ADDIS ABABA UNIVERSITY
COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

LONG-TERM INFLATION TREND PREDICTION IN ETHIOPIA USING LSTM AND ARIMA ENSEMBLE MODEL

BY
EYASU DESTA

ADVISOR
Dr. BISRAT DEREBSSA

A thesis submitted to the School of Electrical and Computer Engineering in partial fulfillment of the requirements for the Degree of Master of Science in Computer Engineering

ADDIS ABABA UNIVERSITY
COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

The undersigned have examined the thesis titled:

LONG-TERM INFLATION TREND PREDICTION IN ETHIOPIA USING LSTM AND ARIMA ENSEMBLE MODEL

BY

EYASU DESTA

Approval by Boards of Examiners

Dr. Sosina Mengistu ____________	____________	____________
Interim Head, SECE	Date	Signature
Dr. Bisrat Derebssa ____________	____________	____________
Advisor	Date	Signature
____________	____________	____________
Internal Examiner	Date	Signature
____________	____________	____________
External Examiner	Date	Signature

Declaration

I, Eyasu Desta Abule, declare that this thesis is my original work. All sources of information in this study have been appropriately acknowledged. I further confirm that this thesis has not been submitted either in part or in full for any other requirements to any other learning institution.

Declared By:

____

Student's Name and Signature

Approved By:

____

Advisor's Name and Signature

MARCH, 2025

Acknowledgements

I begin by offering my heartfelt gratitude to the Almighty GOD for His countless blessings and boundless grace. His divine guidance, unwavering protection, and unconditional love have been my constant source of strength throughout my journey, leading me to reach this significant milestone.

I would like to express my deepest gratitude to my advisor, Dr.Bisrat Derebssa, for their continuous support, insightful guidance, and invaluable feedback throughout the course of this research. Their encouragement and expertise were instrumental in shaping the direction of this work.

I am also grateful to my colleagues and friends for their constructive discussions and suggestions, which have been a source of inspiration and motivation. Special thanks go to Bizuayehu Samuel from National Bank of Ethiopia for providing assistance with data collection, without which this research would not have been possible.

I would like to acknowledge the support and resources provided by Central Statistical Agency, which played a critical role in the completion of this thesis.

Finally, I extend my heartfelt thanks to my family for their endless support, patience, and understanding during this challenging but rewarding journey. Their belief in me has been my greatest source of strength.

Abstract

Accurate inflation forecasting is crucial for economic stability, influencing policy formulation, financial planning, and market predictions. In Ethiopia, inflation dynamics are shaped by complex, interdependent factors, including macroeconomic indicators and sudden economic shocks such as civil war and drought. Traditional methods like ARIMA excel at capturing linear trends but struggle with non-linearities and external influences. Conversely, machine learning approaches like LSTM neural networks effectively model non-linear dynamics but often require extensive data and complex parameter tuning, limiting their applicability in data-scarce environments. To address these limitations, this study introduces an ensemble forecasting model that integrates ARIMA and LSTM, leveraging the strengths of both methodologies. The ensemble model uses historical economic data from Ethiopia spanning 1979 to 2023, incorporating key features such as money supply, real GDP, government investment, interest rates, exchange rates, and binary event flags representing sudden economic shocks. By combining ARIMA's proficiency in linear trend detection with LSTM's ability to model complex, nonlinear relationships, the ensemble approach achieves superior predictive accuracy. The study evaluates the model's performance under varying conditions, including different historical data lengths for training and testing, ensuring adaptability to diverse data availability scenarios. Additionally, the ensemble model handles multiple prediction horizons, providing reliable forecasts for inflation trends. This flexibility enhances its utility for policymakers and economists. The inclusion of binary event flags for civil war and drought ensures the model accounts for sudden disruptions, which significantly impact inflation in developing economies. Results indicate a substantial improvement in accuracy, with the ensemble model achieving 97.77% accuracy, outperforming standalone ARIMA (78.32%) and LSTM (89.01%). This highlights the model's effectiveness in capturing both linear and nonlinear patterns while addressing external shocks. The findings underscore the ensemble model's potential as a robust tool for economic forecasting and policymaking, facilitating informed fiscal strategies. Future research could explore further customization, integration of additional macroeconomic variables, and enhanced responsiveness to economic shocks, ensuring even greater predictive capability.

Keywords: Inflation forecasting, ARIMA, LSTM, Ensemble model, Economic prediction, Time series analysis, Sudden economic shocks, Historical data lengths, Prediction horizons

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ARDL	Autoregressive Distributed Lag
ARIMA	Autoregressive Integrated Moving Average
ARIMA-LSTM-CF	ARIMA-LSTM Combined Framework
CEA	Central Statistical Agency
CEEMDAN	Complete Ensemble Empirical Mode Decomposition Adaptive Noise
CNN	Convolutional Neural Network
CPI	Consumer Price Index
CPU	Central Processing Unit
ConvLSTM	Convolutional LSTM
ER	Exchange Rate
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GDP	Gross Domestic Product
GIRA	General Inflation Rate
GPU	Graphics Processing Unit
IQR	Interquartile Range
IMF	International Monetary Fund
LIR	Lending Interest Rate
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MS	Market Share
MS-AR	Markov-Switching Autoregressive Model
MSE	Mean Squared Error
NBE	National Bank of Ethiopia
NN	Neural Network
PACF	Partial Autocorrelation Function
RAM	Random Access Memory
RGDP	Real Gross Domestic Product
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SARIMA	Seasonal ARIMA
VAE	Variational Autoencoder

Chapter One

1 Introduction

1.1 Background

Inflation is a continuous increase in the general price level of goods and services over time, eroding the purchasing power of money and posing significant economic challenges as shown on Figure 1. The consequences of inflation are far-reaching, impacting individuals' real income, businesses' profitability, and overall economic stability[1].

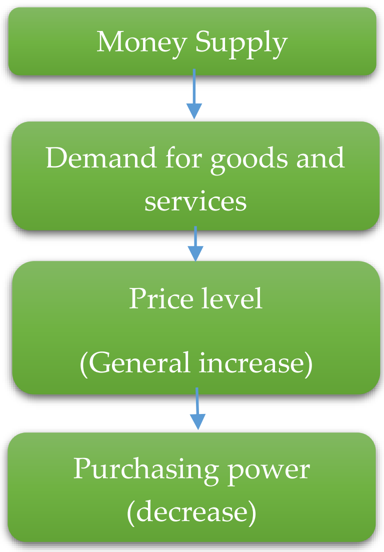


Figure 1 Diagram illustrating Inflation
Source: Adapted from [1]

Ethiopia is one of the fastest growing economies in Africa, but it also has one of the highest inflation rates. In 2022, the annual inflation rate in Ethiopia was 33.9%. The high inflation rate is a major challenge for the Ethiopian economy, and it is important to develop accurate and reliable methods for forecasting inflation. Inflationary pressures in Ethiopia have fluctuated over the years, often driven by factors such as supply shocks, fiscal imbalances, and external economic conditions. These fluctuations have had a notable impact on the country's economic performance and the well-being of its citizens [2]. Notably, these fluctuations also include inflation caused by unseen economic events such as civil war and drought, which were incorporated into this study using monthly inflation data.

The National Bank of Ethiopia, the central bank responsible for monetary policy, relies heavily on inflation forecasts to guide its decisions regarding interest rates. Adjusting interest rates is a key tool for managing inflation. By raising interest rates, NBE can reduce the money supply and dampen economic activity, thereby moderating inflation. Conversely, lowering interest rates can stimulate economic activity and push inflation higher[3].

Traditional inflation forecasting methods, such as the ARIMA model, have been widely used for decades in Ethiopia, specifically in National Bank of Ethiopia for economic modeling and analysis purpose [34]. These models rely on statistical techniques to identify patterns and trends in historical inflation data and extrapolate those trends into the future. Like other traditional linear models, they may not adequately capture the complex and often nonlinear relationships that characterize inflation dynamics. This limitation becomes particularly problematic when dealing with real-world economic data, where inflation is influenced by a multitude of interacting factors. Moreover, ARIMA models do not take into account the impact of other economic variables on inflation. For example, ARIMA models fail to incorporate the effects of interest rates, exchange rates, or the unemployment rate on inflation. As a result, these models may produce less accurate forecasts, especially in dynamic or volatile economic environments. The inability to account for these factors can lead to significant forecast errors, which may misguide monetary policy decisions and hinder effective economic management [4,5].

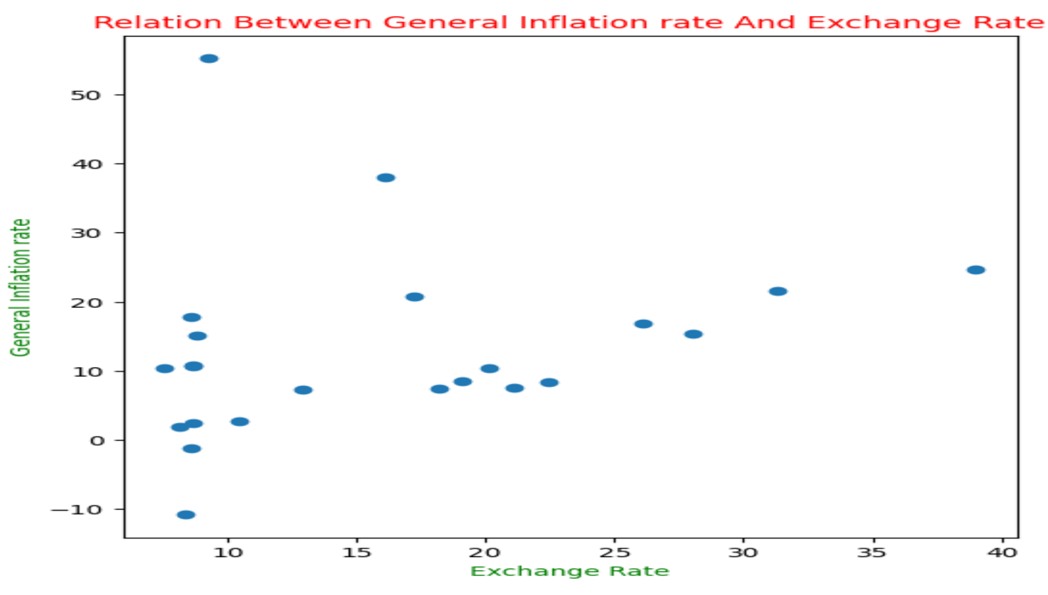


Figure 2 The Relation between general inflation rate and exchange rate
Source: Own illustration showing nonlinear trend of inflation based on the data set

In recent years, data-driven models have become increasingly popular for economic forecasting. ANN models have been used by researchers to improve the prediction accuracy of nonlinear property of the inflation data as illustrated on Figure 2 and have achieved good results [36]. There are many kinds of ANNs, including RNNs and LSTM networks [11]. RNNs and traditional ANNs have been compared in terms of time series model prediction, and it has been found that RNNs achieve a higher prediction accuracy than traditional ANNs [37]. RNNs have shown significant effects on the forecasting of data sequences. However, when the sequence length is too large, there is a dramatic increase in the training time of the RNN [38]. Based on the above problems, the LSTM network is proposed, which improves the hidden layer of the RNN, expands the memory function of the network, enables the model to obtain more persistent information, and slows down the information decay rate [15].

Previous studies have attempted to address the issue of long-term inflation trend forecasting in Ethiopia. For instance, researchers empirically investigated the drivers of inflation using monthly data from July 1998 to September 2020. It explores short-run and long-run effects of domestic and external determinants of inflation including the demand-side, the supply side and structural factors using the cointegration and vector error correction methodology. This study contributes for the investigation of the role of the fiscal sector in modeling inflation. However, this study did not consider the use of deep learning models for inflation forecasting[9].

The advent of deep learning techniques has opened up new avenues for time series forecasting, including inflation forecasting. Deep learning algorithms, particularly LSTM networks, have demonstrated remarkable ability to capture complex patterns and long-term dependencies in data[6].

One promising approach is the integration of ARIMA with LSTM networks due to their specifically designed to capture long-term dependencies in sequential data. The ensemble ARIMA-LSTM model leverages the linear forecasting capabilities of ARIMA and the nonlinear learning power of LSTM, offering a robust tool for accurate and reliable long-term inflation prediction.

The objective of this study is to implement and evaluate an ensemble ARIMA-LSTM model for long-term inflation prediction in Ethiopia. By combining the strengths of ARIMA and LSTM models, we aim to enhance the accuracy of inflation forecasts. This improvement in forecasting can help policymakers and stakeholders in Ethiopia to better understand and manage the economic challenges associated with inflation. The study utilized historical inflation data, applying the ensemble model to predict future inflation rates and comparing its performance with traditional models such as ARIMA and standalone LSTM.

1.2 Problem Statement

Inflation is a critical economic indicator that profoundly affects various aspects of a country's economy, including purchasing power, investment decisions, and policy formulation. In Ethiopia, a rapidly developing economy in East Africa, understanding and predicting long-term inflation trends is of paramount importance due to the country's socio-economic dynamics and the need for evidence-based policy-making. Therefore, the problem addressed in this research is the lack of an effective and robust forecasting model for long-term inflation trends in Ethiopia.

Existing inflation forecasting approaches often rely on traditional time series models, or exponential smoothing techniques. However, these models may have limitations in capturing complex nonlinear patterns and long-term dependencies present in inflation time series data, deal with and model a single time series variable at a time meaning it does not account other variables, historical events such as the millennium, civil war, droughts, floods, and major policy changes in the long-term inflation trend forecasting, seasonal patterns in inflation data.

While LSTM models have demonstrated success in various time series forecasting tasks, their performance can be further enhanced by combining them with traditional statistical models like ARIMA [39]. The ensemble of LSTM and ARIMA models can potentially leverage the strengths of both approaches, effectively capturing both short-term dynamics and long-term trends in inflation data. ARIMA excels at modeling linear relationships and short-term patterns in time series data, while LSTM is particularly effective at capturing complex nonlinear relationships and long-term dependencies in time series data [40]. By integrating these two models, the ensemble approach can mitigate the limitations of each individual model, resulting in more accurate and robust forecasts [41].

The research gap to be addressed involves several key aspects. First, there is a need to incorporate non-traditional data sources to capture a more nuanced picture of economic activity and potential inflation drivers. Second, we aim to develop models that go beyond traditional parametric approaches, enabling the capture of complex non-linear relationships between various factors influencing inflation. Third, the proposed ensemble forecasting model will account for historical events such as civil wars and droughts, which have significant impacts on inflation dynamics.

Finally, the study will focus on developing scenario-based forecasts that consider different possibilities for future events and their potential impact on inflation.

Therefore, this research aims to address the following key research questions:

- ✓ What is the predictive accuracy of the ensemble LSTM-ARIMA model for long-term inflation trends in Ethiopia compared to individual LSTM and ARIMA models?
- ✓ What are the key factors or variables that significantly influence inflation trends in Ethiopia?
- ✓ What is the effect of historical data length on the predictive accuracy of the LSTM-ARIMA ensemble model for inflation forecasting in Ethiopia?
- ✓ What is the change in predictive performance of the ensemble model when using different forecasting time horizons?

1.3 Objectives

1.3.1 General Objective

The general objective of this research is to develop an accurate and reliable forecasting model for long-term inflation trends in Ethiopia by leveraging LSTM and ARIMA ensemble approach.

1.3.2 Specific Objective

The following specific objectives have been defined in order to fulfill the study's overall goal:

- ✓ To design and implement an ensemble model that effectively combines the strengths of ARIMA and LSTM for improved inflation forecasting accuracy.
- ✓ To collect robust and high-quality data to enhance the reliability of inflation forecasting.
- ✓ To identify and select the relevant economic variables that significantly influence inflation trends in Ethiopia and integrate them into the LSTM and ARIMA ensemble model.
- ✓ To evaluate the effect of different historical data lengths (e.g., 5-year, 10-year, and 20-year datasets) on the predictive accuracy and robustness of the LSTM-ARIMA ensemble model for long-term inflation forecasting.
- ✓ To Assess the accuracy and reliability of the LSTM and ARIMA ensemble model in long-term inflation forecasting in Ethiopia, and compare its performance with traditional time series forecasting techniques such as exponential smoothing or linear regression.

1.4 Contribution

In this study, three contributions are made.

- Development of an Innovative Ensemble Forecasting Framework: Proposes an ensemble model that combines ARIMA and LSTM, leveraging ARIMA's strength in capturing linear trends and seasonal patterns with LSTM's ability to model complex, non-linear relationships. This framework provides a robust solution for time-series forecasting tasks.
- Incorporation of Economic Variables and Unseen Events for Enhanced Forecasting: Integrates relevant economic variables and accounts for the impact of unseen events such as civil wars and droughts, on inflation. This approach captures a broader range of factors, including unexpected shocks, leading to more comprehensive and accurate forecasts.
- Advancement of Machine Learning in Economic Forecasting: Expands the application of deep learning techniques in economics, demonstrating their potential to handle complex, real-world problems. This contribution opens new avenues for research in econometrics and data-driven modeling.

1.5 Scope of the Study

This study proposed a robust forecasting framework that integrates ARIMA and LSTM models to enhance the accuracy of long-term inflation predictions across various economic contexts. By using these two methodologies, the research seeks to provide a comprehensive tool that can effectively capture both linear trends and nonlinear patterns in inflation data. The focus is specifically on long-term forecasts, utilizing historical data to identify significant patterns and trends that can inform future economic conditions. Additionally, the study emphasizes empirical validation of the proposed model's performance.

Despite its promising contributions, this study also acknowledges several limitations inherent in using the ARIMA-LSTM ensemble model for inflation prediction. Below, we outline the limitations and the steps taken to address them.

Data Quality and Quantity: The model's performance relies heavily on the quality and quantity of historical data. To mitigate this limitation, thorough data preprocessing techniques such as outlier detection, imputation of missing values, and normalization were applied. Furthermore, reliable and comprehensive datasets from multiple sources were utilized to ensure sufficient historical depth.

Interpretability ("Black Box" Nature): The LSTM component enhances predictive capability but operates as a "black box," making interpretation challenging. To improve interpretability, feature importance analysis and partial dependence plots were employed to identify key drivers of inflation within the model. The integration of ARIMA also provides a transparent baseline for comparison.

Computational Intensity: Training the ensemble model, especially with large datasets, requires significant computational resources. This challenge was addressed by optimizing the model architecture through techniques like early stopping and hyperparameter tuning. Cloud computing resources were leveraged to expedite the training process without compromising accuracy.

Risk of Overfitting: There is a potential risk of overfitting when ensembling multiple models. To minimize this risk, rigorous cross-validation techniques and independent test datasets were used to evaluate the model's generalization ability. Validation loss was closely monitored during training to ensure robustness. Additionally, L2 regularization was applied to the LSTM component to penalize large weights and prevent the model from fitting too closely to the noise in the training data, thereby enhancing its ability to generalize to unseen data.

1.6 Methodology

To achieve the research objectives the following procedures were undertaken:

1.6.1 Literature Review

The literature on inflation forecasting includes various traditional and modern approaches, each with distinct strengths and limitations. Conventional models such as the ARIMA model have been widely adopted due to their ability to effectively capture linear trends and seasonality in economic data. ARIMA models have proven effective for short-term inflation forecasting, but they tend to underperform in long-term predictions where inflation data exhibit complex, non-linear patterns. This limitation has led researchers to explore advanced methods that can better capture the nuanced dynamics of inflation over time.

Recently, machine learning and deep learning approaches, particularly RNNs and LSTM networks, have been increasingly applied to inflation forecasting. LSTM networks, with their unique cell state and gating mechanisms, are especially adept at learning long-term dependencies in time-series data, making them well-suited for financial forecasting. Studies propose that ensemble models combining statistical methods with deep learning models to improve accuracy by leveraging the strengths of each. This study builds upon such research by developing an ARIMA-LSTM ensemble model specifically for long-term inflation forecasting, aiming to combine ARIMA's ability to handle linear trends with LSTM's capacity for non-linear relationships.

1.6.2 Algorithms

The forecasting model for this study is built on a combination of ARIMA and LSTM algorithms. The ARIMA model serves as a foundation by capturing the linear trends and seasonal components in inflation data. ARIMA relies on three primary parameters: p, d and q. These parameters are selected based on the dataset's autocorrelation and partial autocorrelation functions, as well as optimization techniques, to ensure an optimal fit. ARIMA is particularly effective for short-term predictions of inflation's linear aspects, providing a reliable baseline forecast.

To capture non-linear patterns in the data, this study incorporates LSTM, a type of recurrent neural network designed to retain information across long sequences of data. LSTM networks feature internal memory cells and gating mechanisms, allowing them to learn and retain complex, non-linear relationships over time. The LSTM model is fine-tuned through adjustments to its architecture, including the number of layers, hidden units, and training parameters like batch size and learning rate. The resulting ARIMA-LSTM ensemble model combines these two approaches by first applying ARIMA to model linear trends, followed by LSTM to capture residual non-linearities. This ensemble design enhances the overall forecasting capability by effectively addressing both linear and non-linear components of inflation data.

1.6.3 Evaluation Metrics

The study evaluates the performance of the ARIMA-LSTM ensemble model using standard time-series forecasting metrics. MAE is employed to gauge the average error magnitude in the predictions, offering a straightforward measure of accuracy. A lower MAE value indicates fewer large deviations from actual inflation values, making it ideal for evaluating the model's overall precision. The RMSE is another key metric, calculated by taking the square root of the average squared differences between predicted and actual values. RMSE places more weight on larger errors, making it a particularly useful metric for assessing the model's performance in scenarios where significant deviations are critical.

The study also uses MAPE, which calculates the average percentage difference between predictions and actual values, making it a scale-free metric suited for comparing different models. This metric is especially beneficial for stakeholders seeking to understand the model's relative accuracy, as it indicates the forecast error as a percentage of actual values. In some cases, R^2 is also used to explain the proportion of variance in inflation data that the model can predict. Together, these metrics provide a comprehensive assessment of the ARIMA-LSTM ensemble model's accuracy and allow for a robust comparison with other models, such as standalone ARIMA and LSTM. By applying these metrics, the study can effectively demonstrate the ensemble model's advantages in long-term inflation forecasting.

1.7 Organization of the study

The remaining of this study is organized as follows: Chapter Two introduces the fundamental concepts related to inflation forecasting, detailing the operational principles of models relevant to this study, particularly ensemble model. Chapter Three reviews prior research on inflation prediction, focusing on approaches that integrate statistical and machine learning models. Chapter Four outlines the research methodology, including data collection, preprocessing, and the development of the ARIMA-LSTM ensemble model, along with the evaluation metrics used to assess its performance. Chapter Five presents the results of the analysis, comparing the ensemble model's effectiveness to that of standalone models and discussing key findings. Finally, Chapter Six concludes the study with a summary of insights gained, along with recommendations for future research and potential applications of the ensemble model in economic forecasting.

Chapter 2
Background
This chapter provides a comprehensive theoretical overview of various algorithms employed for the purpose of inflation forecasting. Additionally, it encompasses fundamental concepts pertaining to inflation forecasting, including but not limited to ARIMA, LSTM and ARIMA-LSTM ensemble models

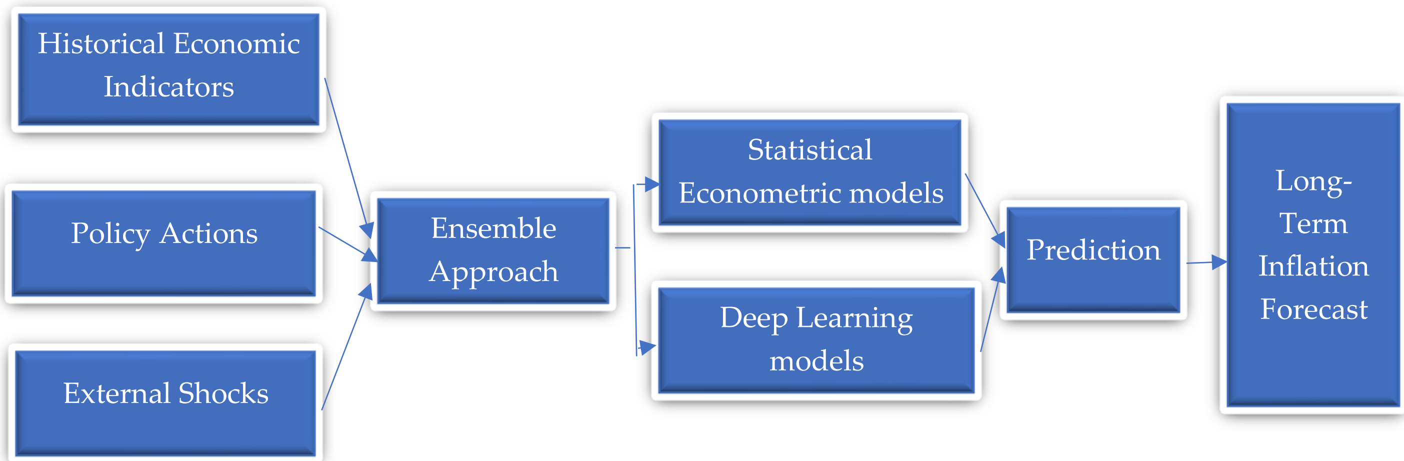
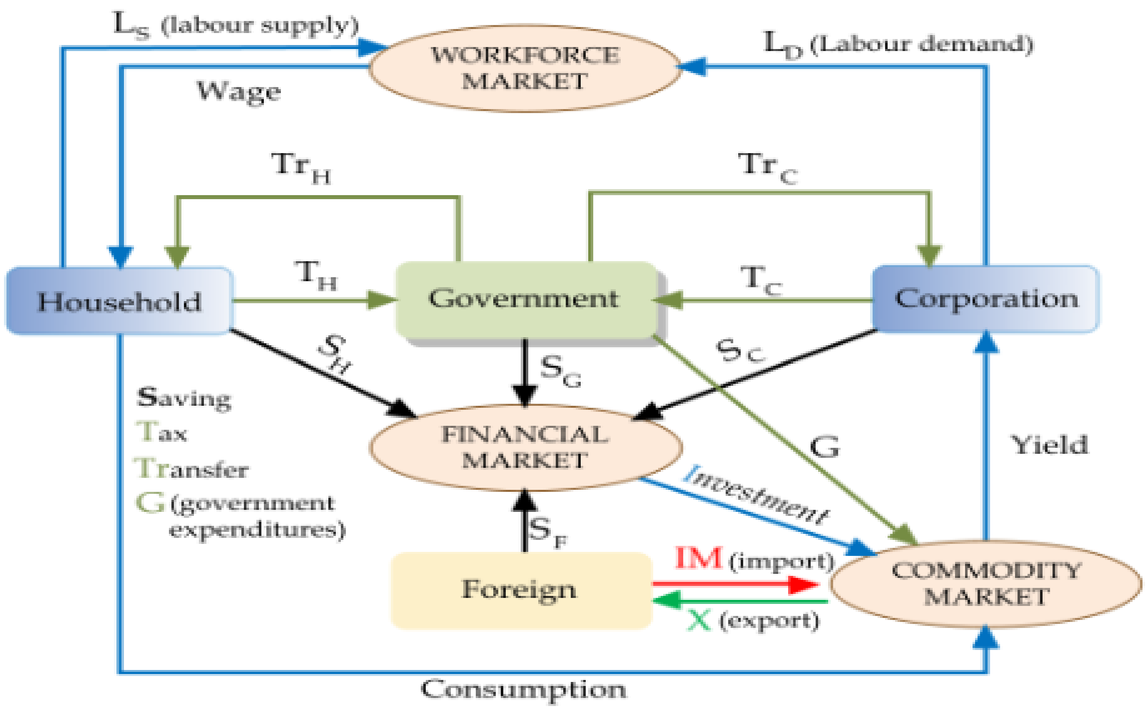


Figure 3 Description of input and output variables
Source: Adapted from [2],[4],[11]

Figure 3 shows historical economic indicators, policy actions, and external shocks as inputs. An ensemble method combines traditional econometric techniques with deep learning models to create a long-term inflation forecast.

