MODELING AND FORECASTING HEADLINE INFLATION IN ETHIOPIA: A MACHINE LEARNING APPROACH

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

BY

TEKLU NEGA

ADVISOR: DEREJE DENBE (PhD)

HAWASSA, ETHIOPIA

JUNE, 2024

MODELING AND FORECASTING HEADLINE INFLATION IN ETHIOPIA: A MACHINE LEARNING APPROACH

Master's Thesis

TEKLU NEGA

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES, SCHOOL OF GRADUATE STUDIES, HAWASSA UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS IN APPLIED STATISTICS

HAWASSA, ETHIOPIA

JUNE, 2024

Declaration

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been duly acknowledged.

Name: Teklu Nega

Signature: ___________________ Date: __________________

HAWASSA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ADVISORS' APPROVAL SHEET

This is to certify that thesis entitled "**Modeling and Forecasting Headline Inflation in Ethiopia: A Machine Learning Approach**" submitted in partial fulfillment of the requirements for the degree of Master's with specialization in Applied Statistics, the graduate program of the department of Statistics, and has been carried out by **Teklu Nega,** ID No: **GpApStR/0003/15** under our supervision. Therefore, we recommend that the student has fulfilled the requirements and hence hereby can submit the thesis to the department.

	Signature	Date
Dereje Danbe (PhD)	________________	________________
Name of Advisor		

HAWASSA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

EXAMINERS' APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by **Teklu Nega** have read and evaluated his thesis entitled "**Modeling and Forecasting Headline Inflation in Ethiopia: A Machine Learning Approach**" and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Masters of Science in Applied Statistics.

	Signature	Date
____________________ Name of Department Head	____________________	________________
Dereje Danbe (PhD)____ Name of Advisor	____________________	________________
____________________ Name of Internal Examiner-I	____________________	________________
____________________ Name of Internal Examiner-II	____________________	________________
____________________ Name of External Examiner	____________________	________________
____________________ SGS Approval	____________________	________________

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the School Graduate Committee (SGC) of the candidate's department.

Stamp of SGS Date:__________________

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to the Almighty God for enabling me to complete this study.

I am immensely thankful to Dr. Dereje Danbe, who has been an exceptional instructor, mentor, and thesis supervisor. His insightful and often humorous advice and unwavering support have been invaluable. I am proud of the experience of working with him and deeply appreciate his patience and willingness to clear all my doubts and address every query. He has been a significant guiding force since the beginning of my academic journey at Hawassa.

Additionally, I would like to extend my heartfelt thanks to my mothers, brother, and uncle for their wise counsel and compassionate care. Their constant readiness to assist me has been crucial. They have supported me mentally and financially throughout this challenging journey, helping me to achieve my dream.

I also want to acknowledge the invaluable help of my friends Mr. Tamiru Dame (MSc in Electrical Engineering, Jiaotong University) and Dr. George (Biostatistics, University of Oslo). Their assistance in providing access to essential study materials was instrumental in completing my thesis.

Moreover, I am grateful to my peers and colleagues who offered their support and encouragement throughout this endeavor. Their companionship and collaboration have made this journey a rewarding experience. Lastly, I thank my academic institution, Hawassa University, for providing a conducive environment for my studies and research.

ABBREVIATIONS AND ACRONYMS

ANN/NN	Artificial Neural Network/ Neural Network
ARCH	Autoregressive Conditional Heterosdesckcity
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back propagation Neural Network
CPI	Consumer Price Index
CSS	Central Statistics Services
CV	Cross Validation
GARCH	Generalized Autoregressive Conditional Heterocedasticity
GDP	Gross Domestic Product
IMF	International Monetary Fund
KNN	K-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operators
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
NBE	National Bank of Ethiopia
NNAR	Neural Network Autoregressive Neural Network
RF	Random Forest
RMSE	Root Mean Squared Error
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Averages
SGD	Stochastic Gradient Descent
SVAR	Structural Vector Aggressive
VAR	Vector Autoregressive model
VECM	Vector Error Correction Model
WB	World Bank

Tables of Contents **Page**

LIST OF TABLES **Page**

LIST OF FIGURES **Page**

ABSTRACT

Inflation is an important indicator of a nation's welfare and has become one of the major economic challenges globally, especially in Ethiopia. Several studies forecasted inflation in Ethiopia using traditional models, as accurate forecasts contributed to a more stable economic environment, even though forecasting with traditional models was challenging. Thus, this study aimed to model and forecast headline inflation in Ethiopia using a machine learning approach. The study was based on secondary data recorded on headline inflation and related factors from January 2000 to December 2023, obtained from various inflation-related organizations. The study utilized multivariable time series data for the past 24 years. The data were transformed, standardized and split into training and testing sets to enhance the forecast accuracy of both the machine learning and time series models. Cross-validation (CV) and grid search were used to tune the machine learning parameters such as the penalty parameter, learning rate, number of trees, maximum iterations, and the number of hidden nodes and layers and also select model with minimum MSE or RMSE. A stochastic gradient decent (SGD) was used to optimize the parameters. The selected models were evaluated based on performance evaluation criteria, including RMSE, MAE, and MAPE tests. The headline inflation in Ethiopia saw slight increases from 2000 to the second quarter of 2007, followed by a sudden shift in the fourth quarter of 2008, and then a rapid increase from the first quarter of 2015 to the fourth quarter of 2023. Food inflation, non-food inflation, export and import prices of goods and services, political stability index, exchange rate, numbers of vehicles, rainfall, world oil price, gross domestic fixed investment, unemployment rate, T-bill sales, agricultural production price were predictors significantly determine the headline inflation in Ethiopia. Furthermore, Food inflation, non-food inflation, export and import prices of goods and services, number of vehicles, gross domestic fixed investment were the most factors that determine the forecasting accuracy of the models. Among various forecasting methods, a specific ANN architecture called NNAR emerged victorious. It outperformed Ridge regression, LASSO, Elastic Net, Random Forest, and even the benchmark model in terms of accuracy for both in-sample and out-of-sample inflation forecasts in Ethiopia. NNAR achieved the lowest RMSE, MAE and MAPE, solidifying its position as the most effective model in this study. Finally, this study recommended that policymakers, financial analysts, investor and stakeholders should give attention to the identified drivers of headline inflation and consider using advanced machine learning models.

Keywords: Headline Inflation, Machine Learning Approach, Forecasting, Ethiopia.

