



NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FORECASTING THE CONSUMER PRICE INDEX IN NAMIBIA: A COMPARATIVE
ANALYSIS OF MACHINE LEARNING AND STATISTICAL METHODS

by

Elago Linea

222107871

Submitted in partial fulfilment of the requirements for the degree

Master of Data Science

in the

Faculty of Computing and Informatics

Department of Informatics

at the

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Supervisor

Dr Richard Maliwatu

02 October 2025

Metadata

STUDY TITLE: Forecasting the Consumer Price Index in Namibia: A Comparative Analysis of Machine Learning and Statistical Methods

STUDENT NAME: Elago Linea

SUPERVISOR: Dr Richard Maliwatu

DEPARTMENT: Informatics

QUALIFICATION: Master of Data Science

SPECIALISATION: Machine learning

MAIN KNOWLEDGE AREA: Data Science

KEYWORDS: Long Short-Term Memory (LSTM) recurrent neural network, Autoregressive Integrated Moving Average (ARIMA), Machine Learning (ML), Statistical methods, Consumer Price Index (CPI), Time Series Forecasting, Support Vector Regression (SVR), and Holt-Winters' exponential smoothing

TYPE OF RESEARCH: Applied Research

METHODOLOGY: Quantitative

STATUS: Thesis

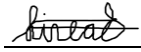
SITE: Namibia University of Science and Technology (NUST)

DOCUMENT DATE: October 2025

Declaration

I, Linea Gongeleni Elago, hereby affirm that this study is an accurate representation of my own work and that it has not been submitted for a master's degree at any other institution of higher education.

Signature



Date

02 October 2025

Abstract

In emerging countries like Namibia, accurate forecasting of the Consumer Price Index (CPI) is important for evidence-based policy development, effective economic planning, and inflation management. However, despite its importance, there has been limited application of advanced forecasting methods within Namibia's context. The study addressed this gap by investigating and comparing the performance of traditional statistical methods, AutoRegressive Integrated Moving Average (ARIMA), Holt-Winters exponential smoothing with machine learning methods, Long Short-Term Memory (LSTM) recurrent neural network, and Support Vector Regression (SVR) in forecasting Namibia's CPI using monthly data from 2013 to 2023. The findings revealed that SVR yielded the lowest Root Mean Square Error (RMSE), indicating higher forecast accuracy compared to other models. The study recommends a shift towards machine learning methods, particularly SVR, for CPI forecasting, given its capability to capture nonlinear trends in economic data. The study further recommends enhancing forecasting methods by incorporating relevant economic indicators such as interest rate, GDP growth, unemployment rate, and government expenditure. The study also emphasises the importance of investing in capacity-building for data science and developing real-time data systems to support policy formulation and CPI monitoring. The research contributes to the growing body of evidence that machine learning can enhance CPI forecasting and decision-making processes in developing countries.

Acknowledgements

I would like to express my gratitude to God for His numerous blessings during my academic journey. I wish to convey my immense appreciation to my supervisor, Dr Richard Maliwatu, for his exceptional assistance during the research endeavour. I extend my gratitude to all my lecturers for their invaluable contributions and support in my academic journey. I am very thankful to my family for their consistent support during my studies. Finally, I wish to extend my deep gratitude to my friends for their priceless support throughout my life.

Table of Contents

| | |
|--|------|
| Metadata..... | i |
| Declaration..... | ii |
| Abstract..... | iii |
| Acknowledgements..... | iv |
| List of Figures | viii |
| List of Tables | viii |
| List of Abbreviations | ix |
| CHAPTER ONE: INTRODUCTION..... | 1 |
| 1.1. Background..... | 1 |
| 1.2. Problem statement..... | 3 |
| 1.3. Objectives of the study..... | 4 |
| 1.4. Research questions..... | 4 |
| 1.5. Significance of the study..... | 5 |
| 1.6. Scope of the study | 5 |
| 1.7. Organisation of the dissertation..... | 5 |
| 1.8. Chapter summary | 6 |
| CHAPTER TWO: LITERATURE REVIEW | 7 |
| 2.1. Importance of CPI forecasting and implications of forecast errors | 7 |
| 2.2. Implications of CPI forecasting errors | 8 |
| 2.3. Traditional statistical methods in CPI forecasting | 9 |
| 2.4. Machine learning methods in CPI forecasting..... | 10 |
| 2.5. Comparative analysis of forecasting methods | 11 |
| 2.6. Forecast accuracy in developing countries | 13 |
| 2.7. Namibia-specific context | 14 |
| 2.8. Chapter summary..... | 15 |
| CHAPTER THREE: RESEARCH METHODOLOGY | 17 |
| 3.1. Research philosophy | 17 |

| | |
|---|----|
| 3.2. Research approach..... | 17 |
| 3.3. Research strategy..... | 18 |
| 3.4. Choices | 18 |
| 3.5. Time horizon | 18 |
| 3.6. Techniques and procedures..... | 18 |
| 3.6.1. Data description..... | 18 |
| 3.6.2. Data preparation..... | 19 |
| 3.6.3. Model development and data analysis..... | 19 |
| 3.6.4. Forecasting models used in the study | 20 |
| 3.7. Chapter summary..... | 27 |
| CHAPTER FOUR: RESULTS AND DISCUSSION OF FINDINGS | 29 |
| 4.1 Descriptive analysis..... | 29 |
| 4.2. Forecasting CPI for Namibia using traditional statistical methods..... | 32 |
| 4.2.1. ARIMA model | 32 |
| 4.2.2. Holt-Winters Exponential Smoothing model | 35 |
| 4.3. Forecasting CPI for Namibia using machine learning methods | 36 |
| 4.3.2. Support Vector Regression model | 38 |
| 4.4. Comparison of the traditional statistical methods and machine learning models..... | 39 |
| 4.5. Discussion of findings..... | 40 |
| 4.6. Chapter summary..... | 42 |
| CHAPTER FIVE: CONCLUSION..... | 43 |
| 5.1. Reflecting on research questions..... | 43 |
| 5.2. Ethical considerations..... | 44 |
| 5.3. Contribution of the study | 44 |
| 5.4. Limitations of the study..... | 44 |
| 5.5. Future work | 45 |
| 5.6. Closing remarks | 46 |
| References | 47 |

| | |
|--|----|
| Appendix: Python's implementation details..... | 52 |
|--|----|

List of Figures

| | |
|--|----|
| Figure 1.1: Namibia CPI from 2015 to 2023..... | 2 |
| Table 2.1: Summary of the reviewed literature..... | 15 |
| Figure 3.1. Recurrent Neural Network architecture..... | 22 |
| Figure 3.2 The architecture for the LSTM recurrent neural network..... | 24 |
| Figure 4.1: CPI Time Series for Namibia (2013-2023)..... | 29 |
| Figure 4.2: Histogram of CPI for Namibia (2013-2023)..... | 30 |
| Figure 4.3: Q-Q Plot of CPI for Namibia (2013-2023)..... | 30 |
| Figure 4.4: Boxplot of CPI for Namibia (2013-2023)..... | 31 |
| Figure 4.5 Correlation matrix of CPI and other economic indicators..... | 31 |
| Figure 4.6: First-order differenced CPI of Namibia (2013–2023)..... | 33 |
| Figure 4.7: Second-order differenced CPI of Namibia (2013–2023)..... | 33 |
| Figure 4.8: Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots..... | 34 |
| Figure 4.9: Forecasting CPI for Namibia using the ARIMA Model..... | 35 |
| Figure 4.10: Forecasting CPI for Namibia using the Holt-Winters Exponential Smoothing model..... | 36 |
| Figure 4.11: Forecasting CPI for Namibia using the LSTM model..... | 37 |
| Figure 4.12: Forecasting CPI for Namibia using the SVR model..... | 38 |

List of Tables

| | |
|--|----|
| Table 2.1: Summary of Literature Review..... | 15 |
| Table 3.1: Summarised study methodology..... | 27 |
| Table 4.1. Comparison of the forecasting models..... | 39 |

List of Abbreviations

| | |
|--------|---|
| ANN | Artificial Neural Network |
| ARFIMA | Autoregressive Fractionally Integrated Moving Average |
| ARIMA | Autoregressive Integrated Moving Average |
| BAVART | Bayesian Additive Vector Autoregressive Tree |
| CBS | Central Bureau of Statistics |
| CPI | Consumer Price Index |
| DES | Double Exponential Smoothing |
| GDP | Gross Domestic Product |
| HRNNs | Hierarchical Recurrent Neural Networks |
| ICPI | Interim Consumer Price Index |
| IFS | International Financial Statistics |
| ILO | International Labour Organisation |
| IMF | International Monetary Fund |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LSTM | Long Short-Term Memory |
| MAE | Mean Absolute Error |
| ML | Machine Learning |
| NSA | Namibia Statistics Agency |
| RMSE | Root Mean Square Error |
| RNN | Recurrent Neural Network |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| SVM | Support Vector Machine |
| SVR | Support Vector Regression |
| VAR | Vector Autoregressive |

CHAPTER ONE: INTRODUCTION

1.1. Background

The Consumer Price Index (CPI) is an essential economic measure that assesses the mean change in prices that consumers have paid overtime for a basket of similar goods and services (Fernando, 2023). The CPI indicates the expenditure of consumers in a specific country or region for a basket of goods and services (Fernando, 2023). Furthermore, it indicates how the cost of living is changing over time. CPIs are statistics often generated monthly by National Statistics Offices on a monthly basis (Fernando, 2024). Governments, financial institutions, and businesses use CPI as a major indicator to forecast inflation and determine the status of the economy (Zahara et al., 2020). According to Konarasinghe (2022), inflation is defined as a persistent phenomenon characterised by a steady rise in prices. When prices rise, consumer purchasing power declines, resulting in diminished money to spend and potentially decreased consumer expenditure (Konarasinghe, 2022). This, in turn, can have a ripple effect on the broader economy. Fernando (2024) proffers that understanding inflation is important for formulating efficient economic policy and for making decisions in any country. Additionally, the periodic publication of CPI allows for the analysis of price changes in individual items and the overall basket, giving insight into inflation trends.

Central banks closely monitor the CPI to adjust monetary policy decisions, given its critical role in tracking inflation (Coibion et al., 2020). According to Castillo-Martinez et al. (2024), central banks use CPI trends to adjust interest rates to control price increases and stabilise the economy. Additionally, correct predictions of CPI enable firms to make informative decisions on price settings and budget estimates. On the other hand, households may rely on such estimations to better plan their financial expenditures (Madudova & Corejova, 2024). All these help to stabilise the economy and growth in the long run. Therefore, CPI is a vital tool for economic planning, influencing monetary policy and cost-of-living adjustments in the world.

In Namibia, the Bank of Namibia relies on CPI data compiled by the Namibia Statistics Agency (NSA) to set inflation targets and guide monetary policies (Bank of Namibia, 2021). The CPI is produced every month, with prices collected from different stores across the country (Namibia Statistics Agency, 2023). Before the establishment of the NSA in 2011, the CPI was produced by the Central Bureau of Statistics department within the National Planning Commission of Namibia. The index was called the Interim Consumer Price Index (ICPI) and was first produced in 1993. The ICPI was derived from the spending of 800 households included in the 1984/5 Namibia Household Income and Expenditure Survey (NHIES). The basket had 183 items with prices gathered from roughly 93 retailers. Approximately 570 price requests were processed monthly. As consumption patterns evolved, the ICPI basket of goods and services became outdated, leading to the discontinuation of its calculation in January 2005. The current CPI index was recalibrated utilising the 2009-2010 NHIES and has exhibited an upward trend over the past decade, increasing from 105.88% in 2014 to 166.82% in 2023 (Namibia Statistics Agency, 2024). The higher levels of the Consumer Price Index (CPI) negatively impact people's welfare and overall economic stability, indicating the presence of inflation. Forecasts can then be used to implement measures to mitigate these effects (Coibion et al., 2020). The following graph shows the trend of Namibia's CPI from 2015 to 2023.



Figure 1.1: Namibia CPI from 2015 to 2023

Adopted from Tradingeconomics.com (2024)

Over time, traditional statistical methods such as Vector Autoregression, exponential smoothing models, Bayesian Forecasting, Autoregressive Integrated Moving Average (ARIMA), and Decomposition Methods have been used for forecasting CPI (Makridakis et al., 2018). The methods rely on past CPI values to detect trends and predict expected values. Seasonal adjustments and trends often enhance these models to improve their accuracy. Although the models have proven useful for capturing fundamental trends and patterns in data, they frequently struggle to account for non-linear relationships, complex seasonal changes, and over time interdependence present in the data [Click or tap here to enter text.](#)(Fattah et al., 2018; Makridakis et al., 2018). Therefore, the models may fall short in predicting accurate CPI trends (Zhou et al., (2017). However, researchers have employed ML to enhance precision when making predictions in different fields of such as economics, education, healthcare, agriculture, and veterinary medicine (Zahara et al., 2020). Specifically, according to Aaron (2018), Zahara et al. (2020), Paranhos (2022), and Barkan et al. (2023), economists are actively exploring ML techniques such as Support Vector Regression (SVR), multivariate linear regression, as well as neural networks for forecasting time series data. The researchers emphasise that these methods have emerged as alternatives for forecasting economic indicators due to the growing amount of data on consumer purchasing behaviour. ML may adjust to evolving data trends, alleviating challenges associated with fluctuations in consumer preferences and product quality, which traditional statistical methods find difficult to address (Zhou et al. (2017). Zhou et al. (2017) and Alsharif et al. (2020) also state that as big data keeps increasing, ML approaches are more and more prevalent because they can efficiently deal with these massive and complex data, thereby identifying trends that traditional statistical methods may struggle to analyse due to the complexity of the data, and do not make assumptions about the data.

1.2. Problem statement

Namibia's CPI has shown a consistent upward trend throughout the years, necessitating a quest to understand these trends for policy formulation and analysis of the economy. Traditional time series methods have been used to forecast CPI. However, these methodologies rely on stationarity assumptions, suggesting that both the average and variance of the data remain constant over time, which may result in erroneous projections

when dealing with non-stationary data. Additionally, numerous economic relations are not linear; however, these approaches assume linear relationships in the data, which may lead to erroneous relationship representations. Furthermore, the traditional approaches are inadequate to deal with large datasets. ML methods can capture dependencies that last over time, manage non-stationarity, predict non-linear relationships in the data, and do not presume a predefined data distribution. Therefore, the present study aimed to determine how well ML methods, LSTM recurrent neural networks, and SVR, as well as traditional statistical methods, the ARIMA model, and Holt-Winters' exponential smoothing perform in forecasting CPI for Namibia.

1.3. Objectives of the study

Main objective

To investigate the performance of ML methods and traditional statistical methods in forecasting the CPI for Namibia.

Sub-objectives

- i) To assess Namibia's CPI trend over the past ten years
- ii) To assess the performance of traditional statistical methods on the CPI data for Namibia.
- iii) To assess the performance of ML methods on the CPI data for Namibia.

1.4. Research questions

Main question

How do ML methods and traditional statistical methods perform in forecasting CPI for Namibia?

Sub-questions

- i) What has been the trend in Namibia's CPI over the past ten years?
- ii) How accurately do traditional statistical methods forecast the CPI for Namibia?
- iii) How accurately do machine learning methods forecast the CPI for Namibia?