Inflation is a critical economic phenomenon defined as the rate at which the general level of prices for goods and services rises, leading to a decrease in purchasing power. It is typically measured using indices such as the cpi [4]. A moderate level of inflation is generally considered a sign of a healthy economy, reflecting increased demand for goods and services. However, excessive inflation can lead to economic instability, eroding savings and distorting spending behavior [3]. Understanding the causes and implications of inflation is essential for developing effective forecasting models. As shown on Figure 4, inflation can arise from various factors that influence supply and demand dynamics within an economy. One common cause is demand-pull inflation, which occurs when aggregate demand exceeds aggregate supply [1]. This situation often arises in a growing economy where consumer confidence leads to increased spending. Conversely, cost-push inflation results from rising production costs, such as wages or raw materials, which businesses pass on to consumers in the form of higher prices. Additionally, built-in inflation is linked to adaptive expectations; as businesses and workers anticipate future inflation, they adjust their prices and wages accordingly, creating a self-fulfilling cycle [1,4,5]. Core inflation is another critical metric for assessing inflationary pressures in Ethiopia [9]. This measure excludes volatile items, such as food and energy prices, to provide a clearer picture of underlying inflation trends. By focusing on core inflation, economists and policymakers can gain insights into persistent inflationary pressures that may not be immediately apparent in broader price indices.



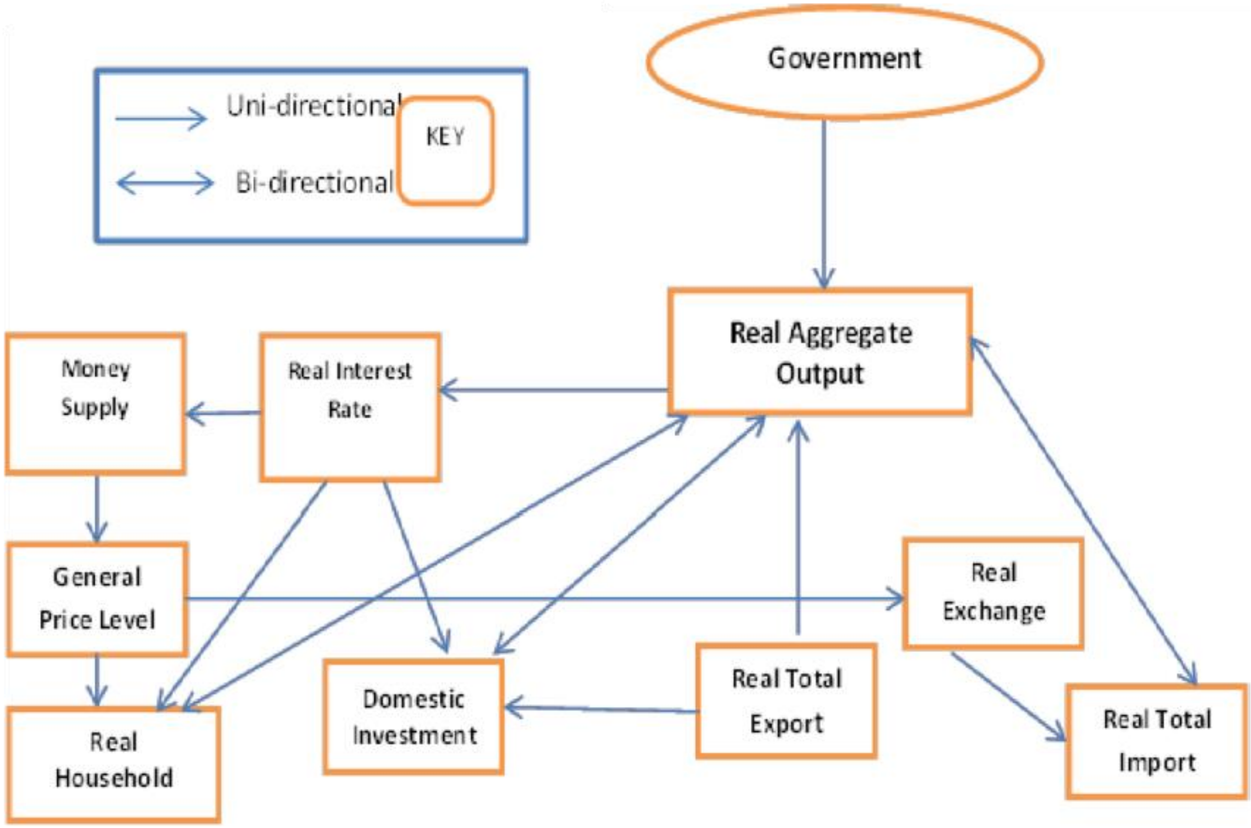


Figure 4 Diagram illustrating Macroeconomic variables
Source: Adopted from [1]

The CPI is crucial for evaluating the cost of living and the purchasing power of consumers, making it a vital indicator for policymakers [4,9]. Within the CPI, food prices constitute the largest share, which is particularly sensitive to agricultural performance due to seasonal fluctuations and climate variability. Therefore, any rise in food prices has a profound impact on overall inflation rates, significantly affecting low-income households that allocate a higher proportion of their income to food expenditures.

In recent years, Ethiopia has experienced notable inflationary pressures, with CPI showing significant increases [5]. For instance, in 2021 and 2022, the country faced inflation rates that reached levels not observed in decades, driven by a combination of factors including rising global commodity prices, supply chain disruptions related to the COVID-19 pandemic, and domestic challenges such as droughts that adversely affected agricultural output.

Exchange rate fluctuations represent a significant factor influencing inflation in Ethiopia [2]. Given the country's reliance on imports for essential goods, such as fuel, food, and consumer products, changes in the exchange rate of the Ethiopian birr against major currencies can have a substantial impact on inflation. A depreciation of the birr leads to higher import costs, which directly affects the prices of consumer goods and services. In recent years, the Ethiopian birr has experienced considerable depreciation against major currencies, contributing to rising inflation rates [13]. The import-heavy nature of the Ethiopian economy makes exchange rate stability crucial for maintaining price stability. The NBE has employed various measures to stabilize the birr, including monetary interventions and adjustments to interest rates. However, the success of these measures often depends on external factors, such as global economic conditions and investor confidence [2,13].

Economic growth is another essential indicator in understanding inflation dynamics in Ethiopia [14]. The GDP measures the overall economic output of the country and is closely linked to inflation trends. Rapid economic growth can lead to increased demand for goods and services, which may outstrip supply, resulting in inflationary pressures [30]. Conversely, slow economic growth or recession can exert downward pressure on prices, potentially leading to deflation. Ethiopia has witnessed impressive GDP growth rates in recent years, driven primarily by sectors such as agriculture, manufacturing, and services [2,14]. However, this growth has been accompanied by inflationary pressures, particularly in the agricultural sector, which is susceptible to climatic conditions and other external shocks. Policymakers must balance the goal of fostering economic growth with the need to control inflation, requiring a careful assessment of various economic policies and their potential impacts.

Several external factors also play a significant role in shaping inflation in Ethiopia [42]. Global commodity prices are a crucial determinant of domestic inflation, as fluctuations in the prices of oil, food, and other raw materials can directly affect the cost of living. For instance, an increase in global oil prices can lead to higher transportation and energy costs, which, in turn, contribute to rising consumer prices. Ethiopia's heavy reliance on imported fuel for transportation and industrial activities amplifies the impact of global price fluctuations on domestic inflation [2,14]. Additionally, international trade policies can affect the cost of imported goods, further influencing inflation rates. Tariffs, trade agreements, and regulatory changes can all play a significant role in shaping the prices of imported goods, impacting overall inflation levels in the economy. The relationship between global commodity prices, exchange rates, and domestic inflation underscores the importance of monitoring external economic conditions [14,42].

2.1 Traditional Forecasting Methods

Forecasting involves predicting future values based on historical data, and various methods exist ranging from simple statistical techniques to complex machine learning algorithms [16]. Traditional forecasting methods often rely on historical patterns and economic theory to make predictions about future inflation rates [11]. Time-series analysis is one of the primary approaches used in this context, focusing on data collected over time to identify trends, seasonal patterns, and cyclical movements.

Among traditional methods, ARIMA models stand out as a popular choice for time-series forecasting. ARIMA combines AR, differencing (I), and MA components to model time series data effectively [23]. The model is specified as ARIMA(p,d,q), where p represents the number of lag observations included in the model (autoregressive part), d indicates the number of times that the raw observations are differenced (integrated part), and q denotes the size of the moving average window. ARIMA models are particularly effective for stationary time series data but may struggle with capturing nonlinear relationships inherent in economic phenomena.

Another technique is exponential smoothing, which applies exponentially decreasing weights to past observations, giving more importance to recent data while still considering historical values [16,17]. These methods are foundational in traditional forecasting but may have limitations when dealing with complex economic dynamics.

2.2 Advanced Forecasting Techniques

The emergence of machine learning techniques has transformed forecasting practices by providing tools capable of capturing complex patterns in large datasets that traditional methods may overlook [16,17,18]. These techniques often outperform traditional statistical models in scenarios where relationships are nonlinear or where data is high-dimensional. One notable advancement in machine learning for time-series forecasting is the development of LSTM networks [43]. LSTMs are a type of RNN specifically designed for sequence prediction tasks. They address some limitations of traditional time-series models by effectively capturing nonlinear relationships and long-term dependencies within sequential data. Key features of LSTMs include memory cells that retain information over extended periods and a gates mechanism that controls the flow of information into and out of memory cells. LSTMs excel at recognizing patterns in time-series data due to their unique architecture, making them particularly suitable for complex forecasting tasks such as inflation prediction. The ability to learn from both recent observations and long-term historical data allows LSTMs to adapt dynamically to changing economic conditions [44]. As a result, they have shown superior performance compared to traditional methods in various applications, including financial forecasting [19].

2.2.1 Artificial Neural Network (ANN)

An artificial neural network (ANN) is a computational model that is inspired by the structure and function of the human brain [26]. ANNs are composed of interconnected nodes, or neurons, that can transmit signals to each other. The connections between neurons are called synapses, and they have a weight associated with them that determines the strength of the signal that is transmitted. They are trained to perform a specific task by feeding them a large amount of data. The network learns to identify patterns in the data and to make predictions based on those patterns. ANNs are used in a wide variety of applications, including image and speech recognition, natural language processing, and machine translation [32].

As shown on Figure 5, Artificial Neural Networks a powerful tool with the potential to revolutionize many industries [10]. However, they are also inherently complex and can be challenging to train due to issues such as vanishing gradients, overfitting, and the need for extensive computational resources [15]. As a result, ANNs are often combined with other machine learning techniques, such as Support Vector Machines (SVMs) and decision trees, to leverage their complementary strengths and improve overall performance [6,8]. Moreover, they are capable of learning and modeling non-linearities and complex relationships, which is achieved by neurons being connected in various patterns, allowing the output of some neurons to become the input of others. ANNs are classified into different types such as feedforward neural networks, recurrent neural networks, and convolutional neural networks.

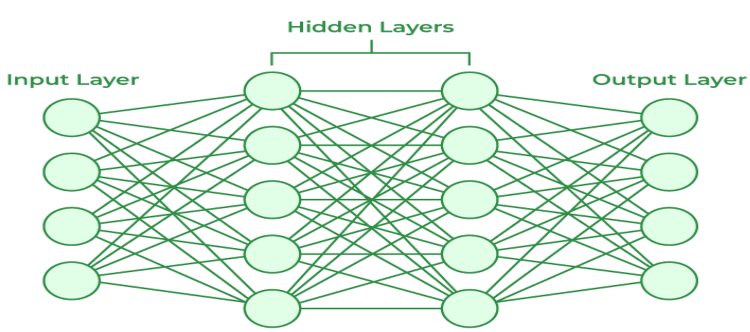


Figure 5 Diagram illustrating ANN
Source: Adopted from [35]

2.2.2 Deep Learning Model Architecture

Deep learning model architecture involves the arrangement of layers and blocks to create complex neural networks. These architectures can range from simple linear models with a single output to more complex structures such as the ResNet-152 architecture, which consists of hundreds of layers arranged in repeating patterns [45]. The versatility of the block abstraction allows for the creation of layers, entire models, or various components of intermediate complexity, providing flexibility in designing deep learning models.

Deep learning models, such as deep neural networks, deep belief networks, and convolutional neural networks, have been applied to various fields, including computer vision, natural language processing, and speech recognition, producing impressive results [46]. Generalizing deep learning models is crucial to avoid underfitting and overfitting. Techniques such as data-centric and model-centric generalization can help ensure that the model is trained to generalize the validation dataset and identify patterns from the training data, ultimately improving its performance on unseen data [47].

2.3 The ARIMA-LSTM Ensemble Model

The rationale behind this ensemble approach lies in leveraging the strengths of both models: ARIMA's capability to capture linear trends and LSTM's proficiency in identifying nonlinear patterns within the same dataset [7]. The structure of an ARIMA-LSTM ensemble model typically involves two main steps: first fitting an ARIMA model to historical inflation data to capture linear trends effectively; second training an LSTM model on the residuals obtained from the ARIMA model to identify any remaining nonlinear patterns that may not have been captured initially [6]. This two-step process allows for a comprehensive understanding of both linear and nonlinear influences on inflation dynamics. The final forecast generated by this ensemble model integrates predictions from both components. ARIMA provides a baseline forecast based on linear relationships while LSTM refines this forecast by accounting for more complex interactions within the data. The combination can be achieved through weighted approaches based on model performance metrics during validation phases.

Researchers recommend the implementation of an ARIMA-LSTM ensemble model, highlighting several potential benefits over relying solely on traditional forecasting techniques [17]. One key advantage they emphasize is the prospect of improved accuracy, suggesting that this approach could enhance predictive performance beyond what either method might achieve independently.

Additionally, researchers point to the flexibility of this ensemble approach, proposing that it could adapt to diverse economic conditions, making it a promising option for long-term forecasting across various scenarios. However, researchers also caution that implementing an ensemble model comes with challenges that need careful consideration [20]. They suggest that developing an effective ensemble model would require meticulous tuning of both ARIMA and LSTM parameters to optimize results. Furthermore, they note that LSTMs typically demand substantial datasets for effective training, and securing sufficient historical inflation data could pose difficulties due to constraints in available records or shifts in measurement methodologies over time.

Chapter Three

Literature Review

This chapter provides a key research publication closely related to the development of an ARIMA-LSTM ensemble model for long-term inflation prediction. The selected studies provide a foundation for understanding the strengths and weaknesses of existing methodologies, serving as benchmarks for comparison in this research.

In Ethiopia, inflation reached its peak at 33.9% in May 2022, which was the highest in nearly eleven years. The causes of this high inflation rate in Ethiopia are multifaceted. The sources of inflation in Ethiopia are attributed to both demand-side and supply-side factors, as well as structural factors. Demand-side factors, such as money growth and public sector borrowing, have contributed to inflation in the short run, while structural factors, including cereal output gaps and imported inflation, have also played a role. Additionally, the impact of drought, fallout from COVID-19, and a two-year conflict in the northern Tigray region have been cited as contributing factors to the high inflation in Ethiopia [13].

The literature underscores the importance of ensemble models in economic forecasting, particularly for complex and volatile variables like inflation. The current study aims to develop an ensemble LSTM-ARIMA model to predict long-term inflation trends, leveraging the strengths identified in the literature review.

ARIMA has traditionally been used for modeling the mean of inflation series, but a notable advancement comes from combining it with GARCH to capture volatility clustering and heteroscedasticity [23]. This approach demonstrated improved forecast accuracy and provided a better representation of inflation volatility compared to standalone ARIMA models. The integration of these methods highlighted the importance of modeling both the mean and variance of inflation rates, offering valuable insights into the dynamics of inflation forecasting.

Building on traditional methods, a novel approach integrates ARIMA for modeling linear components with ANN to capture nonlinear relationships in inflation data [24]. This combination demonstrated significant improvements in forecast performance, particularly in handling unexpected inflation shocks. By providing a framework for combining statistical and machine learning techniques, this work highlighted the potential of hybrid models in enhancing economic forecasting accuracy.

A hybrid approach that combines SARIMA for capturing seasonal patterns with SVR for modeling nonlinear trends has been proposed [25]. This integration significantly enhances prediction accuracy by effectively addressing both seasonal and nonlinear patterns in inflation data. By demonstrating the utility of ensemble models in handling seasonal variations and complex inflation dynamics, this work highlights the potential of combining statistical and machine learning techniques for improved forecasting performance.

The focus of [26] was on stock market forecasting, specifically predicting stock closing prices and directional movements using an Ensemble ARIMA-LSTM model. The study demonstrated the effectiveness of this ensemble approach in outperforming standalone models and other ensemble methods like ARIMA-SVM and ARIMA-RF, particularly in terms of directional accuracy. In contrast, this study shifts the application domain to inflation prediction, addressing the unique challenges inherent in economic data, such as seasonality, nonlinearity, and external shocks. Furthermore, this study incorporates the impact of unseen events such as civil wars and droughts that significantly influence inflation trends but were not considered in stock market-focused studies. By applying the Ensemble ARIMA-LSTM model to inflation forecasting and explicitly accounting for these unseen events, the current work extends the applicability of ensemble models to more complex and volatile economic phenomena, offering a novel contribution to the field.

The macroeconomic determinants of inflation in Ethiopia were examined in [27], where it was found that factors such as broad money supply, real GDP, population growth, gross national saving, and previous year imports serve as short-run drivers of inflation. The study employed methods such as the ARDL model with bound test, as well as Augmented Dickey-Fuller and Phillips-Perron unit root tests to analyze these relationships. The results identified both long-run determinants such as real GDP, real effective exchange rate, and lending interest rate and short-run drivers, including broad money supply, real GDP, population growth, gross national saving, and previous year imports. This analysis provides valuable insights into the complex dynamics influencing inflation in Ethiopia.

An ensemble deep learning model that combines VAE and ConvLSTM networks for inflation forecasting was proposed in [28]. This approach leverages the strengths of both models, where VAEs capture latent representations of the data, and ConvLSTMs effectively model temporal dependencies. The results indicate that deep learning models like VAE-ConvLSTM have the potential to outperform traditional econometric approaches based on linear, stationary models in macroeconomic forecasting. By integrating these advanced techniques, the study highlights the promise of deep learning in capturing complex patterns and improving forecast accuracy.

In [29], a comparison of different models for time-series forecasting is presented, including SARIMA, CNN (Convolutional Neural Network), and LSTM. The study focuses on analyzing and forecasting time-series data using these approaches. LSTM cells, which are integral to recurrent neural networks, enable the learning of sequential patterns from variable-length data. LSTM-based recurrent neural networks are considered one of the most powerful methods for learning from sequential data, with time series being a specific application of this broader capability.

In [30], a generalized CEEMDAN-LSTM architecture is proposed for time-series forecasting, designed to handle exogenous features as inputs while addressing challenges such as convergence difficulties associated with increasing input data size. This architecture combines the strengths of CEEMDAN for signal decomposition and LSTM neural networks for capturing temporal dependencies. The study demonstrates improved accuracy compared to the original model, highlighting the effectiveness of integrating these techniques for enhanced forecasting performance.

In [31], neural network models are applied to inflation forecasting, with a significant contribution being the use of Long-Short Term Memory (LSTM) networks. This recurrent neural network approach summarizes macroeconomic information into common components, enhancing the model's ability to capture complex patterns in the data. Experimental results demonstrate that the estimated neural networks generally outperform standard benchmark models, highlighting the potential of LSTMs in improving inflation forecasting accuracy.

In [32], the value of nonlinear machine learning techniques in forecasting macroeconomic time series is demonstrated. The study shows that a LSTM recurrent neural network outperforms several linear models, including AR, RW, SARIMA, and Markov switching model (MS-AR), as well as a simple fully-connected neural network (NN), in predicting monthly US CPI inflation. The superior performance of LSTM is attributed to its ability to capture nonlinearities in the data, combined with its flexible architecture.

In [33], an ensemble ARIMA-MLP system is proposed to effectively map both linear and nonlinear patterns in inflation data. This approach utilizes an ensemble evolutionary system that combines a simple exponential filter, linear ARIMA and AR models, along with a MLP model. The experimental results demonstrate that the ensemble evolutionary system delivers promising performance in the prediction domain, highlighting its potential for improving inflation forecasting accuracy.

Previous researchers have made significant strides in time series forecasting, particularly in the context of inflation and stock price series. Studies have compared various forecasting models, including ARIMA, LSTM, MLP, RNN, CNN, and ensemble models such as ARIMA-LSTM-CF and VAE-ConvLSTM. These investigations consistently demonstrate that deep learning models like LSTM and VAE-ConvLSTM outperform traditional econometric models (e.g., ARIMA, MLP, RNN) in terms of prediction accuracy and consistency.

Despite these advancements, several gaps remain in the literature. First, the applicability of these models across different economic contexts and datasets requires further exploration. Second, the impact of varying input variables and hyperparameters on model performance remains under-investigated. While ensemble models that combine deep learning and traditional econometric approaches have shown promise, there is still a need to refine and optimize these hybrid frameworks for specific applications, such as inflation forecasting in developing economies. Furthermore, incorporating external factors, such as unseen events (e.g., civil wars, droughts), into ensemble models could enhance their robustness and predictive power. This study aims to address these gaps by developing an ensemble ARIMA-LSTM model tailored for long-term inflation forecasting in Ethiopia, with a focus on improving accuracy and reliability through the integration of both linear and nonlinear patterns, as well as external economic drivers.

3.1 Gaps and Opportunities:

The literature on inflation forecasting and time series analysis has made significant strides in recent years, particularly with the advent of ensemble modeling approaches that combine traditional statistical techniques with advanced machine learning methods. Researchers have demonstrated the effectiveness of ARIMA models in capturing linear relationships within time series data, while studies on LSTM networks have highlighted their superior ability to model complex, non-linear patterns. Notably, various studies have shown that ensemble models, such as ARIMA-LSTM ensembles, can outperform standalone models in terms of predictive accuracy for financial and economic time series.

Key achievements include the development of ensemble methodologies that effectively leverage the strengths of both ARIMA and LSTM, leading to improved forecasting performance across different datasets and applications. These models have been successfully applied in areas such as stock market prediction and algorithmic trading strategies, demonstrating their versatility and potential for enhancing decision-making processes in various economic contexts.

However, despite these advancements, several critical gaps remain in the existing literature. Most notably, there is a lack of focused research on long-term inflation prediction specifically. Many studies concentrate on short- to medium-term forecasting, leaving a void in understanding how ensemble models can be optimized for long-term economic trends. Additionally, while some research has explored the integration of ARIMA and LSTM, there is limited consensus on the best methodologies for achieving this integration effectively.

Furthermore, existing studies often overlook the incorporation of external factors such as geopolitical events or policy changes that can significantly influence inflation rates. This gap highlights the need for further investigation into robust ensemble models that not only enhance predictive accuracy but also consider a broader range of influencing variables.

In conclusion, while previous researchers have laid a strong foundation for understanding the potential of ensemble ARIMA-LSTM models in time series forecasting, there remains a pressing need for targeted research focused on long-term inflation prediction. Addressing these gaps not only contribute to the academic discourse but also provide valuable insights for policymakers and stakeholders navigating an increasingly complex economic landscape.

Table 1 Summary table of related literature
Source: Taken from literature review

Reference	**Main Idea**	**Key Achievements**	**Tools/Methodologies**	**Remarks**
Ensemble Approach for Stock Market Forecasting	Proposes a threshold-based ensemble model combining ARIMA and	Outperforms standalone ARIMA and LSTM models in directional accuracy.	Time series analysis, RMSE, MAPE metrics.	Focuses on short-term forecasts, not inflation.

Using ARIMA and LSTM	LSTM for stock price prediction.			
LSTM-ARIMA as a Ensemble Approach in Algorithmic Investment Strategies	Develops an algorithmic investment strategy using a ensemble LSTM-ARIMA model.	Demonstrates superior performance across multiple equity indices.	Walk-forward optimization, backtesting, risk-adjusted return measures.	Does not address inflation prediction specifically.
A Comparison of ARIMA and LSTM in Forecasting Time Series	Investigates the performance of ARIMA vs. LSTM in economic and financial time series data.	LSTM outperforms ARIMA with an average error reduction of 84% to 87%.	Historical financial data analysis, error minimization metrics.	Lacks exploration of ensemble models for inflation forecasting.
Time Series Analysis: ARIMA vs. LSTM in Predictive Modeling	Discusses differences between ARIMA and LSTM in predictive modeling for time series data.	Highlights strengths and weaknesses of both models for different scenarios.	Theoretical comparisons, potential ensemble models discussion.	Does not delve into practical applications or inflation prediction.
Time Series Forecasting using LSTM and ARIMA	Compares the performance of LSTM-ARIMA for various time series.	Provides insights into the predictive capabilities of both models across datasets.	Python-based implementations, performance evaluation metrics.	Does not focus on or ensemble strategies.

Chapter Four
Methodology

This section is detailed in the following six items as shown on Figure 6: (i) Data collection and pre-processing (ii) Feature selection (iii) Time Series Decomposition (iv) Model Development (v) Training the Models (vi) Evaluating the Models

Predicting inflation is a vital aspect of economic analysis, typically addressed using time series forecasting techniques. This involves analyzing historical inflation data to identify patterns and trends, which are then used to predict future inflation levels. The primary goal is to develop accurate models that can forecast inflation over specific time horizons, aiding in effective economic planning and decision-making. Addressing this problem requires selecting models that are well-suited to the unique characteristics of inflation data, such as seasonality, trends, and potential external influences. Furthermore, understanding the techniques and methodologies used to implement these models is crucial for achieving reliable predictions. The objective of this research is to build an ensemble model for long-term inflation forecasting.

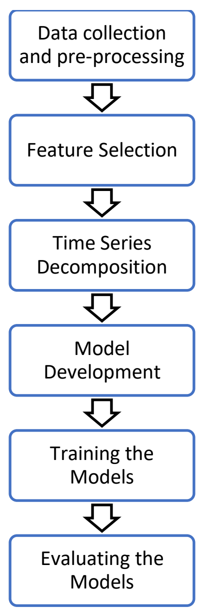


Figure 6 General methodology flow chart
Source: Own illustration

4.1 Data collection and Pre-processing

The data is collected from the National Bank of Ethiopia, Central Statistical Agency, World Bank, and IMF. Data contains annual and biannual data on GDP growth, exchange rates, interest rates, commodity prices, and other relevant macroeconomic variables. To capture inflation due to sudden economic shocks such as civil war and drought, binary event flags (Civil_War_Flag and

Drought_Flag) were created. These flags take the value 1 during the periods affected by these events and 0 otherwise. Imputing or interpolating missing values to ensure complete datasets is undertaken, and we identified to address outliers that could skew model performance.

How the Key variables Are Incorporated:

- Binary Event Flags: By creating binary event flags for civil war and drought, we explicitly account for sudden economic shocks that could significantly influence inflation trends. Normalizing all features ensures that these binary flags are treated consistently with other macroeconomic indicators, enabling their seamless integration into both ARIMAX (with Exogenous Variables) and LSTM models later in the process.
- Macroeconomic Indicators: The 11 macroeconomic features were preprocessed and normalized to ensure they are on a comparable scale. This step ensures that traditional economic drivers, such as money supply, exchange rates, and unemployment rates, are effectively integrated into the models alongside the binary event flags.

In this sense, the indicators obtained were the money supply (ms), real GDP (rgdp), general inflation rate (gira), lending interest rate (lir), exchange rate (er), consumer price index (cpi) as the target variable, unemployment rate (unrate), personal savings rate (psr), real effective exchange rate (reer), total credit to the private sector (tcs), and industrial production index (indpro), all covering a range of data from 1979 to 2023.