CHAPTER ONE

1. INTRODUCTION

1.1 Background of Study

Inflation is an important indicator of the welfare and well-being of a nation and is undeniably an integral component of economic development. According to Salim (2019) inflation is a state in which the value of money is falling or prices are rising. As a result, inflation lowers the purchasing power of the people as it requires more money to afford one unit of goods or service (Ruzima, 2018). For monetary policy purposes, inflation is measured in two ways: headline inflation and core inflation. The headline inflation is a measure of the total inflation in an economy including commodities such as food and energy which tend to be much more volatile and prone to inflationary spikes; whereas core inflation is the total inflation excluding food and energy prices based on the assumption of stable price of food and energy items (James, B. 2011).

Inflation has been increasing globally, with noticeable rises in 2021 compared to 2020. According to the IMF reports (2021), the inflation rates for the world, emerging and developing economies, and Sub-Saharan Africa were 2.4%, 7.5%, and 8.6% respectively in 2020 and by 2021 these rates had grown to 4.9%, 7.6%, and 11% respectively. According to the IMF reports (2023), a combination of climate shocks and the epidemic disintegration of food and energy product and distribution are driving up a cost of living around the world. For numerous member countries, Russia's irruption of Ukraine worsened a formerly delicate situation by pushing the prices of energy, food, and diseases indeed higher aggravating energy and food shortage. More broadly, although inflation has been declining in response to numerous central banks ' interest rate hikes, most countries still face elevated headline and core inflation (IMF, 2021).

Inflation is a complex phenomenon that is influenced by a wide variety of factors, including economic growth, changes in interest rates, and shifts in the supply and demand for goods and services (Ball, R. J., 2017). Numerous studies on the sources of inflation in sub-Saharan Africa including money growth, change in exchange rate, commodity prices, and supply shocks have been reported the major determinants of inflation (Prakash and Maiti, 2016).

According to the World Bank report (2022), Ethiopia has become one of the top ten inflationary countries in the world. Similarly, the Central Statistical Service of Ethiopia (CSS, 2023), has reported that annual average headline inflation rose to 35.1% in November 2023 as compared to in November 2022 which was 33%. Non-food inflation increased 25.2% to 36.5%, but food inflation decreased 38.9% to 34.25 from November 2023 to November 2022 and the highest food inflation, 43.9%, was recorded in May 2022.

Non-food inflation rate increased to 30.4% in January 2023 as compared to the similar period January 2022. Additionally, the inflation rate has remained stubbornly high in recent years, fueled by drought and a two-year conflict in Tigray region. The mean annual headline inflation in recent five years were 13.83%, 15.81%, 20.36%, 26.84% and 33.9% in 2018, 2019, 2020, 2021 and 2022 respectively and increased more than double-digit inflation. During the COVID-19 pandemic, Ethiopia has faced rising inflation that has complicated the government's economic and health policy responses to the virus (Negatu, 2023). There were also other causes of Ethiopia inflation such as flooding and desert (CSS, 2022), broad money supply, government budget deficit, exchange rate, real gross domestic product, rainfall, the nominal exchange rate and inflation expectation, oil price, and world food prize, treasury bills, lending rate, unemployment rate, price of imported and exported good and service, private and government consumptions are the major determinants of inflation (Kashay, 2017; Tekeber N., et al., 2019; Mulugeta, 2020; D.Bedada, et al., 2020; Demeke, 2020; Melaku, 2020; Nakorji and Aminu, 2022; Mohamed F., 2023; Baldwin, 2018).

The financial policy in most central banks is designed for controlling inflation at a low. As a central bank, the National Bank of Ethiopia (NBE) has an aim to maintain price stability by achieving a single digit inflation rate (Mihretu, 2023). Kibrom (2008) argued that the significant rise in global food and energy prices over the past decade has increased the importance of the headline inflation rate for policy decisions. Monetary policy primarily targets the headline inflation rate to maintain price stability (Nyoni and Bonga, 2018). Maintaining price stability, essential for macroeconomic stability, requires minimizing the impact of demand-side variables driven by the financial aspect of inflation (Gebremeskel A., 2020). The dynamics of Ethiopian inflation are predominantly influenced by agriculture and food and are nonlinear in nature (Durevall *et al.,* 2013; Abebe *et al.,* 2023). Considering nonlinearities offers a more nuanced understanding of inflation's behavior, potentially leading to better forecasts (Silva G., *et al.,* 2023).

Accurately forecasting inflation is essential not only for guiding inflation targeting policies to achieve price stability but also for making monetary policy transparent and credible. Most national banks regularly release their inflation forecasts to improve economic decision-making. Although forecasting inflation is challenging, accurate forecasting contribute significantly to a stable economic environment (Özgür and Akkoç, 2021). The role of inflation forecasting in economic decisions is crucial, benefiting investors, policymakers, and households alike. Therefore, this study was focused on machine learning methods to improve the forecast accuracy of the models in order to get accurate forecast of inflation which used for policy making.

1.2 Statement of the Problem

The recent historical evidence indicates that Ethiopia has experienced a rapid increase in inflation due to various factors including weather shocks, disruptions caused by the COVID-19 pandemic and ongoing internal conflict (World Bank, 2022). The inherent instability of Ethiopia inflation and its multifaceted nature pose significant challenges for accurate forecasting, impacting monetary policy decisions and economic planning efforts (Abebe, *et al.*, 2023; Arsovski, *et al.*, 2023; Mirza, *et al.*, 2024).

Even though forecasting inflation is a valuable tool for planning, there are inherent limitations and challenges that can lead to errors. Common problems with forecasts include incomplete data, the selection of the correct forecasting model, and human bias (Rosanne, 1967; Hyndman, 2006; Peter, 2009). These issues can result in inaccurate forecasts. Inaccurate inflation forecasts can have significant and far-reaching consequences, affecting various aspects of the economy, as well as social and financial stability. Some of these risks include monetary policy errors, poor business planning, misguided investment decisions, and flawed governmental fiscal policies (IMF, 2013).

Several studies have conducted in Ethiopia used traditional econometric and time series models such as VAR, ARIMA, VECM, ARMA-GARCH and ARDL to model, forecast and identify determinants of inflation in Ethiopian. These studies mainly focused on the demand and supply side determinants of the inflation excluding others important variables (D. Bedada., et al., 2020; Kashay, 2017; Tekeber N.,et.al., 2019; Mulugeta, 2020; Abebe, 2021; Hagos, 2014, Getachew, 2020; Mihretu M., 2023).

