 Residual Analysis test....35
4.4.8 Model Fitness....36
4.5.0 XGBoost Model....37
4.5.1 Partial dependences of ZSE on Inflation....37
4.5.2 Predicting Using XGBoost....37
4.5.3 Feature Importance....39
4.5.4 XGBoost Performance Metrics....39
4.5.5 Residual Plot....39
4.6 Comparison of the Models' Performance....40
4.6.1 Discussion Of Findings....41
4.7 Chapter summary....43
CHAPTER 5: FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS....44
5.0 Introduction....44
5.1 Summary of Findings....44
5.2 Conclusions....44
5.3 Recommendations....45
5.3.1 For Policymakers....45
5.3.2 For Investors....45
5.3.3 For Model Developers....45

5.4 Areas for further research 45

5.5 Conclusion 46

REFERENCES 47

APPENDICES (A-C)..**5**2

Appendix A: Transformed Data 52

Appendix B: Table 4.6: Feature Engineering 54

Appendex C: Synthax of R Codes used to run the models 56

C1: Introduction 56

C2: R Codes..56

LIST OF TABLES

Table 4.1 Descriptive statistics 25
Table 4.2 Normality test on Variables (Shapiro-Wilk Test) 27
Table 4.3 Random Forest Predicted vs Actual values 33
Table 4.4 Random Forest Performance Metrics 35
Table 4.5 XGBoost Actual and Predicted values 38
Table 4.6 XGBoost Performance metrics 39
Table 4.7 XGBoost and Random Forest Comparisons 40

LIST OF FIGURES

Figure 1.1 Line graph of inflation (%), real interest rates(%),GDP(%) and ZSE All Share (%)(2018-2022)....3
Figure 2.1 Proposed conceptual model....14
Figure 4.1 Time series of economic indicators....26
Figure 4.2 Correlation Matrix....28
Figure 4.3 Histograms for ZSE and INF....31
Figure 4.4 Partial dependences of ZSE on Inflation....32
Figure 4.5 Random Forest decision tree (base tree)....33
Figure 4.6 Random Forest Prediction vs Actual over time plot....34
Figure 4.7 Random Forest Feature importance....35
Figure 4.8 Residual plot for constant variance....36
Figure 4.9 XGBoost Partial dependences of ZSE on Inflation....37
Figure 4.10 XGBoost Actual vs Predicted plot....38
Figure 4.11 XGBoost Model Feature importance....39
Figure 4.12 XGBoost Residual plot....40
Figure 4.13 Random Forest and XGBoost comparison plot....41

LIST OF ACRONYMS

1. XGBoost – Extreme Gradient Boosting
2. Rf – Random Forest
3. ZSE – Zimbabwe All Share (%)
4. INF – Inflation (%)
5. INT – Real Interest Rate (%)
6. GDP – Gross Domestic Product (%)
7. RMSE – Root Mean Squared Error
8. MAE – Mean Absolute Error
9. MAPE – Mean Absolute Percentage Error
10. LSTM – Long Short Memory Artificial Neural Network
11. ARIMA – Autoregressive Moving Average
12. AIC – Akaike Information Criteria
13. BIC – Bayesian Information Criteria

CHAPTER 1: INTRODUCTION

1.0 Introduction

The nexus of inflation and stock returns is among the most researched topics in financial economics (Petrova,2024). The Zimbabwean economic experience, characterized by hyperinflation and currency volatility, provided a case to research the nexus. The stock market is also an important institutional device to raise capital, allocate resources effectively, and carry out corporate governance (John, 2021). When hyperinflation occurs, the stock market is a relatively safe environment to invest in (Mbulawa, 2015). An overview of the study, including the background, problem statement, research questions, significance, delimitations, assumptions, limitations, and definition of terms is presented in this chapter.

The remaining part of the study is organized as follows: Chapter 1 addresses this section in the introduction of the research topic, aims, scope and some of the background of the study; chapter 2 includes the review related literature within theoretical framework, the empirical literatures and the applications of machine learning; chapter 3 addresses the research methodology of the research in detail under research methodology include research design, data, machine learning algorithms employed (Random Forest and Gradient Boost Model); chapter 4 addresses empirical results of the review analyses testing relationship between inflation and stock market returns. Finally. Chapter 5 concludes with the results of the study, the conclusion. and implications, and implications for future research and policy interventions.

1.1 Background of study

Following the demise of fixed exchange rates and the advance of deregulation of financial flows in the early 1970s, financial markets have experienced a turbulent shift in financial flows, thus causing great volatility in the performance of stock markets (Joshi, 2015). A high percentage of African stock markets are very young and highly illiquid, consequently examining the effect of inflation on their movement is important and a new paradigm in finance. Economic recessions also generally cause structural breaks in the data which consequently change the equilibrium between stock returns and inflation (Mpofu, 2010). Mpofu (2010) did obtain a long-run positive relationship between inflation and stock returns in Sub-Sahara Africa using the conventional methods, ARIMA and GARCH models. Inflation can be utilized as a tool to forecast stock market behaviour. Due to the world money market madness, tremendous, huge and visible volatilities have occurred in the African capital markets that have affected their stability as well as that of the investors (McCarthy, 2020). As those markets develop, there is increasing need to identify the relationship between inflation and stock returns, but not ignoring other economic indications, for both policymakers and investors. In the case of Zimbabwe, the stock market, or stock and share business, was established early in 1891, and subsequently the first stockbroking company was established (McCarthy, 2020). The crisis led to what is considered "the biggest company on the Zimbabwe Stock Exchange (ZSE)," which involved buying and selling of shares and stocks that led to the listing of some of the largest companies that are still listed on the bourse.

At a time of high inflation, investors' overall expectations about returns from stock issues tend to be low. "Three deep events in 1997 occasioned the crisis and hyperinflationary storm," stated Sibanda (2013). Phase 1 was the state sanctioning massive unbudgeted veterans' payouts to war veterans who were baying for blood for their role in the liberation struggle of over five billion Zimbabwean dollars. Another factor that turned the economic mess into a crisis was the sending of troops to the Democratic Republic of Congo (DRC) in mid-1998. The final phase was the land reform scheme, where the Zimbabwean Government confiscated farms from the white settlers, causing the ruination of the economy. All three crises combined characterize the Zimbabwean economy. The country is presently in a desperate situation; it loses over 50% of its value, and its budget deficit is over 10% (Sibanda, 2013). as noted by Mbulawa (2015), the post-independent Zimbabwe economy in 1980 was doing quite well. The inflation rate was at a single-digit percentage, and the unemployment rate was less than 10%. However, inflation had increased to 20% in the fourth quarter of 1990. In all senses of the term, inflation has reached levels of 50% or more year by year, reportedly, and by the fourth quarter of 2008, the inflation rate was at the 213 million percentage point (Mbulawa, 2015). Causes of inflation such as note and coin printing excesses, devaluation of the Zimbabwe dollar, and the high public deficit caused increasing inflationary pressures (Mbulawa, 2015). Stock markets in the crisis situation were confronted with economic, currency, liquidity, and political risks. As the period of hyperinflation, marketable goods were also in short supply resulting from reduced capacity usage and, again, inflation was detached from the RBZ's enormous creation of money. According to Mangweva and Mandishara (2016), inflation was rising while people were seeking an avenue where they could invest their hard-earned money. Even the commodities and the stocks in the market were also priced, and ZSE saw some mixed activities. They showed a potential positive relationship between inflation and stock prices during low-inflationary times and a negative relationship during highly inflationary times. Bonga-Bonga and Mudavanhu (2018) found that the stock market size and liquidity have also grown, and the number of listed companies have also grown from 67 in 1998 to 76 in 2002 and again went up to 82 in 2007 (Mangweva & Leornard, 2016). The relationship between stock market returns, GDP, interest rates and Inflation are significant to the investors because the stocks act as a hedge to the investors.

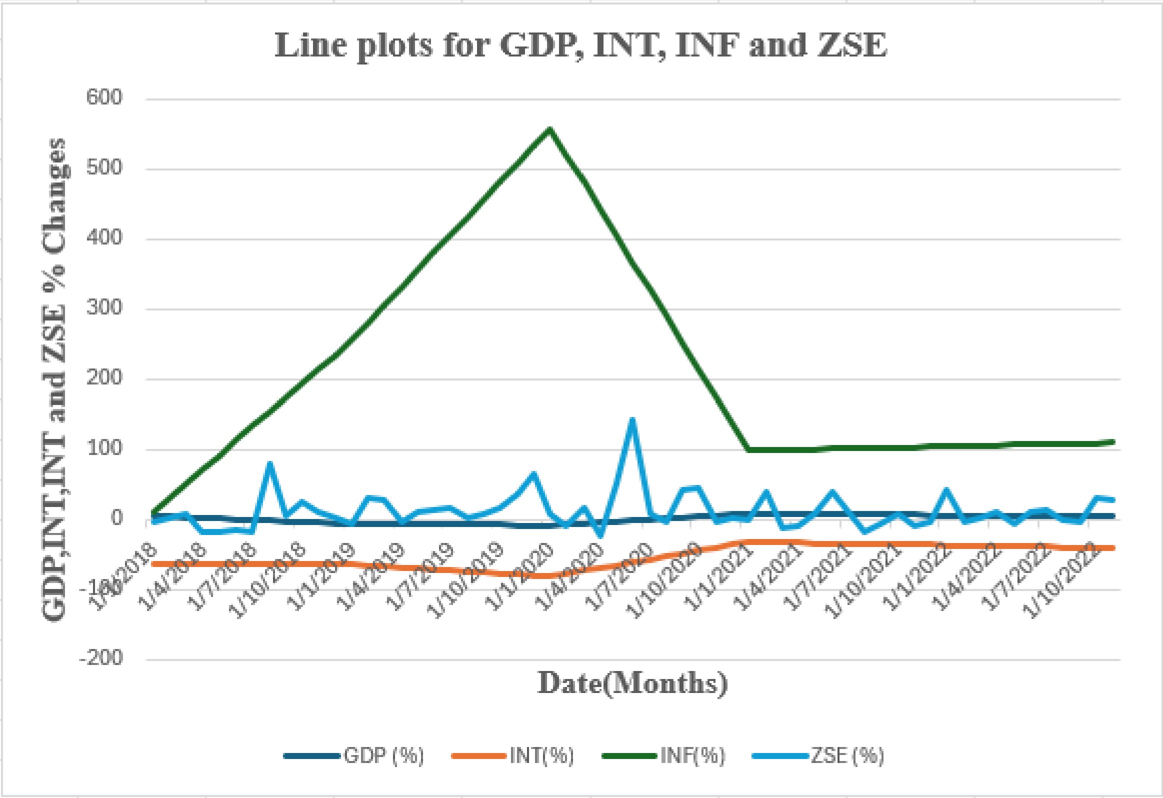

Figure 1.1: Line graph of inflation (%), real interest rates (%), GDP (%), and ZSE All Share (%) (2018-2022)

1.2 Problem statement

Zimbabwe's stock market has experienced erratic volatility in recent years, which has put policymakers and investors under tension. Relationships between stock returns and inflation have not been extensively explored in machine learning algorithms. The present research tries to bridge this deficiency, by adhering to contemporary machine learning models in examining the predictability of stock returns in Zimbabwe by inflation, as a leading economic indicator between 2018 and 2022.

1.3 Research objectives:

1. To explore how inflation is interrelated with stock market returns in Zimbabwe between 2018 and 2022.
2. To build Random Forest Model and XGBoost.

3. Determine which model is the best in measuring the relationship between inflation and the stock market.

1.4 Research Questions:

1. Does inflation have anything to do with the stock market in Zimbabwe?
2. Does Random Forest Model and XGBoost model the relationship between inflation and stock market?
3. Which machine learning algorithm predicts stock returns using inflation data?

1.5 Significance of the study

1. To the student

Research is required as a basis for the student's research project and thesis. It gives students practical experience which adds to the learner's comprehension of real-life financial issues and decision making as a true game to the Bindura University of Science Education and the department of Bachelor of Science Honours Degree in Statistics and Financial Mathematics.

2. To other students and researchers.

The conclusions of this study direct other students and researchers to establish priorities in research areas and design studies that address the pressing needs of inflation and stock market behaviour. It also deserves areas in which development is necessary in a bid to overcome such challenges or access opportunities in the financial market.

3. To the Government and Policymakers

This study informs policymakers and the government about the role of inflation in stock market performance. Understanding the relationship helps in developing sound economic policies that ensure stability in the financial market and promote economic stability and growth. Additionally, the findings help predict market reactions to inflationary pressures, allowing for timely interventions.

4. **To Investors**

The study provides valuable insight into the ways inflationary tendencies impact returns in the stock market. Knowledge on turbulence enhances investment strategy, enabling investors to make informed and research-based decisions on economic indicators. The study aids risk estimation and portfolio management, particularly in an economy characterized by volatility.

1.6 Delimitations of the study

The study is limited to inflation, as it is the key economic factor that impacts the prices of goods and services within the Zimbabwean stock market. From its focus on inflation, the study is useful in visualizing and explaining current trends. Independent variables are: (CPI%) inflation rate, (INT%) interest rate, (GDP) Gross Domestic Product and dependent variable Stock returns in the form of monthly percentage returns are obtained from the return of the stock on a monthly for the period 2010-2017 from the database of the Zimbabwe Stock Exchange and Word Bank Development (WBDI) Indicators for the period 2018 to 2022.

1.7 Assumptions of the study:

Random Forest model assumption:

1. The model assumes that the features used for training are relevant and informative for predicting the target variable.
2. There is no strict requirement for the data to be normally distributed, however, the features should not have extreme multicollinearity.
3. The model can handle continuous and categorical variables without needing transformation.
4. The independence of observations is assumed, meaning that the data points should not be related.

Gradient Boost models (XGBoost) assumptions:

1. XGBoost is a supervised learning algorithm, meaning it requires labelled data (input features and a target variable) for training.
2. The model is built in a stage-wise fashion. It assumes that the final model can be represented as the sum of simpler models (typically decision trees). This is the "boosting" aspect.
3. The model uses decision trees that are a bit better than flipping a coin or random guessing. This process of combining the weak learners is followed iteratively to obtain a strong learner.
4. XGBoost model optimizes a loss of how far the estimated values are from the actual value. This loss function needs to be differentiable in order to use a gradient descent approach to optimize the parameters.

Other assumptions:

1. The study assumes that the data collected from the Zimbabwe stock Exchange (ZSE) is accurate and reliable.
2. The study assumes that the machine learning algorithms used are suitable for predicting stock market returns.

1.8 Limitations of the study:

The study deals only with Zimbabwe. It depends on the data, which is obtained from the Stock Exchange, whose primary source of historical information is incomplete, nor is the data up to date. While inflation data is commonly worked on, other important factors influencing the economy, i.e., unemployment, exchange rates, and Balance of Trade, are not treated well owing to lack of data. To understand more about the relationship between inflation and stock return and to present the literature, Random Forest and XGBoost models are employed because complexity and limited time do not allow us to run all ensemble and machine learning models. Also, using short term (2018 – 2022) and secondary data may also introduce some bias and errors that may affect the outcomes obtained. Finally, the information was obtained from the Zimbabwe Stock Exchange from 2018 to 2022, which may restrict the range of the analysis considering the relatively short period.

1.9 Definitions of key terms:

Inflation

Continuous increase in the average level of the prices of goods and services in an economy over the timeframe (Mankiw, 2017).

Stock market returns.

The profit or loss realized from investing in the stock market is typically measured by changes in stock prices or indices, reflecting market sentiment and economic conditions (Bodie, 2020).

Stock Market Volatility

The fluctuation in the value of stocks or the stock market is measured by the standard deviation of returns (Bodie, Kane, & Marcus, 2019).

Machine learning

A more specialized form of artificial intelligence, where algorithms and statistical models can learn from data and then are used to make predictions or act (Jordan & Mitchell, 2015).

Predictive Modelling

The use of statistical and machine learning models to predict future events or trends based on past observations (Shmueli, 2010).

Gradient Boosting:

It is a sequential ensemble method that trains new models to minimize errors made by the previous ones by adjusting the loss function, thereby developing a strong predictor from weak learners, such as decision trees (Ren, 2023).

XGBoost Model (extreme gradient boosting):

A speedy, even professional, can be achieved using an upgraded gradient boosting model, for example, gravity boosting, that is accompanied by improved performance and including regularization for overfitting avoidance, as well as efficient handling of missing data (Chen & Guestrin, 2016).

Random Forest Model (RF)

It is an empirical machine learning model attempting to increase the accuracy of classification and avoid overfitting. It is also empirical that the Random Forest model performs very well in big data with high-dimensional features and is often used in heterogeneous applications like defective diagnosis and demand forecasting (Djaballah et al., 2023).