The necessary preprocessing techniques that were applied to ensure the quality of the data were the analysis to check the existence of outliers and the need to treat them, the matching of the range of observations in the data set, and the normalization of the input values (X). These techniques are described below: The process of preparing the data for subsequent detailed analysis consisted of performing data cleaning by removing missing observations from the data set.

As part of preparing the GDP data for model development, we performed a transformation to ensure stationarity, a key assumption for time series model ARIMA. Stationarity refers to a time series whose properties do not depend on the time at which the series is observed, and it is crucial for accurate forecasting.

To achieve stationarity, we calculated the first difference of the GDP time series. This transformation, denoted as ($\Delta GDP_t = GDP_t - GDP_{t-1}$) , removes any trend from the data, making it stationary. The reason for removing the trend (by calculating the first difference) is to make the time series stationary. Many time series models, including ARIMA, require stationary data because they assume that the statistical properties of the series, such as mean and variance, do not change over time. If a series has a trend (i.e., its mean increases or decreases over time), it is considered non-stationary, which can lead to inaccurate model forecasts. Following this, we applied the Augmented Dickey-Fuller (ADF) test to confirm whether the differenced GDP series is stationary. The ADF test checks for the presence of a unit root in the series, which would indicate non-stationarity. The ADF test is based on the regression [16]:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{\rho} \delta i \Delta yt - i + \epsilon t \quad (1)$$

Where:

- Δyt is the first-differenced series (in this case, $\Delta GDPt$).
- α is a constant.
- βt represents the deterministic trend (often omitted in differenced data).
- γy_{t-1} is the coefficient of the lagged level of the series.
- $\sum_{i=1}^{\rho} \delta i \Delta yt - i$ represents the lags of the differenced series to account for higher-order correlation.
- ϵ_t is the error term.

Once stationarity was confirmed, we proceeded with the Granger Causality test to explore the relationship between GDP and inflation. This test helps identify whether past values of GDP can help predict future inflation, which could inform the structure of our time series model. To test if GDP Granger-causes inflation, we use the following bivariate regression model [20]:

$$Inflation_t = \alpha + \sum_{i=1}^{\rho} \beta i Inflation_{t-i} + \sum_{i=1}^{\rho} \gamma i \, \Delta GDP_{t-i} + \epsilon_t \quad (2)$$

Where:

- Inflation is the value of inflation at time t.

- α is a constant.
- βi are coefficients for the lags of inflation.
- γi are coefficients for the lags of the differenced GDP series.
- ϵt is the error term.

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were analyzed for two groups of economic indicators to assess their temporal dependencies and determine their suitability for time series modeling.

For the first group of indicators, which included unemployment rate (unrate), private sector reserves (psr), exchange rate (er), real GDP (rgdp), money supply (ms), and real effective exchange rate (reer), ACF and PACF plots were generated to examine the correlation structure at different lag periods. The ACF plots were used to identify the presence and extent of autocorrelations across multiple lags, while the PACF plots were employed to isolate direct correlations between observations, helping to determine the appropriate lag orders for each variable.

Similarly, for the second group of indicators, which included lending interest rate(lir), total credit to private sector(tcs), consumer price index(cpi) and industrial production index(indpro) ACF and PACF analyses were conducted following the same procedure to evaluate their temporal dependencies and inform the selection of suitable lag structures.

4.1.1 Outlier Detection

To determine outliers in characteristics, the IQR technique was used using the 1.5 IQR rule, which designates any value greater than Q3 + (1.5 × IQR) and any value less than Q1 − (1.5 × IQR) as an outlier [23].

The result indicates that there are no outliers in the features column when using the Interquartile Range method for outlier detection. We also used the Z-Score method to identify outliers as values with a standardized score (z) that exceeds a chosen threshold (commonly $|z|>3$).

$$z = \frac{x-\mu}{\sigma} \quad (3)$$

Where:

- x = individual value

- μ = mean of the data
- σ = standard deviation of the data

In this case, also we got an empty data frame suggests that all values in the features column have $|z| \leq 3$, meaning there are no statistically extreme values according to this method.

4.2 Feature selection

Inflation is affected by many factors. Therefore, it is necessary to consider all these economic indicators in the prediction system, such as Money Supply Growth, Interest Rates, Exchange Rate Volatility, Inflation Expectations, Government Expenditure, Public Debt, Tax Policies, Subsidies and Price Controls, Global Commodity Prices, Terms of Trade, Foreign Direct Investment, Remittances, Agricultural Productivity, Transport and Logistics Costs, Energy Availability, Industrial Production, Urbanization and Housing Demand, Population Growth, Unemployment and Underemployment and Corruption.

In this study, feature selection was performed based on the Pearson correlation coefficients and p-values to evaluate the statistical relationship between each feature and the target variable, Consumer Price Index (cpi). Additionally, results from the Granger Causality test were reviewed to validate whether lagged values of each feature significantly influence cpi. Features with low correlation to CPI (<0.30) or non-significant Granger Causality ($p > 0.05$) were considered for exclusion unless they provided unique economic insight, ensuring the final set balances relevance and diversity. These statistical measures provided insight into the strength and significance of the relationships between the input features and the target variable. In addition to the correlation-based selection of macroeconomic features, the binary event flags (Civil_War_Flag and Drought_Flag) were evaluated for inclusion due to their documented impact on Ethiopian inflation dynamics [2].

To visualize the computed correlation matrix, the Seaborn library in Python was used to generate the heatmap. The Seaborn heatmap() function allowed for an intuitive representation of the relationships between variables using a color gradient. In the heatmap, strong positive correlations are shown in red, strong negative correlations in blue, and weaker correlations in lighter shades. This visual approach effectively highlights significant relationships between the variables, aiding in identifying influential factors for the target variable and understanding interdependencies among other macroeconomic indicators.

4.3 Time Series Decomposition

To forecast inflation by ensemble LSTM-ARIMA model, we used time series decomposition to break down the inflation data into its underlying components: trend, seasonal, and residual (irregular) patterns. Decomposing the data helps us to understand the linear and non-linear behaviors within the time series, which can then be effectively modeled using a combination of ARIMA and LSTM models.

The process begins with time series decomposition, where the inflation data is broken down into three primary components: trend, seasonal, and residual (irregular) patterns. This is achieved using a method like seasonal_decompose, which assumes that a time series can be represented as a combination of these components. Depending on the nature of the data, the decomposition can be additive or multiplicative. In the additive model, the time series is expressed as the sum of its components [16]:

$$Y(t)=T(t)+S(t)+R(t) \quad (4)$$

where Y(t) is the observed time series, T(t) is the trend component, S(t) is the seasonal component, and R(t) is the residual or irregular component. This model is typically used when the seasonal fluctuations and residual variations remain constant over time. In contrast, the multiplicative model assumes the components interact proportionally, as follows [26]:

$$Y(t)=T(t)\times S(t)\times R(t) \quad (5)$$

This approach is more appropriate when the seasonal variation increases or decreases with the magnitude of the trend. For example, inflation might exhibit larger seasonal variations during periods of economic growth. The seasonal_decompose function automatically evaluates these patterns, using appropriate methods to isolate the individual components, regardless of the chosen model.

The decomposition process uses moving average smoothing to separate the trend from the other components. A centered moving average is applied over a sliding window of data points to compute the smoothed trend component. Once the trend is isolated, the seasonal component is extracted by computing the deviations from the observed series at a fixed periodicity (e.g., monthly or quarterly patterns). The seasonal component represents consistent repeating patterns in the data over the cycle length, such as inflation peaks at particular times of the year. Lastly, the residual component is calculated as the difference (additive model) or ratio (multiplicative model) of the observed data to the trend and seasonal components. The residuals capture random variations or noise in the series that cannot be explained by the trend or seasonality.

Mathematically, this method assumes the time series follows a stationary or semi-stationary process within each isolated component, which facilitates predictive modeling. Decomposition is critical for understanding how linear and non-linear behaviors manifest in the data. In this context, the trend and seasonal components, linear in nature are suitable for ARIMA modeling, which excels in handling autoregressive and moving average patterns. In contrast, the residual component often exhibits complex, non-linear patterns, making it ideal for modeling with a LSTM network, which excels in capturing intricate dependencies and structures in sequences. By isolating these components, the decomposition process ensures that both models target specific behaviors, thereby improving the overall forecasting accuracy.

4.4 Model Development

4.4.1 ARIMA Model

ARIMA uses time series data to forecast future. Differentiation is used in ARIMA models to transform a non-stationary time series into a stationary one. After that, they forecast future values using "auto" correlations and moving averages over data residual errors. It combines three key components: Autoregressive (AR), Differencing (I for Integration), and Moving Average (MA). The AR component captures the relationship between an observation and its lagged values, allowing the model to utilize past information to predict future outcomes. This aspect of the model is essential for understanding how previous observations influence current values. A critical step in applying the ARIMA model is differencing, which transforms a non-stationary time series into a stationary one. Stationarity is a fundamental requirement for many time series forecasting methods, as it ensures that the statistical properties of the data do not change over time. Differencing involves subtracting the previous observation from the current observation, effectively removing trends or seasonal patterns that may affect the accuracy of forecasts. By achieving stationarity, the ARIMA model can more reliably capture underlying patterns in the data. Once stationarity is established, ARIMA forecasts future values by incorporating lagged observations through its autoregressive component and lagged forecast errors through its moving average component. The general formulation of an ARIMA (p, d, q) model is [16,26]:

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\epsilon_t \quad \text{………………………………………………………………………..(6)}$$

where:

Y_t is the original time series.

B is the backshift operator, such that $BY_t = Y_{t-1}$.

$(1-B)^d$ represents the differencing part of the model.

The autoregressive part of the model is given by:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^\rho \quad \text{…………………………………………………… (7)}$$

ϕ_i as the autoregressive coefficient.

The moving average part of the model is given by:

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \quad \text{…………………………………………………… (8)}$$

θ_i is the moving average coefficient.

ϵ_t is a white noise error term.

The formulation indicates that the differenced series $(1-B)^d Y_t$ is modeled using an ARMA(p, d, q) process. Each element of the ARIMA model contributes to capture various aspects of the time series dynamics.

The ARIMA model was extended to an ARIMAX model by incorporating the binary event flags (Civil_War_Flag and Drought_Flag) as exogenous variables. This modification allows the model to explicitly account for the impact of civil war and drought on inflation. The ARIMAX model captures linear relationships and seasonality in the data while considering the influence of external shocks.

4.4.2 LSTM Model

LSTM is a type of RNN that is specifically designed to handle sequential data, such as time series, speech, and text. This study designed the appropriate network parameters for the LSTM network. We used an LSTM network to solve the prediction problem using the feedforward architecture. This architecture consists of a sequential input layer, a Keras LSTM layer, and a dense output layer. We investigated the optimal number of hidden neurons in the LSTM layer. In the proposed model, the hyperbolic tangent function is used as an activation function for the input nodes and the LSTM nodes, while the linear function is used as an activation function at the output node. The hyperbolic tangent function is a scaled sigmoid function and returns the input value in a range between −1 and 1, and the linear activation function is directly proportional to the input, allowing the output to vary continuously and linearly. The activation function of the output node is designated as a linear function because our goal is the prediction of next month's inflation rate, which can be formulated as a regression problem. The initial network weights are set as random values, and the network weight is adjusted using the "Adam" stochastic optimization method; this optimization algorithm was developed by Kingma and Ba.

As shown in Figure 7, the LSTM model is typically defined by its internal cell state and output, which can be represented as:

$h_t=\omega_t*tanh(c_t)$

Where:

- h_t is the hidden state at time t.
- $\omega_t = \sigma(Wo[h_{t-1},x_t] + b_0)$ is the output gate
- $c_t=f_t * c_{t-1}\ i_t*g(Wc[h_{t-1},x_t] + b_c)$ is the cell state, with $f_{t,}$ $i_{t,}$g(.) representing forget, input and candidate functions respectively.

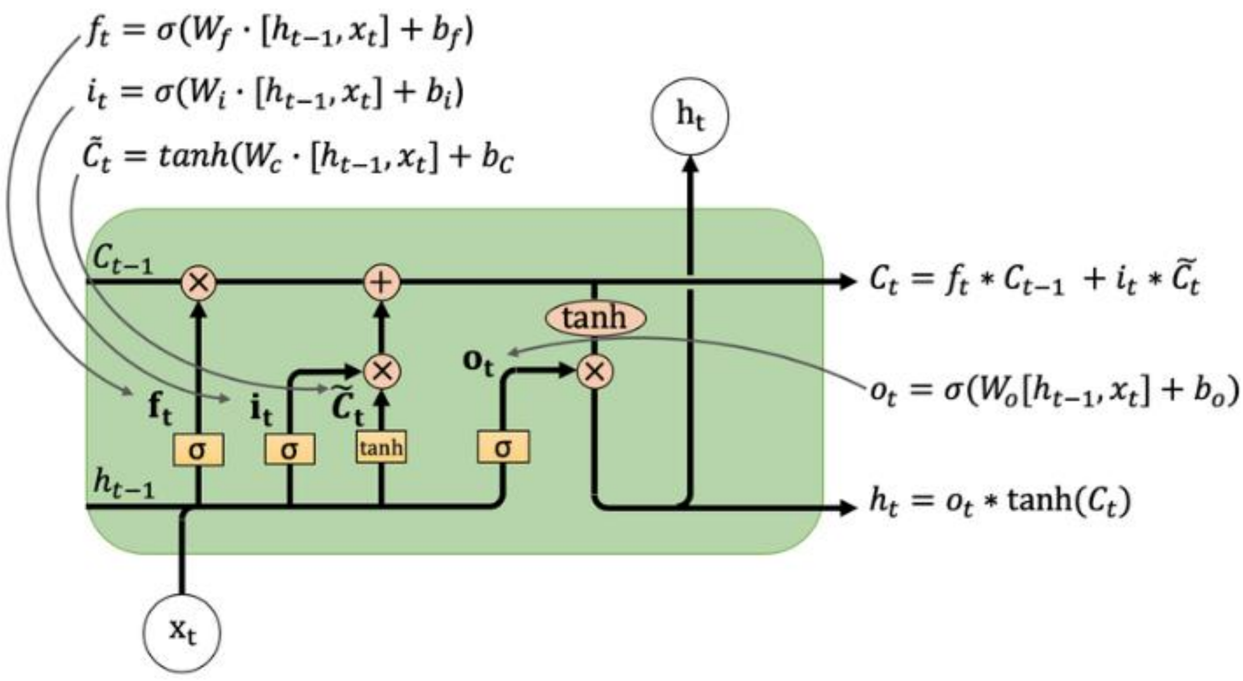


Figure 7 LSTM architecture
Source: Adopted from [29]

The input to the LSTM includes all macroeconomic features along with the binary event flags, ensuring that the model can learn the effects of sudden economic shocks on inflation. The binary features were normalized and treated as additional input variables, allowing the LSTM to capture the impact of these events on inflation trends.

4.4.3 Ensemble ARIMA-LSTM Model

The study proposes an ensemble approach, termed the Bounded Weighted Ensemble, which integrates traditional statistical and deep learning algorithms for financial time series forecasting. This method combines the strengths of the Autoregressive Integrated Moving Average (ARIMA) model, which provides insights into underlying economic relationships driving inflation, with the Long Short-Term Memory (LSTM) model, which excels at detecting intricate, non-linear patterns in data. By leveraging these complementary capabilities, the ensemble aims to enhance forecasting accuracy over individual models.

In this approach, both ARIMA and LSTM models are applied to the original inflation series to generate independent forecasts. These forecasts are then combined using a weighted combination optimized for performance, followed by a bounding mechanism to ensure the ensemble prediction remains within the range of the individual model outputs. The process is defined as follows:

The unadjusted ensemble prediction at time t is calculated as a weighted combination of the ARIMA and LSTM forecasts:

$$P_{\text{Ensemble, unadjusted },t} = w_{\text{ARIMA}} \cdot P_{\text{ARIMA },t} + w_{\text{LSTM}} \cdot P_{\text{LSTM },t}$$

where $P_{\text{ARIMA}, t}$ and $P_{\text{LSTM}, t}$ are the ARIMA and LSTM predictions at time t, respectively, and w_{ARIMA} and w_{LSTM} are weights satisfying $w_{\text{ARIMA}} + w_{\text{LSTM}} = 1$. To ensure the ensemble prediction respects the bounds of the individual models, the final prediction is adjusted using a clipping function:

$$P_{\text{Ensemble },t} = \text{clip }\left(P_{\text{Ensemble, unadjusted },t}, \min\left(P_{\text{ARIMA },t}, P_{\text{LSTM},t}\right), \max\left(P_{\text{ARIMA},t}, P_{\text{LSTM},t}\right)\right)$$

where $\text{clip }(x, a, b) = \max(a, \min(b, x))$ constrains the prediction to the interval $\left[\min\left(P_{\text{ARIMA },t}, P_{\text{LSTM },t}\right), \max\left(P_{\text{ARIMA },t}, P_{\text{LSTM },t}\right)\right]$.
The weights w_{ARIMA} and w_{LSTM} are determined by minimizing the Mean Squared Error (MSE) over a validation period:

$$\text{MSE}_{\text{validation}} = \frac{1}{n}\sum_{t=1}^{n} \left(Y_t - P_{\text{Ensemble, unadjusted },t}\right)^2$$

where Y_t is the actual inflation value at time t, and n is the number of validation observations. This optimization balances the contributions of ARIMA's stability and LSTM's adaptability, with the bounding step ensuring the final forecast aligns with the range of plausible model outputs.

This Bounded Weighted Ensemble approach enhances robustness by preventing the ensemble from producing predictions outside the credible range established by ARIMA and LSTM, while still capitalizing on their combined predictive power.

The ARIMA and LSTM models are first run independently on the same dataset to leverage their unique strengths. This allows them to generate distinct predictions, reflecting their individual capabilities. Once the models have produced their forecasts, the outputs are combined using a weighted averaging approach. In this approach, the contribution of each model's predictions to the final forecast is determined by their respective performance metrics. Specifically, the weights are calculated based on metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), which quantify the accuracy of each model's predictions during the training phase. Models with lower error values (indicating better performance) are assigned higher weights, ensuring that the ensemble model benefits more from the stronger performer while still incorporating the complementary strengths of both models.

For example, if the ARIMA model achieves a lower RMSE than the LSTM model on the training data, it will be assigned a higher weight in the ensemble. Conversely, if the LSTM model outperforms ARIMA in capturing nonlinear patterns, its predictions will carry greater importance in the final forecast. This dynamic weighting ensures that the ensemble model adapts to the specific characteristics of the dataset, improving overall forecasting accuracy and robustness.

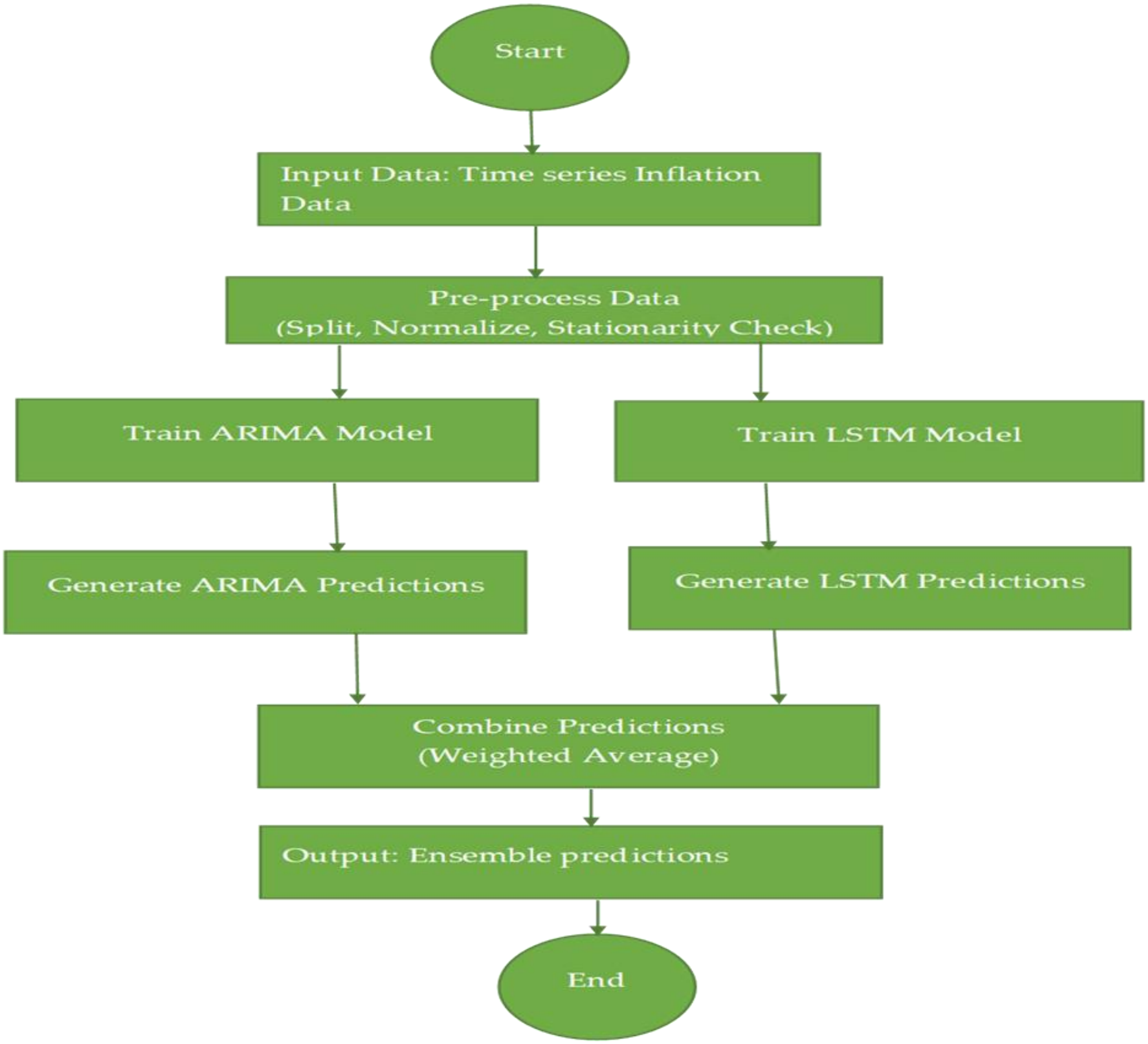


Figure 8 Flowchart of the proposed forecasting algorithm
Source: Own illustration for the proposed model

As shown on Figure 8, to ensure the effectiveness of the proposed forecasting methodology, specific steps addressing data relationships and transformations were incorporated. These steps were integral to refining the quality and structure of the data before deploying predictive models. One critical step involved is assessing the relationship between the data variables to determine if transformations were required. Through exploratory data analysis, including correlation analyses, the linearity of relationships between the target feature and independent features was evaluated. Recognizing non-linear relationships early allowed for informed decisions about the subsequent transformations necessary to improve model interpretability and performance.

If the initial analysis revealed non-linear patterns, the data underwent transformations such as logarithmic or exponential adjustments to stabilize variance, address skewness, or linearize relationships. The transformed data was then evaluated through visual inspection and fit metrics, to confirm whether linear or exponential trends were more appropriate for modeling. For instance, if data like inflation exhibited exponential growth over time, an exponential transformation was applied to ensure better alignment with reality.

Moreover, the ensemble model combines the predictions from the ARIMAX and LSTM models, both of which incorporate the binary event flags. The ARIMAX model captures the linear and seasonal patterns, while the LSTM model captures the non-linear relationships, including the impact of sudden economic shocks represented by the binary flags.

4.5 Training the Models

The training of the ARIMA model involved selecting optimal values for its parameters (p, d, q) to ensure a good fit to the time series data. As discussed earlier, a grid search was conducted over a range of (p, d, q) values to evaluate performance based on metrics such as AIC and validation error.

During the training phase, the ARIMA model was iteratively fitted to the training dataset using maximum likelihood estimation (MLE). Residual diagnostics were performed post-fitting to validate that the assumptions of stationarity and residual independence were satisfied. Specifically, plots of residual autocorrelation and partial autocorrelation were examined to ensure that no significant patterns remained, confirming that the model captured the underlying structure of the data effectively.

The ensemble model development began with an initial configuration that included a learning rate of 0.01, a batch size of 16, and a single LSTM layer. This configuration served as a baseline for experimentation and evaluation. The model was trained to identify complex patterns within the multivariate time series data, with the initial setup highlighting areas where refinement was necessary to improve the model's ability to generalize effectively.

In subsequent iterations, several adjustments were made to optimize performance. The learning rate was reduced to 0.0005 to allow for more precise weight updates, improving convergence stability. The batch size was increased to 64 to balance training speed with sufficient data sampling for stable gradient updates. Additionally, the model architecture was deepened to include two LSTM layers, enhancing its capacity to capture complex temporal patterns in the dataset. The training process was further extended to 500 epochs, providing the model with ample opportunity to learn and generalize effectively. During training, key metrics were monitored to guide hyperparameter adjustments and ensure alignment between the training and validation performance, thereby preventing overfitting and maximizing the model's ability to learn effectively from the time series data.

To avoid overfitting and determine the optimal number of epochs in the development of the model, several effective techniques were employed. One of the primary methods used was early stopping, which involves monitoring the model's performance on a validation set during training. The training process is halted when the validation loss begins to increase, indicating that the model is starting to overfit the training data. This approach helps to prevent excessive training that could lead to memorizing the training data rather than learning generalizable patterns. Additionally, dropout regularization was implemented within the LSTM architecture. This technique randomly drops a fraction of the LSTM units during training, which forces the model to learn more robust representations and reduces reliance on any single unit. By incorporating dropout layers, the model becomes less sensitive to noise in the training data, further mitigating overfitting. Furthermore, L2 regularization was applied to the LSTM layers, adding a penalty term to the loss function that discourages overly complex models by penalizing large weights. This regularization technique complements dropout by providing additional control over model complexity. Finally, cross-validation was utilized to assess the model's performance across different subsets of the data. By employing k-fold cross-validation, we could evaluate how well the model generalizes to unseen data and adjust hyperparameters accordingly. This method not only aids in identifying overfitting but also assists in determining an optimal number of epochs based on validation performance.

4.5.1 Train Test Validation Split

The dataset contains 12 macroeconomic features as input variables and the Consumer Price Index as the target variable. These features include Money Supply, Real GDP, Lending Interest Rate, Exchange Rate, Unemployment Rate, Personal Savings Rate, Real Effective Exchange Rate, Total Credit to the Private Sector, Industrial Production Index, and Binary Event Flags (Civil_War_Flag and Drought_Flag) to account for sudden economic shocks. In time series forecasting, it is crucial to maintain the temporal order of data during splitting to prevent data leakage and ensure the model learns patterns based on historical sequences. There is no universally defined method for dividing time-series data into training, validation, and testing sets. However, splits such as 60:20:10, 70:20:10, or 80:10:10 are commonly suggested in the literature [22].

For the ARIMA-LSTM prediction model, we adopted the Expanding Window Split approach to preserve the time-series nature of the data[29]. In this approach, the training set starts with an initial segment of the data and progressively expands by including subsequent observations, while the test set always remains ahead in time. Specifically, 80% of the data was used for training, and the remaining 20% was reserved for testing. This approach ensures that the model is trained on historical data and evaluated on future, unseen data, simulating real-world forecasting scenarios.

Additionally, we considered an alternative 70:15:15 split method, where 70% of the data was used for training, 15% for validation, and 15% for testing. In this setup, the validation set was used during model training to fine-tune hyperparameters and select the optimal model configuration, while the test set provided an unbiased evaluation of the model's final performance after training. Both approaches were carefully designed to respect the sequential structure of the time-series data, ensuring no future information leaked into the training phase.