Another recent study conducted by Abebe et.al (2023) shows that nonlinear models, threshold autoregressive (TAR) and SETAR models capture features of macroeconomic variables like asymmetry or nonlinearity which unable to capture by traditional econometrics and time series models mentioned above. Even though researchers seem to capture the pattern of the macroeconomic variables given common market conditions, these models are not effective in providing accurate result for the periods of economic shocks and recessions. During such turbulent times, the relationships between variables can change dramatically; rendering the assumptions and relationships built into these models less reliable (Maehashi and Shintani, 2020). The aforementioned studies in Ethiopia did not take in account the forecast accuracy of the models and primarily relied on the aggregate of the general Consumer Price Index (CPI). This aggregate represents the average change in prices over time for a market basket of consumer goods and services. Aggregate of general CPI does not clarify the percentage change in CPI yearly, quarterly or monthly (Tiedemann J., 2024). Finally, to the best of our knowledge, there was no study conducted in Ethiopia to forecast headline inflation using machine learning approach. Thus, this study was focused to model and forecast headline inflation in Ethiopia and to identify determinant factors using machine learning approaches.

Therefore, this study was intended to answer the following research questions:

i. What was the trend of headline inflation in Ethiopia from 2000 to 2023?

ii. Which independent variables were major contributors of headline inflation in Ethiopia?

iii. Which variable mainly determines the forecasting accuracy of the models?

iv. Which machine learning methods best fit and forecast Ethiopia inflation?

1.3 Objectives of the Study

1.3.1 General Objective of the Study

The general objective of this study was to model and forecast headline inflations in Ethiopia using Machine Learning methods.

1.3.2 Specific Objectives of the Study

Specifically, this study was intended to:

- Examine the trend of headline inflation in Ethiopia.
- Identify covariates significantly contributing to the headline inflation in Ethiopia.
- Identify covariates that significantly determine the forecast accuracy of the models.
- Compare the forecasting performance of machine learning models fitted to the data.

1.4 The Significance of Study

The findings of this study may add a new approach to forecasting inflation by including important financial policy instruments yet considered in other literatures. In addition, this study would highlight importance of inflation as a crucial policy and program design, and in the targeting and monitoring of public goals. Identifying the optimal model that has robust forecasting ability would be obtained in the planning conditioning of the government, businesses and the public in general. Furthermore, it would help other researchers to conduct further research with this approach as a benchmark model.

1.5 The scope of the Study

This study mainly focused on modelling and forecasting headline inflation in Ethiopia using the machine learning approaches such as shrinkage models, ANN and Random forest models only. This study was limited to a potential driver’s of headline inflation: demand side, supply side and structural factors of inflation.

1.6 Organization of the Study

This research paper is divided into five chapters: chapter one deals with introduction that includes the background, statement of the problem, objective of the study, scope and significance of the study. Chapter two addresses review of literature (both theoretical and empirical) and chapter three focuses methodological aspect of the study which includes: variable definition, model speciation, estimation and diagnostic procedures. Chapter four of the study deals with the results and the discussion. The last chapter, chapter five, deals with the conclusion and policy recommendation of the study.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 General Definition of Inflation

Inflation is defined as a process of continuously rising prices and falling purchasing power. In other words, inflation is a general and broad-based increase in the price of goods and services over an extended period (Martin, 2022). Definitions given by the economists like Crowther (1958), Gardner Ackley(1961), and H.G. Johnson(1972) regarded inflation as a phenomenon of rising prices. According to Crowther (1958), inflation is a" state in which the value of money is falling, i.e., the prices are rising." Johnson (1972,) countries," I define inflation as substantial rise in prices. For monetary purpose this inflation divided into two; headline inflation and core inflation. Headline inflation is the total inflation in an economy and it also called general inflation. It is different from core inflation, which excludes food and energy prices while calculating inflation (Corporate Finance Institute, 2022). Neo-classical economists defined headline inflation as a galloping rise in prices resulting from an excessive price of food and energy, as well as rise in the volume of money in circulation; assuming full employment level, thus, inflation is a financial phenomenon. To Keynes, Inflation is the patient rise in the broad spectrum of goods and services because of a rise in aggregate demand. Inflation compactly is a sustained rise in the general price level of goods and service.

2.2 Theories on Causes of Inflation (Headline inflation)

2.2.1 Demand Pull Theory

Demand-Pull inflation theory is the most frequent categorization of inflation (Jhingan, 1997). According to Keynesians, the main source of Demand-Pul1 inflation is the increase in aggregate demand which sums up consumption, investment and government expenditure. To them, inflation is caused by demand-pull and not cost-push. When the economy experiences a large imbalance between aggregate demand and supply in which excessive demand manifests, the faster is the inflation (Totonchi, 2015). Bent Hasen's excess demand theory presented an explicit dynamic inflation model, which incorporates two separate price levels, the goods, and labor market prices (Day, 1952; Claes-Henric, 2020).

The Keynesian General Theory of Employment, Interest, and Money, policy issue that targeted reduction in each section of aggregate demand is successful in decreasing pressure on demand and inflation. The theory posited that when both excess demand for goods and the excess demand for factors are positive, prices and wage rates will increase. Government tax increment is one of the policy instruments in which expenditure is reduced and to control the size of money alone or jointly, can be efficient in dropping effective demand controlling inflation (Keynes, 1936). The recent rise in headline inflation following surges in food and energy prices has prompted concerns that high inflation could persist (Zheng Liu, 2011).

2.2.2 Cost Push Theory

Cost-push inflation occurs when we experience rising prices due to higher costs of production and higher costs of raw materials. Cost-push inflation is determined by supply-side factors, such as higher wages and higher oil prices (Tejyan P, 2022). Cost-push inflation may be further aggravated by upward adjustment of wages to compensate for rise in cost of living. A few sectors of the economy may be affected by increase in money wages and prices of their products may be rising. In many cases, their products are used as inputs for the production of commodities in other sectors. As a result, cost of production of other sectors will rise and thereby push up the prices of their products. Thus wage-push inflation in a few sectors of the economy may soon lead to inflationary rise in prices in the entire economy. Further, an increase in the price of imported raw materials may lead to cost-push inflation (Totonch 2011). In other words, cost push inflation may be further provoked by upward alteration of wages to compensate for the adverse effect caused by increase in cost of living (Balami, 2006).

Moreover, another cause of Cost-Push inflation is profit-push inflation in which oligopolistic and monopolist firms charge the price of their products so as to compensate for the rise in wage and other cost of production to make higher profits. This is known as administered-price inflation or price- push inflation (Totonchi, 2015). Some economists such as Lowe (2017) and Yellen (2017) opined that technological advancement has transformed the process of production such that it affects the prices of commodities.