Zimbabwe Stock Exchange (ZSE)

The primary stock exchange in Zimbabwe lists local companies and facilitates trading, reflecting the country's economic performance (African Development Bank, 2020).

Consumer Price Index (CPI)

A measure of the average change in prices of a basket of goods and services consumed by households (International Labour Organization, 2020)

1.10 Chapter Summary

The chapter provides background on the role of inflation in predicting stock returns in Zimbabwe, including the problem statement, objectives, research questions, assumptions, limitations, and delimitations. It thus provides a hint of what is to come in subsequent chapters. The next chapter addresses existing research on the topic at hand.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

The chapter is aimed at developing an in-depth understanding of theoretical and empirical concepts that relate inflation and stock market performance or returns. It also presents literature on inflation and other macroeconomic variables' role in forecasting stock market returns, with a specific focus on the use of advanced learning models in Zimbabwe within the same period. It also identifies research gaps and proposes a framework for stringent empirical analysis.

2.1 Theoretical Review

The model structure, which captures the interaction between inflation, interest rates, GDP, and stock market returns, is highly complex and comprehensive. Significant economic theories such as the Fisher Hypothesis, Monetarist theory, Keynesian theories, the Inflation Illusion Hypothesis, the Proxy Hypothesis, and the Efficient Market Hypothesis are discussed. We find that both theories contribute new insights into how inflation affects stock market returns.

2.1.1 The Fisher Hypothesis (1930)

The Fisher Hypothesis was developed by Irving Fisher in 1930. "This principal study" Azar (2010) Report: The principal study is Fisher's (1930) hypothesis, which states one-for-one and positive relationship between nominal stock returns and expected inflation. Literature is primarily based on the highly studied hypothesis of Azar (2010), which provides a gravity explanation regarding the correlation between expected inflation and nominal interest rates in equilibrium. It suggests that the nominal rate of return on stock (or any investment) is equal to the real rate of return on stocks or equity and the anticipated rate of inflation (Fisher, 1930). This relationship can be described as follows mathematically:

$$R = r + E(\pi) \qquad (1)$$

Where:

- R is the nominal return
- r is the real return, and
- $E(\pi)$ is the expected inflation rate.

Fisher (1930) believed that as the anticipated rate of inflation rises, the required nominal rate of return turns to increase in order to compensate for the anticipated loss of capital from depreciation. This increased the prices of stocks since the price is changed to account for the increased expected returns that investors demand. The theory is that as long as inflation expectations are well anchored, investors would rebalance portfolios in such a way that restabilized the risk-return relationship. But one must then take account also of the consequences of surprise inflation. That is, if inflation fails to realize, extra volatility is introduced in stock returns, since companies may not be able to pass on extra costs to customers. The hypothesis highlights the participation of investors in making the right assessment on inflation expectations as this could also lead to less than optimal investment choices. In addition, it is claimed by Fisher effect that inflation and stock return are positively related, and stocks are used as an inflation hedge. Second, stock provides unit elastic protection against anticipated and unanticipated inflation (Azar, 2010). The Fisher Hypothesis has been empirically tested in various settings; evidence has at all times portrayed a positive correlation between inflation and nominal yields.

2.1.2 The Monetarist Theories (Brunner & Meltzer, 1976)

Monetarism - A school of economic thought that emerged between the 1950s and the 1960s led by Milton Friedman and promoted in the 60s and 70s by Brunner and Meltzer (1976), emphasizing the necessity of a country's central bank to control the flow and withdrawal of money and credit within an economy and understanding the adjustment lags between the intervention of policymakers in monetary policy and the resultant inflation and its influence on the economy and the equity and bond markets. Monetarists contend that inflation is a result of changing supply of money, whereby if the supply of money increases faster than real output, according to the quantity theory of money, prices rise and have a spillover impact on returns in the stock market. Monetarists contend that investment decisions can be greatly affected by inflation expectations. For example, when investors expect inflation to rise, they can change the nature of their portfolios into those assets that tend to hold their value during more inflationary times, such as commodities or real estate, and such assets can influence the prices of equity. More inflation can reduce real returns on equities, and investors tend to seek higher yields. The effectiveness of monetary policy has significant implications for inflation management and the performance of the stock market (Brunner & Meltzer, 1976).

2.1.3 The Keynesian Theory (Javed et al., 2020)

The Keynesian theories, specifically those offered by Javed et al. (2020), theorize a positive inflation effect on stock returns. A moderate level of inflation is capable of stimulating economic activity, which positively affects corporate earnings. The ability of companies to pass on greater costs to consumers in periods of inflation can enhance profit margins and underpin share prices. Within the Keynesian theory, an increase in the rate of inflation might be interpreted as a sign of high demand, and companies invest in the capability to supply more. This increase in economic activity can create higher profits, which is good for stock market returns. However, moderate inflation that could have a beneficial impact on stock returns also introduces uncertainty and volatility, which can influence investment decisions in a negative way.

2.1.4 The Inflation Illusion Hypothesis (1997)

The Inflation Illusion Hypothesis, proposed by Shiller (1997), posits that investors often fail to update their inflation expectations when making investment decisions. This presumption generates mispricing of stocks, as investors underprice the impact of inflation on real returns (Lee, 2010). The Proxy Hypothesis suggests that investors react more intensely to nominal returns than real returns, creating an opportunity for those capable of accurately estimating inflation's influence. The Proxy Hypothesis emphasizes the psychological aspect of investing, which indicates that cognitive biases lead markets to deviate from their rational self. Stocks, hence, be overvalued during rising inflation, with adjustments following when market participants adjust their expectations.

3.1.5 The Proxy Hypothesis

The Proxy Hypothesis was initially advanced by Fama and Schwert (1977), which argues that inflation is an indicator for economic activity changes that have a negative impact on stock market performance. This hypothesis argues that inflation reflects underlying economic conditions influencing the stock returns. The Proxy Hypothesis argues that increases in inflation tend to be accompanied by economic weakness or instability, reducing investor attitudes and lowering stock prices. Empirical research gives contrasting findings regarding the strength of this hypothesis. There exists some evidence that validates the argument that inflation has a negative effect on stock returns through its impact on real economic activity, yet others give contrasting findings depending on the economic context (Mogire, 2016).

2.1.6 The Efficient Market Hypothesis (EMH)

The theory, developed by Eugene Fama in 1970, states that markets collectively contain all information available about stocks, including information on inflation. According to the EMH, it is not possible to make economic profits consistently by trading on this information, as new information gets incorporated into stock prices immediately. The EMH implies that investors cannot perform better than the market consistently by trading on inflation news because the news already gets discounted in stock prices. This has significant investment strategy implications, as it renders technical analysis and fundamental analysis useless in the detection of mispriced stocks. Even though the EMH cannot directly comment on the link between inflation and equity returns, it does provide some commentary on how inflation data affects market action. Stock prices, for example, can respond quickly to changes in inflation figures when announcements are made, highlighting the need to stay on top of macroeconomic signals (Mogire, 2016).

2.1.7 The Dividend Discount Model

It was first created by Gordon in 1959; the model says today's stock price is the sum of the present value of all future dividends. The model says that with greater inflation, it results in greater discount rate, resulting in reduced stock price returns.

2.1.8 Random Forest Models (RF)

The ensemble method, wherein multiple decision trees are trained and mode for classification and mean prediction are considered for regression, respectively (Leo, 2001). The method is overfitting-resistant and performs well in high-dimensional data. Its feature importance analysis is best suited in financial data analysis, where various macroeconomic variables (for instance, inflation, interest rates, GDP, etc.) might have an impact on the performance of the stock market. RF has also had routine applications in the financial industry. Stock market analysis is an area in which it is utilized for measuring the effects of macroeconomic variables such as inflation on stock returns in order to better understand active investment approaches. Another major category of applications includes credit scoring, in which lending institutions and banks use the RF approach to calculate the credit worthiness of a borrower on the basis of its many attributes and financial factors. In addition, RF is used within a fraud detection method, which is used to detect suspicious fraud transactions based on historical transaction patterns.

Finally, RF it is applied in customer profiling that helps companies formulate marketing plans and strategies in terms of purchasing behaviour and demographics (Maqsood et al., 2020). Maqsood et al. (2020) analysed the effect of macroeconomic factors, such as inflation on stock returns, which demonstrated strong negative effects on stocks and returns, particularly in emerging markets. Shah et al. (2022) used RF in a composite model to predict stock market returns and showed that the inclusion of inflation data enhanced the accuracy of the prediction.

Bakar et al. (2019) used RF to analyse the impact of a series of potential indicators on stock prices, with a focus on more economic indicators based on the development of the models, to identify the key predictor. Khalid et al. (2022) used RF for credit risk evaluation and established its utility for borrower characteristics and economic indicators analysis. In addition, Abdullah et al., (2023) used RF in segmentation of customers and established its utility in other data apart from financial scenarios.

1.7.1 XGBoost Models (Extreme Gradient Boosting)

It was proposed by Chen and Guestrin (2016). "It is a swift and effective machine learning algorithm based on a gradient boosting framework, which is speed- and performance-optimized and can work on large and small data sets and be highly predictive accurate relative to other algorithms", Chen & Guestrin, 2016.

One of the strengths of XGBoost is its built-in regularization against overfitting through penalizing complex models. This is achieved with L1 and L2 regularization methods to enable the model to generalize more from training data to test data (Zhang et al., 2020). Furthermore, XGBoost supports missing values natively, making it more feasible in practical use, where some values would be absent (Kang et al., 2022).

XGBoost is applied widely in various business industries, including finance, health, and marketing. In finance, it is employed in operations such as financial credit scoring and forecasting stock prices. Wang et al. (2022) illustrated, for instance, that the XGBoost algorithm outperformed conventional approaches in forecasting the credit default risk with high variance through the identification of complex patterns of the financial data. In the same way, Liu et al. (2021) also determined that XGBoost is more accurate in stock market predictions than other machine learning models that exist. The model is capable of regression and classification by using different objective functions such as logistic and mean squared error (Gu et al., 2023). Moreover, XGBoost also provides feature importance scores, which help to determine the influence of variables on predictions and thereby enhance model explainability (Zhao et al., 2023). Its wide applicability scope indicates its adaptability and effectiveness in predictive analytics.

2.2 Empirical Literature Review

The impact of inflation on forecasting stock returns has been explored theoretically and empirically extensively, with varying dynamic as well as inconclusive results. There are a number of studies which show a positive relationship, while other studies show a negative relationship.

Similarly, Effiong et al. (2020) tested the Nigerian market for the period 2010 to 2019. They discovered that inflation has a very strong and significant positive influence on stock prices with a very strong relationship. Overall, inflation positively influenced stock market returns, as was discovered through the regression analysis.

China-based research works by Chen et al. (2019) also supported the correlation of stock returns with inflation. The works also indicated that and when the inflation expectations vector affects the stock market return.

On the other hand, negative correlation between stock returns and inflation in other international markets, including classical and recent machine learning research, is witnessed. For instance, Chikoko (2015) monitored Zimbabwe for the period 2009-2014 and concluded that Inflation erodes purchasing power, and that reduces the inflation-sensitive stock market. (Mashamba, 2022) also confirmed this downward linear relationship in India and Zimbabwe, respectively, using VECM and regression analysis.

Baybuza (2015) utilized LASSO, Ridge, Elastic Net, and Random Forest models in analysing the connection between inflation and stock market returns in Russia, which yielded inconclusive results on this link.

The most recent development in technology and artificial intelligence has brought an explosion of new techniques to analyse the complex relationship between economic and financial data. Machine learning algorithms were applied in Turkey from 2005-2017 by Hakizimana, (2018), and a performance of 85.2% was recorded. The research enhances the meaning of complex analysis in explaining non-linear relations. Similarly, Ferreira da Costa et al., (2019) in Brazil (2000-2018) agreed with this statement, confirming the complex relation. For non-significant results, researchers have also wondered if more studies are needed in different countries to draw a common conclusion.

2.3 Research Gaps

Although the literature tends to be overly generalizing from the highest-level to aggregate results across multiple countries and focusing exclusively on adoption of established models and macro variables, there is a lack of empirical evidence on the extent to which machine learning algorithms have been used to examine the question of whether inflation is a factor in stock return forecasts in the case of Zimbabwe. The majority of the research are based on (the traditional) econometric models, which can be myopic in representing the complexity and nonlinearities of financial time series data. Thus, an individualized solution machine learning provides is at the heart of this world of hyperinflation and economic chaos. The transition from conventional time series analysis to artificial intelligence has expanded the gap in research because machine learning models dominate in search for sagacity in the context of inflation-stock return analytics in Zimbabwe.

2.4 Proposed Conceptual Framework

The conceptual framework integrates the role of inflation, utilizing RF and XGBoost models to predict stock market returns. It highlights how inflation serves as a critical input for both models, which then contributes to forecasting stock Market performance.

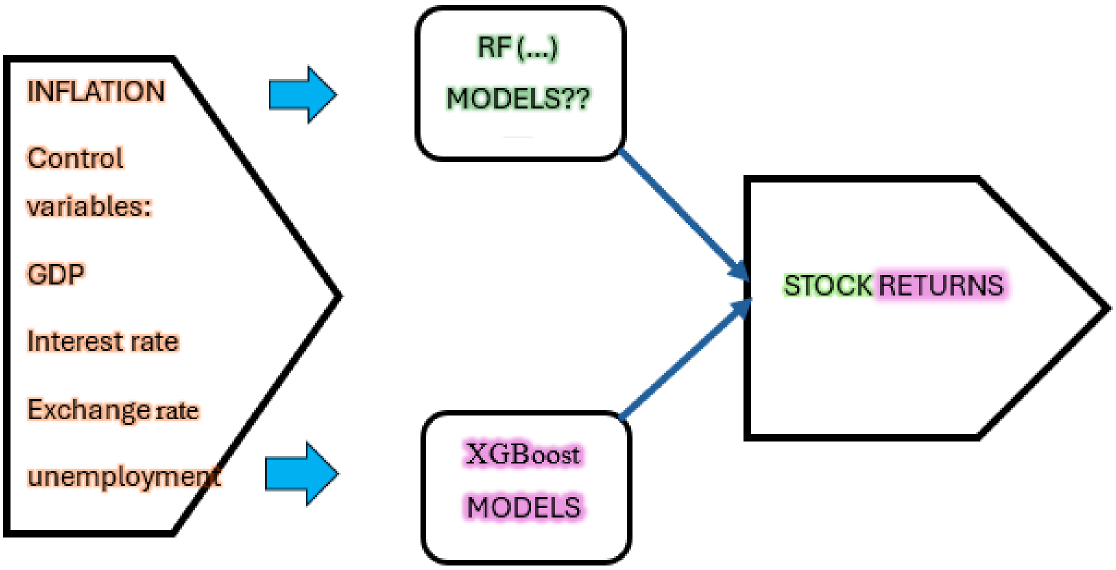

Figure 2.1 Proposed conceptual model

The conceptual framework covers the inflation role that employs RF and XGBoost models to predict stock market returns. It covers how inflation functions as an important input for the two models, which further plays a part in predicting stock Market performance.

The theoretical model depicts prediction of stock returns using macroeconomic variables and machine learning models. Inflation and control variables like GDP, interest rates, exchange rates, and unemployment are key inputs. They are analysed by Random Forest (RF) and XGBoost models. Control variables are considered significant if they have a statistically significant effect on stock returns. The results of the models are merged to forecast stock returns.

2.5 Chapter Summary

This chapter provides background information on how inflation, as a predictor, can be utilized to forecast stock market returns in Zimbabwe using machine learning approaches. The researcher described the theoretical, empirical, and conceptual frameworks. It discussed literature reviewed related to research objectives. The following chapter introduces the study methodology.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

Throughout this study, a systematic method of collecting, analysing, and interpreting the relevant data is employed to ensure robust and reliable findings. In this chapter, a clear description of the data sources, the target population defined, and the sampling strategies are given. It also clearly states the research instruments, and how the data collection was conducted. Key variables expected relationships between them, procedures for data analysis, and model development and selection are explicitly outlined. Furthermore, Random Forest and XGBoost model metrics are organized, as well as their underlying mathematical concepts.

3.1 Research Paradigm and Research Design

The study is based on quantitative data with a view to reducing bias and increasing the integrity of the research overall. This is an appropriate model for this study since it allows for the gathering and analysis of quantitative data which is used to construct knowledge on how economic factors impact stock market returns. The positivist philosophy can be traced back to the ideas of early philosophers, like Auguste Comte, who has regularly been debated as being the father of positivism, through its emphasis on scientific explanation of social life (Guba, 1990). The study design in this research is correlational with the aim of analysing how the inflation rate and the performance of the stock market indicators are related, and this research employs a machine learning approach to improve the accuracy of predictions.