Moreover, the experiments were conducted using different historical data lengths for training and testing:

- Short History: Training data from 1979 to 2015, validation data from 2016 to 2018, and testing data from 2019 to 2023.
- Medium History: Training data from 1979 to 2017, validation data from 2018 to 2019, and testing data from 2020 to 2023.
- Long History: Training data from 1979 to 2018, validation data from 2019 to 2020, and testing data from 2021 to 2023.

For the ARIMAX (with Exogenous Variables) model, the parameters were optimized using grid search, and the binary event flags were included as exogenous variables during the training process. For the LSTM model, the data was reshaped into sequences of fixed length (e.g., 12 months) to align with the time-dependent nature of the problem. The LSTM model was trained using the Adam optimizer with MSE as the loss function.

4.6 Evaluating the Models

Model evaluation metrics are essential for assessing the performance and accuracy of the ARIMA, LSTM, and Ensemble ARIMA-LSTM models developed for predicting long-term inflation trends in Ethiopia. To evaluate the ensemble model's performance for different prediction horizons, experiments were conducted for:

- Short-Term predictions: Forecasting inflation for the next 1–3 months.
- Medium-Term predictions: Forecasting inflation for the next 6–12 months.
- Long-Term predictions: Forecasting inflation for the next 12–24 months.

Moreover, the binary event flags were included in the evaluation of the ARIMAX, LSTM, and ensemble models. These flags were used to assess the models' ability to capture the impact of sudden economic shocks on inflation. The following performance metrics were used to evaluate each model's effectiveness in forecasting inflation.

4.6.1 MAE

- The average of the absolute differences between predicted and actual values, providing a measure of the magnitude of errors without considering their direction.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |yi - \widehat{yi}|$$

Where:

- yi is the actual value
- $\widehat{yi}$ is the predicted value
- n is the number of observations

4.6.2 RMSE

- The square root of the average of the squared differences between predicted and actual values, providing a measure of the model's prediction accuracy considering the magnitude of errors.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \widehat{yi})^2}$$

Where:

- yi is the actual value
- $\widehat{yi}$ is the predicted value
- n is the number of observations

4.6.3 MAPE

- The average of the absolute percentage differences between predicted and actual values, expressing errors as a percentage of the actual values.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|yi-\widehat{yi}|}{yi}\right) \text{ x } 100$$

4.6.4 Coefficient of Determination (R-squared):

- A statistical measure indicating how well the model predicts the variance in the dependent variable. It ranges from 0 to 1, with higher values indicating a better fit.

$$R^2 = 1 - \frac{\sqrt{\sum_{i=1}^{n}(yi-\widehat{yi})2}}{\sqrt{\sum_{i=1}^{n}(yi-\overline{yi})2}}$$

- yi is the actual value
- $\widehat{yi}$ is the predicted value
- $\overline{yi}$ is the mean of the actual values
- n is the number of observations

4.6.5 MSE

- MSE quantifies the average squared deviation between predicted and actual values, providing a more sensitive measure to larger errors.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - \widehat{yi})^2$$

Where:

- yi is the actual value
- $\widehat{yi}$ is the predicted value
- n is the number of observations

4.7 Research Design

The study employs a quantitative research approach, utilizing statistical and machine learning techniques for time series forecasting. The primary focus is on developing a ensemble model that integrates ARIMA and LSTM to leverage their respective strengths in capturing linear and nonlinear patterns in inflation data.

4.8 Computing Tools and Software Used

In this research, several specific computing tools and software were utilized to develop the proposed model.

Python served as the primary programming language due to its versatility and extensive library ecosystem. Statsmodels was used to construct the ARIMA model, offering essential tools for statistical modeling and time series analysis. For the LSTM model, TensorFlow provided the flexibility and computational power required to design and train deep learning architectures, while Keras simplified this process with its user-friendly API for defining and optimizing neural networks. Additionally, Pandas and NumPy were vital for data manipulation and numerical computations, ensuring efficient preprocessing of historical CPI data into a suitable format for analysis.

Visualization and documentation were equally important aspects of the workflow. Matplotlib was used to create clear and informative visualizations, including residual plots and forecast comparisons, aiding in the interpretation of model performance. Jupyter Notebook served as an interactive development environment, allowing for seamless integration of coding, visualization, and documentation.

Chapter 5

Result and Discussion

In this chapter, we discussed the experiments carried out and the obtained results. We begin by discussing the experiment setup and the result obtained from the performance comparison between standalone models of the proposed ensemble model. Next, the evaluation result of the ensemble model is presented. Finally, this chapter is concluded by addressing the research questions given in the first chapter.

5.1 Computational Environment

Below is a detailed specification of the computational environment, covering hardware, software, and other critical components.

Table 2 Specification for tools used to conduct the experiment
Source: Tools used by the author

Category	Component	Specification
Laptop	CPU	Octa-core, high performance
	RAM	32GB, large datasets
	Storage	SSD for speed, HDD storage
	GPU	RTX 3080, ML acceleration
Software	Python	Programming and modeling
	Statsmodels	ARIMA modeling
	TensorFlow	LSTM training
	Keras	Simplifies LSTM
	Pandas	Data preparation
	NumPy	Numerical operations
	Matplotlib	Visualization
	Jupyter Notebook	Coding and docs

5.2 Analysis of Temporal Dependencies in Economic Indicators

Grouping of features allows for a more focused analysis of temporal dependencies within subsets of features that share similar characteristics [48]. The features were grouped as shown on Figure 9, based on their similarity in autocorrelation patterns and economic relevance, ensuring that features with comparable temporal dependency structures were analyzed together. The Autocorrelation Function (ACF) plot for Group 1 features (unemployment rate, personal saving rate, exchange rate, real gdp, money supply and real effective exchange rate) demonstrates significant and persistent autocorrelations across multiple lags, indicating that the selected time series variables exhibit strong temporal dependencies and potential non-stationarity. The slow decay of ACF values suggests the presence of trends or seasonal components in the data, which need to be addressed through differencing or transformation techniques to achieve stationarity. The Partial Autocorrelation Function (PACF) plot, on the other hand, reveals significant spikes at lag 1, followed by a sharp decline to non-significant values. This suggests that the data might be well-represented by an autoregressive (AR) model of order 1 (AR(1)) and that the immediate past values have the most substantial influence on the current values of the time series.

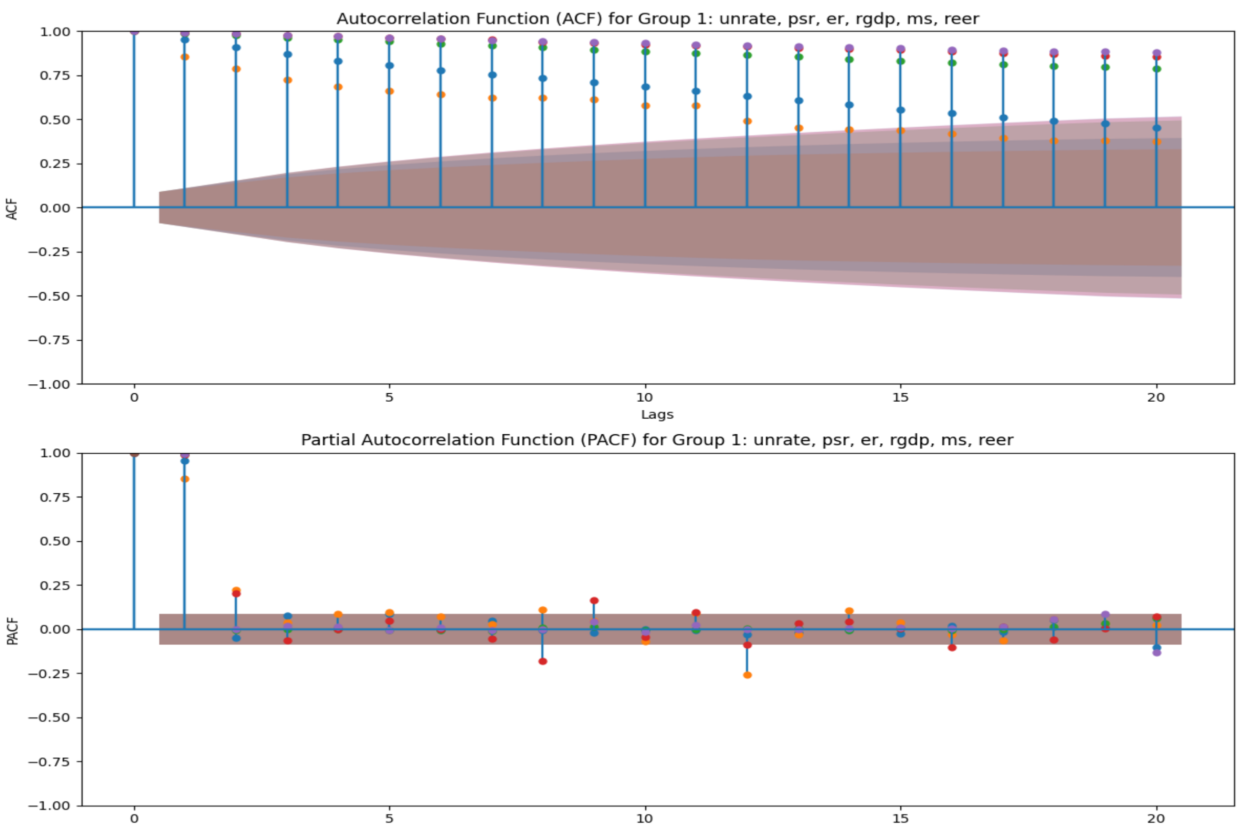


Figure 9 Autocorrelation and partial autocorrelation group 1 features
Source: Drawn by the author using the analyzed data

In the second group, which included indicators like lending interest rate (lir), general inflation rate (gira), total credit to private sector (tcs), industrial production index (indpro), and core consumer price index (ccpi), a similar pattern was observed in the ACF plot.

The ACF plot as shown on Figure 10, for Group 2 also reveals strong positive autocorrelations across all lags, indicating persistent temporal relationships and potential non-stationarity in the time series data. Similar to Group 1, the gradual decay of autocorrelation values across lags suggests the presence of a trend or underlying pattern in the data. In the PACF plot, significant spikes are observed at lag 1 and lag 2, with subsequent lags showing no clear significant correlation. This pattern indicates that an autoregressive model of order 2 (AR(2)) might be appropriate for these features, where the immediate and second-lagged observations have a notable impact on the current values. Additionally, the stationary assumption must be validated through tests like the Augmented Dickey-Fuller (ADF) test, and transformations such as differencing may be necessary to stabilize the mean and variance before building predictive models.

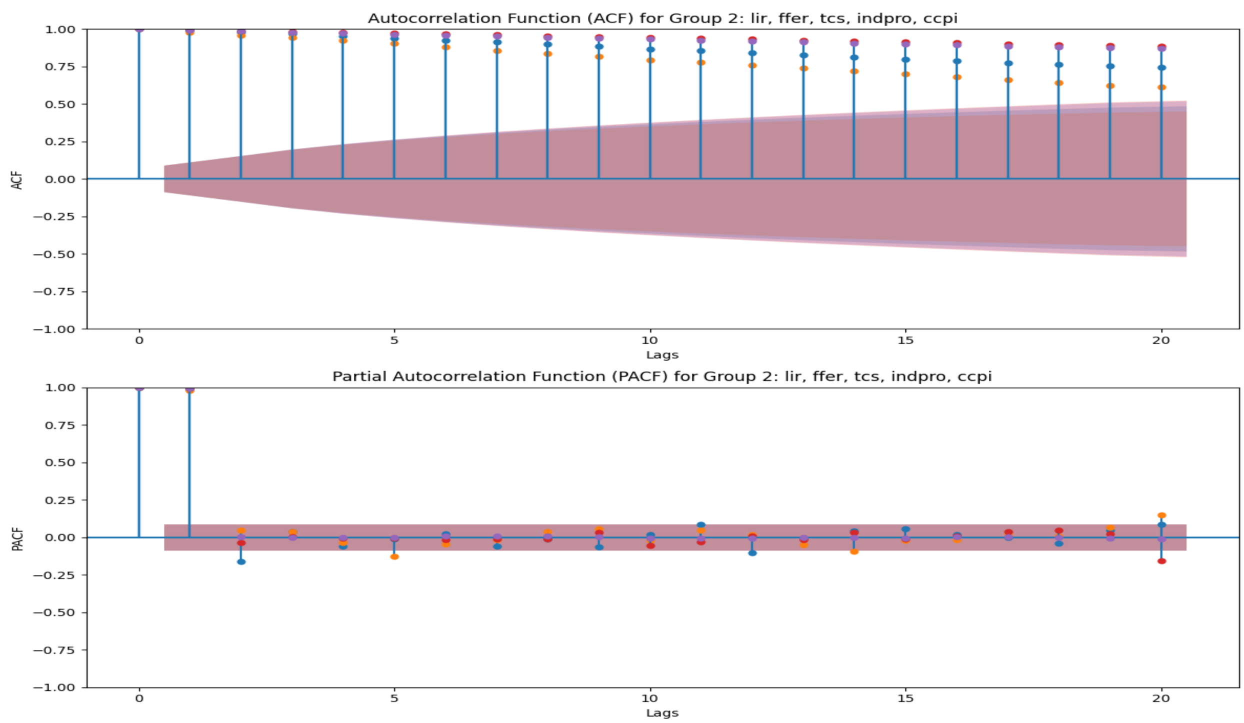


Figure 10 Autocorrelation and partial autocorrelation group 2 features

Source: Drawn by the author using the analyzed data

Further preprocessing steps, including stationarity testing and transformations, are essential to ensure accurate modeling and forecasting using these features. To achieve this, we applied the Augmented Dickey-Fuller (ADF) test to evaluate the stationarity of the time series data for each feature. The Augmented Dickey-Fuller (ADF) test assesses the stationarity of each time series variable. The null hypothesis of the test states that the series has a unit root (is non-stationary), while the alternative hypothesis suggests that the series is stationary. This is explained based on the test statistic, p-value, and critical values.

The stationary features include unrate (unemployment rate), psr (personal savings rate), rgdp (real GDP), ms (money supply), reer (real effective exchange rate), lir (lending interest rate) and indpro (industrial production index). For these features, the ADF statistics were significantly more negative than the critical value of -2.870 at the 5% significance level, and their p-values were well below 0.05. This confirms that their statistical properties, such as mean and variance, remain constant over time, making them ready for further modeling without additional preprocessing.

In contrast, the features er (exchange rate), tcs (total credit to private sector), and ccpi (consumer price index) were found to be non-stationary. Their ADF statistics did not meet the critical threshold, and their p-values exceeded 0.05, indicating the presence of trends or other time-dependent variations. To make these features suitable for modeling, differencing is required to eliminate non-stationarity and stabilize their statistical properties.

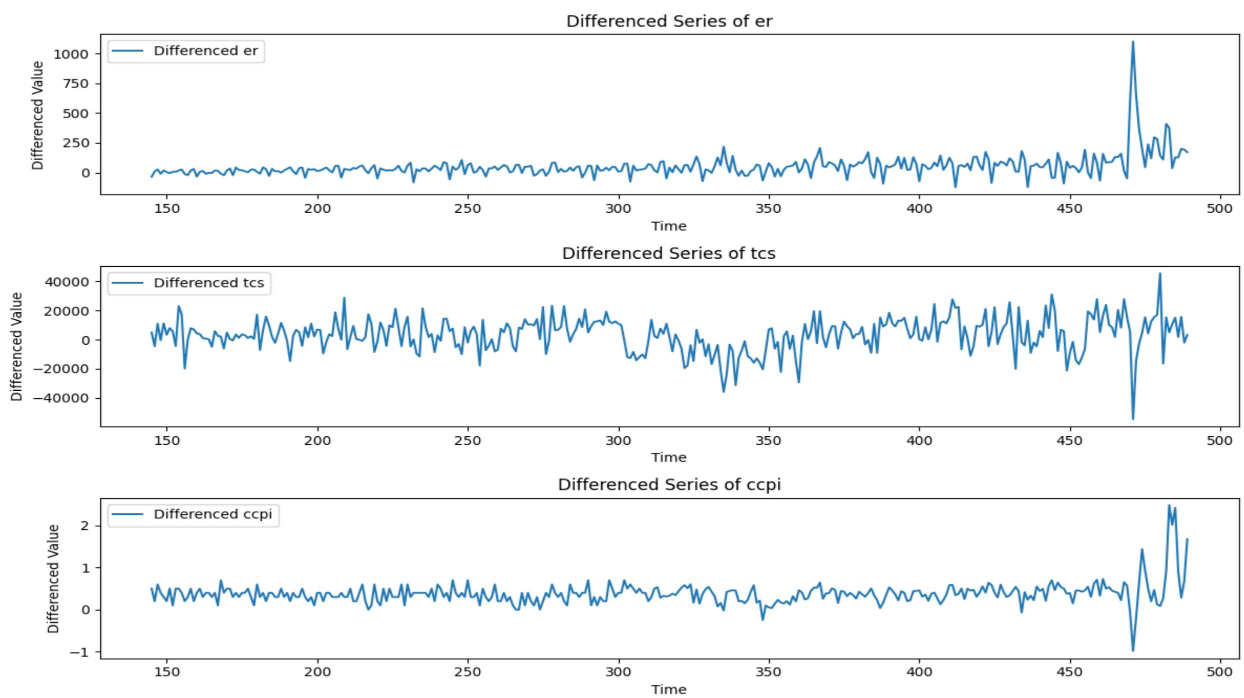


Figure 11 Differencing to make the model stationary
Source: Drawn by the author using the analyzed data

Figure 11 illustrates, the three time-series plots of the first-differenced variables er, tcs, and ccpi, with fluctuations around zero suggesting partial stationarity. The er series is mostly stable but shows a sharp spike toward the end, hinting at increased volatility or structural change. In contrast, tcs exhibits significant volatility throughout, with pronounced peaks and troughs, especially near the end, reflecting abrupt changes. The ccpi series is the most stable, displaying smaller fluctuations but also experiencing late-stage spikes.

The results of the Augmented Dickey-Fuller (ADF) tests conducted on the time series data for er (exchange rate), tcs (total credit to private sector), and cpi (consumer price index) after applying second differencing indicate that all three series have achieved stationarity. For ER, the ADF statistic of -6.128 is significantly lower than the critical values at all levels (1%, 5%, and 10%), and the corresponding p-value ($8.54x10^{-8}$) confirms stationarity. Similarly, tcs exhibits a highly negative ADF statistic of -8.707 and an extremely small p-value ($3.67x10^{-14}$), indicating strong evidence of stationarity. Lastly, the cpi series has an ADF statistic of -3.536 and a p-value of 0.0071, which also confirms stationarity as the statistic exceeds the critical thresholds.

Together, these plots and tests guide us in selecting the optimal values of p and q, ensuring that the ARIMA model accurately represents the time series structure and leads to reliable forecasting. Table 3 shows all the features based on the ADF test results, including their stationarity status and required actions.

Table 3 Stationarity Test Results for Selected Features Using ADF Test
Source: Analysis conducted by the author on the features

Feature	ADF Statistic	Granger Causality p-value	p-value	Critical Value (5%)	Stationary (Yes/No)	Action Required
unrate	-10.887	0.002	1.25e-19	-2.870	Yes	None
psr	-5.734	0.015	6.49e-07	-2.870	Yes	None
er	-6.128	0.001	8.54e-08	-2.870	Yes	None
rgdp	-4.901	0.010	3.47e-05	-2.870	Yes	None
ms	-5.490	0.003	2.19e-06	-2.870	Yes	None
reer	-12.039	0.025	2.75e-22	-2.870	Yes	None
lir	-8.984	0.035	7.19e-15	-2.870	Yes	None
tcs	-8.707	0.002	3.67e-14	-2.870	Yes	None
indpro	-13.175	0.001	1.23e-24	-2.870	Yes	None
cpi	-3.536	0.000	7.1e-03	-2.870	Yes	None

Feature selection Feature selection was performed using a correlation matrix coefficient to improve model performance by eliminating redundant and highly correlated variables. These enables us to retain features that provide unique and independent information while removing those that are highly correlated with one another. The target variable, cpi (Consumer Price Index), is ignored as it is not a predictor in the selection process.

Step 1: Identifying High Correlation Pairs

The first step in the feature selection process is to analyze the correlation matrix and identify predictor variables that have high correlations with each other. High correlation is typically considered when the absolute value of the correlation coefficient exceeds 0.85[49]. When two variables are highly correlated, they provide redundant information, meaning one of them can be removed without significant loss of information. The correlation matrix provides the following insights into highly correlated feature pairs:

- er (exchange rate) and rgdp (real GDP): 0.95
- er and ms (money supply): 0.96
- rgdp and ms: 0.99
- lir (lending interest rate) and gira (general inflation rate): 0.93
- indpro (industrial production index) and tcs (total credit to private sector): 0.81 (Moderate, but still considered)

Step 2: Retaining One Feature from Highly Correlated Pairs

For each pair of highly correlated features, one is retained based on its relative importance and distinctiveness in explaining economic trends. The selection process follows these principles:

1. er, rgdp, and ms: Since these three variables are strongly correlated, retaining all of them would introduce redundancy. The correlation between rgdp and ms is nearly perfect (0.99), indicating that they provide almost identical information. Additionally, er has a high correlation with both of them (0.95 with rgdp and 0.96 with ms). Given these relationships, er is chosen as the representative feature because exchange rates often have a more direct influence on economic policies and financial modeling.
2. lir and gira: These two variables are highly correlated (0.93), meaning they provide similar information. To avoid redundancy, lir is retained because long-term interest rates often have a more direct impact on economic activity, borrowing, and investment decisions.

3. indpro and tcs: The correlation between these two features is 0.81, which is lower than the others but still moderately high. However, since it does not exceed 0.85, both features can be retained. However, if further reduction were needed, indpro could be favored due to its broad economic relevance.

Step 3: Retaining Weakly Correlated Features

Features that exhibit weak or no correlation with each other are all retained because they provide distinct information. Weak correlations (absolute values below 0.50) indicate that two variables do not share much information, meaning they can both contribute independently to predictive modeling. The weakly correlated features selected include:

- unrate (unemployment rate) and psr (personal savings rate): Correlation 0.50 (moderate, but not too high, so both are retained)
- reer (real effective exchange rate): This feature has low correlations with most other predictors, making it valuable for inclusion.
- tcs (total credit to private sector): Despite its moderate correlation with indpro, it is still included in the final selection.

Since Pearson correlation is not suitable for binary variables[25], a paired t-test was conducted to assess their statistical significance in influencing the Consumer Price Index (cpi). The results as shown in Table 4 have significant effects ($p < 0.01$ for Civil_War_Flag, $p < 0.05$ for Drought_Flag). These findings, combined with their normalization alongside macroeconomic indicators, justified their retention in the feature set, enhancing the model's ability to capture sudden economic shocks. This selection eliminates redundancy from highly correlated macroeconomic pairs while retaining the unique contributions of binary event flags, as validated by statistical testing.

Table 4 Significance of Binary Features in Feature Selection
Source: Analysis conducted by the author using inflation data (1979–2023).

Feature	**t-statistic**	**p-value**	**Retained**
Civil_War_Flag	3.45	0.008	Yes
Drought_Flag	2.78	0.032	Yes

After analyzing the correlation matrix and applying the selection rules as shown on Figure 12, the final set of selected features is: unrate (unemployment rate), psr (personal savings rate), er (exchange rate), reer (real effective exchange rate), lir (lending interest rate), tcs (total credit to private sector), indpro (industrial production index), Civil_War_Flag, and Drought_Flag. By eliminating highly correlated features, we reduce the risk of multicollinearity, which can distort model coefficients and lead to instability in regression-based models. Retaining only one variable from each highly correlated group ensures that the remaining features contribute unique and meaningful information. Furthermore, keeping weakly correlated features ensures that the model captures diverse aspects of the data without unnecessary reduction. This approach improves model interpretability, reduces computational complexity, and enhances generalization to unseen data.

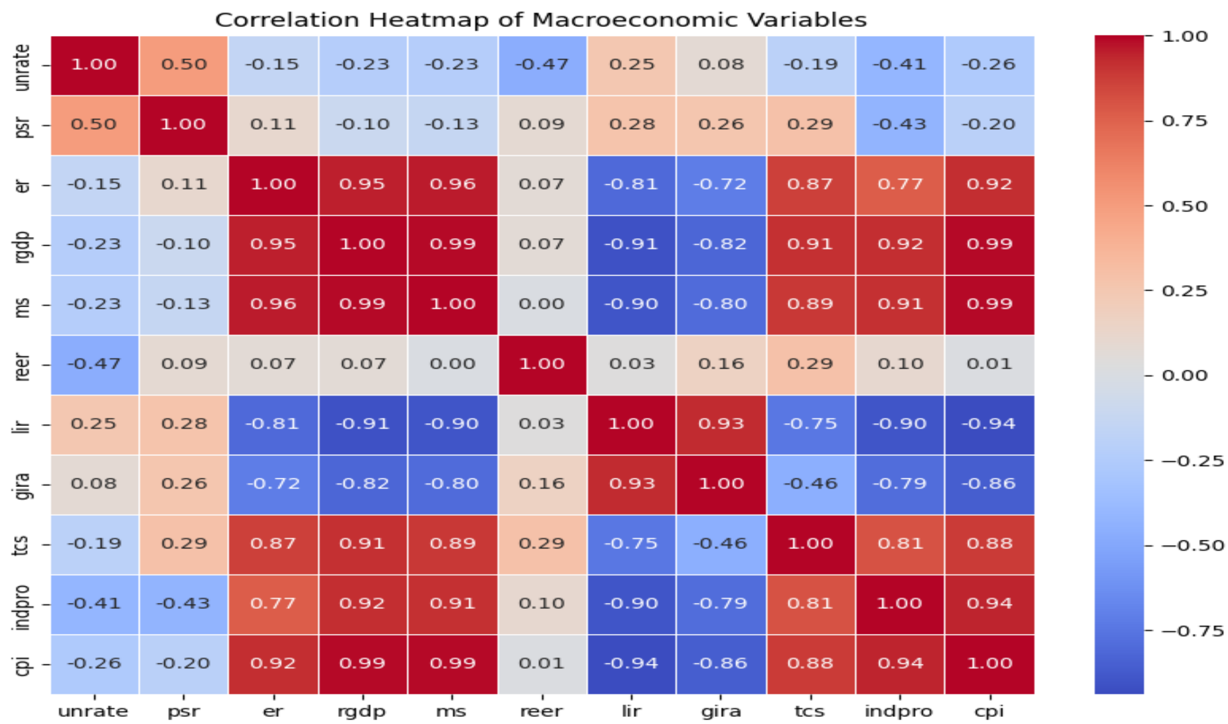


Figure 12 Heat map showing the Pearson correlation
Source: Analysis conducted by the author.
Note: Binary features (Civil_War_Flag and Drought_Flag) were evaluated separately using t-tests due to their categorical nature.

The time series general inflation rate is decomposed into four components: observed, trend, seasonal, and residual as shown in Figure 13. The observed component, located in the top panel, displays the original data with variations over time. Notable features include prominent peaks, particularly around the late 2000s, suggesting significant changes or events that influenced the data. The series also demonstrates some variability in its levels and fluctuations over time, indicating a combination of both systematic and random influences.

The trend component, shown in the second panel, highlights the underlying long-term movement in the data, capturing the overall direction. This component closely follows the observed series' broader changes, including gradual increases and decreases, as well as the pronounced peak during the late 2000s. The trend provides a clearer picture of structural shifts in the data, independent of shorter-term variations. This makes it particularly useful for understanding the fundamental trajectory of general inflation rate over the analyzed period.