2.2.3 Quantity Theory of Money

Quantity theory of money is the earliest economic theory. The theory states that the general level of price change is mainly caused by changes in the quantity of money circulated at the hand of the public (Fisher,1947). The quantity theory of money created the main dialogue issues of classical monetary analysis of the 19th century, in line with the leading conceptual framework for infer in modern financial events and emergence of different scholars towards main department policy instruction intended to safeguard the gold standard. According to David Ricardo (1772-1823) quantity theory of money indicates that any price change must be overwhelmed by the same proportional variation in the Quantity of money. The monetarists adopted Fisher’s (1876-1947) quantity theory of money in explaining inflation. Fisher’s equation of exchange is thus:

$$\mathbf{MV = PQ}$$

Where M represents the money supply, V is the velocity of money in circulation, P represents the price level of goods and services and Q is the level of real output. The theory works on the assumption that V and Q are constant, and price level (P) varies proportionately with the supply of money (M). This implies that inflation, measured by price level is caused by a rise in money supply in an economy, therefore, inflation is a monetary phenomenon). Moreover, monetarists' superiority over Keynesian in policy effectiveness could be approved by employing the common identity exchange equation of Fisher (Totonchi, 2015). Generally, the quantity theory of money which provides an equation of money supply emphasizes on the role of excess money supply in explaining inflation (Cukierman, A, and S Gerlach, 2003).

2.2.4 Structural Inflation Theory

Structuralist inflation theory is based on an identity relation that the price of output is equal to its costs of production. The total cost can be broken down into gross profits, total wages, and intermediate inputs. Gross profits include depreciation, taxes, and productive/ancillary/overhead cost. Total wages also include taxes and social contributions payable by employers. Intermediate inputs can be domestic or imported, with the latter determined by international prices and the nominal exchange rate. In this model, the final price of each product depends on technology (how much labor and inputs are needed per unit of production), import prices, and social conflict (how much of the income created goes to taxes, profits, and wages (ILO, 2015; Taylor and Ömer, 2020; Barbosa-Filho, 2021)

There are many studies that preceded this study in using econometrics and statistical methods, as well as machine learning models that focused on the modelling and forecast the inflation. The former models used when the inflation is linear and stationary, but latter used in both non-linear and non-stationary (Rita, 2019; Adolfo R., 2020; Bruckheimer, 2022)

2.3 Machine Learning Methods

In the last two decades, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the forecasting community. Time series data are not always linear and requires more complex approaches for data forecasting (Appiah, K, *et al.*, 2009). The unique patterns in these series make forecasting difficult using linear traditional models, therefore; the application of machine learning has been growing rapidly in forecasting inflation. The Nakamura (2005) is an early attempt to apply neural networks for forecasting U.S inflation. Inoue and Kilian (2008) considered U.S. inflation forecasts from LASSO and Ridge regression. The recent popular papers, in which machine learning methods have been used to predict inflation, include (Chakraborty and Joseph 2017; Garcia et al., 2017; Maehashi and Shintani 2020; Adolfo R., 2020; Baybuza, 2018; Bruckheimer, 2022, Yuniar I., *et al.*, 2020) among others. The results of the studies show that machine learning models methods are able to produce more accurate inflation forecasts than benchmark models.

2.3.1 Introduction to Machine Learning

Machine learning is a branch of artificial intelligence often described as the art and science of pattern recognition. It is essentially a data-driven approach with mild assumptions about the underlying statistical relationships in the data, and it entails a large variety of methods. It usually consists of two core elements, a learning method and an algorithm, enabling one to automate as many of the modeling choices as possible in a manner that is not subject to the discretion of the forecaster (Hall, 2018).

One of the most common machine learning techniques is the Artificial Neural Network (ANN). The term "Artificial Neural Network" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another; artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes (Stahl and Jordanov, 2012) (See Figure 2.1).

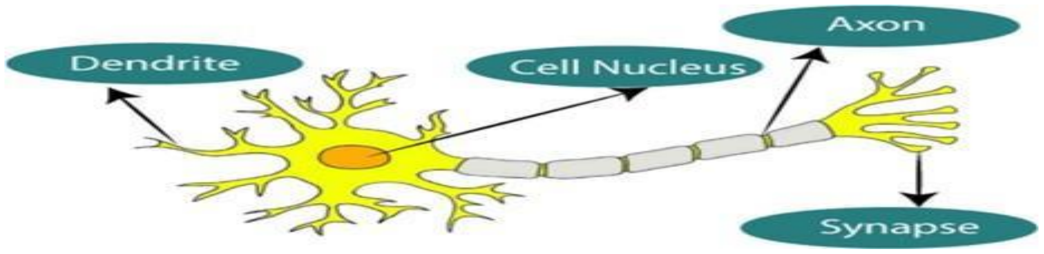


Figure 2.1: Diagram of Biological Neural Network

The typical Artificial Neural Network looks something like the given figure 2.2. Dendrites from Biological Neural Network represent inputs in Artificial Neural Networks, cell nucleus represents nodes, synapse represents weights, and axon represents output. Relationship between Biological neural network and artificial neural network (Rei and Wang, 2020)

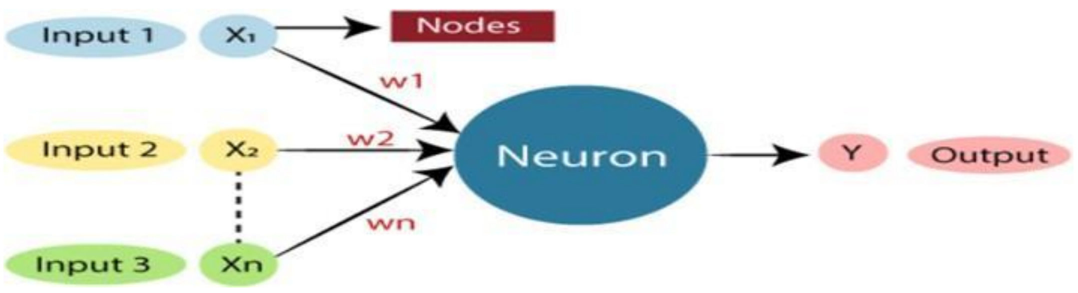


Figure 2.2: Artificial Neural Network (ANN) Model

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell nucleus	Nodes
Synapse	Weights
Axon	Output

Figure 2.3: Relationship between Biological neural network and artificial neural network

2.3.2 Machine Learning Methods in Forecasting

Time series and machine learning approaches are two sides of a coin. They can be combined together in order to give you the benefits of each approach. Time series does a good job at decomposing data into trended and seasonal elements. This analysis can then be used as an input into a machine learning model, which can incorporate the trend and seasonal information into its algorithm, giving you the best of both worlds (Sorjamaa, *et al.,* 2006). The forecasted cannot be reliably, if the actual outcome is affected by the forecasts, the reliability of the forecasts can be significantly lower (Casdagli, *et.al.,* 1991). Econometrics and time series models are some of the most popular methods for inflation forecasting; however, they have the limitation that it works well in a linear model. The problem arises that inflation is non-linear (Abebe, *et al.,* 2023).