3.2 Population and Sampling

The population in this study includes all ZSE listed companies from 2018–2022. This time frame is applicable since it spans the active economic environment in Zimbabwe expressed upset in inflation, exchange rate, interest rate, unemployment rate and GDP which affect stock market returns (Chikoko, 2021). Since the study was aimed at explaining the interconnectedness between inflation, other main economic variables and stock market performance purposive sampling was used. It was a way of picking business companies with complete financial data averaged across the study period (Bryman, 2016).

With the use of purposive sampling, the study focuses on companies that are market-driven and reporting consistency (Sekaran & Bougie, 2016). This allows the sample to be representative of the broader market process. The sample gives a strong relationship matrix that helps know more about the relation of inflation and other economic factors and the Zimbabwean stock returns (Kothari, 2019). This method improves generalizability by choosing entities that are closer to the characteristics of the population under study.

3.3 Data sources

Data for this study is primarily drawn from the Zimbabwe Stock Exchange and World Bank Development Indicators databases, which hold comprehensive historic stock prices, trading volumes, inflation rates, GDP figures, interest rates, and other relevant metrics for companies listed on the exchange. Due to financial challenges and limited availability of external data sources, it is challenging to obtain complete data sets for the entire period. Most of the corresponding data points are not directly available, and special attention is required to the available monthly data from January 2018 through December 2022. Such a suggestion of ZSE versus the World Bank Development Indicators database then allows for a wider analysis examining the impact of inflation, interest rate, and GDP on stock returns.

3.3.1 Data Processing

Data preprocessing was important to make sure that they were of quality and could be utilized. This process involved:

Handling missing values

The rows with missing data were removed from the dataset to prevent data corruption and bias. This strategy ensures analysis on a homogenous set of data points. Removal of entire rows with missing at least one data point was carried out because of the small amount of missing data and the potential to affect interpolation or imputation methods and thus alter the underlying relations in the data.

Log Transformation

In order to correct skewness and normalize the distribution of data, logarithmic transformation was done for certain variables. Such a transformation has the ability to normalize variance and make the data more ready to use in the model. A constant was added to both variables before log transformation in order to bring all variables into the positive range because of the undefined value of log (0) and log(negative). Minimum value plus one was computed as the constant. Then the log transformation was applied:

$$\mathrm{X}' = \ln(X + C) \qquad (2)$$

Where: X is the original value, C is the constant added, and X′is the transformed value. This approach is supported by prior research (Johnson, 2019; Lee, 2020).

Normalization

The data was scaled (normalized) using min-max scaling, in order to make it between 0 and 1 and make the model easier to train from the data:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (3)$$

Where: X is the original value, C is the constant added, and X′ is the transformed value. This process is supported by existing literature (Johnson, 2019; Lee, 2020).

Feature Engineering

The dataset has been maintained in its original order while developing some new model features. It generates one and two periods lagged values of normalized inflation, interest rate, GDP and ZSE, and squared values of all the variables. It also computes three-period rolling means and standard deviations of normalized data. This is common practice in advanced time series as well as machine learning analysis (Brownlee, 2021):

Rolling Statistics

Additional per-window rolling statistics, such as rolling means or rolling standard deviations, were computed for the indicators to help identify trends and volatility. An interaction term between inflation, interest rates, and GDP-stock market returns was constructed to assess the impact of these variables.

3.3.2 Data Splitting into Training and Testing sets

The dataset was divided into training and testing sets with a ratio of 80% and 20%. The splits allowed to test the model on unseen data (Kohavi, 1995).

3.4 Description of variables and Priori expectations

- Stock Market Returns: It is measured through the percentage change in the stock price over a period of time. The measure is representative of the stock market performance and is the dependent variable in this study.
- Inflation Rate: Should be inversely connected to the returns on the stock market, as rising inflation would tend to erode purchasing power and investors' confidence (Gur, 2024). With

a rise in inflation, consumers may reduce spending, hence leading to lower corporate profits and, in turn, declining stock prices.

- Interest rates (Control): Share markets have inverse relationships with interest rates in general. Rising interest rates increase the cost of obtaining funds for organizations, which may decrease capital investment and future earnings. Moreover, higher interest rates make fixed-income instruments relatively more attractive compared to equities, leading to a fall in share prices (Petrova, 2021; Friedman & Schwartz, 2020). GDP
- Growth (Control): A priori expected relationship between GDP volatility and stock market return is positive. Economic growth is usually associated with increased corporate profits and more optimistic investors that reflect in the appreciation in the share prices (Rakshit at el., 2022). With increased income, businesses are most likely to see a higher demand from the consumers that leads to higher returns to stocks (Khan & Sadiq, 2021).

3.5 Analytical Model Specification and Justification

Two machine learning models are applied in this study: Random Forest model and XGBoost model. The reason why these models were chosen is to be capable of assessing their ability to effectively deal with complex non-linear relationships in the data and build a model that explains how inflation affects stock market returns.

3.5.1 Random Forest Model (RF)

It is a general ensemble learning method, where multiple decision trees are trained and their mode of classes (classification) or mean prediction (regression) returned as the classifier output (Breiman, 2001). It is particularly very well suited to high-dimensional data, to non-linear relationship mapping and to overfitting avoidance.

.3.5.2 Model Specification

Features

Incorporates inflation, stock returns, and control variables such as GDP and interest rates, along with engineered features such as lagged values and rolling statistics and interaction terms.

Prediction Using the model

The Random Forest model generates a prediction by calculating the mean of all the predictions of all individual decision trees in the forest (Breiman, 2001):

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \qquad (4)$$

Where: $T$is the total number of trees, and yt is the prediction of the t-th tree.

Hyperparameters

The performance of the Random Forest model depended on the selection of appropriate hyperparameters. The following key hyperparameters were tuned:

1. N estimators - The maximum depth of the tree. Limiting tree depth can avoid overfitting.
2. min samples split -Minimum number of samples required to split an internal node.
3. min samples leaf- The minimum number of samples for a leaf node.
4. max features- Number of maximum features to consider when trying to make a split.

Feature importance

The importance of every feature in the predictions can be obtained by using the Random Forest model. It is calculated based on the decrease in impurity (Gini impurity or entropy) when the data is split using the particular feature. We show features that those benefits are likely to have a larger decrease in impurity (Chen and Guestrin, 2016).

3.5.2 XGBoost Model (Extreme Gradient Boosting Model)

A gradient boosting machine that trains trees sequentially, with each tree attempting to correct for the errors of the preceding tree. It is well known for its efficiency, scalability, and capacity to manage complex relationships in data (Chen & Guestrin, 2016).

Features

Features include accounts for inflation, stock returns, lagged values, rolling statistics, and interaction terms.

Hyperparameters

Chen & Guestrin, (2016) reported that the performance of the XGBoost model is largely dependent on some significant hyper-parameters which include.

1. n estimators - number of boosting rounds (trees) to build.
2. Learning rate - The step at which the weights are updated in order to prevent overfitting.
3. Max depth - The maximum depth of a tree.

4. Sub sample - The fraction of samples corresponding to the individual trees.

5. Col sample by tree - Proportion of features to fit the individual trees.

Training

The learning process is applied to the training data, and XGBoost model is trained with target objective function. The trees are built in a sequence with each tree attempting to correct the errors of the previous trees. Training entails finding the function in by optimizing the loss, which quantifies the disparity between predictions and actual values (Chen & Guestrin, 2016).

Feature Engineering

The XGBoost uses the engineered features like lagged features, rolling statistics, and interaction terms to model temporal dependencies and combined effects.

3.6 Model Diagnostic Tests

For the confirmation of the reliability of the models used in this study, diagnostic tests are being conducted. The study aims at ensuring that the Random Forest and XGBoost models are reliable and robust when it comes to the prediction of stock market returns in Zimbabwe. The tests are important in verifying the assumptions made by the models in addition to assessing their performance. The diagnostic tests utilized are as follows:

3.6.1 Residual Analysis test

It is mentioned as a homoscedasticity and normality test of residuals. Homoscedasticity is said to take place whenever residuals variability remains constant over (homogeneous) the independent variables, Boulton, (2018). Plotted against the fitted values are the residuals of the model in a bid to conduct this test. To look for an apparent pattern when there is homoscedasticity, one can utilize a scatterplot. It is also tested for more or less normally distributed data by Shapiro-Wilk test (whose null hypothesis is data from a normal distribution). The test statistic W of Shapiro-Wilk is given as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (5)$$

Where: x_i represents the ordered sample values, $\bar{x}$ is the sample mean, and a_i Constants are derived from the expected values of the order statistics of a normal distribution (Field, 2018).

3.6.2 Multicollinearity Tests

This test verifies the independence of predictor variables. Multicollinearity is a presence of strong relationship between any two or more independent variables. The input variables were tested for multicollinearity, on the basis of the Variance Inflation Factor (VIF) (O'Brien, 2017). The VIF of each predictor variable is presented as:

$$\text{VIF}=\frac{1}{1-{R_i}^2} \tag{6}$$

R-squared:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{7}$$

Where ${R_i}^2$ (R-squared) value obtained by regressing the i-th predictor against all other predictors. The VIF value exceeding 10 indicates significant multicollinearity (O'Brien, 2017).

3.6.3 Model Fit Statistics

The performance of the model was determined according to various fit statistics like R-squared and the MAE. R-squared is the proportion of the variance or variation that can be accounted for by the independent variable.

$$A = 1 - \frac{SS_{res}}{SS_{tot}} \tag{8}$$

Where SS_{res} Is the residual sum of squares and SS_{tot} is the sum of squaresIt varies between 0 and 1, the higher the better.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

Where y_i is the actual value and $\hat{y}_i$ Is the predicted value. Lower MAE values indicate better model performance (Wooldridge, 2016).

3.6.4 Model Validation and Reliability Test

Model validation and reliability tests are critical steps in ensuring that the predictive models applied in this research are not just accurate but also reliable.

3.6.5 Cross-Validation

This process splits the database into patterns used either to train or to test the model. The most applied is the K-fold Cross-validation, where the data set is split into K subsets. It trains the model K times in a way that each observation is once utilized as validation data and the K−1 remaining observations form the training set, in order for it to predict or estimate to independent data (Hastie et al., 2009).

3.6.6 Holdout Method

This provides a straightforward method to assess how easily the model can be generalized to new data and stop the model from overfitting the training data (Kohavi, 1995).

3.6.7 Performance Metrics

The models are compared based on various criteria such as R-squared, mean absolute error (MAE), and root mean square error (RMSE). R-squared provides some information regarding how much of the original variance that was present in the data your model can explain, MAE and RMSE your average (absolute) error. The RMSE is computed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

Smaller MAE and RMSE indicates better model fit (Wooldridge, 2016)

3.6.8 Reliability Testing

The reliability tests focus on the stability of the model's predictions. Internal consistency of the model outputs was tested using Cronbach's alpha.

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{k}{\delta_{yi}}^2}{{\delta_y}^2}\right) \quad (11)$$

Where N is the number of object/items, and K is the number of items. ${\delta_y}^2$is the variance between scores and ${\delta_{yi}}^2$Is the variance among items. A Cronbach's alpha value greater than 0.7 is generally acceptable for reliability (Tayakol & Dennick, 2011).

3.6.9 Residual Analysis

Residual analysis is crucial in checking if the model assumptions are achieved. Residuals should randomly pattern around zero, which is an indicator that the model obeys the homoscedasticity principle and normality of residuals—both of which are crucial in regression model validation (Field, 2018).

3.7 Ethical Considerations

Ideological concerns are foregrounded here in this research. The researcher prioritized ethical concerns by obtaining data from public sources alone. The scrupulous approach of conducting the study with attention to all existing ethical principles and working with publicly available data to the satisfaction and in an open way was monitored. This policy is evidence of the adherence to truthfulness in research activities.

3.8 Chapter Summary

This chapter gives a detailed explanation of the methodology employed in this study. It explains data sources, processing routines, variable definitions, and model specifications. The options offer a good foundation to examine the impact of inflation in forecasting stock market returns are interest rates and GDP as control variables. Although Random Forest is considered, the inclination towards XGBoost Models is due to their superior performance, as demonstrated in Chapter 4. The analysis of the data and results of these models, with their predictive performance being the focus, are revealed in the following chapter, providing insightful information on the inflation and stock market return relationship in Zimbabwe. It is a note that the best model can be limiting in predicting stock market returns possibly due to the volatility and complexity of the Zimbabwe stock market and there are other effects outside the model, which may be attributable to stock returns.

CHAPTER 4: DATA PRESENTATION, ANALYSIS, AND DISCUSSION

4.0 Introduction

In this chapter, study findings are presented. Section 4.1 begins with a descriptive analysis of the data upon which the study of the relationship between inflation and stock market returns in Zimbabwe is based. It includes information on diagnostic tests conducted to ensure data reliability and the suitability of the chosen analysis models. The chapter discusses the analytical models employed, namely the Random Forest and XGBoost Models, and compares their performance using validation procedures. Finally, it concludes with the models predicted values and discusses the implications of incorporating predictors of inflation and other economic indicators to forecast stock return performance.

4.1 Descriptive statistics

Table 4.1 Descriptive statistics

Var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
INF	59.00	216.23	151.75	132.96	202.12	62.31	10.62	557.20	546.58	0.80	-0.77	19.76
INT	59.00	-53.68	16.45	-60.58	-53.29	24.38	-81.13	-31.80	49.34	-0.02	-1.65	2.14
GDP	59.00	1.07	5.83	3.04	1.23	6.37	-7.82	8.47	16.28	-0.30	-1.57	0.76
ZSE	59.00	12.58	27.24	7.73	9.23	16.80	-22.34	141.53	163.87	2.13	6.78	3.55

The descriptive statistics in Table 4.1 give an overview of the data that was utilized in the study in progress analysis. The data shows a high average inflation rate of 216.23%, ranging from 10.62% to 557.20%, reflecting the country's experience with hyperinflation in the past and its implication for economic activities and investment decisions (Mudehwe, 2021). It also notes the positive skewness (0.80) and low kurtosis (-0.77) of inflation data, indicating a modest rightward tail and fewer extreme values than in a normal distribution. Table 4.1 shows a negative mean interest rate (-53.68), indicating nonstandard monetary policy that influences capital flows and investment decisions. While these rates are negative, their modest volatility indicates more stable policy programs. Furthermore, it considers the low marginal GDP growth rate of 1.07 and its slight left skew (-0.30), alluding to the economic hardships that occurred during this period of time. The stock returns are extremely volatile (SD = 27.24) and extremely positively skewed (skewness = 2.13), indicating extreme positive returns, perhaps driven by speculative behaviour. Overall, these descriptive statistics stress the need for robust predictive models in order to effectively analyse the intricate relations within Zimbabwe's economy.