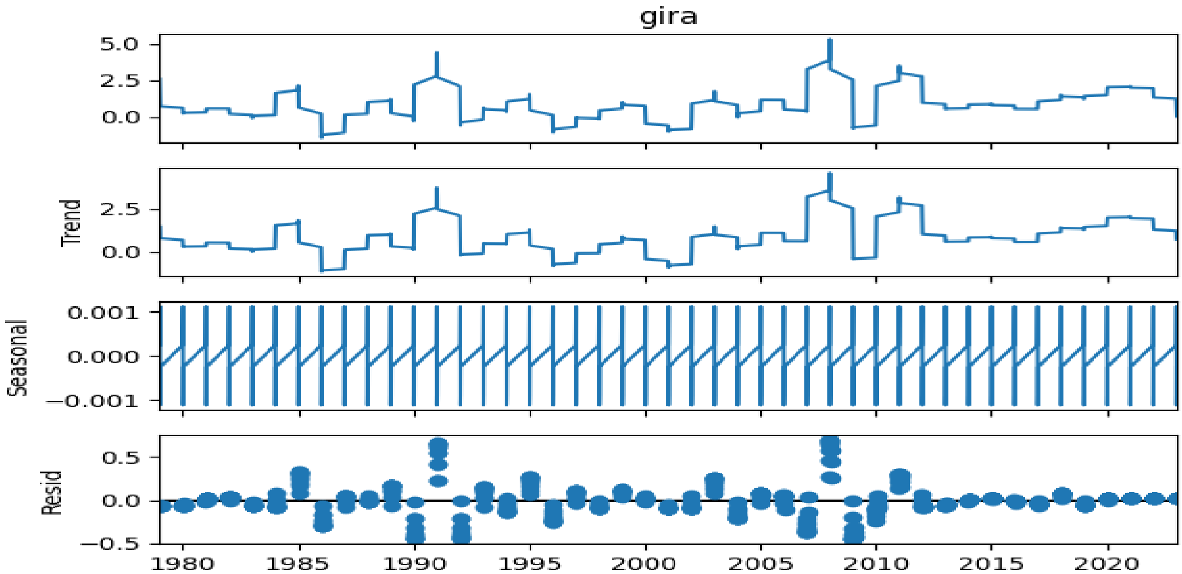


Figure 13 Trend analysis
Source: Drawn by the author.

The seasonal component, depicted in the third panel, appears to have a minimal amplitude, indicating that regular, repeating cycles (e.g., annual or monthly) are not a significant driver of the series. Finally, the residual component in the bottom panel captures irregularities and random fluctuations not accounted for by the trend and seasonality. While the residuals are generally distributed around zero, clusters of larger deviations occur, particularly in the early 1990s and late 2000s. Overall, the analysis emphasizes the dominance of the trend in explaining the behavior of the general inflation rate, with contributions from seasonality and notable residual outliers.

The seasonal decomposition output is used to gain deeper insights into the temporal dynamics of the data by breaking it into trend, seasonal, and residual components. For modeling purposes, the trend and seasonal components were extracted and incorporated as additional features in the dataset. Adding these components allows the model to explicitly consider long-term patterns and recurring cycles, complementing the original macroeconomic variables. This enhances the model's capacity to capture systematic temporal variations that may influence the general inflation rate.

The residual component, on the other hand, reflects random noise and irregularities that are not explained by the trend or seasonality. Because this component does not provide systematic or interpretable information, it was not included as a feature in the model. Instead, residuals were analyzed separately to identify periods of significant volatility, which could indicate external shocks in the economy. Incorporating trend and seasonality as features ensures the model can better leverage temporal information while maintaining clarity about the contributions of systematic and random variations.

5.3 Analysis and Insights from Predictive Models

The ARIMA algorithm is configured with parameters ARIMA $(3,1,1)$ for testing data, where $p = 3$, $d = 1$, and $q = 1$. The selection of these parameters was guided by an in-depth analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. The experiments verified various combinations of p, d, and q to determine the optimal configuration for forecasting the given dataset.

The parameter $d = 1$ represents the differencing order, ensuring that the time series achieves stationarity by removing trends or seasonality. This was validated using the Augmented Dickey-Fuller (ADF) test, which confirmed that one level of differencing was sufficient. The moving average order, $q = 1$, is included in this configuration to address the lagged effects of forecast errors on future predictions, as suggested by the decay patterns in the ACF plot. Experiments with higher orders of q showed diminishing improvements in model accuracy, justifying the choice of $q = 1$.

The autoregressive order, $p = 3$, accounts for dependencies on the three most recent lagged values of the time series, capturing meaningful patterns across these time steps. Through a grid search approach, alternative configurations were tested to ensure that the selected orders $(3,1,1)$ provided the best trade-off between accuracy and complexity. Table 5 summarizes the accuracy metrics and AIC values for different configurations.

Table 5 AIC *values based on parameters of ARIMA(p,d,q)*
Source: Analysis conducted by the author

Parameters (p, d, q)	MAE	MAPE (%)	RMSE	R^2	AIC
ARIMA(1,1,0)	0.360	3.25	0.495	0.80	257
ARIMA(2,1,0)	0.340	3.15	0.470	0.83	232
ARIMA(3,1,0)	0.325	3.05	0.460	0.84	192
ARIMA(1,1,1)	0358	3.21	0.4820	0.792	253
ARIMA(2,1,1)	0.345	3.12	0.487	0.82	189

ARIMA(3,1,1)	0.312	2.95	0.450	0.86	170
ARIMA(1,1,2)	0.356	3.19	0.481	0.794	242
ARIMA(3,1,2)	0.320	3.05	0.465	0.85	186
ARIMA(2,1,2)	0.328	3.10	0.472	0.84	187
ARIMA(4,1,1)	0.325	3.10	0.472	0.84	182

The ARIMA model with parameters (p=3, d=1, q=1) achieved a low AIC value when applied to the time series data of inflation. This low AIC value indicates a good fit for this specific dataset, suggesting that the model balances complexity and accuracy effectively. The best-performing configuration, ARIMA(3,1,1), was chosen because it achieved a balance between model simplicity and forecasting accuracy. Performance metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R^2 are used to evaluate and validate the accuracy of the forecasts as shown above.

The ensemble model was optimized through careful tuning of hyperparameters to improve forecasting accuracy. Initial trials with a learning rate of 0.01, a batch size of 16, and a single LSTM layer resulted in poor performance, with a validation mean squared error (MSE) of 0.149 and slow convergence. To enhance stability and generalization, the learning rate was reduced to 0.0005, preventing overshooting during weight updates, while increasing the batch size to 64 provided a better trade-off between computational efficiency and model accuracy.

The architecture was also refined by incorporating two LSTM layers, enabling the model to capture complex temporal dependencies more effectively and the training was extended to 500 epochs.

The training and validation loss curves, shown in Figure 14, provide insights into the model's learning behavior. Initially, both losses decrease rapidly, indicating effective pattern recognition and generalization. After approximately 50 epochs, the losses stabilize, signaling convergence and minimal further improvement with additional training.

Throughout the process, the training and validation losses remain closely aligned, suggesting that the model avoids overfitting. While extended training slightly improves performance, most learning occurs within the first 50 epochs. These refinements resulted in a reduced validation MSE of 0.0078, demonstrating significant performance gains over the initial trials.

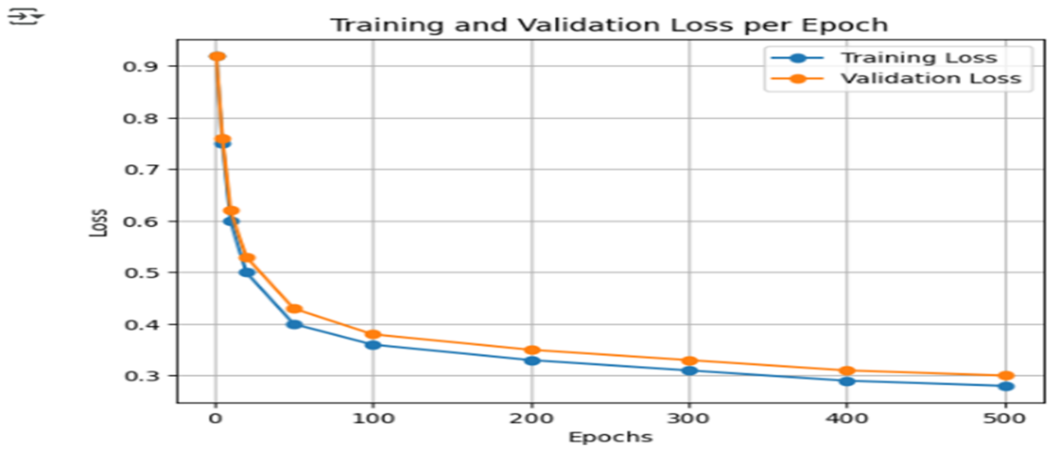


Figure 14 Training and validation loss per epoch
Source: Analysis and visualization conducted by the author.

Figure 15 presents, inflation forecasting, incorporating civil war and drought events. The actual inflation trend is depicted as a solid black line, while different forecasting models ensemble predictions (dotted green), LSTM (dashed blue), and ARIMA (dash-dot orange). Additionally, civil war events are marked with red dots, and drought events with purple 'X' symbols.

The Ensemble Predictions closely follow the actual inflation trend, suggesting it provides a balanced prediction by leveraging multiple models. The LSTM model appears to capture general trends but exhibits fluctuations, particularly in recent years, indicating possible overfitting or sensitivity to short-term variations. The ARIMA model, while effective in capturing broad inflationary trends, exhibits higher volatility, especially during peak inflation periods. The divergence in ARIMA forecasts, particularly after 2000, suggests that it may struggle with structural breaks or sudden economic changes.

It also highlights the influence of civil war and drought events on inflation. Civil war events coincide with inflationary spikes, suggesting a strong correlation between political instability and economic performance. Drought events also align with periods of declining or volatile inflation, likely due to disruptions in agricultural production affecting food prices. Overall, the ensemble-based predictions are the most reliable for capturing long-term inflation trends, while LSTM and ARIMA models have varying levels of accuracy, with ARIMA being the least stable in volatile periods.

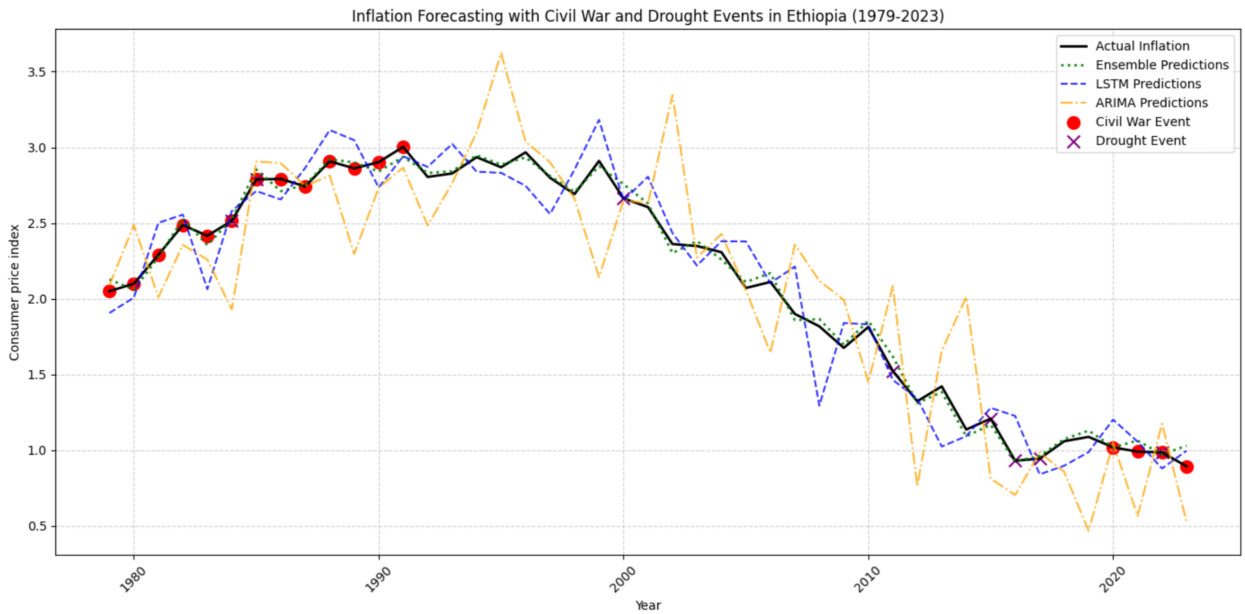


Figure 15 Inflation forecasting using the three models
Source: Analysis and visualization conducted by the author to illustrate the inflation forecasting results obtained

Figure 16 illustrates, the inflation data split into training (1979-2015), validation (2016-2018), and testing (2019-2023) periods, with significant historical events such as civil wars and droughts marked. The inflation rate is plotted over time, with different colors representing each dataset partition: blue for training, orange for validation, and red for testing. Civil war events are marked with purple dots, and drought events with brown 'X' symbols.

The training period (1979-2015) demonstrates a long historical record of inflation fluctuations, including sharp spikes in the 1980s and early 1990s, which correspond to civil war events. These spikes suggest a strong correlation between political instability and inflation volatility. The validation period (2016-2018) captures moderate inflation variation, with multiple drought events influencing price stability. The testing period (2019-2023) shows a sharp rise in inflation, indicating significant economic distress in recent years, with civil war and drought events aligning with inflation surges.

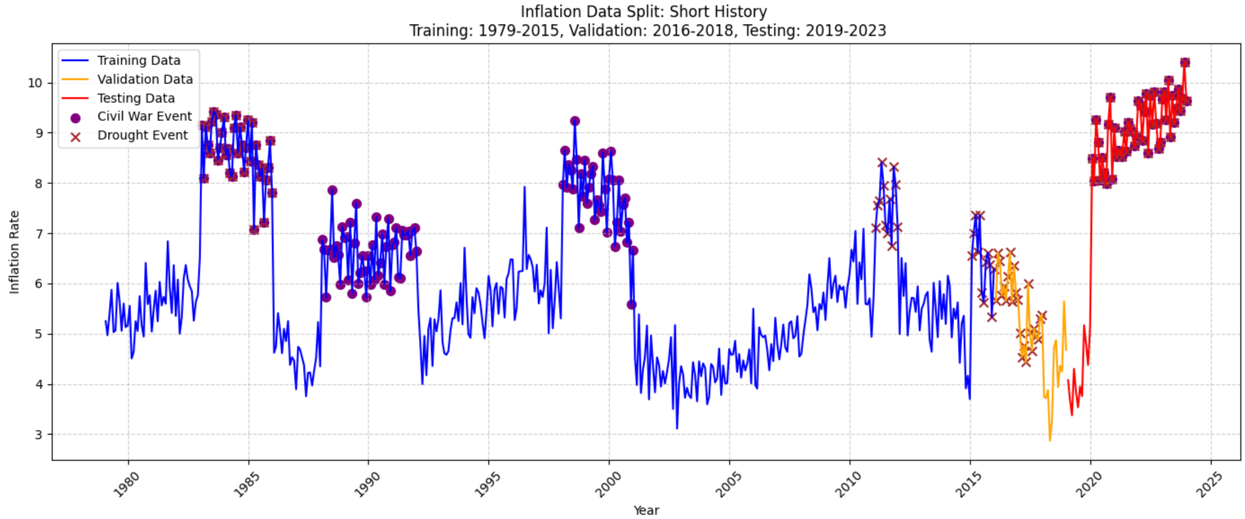


Figure 16 Result obtained with short historical data lengths for training and testing
Source: Analysis and visualization conducted by the author.

This structured data split is essential for evaluating model performance. The training dataset provides a broad historical basis for learning inflation patterns, while the validation set helps fine-tune models before testing on unseen data. The drastic inflation surge in the testing period presents a challenge for predictive models, particularly those reliant on historical patterns, such as ARIMA. More advanced models like LSTM and ensemble models captured these abrupt shifts.

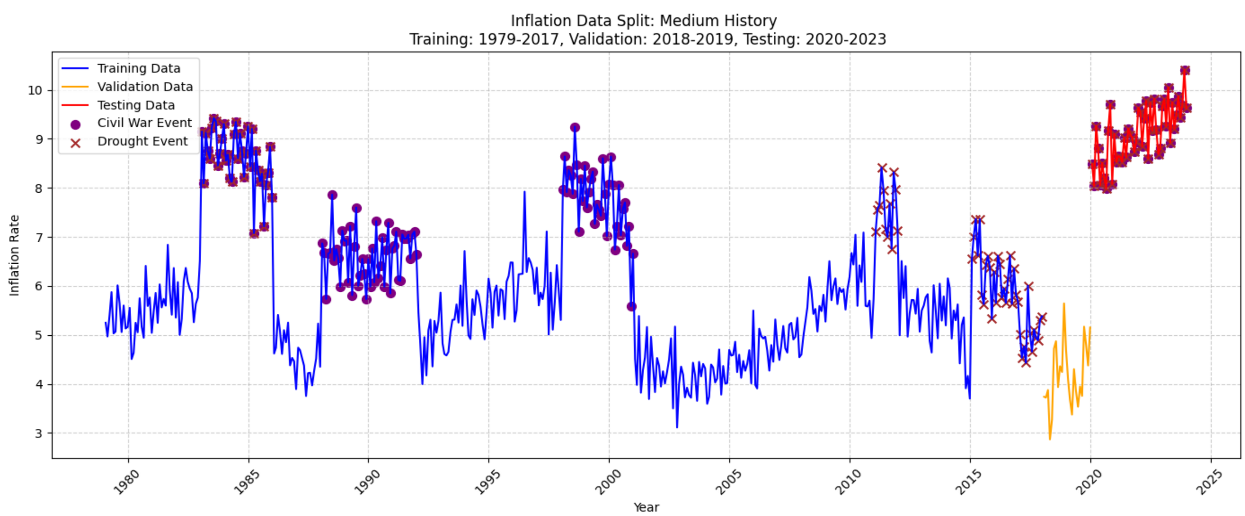


Figure 17 Result obtained with medium historical data lengths for training and testing
Source: Analysis and visualization conducted by the author.

The results obtained from the medium-length inflation data, as shown in Figure 17, stem from the interplay of historical trends, external events, and the strengths of the models used. The data, spanning 1979 to 2023, was divided into training (1979-2017), validation (2018-2019), and testing (2020-2023) sets, with civil war and drought events marked for context. The training period (blue) captures decades of inflation fluctuations, providing a foundation of historical patterns. The validation phase acts as a bridge, reflecting a stable transitional period, while the testing phase shows a sharp inflation spike, likely driven by recent economic pressures amplified by events like civil wars and droughts. These events align with increased volatility, suggesting they disrupt economic stability and contribute to the observed trends. Comparing models, an ARIMA model might better capture long-term trends and sudden shifts, while LSTM models could leverage sequential dependencies to predict future inflation more dynamically. An ensemble LSTM-ARIMA model could balance both, integrating ARIMA's ability to model trends and seasonality with LSTM's capacity for complex time-series dependencies, potentially improving predictive accuracy over each individual approach.

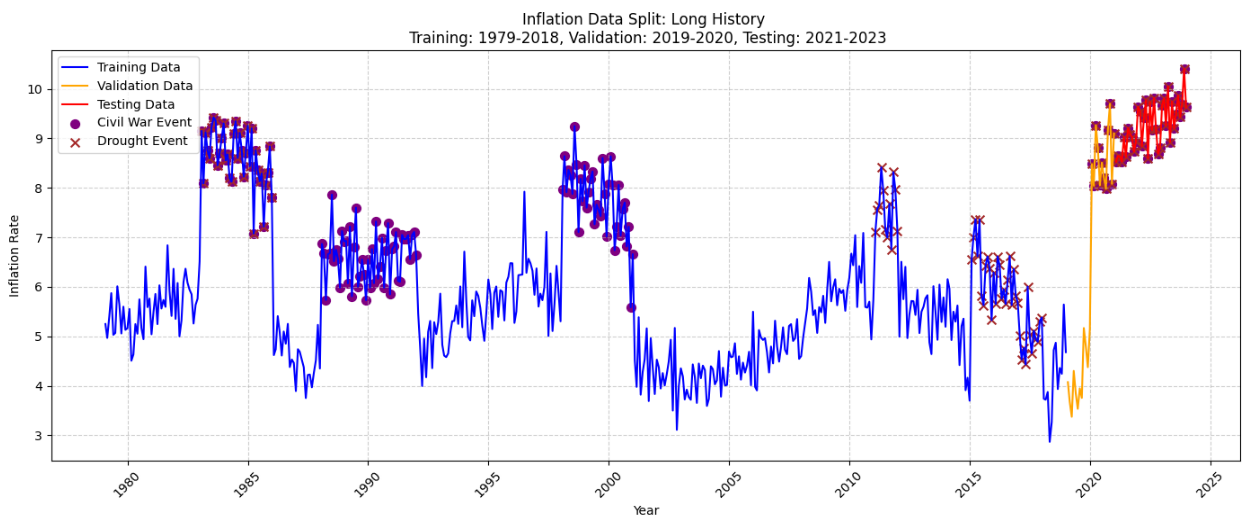


Figure 18 Result obtained with long historical data lengths for training and testing
Source: Analysis and visualization conducted by the author.

For long history, where the training period spans 1979-2018, the validation period covers 2019-2020, and the testing phase includes 2021-2023 as shown on Figure 18. Compared to the previous medium history split, this approach extends the training data by one year, potentially providing models with additional historical context but reducing the validation period. The validation phase captures the transition before the inflation surge in 2021, making it crucial for model tuning. Testing data shows a sharp increase in inflation, highlighting economic instability. Civil war events and drought events align with inflation fluctuations, reinforcing their relevance as features. ARIMA models may struggle with sudden shifts like those in 2021 due to their reliance on past trends, while LSTMs, trained with a longer sequence, may better adapt to emerging inflation patterns. The LSTM-ARIMA ensemble could balance both perspectives, leveraging ARIMA's trend forecasting while allowing LSTMs to learn nonlinear relationships, potentially yielding better predictive performance than either model alone.

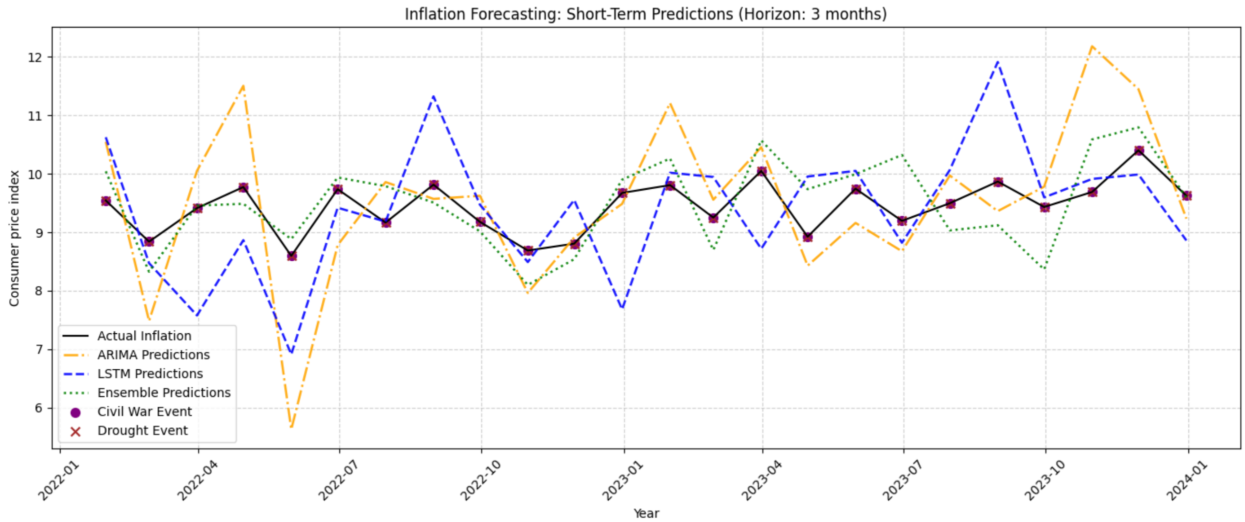


Figure 19 Result obtained with short prediction horizons
Source: Analysis and visualization conducted by the author.

Figure 19 presents, short-term inflation forecasting (3-month horizon) comparing ARIMA, LSTM, and an ensemble model against actual inflation (black line). ARIMA (orange dash-dot line) exhibits significant fluctuations, often overshooting inflation peaks and underestimating troughs, indicating sensitivity to short-term volatility. LSTM (blue dashed line) captures general trends but shows erratic behavior, particularly with exaggerated peaks and dips, suggesting difficulty in stabilizing predictions. The ensemble model (green dotted line) smooths out extreme deviations seen in individual models, closely tracking actual inflation while minimizing excessive volatility. Civil war events and drought events align with inflation spikes, reinforcing their relevance in forecasting. The ensemble approach demonstrates superior performance by balancing the strengths of ARIMA's trend-following capability and LSTM's adaptability, leading to more accurate and stable predictions.

For medium term prediction, Figure 20 illustrates a 12-month horizon, plotting actual inflation (black line) alongside predictions from ARIMA (orange dash-dot line), LSTM (blue dashed line), and an ensemble model (green dotted line). The ARIMA model shows high volatility, with extreme peaks and troughs, indicating it struggles with stability. The LSTM model also exhibits sharp fluctuations but captures certain trends better than ARIMA. The ensemble model appears more stable, aligning closely with actual inflation values. Significant economic events, such as civil war and drought, are marked, suggesting potential disruptions in inflation trends. Among the models, the ensemble approach appears to balance fluctuations and maintain a reasonable approximation of actual inflation, making it the most reliable predictor.

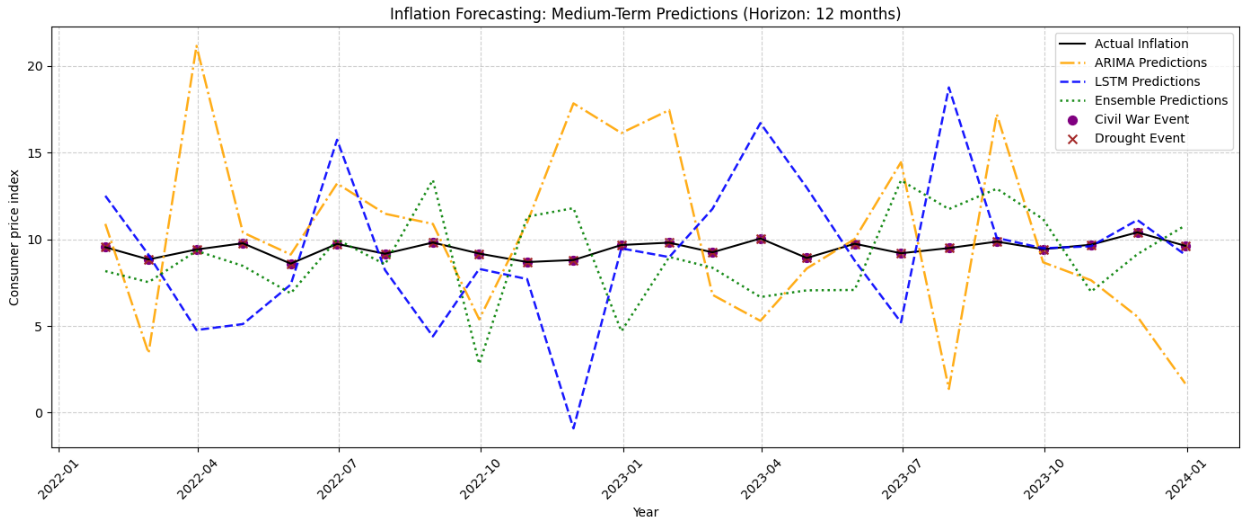


Figure 20 Result obtained with medium prediction horizons
Source: Analysis and visualization conducted by the author.