1.5. Significance of the study

For policymakers to make well-informed decisions, accurate CPI estimates must be produced. To decide on monetary policy, including interest rates, policymakers must have a thorough understanding of how prices for goods and services are changing. However, investors must be aware of this information to help them decide whether to purchase stock or bonds, among other investment-related decisions. For this reason, having precise CPI projections is critical. The present study intended to find recurring patterns, possible contributing variables, and the implications of these trends for consumers, businesses, and policymakers by analysing historical CPI data. Therefore, the study advances knowledge of Namibia's economic processes and offers guidance for making informed decisions.

1.6. Scope of the study

The study covered the period of 2013 to 2023 and focused on the application of LSTM and ARIMA models to forecast future CPI values for Namibia. The data of the study were obtained from the International Financial Statistics (IFS) database and the Namibia Statistics Agency data portal, where it is available in the public domain. The primary variable is CPI, and the features of the study are interest rate, unemployment rate, GDP growth rate, real gross domestic product, real government consumption expenditure, real exports of goods and services, and real imports of goods and services.

1.7. Organisation of the dissertation

The structure and organisation of the study conform to a cohesive order. The study began with a comprehensive introduction that detailed the problem statement, research questions, aims, objectives, aims, significance and scope of the study. The literature review highlights the work done by other researchers in economic forecasting, followed by a detailed explanation of the chosen methodology. The heart of the research lay in the evaluation of the application of models of LSTM, SVR, ARIMA, and Holt-Winters' exponential smoothing for Namibia's CPI prediction. This approach enabled a detailed analysis and comparison between the statistical method and ML methods to help understand which model is best suited for

forecasting CPI for Namibia. Finally, the study concludes with a comprehensive discussion and closing remarks.

1.8. Chapter summary

Chapter One introduces the study on forecasting the CPI in Namibia using both ML and traditional statistical methods. It highlights CPI as a vital economic indicator that measures changes in the average prices of goods and services over time, guiding decisions on inflation, monetary policy, and overall economic planning. The chapter traces the evolution of Namibia's CPI from its initial compilation by the Central Bureau of Statistics to its current production by the NSA, noting a consistent upward trend that reflects inflationary pressures. Traditional forecasting methods such as ARIMA and Holt-Winters' exponential smoothing are discussed for their usefulness but also their limitations in handling non-linear and non-stationary data. In contrast, ML techniques such as LSTM networks and SVR are presented as more advanced tools capable of capturing complex patterns and improving forecast accuracy. The chapter defines the problem as the need for more reliable CPI forecasting methods to support effective policy formulation and economic analysis. The study's objectives and research questions focus on examining CPI trends, assessing the performance of traditional and ML models, and comparing their predictive accuracy. Its significance lies in providing accurate CPI projections to assist policymakers, investors, and businesses in decision-making. The study covers the period from 2013 to 2023, using data from the International Financial Statistics and NSA databases, with CPI as the main variable and several macroeconomic indicators as explanatory features. The chapter concludes by outlining the structure of the dissertation, which proceeds from the introduction to the literature review, methodology, results, and conclusion, laying the foundation for identifying the most effective model for forecasting Namibia's CPI.

CHAPTER TWO: LITERATURE REVIEW

The study examined existing empirical research on forecasting CPI, highlighting a variety of methodological approaches, including traditional statistical methods and advanced ML techniques. Furthermore, the review critically assesses studies dedicated to identifying and analysing the key determinants influencing CPI dynamics across diverse analytical frameworks.

2.1. Importance of CPI forecasting and implications of forecast errors

For well-informed economic planning, policymaking, and investment decision-making, accurate CPI forecasting is essential (Metu et al., 2024). As per the Organisation for Economic Co-operation and Development Standards (2025), no forecast is error-free; a reasonable margin of error for CPI forecasting usually ranges from ± 1 to ± 2 percentage points, based on the planning scope and context. Errors that go beyond this range may have significant economic consequences.

CPI forecasting is important for efficient economic planning and policymaking as it helps governments manage inflation, guide monetary policy, and ensure fiscal stability (Ben, 2024). By predicting inflation trends, policymakers can adjust interest rates, control public spending, and plan social security and wage adjustments to protect citizens' purchasing power. Accurate CPI forecasts also support cost-of-living adjustments and price stabilisers, thereby fostering consumer and investor confidence in the economy. Additionally, CPI forecasting informs trade and foreign exchange policies, thus helping governments maintain economic competitiveness and aiding in long-term investment and infrastructure planning by offering an overview of future costs. Overall, CPI forecasting ensures that governments can respond proactively to inflationary pressures and maintain economic stability.

However, errors in CPI forecasting carry serious risks. According to Ben (2024) and Mishkin (2018), if CPI is over-forecasted, it may lead to excessively tight monetary policies, such as

sudden interest rate hikes, which may hinder economic activity, limit investment, and raise unemployment if the CPI is overestimated.

On the other hand, underestimating CPI could cause policymakers to take longer to respond, which would allow inflation to get out of control. This reduces buying power, especially for people with fixed and low incomes (ILO, 2022). Additionally, wage negotiations, price regulation, and government budgeting are all made more difficult by inaccurate estimates, particularly in industries that are vulnerable to inflation, such as food, housing, transportation, and energy. Therefore, in order to ensure macroeconomic stability and social equality, the need for accurate and flexible forecasting tools like machine learning models that reduce error was not only a technical endeavour but also a policy requirement (Araujo et al., 2023; Makridakis et al., 2018).

2.2. Implications of CPI forecasting errors

Errors in forecasting the CPI could lead to substantial economic and social consequences. An underestimation of the CPI forecasts may lead central banks to lower their interest rates, which might increase inflationary pressures and destabilise the economy (Mishkin, 2018). However, an overestimation of the CPI forecasts may result in excessively elevated interest rates, which could hinder economic growth and investments (Taylor & Barbosa-Filho, 2021). In fiscal policy, inaccurate CPI forecasts may mislead the development of government budgets, subsidies, or tax policies, potentially overburdening or under-supporting some sectors (Auerbach & Gale, 2023). Additionally, wage adjustments are influenced by the accuracy of CPI estimations; underestimation might decrease workers' real purchasing power, whereas overestimation could elevate labour costs, thereby impacting business profitability and employment. Furthermore, forecasting errors can undermine investor confidence, potentially leading to misinterpretations of inflation trends and poor investment decisions, which in turn affect financial markets. At a social level, underestimations can reduce household purchasing power, greatly affecting vulnerable populations, whereas overestimations may induce unnecessary panic regarding inflation, which negatively impacts consumer behaviour (Lakshmanamoorthy & Athinarayanan, 2024).

2.3. Traditional statistical methods in CPI forecasting

The study done by Dhamo et al. (2018) was about forecasting CPI and determining the association of CPI with other economic factors in Albania using SARIMA and multiple regression models. The variables of the study were the number of Albanian people travelling abroad and the exchange rate. The study found that there was a positive association between the exchange rate and the number of Albanians travelling abroad, explaining 75% of the CPI's performance. The data visualisation revealed that SARIMA models outperformed multiple regression models in short-term predictions. The researchers recommended future research to further investigate the correlation between CPI and other economic indicators such as GDP, unemployment rate, and interest rate. Furthermore, they stressed the analysis of the influence of external factors on CPI, such as changes in exchange rates and international trade policies. This would help to understand how external economic factors can influence the CPI and supply an overview of the vulnerability of the CPI to these factors.

Shinkarenko et al. (2021) studied the behaviour of the CPI in Ukraine from January 2010 to September 2020 using Holt-Winters and ARIMA models to forecast CPI. The Holt-Winters model was found to have a minimum error and was used for the inflation forecast in Ukraine. The research determined that an inflation rate of 5-10% annually was normal for the economy of Ukraine, and there were no grounds to expect it to rise in the near future. The researchers suggested that the causes of the index's growth should be studied in relation to the CPI's components, the exchange rate, and GDP.

Konarasinghe (2022) tested the Holt-Winters' three-parameter model, the Double Exponential Smoothing model (DES), and multiplicative models to predict CPI in Malaysia from 2020 to 2024. The findings indicated that the DES model had significantly superior predictive ability compared to the other models used in the study. Furthermore, the research suggested that exponential smoothing models are potential tactics for forecasting CPI, notwithstanding their infrequent use in CPI forecasting as per existing literature. However, it was determined that the DES prediction yielded a linear trend and failed to effectively capture intricate seasonal patterns or non-linearities in the data, which are critical for accurate long-term CPI forecasts.

2.4. Machine learning methods in CPI forecasting

Akande et al. (2022) conducted a study that placed an emphasis on the significance of precise identification and comprehension of the factors influencing inflation components, especially within the Nigerian food sector. The research employed two ML methods, random forest and gradient boosting. The research utilised monthly CPI data. The random forest was identified as the most effective method, highlighting the capability of non-linear models to accurately predict specific inflation subgroups. The findings indicate that the food CPI was the primary factor influencing inflation in the urban and rural areas of Nigeria. Bread and cereals were the greatest substantial contributors to food inflation in the country. Furthermore, biscuits, agricultural rice, and white garri were identified as the primary contributors to the inflation of bread and cereal. This corresponds with the established view that food prices have a substantial effect on inflation, given their considerable share of household income.

LSTMs have also been used to forecast currency exchange rates, where historical data is crucial for making accurate predictions. Paranhos (2022) applied LSTM models to forecast exchange rates and demonstrated the model's effectiveness in capturing volatile market trends. The study showed that LSTM's ability to handle non-linear relationships allowed for more accurate forecasting compared to traditional models. Economic growth and inflation forecasting have also benefited from LSTM models. LSTMs help capture complex interrelationships between variables such as GDP, interest rates, and inflation levels.

Zahara et al. (2020) studied CPI prediction using LSTM-based cloud computing. The output of the network model is the CPI value, whereas the input comprises 34 variables related to the price of commodities in Surabaya. The study used optimisation algorithms of stochastic gradient descent, root mean square propagation, adaptive gradient, adaptive moment, and Nesterov Adam to improve the accuracy of the forecasts. The study concluded that Nesterov Adam was the best optimisation algorithm that improved the accuracy of the CPI forecasts, with a lower RMSE of 4.088. The researchers recommended the application of the LSTM model for CPI forecasting.

2.5. Comparative analysis of forecasting methods

Several studies have conducted comparative analyses of machine learning and traditional statistical methods for forecasting CPI and related economic indicators. Barkan et al. (2023) conducted a study on forecasting disaggregated inflation components of the monthly US CPI data (1994–2019). The study compared the Hierarchical Recurrent Neural Networks (HRNNs) method with ARIMA, Phillip's curve models, random walk, random forest, gradient, and boosting deep neural networks. The researchers found that HRNNs significantly outperform all the other methods across multiple forecasting horizons in terms of RMSE and Pearson correlation. The study concluded that HRNNs are effective for component-level inflation forecasting, as they provide valuable tools for policymakers and financial analysts. Additionally, performance gains are largest in sectors with deeper hierarchical structures, such as food and beverages.

Almosova and Andresen (2019) compared the performance of ML and statistical methods in forecasting monthly United States CPI inflation. The LSTM recurrent neural network was evaluated against the linear autoregressive model, the random walk model, the Seasonal Autoregressive Model (SARIMA), a Markov switching model, and the simple fully connected neural network to predict monthly US CPI inflation. The research indicated that the LSTM recurrent neural network outperformed all the other models. Its capacity to handle non-linear assumptions contributed to its good performance. The study further indicated that the flexibility of LSTM models makes them a capable tool for forecasting CPI; however, they need a lot of training data and can overfit if not adequately regularised.

Araujo and Gaglianone (2023) predicted inflation in Brazil using ML methods. The study's objective was to evaluate whether ML techniques could increase prediction accuracy in applied macroeconomics by comparing them with econometric techniques. The methods used in the study were random walk (Atkeson-Ohanian), ARIMA, Vector Autoregressive (VAR) model, Phillip's curve, elastic net, lasso, adaptive lasso, ridge regression, random forest, quantile regression forest, XGBoost, and RNN. The research covered a wide range of factors, including financial markets, price indices, economic activity, interest rates, government debt, imports and exports of goods and services, and international variables that are related to the

Brazilian economy. The study found that ML approaches had a lower RMSE than econometric methods.

Zhou et al. (2017) used LSTM recurrent networks and ARIMA models to forecast stock market patterns. The findings found that the LSTM recurrent method significantly outperformed the ARIMA model and concluded that LSTM's ability to capture long-term dependencies made it an important tool for financial forecasting.

Gur (2024) conducted a study to assess the forecasting performance of various statistical and ML models on monthly U.S. CPI data. The models evaluated included ARIMA, the VAR model, LSTM, Random Forest (RF), Gradient Boosting Machines (GBM), and hybrid approaches. The results showed that hybrid models incorporating LSTM and statistical methods outperformed the individual models, especially in determining the complex and non-linear structure of inflation data. Among the standalone models, LSTM demonstrated the lowest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) across different forecasting horizons. Furthermore, the VAR model indicated that variables such as CPI, oil prices, and GDP had a strong and significant effect on CPI. The study concluded that advanced machine learning models, particularly those with memory capabilities like LSTM, provide superior accuracy and robustness in forecasting inflation under uncertain economic conditions. These findings showed that hybrid methods can greatly enhance economic focus and offer valuable insights for policymakers, investors, and market analysts.

Barhoumi et al. (2022) developed a machine learning nowcasting framework to overcome the lag in official GDP releases and track real-time economic activity in data-scarce sub-Saharan African economies during the COVID-19 pandemic. Their three-stage approach first selected timely predictors, such as commodity prices, trade flows and mobility indicators, based on statistical relationships and data availability; then ran horseraces over thirty ML models (ridge and Lasso regressions, random forests, gradient boosting and support vector machines) using cross-validated RMSE to identify the best performer; and finally re-estimate the winning model on the full sample to produce nowcasts. It was applied to Nigeria and Botswana, and the framework consistently yielded lower out-of-sample errors than traditional methods, accurately capturing the sharp COVID-19-induced contractions in 2020 and the subsequent

recoveries, with Shapley decompositions illuminating each predictor's contribution to the nowcasts. Aggregating country nowcasts proxies regional GDP growth, revealing an unprecedented 8.5 percent contraction in quarter 2, 2020, and a spatial rebound thereafter. The study concluded that machine learning methods, when compared to statistical methods for nowcasting, can greatly enhance the surveillance capabilities of the International Monetary Fund (IMF) and policymakers in environments with delayed or sparse official statistics.

Kubheka (2023) conducted a study focused on modelling South Africa's inflation using bootstrapped LSTM methods. The study aimed to evaluate whether LSTM, when combined with clustered bootstrap sampling, could outperform the traditional statistical model Autoregressive Fractionally Integrated Moving Average (ARFIMA). The author used monthly inflation data from 1978 to 2022, applied several error metrics (RMSE, MAE, MAPE, RMSPE), and conducted the Diebold-Mariano test to compare model accuracy. Results demonstrated that the clustered LSTM model consistently outperformed both ARFIMA-based models across all error metrics. The findings suggested that deep learning techniques like LSTM, especially when enhanced with bootstrap methods, offer promising alternatives for forecasting inflation in complex and volatile economic environments such as South Africa. The study also highlighted the limitations of using univariate models and suggested future extensions with multivariate and alternative bootstrapping techniques.