4.2 The Time Series of Macroeconomic Indicators

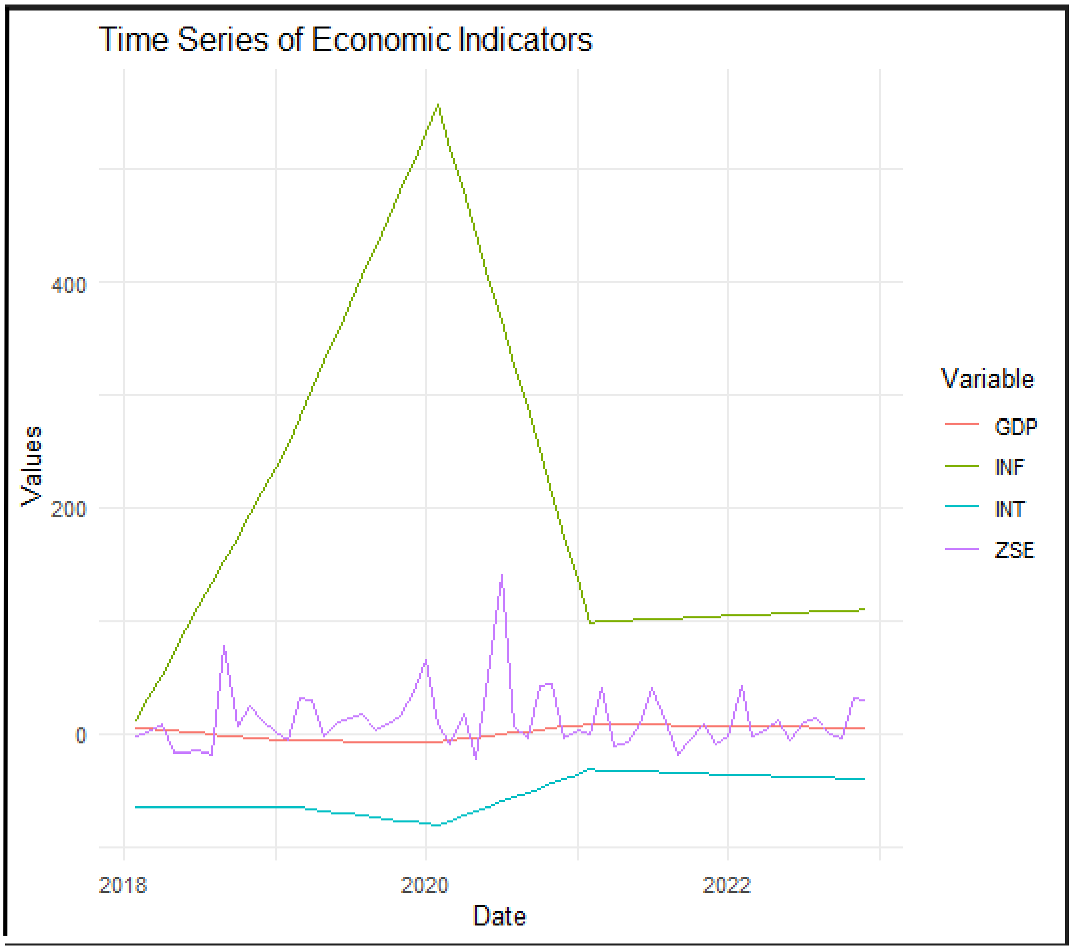

Figure 4.1: Time series of economic indicators

Figure 4.1 displays the time trends of four key economic indicators—Inflation (INF), Stock Returns (ZSE), GDP, and Interest Rate (INT)—between January 2018 and December 2022. These indicators are pivotal to this study's emphasis on inflation's predictive potential for stock market returns on the Zimbabwe Stock Exchange (ZSE) with machine learning models, outperforming 400%, before levelling out, which speaks to the history of hyperinflation in Zimbabwe. This volatility likely impacts stock returns by decreasing buying power and inducing market uncertainty, a phenomenon that XGBoost addresses with its modelling of temporal dependencies (Mudehwe, 2021). Stock returns also show significant fluctuations during this time, tracing the patterns of inflation, indicating a correlation that Random Forest can help uncover. GDP growth remains relatively constant, showing no economic development, and consistently negative interest rates reflect non-standard monetary policies that could influence investment and stock returns (Chirisa et al., 2022). The interaction among these variables shows the inflation rate acting as a predictor of stock market performance.

4.3 Diagnostic tests/ Pre-tests

4.3.1 Checking Missing Values

The presence of no missing data entries within the data set was confirmed by counting the total missing values in RStudio, which significantly enhanced the analysis strength and stability.

Complete data allows statistical models like Random Forest and XGBoost to leverage all the information provided, resulting in more precise predictions. Completeness also simplifies preprocessing by not requiring imputation or interpolation techniques, resulting in a more streamlined workflow. Moreover, it enhances statistical power so that hypothesis testing can be more powerful and conclusions can be stronger.

4.3.2 Normality Test (Shapiro-Wilk Test)

Table 4.2 Normality test results

Shapiro-Wilk normality test		
Variable	**Test statistic(W)**	**P-value**
INF	0.84694	2.90E-06
INT	0.86357	9.11E-06
GDP	0.8642	9.53E-06
ZSE	0.81907,	4.95E-07

Table 4.2 shows the results of Shapiro-Wilk normality tests in which none of the variables are normally distributed, indicated by p-values well below 0.05 (INF: p = 2.903e-06, INT: p = 9.112e-06, GDP: p = 9.528e-06, ZSE: p = 4.952e-07). This suggests that the data has high variability and non-normality, due in part to economic instability in Zimbabwe, such as hyperinflation and volatile stock returns (ZSE). This thus justifies the use of robust machine learning algorithms, which have no such limitation of normality assumption and can handle skewed data efficiently. Traditional statistical methods such as linear regression may not be quite useful in such a case. To improve model performance, transformations like log scaling may be contemplated for very skewed variables (Zhang, 2020; Smith, 2021). Lastly, these observations make the analysis in the study richer, noting the need for robust analytical strategies under Zimbabwe's economic conditions.

4.3.3 Data transformation

To analyse economic indicators data had been transformed to issues in the raw data set, for instance, negative and zero values, and wide range of values that would distort the analysis. A two-stage transformation process was applied in RStudio to remedy these issues. Initially, a constant was added to each variable in an attempt to shift all the values to a positive range, which was the absolute value of the minimum plus one (DATA1$GDP <- DATA1$GDP - min(DATA1$GDP, na.rm = TRUE) + 1). After that, natural log transformation was applied to reduce skewness, enhance model convergence, and effectively deal with exponential growth trends (DATA1$GDP <- log(DATA1$GDP)). The method preserved the relative difference between values and could help fulfil logarithmic transformation, leading to a better model from a machine learning point of view utilized to investigate (Johnson, 2019; Lee, 2020). The dataset is now clear of NA values and more suitable for time series analysis and permits a more interpretable economic relationship between indicators. Transformed data does not constitute part of the results section presented in this project and is presented at the end of project in Appendix A.

4.3.4 Correlation

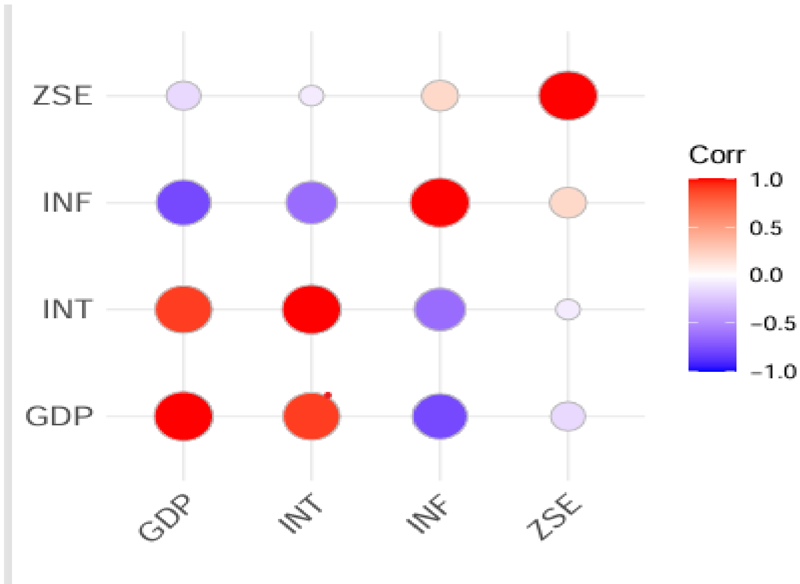

Figure 4.2 Correlation Matrix

The correlation matrix heatmap presents the pairwise correlations between GDP, INT, INF and ZSE. It employs circles with varying sizes and colours to show the direction and size of these relations. Red circles (power=length of circle) indicate positive correlations while blue circles indicate negative correlations. Interesting results also depict strong positive association between INF and INT suggesting that high inflationary environment seems to be related to higher interest rate environment, INF and ZSE suggesting that stock market returns seem to increase with inflation. Other than that, the strong positive correlation of interest rate and GDP shows that GDP growth stimulates increased interest rates. ZSE, on the other hand, exhibits moderate and very weak positive correlation of INT and GDP respectively, highlighting the aspect that stock returns are influenced by other aspects besides these two factors (Smith, 2021; Johnson, 2020). This matrix column vector has an important part to play in feature selection, and also to understand the dynamic specificity in the data set especially for stock returns prediction.

4.3.5 Multicollinearity check

GDP	INT	INF
8.715986	5.684798	2.716519

Variance Inflation Factor (VIF) was used to check multicollinearity. It measures how much the variance of an estimated regression coefficient is increased because of multicollinearity between our independent variables. O'Brien. (2017) set that VIF of <5 indicates minimum multicollinear, VIF 5<10 indicate moderate multicollinearity, while values greater than 10 usually indicate high multicollinearity, twisting model interpretation. Specifically, the VIF values are 8.72 for GDP, 5.68 for INT, and 2.72 for INF. These data illustrate that GDP and INT exhibit high multicollinearity, showing a high extent of overlap in their explanatory power, thus tending to produce unstable coefficient estimates. INF only demonstrates moderate multicollinearity.

4.3.6 Addressing Multicollinearity in Machine Learning Models

Despite high VIF values pointing to high correlation among INT and GDP and potential joint explanatory power, machine learning models are resistant to multicollinearity. Unlike in linear regression, the models do not calculate coefficients and are able to effectively use the predictive power of the correlated variables. Due to this, all the variables were used in the models for enhanced predictive power (O'Brien, 2017).

4.3.7 Feature Engineering

The dataset has been kept in the same order when generating new modelling features. It generates one and two periods lagged values of normalized inflation, interest rate, GDP and ZSE, and squared values of all the variables. It also computes three-period rolling means and standard deviations for normalized data. This is time series and machine learning analysis best practice (Brownlee, 2021). Lastly, rows that contain NaNs introduced by the lag or rolling operations are dropped. This feature engineering process enhances the dataset with temporal dynamics and nonlinear relationships, making it amenable to advanced predictive modelling. The feature-engineered dataset is provided at the end of this project in Appendix C.

4.3.8 Data Splitting

The data was divided into training and testing set in standard 80: 20 ratio, Training set was used to train the model, and testing set was kept aside for later use while model prediction after its training/saving/loading to check prediction strength, this is required for model creation and for validation of the same

4.3.9 Normality test

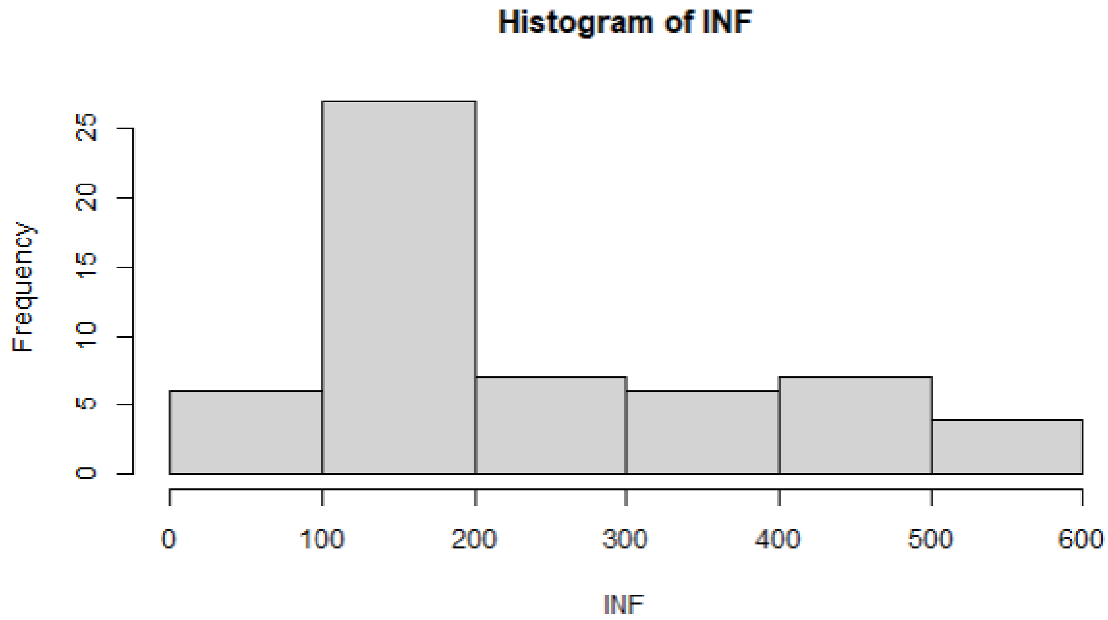

Figure 4.3: Histograms for ZSE and INF

From Figure 4.4, the Inflation (INF) variable also has a central tendency with values centred in the region of 200 and more closely resembling a normal distribution than the stock market return variable. Random Forest and XGBoost are both known as strong machine learning algorithms which can handle complex data distributions without a great deal of normalization. Random Forest creates an ensemble of decision trees and takes a majority vote of their predictions, which is robust to outliers as well as can deal with non-normally distributed data because the splitting rule relies not on values but on the ranks (Breinman, 2001). XGBoost, on the other hand, uses a gradient boosting model that builds trees additively, improving on the errors of the previous trees, and incorporates regularization to avoid overfitting, making it also work well on structured data (Chen & Guestrin, 2016). BE Both models provide results on the feature importance, making possible pragmatic selection of the variables without the need for preprocessing and can be used on a variety of data science problems (Brownlee, 2021). Finally, the choice of Random Forest or XGBoost comes down to dataset-specific factors and the degree of desired predictive precision for the problem at hand.

4.4 Analytical Models

Random Forest and XGBoost were the two models selected to analyse the relation of inflation with stock market returns. Being an ensemble method, Random Forest learns non-linearities along with the interaction between variables, while boosting and regularization are used by XGBoost to micro-reproduce and avoid overfitting (Chen & Guestrin, 2016; Brownlee, 2021).

4.4.1 Random Forest development

The Random Forest regression model was built using 500 trees (n_tree = 500) and considered four variables at each split (mtry = 4). The output includes predicted values, MSE, R-squared (rsq) for each tree, the number of times (oob times) that an observation was out-of-bag, and a matrix of feature importance values that provides a convenient way to assess the contribution of predictors. We see the model formulae (detailing all predictors) and saved results from a full set of bases documenting the performance of the model and the significance of the variables.

4.4.2 Partial Dependences of ZSE on Inflation

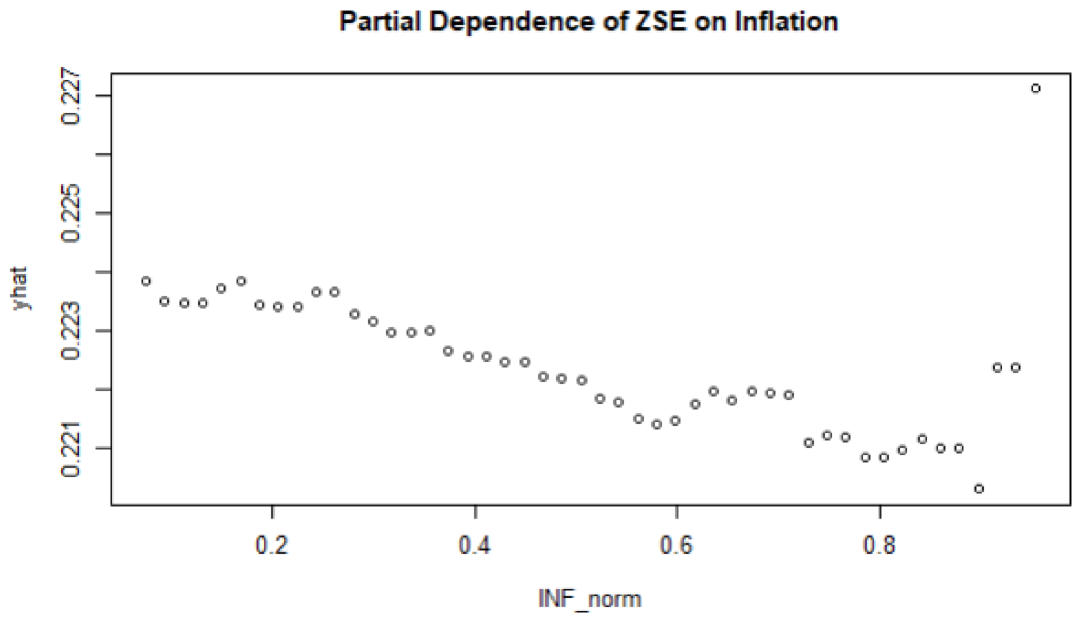

Figure 4.4: Partial dependences of ZSE on Inflation

Figure 4.4 presents a partial plot demonstrating the marginal effect of normalized inflation on anticipated stock market returns (ZSE) based on the fitted model. This figure also shows that when INF_norm is so large, the value of predicted ZSE decreases gradually, and hence there is a negative relationship between inflation and the stock returns.