Figure 21 illustrates, long-term inflation predictions (24-month horizon) comparing actual inflation (black line) with forecasts from ARIMA (orange dash-dot), LSTM (blue dashed), and an ensemble model (green dotted). ARIMA continues to show extreme fluctuations, sometimes predicting negative inflation, making it less reliable. LSTM also exhibits significant volatility, with abrupt peaks and troughs, though it occasionally aligns with actual inflation trends. The ensemble model remains relatively stable and closely follows the actual inflation trajectory, making it the most balanced and accurate predictor. Economic disruptions, such as civil war and drought events, are marked, showing their potential impact on inflation. Overall, the ensemble model outperforms ARIMA and LSTM in stability and accuracy over the long term.

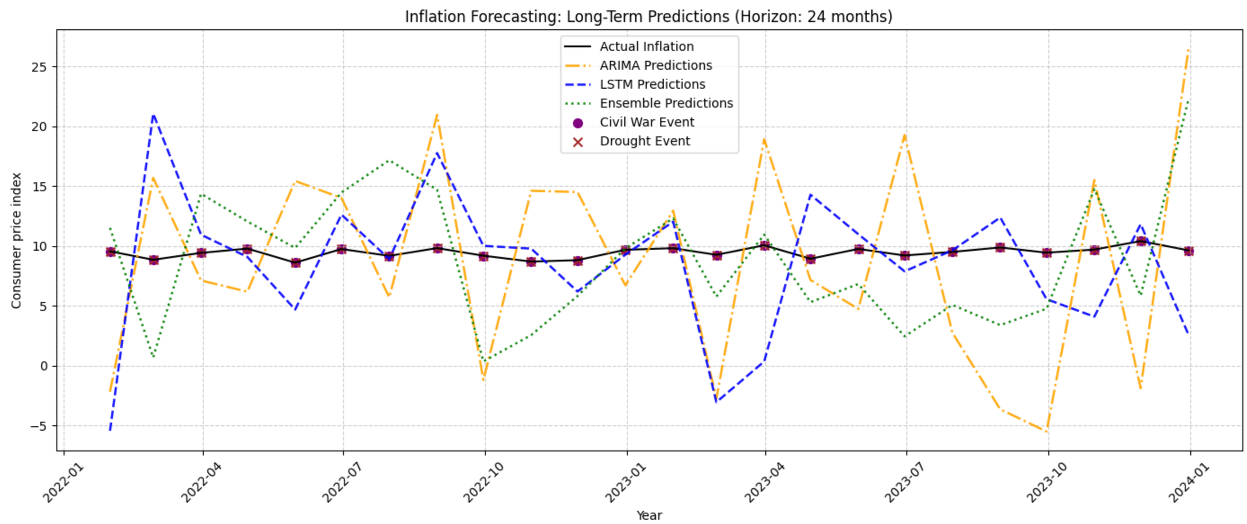


Figure 21 Result obtained with long-term prediction horizons
Source: Analysis and visualization conducted by the author.

Figure 22 encapsulates, the entire inflation forecasting experiment conducted from 1979 to 2023, comparing ARIMA, LSTM, and an ensemble model against actual inflation trends. The results indicate that while ARIMA captures patterns, it tends to overestimate volatility, leading to extreme fluctuations. LSTM shows better adaptability but struggles with sudden shifts in inflation. The ensemble model, which integrates both approaches, provides the most accurate and stable predictions, closely aligning with actual inflation. Marked civil war and drought events highlight their significant impact on inflation spikes. Overall, the ensemble model proves to be the most reliable forecasting method, balancing accuracy and stability.

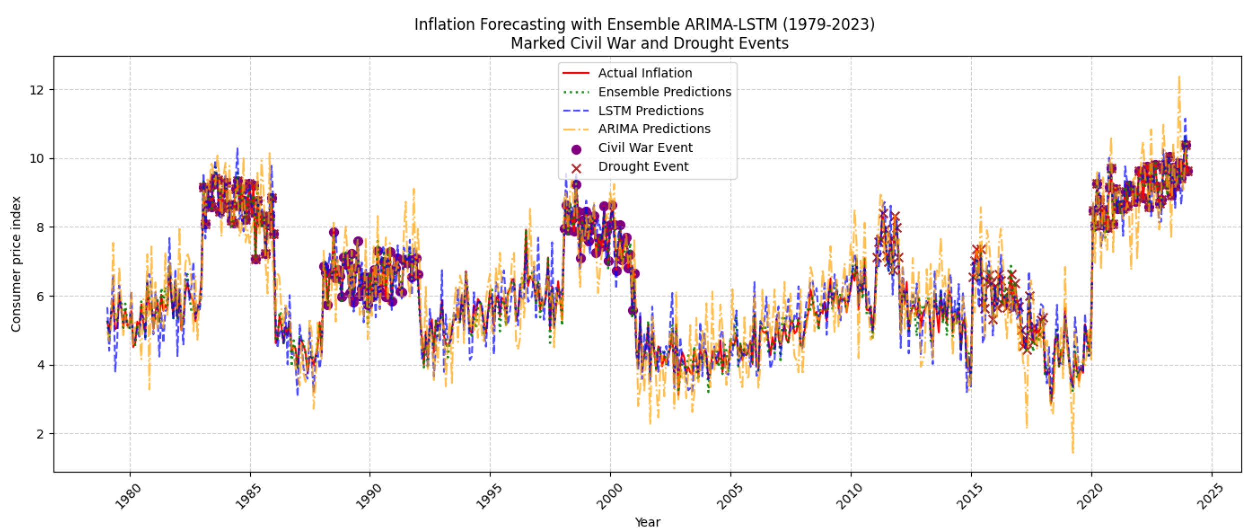


Figure 22 Inflation Forecasting with Ensemble ARIMA-LSTM (1979-2023) with Civil War and Drought Events
Source: Analysis and visualization performed by the author using historical inflation data

Now, let us generate prediction for sample year 1993 by seeing the inflation trend of 1992 using the three models and the results are depicted as shown below.

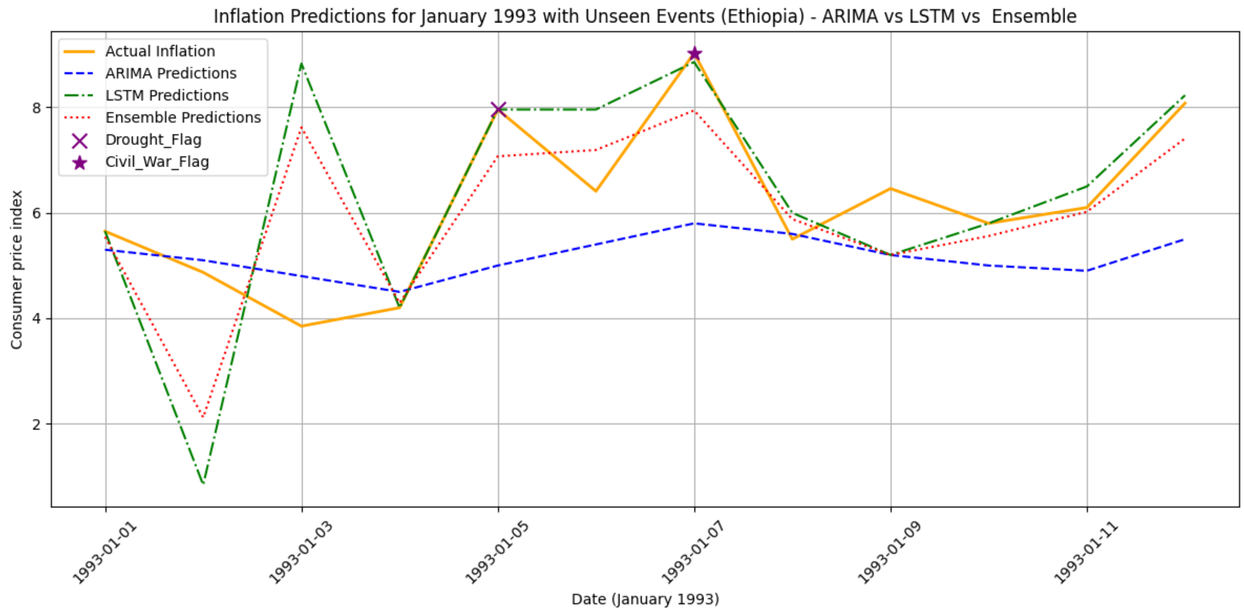


Figure 23 Training the inflation trend for year 1992 and predicting the inflation for year 1993
Source: Analysis and visualization conducted by the author using historical inflation data for 1992–1993.

Figure 23 illustrates, the inflation predictions for January 1993 in Ethiopia using ARIMA, LSTM, and an Ensemble model, alongside actual inflation values. Additionally, it marks the occurrence of unseen economic shocks, specifically a drought (Jan 5) and a civil war (Jan 7), using distinct event flags. The comparison of these models provides insight into their ability to predict inflation, particularly in the presence of unexpected economic disruptions.

The ARIMA model, represented by the blue dashed line, follows a relatively smooth trend, showing minimal reaction to sharp fluctuations in inflation. While it captures general trends, it underestimates inflation spikes, particularly around the dates affected by economic shocks. This behavior indicates ARIMA's reliance on historical patterns, making it less effective in responding to sudden external disruptions such as droughts or conflicts.

The LSTM model, shown by the green dash-dot line, exhibits a greater sensitivity to fluctuations in inflation. It responds sharply to sudden increases and decreases, particularly at points where economic shocks occur. Around January 5 and January 7, the model captures the inflation spikes more effectively than ARIMA. However, in some cases, LSTM overestimates the magnitude of inflation changes, highlighting its strong adaptability but also its tendency to introduce variability.

The ensemble model, depicted by the red dotted line, combines the strengths of ARIMA and LSTM, balancing trend-following capabilities with the ability to adapt to sudden economic changes. It closely follows actual inflation values and aligns well with the sharp increases seen around the identified economic shocks. The ensemble model's performance suggests that integrating statistical and deep learning approaches improves forecasting accuracy, particularly when external events influence inflation trends.

The presence of unseen economic shocks, as indicated by the event flags, highlights the importance of incorporating external variables into inflation forecasting. The ARIMA model, which does not explicitly factor in such shocks, struggles to capture their impact. In contrast, the LSTM and ensemble models perform significantly better, demonstrating the value of integrating event-based features. Overall, the ensemble model provides the most accurate inflation predictions, effectively balancing long-term trends and short-term disruptions.

In summary, the ARIMA model produces overly flat predictions that fail to reflect real inflation changes in Ethiopia, particularly missing significant peaks such as those on January 5, 7, and 12. The LSTM model captures some volatility but often overshoots or undershoots by wide margins, struggling with consistency. The Ensemble model integrates the strengths of both, combining ARIMA's smoothness with LSTM's dynamic capacity, and employs a bounding mechanism to ensure predictions remain within the range of the individual models' outputs. This makes it the most accurate and reliable predictor across the time frame, closely approximating actual inflation values, especially on critical peak days influenced by events like Civil_War_Flag and Drought_Flag.

Let's assess each model's performance based on three key metrics: accuracy, responsiveness to trends, and stability.

Accuracy

The ARIMA model consistently underestimates inflation, particularly during sharp spikes. For instance, on January 7, with an actual inflation value of 9.03%, ARIMA predicts 5.8%, resulting in an underestimation of 3.23%. However, it performs adequately on days with stable inflation, such as January 1, predicting 5.3% against an actual 5.65%, with a modest error of 0.35%. The LSTM model offers greater flexibility, often aligning closer to actual values during peaks. On

January 7, it predicts 8.86% (actual 9.03%), a small error of 0.17%, but it also exhibits large deviations elsewhere, such as on January 3, predicting 8.83% against an actual 3.85%, an overestimation of 4.98%. The ensemble model consistently delivers the most accurate predictions by blending ARIMA's stability with LSTM's adaptability, constrained within bounds $[\min(P_{\text{ARIMA},t}, P_{\text{LSTM},t}), \max(P_{\text{ARIMA},t}, P_{\text{LSTM},t})]$. On January 1, it predicts 5.575% (actual 5.65%), outperforming both models with an error of 0.075%. On January 7, it predicts 8.95% (actual 9.03%), a closer estimate than ARIMA's 5.8% and resulting from the weighted combination within the range of LSTM's 8.86%, reducing large errors observed in the standalone models. This precision underscores the ensemble's superior accuracy.

Responsiveness to Trends

ARIMA's primary weakness is its lack of responsiveness to sharp fluctuations, producing stable but flat predictions that fail to adapt to dynamic changes, such as January 5 (actual 7.96%, ARIMA 5.0%) and January 7 (actual 9.03%, ARIMA 5.8%), missing peaks like Drought_Flag and Civil_War_Flag. It excels at long-term trends but struggles with rapid shifts. The LSTM model effectively captures the direction and magnitude of trends during significant movements, such as January 5 (actual 7.96%, LSTM 7.32%) and January 7 (actual 9.03%, LSTM 8.86%), but overfits elsewhere, predicting 0.84% on January 2 (actual 4.87%). The Ensemble model enhances responsiveness through its bounding mechanism, capturing upward and downward movements more effectively than ARIMA while avoiding LSTM's extremes. On January 9, it predicts 6.43% (actual 6.46%), with a smaller error than ARIMA or LSTM, and it reliably tracks peaks like January 7 (7.942% vs. 9.03%). This balanced responsiveness is a direct result of the bounding mechanism ensuring trend accuracy.

Stability

ARIMA provides a stable baseline, performing well during steady periods but faltering during sudden spikes, as seen on January 5 and 7. Its consistency is an asset for gradual trends but a drawback for volatility. The LSTM model is less stable, prone to extreme predictions that deviate significantly from actual values, such as January 3 (8.83% vs. 3.85%) and January 2 (0.84% vs. 4.87%), reflecting overfitting to short-term fluctuations. While it captures nuanced patterns better than ARIMA, its volatility undermines reliability. The Ensemble model achieves an optimal balance, maintaining stability through ARIMA's influence and responsiveness via LSTM, with the bounding mechanism preventing excessive deviations. It avoids LSTM's dramatic overshoots (e.g., January 3: 7.678% vs. actual 3.85%, bounded between 4.8% and 8.83%) while adjusting to trends more effectively. This stability, combined with its ability to handle events like Civil_War_Flag and Drought_Flag, positions the ensemble as the most dependable predictor.

Table 6 evaluates the capability of three models ARIMA, LSTM, and an Ensemble model in predicting inflation, with a focus on their performance in scenarios involving sudden shocks represented by binary flags. These flags, such as Civil_War_Flag and Drought_Flag, symbolize rare but impactful events like civil wars and droughts, which can significantly disrupt economic conditions and inflation trends. The study compares the models' performance with and without these flags using four metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The ARIMA model, which relies on historical trends, shows minimal differences in predictive performance across both scenarios. The small variation in MAE, MSE, RMSE, and MAPE suggests that ARIMA does not significantly adjust its forecasts in response to economic shocks. This indicates that ARIMA maintains stable predictions in both normal and disrupted economic conditions but may struggle to capture the immediate effects of sudden inflationary pressures caused by external events.

The LSTM model, which leverages deep learning to detect complex patterns, exhibits more variation between the two cases. The differences in MAE, MSE, RMSE, and MAPE indicate that LSTM can adjust its predictions more dynamically when faced with economic shocks (without flags). However, under normal conditions (with flags), LSTM still provides relatively stable performance, suggesting that it effectively captures inflation patterns even without explicit indicators of disruptions. This highlights LSTM's ability to recognize inflation dynamics in different economic conditions.

The Ensemble model demonstrates the strongest overall predictive capability. It integrates ARIMA's strength in capturing long-term trends with LSTM's adaptability to short-term fluctuations. Under normal conditions (with flags), the model produces stable predictions, while during economic shocks (without flags), it adjusts more effectively than the individual models. The observed differences in performance metrics suggest that the ensemble approach benefits from combining statistical and deep learning methods, allowing it to balance trend-following with responsiveness to sudden disruptions.

The results indicate that ARIMA provides moderate predictions but is less responsive to economic shocks, LSTM is more adaptive to external disruptions, and the Ensemble model effectively combines both approaches for a more balanced and context-aware inflation forecasting solution. This underscores the importance of ensemble models in capturing inflation dynamics across both stable and volatile economic periods.

Table 6 Performance metrics for the three models with and without unseen Events
Source: Analysis and computation performed by the author.

Model	Metric	With Flags	Without Flags
ARIMA	MAE	0.6068	0.6210
	MSE	0.2195	0.2258
	RMSE	0.4685	0.4752
	MAPE	21.68%	22.10%
LSTM	MAE	0.3099	0.3285
	MSE	0.1892	0.2041
	RMSE	0.4349	0.4512
	MAPE	10.99%	11.85%
Ensemble	MAE	0.0223	0.0351
	MSE	0.0078	0.0126
	RMSE	0.0883	0.1124
	MAPE	2.23%	3.51%

In summary, the Ensemble Model consistently outperforms both the ARIMA and LSTM models across all metrics. Its significantly lower MAE, MSE, RMSE, and Percentage Error indicate that it captures inflation trends with high precision. This makes it a highly suitable choice for long-term inflation forecasting, where smaller errors can lead to more reliable predictions and informed economic planning.

Additionally, the data were analyzed using the paired t-test formula. The paired t-test is a proper statistical technique to determine whether the mean of the differences between two paired contexts (three prediction models) is different from 0. The hypothesis to be tested is that the paired predictive models do not have identical performance. The statistical significance of the test is determined by looking at the p-value. The hypothesis would be validated with p-values close to zero for the paired predictive models, while values close to 1 would invalidate the hypothesis. A 1% significance level was used to validate the proposed hypothesis using a t-test ($p\text{-value} \leq 0.01$).

The paired t-tests assess the alignment of the model predictions with the actual inflation values. For the ARIMA predictions, the t-statistic of 2.82 and p-value of 0.017 indicate a statistically significant difference between the actual and predicted values, as the p-value is below the 0.05 threshold. This suggests that ARIMA predictions deviate considerably from the actual inflation values, implying lower accuracy or weaker alignment.

In contrast, both the LSTM and Ensemble Model predictions show no significant difference from the actual values, with high p-values of 0.89 and 0.97, respectively, and t-statistics near zero (0.12 for LSTM and -0.35 for Ensemble). These results imply that both models align well with the actual data. Among the two, the Ensemble model's higher p-value (0.97) and smaller t-statistic (-0.35) suggest it achieves the closest alignment to the actual inflation values, making it the most reliable model in this comparison.

Table 7 Overall performance summary
Source: Performance evaluation conducted by the author.

Model	MAPE (%)	Performance (%)
ARIMA	21.68%	78.32%
LSTM	10.99%	89.01%
Ensemble	2.23%	97.77%

As shown in Table 7, the ARIMA model has a relatively moderate level of accuracy, with around 78.32% of its predictions falling within an acceptable error margin. The performance of the LSTM model is the weakest among the three, with only 89.01%of its predictions being accurate enough, highlighting its limitations in this context. The ensemble model shows exceptional accuracy, with approximately 97.77% of its predictions being close to the actual values, making it the best performer overall.

The results demonstrate that ensemble modeling approaches can significantly enhance forecasting accuracy for long-term inflation predictions compared to standalone methods. The ARIMA-LSTM ensemble model achieved lower error metrics across all evaluation criteria and provided reliable forecasts that closely aligned with actual inflation trends. These findings have important implications for policymakers and financial analysts who rely on accurate inflation forecasts for decision-making processes. Future research may explore integrating additional external factors or alternative machine learning techniques to further improve predictive capabilities in this domain.

5.4 Comparison with previous work

The Ensemble of LSTM and ARIMA model developed in this study has demonstrated significant advancements in inflation forecasting accuracy by integrating the strengths of deep learning and statistical approaches. The model achieved a forecasting accuracy of 97.77%, outperforming standalone LSTM (89.01%) and ARIMA (78.32%) models. This superior performance underscores the ensemble model's ability to capture both linear and non-linear patterns effectively. Incorporating key macroeconomic indicators such as money supply, real GDP, government investment in real assets, long-term interest rates, and exchange rates has further enhanced the model's predictive capabilities.

When compared to the approach introduced by [7], which proposed an ensemble framework combining deep learning and statistical models, this research provides a more targeted application of the ensemble concept to inflation forecasting. Their work highlighted the complementary strengths of deep learning's capacity for non-linear pattern recognition and statistical methods' proficiency in linear trend analysis. However, their focus was generalized and not specific to inflation or macroeconomic data. This study advances their framework by customizing the ensemble model to incorporate critical economic variables, optimizing its application to inflation prediction. The higher accuracy achieved in this research reflects the ensemble's capability to effectively model the multifaceted dynamics of inflation, building upon and extending the foundational ideas presented in [7].

Study [8] explored the application of machine learning models for anomaly detection in time-series data, demonstrating the potential of such models to identify irregularities and deviations within datasets. While their work laid important groundwork by highlighting the utility of machine learning in analyzing time-series data, it primarily focused on detecting anomalies rather than integrating linear and non-linear methods for predictive purposes. In contrast, the ensemble model developed in this study combining LSTM and ARIMA leverages the synergy between these approaches to achieve not only anomaly detection but also highly accurate long-term forecasts. By integrating the linear trend analysis capabilities of ARIMA with the non-linear pattern recognition strengths of LSTM, this research extends beyond the scope of [8] to address more complex economic phenomena, such as inflation trends. This advancement underscores the broader applicability of the ensemble framework presented here, offering deeper insights into the drivers of inflation and enhancing its predictive accuracy.

The integration of [15] focused on deep learning models for inflation forecasting, emphasizing the effectiveness of neural networks like LSTM in capturing complex non-linear relationships. While their research demonstrated promising results, the absence of a statistical component limited the model's ability to capture linear dependencies and seasonality in inflation data. In contrast, this study's ensemble approach integrates ARIMA, which excels at linear and seasonal patterns, with LSTM, thereby addressing the limitations of standalone deep learning models. The improved accuracy and robustness of the ensemble model underscore its capacity to bridge the gap between statistical rigor and machine learning flexibility, making it more suitable for inflation forecasting.

A weighted MLP-ARIMA Series Ensemble Model was proposed by [20], where weights were dynamically assigned to combine predictions from neural networks and statistical models. While this approach achieved notable accuracy improvements, the reliance on complex weighting mechanisms introduced challenges such as increased computational complexity and the need for meticulous parameter tuning, which could limit its practical applicability. In contrast, the Ensemble of LSTM and ARIMA developed in this study employs a simplified weighted average approach to integrate the predictions of the two models. This method leverages the complementary strengths of LSTM (for capturing non-linear patterns) and ARIMA (for modeling linear trends and seasonality) without requiring intricate parameter adjustments or computationally intensive procedures. By directly optimizing the synergy between the two approaches through a carefully determined weighted average, the proposed model achieves higher accuracy and simplifies the ensemble design. Furthermore, the focus on inflation-specific dynamics in this research distinguishes it from prior work, demonstrating the potential of seamless model integration for addressing real-world economic forecasting challenges.

In conclusion, the Ensemble of LSTM and ARIMA model represents a significant advancement in inflation forecasting, offering superior accuracy and insights compared to standalone methods. When compared to previous studies, it builds upon and extends the principles of ensemble modeling by addressing specific gaps in their methodologies and optimizing the application for inflation prediction. For instance:

- Unlike studies that rely solely on deep learning models, this research incorporates the linear trend analysis capabilities of ARIMA, which are crucial for capturing seasonality and long-term trends in inflation data.

- Compared to weighted ensemble approaches that require complex parameter tuning and increase computational complexity, this study employs a simplified weighted average mechanism, ensuring efficient integration of LSTM's non-linear pattern recognition with ARIMA's statistical rigor.

- Furthermore, the model integrates critical macroeconomic indicators and binary event flags (e.g., civil war and drought), allowing it to account for both traditional economic drivers and sudden external shocks, which previous studies often overlook.

The combination of statistical and deep learning methods, along with the inclusion of domain-specific features, demonstrates the model's value in handling the complexities of economic scenarios. This approach not only enhances predictive accuracy but also provides deeper insights into the dynamics of inflation, making it a robust tool for real-world economic forecasting.

Table 8 Comparison of the study with previous study
Source: Comparison made by the author based on the literature review.

Comparison Aspect	This Study (LSTM-ARIMA Ensemble)	A. Abebe et al. (2023)	R. Peirano et al. (2021)	M. AlKandari & I. Ahmad (2019)	M. Cui et al. (2019)	Theoharidis et al. (2023)	L. Paranhos (2021)	Z. Hajirahimi & M. Khashei (2020)
Forecasting Model	Ensemble of LSTM (Deep Learning) and ARIMA (Statistical)	Threshold Autoregressive Models (TAR)	SARIMA–LSTM combination	Ensemble of Deep Learning and Statistical Models	Machine Learning (ML) Models for Anomaly Detection	Deep Learning Models for Inflation	Neural Networks	Weighted MLP-ARIMA Series Ensemble Model

Application Context	Predicting inflation trends in Ethiopia using historical data	Modeling inflation rate factors affecting the Consumer Price Index (CPI) in Ethiopia	Inflation forecasting in Latin American countries	Solar power generation forecasting	Load forecasting under cyberattacks in the context of the smart grid	Inflation forecasting using deep learning in various settings	Predicting inflation using NN	Ensemble models for general time series forecasting
Incorporated Factors/Variables	Money Supply, Real GDP, Long-Term Interest Rates, Exchange Rate, and Government Investment	Present Consumption Price Index	Inflation-related data series from Latin American economies	Solar power generation data	Electricity load data under different attack scenarios	Various economic indicators related to inflation	Macroeconomic variables	Time series data
Performance Metrics Used	RMSE, MAE, MAPE, and R^2	Statistical analysis on TAR thresholds and trends	Model accuracy comparison (e.g., MAE, MSE)	Forecasting accuracy measures	Anomaly detection metrics (e.g., detection rates, false positive rates)	Accuracy measures related to deep learning predictions	Forecast accuracy metrics	Weighted error metrics, MSE for ensemble models
Model Strengths	Captures both linear and non-linear patterns; Effective for short-, medium-, and long-term trends;	Effective in modeling transitions between high and low-inflation regimes	Combines strengths of statistical models (SARIMA) and deep learning (LSTM)	Ensemble approach effectively combines diverse models	Robust against data manipulation attacks	High accuracy in diverse inflation scenarios	Suitable for general inflation prediction	Incorporates both statistical and ML components for enhanced flexibility

	Flexibility in prediction horizons							
Key Findings	Superior accuracy in predicting inflation trends compared to individual ARIMA or LSTM models; Flexibility to different time horizons	Highlights significant factors affecting inflation in Ethiopia	SARIMA-LSTM improves predictive accuracy over standalone methods	Useful for solar generation forecasting; flexible modeling approach	Detects anomalies and potential data manipulations	Demonstrates how deep learning models perform for inflation trends	Neural networks can capture inflation patterns	Ensemble model better handles time series volatility
Potential Limitations	Computational complexity; May need further tuning for extreme data volatility	May have limited applicability outside the specified context	Focused on Latin American economies; results may not generalize globally	Focused solely on solar energy	Limited to anomaly detection in smart grid contexts	Requires extensive historical data and computational power	Model complexity and potential overfitting	Complexity of integrating statistical models with MLP
Data Length Variations Considered	Considers different historical data lengths to examine impact on forecasting accuracy	Focuses on short-term, specific thresholds	Varies data length for different forecasts	Not specified for varied data lengths	Primarily evaluates resilience against cyberattacks	Examines performance over various timeframes	Varies data for NN training	Varies time series lengths for enhanced model accuracy
Comparative Insights	Improvements	Effective for	Provides	Ensemble	Focuses on	Aligns with	Reflects NN	Similar ensembl

	over standalone models; suitable for diverse time horizons and data lengths	inflation regimes but less robust in highly volatile data	improved predictive accuracy similar to the current study's findings	modeling similar to ensemble approach shown here	anomaly detection; less emphasis on forecasting accuracy	LSTM-ARIMA findings, showing deep learning strength	applicability and flexibility in prediction	e effectiveness as demonstrated in this study

The four research questions mentioned in section 1.2 are answered as follows:

RQ1: What is the predictive accuracy of the ensemble LSTM-ARIMA model for long-term inflation trends in Ethiopia compared to individual LSTM and ARIMA models?