2.6. Forecast accuracy in developing countries

Botha et al. (2022) investigated whether big data and modern statistical learning techniques can enhance near-term inflation forecasts in South Africa by assembling a comprehensive dataset of 216 disaggregated consumer price indices from January 2009 to March 2021 and benchmarking a diverse array of forecasting approaches. These included traditional univariate models (random walk, first-order autoregressive (AR(1)), large Bayesian VARs, stochastic volatility models, the SARB's Disaggregated Inflation Model (DMI), policymaker forecasts, dynamic factor models (both principal component-based and three-pass regression filters), various penalised regression techniques (Least Absolute Shrinkage and Selection Operator (LASSO), adaptive LASSO, ridge, elastic net, Smoothly Clipped Absolute Deviation, and L-penalized models), Bayesian model averaging, and nonlinear methods such as random

forests, gradient boosting, complete subset regression ensembles, and a basic three-layer neural network. Their recursive out-of-sample evaluation revealed that the official SARB's Disaggregated Inflation Model (DMI) forecasts perform best at very short horizons (1-3 months), likely due to the incorporation of judgement inputs and within-month data updates, but that statistical learning models, particularly gradient boosting on four-digit data and the neural network on the fully disaggregated series, outperform both traditional benchmarks and dynamic factor models at medium to long horizons (6-24 months). Importantly, greater data granularity enhances the effectiveness of sparse methods like LASSO by enabling the selection of highly informative predictors, thereby reducing noise. Shapley value decompositions further contribute to interpretability by quantifying each variable's contribution. The authors concluded that, especially during periods of typical inflation dynamics such as the COVID-19 pandemic, machine learning forecasts offer significant accuracy gains and valuable transparency, thus making them promising tools for central banks operating in data-scarce environments.

2.7. Namibia-specific context

Currently, there is insufficient specific research on CPI forecasting in Namibia. To understand the importance of diesel fuel prices in Namibia and how their varying prices impacted the food and non-alcoholic beverage sectors, Pazvakawambwa and Isaak's (2022) study on forecasting CPI for diesel prices, food, and non-alcoholic beverages was conducted. The association between average diesel costs and the CPI for food and non-alcoholic drinks was investigated using a time series regression analysis tool. Future values were predicted using the ARIMA model with monthly data from 2017 to 2021. Both diesel prices and the CPI for food and non-alcoholic beverages showed an upward trend over the studied period. According to the study, average diesel costs and the CPI for food and non-alcoholic beverages showed a weak positive correlation. The impact of average diesel prices was not statistically significant, according to the regression equation for the time series. As such, it was recommended that future food and non-alcoholic beverages CPI projections should employ the ARIMA model, which was determined to be statistically significant.

2.8. Chapter summary

Chapter Two reviewed literature on CPI forecasting and emphasised its importance for economic planning and policy, especially in developing countries. It compared traditional methods like ARIMA with ML techniques such as LSTM, thereby showing that ML models generally provide more accurate forecasts. Comparative studies from various countries highlight the strengths of ML in capturing non-linear patterns and improving forecast performance. Despite the growing global literature on CPI forecasting, few studies have focused specifically on Namibia. Botha et al. (2022) and Kubheka (2023) emphasise the importance of strong forecasting methods in countries with developing statistical systems, demonstrating that ML models such as gradient boosting and LSTM offer significant forecasting improvements in volatile economies. This study contributes to this gap by exploring Namibia CPI dynamics using models tested and validated in similar contexts.

Table 2.1: Summary of the reviewed literature

| Study | Model approach and scope | Key findings |
|--|---|--|
| Pazvakawambwa and Isaak (2022), | ARIMA; diesel prices and food CPI; Namibia; monthly (2017–2021) | The ARIMA model was found to be statistically significant. |
| Gjika Dhamo et al. (2018) | SARIMA and multiple regression; exchange rate and tourism; Albania; quarterly (2017–2018) | SARIMA models provided better short-term predictions than multiple regression models |
| Shinkarenko et al. (2021); Konarasinghe (2022) | Holt-Winters, DES; Economic variables; Ukraine and Malaysia; monthly (2020–2024) | The Holt-Winters model performed better than DES |

| | | |
|--|---|---|
| Barkan et al. (2023); Almosova and Andresen (2019); Gur (2024) | LSTM, HRNN, US CPI, deep learning vs traditional models, monthly (1994–2019, US) Hybrid approaches, LSTM, VAR, ARIMA | LSTM and HRNNs performed better than the traditional statistical methods Hybrid approaches outperformed all the other models |
| Araujo and Gaglianone (2023) | ML methods versus statistical methods; CPI and macro indicators; Mexico & Brazil | ML methods are more accurate during volatility with lower RMSE |
| Akande et al. (2022) | Ensemble ML; Nigeria; food CPI and inflation drivers; monthly (2010 – 2020) | The Random Forest (RF) base learner emerges as the best performing |
| Zahara et al. (2020) | LSTM with optimizers; Indonesia; commodity price inputs; monthly (2016–2018) | Nesterov Adam yielded lowest RMSE |
| Barhoumi et al. (2022) | ML methods and Statistical methods | ML outperformed statistical methods |
| Kubheka (2023) | Bootstrapped LSTM methods and statistical methods | Bootstrapped LSTM methods outperformed statistical methods |

CHAPTER THREE: RESEARCH METHODOLOGY

The chapter highlights the methodology that was utilised to fulfil the study objectives and questions. The study methodology was inspired by Saunders's research onion framework (Saunders et al., 2007).

3.1. Research philosophy

Informed by Saunders's Research Onion framework, the study utilised a positivist research philosophy (Saunders, 2007). The framework states that positivism assumes that reality is objective, measurable, and independent of the researcher; furthermore, it supports the use of quantitative methods to identify patterns and test hypotheses. The philosophy was deemed appropriate for the study because the research supports the comparison of forecasting models on measurable performance criteria, which is central to the research objectives. To evaluate forecasting accuracy, the study used the model evaluation of Root Mean Squared Error (RMSE). The approach aligns with the study's goal of producing generalisable and replicable findings that can inform economic forecasting and policymaking in Namibia.

3.2. Research approach

The study followed an abductive research approach that combines both data-driven insights and theoretical understanding. The approach is particularly appropriate for addressing complex real-world problems where existing theories may not fully explain observed phenomena (Dubois & Gadde, 2002). In this research, initial assumptions about the determinants of Namibia's CPI, such as interest rates and GDP growth, were drawn from established economic theory. These guided the initial stages of model development using traditional statistical forecasting techniques. However, observed nonlinearities and structural breaks in the data justified the integration of ML techniques to enhance predictive performance. The performance of these models offered new empirical insights, which were then used to refine or reinterpret theoretical expectations. This reasoning process aligns with the principles of abduction, where theory and data continuously inform one another to reach

the most plausible explanation (Dubois & Gadde, 2002). By combining theoretical knowledge with empirical exploration, the abductive approach enabled this study to deliver both accurate forecasts and a deeper understanding of CPI dynamics in Namibia.

3.3. Research strategy

A comparative strategy to assess the performance of ML and traditional statistical methods for forecasting CPI in Namibia was used. The strategy was adopted in line with research on Saunders' onion framework, which supported the comparison of different phenomena to identify the most effective method (Saunders et al., 2007).

3.4. Choices

The study utilised a quantitative approach. According to Mohajan (2020) and Xiong (2022), researchers use quantitative research methods to predict future outcomes, as well as to generalise findings to the entire population. Quantitative methods were selected due to their strength in handling structured numerical data and enabling statistical inference.

3.5. Time horizon

The study used a longitudinal approach where data were collected at different points to forecast CPI. Secondary monthly data from 2013 to 2023 were used to capture meaningful trends and patterns. Longitudinal studies are most suited for time series analysis and model forecasting (Saunders et al., 2007).

3.6. Techniques and procedures

The steps employed for data preparation, development of the model, and analysis are described below.

3.6.1. Data description

The study used secondary data obtained from the International Financial Statistics (IFS) database and the Namibia Statistics Agency data portal, where it was available in the public domain. The dataset consisted of 132 monthly CPI observations and was divided into 70%

training and 30% testing sets. The predicted variable of the study is CPI, and the input features of the study are interest rate, deposit interest rate, lending rate, unemployment rate, gross domestic product growth rate, real gross domestic product, real government consumption expenditure, real exports of goods and services, and real imports of goods and services.

3.6.2. Data preparation

Data preparation entailed the extraction of relevant features from the two sources and combining them into a single dataset. The data were cleaned and converted to a time series format. The features of the study had different production frequencies, with some of the features produced monthly, quarterly, or annually. Quarterly values, assuming constant activity within the quarter, were divided by three and distributed evenly over three months to align the data. Monthly estimates for every year were also produced by dividing annual values by 12. There were no missing values discovered when the dataset was examined. The Year and Month features were combined into a single feature called Date to enable the dataset to be in a time series format. To ensure that the data were ordered by time, the Date feature was set to be an index. This allowed for easier time series operations and visualisations. After all data were converted into monthly data, the dataset was divided into a training and a testing set. Feature scaling and time step creation play a part in turning raw data into a well-organised and useful set of features. This allows the ML models to use them to make spot-on forecasts. Therefore, the features were scaled, and time-step features were created to convert the data into the appropriate format for the LSTM model architecture.

3.6.3. Model development and data analysis

Python, a programming language, was used for model development and analysis. All the steps and codes are provided in the appendix. An exploratory analysis was done in the form of data visualisation to identify patterns. Correlation analysis was done to study relationships between CPI and the features of interest. We chose the features that had a higher positive correlation with CPI for the prediction. Features that indicated high correlation with other variables were excluded from the study to enhance model performance. Finally, the features of interest were fitted to the models to predict CPI values. Models' performances were evaluated with the RMSE to determine the best model for forecasting CPI for Namibia.

3.6.4. Forecasting models used in the study

a) AutoRegressive Integrated Moving Average (ARIMA)

The ARIMA model is one of the statistical methods that has been extensively used for time series analysis (Al-Chalabi et al., 2018; Fattah et al., 2018; Imron et al., 2022). It combines autoregression differencing, which is done to handle non-stationary data, and moving averages to predict future values based on previous values (Fattah et al., 2018). However, according to Fattah et al. (2018), the major limitation of ARIMA was its assumption of stationarity. CPI data, especially in volatile economies, was often non-stationary, thus making the ARIMA model less effective without substantial data preprocessing. Additionally, ARIMA struggled with nonlinear relationships in the data. The model has three parameters as follows:

ARIMA (p, d, q) model whereby:

- p : Number of autoregressive terms
- d : Number of differences
- q : Number of moving averages

ARIMA (p, d, q) model is given by:

$$y_t = \mu + \sum_{i=1}^p (\sigma y_{t-i}) + \sum_{i=1}^q (\theta \varepsilon_{t-i}) + \varepsilon_t \quad (\text{Eqn: 3.1})$$

The model's parameters are described as follows:

μ : The mean value of the time series data

σ : Autoregressive coefficients

θ : Moving average coefficients

ε : The white noise of the time series data

The model underperforms when dealing with seasonal variation and non-stationary data and assumes a linear relationship in the data (Al-Chalabiet al., 2018). However, the ARIMA model works well on small datasets, and it is easy to interpret the results (Al-Chalabiet al., 2018).

b) Holt-Winters Exponential Smoothing Method

This study used Holt-Winters Exponential Smoothing, a traditional statistical model, to forecast the CPI. According to Kumar et al. (2024), the model is widely used due to its capability of capturing both pattern and seasonal variations in time series data, which was especially effective in situations where economic indicators exhibit regular cyclical behaviour. The Holt-Winters forecast takes the following general form:

$$\hat{y}_{t+h} = (l_t + h \cdot b_t) + s_{t+h-m(k+1)} \quad (\text{Eqn 3.2})$$

The parameters are defined as follows:

\hat{y}_{t+h} : forecast for h periods ahead

l_t : estimated level at time t

b_t : estimated trend

s_t : seasonal component

m : length of seasonality

The model has three smoothing parameters, namely alpha (α) for the level, beta (β) for the trend, and gamma (γ) for the seasonal component. Based on the available new data and past predictions, these parameters allow the model to continuously update the level, trend, and seasonality at each time point. Adaptability to change trends in time series data, interpretability, and simplicity are some of the model's benefits. Nevertheless, the approach is sensitive to the initial values of the components and the choice of the smoothing parameters, requiring optimisation or experimentation.

c) Long Short-Term Memory (LSTM) recurrent neural network

Over the years, ML has become an effective alternative to traditional statistical methods for forecasting CPI due to its ability to manage enormous datasets, adapt to nonlinear relationships, and deal with complex time-series data. According to Barkan et al. (2023) and et al. (2018), Recurrent Neural Networks (RNNs) are a particular kind of artificial neural

network designed to analyse sequential input by identifying temporal relations. Time series forecasting, speech recognition and natural language processing, and other sequential data can be handled by RNNs (Javaid, 2019). Additionally, RNNs are suitable for time series analysis because of the repeating neural network modules, each of which processes one element of the input sequence while sustaining a hidden state representing data about prior elements. The hidden state acts as the memory of the network, allowing it to retain context from past inputs. The recurrent connection allows information to persist over time by relating the output of one module with the input of the next. Therefore, an RNN's output depends on the current input and all previous inputs (Javaid, 2019). RNNs excel at identifying patterns and relationships present in time series data, which makes them appropriate for predicting economic indicators (Paranhos, 2022).

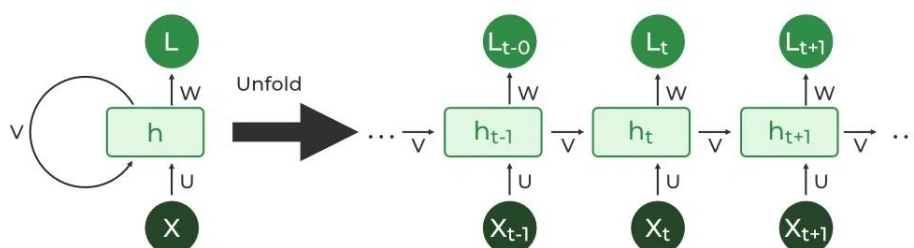


Figure 3.1. Recurrent Neural Network architecture

Adopted from <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network>, 2024

X_t represents the network's input at time t . The hidden state at time t is denoted by h_t , while the model's output is represented by L_t . It is calculated with the current input and the hidden state of the previous time step h_{t-1} and acts as the network's memory.