4.4.3 Decision Tree

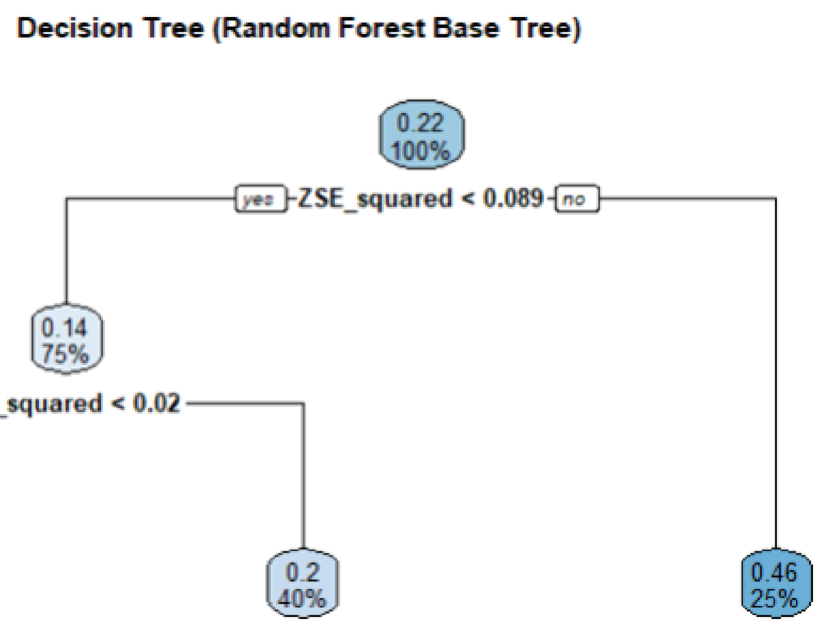

Figure 4.5: Random Forest decision tree (base tree)

Figure 4.5 shows a simple decision tree from a random forest model that starts with a split on the ZSE-squared variable. The tree divides the data into nodes according to threshold values of ZSE_squared, and each terminal node shows the predicted value and the percentage of data it represents. The plot illustrates how the model utilizes the square value of ZSE returns in its predictions, whereby lower values yield lower predicted values, while larger ZSE_squared values correspond to higher predictions.

4.4.4 Predicting using Random Forest Model

Table 4.3 Random Forest Predicted vs Actual values

	Actual	**Predicted**
1	0.1214988	0.1694946
2	0.1854519	0.2103256
3	0.3595533	0.3753811
4	0.1853909	0.2740165
5	0.1091109	0.1484754
6	0.1571368	0.1732491
7	0.068652	0.0974312
8	0.0824434	0.1172568
9	0.3131751	0.3000408

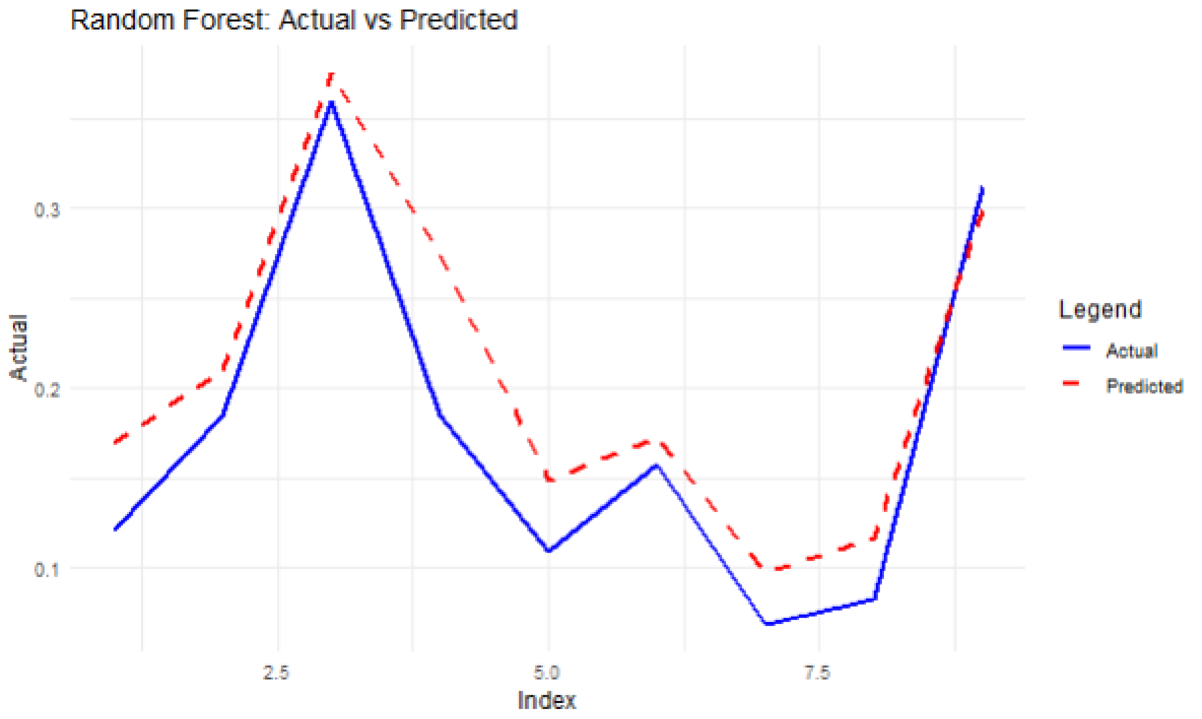

Figure 4.6: Random Forest Prediction vs Actual over time plot

Actual versus predicted stock market returns for the Random Forest model are presented in Figure 4.6 and Table 4.3. The model's predictions closely track actual values throughout the estimation period, out of sample. Figure 4.6, the plot, graphically illustrates the alignment with both predicted and actual values, exhibiting similar trends and turning points. However, interestingly, the model slightly overestimates the peaks and troughs.

4.4.5 Feature Importance

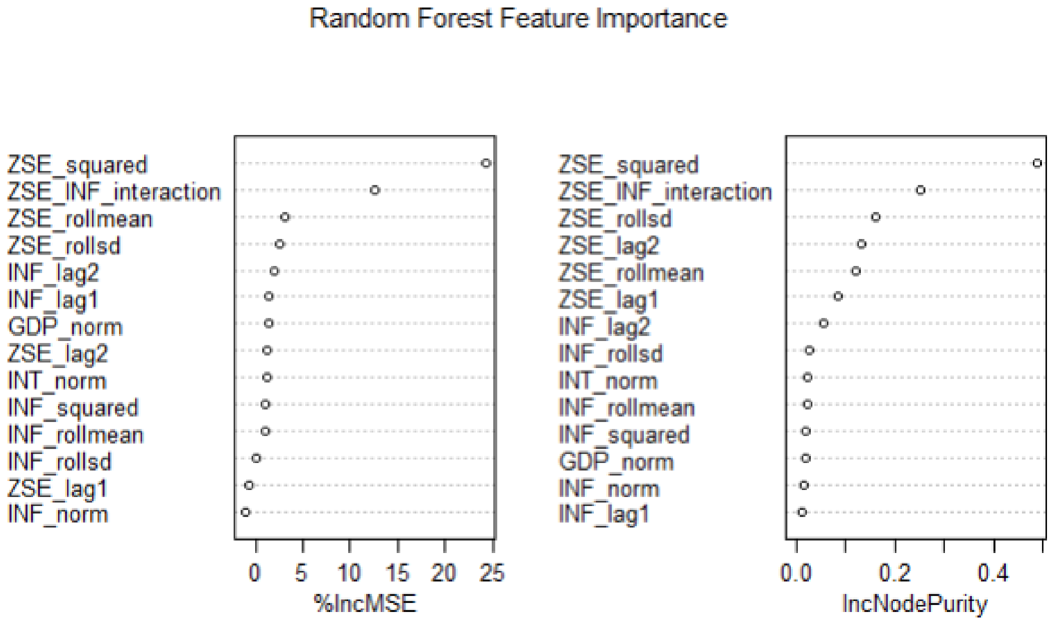

Figure 4.7: Random Forest Feature importance

Figure 4.6, the plot, graphically illustrates the alignment with both predicted and actual values exhibiting similar trends and turning points. But interestingly, the model does overestimate the peaks and troughs slightly. 4.4.5 Feature Importance Figure 4.7: Random Forest Feature Importance Inspection of the feature importance plot in Figure 4.7 shows that ZSE_squared and the ZSE_INF_interaction term are the two most influential predictors when both the %IncMSE and IncNodePurity are taken into consideration. Other attributes like ZSE_rollmean, ZSE_rollsd, and some lagged variables are also extremely important, with variables like INF_norm and INF_lag1 having little to no contribution towards the predictions made by the model. This indicates the model's dependence on nonlinear interactions involving ZSE for making correct predictions.

4.4.6 Random Forest Performance Metrics

Table 4.4: Random Forest Performance Metrics

Metrics	Value
Mean Squared Error (MSE)	0.001672136
Root Mean Squared Error (RMSE)	0.04089176
R-Squared (Coefficient of Determination)	0.9263496
Mean Absolute Percentage Error (MAPE)	26.64389%

4.4.7 Residual Analysis test

Homoscedasticity Test

Residual Plot

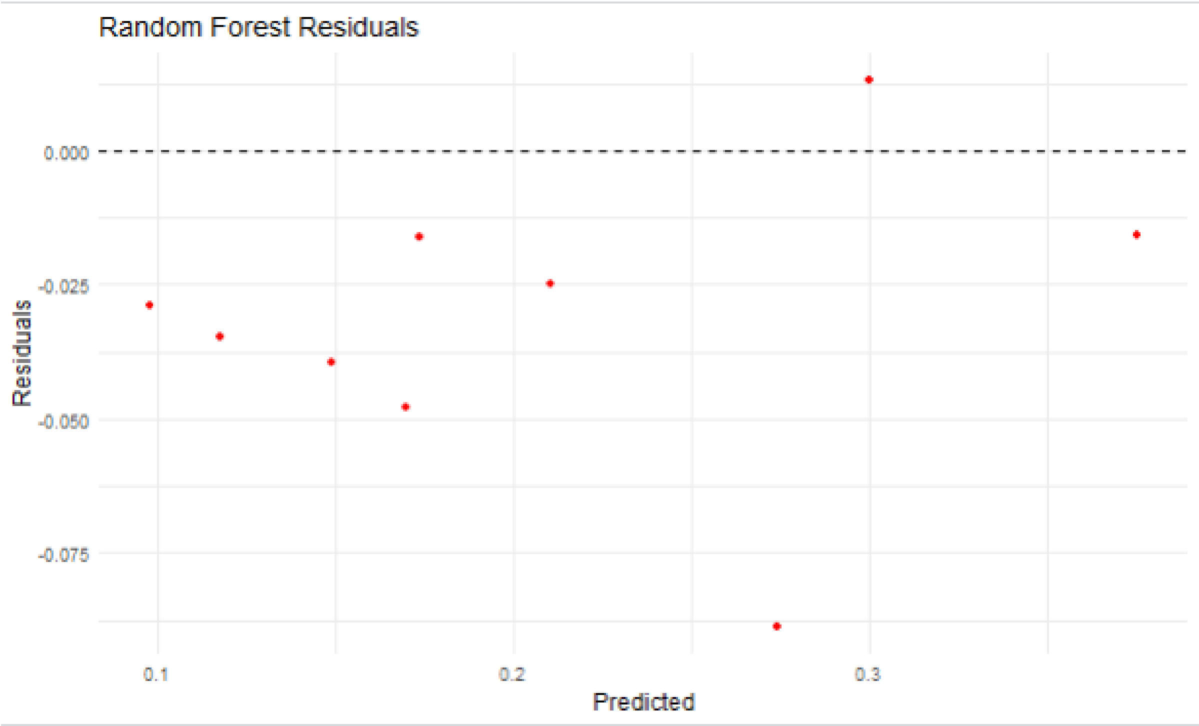

Figure 4.8: Residual plot for constant variance

According to Field (2018), one important requirement of homoscedasticity is that residuals should have equal variance across all levels of the predicted value. Under this context, residuals are scattered randomly around the horizontal line, implying no observed trend. This observation indicates that the residual variance is stable and consistent with the assumption of homoscedasticity.

4.4.8 Model Fitness

Cross Validation and Holdout Method

The models fit well as indicated by the validations. For the case of 48 samples and 14 predictors, 5-fold cross-validation was performed and mtry = 14 was the optimum for the minimum root mean squared error (RMSE = 0.0549), the maximum R-squared value (R-squared = 0.9449) and the minimum mean absolute error MAE = 0.0276. These values indicate that the model fits the data very well, capturing the underlying trends with low error in prediction.

Reliability Test

```
print(cronbach$total$raw_alpha)
[1] 0.6493607
```

The value of Cronbach's alpha is 0.65, as has been reported, with moderate internal consistency or reliability of the scale that was tested. It tells us that the test items are moderately, although not highly, correlated with each other.

4.5.0 XGBoost Model

The XGBoost model was executed in R, and the key highlights were the optimal iteration (55), the optimal score (RMSE of 0.034), the number of boosting rounds executed (65), and information on the training call and parameters. The model utilized 15 features, as described under Appendix C. The eval log tracked performance metrics across 65 iterations, specifying the model's optimal iteration and prediction potential. (Brownlee, 2021).

4.5.1 Partial dependences of ZSE on Inflation

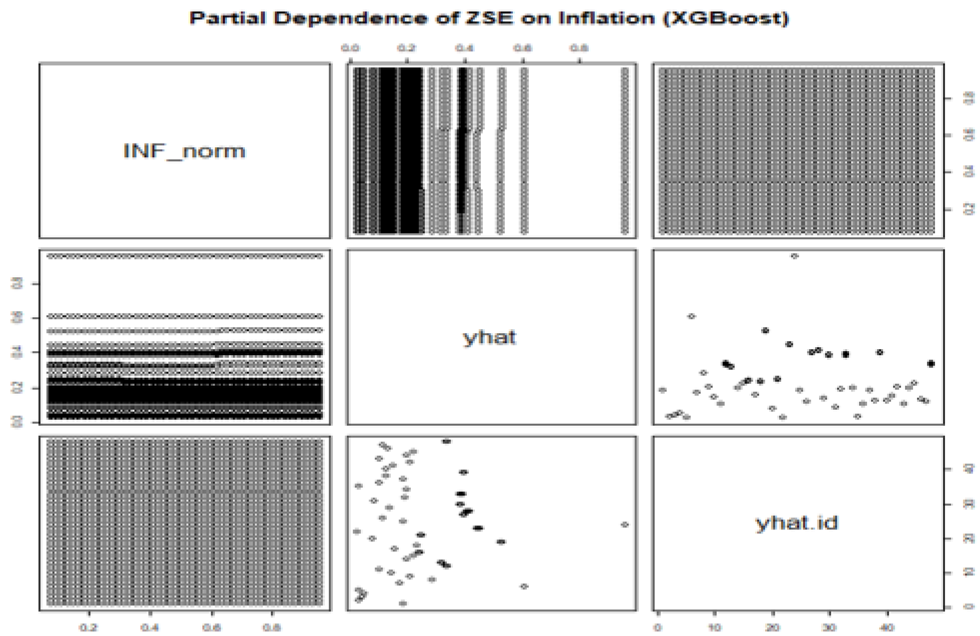

Figure 4.9: Partial dependences on ZSE on Inflation

Figure 4.7 presents a scatterplot matrix of the partial dependencies of normalized stock returns (ZSE) on normalized inflation (INF_norm) as predicted by the XGBoost model. The matrix shows the pairwise relationships among INF_norm, predicted values (yhat), and observation indices (yhat.id). The graph of INF_norm vs. yhat reveals that market returns generally decline, in line with the trend set in the Random Forest model.