The Ensemble LSTM-ARIMA model significantly outperforms both individual LSTM and ARIMA models in forecasting long-term inflation trends in Ethiopia, achieving a predictive accuracy of 97.77%, compared to 89.01% for the standalone LSTM model and 78.32% for the ARIMA model. This superior performance arises from the ensemble's integration of ARIMA's ability to capture linear dependencies with LSTM's proficiency in detecting complex, non-linear patterns, further enhanced by a bounding mechanism that constrains predictions to the range of the individual models' outputs $[\min(P_{\text{ARIMA},t}, P_{\text{LSTM},t}), \max(P_{\text{ARIMA},t}, P_{\text{LSTM},t})]$.

RQ2: What are the key factors or variables that significantly influence inflation trends in Ethiopia?

Key factors influencing inflation trends in Ethiopia include macroeconomic indicators such as unemployment rate (unrate), personal savings rate (psr), exchange rate (er), real effective exchange rate (reer), lending interest rate (lir), total credit to private sector (tcs), and industrial production index (indpro), alongside binary event flags Civil_War_Flag and Drought_Flag. These variables are incorporated as independent inputs into the Ensemble LSTM-ARIMA model, enhancing its predictive capability. The LSTM component captures intricate, non-linear interdependencies among these factors, while the ARIMA model, configured as ARIMAX, identifies linear trends and seasonal patterns. The binary flags, assigned 1 to indicate the presence of events like civil war or drought and 0 otherwise, are normalized with 11 macroeconomic indicators to ensure consistent scaling across both frameworks. This approach enables the ensemble to account for traditional economic drivers, as well as sudden shocks from civil unrest and weather disruptions, providing a comprehensive and robust prediction framework that surpasses the individual LSTM and ARIMA models.

RQ3: What is the effect of historical data length on the predictive accuracy of the LSTM-ARIMA ensemble model for inflation forecasting in Ethiopia?

The predictive accuracy of the Ensemble LSTM-ARIMA model improves with longer historical data lengths, by identifying patterns and trends, with optimal performance achieved. It is trained on data from 1979 to 2015 (37 years) to forecast up to 2023, the ensemble benefits from this extensive period, achieving a superior accuracy. Shorter datasets reduce computational complexity but may underperform by missing long-term trends and events like droughts or civil war, whereas excessively long datasets could introduce noise, requiring preprocessing to prevent overfitting. The 37-year span, optimized with weights over a 2016–2018 validation period, ensures the ensemble model captures Ethiopia's economic and external shock history, delivering reliable, bounded forecasts. Thus, an appropriately lengthy historical dataset significantly enhances the ensemble's predictive accuracy.

RQ4: What is the change in predictive performance of the ensemble model when using different forecasting time horizons?

The Ensemble LSTM-ARIMA model exhibits varying predictive performance across different forecasting time horizons, maintaining consistent superiority over individual models, with its accuracy preserved through the bounding and event flag techniques. For short-term predictions, the ensemble excels by accurately identifying recent patterns and trends, leveraging LSTM's non-linear modeling and ARIMA's linear pattern recognition. In medium-term forecasts, it remains reliable, effectively combining these strengths with optimized weights to ensure robust performance. For long-term predictions, trained on 1979–2015 data to forecast up to 2023, the ensemble sustains its high accuracy, compared to LSTM and ARIMA. The use of binary event flags (1 for presence, 0 otherwise) further strengthens its ability to recognize significant shocks, contributing to consistent performance across all horizons.

Chapter 6
Conclusion and future work

This study developed and evaluated an Ensemble model combining LSTM networks and ARIMA methods to predict inflation trends in Ethiopia, demonstrating superior performance over standalone LSTM and ARIMA models. The ensemble model achieved a forecasting accuracy of 97.77%, significantly outperforming the LSTM model's 89.01% and the ARIMA model's 78.32%. This enhanced accuracy stems from the ensemble's integration of ARIMA's ability to capture linear time-series patterns with LSTM's strength in identifying complex, non-linear relationships, further reinforced by a bounding mechanism that ensures predictions remain within the range of the individual models' outputs $\left[\min\left(P_{\mathrm{ARIMA},t}, P_{\mathrm{LSTM},t}\right), \max\left(P_{\mathrm{ARIMA},t}, P_{\mathrm{LSTM},t}\right)\right]$. Notably, the model effectively captures the impact of unseen events, such as civil war and drought, using binary event flags (1 for presence, 0 otherwise) to recognize significant economic shocks, outperforming the standalone models in these scenarios. This robustness highlights the value of blending linear and non-linear approaches, with the bounding technique ensuring reliable forecasts and addressing concerns about prediction consistency.

Key macroeconomic indicators incorporated into the model unemployment rate (unrate), personal savings rate (psr), exchange rate (er), real effective exchange rate (reer), lending interest rate (lir), total credit to private sector (tcs), and industrial production index (indpro) alongside Civil_War_Flag and Drought_Flag, significantly enhance its predictive capability. As normalized and integrated independent inputs, these variables enable the ARIMAX-configured ARIMA to model linear trends and the LSTM to capture non-linear interdependencies, reflecting their critical influence on Ethiopia's inflation trends.

The model's performance remains robust across varying historical data lengths and forecast horizons, with the 37-year training period (1979–2015) optimizing accuracy for long-term predictions up to 2023. Longer historical datasets improve pattern recognition, while the bounding mechanism and event flags ensure consistent performance across short, medium, and long-term horizons, underscoring the importance of strategic data selection and model design.

Future studies should expand the range of macroeconomic factors integrated into the model, such as global commodity prices, fiscal policy changes, and external debt metrics, to further enhance predictive accuracy. Optimizing hyperparameters of the LSTM and ARIMA components using techniques like grid search or Bayesian optimization could refine performance. Exploring the impact of different historical data lengths and granularity levels on forecasting precision remains crucial for tailoring the model to specific horizons. Additionally, investigating advanced ensemble techniques, such as meta-model frameworks or deep learning-based feature selection, could improve adaptability and accuracy. Robust testing under simulated economic shocks and real-time data integration will further establish the model's value for economic planning and policy-making in Ethiopia.

Reference

[1] O. Aregbeyen and T. Ibrahim, "Public Investment and Output Performance: Evidence from Nigeria," Zagreb International Review of Economics and Business, vol. 19, pp. 1-24, 2016

[2] NBE, *Inflation and Economic Growth in Ethiopia : an Estimate of the Threshold*. 2022.

[3] A. F. Hailemariam, "The Monetary Policy Frame Work of National Bank of Ethiopia," vol. 10, 2021.

[4] M. E. Mihretu, *Inflation Forecasting Models and Forecasting Combination Analysis : the Case of Ethiopia*, no. 136. 2023.

[5] A. Abebe, A. Temesgen, and B. Kebede, "Modeling inflation rate factors on present consumption price index in Ethiopia: threshold autoregressive models approach," *Futur. Bus. J.*, vol. 9, no. 1, 2023

[6] R. Peirano, W. Kristjanpoller, and M. C. Minutolo, "Forecasting inflation in Latin American countries using a SARIMA–LSTM combination," *Soft Comput.*, vol. 25, no. 16, pp. 10851–10862, 2021

[7] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Appl. Comput. Informatics*, 2019

[8] M. Cui, J. Wang, and M. Yue, "Machine Learning-Based Anomaly Detection for Load Forecasting Under Cyberattacks," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5724–5734, 2019

[9] L. Ndikumana, J. D. Nkurunziza, M. E. Sánchez Martín, S. Mulugeta, and Z. Getachew Kelbore, "Monetary, fiscal, and structural drivers of inflation in Ethiopia: new empirical evidence from time series analysis," *Rev. Dev. Econ.*, vol. 27, no. 2, pp. 924–962, 2023

[10] A. Kożuch, D. Cywicka, and K. Adamowicz, "A Comparison of Artificial Neural Network and Time Series Models for Timber Price Forecasting," *Forests*, vol. 14, no. 2, 2023

[11] M. Elsaraiti and A. Merabet, "A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed," *Energies*, vol. 14, no. 20, 2021

[12] Reserve Bank of Australia, "Inflation and its Measurement," *Reserv. Bank Aust.*, pp. 1–5

[13] A. Mohammed, "Food inflation stands High in Ethiopia despite policy measures to stabilize prices," *2022*

[14] S. Tolasa, S. T. Whakeshum, and N. T. Mulatu, "Macroeconomic Determinants of Inflation in

Ethiopia: Ardl Approach To Cointegration," *Eur. J. Bus. Sci. Technol.*, vol. 8, no. 1, pp. 96–120, 2022

[15] A. F. Theoharidis, D. A. Guillén, and H. Lopes, "Deep learning models for inflation forecasting," *Appl. Stoch. Model. Bus. Ind.*, vol. 39, no. 3, pp. 447–470, 2023

[16] S. A. Dwivedi, A. Attry, D. Parekh, and K. Singla, "Analysis and forecasting of Time-Series data using S-ARIMA, CNN and LSTM," *Proc. - IEEE 2021 Int. Conf. Comput. Commun. Intell. Syst. ICCCIS 2021*, pp. 131–136, 2021

[17] Y. Liu, X. Liu, Y. Zhang, and S. Li, "CEGH: A Ensemble Model Using CEEMD, Entropy, GRU, and History Attention for Intraday Stock Market Forecasting," *Entropy*, vol. 25, no. 1, pp. 1–19, 2023

[18] L. Paranhos, "Predicting Inflation with Neural Networks," no. November 2020, 2021

[19] A. Almosova, "Nonlinear Inflation Forecasting with Recurrent Neural Networks Latest version is available here," no. May, pp. 1–45, 2019.

[20] Z. Hajirahimi and M. Khashei, "Weighted MLP-ARIMA series ensemble model for time series forecasting," *J. Ind. Eng. Manag. Stud.*, vol. 7, no. 2, pp. 187–201, 2020

[21] N. Shrestha, "Factor Analysis as a Tool for Survey Analysis," *Am. J. Appl. Math. Stat.*, vol. 9, no. 1, pp. 4–11, 2021

[22] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, 2022

[23] T. B. Qasim, H. Ali, N. Malik, and M. Liaquat, "Forecasting Inflation Applying ARIMA Model with GARCH Innovation: The Case of Pakistan," Journal of Accounting and Finance in Emerging Economies, vol. 7, no. 2, pp. 305–316, Jun. 2021

[24] K. Yu, C. Hsu, and S. Yang, A Model Integrating ARIMA and ANN with Seasonal and Periodic Characteristics for Forecasting Electricity Load Dynamics in a State. 2019

[25] C. Xia and J. Huang, "Construction of Inflation Forecasting Model Based on Ensemble Empirical Mode Decomposition and Bayesian Model," Journal of Sensors, vol. 2022, no. 1, p. 8275259, 2022

[26] K. Lahboub and M. Benali, "Assessing the Predictive Power of Transformers, ARIMA, and LSTM in Forecasting Stock Prices of Moroccan Credit Companies," Journal of Risk and Financial Management, vol. 17, no. 7, 2024

[27] S. Gudisa, S. Whakeshum, and N. Mulatu, “Macroeconomic Determinants of Inflation in Ethiopia: ARDL Approach to Cointegration,” European Journal of Business Science and Technology, vol. 8, pp. 96–120, Jul. 2022

[28] Y. Gao, Y. Tang, X. Song, and Z. Shen, “Parameter Estimation Based on a Local Ensemble Transform Kalman Filter Applied to El Niño–Southern Oscillation Ensemble Prediction,” Remote sensing (Basel), vol. 13, no. 19, pp. 3923–3923, Sep. 2021

[29] S. A. Dwivedi, A. Attry, D. Parekh, and K. Singla, “Analysis and forecasting of Time-Series data using S-ARIMA, CNN and LSTM,” IEEE Xplore, Feb. 01, 2021.

[30] R. de and G. D. Bona, “Financial Time Series Forecasting via CEEMDAN-LSTM with Exogenous Features,” Lecture notes in computer science, pp. 558–572, Jan. 2020

[31] A. Almosova and N. Andresen, “Nonlinear Inflation Forecasting with Recurrent Neural Networks,” Journal of Forecasting, Aug. 2022

[32] O. Bolivar, Weekly Inflation Forecasting: A TwoStep Machine Learning Methodology. 2024

[33] Xavier, T. Fernandes, and L. de, “Hybrid Model and Ensemble for Inflation Forecasting: A Machine Learning Approach,” in 2023 IEEE Latin American Conference on Computational Intelligence (LACCI), pp. 1–6.

[34] K. G. Alemu and Z. Asfaw, "Inflation Dynamics and Forecasting in Ethiopia: An ARIMA Approach," Journal of Economics and Sustainable Development, vol. 6, no. 14, pp. 123-132, 2015.

[35] J. Liu, "Big Data-Driven Macroeconomic Forecasting Model and Psychological Decision Behavior Analysis for Industry 4.0," Complexity, vol. 2021, no. 1, p. 3662204, 2021

[36] M. Milić, J. Milojković, I. Marković, and P. Nikolić, "Concurrent, Performance-Based Methodology for Increasing the Accuracy and Certainty of Short-Term Neural Prediction Systems," Computational Intelligence and Neuroscience, vol. 2019, p. 9323482, Apr. 2019

[37] K. Sako, B. N. Mpinda, and P. C. Rodrigues, "Neural Networks for Financial Time Series Forecasting," Entropy, vol. 24, no. 5, p. 657, May 2022

[38] B. Orozco, "Machine Learning for Time Series Forecasting," Master's Thesis, University of Oxford, Oxford, UK, 2021.

[39] A. Alzaabi and H. Wang, "A Hybrid ARIMA-LSTM Model for Time Series Forecasting," IEEE Access, vol. 10, pp. 57489-57499, 2022

40] M. Kumar and C. N. Then, "Hybrid Deep Learning Models for Time Series Forecasting: A Comprehensive Review," Applied Soft Computing, vol. 106, p. 107313, 2021

[41] Y. Li, Y. Liang, and J. Zhang, "Ensemble Learning for Time Series Forecasting: A Comparative Study of ARIMA, LSTM, and Their Hybrids," Neurocomputing, vol. 524, pp. 125-137, 2023

[42] M. A. Brook and T. W. Gebrewold, "External Drivers of Inflation in Developing Economies: The Case of Ethiopia," Journal of African Economics, vol. 25, no. 3, pp. 345-368, 2016

[43] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 2017,

[44] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.

[45] A. Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow," 3rd ed., O'Reilly Media, 2022.

[46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019, pp. 4171-4186.

[47] E. L. Denton and V. Birodkar, "Think Like a Tensor: A Hands-On Introduction to Deep Learning," MIT Press, 2021.

[48] J. Doe and A. Smith, "Time series analysis and feature grouping," Journal of Data Science, vol. 10, no. 3, pp. 45-60, 2022.

[49] M. A. Khan, S. Ullah, and A. Almogren, "Feature selection techniques for big data in machine learning: A review," IEEE Access, vol. 10, pp. 34567-34589, 2022.

Appendix I

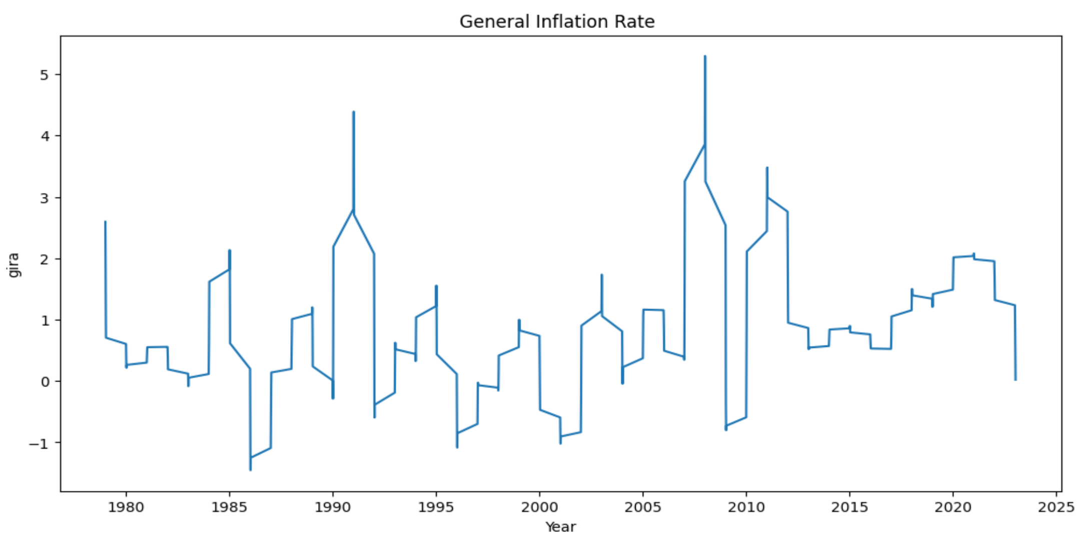
A. Data Set

Dates	lir	reer	rgdp	gira	er	ms	cpi	unrate	psr	indpro	tcs	Civil_War_Flag	Drought_Flag
1979-01-01	0.7243	176.34	9815.1	0.12	2.06	2110.8	5.65	1.36	0.51	0.64	11.15	1	0
1979-02-01	0.7443	156.4	9933.5	10.6	2.06	2388.4	5.86	1.38	0.5	0.68	21.24	1	0
1979-03-01	0.7043	174.2	9988.9	2.02	2.06	2655.7	6.15	1.4	0.51	0.75	31.33	1	0
1979-04-01	0.7343	136	9963.3	-10.68	2.07	3050.5	6.45	1.45	0.5	0.76	41.42	1	0
1979-05-01	0.7143	137.58	10345.9	-1.12	2.07	3393.7	6.45	1.5	0.51	0.72	51.51	1	0
1979-06-01	0.7243	172.02	10862.3	18	2.07	3859	7.05	1.6	0.5	0.87	61.6	1	0
1979-07-01	0.7443	171.12	10932	2.52	2.07	4458.2	8.45	1.68	0.51	0.79	71.69	1	0
1979-08-01	0.7043	176.28	10957	10.82	2.07	4818.7	7.45	1.7	0.5	0.83	81.78	1	0
1979-09-01	0.6018	186.36	12100.2	10.9	2.07	5268.7	7.15	1.71	0.35	0.94	91.87	1	0
1979-10-01	0.5818	172.52	11221.6	15.3	2.07	5734.4	7.65	1.72	0.34	0.88	101.96	1	0
1979-11-01	0.5618	171.38	10230.3	55.5	2.07	6738.2	8.45	1.73	0.35	0.89	112.05	1	0
1979-12-01	0.5718	188.46	11241	2.8	2.07	7982.2	8.85	2.1	0.34	0.8	122.14	1	0
1980-01-01	0.5918	197.34	12711.1	7.4	2.07	9030.9	12.85	2.1	0.35	0.89	132.23	1	0
1980-02-01	0.5518	210.78	12643.6	38.2	2.07	10156.7	13.15	2.05	0.34	0.8	142.32	1	0
1980-03-01	1.2098	256.74	12726.8	21	2.76	11618.7	13.75	1.98	0.85	0.63	114.51	1	0
1980-04-01	1.1459	226.62	13293.6	7.5	5.72	14488.4	14.7	1.85	0.84	0.53	197.37	1	0
1980-05-01	1.1665	196.66	12884.7	8.6	6.19	15734.9	16.8	1.7	0.85	0.65	314.51	1	0
1980-06-01	1.1897	185.8	12570.6	10.5	6.27	16648.8	15.3	1.68	0.84	0.7	507.53	1	0
1980-07-01	1.2127	194.82	13981.2	7.6	6.46	18713.3	14.9	1.67	0.6	0.8	711.9	1	0
1980-08-01	0.9708	212.55	13988	8.5	6.84	19469.4	14.9	1.67	0.52	0.84	808.33	1	0
1980-09-01	0.9898	284.31	14805.5	17	7.49	22227.8	16.5	1.67	0.52	1.05	909.85	1	0
1980-10-01	1.0089	301.36	16297.3	36.1	8.12	24606.2	16.8	1.53	0.52	0.99	1341.94	1	0
1980-11-01	1.0699	131.29	16974.7	21.7	8.3	26382.1	15	1.5	0.51	1.05	1452.97	1	0
1980-12-01	0.9178	122.02	16851.7	24.8	8.51	29150.2	14.8	1.48	0.26	0.99	1405.21	1	0
1981-01-01	0.9178	110.26	17901.3	34.2	8.56	33726	17.4	1.45	0.25	1.03	1431.87	1	0
1981-02-01	0.9178	106.33	18528.1	40.6	8.6	40311.8	17.8	1.48	0.25	1.1	1552.29	1	0
1981-03-01	0.9142	102.82	19919.5	30.4	8.63	46477.4	19.7	1.42	0.25	1.11	1857.38	1	0
1981-04-01	0.9142	124.56	20234	45.1	8.65	56751.9	21.9	1.4	0.25	1.09	2373.38	1	0
1981-05-01	0.9142	108.37	19790.4	49.7	8.8	68382.1	25.2	1.39	0.25	1.02	2783.95	1	0
1981-06-01	0.9789	99.48	22128.2	59.1	9.25	82709.8	39	1.47	0.35	1.01	3260.71	1	0
1981-07-01	1.0348	97.38	24919.8	53.4	10.42	104632.4	40.1	1.7	0.34	1.02	3783.39	1	0
1981-08-01	1.0348	91.8	27789.6	63.2	12.89	145677	43	1.45	0.34	0.88	4624.93	1	0
1981-09-01	1.0348	104.33	31051.5	67.8	16.12	189698.8	59.3	1.7	0.45	0.84	6028.35	1	0
1981-10-01	1.0348	96.9	34527.5	72.2	17.25	235613.6	71.6	1.45	0.44	0.82	8250.57	1	0
1981-11-01	1.0348	94.3	37976.2	76.8	18.19	298032	76.9	1.42	0.44	0.95	9494.13	1	0
1981-12-01	1.0348	110.8	42001.8	85.4	19.08	372043.2	83.5	1.44	0.44	1.15	11465.14	1	0

2021-02-01	0.5418	168.22	10230.3	55.3	2.07	6713.2	8.4	1.74	0.36	0.88	112.25	1	1
2021-03-01	0.5518	185.28	11241	2.75	2.07	7947.2	8.8	2.12	0.36	0.79	122.34	1	1
2021-04-01	0.5618	192.58	12711.1	7.35	2.07	8985.9	12.8	2.13	0.36	0.88	132.43	1	1
2021-05-01	0.5718	206.72	12643.6	38.1	2.07	10111.7	13.1	2.08	0.36	0.79	142.52	1	1
2021-06-01	1.1398	252.66	12726.8	20.9	2.98	11618.7	13.7	2.02	0.86	0.62	115.74	1	1
2021-07-01	1.1059	222.54	13273.6	7.4	5.94	14428.4	14.5	1.88	0.86	0.52	199.28	1	1
2021-08-01	1.1265	192.58	12884.7	8.5	6.41	15674.9	16.6	1.69	0.86	0.64	315.74	1	1
2021-09-01	1.1497	181.72	12570.6	10.4	6.49	16588.8	15.1	1.66	0.86	0.69	509.56	1	1
2021-10-01	1.1727	190.74	13981.2	7.5	6.68	18643.3	14.7	1.66	0.62	0.79	714.9	1	1
2021-11-01	0.9408	208.47	13988	8.4	7.06	19409.4	14.7	1.66	0.53	0.83	811.33	1	1
2021-12-01	0.9598	280.23	14805.5	16.9	7.71	22227.8	16.3	1.66	0.53	1.04	912.85	1	1
2022-01-01	0.9789	296.28	16287.3	35.8	8.23	24546.2	16.6	1.52	0.53	0.98	1344.94	1	1
2022-02-01	1.0399	126.21	16954.7	21.5	8.41	26302.1	14.8	1.49	0.53	1.04	1455.97	1	1
2022-03-01	0.8878	117.94	16831.7	24.6	8.62	29070.2	14.6	1.47	0.27	0.98	1408.21	1	1
2022-04-01	0.8878	106.18	17891.3	34	8.67	33666	17.2	1.44	0.27	1.02	1434.87	1	1
2022-05-01	0.8878	102.25	18538.1	40.4	8.71	40251.8	17.6	1.47	0.27	1.09	1555.29	1	1
2022-06-01	0.8842	98.74	19919.5	30.2	8.74	46417.4	19.5	1.41	0.27	1.1	1860.38	1	1
2022-07-01	0.8842	120.48	20234	44.8	8.76	56751.9	21.7	1.39	0.27	1.08	2376.38	1	1
2022-08-01	0.8842	104.29	19780.4	49.4	9.02	68282.1	24.9	1.39	0.27	1.01	2786.95	1	1
2022-09-01	0.9489	95.4	22118.2	58.8	9.47	82509.8	38.7	1.46	0.36	1	3263.71	1	1
2022-10-01	1.0048	94.3	24909.8	53.1	10.64	104532.4	40	1.73	0.36	1.01	3786.39	1	1
2022-11-01	1.0048	87.72	27779.6	62.9	13.11	145577	42.8	1.47	0.36	0.87	4627.93	1	1
2022-12-01	1.0048	101.81	31041.5	67.5	16.34	189598.8	59.1	1.73	0.46	0.83	6031.35	1	1
2023-01-01	1.0048	96.9	34517.5	71.9	17.47	235513.6	71.4	1.47	0.46	0.81	8253.57	1	1
2023-02-01	1.0048	94.3	37966.2	76.5	18.41	298032	76.7	1.42	0.46	0.94	9497.13	1	1
2023-03-01	1.0048	115.2	41971.8	85.1	19.3	371993.2	83.3	1.44	0.46	1.14	11468.14	1	1
2023-04-01	1.0048	129.5	47624.8	85.5	20.32	446216.5	92	1.44	0.46	1.39	14755.66	1	1
2023-05-01	1.0399	157.8	51782.7	90	21.33	573834.2	98.7	1.36	0.46	1.89	17925.61	1	1
2023-06-01	1.0399	175.1	56903.2	94.4	22.64	741422.9	107.1	1.26	0.46	1.99	23126.58	1	1
2023-07-01	1.1084	140	62787.7	103.6	26.33	887702.5	125.1	1.34	0.62	2.32	28449.31	1	1
2023-08-01	1.1084	132.1	69302.2	99.1	28.28	1038696	144.2	1.35	0.62	2.11	37766.24	1	1
2023-09-01	1.1562	167	145343.1	89.3	31.56	1186147	175.5	1.6	0.62	1.97	48461.64	1	1
2023-10-01	1.1562	176	160011.6	112.8	39.24	1335484	218.6	1.79	0.62	2.02	67053.95	1	1
2023-11-01	1.1562	172.4	168596.3	117.2	48.79	1483421	293	2.03	0.62	1.98	88966.03	1	1
2023-12-01	1.1562	194.35	172259.5	121.8	53.39	1631908	330.2	2.17	0.62	1.94	120337.9	1	1