The equation of a basic RNN is defined as follows:

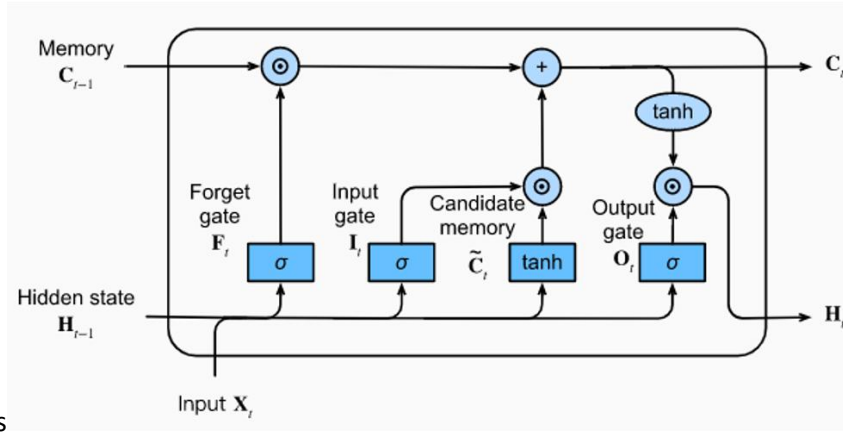
$$S(t) = f(U X_t + W h_{t-1} + b) \quad (\text{Eqn: 3.3})$$

The function f that enables the model to manage non-linearity can be any activation function, including tanh, sigmoid, ReLU, or any logistic function, which allows the model to handle

nonlinearity. The U and W are weight matrices for the input and the hidden state, which determine how much influence the input has on the current hidden state and control how the past hidden state influences the current hidden state, respectively. The parameter b allows the model to perform better on the fitted data by shifting the activation function (Barkan et al., 2023). The network has many layers connected by the parameters W , U , and B , and learns the same behaviours and correlations from the data irrespective of where they occur in the sequence because these parameters have weight for every time step (Nabi Javaid, 2019).

According to Miguel (2021), basic RNNs fall short in diminishing and expanding gradients, which limit their capacity to detect dependence over time. When the gradients of the loss function circulate backwards through time, they become too small, thereby causing the diminishing gradient issue. This makes it hard for the model to study from long-term dependencies, leading to poor performance on tasks that require the recall of information.

LSTMs are a type of RNN that can inherently capture dependencies that persist in the data by addressing the vanishing gradient (Miguel, 2021). LSTMs have a more complex memory than basic RNNs, which can handle complex non-stationarity and non-linear relationships, as well as perform automatic feature engineering. LSTM includes three kinds of gates: input, output, and forget gates. The input gate acts to control the data that streams in and out of the memory cell. According to Barkan et al. (2023), it is usually computed using a tanh activation function. The forget gate controls whether the current information can be kept in the memory or erased. The output gate regulates whether the output at the current time step should be affected by the memory. Therefore, the mixture of these gates permits LSTMs to keep critical data from long sequences and remove outdated information. These characteristics make LSTM networks particularly suited for forecasting tasks involving time series, such as predicting the CPI, where patterns and trends have similarly been shown to exist over longer periods of time (Miguel, 2021). Therefore, LSTM makes a viable candidate approach for improving CPI forecasting in Namibia. The architecture for the LSTM is as follows:



SS

Figure 3.2 The architecture for the LSTM recurrent neural network

Adopted from: https://d2l.ai/chapter_recurrent-modern/lstm.html

The batch size is n , the number of inputs is i , and there are h hidden units. The hidden state from the previous time step is $H_{t-1} \in \mathbb{R}^{n \times h}$, while the input is $X_t \in \mathbb{R}^{n \times i}$. The gates are defined as follows at time step t : $I_t \in \mathbb{R}^{n \times h}$ for the input gate, $F_t \in \mathbb{R}^{n \times h}$ for the forget gate, and for the output gate is $O_t \in \mathbb{R}^{n \times h}$.

The equations for the gates are computed as follows:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i), \quad (\text{Eqn: 3.4})$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f), \quad (\text{Eqn: 3.5})$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o), \quad (\text{Eqn: 3.6})$$

Where $b_i, b_f, b_o \in \mathbb{R}^{1 \times h}$ are bias parameters and $W_{hi}, W_{hf}, W_{ho} \in \mathbb{R}^{h \times h}$ and $W_{xi}, W_{xf}, W_{xo} \in \mathbb{R}^{i \times h}$ are weight parameters.

The input node is $\tilde{C}_t \in \mathbb{R}^{n \times h}$. It usually uses an activation function of tanh since tanh produces the outputs in the range of -1 and 1, which allows for smooth and continuous gradients and ensures stable gradients, and reduces the risk of exploding values over multiple time steps.

The function is defined as

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c), \quad (\text{Eqn: 3.7})$$

where $W_{xc} \in \mathbb{R}^{i \times h}$ and $W_{hc} \in \mathbb{R}^{h \times h}$ are weight parameters and $b_c \in \mathbb{R}^{1 \times h}$ is a bias parameter.

The input gate I_t regulates the quantity of new data that is taken into consideration, through \tilde{C}_t whereas the forget gate F_t determines how much of the old cell's internal state $C_{t-1} \in \mathbb{R}^{n \times h}$ is retained. Applying the Hadamard (elementwise) product operator \odot , the C_t denotes the memory cell internal state with the following equation:

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (\text{Eqn: 3.8})$$

The internal state of the memory cell, C_{t-1} , stays constant and is carried over to succeeding time steps when the forget gate is set to 1 and the input gate is set to 0. However, the input gate and forget gate allow the model to study when to hold this value constant and when to adjust it in response to future inputs. In reality, this architecture improves the diminishing gradient issue, thus making models considerably simpler to train, particularly when dealing with longitudinal data. Tanh is firstly applied to the internal state of the memory cell, followed by another application of the point-wise multiplication with the output gate, which guarantees the values of H_t to be in the interval $(-1,1)$.

$$H_t = O_t \odot \tanh(C_t) \quad (\text{Eqn: 3.9})$$

d) Support Vector Regression

Support Vector Regression is a supervised ML method derived from Support Vector Machines (SVM), often used for classification and regression tasks (Makridakis et al., 2018). SVR is a regression that determines a hyperplane that minimises the overall error under tolerance and maximises the margin between two sets of data (Makridakis et al., 2018). Furthermore, SVMs have gained attention for their capability to manage large datasets and accurately model nonlinear associations. Therefore, they are particularly useful when traditional methods fail to capture complex data patterns.

The SVR formula is as follows:

Given a training dataset (x_i, y_i) where $x_i \in R^n$ represents the input features and $y_i \in R$ represents the target variable CPI.

$$f(x) = w^T \phi(x) + b$$

Where:

$\phi(x)$ maps the input into a higher-dimensional feature space

w is the weight vector

b is the bias term

The objective is to minimize the following optimization problem:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi + \xi_i^*)$$

Subject to:

$$y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i$$

$$w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Where:

ε defines the margin of tolerance

ξ_i, ξ_i^* are slack variables allowing errors beyond the margin

$C > 0$ is a regularisation parameter controlling the trade-off between model complexity and tolerance for deviations larger than ε . The regularisation prevented the model from fitting every fluctuation in the training data.

SVR leverages the kernel trick to avoid overfitting and handle nonlinear relationships by mapping the input into a higher-dimensional feature space without explicitly computing the transformation (Smola & Scholkopf, 2004). The kernel function used in the study is the radial basis function kernel, denoted by:

$$K(x_i - x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Where γ controls the width of the Gaussian function.

3.7. Chapter summary

The chapter outlined the methodology used, which was based on Saunders's Research Onion model, to forecast Namibia's CPI. The study employed a positivist philosophy and a deductive approach, using a comparative strategy within a quantitative method design. Secondary data covering the period 2013–2023 were collected and pre-processed to align features with varying frequencies in a monthly time series format. Python was used to develop and assess forecasting models, ARIMA, Holt-Winters Exponential Smoothing, SVR, and LSTM. The performance of the model was compared using the RMSE. The study sought to determine the best model for enhancing CPI forecasting procedures in Namibia by combining modern machine learning techniques with traditional statistical methods. The following table summarises the research objectives, the activities undertaken to achieve them, and the evaluation criteria used to assess their outcomes:

Table 3.1: Summarised study methodology

| Research objective | Summary of research activities carried out | Evaluation criteria |
|--|---|----------------------------|
| To assess the performance of traditional statistical methods on the CPI data for Namibia | Implemented ARIMA and Holt-Winters models using Python Conducted exploratory data analysis and stationarity checks | RMSE was used |
| To assess the performance of machine learning methods on the CPI data for Namibia | Developed SVR and LSTM models Performed data normalisation, sequence generation, and hyperparameter tuning | RMSE was used |

| | | |
|---|--|--|
| <p>To compare the performance of traditional statistical and ML methods based on RMSE</p> | <p>Conducted the evaluation of models using RMSE Plotted forecasts vs. actual CPI values for visual comparison</p> | <p>The RMSE of the fitted models were compared to determine the model with the lowest RSME</p> |
| <p>To provide recommendations for improving CPI forecasting practices in Namibia</p> | <p>Provided suggestions depending on the performance of the modes</p> | <p>The improvements were given based on the performance of the models</p> |

CHAPTER FOUR: RESULTS AND DISCUSSION OF FINDINGS

This chapter provides an analysis of CPI trends in Namibia and presents forecasts using four different models: ARIMA, SVR, LSTM, and Holt's winter exponential smoothing. The study used monthly data spanning from January 2013 to December 2023, totalling 132 observations, sourced from the IFS website. The dependent variable, CPI, measures inflation trends in Namibia. Predictor variables included macroeconomic indicators such as interest rates, unemployment rate, GDP growth rate, real GDP, real government consumption expenditure, real imports of goods and services, and real exports of goods and services. These features were selected based on economic theory and their empirical relevance in influencing inflation dynamics.

4.1 Descriptive analysis

The average CPI is 131.43, which indicates the average price of goods and services in Namibia between the period of 2013 and 2023. The standard deviation of 18.60 shows moderate inflation over time. It indicates that inflation did not remain constant, but there were periods of higher or lower inflation, possibly due to changes such as COVID-19 in 2020 or commodity price swings.

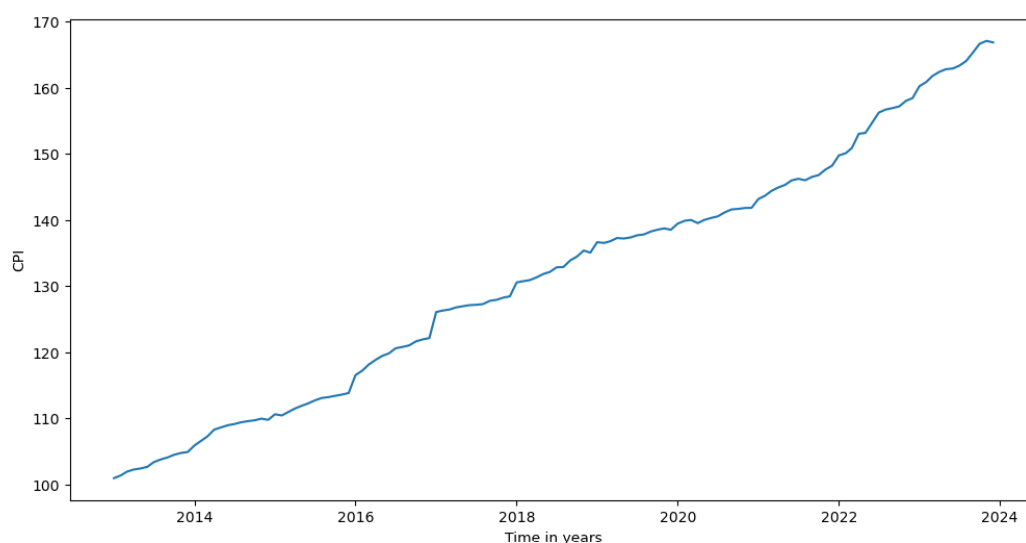


Figure 4.1: CPI Time Series for Namibia (2013-2023)

Figure 4.1 presents the time series plot of the Namibian CPI from 2013 to 2023. CPI ranged from 100.94 in 2013 to 167.05 in 2023, which indicates a steady increase in prices. The consistent upward trend indicates that prices of goods and services have steadily risen from 2013 to 2023. The graph is fairly smooth, with minor fluctuations that may represent seasonal variations or specific economic events impacting the price index. Overall, the cost of living in Namibia has increased over the years.

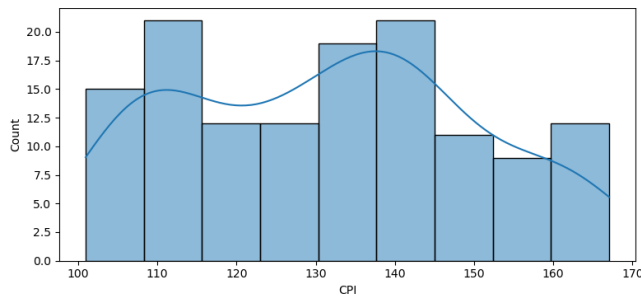


Figure 4.2: Histogram of CPI for Namibia (2013-2023)

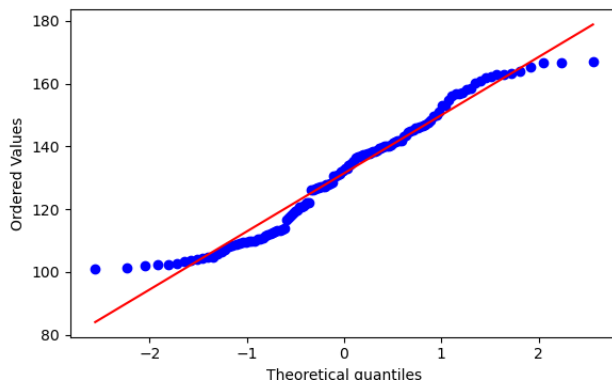


Figure 4.3: Q-Q Plot of CPI for Namibia (2013-2023)

This skewed distribution advocates for the use of forecasting models that are capable of capturing nonlinear patterns, such as machine learning models. Additionally, both the histogram and the Q-Q plot in Figure 4.3 show that CPI is not normally distributed.

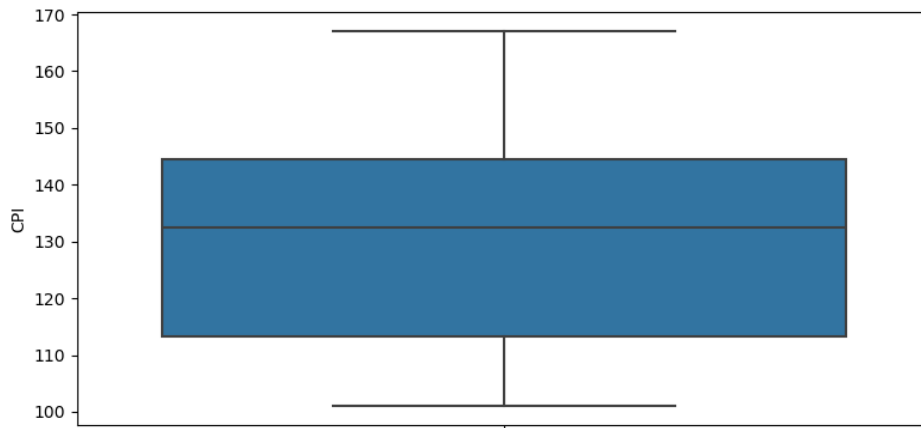


Figure 4.4: Boxplot of CPI for Namibia (2013-2023)

Figure 4.4 shows the boxplot of Namibia’s CPI from 2013 to 2023. The plot presents a compact distribution with no visible outliers beyond the whiskers. This indicates that CPI values were largely stable and constantly distributed over the period. The interquartile range is approximately 120 to 150, capturing the middle 50% of values, while the median lies around 132 to 135. The absence of outliers suggests that inflation increased gradually and steadily without extreme fluctuations.