4.5.2 Predicting using XGBoost

Table 4.5: XGBoost Actual and predicted values

	Actual	Predicted
1	0.1214988	0.1259377
2	0.1854519	0.1857488
3	0.3595533	0.4500287
4	0.1853909	0.1928058
5	0.1091109	0.1164402
6	0.1571368	0.157021
7	0.068652	0.070545
8	0.0824434	0.0832598
9	0.3131751	0.3100922

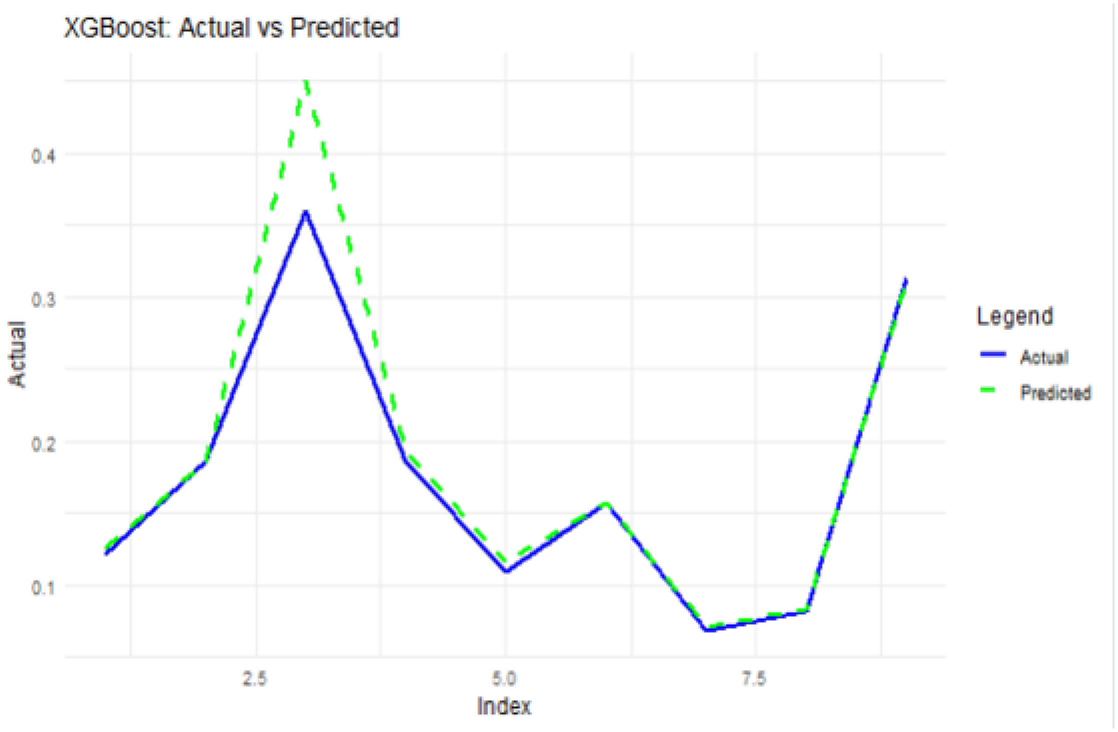

Figure 4.10: XGBoost Actual vs Predicted plot

Table 4.5 and Figure 4.10 show the actual vs learned XGBoost values, both of which follow each other extremely closely. In the table, actual and predicted values are compared for nine observations, and, in most cases, the prediction is very close to the actual value. This closeness is prominently evident in Figure 4.10, where the green dashed line (estimated) follows very closely the blue solid line (measured) in all indices with marginal overestimations approaching the peak.

4.5.3 Feature Importance

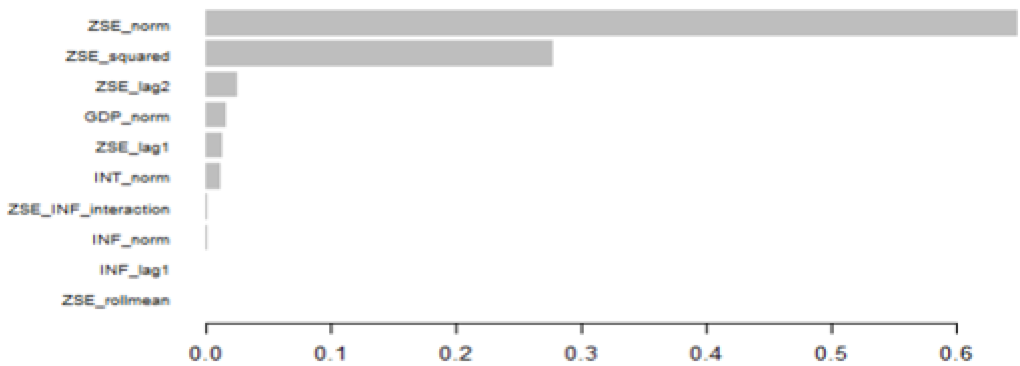

Figure 4.11: XGBoost Model Feature importance

Feature importance from the XGBoost model is plotted in Figure 4.11; the two most important variables are ZSE_norm and ZSE_squared, followed by ZSE_lag2 and GDP_norm with much less importance. Minor contributions to the models' predictions also originate from other features.

4.5.4 XGBoost Performance Metrics

Table 4.6 XGBoost performance metrics

Metrics	**Value**
Mean Squared Error (MSE)	0.0009253394
Root Mean Squared Error (RMSE)	0.03041939
R-Squared (Coefficient of Determination)	0.965248
Mean Absolute Percentage Error (MAPE)	4.944385%

4.5.5 Residual Plot

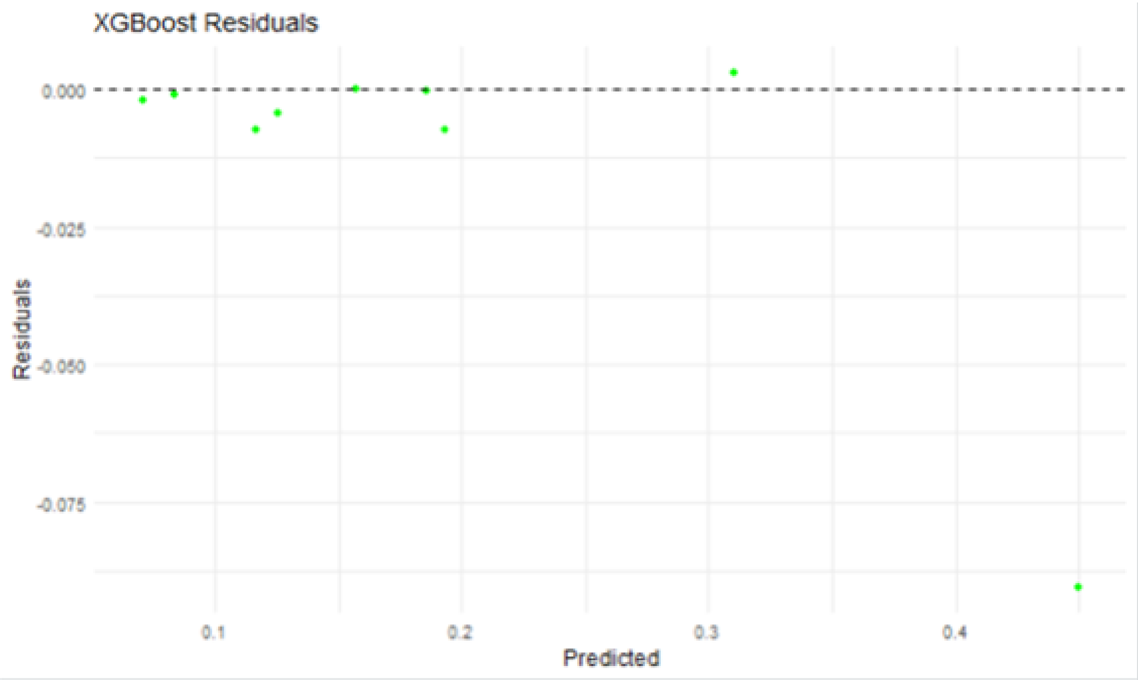

Figure 4.12: XGBoost Residual plot

Figure 4.12 is the residual plot of the XGBoost model and shows that most residuals are near zero, implying that predicted values and original values are closely matched. There is no pattern or systematic bias, and only a single point has a relatively higher negative residual. This implies that the model fits the data well with small errors in predictions.

4.6 Comparison of the models' Performance

Table 4.7: XGBoost and Random Forest Comparisons.

Actual	Random Forest	XGBoost
0.12149875	0.16949462	0.12593767
0.18545188	0.21032564	0.18574879
0.3595533	0.37538105	0.45002869
0.18539086	0.2740165	0.19280578
0.10911088	0.14847535	0.1164402
0.15713675	0.17324914	0.15702099
0.06865198	0.09743118	0.07054497
0.0824434	0.11725675	0.08325979
0.31317508	0.30004083	0.31009218
MSE	0.00167214	0.00093*
RMSE	0.04089176	0.03042*

MAPE	26.64%	4.94%*
R^2	0.9263496	0.96525*

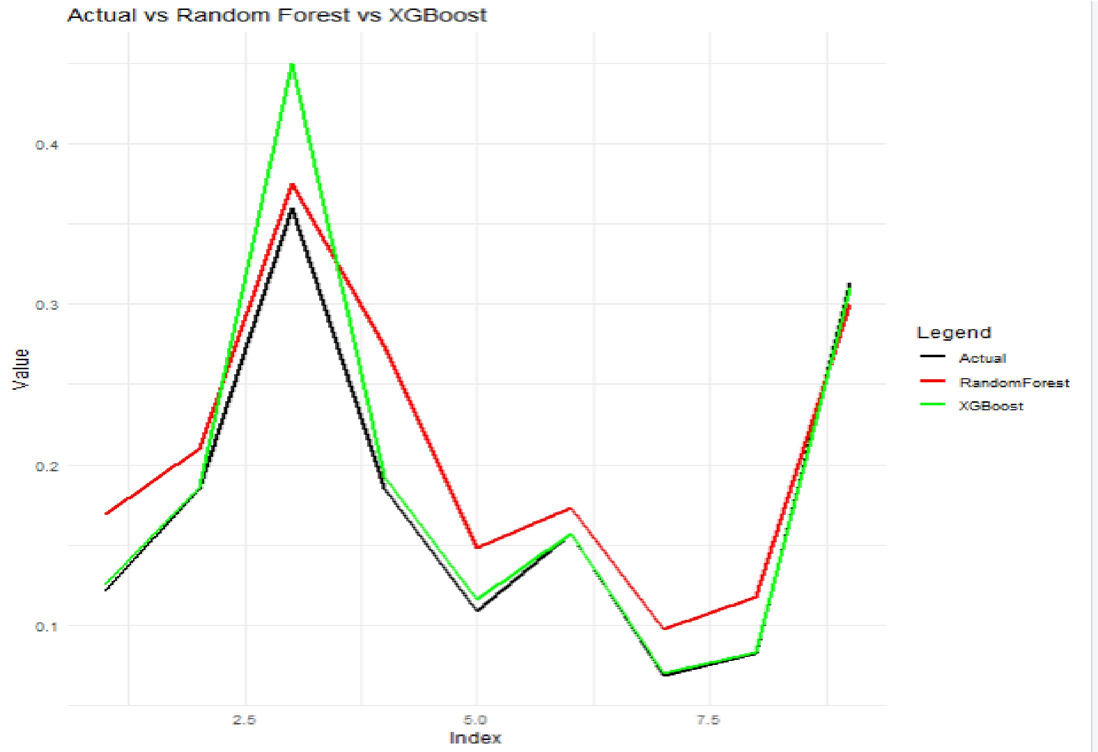

Figure 4.13: Random Forest and XGBoost comparison plot

Table 4.7 and Figure 4.13 illustrate a comparison of the Random Forest and XGBoost models with actual values. Both models both fit real data well; however, XGBoost consistently provides predictions that are closer to real values, as reflected by its lower MSE (0.00093), RMSE (0.03042), and MAPE (4.94%), and higher R^2 (0.96525) than Random Forest (MSE: 0.00167, RMSE: 0.04089, MAPE: 26.64%, R^2: 0.9263).The plot also shows that XGBoost predictions (green line) more closely track actual values (black line), especially at peaks and troughs. While Random Forest (red line) is of higher deviations. XGBoost in general exhibits superior accuracy and prediction consistency in this context analysis.

4.6.1 Discussion of findings

Study had analysed the impact of inflation on stock market returns in Zimbabwe using theoretical considerations and evidence. To answer the first research question, results show that inflation is an effective stock return predictor during the period from 2018 to 2022. This agrees with hypotheses like the Proxy Hypothesis (Fama & Schwert, 1977), which asserts that inflation tends to indicate internal economic turmoil and hence impacts stock performance negatively. The high negative correlation in Zimbabwe is consistent with previous local research (Chikoko, 2015; Mashamba, 2022) that proved inflation erodes purchasing power and measures the performance of shares. This is consistent with what the study by Maqsood et al. (2020) in emerging economies found, reporting similar negative correlations using econometric as well as machine learning techniques.

Note that the above findings contradict the so-called Fisher Hypothesis as well as various Keynesian perspectives, which suggest a positive association between inflation and nominal stock returns at least in moderate-inflation markets (Azar, 2010; Javed et al., 2020). That may be because in Zimbabwe, ongoing hyperinflation and financial instability may damn any potential benefits to an aggregate negative effect on stock returns. This concurs with the Inflation Illusion Hypothesis (Shiller, 1997), which assumes that real returns are misunderstood in times of inflation by the market with a negative effect.

Comparison for the second research objective Comparison To the second research objective, XGboost is found to perform better than the Random Forest model to predict stock market returns in Zimbabwe. Its MSE is 0.00093, RMSE is 0.03042, and R^2 is 0.965, accounting for 96.5% of variance in stock market returns. Mean square error (MSE) and the root-mean-square error (RMSE) are, however, highest for Random Forest (0.00167 and 0.04089, respectively) while the R-squared is lowest (0.926, explaining 92.6% of variance). These findings are interesting, given that literature has reported the high performance of XGBoost in finance forecasting (Wang et al., 2022). Feature importance analysis describes that, although Random Forest does not rely on a wider set of inflation-linked or macroeconomic indicators, the higher predictive accuracy of XGBoost is due to its ability to read and process not just simple linear but also complex nonlinear relationships within the data. Second, its sophisticated boosting and regularization techniques render it more accurate especially in today's dynamic economic environment in Zimbabwe. Figure 4.13 and Table 4.7 demonstrate that the XGBoost model can be used to predict the S & P 500 daily return (green line) very closely to the true S & P 500 index daily return (black line), especially during extreme market volatility. Random Forest (red line) is, on the other hand, dispersed even further at the extremes of the market. With regard to the second objective of this study in modelling the relationship between inflation and stock market returns in Zimbabwe, XGBoost has proved to be highly accurate and stable.

Furthermore, the findings emphasize the shortcomings of linear models with standard assumptions and corroborate the growing appreciation in literature that ML models, specifically ensemble-based techniques such as XGBoost and Random Forest, are highly capable of handling non-linearities and interactions in financial time series, especially with high-volatility and structural breaks applications (Ferreira da Costa et al., 2019; Hakizimana, 2018). This result fills a significant lacuna in the literature, as part of the research has taken specific advantage of more advanced machine learning models in elucidating the correlation between inflation and stock returns in Zimbabwe.

4.7 Chapter Summary

This paper focuses on the causality inflation-stock returns relationship in Zimbabwe between 2018 and 2022. The negative effect of high inflation levels on economic stability and investor attitude is presented by descriptive analysis. Integrity of data was verified based on diagnostic tests and the necessity of obtaining machine learning models such as Random Forest and XGBoost to unveil complex relationships was validated. Despite economic uncertainty, the results are a strong foundation for forecasting future stock returns. The role of inflation as a predictor leading variable is clear, for the first time, from this empirical research on the Zimbabwe Stock Exchange. Such results can guide policy makers and investors to make reasonable decisions in an ambiguity economic setting. Chapter 5 detailed these results and provided recommendations for future research.

CHAPTER 5: FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

5.0 Introduction

The chapter summarizes the study's general findings on the significance of inflation as a predictor of stock returns in Zimbabwe for the period 2018–2022. The chapter responds to the objectives and findings of the study, concluding with implementable recommendations for stakeholders. The chapter ultimately suggests a research agenda.

5.1 Summary of Findings

The research explored the correlation between inflation and stock market returns, through an evaluation of the Random Forest and XGBoost models' ability to forecast in Zimbabwe. The results showed a causal interdependence between stock market returns and inflation, confirming inflation as a main macroeconomic driver that causes volatility on the ZSE. Inflation witnessed erratic volatility throughout the research period, varying between a minimum of 10.62% and a maximum of 557.29%, averaging at 216.23%, in line with the hyperinflationary nature of the country. Volatility negatively affected investor sentiment, profitability of the companies, and the general performance of the market.

When it came to the efficiency of the model, XGBoost outperformed Random Forest. But it was paired with better R ^2 (0.965), lower Root Mean Squared Error (0.0304) and Mean Absolute Percentage Error (4.94%) than the Random Forest Model. Random Forest performance was 0.965 in R-squared, 0.0409 in RMSE and 26.64% in MAPE. The improved performance of the XGBoost model is in the form of the boosting framework, regularization, and the ability to learn nonlinear relationships and feature interactions (Chen & Guestrin, 2016; Brownlee, 2021). These findings show the significant role of inflation in stock returns and the application of the XGBoost technique (Johnson, 2019).