Most traditional forecasting methods rely on fitting data to a prespecified relationship between dependent and independent variables, thus assuming a specific functional and stochastic process. In contrast, machine learning offers a different approach to statistical analysis and forecasting, in particular, because it is to a great extent data-driven, as it makes almost no assumption about the underlying statistical relationships in the data (Samuel, 1959; Hansen, 2019).

Many studies compared traditional forecasting models to machine learning models like Support Vector Regression, Artificial Neural Network, Decision Trees, Random Forest, XG Boost and K-Nearest Neighbor regression. They concluded that machine learning is a powerful tool in forecasting compared to univariate or multivariate time-series models (Özgür and Akkoç, 2021; Rita, 2019; Ülke et al., 2016; Michael V., *et al.,* 2020; Mahajan & Srinivasan, 2019)

2.4 Empirical Review

2.4.1 Empirical Review on Determinant of Ethiopia Inflation

A number of studies were tried to find the main determinants of inflation in Ethiopia. For instance, Mamo B.,(2016) has carried out a research on what causes inflation in a Post-Communist Economy using Evidences from Ethiopia. He has employed Johnson co integration and VEC approach tests to reveal that broad money supply, government budget deficit, exchange rate and inflation expectation are found to be the major determinants of Ethiopia inflation. Result also shows that domestic gross product growth rate has no effect on inflation. Similar study by D. Bedada, *et al.*, (2020) has assessed the determinants of inflation over the period 1974/75 to 2014/15 using Johnson co-integration and VEC approach. The result indicated that in the long run consumer price index has been found to be positively influenced by money supply, real gross domestic product and overall budget deficit. Tekeber N., et al., (2019) employed Autoregressive Distributed Lag (ARDL) to examine the supply and demand side determinant of inflation in Ethiopia based on inflation records from 1985 to 2016. The empirical results imply that there was an evidence of a long-run positive impact of money supply, world oil price, budget deficit and real effective exchange rate on inflation in Ethiopia whereas real gross domestic product insignificantly affected price level. Results also show that, in the short run, real effective exchange rate, money supply, budget deficit and world oil price are the main determinants of inflation in Ethiopia. Other study by (Kashay, 2017) employed OLS and found that GDP, money supply and national saving were significantly and positively contributed to inflation rate both in the short and long-run. Mulugeta (2020) has also used VAR, co-integration analysis and the VECM to identify the major sources of inflation. The result showed that in the short run, the Broad Money Supply, Real GDP, the imported inflation of international petroleum price and the nominal exchange rate has an insignificant effect on Ethiopia's annual price level. In the long run, Ethiopia's inflation is mainly driven by the broad money supply, the real GDP, interest rate, rainfall, the nominal exchange rate, and the budget deficit with high significance level.

On the other hand, Mohamed F. (2023) uses machine learning algorithms like Support vector machine K-nearest neighbor, Random Forest, Artificial Neural Network, Gradient boosting and decision tree to determine the accurate algorithm and analyze the factors affecting Egypt inflation. The study found that major significant variables determining inflation in Egypt are the exchange rate (30.5%), gross fixed formation (24.5%) and government expenditure (12.3%). They also found a positive relationship between the inflation rate and government expenditure, money supply, gross domestic product (GDP) growth, gross fixed formation, foreign direct investment, GDP per capita and exchange rate.

2.4.2 Empirical Review on Inflation Forecasting

The empirical studies related to inflation forecasting from Ethiopia were reviewed to identify gaps for the current study. Many scholars have investigated the relative accuracy of alternative inflation forecasting models. Recently, machine learning methods have been used to compare the accuracy of inflation forecasts relative to classical econometric models and time series models for forecasting inflation. Abebe *et al.,* (2023) conducted research to model inflation rate factors on present consumption price index (CPI) in Ethiopia by using TAR models. They show that the nonlinear SETAR model outperformed the linear ARMA models. The superiority in performance of nonlinear models (TAR) was attributed to their ability to capture the stochastic nature of the monthly rates as evident in the pattern of the forecast errors. However, Teräsvirt T., *et al.,* (2004) examined forecast accuracy of time series models for 47 monthly macroeconomic variables of the G7 economies. The results indicated that the STAR model generally out-performs linear autoregressive models, but neural network models at long forecast horizons by using Bayesian regularization produces more accurate forecasts than a corresponding model specified model. Mihretu M. (2023) compared different inflation forecasting models and combinations of techniques that best fit Ethiopian inflation forecasting. He compared ARIMA, ECM, VECM, Phillips curve and BVAR models using RMSE for both in-sample and out-of-sample forecasting. The empirical result revealed that the ARIMA model performs better than others models whereas VECM performs worse than other models compared to eight period ahead forecasts. Additionally, the Augmented Dickey-Fuller test confirms that Ethiopian CPI data does not follow a random walk.

Several research studies done out of Ethiopia showed the performance of machine learning in forecasting inflation. For instance, Baybuza (2018) forecasts Inflation Using Machine Learning Methods of penalized regression (LASSO, Ridge, and Elastic Net) and Random forest, as well as boosting techniques in Russia. The empirical study revealed that the Random forest and boosting model are the best inflation forecasting models compared with traditional models like random walk and auto-regression. The result shows LASSO mode outperforms Ridge model, benchmark models, but elastic net model are almost identical to the results of the LASSO. Similarly, study by Nakorji and Aminu(2022) have forecasted the inflation rate in Nigeria by machine learning techniques. The result indicated that Ridge regression and ANN are the best in forecasting inflation in Nigeria. The study further revealed that the major determinants of headline inflation in Nigeria were food inflation, core inflation, high lending rate, maximum lending rate, and the inter-bank rate. Özgür and Akkoç (2021) examined the shrinkage model to identify which method performs better than the baseline autoregressive methods. They compare the data from 2007 to 2019 for their forecasting estimate and use five different shrinkage models with ridge, LASSO, adaptive LASSO, Group LASSO, and ElasticNet as the comparative models. They conclude that LASSO and Elastic Net performed best compared to all the other baseline and shrinkage models with root mean squared error of 0.834 and 0.893 with lambda values at 0.065 and 0.710, respectively. Similarly, research done by Popoola, Y. (2023) showed that the machine learning models such as LASSO, Ridge, and Elastic Net regression have the predictive performance over traditional models such as random walk, ARIMA in forecasting Swedish inflation. The result also reveals for all forecasting horizons, the Elastic Net yields more accurate predictions than the benchmark test. Hakizimana, C. (2022), implemented machine learning techniques such Random Forest, Ridge Regression, LASSO Regression and K-Nearest Neighbor (KNN) to forecast forecasting Rwandan inflation. The result revealed that Random Forest model is the best model that works well in forecasting inflation. The study suggested that include other machine learning techniques such as Neural networks methods may improve the forecast accuracy.