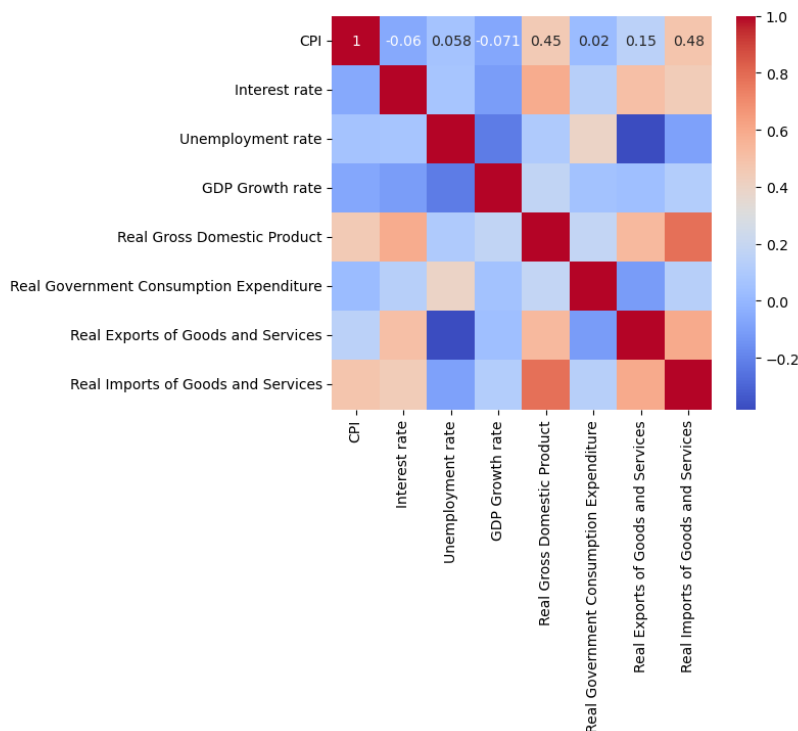


Figure 4.5 Correlation matrix of CPI and other economic indicators

Figure 4.5 displays insights into the linear relationships between CPI and selected economic indicators in Namibia over the period 2013–2023. Among the variables, real imports of goods and services show the strongest positive correlation with CPI, at 0.481, indicating that

increases in imports are moderately associated with rising consumer prices. Real GDP also determines a moderate positive correlation with CPI (0.455), suggesting that periods of economic growth tend to be accompanied by upward pressure on prices, likely due to increased consumer and production activity. Real exports of goods and services correlate positively with CPI at 0.149. Interest rates exhibit a very weak negative correlation with CPI (-0.060), suggesting that changes in interest rates had a limited inverse influence on inflation during the study period. The unemployment rate shows a weak positive correlation with CPI at 0.058, indicating minimal linear association. On the other hand, the GDP growth rate shows a weak negative correlation with CPI (-0.071), implying that a higher growth rate does not necessarily correspond with inflation increases. Lastly, real government consumption expenditure has a 0.020 correlation with CPI, suggesting that public sector spending had a minimal direct impact on consumer price levels during the observed period.

4.2. Forecasting CPI for Namibia using traditional statistical methods

4.2.1. ARIMA model

For the ARIMA model to perform well, the data needs to be stationary; therefore, the Augmented Dickey-Fuller (ADF) test is used to determine if the CPI series for Namibia from 2013 to 2023 is stationary. The findings indicated a p-value of 0.989, which is significantly greater than the 0.05 significance level. This shows that the CPI series is non-stationary, implying that its variance and mean fluctuate over time. To address this, first-order differencing is utilised on the CPI series as shown below in Figure 4.6.

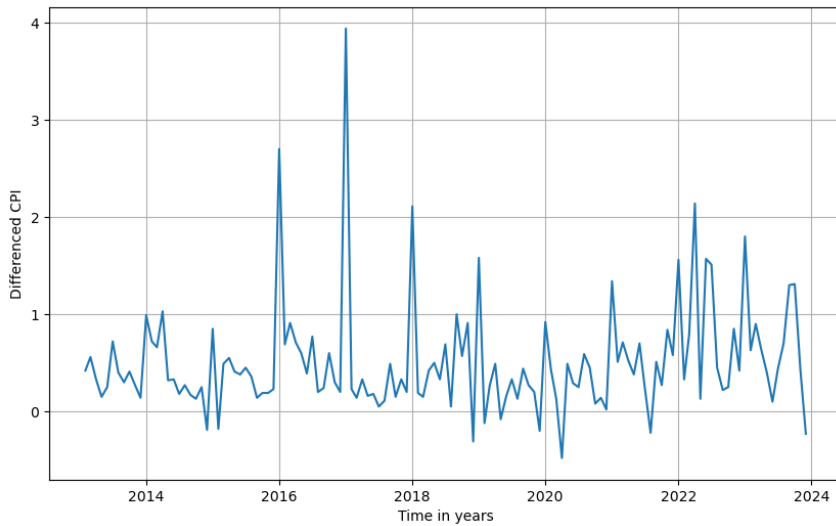


Figure 4.6: First-order differenced CPI of Namibia (2013–2023)

The ADF test for the first-order differencing yielded a p-value of 0.337, exceeding the 0.05 significance level.

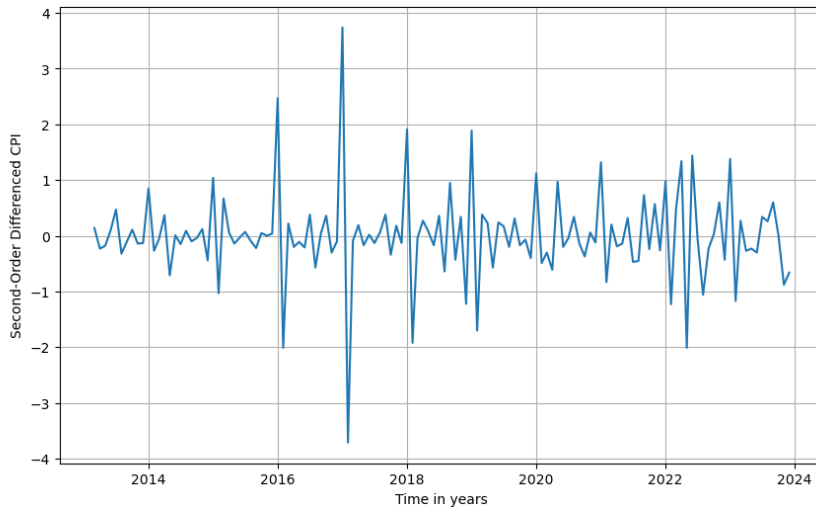


Figure 4.7: Second-order differenced CPI of Namibia (2013–2023)

The second differencing is performed to achieve stationarity, as shown in Figure 4.7, to allow for the elimination of the trend and seasonal patterns in the data, making the data stationary. The variation appeared to be more consistent than in Figure 4.6. The second differencing p-value is 0.00000138, less than the significance level of 0.05, indicating that the data is stationary. Therefore, the ARIMA model can be fitted with the data to forecast CPI.

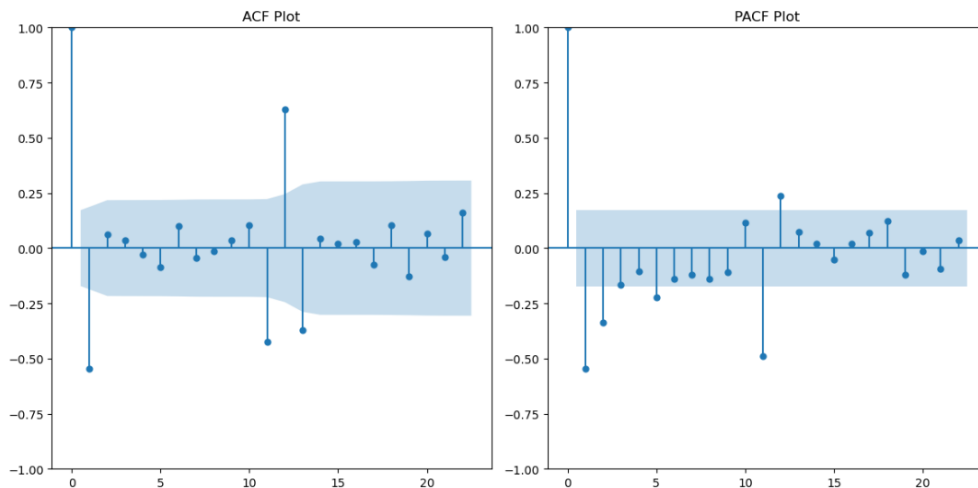


Figure 4.8: Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots

Figure 4.8 confirms that the second-order differencing was effective in achieving stationarity for the CPI series. Both plots display a significant spike at lag 0, and all subsequent autocorrelations fall well within the confidence interval.

ARIMA model summary

```

=====
Dep. Variable:          CPI    No. Observations:          92
Model:                 ARIMA(1, 1, 2)    Log Likelihood              -80.191
Date:                  Wed, 29 Jan 2025    AIC                        168.381
Time:                  17:37:31           BIC                        178.424
Sample:                01-01-2013         HQIC                       172.433
                    - 08-01-2020
Covariance Type:      opg
=====
              coef    std err          z      P>|z|    [0.025    0.975]
-----
ar.L1         1.0000    0.002   439.054    0.000    0.996    1.004
ma.L1        -1.0796    0.371   -2.910    0.004   -1.807   -0.353
ma.L2         0.0836    0.189    0.442    0.658   -0.287    0.454
sigma2        0.3275    0.096    3.399    0.001    0.139    0.516
=====
Ljung-Box (L1) (Q):          0.00    Jarque-Bera (JB):          1024.48
Prob(Q):                     0.95    Prob(JB):                   0.00
Heteroskedasticity (H):      1.96    Skew:                       3.24
Prob(H) (two-sided):         0.07    Kurtosis:                   18.10
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
Train RMSE: 10.539455875325547
Test RMSE: 4.745369454359933

```

The log-likelihood value of -80.191 indicates that the ARIMA model provides an adequate fit to the CPI data. Additionally, the variance of residuals is significant since it has a p-value of 0.001, which is less than the 0.05 threshold. Consequently, the model's estimated uncertainty in CPI fluctuations is reliable. The lower test RMSE compared to the training RMSE signifies that the model has good generalisation ability.

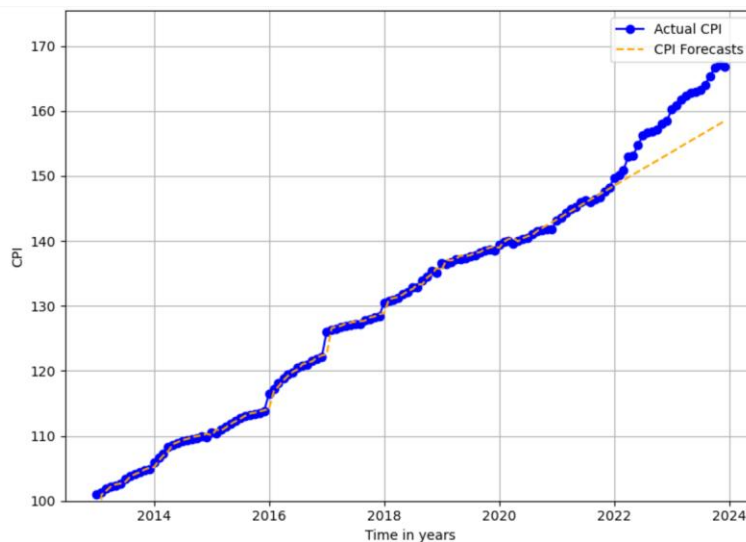


Figure 4.9: Forecasting CPI for Namibia using the ARIMA Model

Figure 4.9 presents the actual and predicted CPI values for Namibia from 2013 to 2023. The orange dashed line indicates the forecasted values, whereas the blue line represents the real CPI values. Throughout the training period, the model performed well, closely tracking the real CPI values. The predicted CPI does, however, deviate a bit from the real trend after 2022, which suggests an underestimation. The ARIMA model may not adequately account for fundamental changes in CPI drivers, which could be the cause of the underestimation throughout the test period (2022–2024). The linear trend assumption of ARIMA may constrain its capacity to represent complex, irregular patterns in the data. Adding more economic variables may also improve the precision of the model's long-term forecasts.

4.2.2. Holt-Winters Exponential Smoothing model

The Holt-Winters Exponential Smoothing model indicates a good fit with the training data, but the larger test RMSE suggests that it has trouble generalising, with a train RMSE of 0.378 and a test RMSE of 4.328. The low train RMSE indicates the model's capability to determine the seasonal and trend elements of the training dataset. The greater test RMSE, however, suggests that the model's forecasts differ more from the test set's actual CPI. This result highlights a limitation of classical time series models, as they rely heavily on the persistence of past seasonal and trend patterns. In Namibia's case, exogenous shocks, such as regional droughts and changes in global commodity prices, introduced irregular variations that Holt-

Winters was unable to capture. Similar findings are reported in other CPI forecasting studies, that Holt-Winters performs well in stable environments but deteriorates under structural breaks or inflationary shocks (Akinyemi & Olayemi, 2021; Fildes et al., 2018).

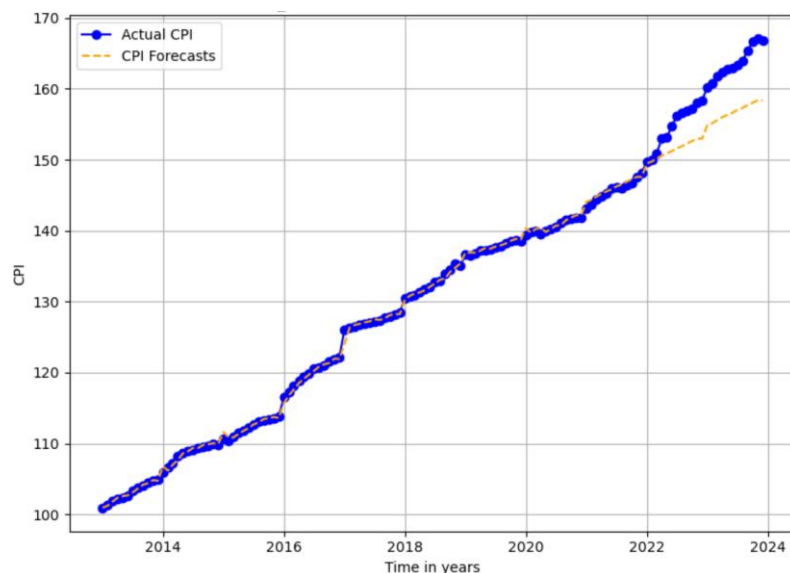


Figure 4.10: Forecasting CPI for Namibia using the Holt-Winters Exponential Smoothing model

The Holt-Winters Exponential Smoothing model seems to accurately track the actual CPI trend during training; however, it deviates from it throughout the anticipated period. The model accurately predicts the seasonality and trend until about 2022, at which point the predicted CPI falls short of the real values.