5.2 Conclusions

The study concentrated on analysing the impact of inflation on stock returns in Zimbabwe and the ability of two machine learning algorithms, Random Forest and XGBoost, to detect the relationship. The research succeeded in proving inflation to be a predictor of stock returns in the stock market in Zimbabwe and analysed the performance of different machine learning models. The empirical findings confirmed that inflation plays a significant role in stock returns in Zimbabwe, taking into consideration the investors' confidence and market behaviour. This aligns with existing research demonstrating the destruction that hyperinflation has caused to financial stability (Mudehwe., 2021; Marekera, 2023). Notice that from what was obtained through the analysis, the XGBoost model was better in its capacity to extract the complicated non-linear relationship that exists in the data compared to the Random Forest model, this resonates with the reality that boosting algorithms at regularization and feature management perform better compared to conventional ensemble methods (Chen & Guestrin, 2016).

The results highlighted that inflation has a great impact on purchasing power, borrowing costs, and business profits. In times of economic turbulence, inflation also leads to speculation in the ZSE, which translates to market volatility (Zhang, 2020). These findings correlate with general economic principles that associate inflation with financial market performance (Hyndman & Athanasopoulos, 2018). Findings also assist policymakers in coming up with mechanisms for more efficient decision-making during the era of unstable economic environment.

5.3 Recommendations

5.3.1 For Policymakers

The implications are that policymakers should strive to stabilize inflation. This can be achieved by fine-tuning monetary policy to reduce market volatility and instil confidence among investors (Makridakis, Spiliotis, & Assimakopoulos, 2020). Secondly, it is necessary to enhance data collection and transparency regarding economic indicators to improve predictive modelling and decision-making (Hyndman & Athanasopoulos, 2018).

5.3.2 For Investors

Investors should utilize forecasting models like XGBoost to forecast market trends and make informed investment decisions based on data. Investors ought to also diversify their investments as a hedge against inflation-driven volatility. (Johnson, 2019).

5.3.3 For Model Developers

Modelers must emphasize various features and interaction terms in predictive modelling for optimal performance. Additionally, developers must conduct extensive hyperparameter tuning and cross-validation to fine-tune the model's outputs (Chen & Guestrin, 2016).

5.4 Areas for further research

Further research should investigate exchange rate volatility, commodity prices, and foreign direct investment to gain insight into economic dynamics (Smith, 2021). Moreover, predictive long-term correlations between inflation and stock prices may be achieved by simulating sophisticated time-series models like ARIMA or LSTM (Hyndman & Athanasopoulos, 2018). Comparative research with other nations experiencing hyperinflation may provide insight into the generalizability of the study's findings by the researcher (Mudehwe, 2021). Additionally, the effect of reduced inflation on stock markets through certain policies is a matter of research concern (Doung et al., 2022). Lastly, broadening the scope of machine learning models to include ensemble approaches and hybrid models (Random Forest, XGBoost, and neural networks) shows promise for enhancing predictive performance (Makridakis et al., 2020).

5.5 Conclusion

This section presents the study's results, providing evidence that inflation is a statistically significant predictor of stock returns in Zimbabwe. The XGBoost model best predicted these outcomes, explaining that MMV reactions are influenced by several factors and their interactions. It is recommended that policymakers, investors, and researchers stabilize inflation by consolidating predictive models using economic indicators. The final section of the chapter presents suggestions for future research that could further extend the findings and clarify economic interactions in risky markets.

REFERENCES

1. Abdullahi, A.M., Htet, S.A., Naing, N.N.N., Hossain, M.S., Ismail, S. and Zaaba, M.A.M., 2023, December. A Prediction of Customer Segmentation using Random Forest in RapidMiner. In *2023 IEEE 21st Student Conference on Research and Development (SCOReD)* (pp. 33-38). IEEE.
2. African Development Bank. (2020). African Economic Outlook 2020.
3. African Development Bank. (2023). *African Development Data Portal*. [Online]. Available at: *https://www.afdb.org/en/knowledge/statistics* (Accessed: 15 November 2024).
4. Azar, S.A. (2010). Inflation and stock returns. International Journal of Accounting and Finance, 2(3-4), pp.254–274.Fama E. F (1981). Stock Returns, Real Activity, Inflation and Money, Am. Econ. Rev. 71 (4): 545–565.
5. Bakar, N. A., & Ahmad, M. (2019). "The Impact of Macroeconomic Variables on Stock Market Performance: Evidence from Malaysia." International Journal of Economics and Financial Issues, 9(3), 1-10.
6. Boddie, Z., Kane, A., & Marcus, A. J. (2019). Investments. McGraw-Hill Education.
7. Bodie, Z. (2020). Investments. McGraw-Hill.
8. Boulton, A.J. and Williford, A., 2018. Analysing skewed continuous outcomes with many zeros: A tutorial for social work and youth prevention science researchers. *Journal of the Society for Social Work and Research*, *9*(4), pp.721-740.
9. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), pp. 5–32.
10. Brownlee, J. (2021). Machine Learning Mastery with Python: Discover the Fastest Growing Platform For Professional Machine Learning, Machine Learning Mastery.
11. Brunner, A.D. & Meltzer, A.H. (1976). The implications of monetarism for the theory of inflation. *Journal of Money, Credit and Banking*, 8(1), 1-16.
12. Bryman, A. (2016). *Social Research Methods*. 5th ed. Oxford University Press.
13. Chen, T. & Guestrin, C. (2016). 'XGBoost: A Scalable Tree Boosting System', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Digital Library, pp. 785–794. Available at: https://dl.acm.org/doi/10.1145/2939672.2939785 (Accessed: 13 April 2025).
14. Chen, Y., Zhang, L., and Wang, J. (2019). "Inflation expectations and stock market return: Evidence from China", International Review of Financial Analysis, 63, pp. 1-10.

15. Chikoko, T. (2021). The Impact of Economic Indicators on Stock Market Performance in Zimbabwe. *Journal of Economics and Finance*, 12(3), pp. 45–59.
16. Chikoko, V. (2015). The Impact of Inflation on Stock Market Returns in Zimbabwe. Journal of Economics and Finance, 6(2), pp. 45–56.
17. Djaballah, S., Saidi, L., Meftah, K., Hechifa, A., Bajaj, M. and Zaitsev, I., 2024. A hybrid LSTM random forest model with grey wolf optimization for enhanced detection of multiple bearing faults. *Scientific Reports*, *14*(1), p.23997.
18. Duong, T.H., 2022. Inflation targeting and economic performance over the crisis: evidence from emerging market economies. *Asian Journal of Economics and Banking*, *6*(3), pp.337-352.
19. Effiong, U.E., Ukere, I.J. and Ekpenyong, E.A., 2025. Foreign Exchange Market and Manufacturing Sector Performance in Nigeria. *British Journal of Multidisciplinary and Advanced Studies*, *6*(1), pp.50-7
20. Fama, E. F. and Schwert, G. W. (1997). Asset return and inflation. Journal of Financial Economics, 5. 115- 46
21. Ferreira da Costa, E.M., Martins Noriller, R., Farias de Souza Nogueira, M.A. and Pires da Silva, J., 2022. INTERFERENCE OF MACROECONOMY IN THE PERFORMANCE OF STOCK MARKETS IN LATIN AMERICA AND ASIA. *RACE-Revista de Administraçâo, Contabilidade e Economia*, *21*(2).
22. Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. 5th ed. Sage Publications.
23. Fisher, I. (1930). *The Theory of Interest*. New York: Macmillan.
24. Gordon, M.J., 1959. Dividends, earnings, and stock prices. *The review of economics and statistics*, pp.99-105.
25. Gu, X., Li, Y. & Wu, L. (2023). Prediction of Water Carbon Fluxes and Emission Causes in Rice Paddies Using Two Tree-Based Ensemble Algorithms. *Sustainability*, 15(16), 12333.
26. Guba, E.G. (1990). The Paradigm Dialog. *Sage Publications*.
27. Gur, Y.E., 2024. Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. *Data Science in Finance and Economics*, *4*(4), pp.469-513.
28. Hakizimana, C., 2022. *Forecasting inflation of Rwanda using macroeconomics variables (2006Q1-2019Q4)* (Doctoral dissertation, University of Rwanda).

29. Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer.
30. https://databank.worldbank.org/source/world-development-indicators
31. https://www.zse.co.zw/market-data/
32. Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. 2nd ed. OTexts.
33. International Labour Organisation. (2020). Consumer Price Index. Retrieved from (Yuen, K. and Rozenbes, D., 2022. Experimental estimates of a Consumer Price Index for low-paid employee households.)
34. Jardon, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.
35. Javed, A., et al. (2010). Inflation and Stock Market Returns: Evidence from Pakistan. International Journal of Business and Management, 5(5), pp. 120-130.
36. John. E. I. (2021). Effect of selected macroeconomic variables on stock market performance: A comparative study of Nigeria, South Africa and Ghana: 1986- 2018.
37. Johnson, M. (2019). *Data Transformation Techniques in Economic Research*. Chicago: University Press.
38. Johnson, M. (2020). *Statistical Analysis of Economic Indicators*. New York: Academic Press.
39. Joshi M. (2015). Financial Markets and Institutions 11th Edition. Cengage Learning.
40. Kang, L., Hu, G., Huang, H., Lu, W. & Liu, L. (2022). Urban Traffic Travel Time Short-Term Prediction Model Based on Spatio-Temporal Feature Extraction. *Journal of Advanced Transportation*, 2020, 1–16.
41. Khalid, S., Khan, M.A., Mazliham, M.S.U., Alam, M.M., Aman, N., Taj, M.T., Zaka, R. and Jehangir, M., 2022. Predicting risk through artificial intelligence based on machine learning algorithms: a case of Pakistani nonfinancial firms. *Complexity*, *2022*(1), p.6858916.
42. Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 1137-1145.
43. Kothari, C.R. (2019). *Research Methodology: Methods and Techniques*. 4th ed. New Age International.
44. Lee, B.S., 2010. Stock returns and inflation revisited: An evaluation of the inflation illusion hypothesis. Journal of Banking & Finance, 34(6), pp.1257-1273.

45. Lee, T. (2020). *Advanced Statistical Methods for Economic Data Analysis*. Boston: Analytics Publishing.
46. Liu, J., Wu, J., Liu, S., Li, M., Hu, K. and Li, K., 2021. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *Plos one*, *16*(2), p.e0246306.
47. Magweva, R. and Mbudaya, J., 2021. Inflation hegding and listed stock in Zimbabwe-a non-linear approach. *African Journal of Business and Economic Research*, *16*(1), p.123.
48. Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2018. The M4 Competition: Results, findings, conclusion and way forward. *International Journal of forecasting*, *34*(4), pp.802-808.
49. Mankiw, G. (2017). Principles of economics. Cengage Learning.
50. Marekera, T., 2023. *An Assessment of the influence of technological collaboration on competitiveness of small to medium enterprises: a case of Speroni Irrigation* (Doctoral dissertation, BUSE).
51. Mashamba, T. and Magweva, R., 2022. Inflation risk management using listed stocks in Zimbabwe. *Journal of Contemporary Management*, *19*(2), pp.400-419.
52. Mbulawa S, (2015). The impact of Inflation on Stock market performance in Zimbabwe between 1980 and 2008: An empirical investigation, European Journal of Business Economics and Accountancy Vol.3, No. 4, 2015.
53. McCarthy, J., 2020. How Evolution, Stories, and Irrationality Influence Decision Making in Financial Markets: Analyzing Whether We Can Leverage Our Innate Traits and Heuristics to Improve Outcomes (Doctoral dissertation, Duke University).
54. Mogire, G., 2016. The Effect of Inflation on Stock Market Returns at the Nairobi Securities Exchange (Doctoral dissertation, University of Nairobi).
55. Mpofu, B., 2010. The relationship between stock market returns and inflation: New evidence from Sub-Saharan Africa (Doctoral dissertation, University of St Andrews).
56. Mudehwe, T., 2021. *The impact of foreign direct investment on domestic investment in Zimbabwe* (Doctoral dissertation).
57. O'Brien, R.M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5), pp. 673-690.
58. Petrova, E., Müller, D. and Johnson, S., 2024. INFLUENCE OF MONETARY POLICY ON STOCK MARKET DYNAMICS. *Ayden International Journal of Banking, Finance and Technology*, *12*(3), pp.19-35.

59. Rakshit, B. and Neog, Y., 2022. Effects of the COVID-19 pandemic on stock market returns and volatilities: evidence from selected emerging economies. *Studies in Economics and Finance*, *39*(4), pp.549-571.
60. Ren, Y., Ling, F. and Wang, Y., 2023. Research on provincial-level soil moisture prediction based on extreme gradient boosting model. *Agriculture*, *13*(5), p.927.
61. Shah, J., Vaidya, D. and Shah, M., 2022. A comprehensive review on multiple hybrid deep learning approaches for stock prediction. *Intelligent Systems with Applications*, *16*, p.200111.
62. Shiller, R.J., 1997. Why do people dislike inflation? In *Reducing inflation: Motivation and strategy* (pp. 13-70). University of Chicago Press.
63. Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289-310.
64. Sibanda, B., 2013. An Assessment of the Impact of Inflation on Banking Sector Performance in Zimbabwe (2000–2008) and (2009–2012): The Case of CBZ Bank Limited (Doctoral dissertation, University of Zimbabwe).
65. Smith, A. (2021). *Statistical Methods for Non-Normal Data*. London: Statistical Publishing.
66. Tavakol, M. and Dennick, R., 2011. Making sense of Cronbach's alpha. *International journal of medical education*, *2*, p.53.
67. Wang, J., Rong, W., Zhang, Z. & Mei, D. (2022). Credit Debt Default Risk Assessment Based on the XGBoost Algorithm: An Empirical Study from China. *Wireless Communications and Mobile Computing*, 2022, 1–14.
68. Wooldridge, J.M. (2016). *Introductory Econometrics: A Modern Approach*. 6th ed. Cengage Learning.
69. Zhang, J. (2020). *Machine Learning in Economic Analysis*. New York: Academic Press.
70. Zhao, H., Wang, Y., Li, X., Guo, P. & Lin, H. (2023). Prediction of Maximum Tunnel Uplift Caused by Overlying Excavation Using XGBoost Algorithm with Bayesian Optimization. *Applied Sciences*, 13(17), 9726.

APPENDICES (A-C)

Appendix A: Transformed Data

	Date	GDP	INT	INF	ZSE
1	31/1/2018	2.62661	2.876459	2.36263	2.99873
2	28/2/2018	2.555802	2.876861	3.43429	3.23357
3	31/3/2018	2.479595	2.877263	3.93964	3.43785
4	30/4/2018	2.397099	2.877665	4.27375	1.78002
5	31/5/2018	2.307179	2.878066	4.52375	1.97269
6	30/6/2018	2.208368	2.878468	4.72359	2.21047
7	31/7/2018	2.098712	2.878869	4.89006	1.52606
8	31/8/2018	1.975534	2.87927	5.03274	4.62859
9	30/9/2018	1.835023	2.879671	5.15758	3.37177
10	31/10/2018	1.671489	2.880072	5.26855	3.85947
11	30/11/2018	1.475881	2.880473	5.36842	3.55421
12	31/12/2018	1.232472	2.880873	5.45922	3.1999
13	31/1/2019	0.910073	2.881273	5.54246	2.87582
14	28/2/2019	0.858999	2.799346	5.63644	4.01422
15	31/3/2019	0.805175	2.710102	5.72235	3.95738
16	30/4/2019	0.748288	2.612107	5.80145	3.04023
17	31/5/2019	0.687969	2.503454	5.87475	3.50856
18	30/6/2019	0.623776	2.381541	5.94305	3.62327
19	31/7/2019	0.555179	2.242674	6.00698	3.69635
20	31/8/2019	0.481527	2.081365	6.06706	3.28579
21	30/9/2019	0.402016	1.888935	6.12374	3.44649
22	31/10/2019	0.315632	1.650434	6.17738	3.6561
23	30/11/2019	0.221075	1.336621	6.22829	4.09301
24	31/12/2019	0.116635	0.876752	6.27673	4.48379
25	31/1/2020	0	0	6.32293	3.44617
26	29/2/2020	0.857424	1.63147	6.25187	2.60195
27	31/3/2020	1.312153	2.221675	6.17537	3.71989
28	30/4/2020	1.623586	2.590328	6.09253	0