Moreover, study conducted in Nigerian and American, Indonesia, Ghana, Kenya, India, China and South Africa and Portuguese were intended to forecasting inflation by using machine learning models. The result show that ANN model outperforms traditional models such as ARMA, VAR, ARIMA, SARIMA, ARCH and GARCH models (Michael V., *et al.,* 2020; Yuniar I., *et al.,* 2020; Nunoo, E. and Kings, I. 2013; Mwangi 2016; Mahajan and Srinivasan 2019; Rita, 2019). Mwangi, (2016) also recommended that other data variables like political stability and security threats can be included to enhance the predictive capabilities of the model.

Besides some traditional econometric approaches to forecast inflation, such as the ARIMA and SARIMA models, this paper considered machine learning techniques, (ANN, LASSO, Ridge, and Elastic Net, Random Forest) well-known inflation forecasts embedded in financial market data. The set of forecasting methods also includes linear and nonlinear machine learning methods and several forecast techniques that have not yet been addressed in Ethiopia.

CHAPTER THREE

3. METHODOLOGY

3.1 Description of Study Area

Ethiopia is located in Eastern Africa between 30 and 180 N Latitude, and 380 and 480 Longitude. Total land area of the country is about 122.2 million hectares. The central part of the country is mostly high plateau, 1500 to 3000 meters above sea-level. The country is divided into 12 regional states and two chartered cities (Addis Ababa and Dire Dawa). The country has a vast potential of agricultural land with favorable climate, soils and vegetation, which are conducive to the production of a number of crops for domestic consumption and export. The current population of Ethiopia has been estimated to be 129,719,179 in 2024 and which has shown a 2.52% increase from 2023. The majority of the population, over 85 percent, lives in rural areas and obtains their livelihood from agriculture. Agricultural commodities, especially cereals and oil seeds, production and their marketing are the means of livelihood for millions of households in the country.

3.2 Source of Data and Variables of the Study

The study was based on secondary data recorded on headline inflation and related factors from January 2000 to December 2023. The data was collected by the Central Statistics Service (CSS), National Bank of Ethiopia (NBE), Ethiopian National Meteorological Agency (ENMA) and World Bank (WB).

Variables in the study

The consumer price index (CPI) is the central series of this study, and choosing an appropriate measure for it is complicated. The most publicly accepted measure of inflation is the headline inflation computed by the Central Statistical Service (CSS). The series for headline inflation, food inflation and nonfood inflation were also obtained from CSS.

3.2.1 Dependent Variable

The dependent variable considered in the study was headline inflation (HI). Headline inflation is the total inflation in an economy and it includes inflation in a basket of goods that includes commodities like food and energy. Headline inflation is more volatile due to strong transitory fluctuations influenced by the official sector or determined outside the Ethiopian economy, prompting many to emphasize core inflation by excluding volatile CPI components like food and fuel. However, in the Ethiopian context, food and fuel are significant in the CPI basket and should not be ignored (Gashaw, 2021).

Most central banks use headline inflation as their target variable. The reason being headline inflation is a broad measure that closely represents the basket of goods and services consumed by most households (Clinton, 2006). Rich and Steindler (2005) one can argue that headline inflation or indeed a broader measure of inflation serves as the ultimate goal in an inflation targeting framework. The most common measure of inflation is the percentage change in the headline consumer price index (CPI), which captures the cost of living of the average consumer. The CPI includes also domestically produced and imported consumer goods and services.

The headline inflation (HI) can be calculated as

$$HI_t = ln\ (CPI_t - CPI_{t-12}) \tag{3.1}$$

where, HI_t headline inflation at time t, $ln\ is\ natural\ logarithm,\ CPI_t$ is general consumer price index at time and CPI_{t-12} is last year's same month general consumer price index calculated by CSS.

3.2.2 Independent Variables

The independent variables included in the study were food inflation and non-food inflation obtained from CSS. Broad and narrow money supply, treasury bills demand and sold, the exchange rate, price of imported and exported good and service, agricultural production price, investment, and national saving deposit, net foreign asset, minimum deposit, gross government consumption, gross private consumption, capital account, government budget deficit, government expenditure, foreign direct investment, numbers of vehicles, real GDP, net foreign demand, lending interest rate and unemployment rate were the predictor variables obtained from National bank of Ethiopia (NBE). Observations on average rainfall collected from Ethiopian National Meteorological Agency; world oil price, world food price, political stability index obtained from WB (See Table Appendix A).

3.3 Data Preparation and Preprocessing

After gathering the data, it needs to undergo with data preparation before analysis. This involves several key steps: cleaning the data by removing any missing or inaccurate entries, ensuring the data is in an appropriate time format, transforming it to a suitable structure for analysis, and dividing it into training and test sets. In the case of monthly headline inflation data, we have used transformation methods to quarterly data avoid the volatility nature in the data set, which could pose challenges for traditional linear models like ARIMA models; similarly, too covariates.

The data frame has been split into training and testing datasets in the ratio of 4:1 (80% train and 20% test datasets) .The model was built on train data and the test data was used in prediction to check for the accuracy (Joesph, *et al.*, 2021). Window functions have been used to perform statistical operations on data subsets, which split data into train data and test data set.

3.4 Statistical Models (Benchmark Model)

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. It is an important step in any data analysis. It used to understanding where outliers occur and how variables are related can help one design statistical analyses that yield meaningful results. The patterns in data can be expressed by descriptive statistics, time series plot, and correlation matrix (Sharma, *et al.,* 2007; Andersen T. *et al.,* 2008).