4.3. Forecasting CPI for Namibia using machine learning methods

4.3.1. LSTM recurrent neural network

The LSTM model was trained for 30 epochs, during which the loss function decreased from 0.0528 in the first epoch to 2.0648e-04, indicating that the model is learning well because a lower loss signifies better performance on the training data. The loss reached very low values (below 0.001) around epoch 10, which means that the model successfully learnt the fundamental trend in the data. After epoch 10, the loss continued to decrease, but at a slower rate, which is a positive indication that the model is stabilising without overfitting. The train

RMSE of 0.8235, which is quite low, indicates good performance of the model on the training data. The test RMSE of 1.2938 is slightly higher than the training RMSE but still close, which shows that the model generalises well.

Globally, LSTM models have been widely used in inflation and CPI forecasting. For instance, Zhang et al. (2020) found that LSTM outperformed ARIMA in forecasting Chinese CPI due to its ability to model nonlinear dependencies. Similarly, Khan et al. (2022) reported strong LSTM performance in Pakistan. However, studies also show that LSTM models often require larger datasets for optimal performance. Given the study's relatively small dataset (132 observations), the LSTM's performance in this study was good but not superior, suggesting that data scarcity constrained the deep learning model's full potential.

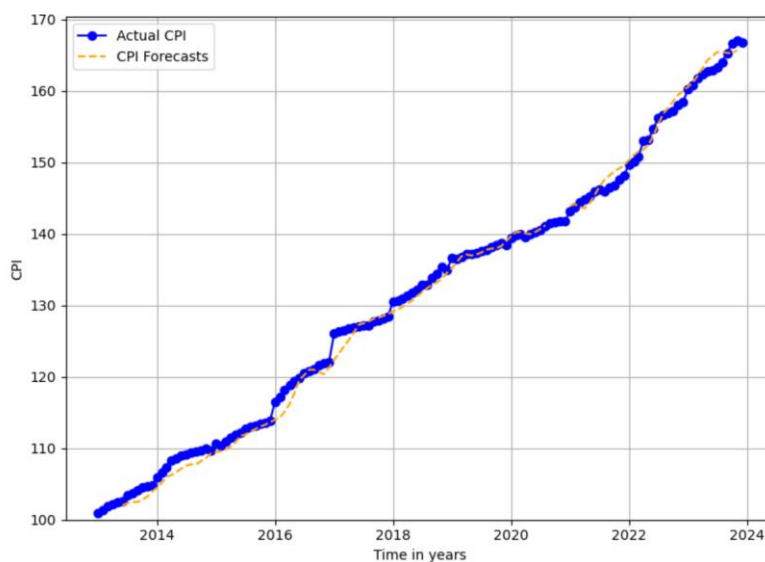


Figure 4.11: Forecasting CPI for Namibia using the LSTM model

Figure 4.11 shows the predicted CPI and the actual CPI over time; the blue dots indicate the actual CPI, and the dashed orange line predicts CPI values. The predicted CPI closely follows the actual CPI, which indicates that the model captures the general trend well. Some deviations occurred, especially between the periods of 2013 and 2018. Overall, the LSTM model performed well in capturing CPI trends with minimal deviation.

4.3.2. Support Vector Regression model

Strong predictive accuracy is demonstrated by the SVR model, which also demonstrates good generalisation to unseen data, with a test RMSE of 0.6737 and a low train RMSE of 0.5866. The model's small generalisation gap of 0.0871 indicates that it is not overfitting and maintains consistency across datasets, pointing to a well-balanced model that makes reliable forecasts with minimal errors. This aligns with findings from global and regional studies, which have consistently highlighted SVR's strengths in small-to-medium-sized datasets due to its reliance on kernel methods rather than deep data-hungry structures (Lu et al., 2019).

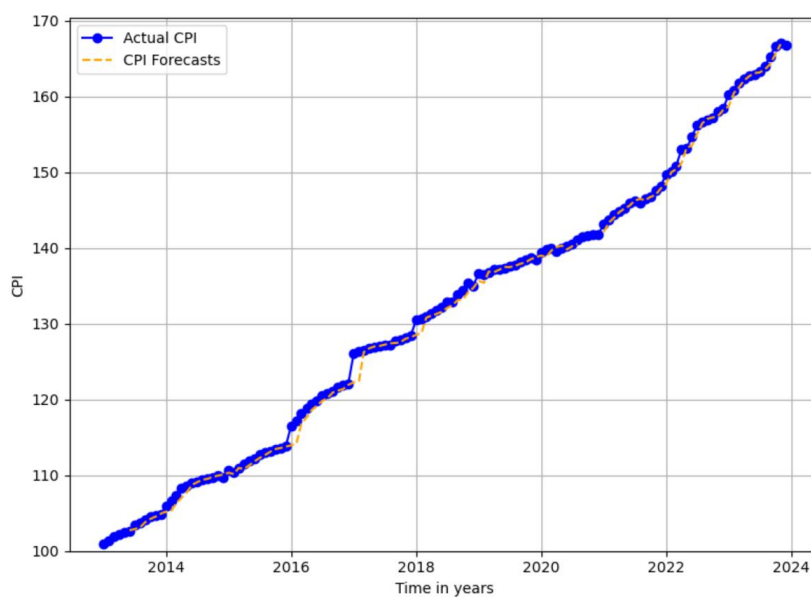


Figure 4.12: Forecasting CPI for Namibia using the SVR model

Figure 4.12 shows the actual versus forecasted CPI for Namibia over time. The actual CPI (blue solid line with dots) and forecast CPI (dashed orange line) follow a similar trend, which determines the model's ability to accurately capture CPI trends. Both lines show a steady upward trend, indicating consistent inflation over the years. The closeness of the forecasted values to the actual CPI highlights the model's accuracy, with only minimal deviations, which are expected in any forecasting approach. Specifically, certain periods, such as 2016 and 2021, show steeper increases in CPI, likely due to economic shocks or inflationary pressures. Overall, the predicted CPI aligned well with actual values, confirming that the model effectively predicted Namibia's CPI trends despite small variations influenced by unforeseen economic events.

4.4. Comparison of the traditional statistical methods and machine learning models

Table 4.1. Comparison of the forecasting models

| Model | Type | Train RSME | Test RSME |
|------------------------------------|-------------------------|------------|-----------|
| Support Vector Regression (SVR) | Machine Learning Method | 0.5866 | 0.6737 |
| LSTM | Machine Learning Method | 0.82 | 1.29 |
| ARIMA | Statistical Method | 10.87 | 4.74 |
| Holt-Winters Exponential Smoothing | Statistical Method | 0.378 | 4.328 |

The SVR model performed well, indicating a high degree of accuracy and capacity for generalisation. It also has the lowest RMSE values for both training (0.5866) and testing (0.6737) of the four models used in the study. The minimal difference between train and test RMSE indicates that the model is not overfitting. The Long LSTM model follows the SVR with a lower training RMSE (0.82) and a higher test RMSE (1.29), which means that it learnt well and can apply what it learnt. Seasonal trends are successfully identified by the Holt-Winters Exponential Smoothing model during training (RMSE of 0.378), but its higher test RMSE of 4.328 indicates that it would have difficulties generalising to new data. The model's assumptions on trend and seasonality, which may not hold true in the test sample, could be the cause of the disparity. ARIMA shows a very high RMSE on the training set (10.87), indicating that it struggled to fit the data accurately. The model suggested that it has difficulties in capturing the underlying format of the CPI data, possibly due to non-stationarity or nonlinear relationships.

These findings echo broader global evidence. In mature economies with stable inflationary patterns, ARIMA and exponential smoothing often remain competitive (Box et al., 2016). However, in emerging and developing economies where inflation drivers are volatile, ML

models, especially SVR, have increasingly been found superior (Adebiyi et al., 2021; Shahid et al., 2021).

4.5. Discussion of findings

The study aimed to forecast Namibia's CPI using a comparative analysis of traditional statistical methods (ARIMA and Holt-Winters exponential smoothing) and ML techniques (LSTM and SVR). The comparative performance of the models is assessed using the RMSE to determine model accuracy.

The study found that CPI in Namibia showed signs of non-stationarity and non-normality, as revealed by the Augmented Dickey-Fuller test and histogram analysis. These statistical properties validate the use of ML models, which do not require assumptions of stationarity or normality. Additionally, the correlation matrix and lag analysis highlighted significant relationships between CPI and macroeconomic variables of interest rate, GDP growth, unemployment rate, and government expenditure, showing the importance of incorporating multivariate inputs in forecasting.

Among the models tested, SVR achieved the best overall performance, showing excellent generalisation ability and high forecasting accuracy with the lowest RMSE values for both training and testing datasets. SVR's outstanding accuracy is primarily because of its capacity to model complex trends and non-linear relationships (Valkenborg et al., 2023). The model's strength is in its use of kernel functions, which translate the input features into higher-dimensional spaces to handle non-linear relationships between variables (Valkenborget al., 2023). Compared to traditional linear models, SVR can recognise and fit more complex trends in the data because of this transformation. Furthermore, by optimising the change between the expected and actual values, SVR avoids overfitting and is resistant to outliers, which improves its capacity for generalisation. These qualities are particularly beneficial in economic forecasting, as data is frequently noisy and subject to the influence of several related macroeconomic variables (Alsharif et al., 2020).

The LSTM model, while slightly less accurate than SVR, still demonstrated strong performance. Its capacity to retain long-term dependencies within sequences made it suitable for capturing temporal patterns in the CPI series. These results align well with existing literature that highlights the effectiveness of LSTM in handling non-linear and volatile economic data (Almosova & Andresen, 2019; Barkan et al., 2023; Zahara et al., 2020).

The Holt-Winters exponential smoothing model showed a very good fit with the training data but failed to generalise well with the test data, indicating poor predictive power for unseen data. This highlights the model's limitation in adapting to abrupt structural changes or nonlinear relationships. However, the model may still offer value for short-term forecasting in stable economic conditions.

The analysis showed that the ARIMA model performed poorly. According to Shinkarenko et al. (2021), ARIMA's ability to predict Namibia's CPI is hindered by its reliance on linear trends and stationary assumptions, especially in a dynamic economic environment. Moreover, ARIMA does not inherently handle seasonality (Ahmar et al., 2018).

All in all, the study found that ML methods (SVR and LSTM) outperformed traditional statistical methods (ARIMA and Holt-Winters) in terms of accuracy and flexibility. Such models offer enhanced reliability for economic planning in Namibia, where both global and domestic factors influence inflation. The findings also align with the emerging body of literature, including Barkan et al. (2023), Almosova and Andresen (2019), Zahara et al. (2020), Barkan et al. (2023), Almosova and Andresen (2019), Gur (2024), Barhoumi et al. (2022), and Kubheka (2023), who advocate for the integration of ML methods in forecasting, particularly in emerging economies. The superior performance of ML models in this study further confirms the shift in the forecasting research towards data-driven flexible models capable of navigating economic volatility and limited data environments.

4.6. Chapter summary

Chapter Four presents the results and discussion of the study, analysing Namibia's Consumer Price Index (CPI) trends and forecasting performance using four models, ARIMA, Holt-Winters exponential smoothing, LSTM and SVR, based on monthly data from 2013 to 2023. The descriptive analysis shows that Namibia's CPI increased steadily from 100.94 in 2013 to 167.05 in 2023, indicating a persistent rise in the cost of living. The data exhibited moderate variability and a non-normal, skewed distribution, suggesting non-linear patterns better captured by machine learning models. Correlation analysis revealed moderate positive relationships between CPI and real imports, real GDP, and exports, while interest rate and GDP growth rate showed weak negative correlations. In forecasting performance, the ARIMA model required second-order differencing to achieve stationarity and moderately fitted the data but underestimated CPI in later years due to its linearity and sensitivity to structural changes. The Holt-Winters model captured seasonal and trend components effectively during training but showed poor generalisation on test data, reflecting its limitations under volatile economic conditions. Among machine learning methods, the LSTM model demonstrated strong performance, with low RMSE values and the ability to learn temporal patterns, though its effectiveness was slightly constrained by the small dataset. The SVR model achieved the best results overall, showing the lowest RMSE for both training and testing, indicating high accuracy and excellent generalisation due to its ability to model non-linear relationships and resist overfitting. The comparative analysis confirmed that ML models, especially SVR, outperformed traditional statistical approaches, aligning with global research findings that advocate for ML adoption in forecasting complex and dynamic economic indicators. Overall, the results demonstrate that machine learning techniques provide more accurate and adaptable CPI forecasts for Namibia, offering valuable insights for policymakers, businesses, and investors in managing inflation and planning economic interventions.

CHAPTER FIVE: CONCLUSION

The purpose of the study was to compare the forecasting capability of ML methods with traditional statistical methods on CPI data for Namibia.

5.1. Reflecting on research questions

This section revisits the study's research questions and discusses how the findings respond to each, based on the analyses and model evaluations conducted.

i. What has been the trend in Namibia's CPI over the past ten years?

The CPI trend in Namibia has shown a steady upward trajectory since 2013. CPI rose gradually between 2013 and 2016, before accelerating in 2016 to 2018, reflecting drought impacts, currency depreciation, and higher import costs. Growth slowed in 2019 to 2020, partly due to reduced demand during the COVID-19 pandemic, before surging again in 2021 to 2022 as global supply chain disruptions, higher food and fuel prices, and the Russia-Ukraine conflict exerted inflationary pressures. By 2023, CPI remained elevated but began to stabilise, indicating possible effects of monetary and fiscal measures.

ii. How accurately do traditional statistical methods forecast the CPI for Namibia?

The study evaluated ARIMA and Holt-Winters exponential smoothing as traditional statistical methods. While ARIMA performed moderately with a test RMSE of 4.74, Holt-Winters exponential smoothing outperformed it with a lower test RMSE of 4.33. This suggests that, among traditional methods, Holt-Winters is more effective in capturing CPI trends in the Namibian context.

iii. How accurately do machine learning methods forecast the CPI for Namibia?

SVR and LSTM are used as ML methods. SVR yielded the best performance with a test RMSE of 0.67, while LSTM achieved a test RMSE of 1.29. The results confirm the potential of ML, particularly SVR, in handling the non-linear patterns present in CPI data.

5.2. Ethical considerations

The study was conducted honestly and transparently, upholding key ethical principles such as integrity, accountability, and respect for intellectual property. All methods were clearly documented to ensure that other researchers can understand, verify, and reproduce the findings. Importantly, the analysis was based entirely on secondary data obtained from publicly available and reputable sources (IMF). The use of publicly accessible data ensured that the models were developed and applied in a responsible and ethically sound manner, maintaining the highest standards of academic research.

5.3. Contribution of the study

The study contributes to a deeper understanding of CPI forecasting by offering a comparative evaluation of traditional statistical and modern ML methods in predicting the CPI for Namibia. It is among the first studies to apply and contrast methods such as ARIMA, Holt-Winters exponential smoothing, SVR, and LSTM recurrent networks utilising Namibian macroeconomic data. The findings provide empirical evidence on the relative accuracy and applicability of these models, with SVR demonstrating superior forecasting performance. From a practical perspective, the study provides beneficial advice to policymakers, economists, and data scientists in Namibia by identifying the most effective techniques for CPI forecasting. This can aid in improved inflation targeting, monetary policy formulation, and resource planning. Additionally, the study encourages the integration of ML approaches in official statistics and economic modelling in the country, paving the way for future innovations in data-driven economic analysis.