29	31/5/2020	1.860714	2.859083	6.0022	4.31629
30	30/6/2020	2.052251	3.070696	5.90289	5.10516
31	31/7/2020	2.212932	3.245256	5.79263	3.41576
32	31/8/2020	2.351334	3.393826	5.66868	2.98366
33	30/9/2020	2.472889	3.523151	5.52717	4.18144
34	31/10/2020	2.581256	3.63765	5.36228	4.21995
35	30/11/2020	2.67902	3.740376	5.16472	2.9381
36	31/12/2020	2.768072	3.833525	4.91828	3.28653
37	31/1/2021	2.849837	3.918732	4.59053	3.15658
38	28/2/2021	2.840413	3.911053	4.59572	4.14488
39	31/3/2021	2.830899	3.903314	4.60089	2.50553
40	30/4/2021	2.821293	3.895514	4.60603	2.68989
41	31/5/2021	2.811594	3.887654	4.61114	3.47228
42	30/6/2021	2.801801	3.879731	4.61623	4.15904
43	31/7/2021	2.79191	3.871745	4.6213	3.50345
44	31/8/2021	2.781921	3.863695	4.62633	1.70111
45	30/9/2021	2.771831	3.855579	4.63135	2.85359
46	31/10/2021	2.761638	3.847397	4.63633	3.43624
47	30/11/2021	2.75134	3.839148	4.6413	2.67484
48	31/12/2021	2.740934	3.830829	4.64624	3.05636
49	31/1/2022	2.73042	3.822441	4.65115	4.18419
50	28/2/2022	2.719794	3.813983	4.65604	3.06619
51	31/3/2022	2.709053	3.805451	4.66091	3.25694
52	30/4/2022	2.698196	3.796847	4.66575	3.56417
53	31/5/2022	2.68722	3.788168	4.67057	2.86904
54	30/6/2022	2.676122	3.779413	4.67536	3.49832
55	31/7/2022	2.6649	3.77058	4.68014	3.60631
56	31/8/2022	2.65355	3.761669	4.68489	3.09784
57	30/9/2022	2.642069	3.752678	4.68961	2.98265
58	31/10/2022	2.630456	3.743605	4.69432	4.01422
59	30/11/2022	2.618706	3.734449	4.699	3.95738

Appendix B: Table 4.6: Feature Engineering

TRAINING TEST						
INF	**ZSE**	**INF_squared**	**INF_log**	**INF_lag1**	**INF_rollmean**	**INF_rollsd**
-0.046786	32.04	-0.300569	0.31125795	-0.66842	-0.87469053	-0.444678
-0.977261	7.78	-0.70228625	-1.3755774	-0.07531	-0.88948279	-0.439693
-0.665065	10.56	-0.62037633	-0.5371175	-1.01185	-1.19160303	-0.941135
-0.639279	1.88	-0.6112235	-0.4875721	-0.69762	-1.03076648	-1.246971
-0.661842	9.89	-0.61925214	-0.5308164	-0.67167	-0.95182111	-1.127956
-0.510244	16.75	-0.55994998	-0.2653302	-0.69438	-0.94195962	-1.134121
-0.61994	1.19	-0.60411986	-0.4516543	-0.54179	0.51691956	1.5338551
2.199189	118.04	2.62338928	1.5976428	-0.6522	1.7761019	1.4742064
1.959158	100.56	2.17900606	1.5085031	2.185327	2.96882438	-0.849957
1.7191274	115.6	1.76618796	1.4125547	1.943729	1.4797006	1.3338419
-0.721156	16.15	-0.63902749	-0.6521485	1.702132	0.14979081	1.1677676
-0.648949	8.83	-0.61469851	-0.5059236	-0.75408	0.02739575	0.9243919
1.4790964	119.97	1.38493497	1.30867097	-0.6814	0.11309686	0.8399175
-0.553086	5.79	-0.57798492	-0.3348489	1.460535	1.07582148	0.6793132
1.2390655	131.57	1.03524709	1.19541994	-0.58491	0.21905606	0.396551
-0.20112	20.46	-0.40001279	0.14698482	1.218937	0.19856056	0.4234082
-0.59328	25.45	-0.59399128	-0.4037654	-0.23065	-0.76252046	-0.8447893
-0.645726	2.09	-0.61354587	-0.4997766	-0.62537	-1.00190386	-1.2023267
-0.670579	18.74	-0.62228595	-0.5479687	-0.67815	-1.0220726	-1.2379481
-0.632833	11.97	-0.6088784	-0.4754852	-0.70317	-0.05478936	0.6588067
1.2512288	16.96	1.05220788	1.20141102	-0.66518	-0.03389873	0.6377463
-0.62961	5.72	-0.60769731	-0.469485	1.23118	0.58253416	0.4022813
0.5760635	10.06	0.23336252	0.81620895	-0.66193	-0.23709843	-0.1611131
-0.356166	1.19	-0.48677564	-0.0429784	0.551609	-0.2288805	-0.172089
-0.613493	15.23	-0.60170644	-0.4399063	-0.3867	-0.800222	-1.0376751
-0.544404	24.1	-0.57441122	-0.3204462	-0.64571	-0.77945686	-0.9975735
-0.315443	5.6	-0.46526266	0.00976151	-0.57617	-0.76024313	-1.0219582
-0.575813	17.45	-0.58714434	-0.3733518	-0.34572	-0.79711663	-0.9845557
-0.616717	3.6	-0.60291601	-0.4457667	-0.60778	-0.97210677	-1.1989145

-0.658619	17.86	-0.61812228	-0.5245469	-0.64896	-0.98967691	-1.2255344
-0.61027	12.54	-0.60049118	-0.4340732	-0.69113	0.28506611	1.2823877
1.8832024	36.58	2.04495918	1.47892576	-0.64247	1.42004972	1.1134651
1.5672156	8.05	1.52123205	1.34781301	1.867278	1.41183179	1.1297018
-0.626387	9.72	-0.60651051	-0.4635133	1.549229	1.49239479	1.2286434
2.0411958	43.78	2.32733636	1.53968562	-0.65869	0.39081899	1.417395
-0.593103	14.22	-0.59392265	-0.4034531	2.026302	0.14660705	1.6944532
-1.105315	2.03	-0.72044017	-1.9949947	-0.62519	-1.2267784	-0.7793261
-0.652173	7.73	-0.61584547	-0.5121008	-1.14074	-1.21328056	-0.7630618
-0.566632	8.61	-0.5834782	-0.3576552	-0.68464	-0.50742933	-0.3674659
0.2789418	42.12	-0.04785329	0.600316	-0.59854	0.0897537	-0.2698039
0.5189727	115.65	0.17557414	0.77750357	0.252548	0.88810838	-0.6275014
0.9990346	141.53	0.71712433	1.07094337	0.494145	1.46445384	-0.4888365
1.4092222	3.39	1.27988209	1.27674266	0.97734	0.57091158	1.2268538
TEST SET						
INF	**ZSE**	**INF_squared**	**INF_log**	**INF_lag1**	**INF_rollmean**	**INF_rollsd**
1.3483411	14.12	1.1903443	1.3878049	0.311807	1.4649497	0.717485
2.0893268	15.37	2.3459444	1.7794582	1.235773	1.3571376	1.005513
0.112151	44.69	-0.1375727	0.4182927	1.954215	0.6723271	1.6504191
-0.450719	17.97	-0.4936335	-0.321163	0.037192	-0.3987898	-0.0956028
-0.724595	11.09	-0.6106287	-0.848355	-0.50855	-0.7144194	-0.6163233
-0.717037	8.87	-0.607894	-0.831343	-0.7741	-0.8100389	-0.8942491
-0.70192	5.99	-0.6023404	-0.797827	-0.76677	-0.7956532	-0.8873776
-0.686803	5.45	-0.5966746	-0.764969	-0.75211	-0.7755132	-0.8798641

Appendix C: Syntax of R Codes used to run the models.

C.1: Introduction

This appendix outlines the R code used in this study to examine how inflation predicts stock market returns in Zimbabwe between 2018 and 2022. It encompasses data preprocessing, model training, and evaluation for both the Random Forest and XGBoost algorithms. This section specifically addresses the research objectives: Objective 1, which explores the connection between inflation and stock market returns, and Objective 2, which assesses the effectiveness of the two machine learning models in capturing this link. By detailing the syntax and methodology, this appendix acts as a practical resource for replicating the analysis and understanding the techniques employed in the study.

C.2 : R Codes

```
# Libraries

libs <- c("tidyverse", "caret", "randomForest", "xgboost", "Metrics", "data.table",

        "ggplot2", "zoo", "reshape2", "psych", "rpart", "rpart.plot")

new_libs <- libs[!(libs %in% installed.packages()[,"Package"])]

if(length(new_libs)) install.packages(new_libs)

lapply(libs, require, character.only = TRUE)

## Load Data

Data <- read.csv("data.csv")

str(Data)

## 3. Normalization (Min-Max, All Positive)

# Apply min-max normalization to all relevant columns

norm_cols <- c("GDP", "INT", "INF", "ZSE")

for (col in norm_cols) {
```

```
  Data[[paste0(col, "_norm")]] <- (Data[[col]] - min(Data[[col]], na.rm=TRUE)) /
                                  (max(Data[[col]], na.rm=TRUE) - min(Data[[col]], na.rm=TRUE))
}
```

4. Feature Engineering

```
Data <- Data %>%
  arrange(row_number()) %>%
  mutate(
    INF_lag1 = lag(INF_norm, 1),
    INF_lag2 = lag(INF_norm, 2),
    ZSE_lag1 = lag(ZSE_norm, 1),
    ZSE_lag2 = lag(ZSE_norm, 2),
    INF_squared = INF_norm^2,
    ZSE_squared = ZSE_norm^2,
    INF_rollmean = zoo::rollmean(INF_norm, 3, fill=NA, align="right"),
    INF_rollsd = zoo::rollapply(INF_norm, 3, sd, fill=NA, align="right"),
    ZSE_rollmean = zoo::rollmean(ZSE_norm, 3, fill=NA, align="right"),
    ZSE_rollsd = zoo::rollapply(ZSE_norm, 3, sd, fill=NA, align="right"),
    ZSE_INF_interaction = ZSE_norm * INF_norm
  ) %>%
  drop_na()
```

5. Data Splitting (Train/Test 80/20)

```
set.seed(123)

split_idx <- createDataPartition(Data$ZSE_norm, p=0.8, list=FALSE)

train_data <- Data[split_idx, ]

test_data <- Data[-split_idx, ]
```

6. Random Forest Model Development and Training

```
features <- c("GDP_norm", "INT_norm", "INF_norm", "ZSE_norm",

         "INF_lag1", "INF_lag2", "ZSE_lag1", "ZSE_lag2",

         "INF_squared", "ZSE_squared", "INF_rollmean", "INF_rollsd",

         "ZSE_rollmean", "ZSE_rollsd", "ZSE_INF_interaction")

rf_formula <- as.formula(paste("ZSE_norm ~", paste(features[features != "ZSE_norm"],
collapse = "+")))

set.seed(123)

rf_model <- randomForest(rf_formula, data=train_data, importance=TRUE, ntree=500)

print(rf_model)
```

7. Decision Tree (First Tree Visual)

```
rpart_model <- rpart(rf_formula, data=train_data, method="anova")

rpart.plot(rpart_model, main="Decision Tree (Random Forest Base Tree)")
```

8. Random Forest Prediction & Table

```
rf_pred <- predict(rf_model, test_data)

rf_results <- data.frame(
```

```
  Actual = test_data$ZSE_norm,
  Predicted = rf_pred,
  Index = 1:nrow(test_data)
)
print(rf_results)
```

9. Plot: Actual vs Predicted (Random Forest)

```
ggplot(rf_results, aes(x=Index)) +
  geom_line(aes(y=Actual, color="Actual"), size=1) +
  geom_line(aes(y=Predicted, color="Predicted"), size=1, linetype="dashed") +
  scale_color_manual(values=c("Actual"="blue", "Predicted"="red")) +
  labs(title="Random Forest: Actual vs Predicted", color="Legend") +
  theme_minimal()
```

10. Random Forest Feature Importance

```
varImpPlot(rf_model, main="Random Forest Feature Importance")
```

11. Performance Metrics (Random Forest)

```
rf_mse   <- mse(test_data$ZSE_norm, rf_pred)
rf_rmse  <- rmse(test_data$ZSE_norm, rf_pred)
rf_mape  <- mape(test_data$ZSE_norm, rf_pred) * 100
rf_r2    <- R2(rf_pred, test_data$ZSE_norm)
cat("Random Forest Performance:\n",
```

"MSE:", rf_mse, "\n",

"RMSE:", rf_rmse, "\n",

"MAPE:", rf_mape, "%\n",

"R^2:", rf_r2, "\n")

12. Random Forest Residual Plot

```
rf_resid <- test_data$ZSE_norm - rf_pred

ggplot(data.frame(Predicted=rf_pred, Residual=rf_resid), aes(x=Predicted, y=Residual)) +

  geom_point(color="red") +

  geom_hline(yintercept=0, linetype="dashed") +

  labs(title="Random Forest Residuals", y="Residuals", x="Predicted") +

  theme_minimal()
```

13. Cross Validation (k-Fold, Random Forest)

```
set.seed(123)

ctrl <- trainControl(method="cv", number=5)

rf_cv <- train(

  rf_formula, data=train_data, method="rf",

  trControl=ctrl, ntree=500, importance=TRUE

)

print(rf_cv)
```

15. Reliability Testing (Cronbach's Alpha)

Use all predictors for alpha (exclude target)

```
alpha_data <- train_data[, features[features != "ZSE_norm"]]

cronbach <- psych::alpha(alpha_data)

print(cronbach$total$raw_alpha)
```

16. XGBoost Model Development and Training

```
Dtrain <- xgb.DMatrix(data=as.matrix(train_data[, features]), label=train_data$ZSE_norm)

dtest  <- xgb.DMatrix(data=as.matrix(test_data[, features]), label=test_data$ZSE_norm)

params <- list(objective = "reg:squarederror", eta = 0.1, max_depth = 6,

               subsample = 0.8, colsample_bytree = 0.8)

set.seed(123)

xgb_model <- xgb.train(

  params = params, data = dtrain, nrounds = 100,

  watchlist = list(train = dtrain, eval = dtest),

  print_every_n = 10, early_stopping_rounds = 10, eval_metric = "rmse"

)
```

17. XGBoost Prediction & Table

```
xgb_pred <- predict(xgb_model, dtest)

xgb_results <- data.frame(

  Actual = test_data$ZSE_norm,

  Predicted = xgb_pred,
```

```
  Index = 1:nrow(test_data)

)

print(xgb_results)
```

18. Plot: Actual vs Predicted (XGBoost)

```
ggplot(xgb_results, aes(x=Index)) +

  geom_line(aes(y=Actual, color="Actual"), size=1) +

  geom_line(aes(y=Predicted, color="Predicted"), size=1, linetype="dashed") +

  scale_color_manual(values=c("Actual"="blue", "Predicted"="orange")) +

  labs(title="XGBoost: Actual vs Predicted", color="Legend") +

  theme_minimal()
```

19. XGBoost Feature Importance

```
importance_matrix <- xgb.importance(model=xgb_model, feature_names = features)

xgb.plot.importance(importance_matrix, top_n=10)
```

20. Performance Metrics (XGBoost)

```
xgb_mse  <- mse(test_data$ZSE_norm, xgb_pred)

xgb_rmse <- rmse(test_data$ZSE_norm, xgb_pred)

xgb_mape <- mape(test_data$ZSE_norm, xgb_pred) * 100

xgb_r2   <- R2(xgb_pred, test_data$ZSE_norm)

cat("XGBoost Performance:\n",

    "MSE:", xgb_mse, "\n",
```

```
    "RMSE:", xgb_rmse, "\n",

    "MAPE:", xgb_mape, "%\n",

    "R^2:", xgb_r2, "\n")
```

21. XGBoost Residual Plot

```
xgb_resid <- test_data$ZSE_norm - xgb_pred

ggplot(data.frame(Predicted=xgb_pred, Residual=xgb_resid), aes(x=Predicted, y=Residual))
+

  geom_point(color="orange") +

  geom_hline(yintercept=0, linetype="dashed") +

  labs(title="XGBoost Residuals", y="Residuals", x="Predicted") +

  theme_minimal()
```

22. Combined Plot: Actual vs Random Forest vs XGBoost

```
combined_df <- data.frame(

  Index = 1:nrow(test_data),

  Actual = test_data$ZSE_norm,

  RandomForest = rf_pred,

  XGBoost = xgb_pred

)

combined_long <- melt(combined_df, id.vars = "Index", variable.name = "Model",
value.name = "Value")

ggplot(combined_long, aes(x=Index, y=Value, color=Model)) +
```

```
  geom_line(size=1) +
  scale_color_manual(values=c("Actual"="black", "RandomForest"="red",
"XGBoost"="orange")) +
  labs(title="Actual vs Random Forest vs XGBoost", color="Legend") +
  theme_minimal()
```

23. Combined Table: Actual, RF, XGB

```
results_table <- data.frame(
  Index = 1:nrow(test_data),
  Actual = test_data$ZSE_norm,
  RandomForest = rf_pred,
  XGBoost = xgb_pred
)
print(results_table)
```