In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. Autoregressive integrated moving average (ARIMA) model is one of the most widely used time series model, in which a time series is expressed in terms of past values of itself (the autoregressive component) plus current and lagged values of a 'white noise' error term (the moving average component). We concerned with the ARIMA and its extension models as a benchmark model to analyze headline inflation based on previous study (Araujo and Gaglianone, 2022; Michael V., *et al.,* 2020; Yuniar I., *et al.,* 2020; Rita, 2019).

3.4.1 Autoregressive Integrated Moving Average (ARIMA) Model

The autoregressive moving average (ARMA) model was first introduced in the 1950s by the statistician Whittle (1951), who developed it as a method for analyzing and forecasting time series data. The model is most widely used in forecasting the univariate time series data analysis (Hyndman and Athanasopoulos, 2018), and also the most common benchmark for neural networks (Gonzalez, 2000).

ARIMA is an extension of the Autoregressive Moving Average (ARMA) model that can also handle non-stationary data. The model is composed of three components: the autoregression (AR) component, the Integration (I) component, and the moving average (MA) component. The autoregression component of an ARIMA model captures the dependence between an observation and a number of lagged observations. The moving average component captures the dependence between the residual errors of the time series at different times. The Integration component, also known as the differencing component, is used to remove non- stationarity from the data. This is done by taking the difference between consecutive observations.

The ARIMA (p,d,q) model is a combination of the three component. The general form of the ARIMA model can be rewritten as:

$$\Delta^d HI_t = \sum_{i=1}^{p} \beta_i \Delta^d HI_{t-i} + \sum_{j=1}^{q} \theta_i \epsilon_{t-j} + \epsilon_t \qquad (3.2)$$

$\Delta^d HI_t = HI_t - HI_{t-d}$, where, $\Delta^d HI_t$ is change or the difference between headline inflation (HI_t) at time t and lag of headline inflation (HI_{t-d}) at time$(t-d)$ and d is order of integration, HI_{t-1} , HI_{t-2} , …. ,HI_{t-p} were previous lags of headline inflation, ϵ_t is error term at time t, assumed to be independently and identically distributed, p and q are the order of AR and MA; $\beta_1, \beta_2, \ldots\ \beta_p$coefficient of autoregressive model and $\theta_1, \theta_2, \ldots\ and\ \theta_q$ coefficient of a moving average model. In summary, ARIMA model is a time series forecasting method that can handle both stationary and non-stationary data by using a combination of autoregression, integration, and moving average components.

3.4.2 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

In many real-world applications, time series data exhibit seasonal patterns that cannot be captured by ARIMA models. SARIMA models an extension of the ARIMA models, are often used for time series forecasting, particularly when data exhibit strong seasonality (Brockwell & Davis, 2016). By accounting for seasonality, SARIMA models can provide more accurate and reliable forecasts for time series data that exhibit seasonal patterns. Basic idea is to separate the model into seasonal and non-seasonal components. Non-seasonal component captures the short-term dynamics of the headline inflation series, while seasonal component captures the longer-term cyclical patterns of headline inflation.

In addition, by modeling seasonal components separately, we can account for any changes in seasonal patterns over time, such as shifts in timing or strength of seasonal effect due to factors such as changes in consumer behavior, economic conditions, or environmental factors.

Basic non-seasonal elements p, d, q corresponding to autoregressive (AR), integrated (I), moving averages (MA) respectively in ARIMA are kept as non-seasonal element. Seasonal elements P, D, Q corresponding to the same properties are introduced to model seasonal trends in time series. The general formula for a SARIMA (p,d,q)(P,D,Q)s model is expressed as follows:

$$\left(1 - \sum_{i=1}^{p} \phi_1 L^i\right)\left(1 - \sum_{i=1}^{P} \Phi_1 L\right)(1-L)^d (1-L^4)^D HI_t = \left(1 - \sum_{i=1}^{q} \theta_1 L^i\right)\left(1 - \sum_{i=1}^{Q} \Theta_1 L^{i4}\right)\varepsilon_t \qquad (3.3)$$

where, $L^i = HI_{t-i}$ is lag operator $\phi_1\ and\ \Phi_1$coefficients of non-seasonal and seasonal AR order p, and P respectively, $\theta_1 and\ \Theta_1$coefficients of non-seasonal and seasonal MA order p, and P respectively. D and d is orders of seasonal and non-seasonal difference of headline inflation, S is number period of seasonal in this case $S = 4$, since our data was quarterly data.

3.4.3 Testing Stationary: Unit Root Test

Before fitting a particular model to time series data, the series must be made stationary. Stationarity occurs in a time series when the mean and auto covariance of the series remain constant over the time series. Therefore, the stochastic process HI_t is said to be stationary if:

$$E(\mathrm{HI_t}) = \mu, constant\ for\ all\ value\ of\ t$$

$$Cov(HI_t,\ HI_{t-j}) = \rho_j = E[(\mathrm{HI_t} - \mu)(HI_{t-j} - \mu)T] = \rho - jT \qquad (3.4)$$

$for\ all\ t\ and\ j = 0,1,2, \ldots$

Where, t is time period from 2000 to 2023. To test for stationarity of a series several procedures have been developed. The most popular ones are the Augmented Dickey Fuller (ADF) test that was proposed by Dickey and Fuller (1979, 1981), and the Phillip-Perron (PP) test that was proposed by Phillips and Perron (1987, 1988) and graphical visualization.

3.4.4 Building ARIMA Models

To identify a perfect ARIMA model for a particular time series data, Box and Jenkins (1976) proposed a methodology that consists of four phases

Model Identification

The purpose of the identification stage is to determine the differencing quire achieving stationarity and also the order of both the seasonal and the non- seasonal AR and MA operators for the residual series. There are a number of identification methods proposed in the literature. The autocorrelations function (ACF) and the partial autocorrelation functions (PACF) are the two most useful tools in any attempt at time series model identification (Granger and Newbold, 1986).

Autocorrelation Function (ACF)

The sample ACF (ρ_k) measures the amount of linear dependence between observations in a time series that are separated by a lag k. An important tool that helps in detecting stationarity and identifying models for time series data is the autocorrelation coefficient. The autocorrelation coefficient is used to measure relation between the value of series in different time period, the function ACF is given by:

$$\rho_k = \frac{cov(HI_t, HI_{t+k})}{\sqrt{Var(HI_t)}\sqrt{Var(HI_{t+k}}} = \frac{\gamma_k}{\gamma_0} \quad where\ k = 1, 2, \dots, \frac{n}{2} \qquad (3.5)$$

Where, γ_0 is the variance of stationary series, which is fixed and equal for all different time period, $\hat{\gamma}_k$ is k^{th} covariance between HI_t series at time t and $t + k$, ρ_k is autocorrelation of headline inflation series.