5.4. Limitations of the study

Despite its valuable contributions, the study has some limitations, particularly concerning the data used. The analysis was based on monthly macroeconomic data covering the period from 2013 to 2023, which, although sufficient for model training, may not fully capture long-term structural trends or rare economic shocks. Additionally, the dataset was limited in scope, focusing primarily on a narrow set of macroeconomic indicators. Important drivers of inflation, such as housing costs, exchange rates, wage growth, and energy prices, were

included due to data availability constraints. The exclusion of these variables may have affected the comprehensiveness and forecasting precision of the models. However, the outcomes indicate that the study still achieved strong predictive performance, demonstrating the feasibility and reliability of using both traditional and machine learning methods to forecast CPI in the Namibian context.

5.5. Future work

To further improve CPI forecasts, future studies may look into hybrid models that integrate the strengths of both statistical and ML approaches. Such approaches, including ensemble techniques, may better capture the linear and nonlinear patterns in CPI data. While the study successfully compared traditional statistical and ML models for CPI forecasting in Namibia, various areas remain open for future research. First, the reliability and precision of the model might be enhanced by including a wider range of economic and social indicators, such as exchange rates, commodity prices, household income, and worldwide inflation trends. Second, employing real-time datasets or extending the time period may offer an improved understanding of how models function in dynamic economic environments. Finally, using different performance metrics like Theil's U-statistic or Mean Absolute Scaled Error (MASE) could provide more detailed assessments of forecasting accuracy. These potential paths could enhance Namibia's economic forecasting capabilities and promote evidence-based policymaking even more. To improve forecasting accuracy and assist in the formulation of economic policy, the study recommended that capacity-building in data science and ML methodologies be undertaken, along with the establishment of real-time data infrastructure. Furthermore, CPI forecasting accuracy may be enhanced by incorporating economic variables that show strong relationships with the CPI. As shown in the correlation matrix in Chapter Four, Figure 4.5, variables such as the interest rate, GDP growth, unemployment rate, and government expenditure are significantly correlated with CPI movements. Additionally, extending the forecast period may offer greater policy planning value and early warning capabilities.

5.6. Closing remarks

In closing, the study has provided important information about the effectiveness of various forecasting methods for predicting Namibia's CPI. By systematically comparing traditional statistical models and modern ML approaches, the research has demonstrated the value of leveraging advanced analytical tools in understanding CPI dynamics. Given the results, it is advised that Namibia's forecasting efforts switch to machine learning techniques, especially the SVR model, because of their better performance. The findings not only serve as a foundation for upcoming research but also as a practical guide for policymakers and analysts seeking more accurate inflation forecasts. ML can improve economic forecasting and provide financial analysts and policymakers with a more accurate tool to predict inflation trends. Organisations that predict the state of the economy should fund analysts' training programmes to help them become proficient at using modern machine learning methods. By offering timely insights into CPI trends, machine learning-based real-time forecasting systems can improve decision-making. Therefore, the study reinforces the importance of evidence-based planning in economic planning and suggests that there must be continued innovation in the use of data science for national development.

References

- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 614342. <https://doi.org/10.1155/2014/614342>
- Ahmar, A. S., Gs, A. D., Listyorini, T., Sugianto, C. A., Yuniningsih, Y., Rahim, R., & Kurniasih, N. (2018). Implementation of the ARIMA(p,d,q) method to forecasting CPI Data using forecast package in R Software. *Journal of Physics: Conference Series*, 1028(1). <https://doi.org/10.1088/1742-6596/1028/1/012189>
- Akande et al. (2022). Predicting infation component drivers in Nigeria a stacked 2022. 10.1007/s34546-022-00384-2
- Akinyemi, O., & Olayemi, A. (2021). *Modelling inflation in Nigeria: A comparative analysis*. <https://www.bing.com/ck/a?!&p=83bf409b39b0801192380a44c34cb2f31158d2815c2e18f33e6995c171626367JmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9zY2hvbGFyLmdvb2dsZS5jb20vY2I0YXRpb25zP3VzZXI9VS01RnliOEFBQUFKJmhsPWVu>
- Al-Chalabi, H., Al-Douri, Y. K., & Lundberg, J. (2018). *Time series forecasting using ARIMA Model A case study of mining face drilling rig*. <https://www.bing.com/ck/a?!&p=6e39d2d223b6dfa8922c354c89ce42fa5aa97f1de90770350895b48fb7ba5238JmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cuZGI2YS1wb3J0YWwub3JnL3NtYXNoL2dldC9kaXZhmjoxMjY2MzM2L0ZVTEURVhUMDEucGRm>
- Almosova, A., & Andresen, N. (2019). *Nonlinear inflation forecasting with recurrent neural networks*. <https://www.bing.com/ck/a?!&p=71b784c247cc2e438803041c4c54ad08d9ad550ff65396b9fac771f34394be5cJmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9vbmVmbmVsaWJyYXJ5LndpbGV5LmNvbS9kb2kvMTAuMTAwMi9mb3luMjkwMT9tc29ja2lkPTMzMdRiODQ1MmM3ZDY4ZTUyODAyYUJmZjYmL2OTA1>
- Alsharif, M. H., Kelechi, A. H., Yahya, K., & Chaudhry, S. A. (2020). Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and research trends. *Symmetry*, 12(1), 88. <https://doi.org/10.3390/SYM12010088>
- Araujo, G. S., & Gaglianone, W. P. (2023a). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2), 100087. <https://doi.org/10.1016/j.latcb.2023.100087>
- Araujo, G. S., & Gaglianone, W. P. (2023b). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2), 100087. <https://doi.org/10.1016/j.latcb.2023.100087>
- Asian Economics, 74, 101317. <https://doi.org/10.1016/j.asieco.2021.101317>
- Auerbach, A. J., & Gale, W. (2023). *The Federal Budget Outlook*. <https://www.bing.com/ck/a?!&p=7aca14404a468dea7726d1ea2e2bf79b9abbc4ff283cce92e1321d368450d92dJmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304>

[b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9lbWwYmVya2VsZXkuZWR1L35hdWVYmFjaC8&ntb=1](https://www.bing.com/ck/a?!&&p=fd4927820a954b7878f0468c53377880efd927c62c5c644526eee201e50b2812JmldtHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9lbWwYmVya2VsZXkuZWR1L35hdWVYmFjaC8&ntb=1)

Bank of Namibia (2021). <https://www.bon.com.na/Bank/Monetary-Policy.aspx>

Barhoumi, K., Choi, S. M., Iyer, T., Li, J., Ouattara, F., Tiffin, A., & Yao, J. (2022). *Overcoming data sparsity: A machine learning approach to track the real-time impact of COVID-19 in Sub-Saharan Africa*, *WP/22/88*, May 2022. <https://www.bing.com/ck/a?!&&p=fd4927820a954b7878f0468c53377880efd927c62c5c644526eee201e50b2812JmldtHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cuaW1mLm9yZy9lbi9QdWJsaWNhdGlvbnMvV1AvSXNzdWVzLzlwMjlvMDUvMDcvU3ViLVNhaGFyYW4tQWZyaWNhLUVjb25vbWljLUFjdGl2aXR5LUdEUC1NYWNoaW5lLUxlYXJuaW5nLU5vd2Nhc3RpbmctQ09WSUQtMTktNTE3NjQ2>

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, N. (2023). Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*, 39(3), 1145–1162. <https://doi.org/10.1016/j.ijforecast.2022.04.009>

Ben B. (2024). *Forecasting for monetary policy making and communication at the Bank of England: A review* | Bank of England. Museum. <https://www.bankofengland.co.uk/independent-evaluation-office/forecasting-for-monetary-policy-making-and-communication-at-the-bank-of-england-a-review/forecasting-for-monetary-policy-making-and-communication-at-the-bank-of-england-a-review>

Botha, B., Burger, R., Kotzé, K., Rankin, N., & Steenkamp, D. (2023). Big data forecasting of South African inflation. *Empirical Economics*, 65(1), 149–188. <https://doi.org/10.1007/s00181-022-02329-y>

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Castillo-Martinez, L., Reis, R., Fazio, M., Ferreira, J. A., & Leonardi, E. (2024). *How do central banks control inflation? A guide for the perplexed*. <https://www.bing.com/ck/a?!&&p=4ee4e3520e54cd6a8a743ca4c72480c27e2886fbf68bd2baa3a92e5f47d0dc27JmldtHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cubHNlMmFjLnVrL0NGTS9hc3NldHMvcGRmL0NGTS1EaXNjdXNzaW9uLVBhcGVycy0yMDI0L0NGTURQMjAyNC0zMy1QYXBldi5wZGY>

Coibion, O., Gorodnichenko, Y., Kumar, S., & Pedemonte, M. (2020). Inflation expectations as a policy tool? *Journal of International Economics*, 124. <https://doi.org/10.1016/j.jinteco.2020.103297>

Dhamo, E., Puka, L., & Zaçaj, O. (2018, September 6). *Forecasting consumer price index (CPI) using time series models and multiple regression models (Albania Case Study)*. <https://doi.org/10.3846/bm.2018.51>

Doojav, G. O., & Narmandakh, B. (2025). Forecasting Inflation in Mongolia using machine learning. *International Economic Journal*. <https://doi.org/10.1080/10168737.2025.2491386>

Dubois, A., & Gadde, L. E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research*, 55(7), 553–560. [https://doi.org/10.1016/S0148-2963\(00\)00195-8](https://doi.org/10.1016/S0148-2963(00)00195-8)

- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10. <https://doi.org/10.1177/1847979018808673>
- Fernando, J. (2023, August 11). *Consumer Price Index (CPI) explained: What it is and how it's used*. <https://www.investopedia.com/terms/c/consumerpriceindex.asp>
- Fernando, J. (2024, January). *Consumer Price Index (CPI): What it is and how it's used*. <https://www.investopedia.com/terms/c/consumerpriceindex.asp>
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2008). *Effective forecasting and forecasting inflation using machine learning for an emerging economy*. (n.d.). <https://www.bing.com/ck/a?!&p=68a1bc3afd510d2a7559bde747d97336f4b40e620e1a24514aa31315a91321ajmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9yZXNlYXJjaHBvcnRhbC5iYXRoLmFjLnVrL2VuL3B1YmxpY2F0aW9ucy9lZmZlY3RpdmUtZm9yZW5kLWp1ZGdtZW50YWwtYWRqdXN0bWVudHMtYW4tZW1waXJpY2FsLWV2YQ>
- Gur Y. (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):469-513. DOI:10.3934/DSFE.2024020
- Hall, A. (2018). *Machine learning approaches to macroeconomic forecasting*. *Federal Reserve Bank of Kansas City Economic Review*. <https://www.bing.com/ck/a?!&p=0c091c9a5253bb4e858c1e20f70293e30d3bef73a8b8109aecd139a7195d2074JmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cuc2NpcnAub3JnL3JlZmVvZW5jZS9yZWZlcmVuY2VzcGFwZXJzP3JlZmVvZW5jZWlkPTM2Njc0ODQ>
- Imron, M., Utami, W. D., Khaulasari, H., & Armunanto, F. (2022). ARIMA model of outlier detection for forecasting Consumer Price Index (CPI). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 16(4), 1259–1270. <https://doi.org/10.30598/barekengvol16iss4pp1259-1270>
- International Labour Organization (2022). *Global Wage Report 2022–23 The impact of inflation and COVID-19 on wages and purchasing power*
- Javaid N. (2019). *Recurrent Neural Networks (RNNs). Towards Data Science*
- Khan, M. H., Shah, S. A., & Ullah, I. (2022). *Forecasting inflation in Pakistan using deep learning*. <https://www.bing.com/ck/a?!&p=e80714939db419930eb6872d8a5a96d6aea95f4ff4b4a201270367b5e7160929JmltdHM9MTc1OTk2ODAwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cucmVzZW5jZmVvZW5jZS9yZWZlcmVuY2VzcGFwZXJzP3JlZmVvZW5jZWlkPTM2Njc0ODQ>
- Konarasinghe, K. M. U. B. (2022). *Modeling consumer price index of Malaysia: Application of exponential smoothers*. www.imathm.edu.lk/publications
- Kubheka, S. (2023). South African inflation modelling using bootstrapped long short-term memory methods. *SN Business and Economics*, 3(7). <https://doi.org/10.1007/s43546-023-00490-9>

- Kumar, L., Khedlekar, S., & Khedlekar, U. K. (2024). A comparative assessment of Holt-Winters exponential smoothing and autoregressive integrated moving average for inventory optimization in supply chains. *Supply Chain Analytics*, 8. <https://doi.org/10.1016/j.sca.2024.100084>
- Lakshmanamoorthy, S., & Athinarayanan, S. (2024). Balancing act: Exploring the relationship between inflation and consumer behaviour. *Shanlax International Journal of Economics*, 12(2), 8–12. <https://doi.org/10.34293/economics.v12i2.7194>
- Lu, C. J., Lee, T. S., & Chiu, C. C. (2019). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 47(2), 115–125. <https://doi.org/10.1016/j.dss.2009.02.001>
- Madudova, E., & Corejova, T. (2024). The issue of measuring household consumption expenditure. *Economies*, 12(1). <https://doi.org/10.3390/economies12010009>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3). <https://doi.org/10.1371/journal.pone.0194889>
- Metu et al., (2024). Economic Planning Theory and Practice. <https://www.researchgate.net/publication/379651991>.
- Miguel, T. (2021, January 1). *How the LSTM improves the RNN: Towards Data Science*. <https://www.bing.com/ck/a?!&&p=d7fd087ffbb8635fb4e6e85dd143905af3622688ba45009bf79e6d2758478d8bJmltdHM9MTc2MDA1NDQwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly90b3dhcmRzZGF0eXNjaWVuY2UuY29tL2hvdj10aGUtbnN0bS1pbXB3Zlcy10aGUtcm5uLTFIZjE1NmI3NTEyMS8>
- Mishkin, F. S. (2018). *Housing and the monetary transmission mechanism*. *Applied Economics Letters*, 29(5), 409–413. <https://doi.org/10.1080/13504851.2020.1859409>
- Mohajan, H. (2020). *Munich personal RePEc archive quantitative research: A successful investigation in natural and social sciences*. <https://www.bing.com/ck/a?!&&p=305ec6b5ff1fbaf1bfae54f10b33b273d0daf7593da92d0564055e67943028c6JmltdHM9MTc2MDA1NDQwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9tcHJhLnViLnVuaS1tdWVuY2h1bi5kZS8xMDUxNDkv>
- Nabi, J. (2019, July 11). *Recurrent Neural Networks (RNNs): Towards Data Science*. <https://www.bing.com/ck/a?!&&p=b30e0a2731a852b4e11d00bf7168d8af1098f2f5b4a8f5b9efc4556f028a7be1JmltdHM9MTc2MDA1NDQwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly93d3cucmVzZWZyY2hnYXRlM5ldC9wdWJsaWNhdGlvbj8zODM0NDk2NDIfUmVjdXJyZW50X05ldXJhbF90ZXR3b3Jrc19BX0NvbXB3Zlcy10aGUtbnN0bS1pbXB3Zlcy10aGUtcm5uLTFIZjE1NmI3NTEyMS8>
- Namibia Statistics Agency (2025). Namibia Statistics Agency. <https://nsa.org.na/>
- Organisation for Economic Co-operation and Development (2025), <https://www.data4sdgs.org/partner/organisation-economic-co-operation-and-development>.
- Paranhos, L. (n.d.). *Predicting inflation with recurrent neural networks*. <https://www.bing.com/ck/a?!&&p=9aaa96e2bcf7d8e27f05bd82b6d8fc28ad0fbc83bc2f8f93e16bb156f032680aJmltdHM9MTc2MDA1NDQwMA&ptn=3&ver=2&hsh=4&fclid=3304b845-2c7d-68e5-2802-abff2dbb6905&u=a1aHR0cHM6Ly9hcnhpdj5vcmcvYWJzLzlxMDQuMDM3NTc>

Appendix: Python's implementation details

The appendix provided a detailed overview of the Python code and tools used to develop, train and evaluate the forecasting models of the study.