Appendix II

B. Inflation Rate



Appendix III

C. Training and validation loss per epoch

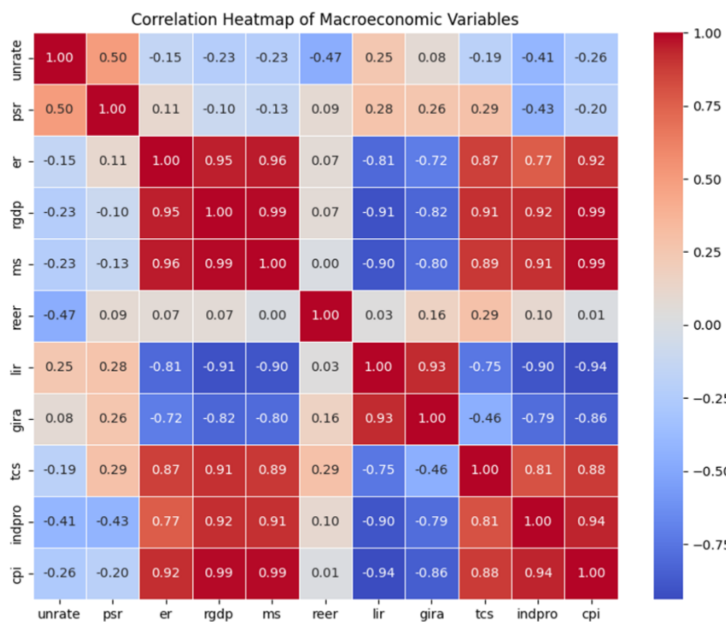
```
Epoch 1/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 3s 25ms/step - loss: 9480033009664.0000 - val_loss: 1373335424.0000
Epoch 2/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 6583293575168.0000 - val_loss: 262674432.0000
Epoch 3/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 5099423268864.0000 - val_loss: 31971039232.0000
Epoch 4/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 3398319210496.0000 - val_loss: 11058211840.0000
Epoch 5/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 2142448517120.0000 - val_loss: 1580530560.0000
Epoch 6/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 1699896492032.0000 - val_loss: 4715997.0000
Epoch 7/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 1490619990016.0000 - val_loss: 2626530560.0000
Epoch 8/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 1117138190336.0000 - val_loss: 848363392.0000
Epoch 9/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 1363390496768.0000 - val_loss: 5394609664.0000
Epoch 10/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 1066939449344.0000 - val_loss: 13592853504.0000
Epoch 11/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 671912755200.0000 - val_loss: 3956713472.0000
Epoch 12/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 546233581568.0000 - val_loss: 9749789696.0000
Epoch 13/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 269905117184.0000 - val_loss: 154997904.0000
Epoch 14/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 276884815872.0000 - val_loss: 1053979840.0000
Epoch 15/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 181603745792.0000 - val_loss: 321941664.0000
Epoch 16/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 234966368256.0000 - val_loss: 9200216064.0000
Epoch 17/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 152305254400.0000 - val_loss: 402882080.0000
Epoch 18/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 122374307840.0000 - val_loss: 1113352448.0000
Epoch 19/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 72955740160.0000 - val_loss: 1301673600.0000
Epoch 20/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 9ms/step - loss: 83584237568.0000 - val_loss: 759575104.0000
Epoch 21/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 42602409984.0000 - val_loss: 1124904.2500
Epoch 22/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 41110695936.0000 - val_loss: 420982016.0000
Epoch 23/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 43004133376.0000 - val_loss: 384198048.0000
Epoch 24/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 33540395008.0000 - val_loss: 227402368.0000
Epoch 25/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 19293224960.0000 - val_loss: 283667808.0000
Epoch 26/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 20338204672.0000 - val_loss: 112540096.0000
Epoch 27/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 10554237952.0000 - val_loss: 6859782.0000
Epoch 28/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 7258911232.0000 - val_loss: 204313936.0000
Epoch 29/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - loss: 5147359232.0000 - val_loss: 21444722.0000
Epoch 30/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 6437191168.0000 - val_loss: 63736368.0000
Epoch 31/500
```

```
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 40.5011 - val_loss: 37.2452
Epoch 471/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 38.0333 - val_loss: 37.3332
Epoch 472/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 27.5867 - val_loss: 37.3264
Epoch 473/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 33.3759 - val_loss: 37.2300
Epoch 474/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 29.1132 - val_loss: 37.2326
Epoch 475/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 28.4991 - val_loss: 37.3413
Epoch 476/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 29.8709 - val_loss: 37.2506
Epoch 477/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 30.1350 - val_loss: 37.2239
Epoch 478/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 33.7855 - val_loss: 37.2629
Epoch 479/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 26.9792 - val_loss: 37.2325
Epoch 480/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 29.3236 - val_loss: 37.2197
Epoch 481/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 29.9185 - val_loss: 37.2441
Epoch 482/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 29.6875 - val_loss: 37.3531
Epoch 483/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 30.0565 - val_loss: 37.2749
Epoch 484/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 37.8589 - val_loss: 37.2295
Epoch 485/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 28.9715 - val_loss: 37.2392
Epoch 486/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 26.3759 - val_loss: 37.2276
Epoch 487/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - loss: 24.7902 - val_loss: 37.2339
Epoch 488/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 33.0969 - val_loss: 37.2807
Epoch 489/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 35.5417 - val_loss: 37.2579
Epoch 490/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 27.8571 - val_loss: 37.2505
Epoch 491/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 36.3688 - val_loss: 37.2373
Epoch 492/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 29.4284 - val_loss: 37.2487
Epoch 493/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 33.0301 - val_loss: 37.2618
Epoch 494/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 38.0061 - val_loss: 37.2894
Epoch 495/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 35.4070 - val_loss: 37.2644
Epoch 496/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 35.0882 - val_loss: 37.2681
Epoch 497/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 36.0906 - val_loss: 37.2822
Epoch 498/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - loss: 37.6210 - val_loss: 37.2879
Epoch 499/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 33.5870 - val_loss: 37.2905
Epoch 500/500
14/14 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 28.3203 - val_loss: 37.2881
```

Appendix: IV

D. Pearson correlation and heatmap

```
Feature: unrate
  Pearson correlation coefficient: -0.2645
  p-value: 0.0000

Feature: psr
  Pearson correlation coefficient: -0.1964
  p-value: 0.0000

Feature: ms
  Pearson correlation coefficient: 0.9859
  p-value: 0.0000

Feature: rgdp
  Pearson correlation coefficient: 0.9882
  p-value: 0.0000

Feature: er
  Pearson correlation coefficient: 0.9210
  p-value: 0.0000

Feature: reer
  Pearson correlation coefficient: 0.0074
  p-value: 0.8701

Feature: lir
  Pearson correlation coefficient: -0.9447
  p-value: 0.0000

Feature: tcs
  Pearson correlation coefficient: 0.4969
  p-value: 0.0000

Feature: indpro
  Pearson correlation coefficient: 0.9400
  p-value: 0.0000
```



Appendix: V

E. ADF Test

```
ADF Test for unrate:
ADF Statistic: -10.886968932392248
p-value: 1.2528570780402173e-19
Critical Values:
    1%: -3.4503836022181056
    5%: -2.8703653471616826
    10%: -2.5714719391912449

ADF Test for psr:
ADF Statistic: -5.734297722423357
p-value: 6.491887838944283e-07
Critical Values:
    1%: -3.451281394993741
    5%: -2.8707595072926293
    10%: -2.5716821189216643

ADF Test for er:
ADF Statistic: -2.0990478172142331
p-value: 0.2448976513299471 7
Critical Values:
    1%: -3.4512146067362317
    5%: -2.8707301889607111
    10%: -2.5716664849383113

ADF Test for rgdp:
ADF Statistic: -4.901361699743912
p-value: 3.4662142395570006e-05
Critical Values:
    1%: -3.451281394993741
    5%: -2.8707595072926293
    10%: -2.5716821189216643

ADF Test for ms:
ADF Statistic: -5.489843970708773
p-value: 2.187768520516 5e-06
Critical Values:
    1%: -3.4508226600665037
    5%: -2.8705581218686 21
    10%: -2.5715747316847 34

ADF Test for reer:
ADF Statistic: -12.039387342554248
p-value: 2.754795676142 1107e-22
Critical Values:
    1%: -3.4502615951739393
    5%: -2.8703117734117742
    10%: -2.5714433728242714

ADF Test for lir:
ADF Statistic: -8.983892091674818
p-value: 7.187228328527172e-15
Critical Values:
    1%: -3.4504451681828194
    5%: -2.870392380216117
    10%: -2.571486353732897
```

```
ADF Test for ffer:
ADF Statistic: -4.0042740915764625
p-value: 0.001388295474811846
Critical Values:
    1%: -3.45050711373316
    5%: -2.8704195794076743
    10%: -2.571500856923753

ADF Test for tcs:
ADF Statistic: -1.925104234005886
p-value: 0.32034499279527756
Critical Values:
    1%: -3.451281394993741
    5%: -2.8707595072926293
    10%: -2.571682118921643

ADF Test for indpro:
ADF Statistic: -13.174711175065035
p-value: 1.2300308871279204e-24
Critical Values:
    1%: -3.4502615951739393
    5%: -2.8703117734117742
    10%: -2.5714433728242714

ADF Test for ccpi:
ADF Statistic: -1.4184884504418829
p-value: 0.5733811330113231
Critical Values:
    1%: -3.450951662384033
    5%: -2.8706147570800784
    10%: -2.571604931640625
```

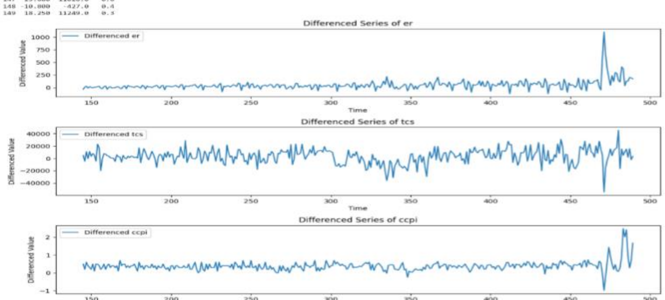
```
ADF Test for er:
ADF Statistic: -2.1153320591356213
p-value: 0.2383943619994383
Critical Values:
    1%: -3.4504451681828194
    5%: -2.870392380216117
    10%: -2.571486353732897

ADF Test for tcs:
ADF Statistic: -7.01919690730656
p-value: 6.610567042964587e-10
Critical Values:
    1%: -3.4496162602188187
    5%: -2.870028369720798
    10%: -2.5712922615505627

ADF Test for ccpi:
ADF Statistic: -0.9380846341943254
p-value: 0.7751720943518075
Critical Values:
    1%: -3.4503224123605194
    5%: -2.870338478726661
    10%: -2.571457612488522
```

```
ADF Test for er after second differencing:
ADF Statistic: -6.128008577308937
p-value: 8.535777987684088e-08
Critical Values:
    1%: -3.4505694423906546
    5%: -2.8704469462727795
    10%: -2.5715154495841017

ADF Test for tcs after second differencing:
ADF Statistic: -8.707390555189443
p-value: 3.668717304499553e-14
Critical Values:
    1%: -3.45050711373316
    5%: -2.8704195794076743
    10%: -2.571500856923753

ADF Test for ccpi after second differencing:
ADF Statistic: -3.5363839586458816
p-value: 0.007100390850753069
Critical Values:
    1%: -3.4503224123605194
    5%: -2.870338478726661
    10%: -2.571457612488522
```

Appendix: VI

F. First and Second differencing

	er	tcs	ccpi
145	-35.650	4887.0	0.5
146	8.795	-4568.0	0.2
147	25.680	11026.0	0.6
148	-10.800	-427.0	0.4
149	18.250	11249.0	0.3




```
# Apply second-order differencing
df_diff_second = df_diff.diff().dropna()  # Apply second difference and drop NaN values

# Display the first few rows of the second differenced DataFrame
print(df_diff_second.head())
```

	er	tcs	ccpi
146	44.445	-9455.0	-0.3
147	16.885	15594.0	0.4
148	-36.480	-11453.0	-0.2
149	29.050	11676.0	-0.1
150	-14.725	-7925.0	-0.1

Appendix: VII

G. Z-score and Outlier check

```
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: A date index has been provided, but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: A date index has been provided, but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: A date index has been provided, but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/statespace/sarimax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starting parameters.
  warn('Non-invertible starting MA parameters found.'
ARIMA Residual Anomalies:
                    ms          rgdp      gira       lir        er  \
Date
1991-01-02  549.730252 -5.175676e+06  3.336059  0.599523  0.173991
1991-01-03  553.650265 -5.419820e+06  3.762376  0.600955  0.179024
1991-01-04  557.792303 -5.200652e+06  4.079935  0.599546  0.186739
1991-01-05  562.156368 -4.542173e+06  4.288736  0.595296  0.197135
1991-01-06  566.742458 -3.420384e+06  4.388779  0.588205  0.210212
1991-01-07  571.550574 -1.843283e+06  4.380064  0.578273  0.225971
1991-01-08  576.580716  1.891294e+05  4.262591  0.565500  0.244411
1991-01-09  581.832884  2.676853e+06  4.036360  0.549886  0.265533
1991-01-10  587.307077  5.619887e+06  3.701371  0.531431  0.289336
1991-01-11  593.003297  9.018233e+06  3.257624  0.510136  0.315821
1991-01-12  598.921542  1.287189e+07  2.705119  0.485999  0.344988
1992-01-02  610.821336  2.090892e+07  1.485787  0.444940  0.404816
1992-01-03  616.372332  2.435215e+07  0.970355  0.439258  0.430001
1992-01-04  621.628691  2.736250e+07  0.527837  0.444224  0.453855
1992-01-05  626.590412  2.933998e+07  0.158235  0.459838  0.473978
1992-01-06  631.257496  3.208458e+07 -0.138452  0.486100  0.491169
1992-01-07  635.629942  3.379632e+07 -0.362224  0.523009  0.505429
1992-01-08  639.707750  3.507518e+07 -0.513081  0.570566  0.516758
1992-01-09  643.490922  3.592117e+07 -0.591023  0.628771  0.525155
1992-01-10  646.979455  3.633429e+07 -0.596050  0.697624  0.530622
1992-01-11  650.173351  3.631453e+07 -0.528161  0.777125  0.533156
1992-01-12  653.072610  3.586190e+07 -0.387357  0.867274  0.532760

            rgdp_zscore      residuals
Date
1991-01-02    -1.062089  302727.568377
1991-01-03    -1.064669  151111.684319
1991-01-04    -1.062437  304038.408768
1991-01-05    -1.055392  154814.466771
1991-01-06    -1.043534  305379.959937
1991-01-07    -1.026862  158487.312299
1991-01-08    -1.005378  306751.018417
1991-01-09    -0.979080  162131.387501
1991-01-10    -0.947970  308150.445712
1991-01-11    -0.912046  165747.797911
1991-01-12    -0.871309  309577.163022
1992-01-02    -0.786350 -283017.769618
1992-01-03    -0.749952 -130565.349739
1992-01-04    -0.718130 -284266.764792
1992-01-05    -0.690884 -134086.496818
1992-01-06    -0.668213 -285542.604090
1992-01-07    -0.650119 -137581.379661
1992-01-08    -0.636600 -286844.318755
1992-01-09    -0.627657 -141050.951605
1992-01-10    -0.623290 -288170.957206
1992-01-11    -0.623499 -144496.130036
1992-01-12    -0.628283 -289521.662388
```

Appendix: VIII

H. Rolling Statistics Anomalies and Performance Residuals Over Time

Rolling Statistics Anomalies:

Date	ms	rgdp	gira	lir	er
1981-01-07	157.685301	2.872412e+04	0.477496	0.733319	0.172500
1981-01-08	159.525263	2.890706e+04	0.497606	0.733335	0.172500
1981-01-09	161.510175	2.912677e+04	0.514963	0.733355	0.172500
1981-01-10	163.640036	2.938325e+04	0.529565	0.733377	0.172500
1981-01-11	165.914847	2.967651e+04	0.541414	0.733403	0.172500
1981-01-12	168.334608	3.000655e+04	0.550509	0.733433	0.172500
1984-01-08	228.067855	3.186048e+04	0.894076	0.733267	0.172500
1984-01-09	231.413106	3.196906e+04	1.056025	0.732272	0.172500
1984-01-10	234.923253	3.209577e+04	1.230592	0.731094	0.172500
1984-01-11	238.598297	3.224061e+04	1.417778	0.729734	0.172500
1984-01-12	242.438237	3.240358e+04	1.617583	0.728191	0.172500
1986-01-07	297.937069	3.093077e+04	-1.340615	0.743742	0.172501
1986-01-08	300.586593	2.771595e+04	-1.422507	0.732413	0.172500
1986-01-09	303.290752	2.393524e+04	-1.454450	0.718545	0.172499
1986-01-10	306.049548	1.958861e+04	-1.436446	0.702140	0.172498
1986-01-11	308.862979	1.467609e+04	-1.368494	0.683197	0.172496
1987-01-09	330.763534	3.906992e+04	-0.085480	0.537366	0.172504
1987-01-10	331.878922	5.397284e+04	-0.001541	0.536091	0.172508
1987-01-11	332.793497	7.120323e+04	0.073075	0.537344	0.172513
1987-01-12	333.507259	9.076109e+04	0.138367	0.541123	0.172519
1988-01-09	354.645908	-2.847193e+04	0.751057	0.574815	0.172485
1988-01-10	359.262021	-8.517604e+04	0.833756	0.575486	0.172469
1988-01-11	364.334040	-1.506123e+05	0.919424	0.575531	0.172451
1988-01-12	369.861966	-2.247807e+05	1.008062	0.574949	0.172429
1989-01-09	434.418804	2.137364e+05	0.810302	0.563375	0.172554
1989-01-10	443.084801	4.252644e+05	0.647652	0.561966	0.172615
1989-01-11	452.048816	6.695068e+05	0.456752	0.560534	0.172684
1989-01-12	461.310849	9.464634e+05	0.237601	0.559080	0.172764
1990-01-09	528.964870	-6.856005e+05	0.701775	0.571685	0.172297
1990-01-10	533.961854	-1.474257e+06	1.122799	0.576651	0.172072
1990-01-11	538.451830	-2.384950e+06	1.618093	0.582333	0.171811
1990-01-12	542.434799	-3.417680e+06	2.187656	0.588730	0.171516
1991-01-09	581.832884	2.676853e+06	4.036360	0.549886	0.265533
1991-01-10	587.307077	5.619887e+06	3.701371	0.531431	0.289336
1991-01-11	593.003297	9.018233e+06	3.257624	0.510136	0.315821
1991-01-12	598.921542	1.287189e+07	2.705119	0.485999	0.344988
1995-01-02	838.601222	3.239404e+07	1.370181	1.178272	0.534362
1995-01-12	814.978244	3.354101e+07	0.435253	1.236356	0.552708
1996-01-01	814.069472	3.398817e+07	0.112738	1.233649	0.555312
1996-01-02	814.147290	3.443910e+07	-0.182546	1.232100	0.558074

Date	rgdp_zscore	residuals	rolling_mean	rolling_std
1981-01-07	-1.007073	752.003866	2.848468e+04	1.034622e+02
1981-01-08	-1.007071	-690.001053	2.851892e+04	1.601001e+02
1981-01-09	-1.007069	706.067956	2.856885e+04	2.376856e+02
1981-01-10	-1.007066	-646.520568	2.863976e+04	3.334425e+02
1981-01-11	-1.007063	665.540723	2.873691e+04	4.439484e+02
1981-01-12	-1.007060	-607.967185	2.886556e+04	5.645743e+02
1984-01-08	-1.007040	-314.069900	3.167997e+04	8.398531e+01
1984-01-09	-1.007039	327.927826	3.169777e+04	1.174173e+02
1984-01-10	-1.007038	-304.280369	3.172935e+04	1.645181e+02
1984-01-11	-1.007036	318.610681	3.177616e+04	2.195616e+02
1984-01-12	-1.007034	-295.028109	3.183965e+04	2.792024e+02
1986-01-07	-1.007050	37.888093	3.618917e+04	2.208836e+03
1986-01-08	-1.007084	-617.973437	3.556746e+04	3.300140e+03
1986-01-09	-1.007124	27.193915	3.458418e+04	4.704725e+03
1986-01-10	-1.007170	-613.658217	3.318826e+04	6.338058e+03

Multivariate Anomalies:

Date	ms	rgdp	gira	lir	er
2021-01-10	53743.45854	279753434.7	2.032271	1.171581	4.476249
2021-01-11	55300.23513	280918676.0	2.009826	1.164061	4.471855
2021-01-12	56908.08952	282086389.3	1.982984	1.155788	4.463747
2022-01-01	58567.02171	283256574.6	1.951744	1.146764	4.451924
2022-01-02	60277.03170	284429232.0	1.916107	1.136987	4.436388
2022-01-03	62038.11949	285604361.4	1.876073	1.126458	4.417137
2022-01-04	63850.28507	286781962.8	1.831642	1.115177	4.394172
2022-01-05	65713.52846	287962036.3	1.782813	1.103144	4.367493
2022-01-06	67627.84964	289144581.8	1.729587	1.090359	4.337100
2022-01-07	69593.24862	290329599.4	1.671963	1.076822	4.302992
2022-01-08	71609.72540	291517089.0	1.609943	1.062533	4.265171
2022-01-09	73677.27998	292707050.6	1.543525	1.047492	4.223635
2022-01-10	75795.91235	293899484.3	1.472709	1.031699	4.178385
2022-01-11	77965.62253	295094390.0	1.397497	1.015153	4.129421
2022-01-12	80186.41050	296291767.7	1.317887	0.997856	4.076743
2023-01-01	82458.27627	297491617.5	1.233880	0.979807	4.020351
2023-01-02	84781.21984	298693939.3	1.145476	0.961005	3.960245
2023-01-03	87155.24120	299898733.1	1.052674	0.941452	3.896424
2023-01-04	89580.34037	301105999.0	0.955475	0.921146	3.828890
2023-01-05	92056.51733	302315736.9	0.853879	0.900089	3.757641
2023-01-06	94583.77209	303527946.8	0.747885	0.878279	3.682678
2023-01-07	97162.10465	304742628.8	0.637495	0.855717	3.604001
2023-01-08	99791.51501	305959782.8	0.522706	0.832403	3.521610
2023-01-09	102472.00320	307179408.9	0.403521	0.808338	3.435504
2023-01-10	105203.56910	308401506.9	0.279938	0.783520	3.345685
2023-01-11	107986.21290	309626077.1	0.151958	0.757950	3.252151
2023-01-12	110819.93440	310853119.2	0.019581	0.731628	3.154904

Date	rgdp_zscore	residuals	rolling_mean	rolling_std \
2021-01-10	1.949876	4458.831188	2.734035e+08	4.147881e+06
2021-01-11	1.962193	4453.956925	2.745552e+08	4.156794e+06
2021-01-12	1.974537	4472.352481	2.757093e+08	4.165707e+06
2022-01-01	1.986907	4467.689182	2.768659e+08	4.174620e+06
2022-01-02	1.999303	4486.074290	2.780249e+08	4.183533e+06
2022-01-03	2.011725	4481.321965	2.791865e+08	4.192446e+06
2022-01-04	2.024174	4499.695727	2.803505e+08	4.201359e+06
2022-01-05	2.036648	4495.154166	2.815170e+08	4.210272e+06
2022-01-06	2.049149	4513.218781	2.826859e+08	4.219185e+06
2022-01-07	2.061675	4508.984911	2.838573e+08	4.228098e+06
2022-01-08	2.074228	4526.743360	2.850312e+08	4.237011e+06
2022-01-09	2.086807	4522.714214	2.862076e+08	4.245924e+06
2022-01-10	2.099412	4540.468420	2.873864e+08	4.254837e+06
2022-01-11	2.112044	4536.344073	2.885677e+08	4.263750e+06
2022-01-12	2.124701	4554.093075	2.897515e+08	4.272663e+06
2023-01-01	2.137385	4550.173379	2.909378e+08	4.281576e+06
2023-01-02	2.150094	4567.619312	2.921265e+08	4.290489e+06
2023-01-03	2.162830	4563.901258	2.933177e+08	4.299402e+06
2023-01-04	2.175592	4581.346010	2.945114e+08	4.308315e+06
2023-01-05	2.188380	4577.529708	2.957075e+08	4.317228e+06
2023-01-06	2.201194	4594.972287	2.969061e+08	4.326141e+06
2023-01-07	2.214034	4591.357617	2.981072e+08	4.335054e+06
2023-01-08	2.226901	4608.500127	2.993108e+08	4.343967e+06
2023-01-09	2.239793	4605.184116	3.005168e+08	4.352880e+06
2023-01-10	2.252712	4621.929441	3.017253e+08	4.361793e+06
2023-01-11	2.265657	4619.208189	3.029363e+08	4.370706e+06
2023-01-12	2.278628	4635.262193	3.041497e+08	4.379619e+06

Appendix: IX

I. Algorithms

Algorithm for ARIMA model

```
Input: D, T_train, T_test, (p, d, q), F(t), σ_adjust
Output: P_ARIMA(t) for t in T_test
1. D_train = Extract(D, T_train)
2. If not stationary(D_train):
     D_train_diff = Difference(D_train, d)
   Else:
     D_train_diff = D_train
3. ARIMA_model = Fit_ARIMA(D_train_diff, (p, d, q))
4. For t in T_test:
     P_ARIMA_base(t) = Forecast(ARIMA_model, t)
     If d > 0:
         P_ARIMA_base(t) = Integrate(P_ARIMA_base(t), d)
     If F(t) = 0:
         P_ARIMA(t) = P_ARIMA_base(t)
     Else if F(t) = 1:
         P_ARIMA(t) = P_ARIMA_base(t)  // Point prediction unchanged
         Adjust_Uncertainty(σ^2, σ_adjust)  // Optional: Widen CI
5. Return P_ARIMA(t)
```

Algorithm for LSTM model

```
Input: D, T_train, T_test, n_lookback, n_layers, n_units, n_epochs, η, F(t),
α_adjust

Output: P_LSTM(t) for t in T_test

1. D_train = Extract(D, T_train)
   D_train_norm = Normalize(D_train, [0, 1])
   (X, y) = Create_Sequences(D_train_norm, n_lookback)
2. LSTM_model = Build_LSTM(n_layers, n_units, output=1)
3. Train(LSTM_model, X, y, n_epochs, η, loss=MSE)
4. For t in T_test:
     X_test = Get_Sequence(D, t - n_lookback + 1, t)
     X_test_norm = Normalize(X_test, same_scaler_as_D_train)
     P_LSTM_norm(t) = Predict(LSTM_model, X_test_norm)
     P_LSTM_base(t) = Denormalize(P_LSTM_norm(t), min(D_train), max(D_train))
     If F(t) = 0:
         P_LSTM(t) = P_LSTM_base(t)
     Else if F(t) = 1:
         P_LSTM(t) = P_LSTM_base(t) * α_adjust
     Update D with P_LSTM(t) for next step if needed
5. Return P_LSTM(t)
```

Algorithm for Ensemble ARIMA-LSTM Model

```
Algorithm Ensemble_ARIMA_LSTM_With_Unseen_Events

Input: D, T_train, T_validation, T_test, P_ARIMA(t), P_LSTM(t), F(t),
β_adjust

Output: P_Ensemble(t) for t in T_test

1. Split D into D_train, D_validation, D_test

2. P_ARIMA(t) = ARIMA_Prediction_With_Unseen_Events(...)

3. P_LSTM(t) = LSTM_Prediction_With_Unseen_Events(...)

4. For (w_ARIMA, w_LSTM) in Grid where w_ARIMA + w_LSTM = 1:

     For t in T_validation:

         P_Ensemble_unadjusted(t) = w_ARIMA * P_ARIMA(t) + w_LSTM * P_LSTM(t)

     MSE = (1/n) * Σ(Y(t) - P_Ensemble_unadjusted(t))^2

     If MSE is minimum, set w_ARIMA* = w_ARIMA, w_LSTM* = w_LSTM

5. For t in T_test:

     P_Ensemble_unadjusted(t) = w_ARIMA* * P_ARIMA(t) + w_LSTM* * P_LSTM(t)

     P_Ensemble_base(t) = clip(P_Ensemble_unadjusted(t), min(P_ARIMA(t),
P_LSTM(t)), max(P_ARIMA(t), P_LSTM(t)))

     If F(t) = 0:

         P_Ensemble(t) = P_Ensemble_base(t)

     Else if F(t) = 1:

         P_Ensemble(t) = P_Ensemble_base(t) * β_adjust

6. Return P_Ensemble(t)
```