Import Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
from statsmodels.tsa.seasonal import seasonal_decompose
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.layers import Input, Dense, LSTM, Dropout
import tensorflow as tf
# Set random seeds
# Set random seeds for reproducibility
seed_value = 123
np.random.seed(seed_value)
tf.random.set_seed(seed_value)
```

Data Loading and Preprocessing

```
# Loading Data:
data = pd.read_csv(r'C:\Users\LElago\Downloads\Monthly1.csv').dropna()
data.head()
# Handling missing values
data = data.dropna()
# Date conversion
data['Date'] = pd.to_datetime(data['Year'].astype(str) + '-' + data['Month'].astype(str) + '-01')
desc_stats = data.describe()
desc_stats
plt.figure(figsize=(8, 4))
sns.histplot(data['CPI'], kde=True)
plt.title("Histogram of CPI")
plt.xlabel("CPI")
```

```

plt.tight_layout()
plt.show()
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import scipy.stats as stats
# Q-Q Plot for CPI
plt.figure(figsize=(6, 4))
stats.probplot(data['CPI'], dist="norm", plot=plt)
plt.title("Q-Q Plot of CPI")
plt.tight_layout()
plt.show()
# Boxplot for outlier detection
plt.figure(figsize=(8, 4))
sns.boxplot(y=data['CPI'])
plt.title("Boxplot of CPI")
plt.tight_layout()
plt.show()
# Setting Date as the index
data.set_index('Date', inplace=True)
# Dropping the original Year and Month columns
data.drop(['Year', 'Month'], axis=1, inplace=True)
# Display the cleaned data
data.head()
import seaborn as sns
import matplotlib.pyplot as plt

```

Correlation and Visualization

```

# Calculate the correlation matrix
correlation_matrix = data.corr()
# Plot the correlation matrix
plt.figure(figsize=(6, 5))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
# Plotting the CPI over time
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['CPI'])
plt.title('Consumer Price Index Over Time')
plt.xlabel('Time in years')

```

```
plt.ylabel('CPI')
plt.show()
```

Stationarity Testing

```
# Apply first-order differencing
cpi_diff = data['CPI'].diff().dropna()
# Plot the differenced series
plt.figure(figsize=(10, 6))
plt.plot(cpi_diff)
plt.title('Differenced CPI Series')
plt.xlabel('Time in years')
plt.ylabel('Differenced CPI')
plt.grid(True)
plt.show()
# Perform ADF test again
adf_test_diff = adfuller(cpi_diff)
print(f"ADF Statistic (Differenced): {adf_test_diff[0]}")
print(f"p-value (Differenced): {adf_test_diff[1]}")
print(f"Critical Values (Differenced): {adf_test_diff[4]}")
#Since the p-value is greater than 0.05, the series is non-stationary. Therefore, second differencing is
performed to achieve stationarity
# Apply second-order differencing
cpi_diff2 = cpi_diff.diff().dropna()
# Plot the second-order differenced series
plt.figure(figsize=(10, 6))
plt.plot(cpi_diff2)
plt.title('Second-Order Differenced CPI Series')
plt.xlabel('Time in years')
plt.ylabel('Second-Order Differenced CPI')
plt.grid(True)
plt.show()
# Perform ADF test again on the second-order differenced series
adf_test_diff2 = adfuller(cpi_diff2)
adf_result_diff2 = {
    'ADF Statistic': adf_test_diff2[0],
    'p-value': adf_test_diff2[1],
}
adf_result_diff2
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
```

```

import matplotlib.pyplot as plt
# Assuming 'cpi_diff2' is your second-order differenced CPI series
plt.figure(figsize=(12, 6))
# Plot ACF
plt.subplot(121)
plot_acf(cpi_diff2, ax=plt.gca())
plt.title('ACF Plot')

# Plot PACF
plt.subplot(122)
plot_pacf(cpi_diff2, ax=plt.gca())
plt.title('PACF Plot')
plt.tight_layout()
plt.show()
# Since the p-value is less 0.05, the null hypothesis can be rejected and the data can be concluded that it is
stationary. Therefore, ARIMA model can be fit on the data.

```

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error

```

ARIMA Model

```

# Define the actual CPI series
actual = data["CPI"] # Adjust column name if necessary
# Split the data into training and testing sets
train_size = int(len(actual) * 0.7)
train_data, test_data = actual[:train_size], actual[train_size:]
# Fit ARIMA model (adjust order as needed)
arima_model = ARIMA(train_data, order=(1, 1, 1))
arima_result = arima_model.fit()
# Print ARIMA model summary before plotting
print(arima_result.summary())
# Forecast on training and testing data
train_predict = arima_result.predict(start=train_data.index[0], end=train_data.index[-1], dynamic=False)
test_predict = arima_result.predict(start=test_data.index[0], end=test_data.index[-1], dynamic=True)
# Calculate RMSE for train and test sets
train_rmse = np.sqrt(mean_squared_error(train_data, train_predict))
test_rmse = np.sqrt(mean_squared_error(test_data, test_predict))

```

```

print(f"Train RMSE: {train_rmse}")
print(f"Test RMSE: {test_rmse}")
# Combine actual and forecasted values for plotting
CPI_Forecasts = pd.Series(np.nan, index=actual.index)
CPI_Forecasts.loc[train_predict.index] = train_predict
CPI_Forecasts.loc[test_predict.index] = test_predict
# Plot actual vs forecasted CPI
plt.figure(figsize=(8, 6))
plt.plot(actual.index, actual, label="Actual CPI", color="blue", marker='o', linestyle='-')
plt.plot(actual.index, CPI_Forecasts, label="CPI Forecasts", color="orange", linestyle='--')
plt.xlabel("Time in years")
plt.ylabel("CPI")
plt.title("Actual vs CPI Forecasts CPI (from 2013 onward)")
plt.legend()
plt.ylim(100) # Adjust y-axis as needed
plt.grid(True)
plt.tight_layout()
plt.show()

```

LSTM Model

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, Input
# Feature scaling
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data)
# Split the data into training and testing sets
train_size = int(len(data) * 0.7)
train_data, test_data = scaled_data[:train_size], scaled_data[train_size:]
# Prepare the data for LSTM
def create_dataset(data, time_step=1):
    X, Y = [], []
    for i in range(len(data) - time_step - 1):
        a = data[i:(i + time_step), :]
        X.append(a)

```

```

        Y.append(data[i + time_step, 0])
    return np.array(X), np.array(Y)

time_step = 4
X_train, y_train = create_dataset(train_data, time_step)
X_test, y_test = create_dataset(test_data, time_step)

# Reshape input for LSTM
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])

# Build the LSTM model
model = Sequential()
model.add(Input(shape=(time_step, X_train.shape[2])))
model.add(LSTM(50, return_sequences=True))
model.add(LSTM(50, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))

# Compile the model
model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
model.fit(X_train, y_train, batch_size=3, epochs=30)

# Make predictions
train_predict = model.predict(X_train)
test_predict = model.predict(X_test)

# Reverse scaling
train_predict = scaler.inverse_transform(np.concatenate((train_predict, np.zeros((train_predict.shape[0],
scaled_data.shape[1] - 1))), axis=1))[:, 0]
test_predict = scaler.inverse_transform(np.concatenate((test_predict, np.zeros((test_predict.shape[0],
scaled_data.shape[1] - 1))), axis=1))[:, 0]
y_train_actual = scaler.inverse_transform(np.concatenate((y_train.reshape(-1, 1), np.zeros((y_train.shape[0],
scaled_data.shape[1] - 1))), axis=1))[:, 0]
y_test_actual = scaler.inverse_transform(np.concatenate((y_test.reshape(-1, 1), np.zeros((y_test.shape[0],
scaled_data.shape[1] - 1))), axis=1))[:, 0]

# RMSE calculation
train_rmse = np.sqrt(mean_squared_error(y_train_actual, train_predict))
test_rmse = np.sqrt(mean_squared_error(y_test_actual, test_predict))
print(f"Train RMSE: {train_rmse}")
print(f"Test RMSE: {test_rmse}")

# Generate forecast array for plotting
CPI_Forecasts = np.empty_like(actual)
CPI_Forecasts[:] = np.nan

```

```

CPI_Forecasts[time_step + 1:len(train_predict) + time_step + 1] = train_predict
CPI_Forecasts[len(train_predict) + 2 * time_step + 1:len(train_predict) + 2 * time_step + 1 + len(test_predict)]
= test_predict
# Plot actual vs forecasted CPI
dates = data.index
plt.figure(figsize=(8, 6))
plt.plot(dates, actual, label="Actual CPI", color="blue", marker='o', linestyle='-')
plt.plot(dates, CPI_Forecasts, label="CPI Forecasts", color="orange", linestyle='--')
plt.xlabel("Time in years")
plt.ylabel("CPI")
plt.title("Actual vs CPI Forecasts (from 2013 onward)")
plt.legend()
plt.ylim(100) # Set the y-axis to start at 100
plt.grid(True)
plt.tight_layout()
plt.show()

```

SVR Model

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.svm import SVR
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
# Feature scaling
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data)
# Split the data into training and testing sets
train_size = int(len(data) * 0.7)
train_data, test_data = scaled_data[:train_size], scaled_data[train_size:]
# Prepare the data for SVR
def create_dataset(data, time_step=1):
    X, Y = [], []
    for i in range(len(data) - time_step - 1):
        a = data[i:(i + time_step), 0] # Use only the CPI column
        X.append(a)
        Y.append(data[i + time_step, 0])
    return np.array(X), np.array(Y)
time_step = 4

```

```

X_train, y_train = create_dataset(train_data, time_step)
X_test, y_test = create_dataset(test_data, time_step)
# Reshape input for SVR (SVR does not require 3D input like LSTM)
X_train = X_train.reshape(X_train.shape[0], time_step)
X_test = X_test.reshape(X_test.shape[0], time_step)
# Build the SVR model
svr_model = SVR(kernel='rbf', C=10000, gamma=0.001, epsilon=0.0000001)
# Train the model
svr_model.fit(X_train, y_train)
# Make predictions
train_predict = svr_model.predict(X_train)
test_predict = svr_model.predict(X_test)
# Reverse scaling
train_predict = scaler.inverse_transform(np.concatenate((train_predict.reshape(-1, 1),
                                                         np.zeros((train_predict.shape[0], scaled_data.shape[1] - 1))), axis=1))[:, 0])
test_predict = scaler.inverse_transform(np.concatenate((test_predict.reshape(-1, 1),
                                                         np.zeros((test_predict.shape[0], scaled_data.shape[1] - 1))), axis=1))[:, 0])
y_train_actual = scaler.inverse_transform(np.concatenate((y_train.reshape(-1, 1),
                                                           np.zeros((y_train.shape[0], scaled_data.shape[1] - 1))), axis=1))[:, 0])
y_test_actual = scaler.inverse_transform(np.concatenate((y_test.reshape(-1, 1),
                                                         np.zeros((y_test.shape[0], scaled_data.shape[1] - 1))), axis=1))[:, 0])
# RMSE calculation
train_rmse = np.sqrt(mean_squared_error(y_train_actual, train_predict))
test_rmse = np.sqrt(mean_squared_error(y_test_actual, test_predict))
print(f"Train RMSE: {train_rmse}")
print(f"Test RMSE: {test_rmse}")
# Generate forecast array for plotting
CPI_Forecasts = np.empty_like(actual)
CPI_Forecasts[:] = np.nan
CPI_Forecasts[time_step + 1:len(train_predict) + time_step + 1] = train_predict
CPI_Forecasts[len(train_predict) + 2 * time_step + 1:len(train_predict) + 2 * time_step + 1 + len(test_predict)]
= test_predict
# Plot actual vs forecasted CPI
dates = data.index
plt.figure(figsize=(8, 6))
plt.plot(dates, actual, label="Actual CPI", color="blue", marker='o', linestyle='-')
plt.plot(dates, forecasted, label="CPI Forecasts", color="orange", linestyle='--')
plt.xlabel("Time in years")
plt.ylabel("CPI")

```

```

plt.title("Actual vs CPI Forecasts (from 2013 onward)")
plt.legend()
plt.ylim(100) # Set the y-axis to start at 100
plt.grid(True)
plt.tight_layout()
plt.show()
import numpy as np
import pandas as pd
from sklearn.metrics import mean_squared_error
from statsmodels.tsa.holtwinters import ExponentialSmoothing
import matplotlib.pyplot as plt
# Split data into training and testing sets
train_size = int(len(data) * 0.7)
train_data, test_data = data.iloc[:train_size], data.iloc[train_size:]
Holt-Winters' Exponential Smoothing model
hw_model = ExponentialSmoothing(train_data['CPI'], seasonal='add', seasonal_periods=12, trend='add')
hw_fit = hw_model.fit()
# Forecast on training and testing data
train_predict = hw_fit.fittedvalues
test_predict = hw_fit.forecast(steps=len(test_data))
# RMSE calculation
train_rmse = np.sqrt(mean_squared_error(train_data['CPI'], train_predict))
test_rmse = np.sqrt(mean_squared_error(test_data['CPI'], test_predict))
print(f"Train RMSE: {train_rmse}")
print(f"Test RMSE: {test_rmse}")
# Create forecasted CPI for plotting
CPI_Forecasts = pd.Series(index=data.index, dtype='float64')
CPI_Forecasts[:len(train_predict)] = train_predict # Add training predictions
CPI_Forecasts[len(train_data):] = test_predict # Add testing predictions
# Plot actual vs forecasted CPI
plt.figure(figsize=(8, 6))
plt.plot(data.index, data['CPI'], label="Actual CPI", color="blue", marker='o', linestyle='-')
plt.plot(CPI_Forecasts.index, CPI_Forecasts, label="CPI Forecasts", color="orange", linestyle='--')
plt.xlabel("Time in years")
plt.ylabel("CPI")
plt.title("Actual vs CPI_Forecasts (Holt-Winters' Exponential Smoothing)")
plt.legend()
plt.grid(True)
plt.tight_layout()

```