Partial Autocorrelation Function (PACF)

A second important tool used in identifying models of time series is the partial autocorrelation. It can also be used for determining the possible order of seasonal autoregressive, non-seasonal autoregressive, moving average and seasonal moving average that should be incorporated in the model. The partial autocorrelation coefficient of order measures the correlation between values k periods apart when the effect of time lags $1,2, \dots\dots, k-1$ kept constant, where k is lag of headline inflation. When the partial autocorrelation coefficient is looked at as a function of k it is called the partial autocorrelation function (PACF) to estimate the partial autocorrelation of order k, fitting an autoregressive model of order AR (k).

$$\phi_k = \frac{\gamma_k - \sum_{i=1}^{k-1}\phi_i\gamma_{k-i}}{1-\sum_{i=1}^{k-1}\phi_i\gamma_i} \quad where\ k = 1, 2, \dots, \frac{n}{2} \qquad (3.6)$$

Where, ϕ_k is the partial autocorrelation at lag k, γ_k is the auto covariance at lag k and $\phi_i$represents the partial autocorrelation at lag i, where i ranges from 1 to k-1.

Model selection criteria

A model selection criterion is used to determine the approximating model from a set of estimated models. In time series analysis there may be several adequate models that can fit a given data set. The traditional model selection criteria such as the Akaike information criterion (AIC) proposed by Akaike (1974) and the Schwarz Bayesian information criterion (SBIC) proposed by Schwartz (1978) were employed to identify the optimal lag specification for the model.

Akaike's Information Criterion (AIC)

Akaike (1974) proposed the goodness of fit for ARMA (p, q) by balancing the error of the fit against the number of parameters in the model. It is defined by:

$$AIC = -2\ log\ L_k + 2k \qquad (3.7)$$

Where, L_k is the maximized log-likelihood and $k = 1 + p + q$, is the number of parameters in the model. The most appropriate model is the one which minimizes the AIC, and there is no requirement for the models to be nested.

Bayesian Information Criterion (BIC)

Schwarz (1978), introduced the BIC and is defined

$$BIC = -2 \, log \, L_k + k \, logT \qquad (3.8)$$

Different simulation studies have tended to identify that BIC does well at getting the correct order in large samples, whereas AIC tends to be superior in smaller samples where the relative number of parameters is large (McQuarrie and Tsai (1998)) as pointed out by Shumway and Stoffer (2010), and others method like RMSE, Log likelihood and MAE are applied to select the good model with minimum forecast error.

3.4.5 Parameters Estimation

Nonlinear estimation procedure is used to estimate ARIMA and its extension model parameters to maximize the likelihood function with respect to the parameters. That is, the maximum likelihood estimation method is used to test the parameters of the ARIMA and SARIMA models.

3.4.6 Diagnostic checking

Once the model is estimated, we run diagnostic tests. Checking the normality of residual by looking at histogram of residual, the independence of residual looking ACFs graphs, and compute the LB (Ljung-Box) test on the residual for test the magnitudes of the autocorrelation of the residuals.

Ljung-Box test of residual

. The Q-statistic is often used as a test of whether the residual is white noise. The Q-statistic at lag m is a test statistic for the null hypothesis that there is no autocorrelation up to order m and is computed as:

$$Q = T(t+2) \sum_{i=1}^{m} \frac{\widehat{\rho_J}}{T-j} \qquad (3.9)$$

Where, $\widehat{\rho_J}$ is the j^{th} lag of autocorrelation of residual and T is the number of observations. If the series is not based upon results of ARIMA estimation, then under the null hypothesis, Q is asymptotically distributed as chi-square with degrees of freedom equal to $(m - k)$, where k denotes the number of parameters.

Correlogram Squared Residuals

The existence of serial correlation can be determined using correlogram squared residuals. The hypotheses to be tested are as follows:

H_0: There is no serial correlation Vs H_1: There is serial correlation

Finally, after diagnosing the model, we forecasted inflation. However, forecasting inflation using traditional linear models like ARIMA or SARIMA is challenging for several reasons. First, forecasts become less reliable the further they are projected into the future, as unforeseen events can significantly impact accuracy (Makridakis, 1977; Baldwin, 2018; Cecchetti and Moessner, 2008; Blanchard and Gali, 2007). Second, when inflation is nonlinear and volatile, traditional linear models result in low forecast accuracy (Munim et al., 2019; Gustavo S. and Wagner P., 2023). Finally, the reliability of an ARIMA model often depends on the forecaster's expertise and experience, introducing subjectivity into the evaluation process, which can hinder forecast accuracy improvement (Kontopoulou, 2023).On other hand, machine learning models are can be handle effect of unforeseen events, non-linearity and it is not subjective to select parameters that more robust to noisy data and make decision without human intervention compared to traditional models (Bahrammirzae, 2010; Cui and Athey, 2022). Moreover it may struggle with highly erratic or irregular patterns in the time series.

3.5 Machine Learning Methods

Machine learning is the combination of automated computer algorithms with powerful statistical methods to learn (discover) hidden patterns in the datasets. In that sense, Statistical Learning Theory gives the statistical foundation of ML. Therefore, this paper is about Statistical Learning developments and not ML in general as we are going to focus on statistical models.

ML methods can be divided into three major groups:

The supervised learning is used when the dependent variables are clearly identified, and labeled, even if the specific relationships in the data are not known (e.g., linear regression, logistic regression). The goal approach is to find a model that relates a target variable with a set of independent variables similar to statistical model (Torgo, 2017; Ngai and Wu, Y., 2022). Unsupervised learning is classes of ML methods that uncover undetected patterns in a data set with no pre-existing labels. It used where there is no specific output defined beforehand, and the goal is to recognize data patterns and group data without having a target output(e.g., cluster analysis, principal components) (Jung *et al*., 2018; Ngai and Wu, Y., 2022). Reinforcement learning is a dynamic environment is set in order to interact with the learning algorithm with the goal to provide a valuation about the response of the system. It does so by exploration and exploitation of knowledge it learns by repeated trials of maximizing the reward (Ngai, and Wu, Y., 2022).

Supervised learning is the process of algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. It is the machine learning task of learning a function that maps an input to an output based on example input-output pairs (Asongo, *et al,* 2021). Therefore, this thesis was concerned with this method since we have identified dependent variable and training the model with predefined training dataset.

The supervised ML methods presented here can be roughly divided in two groups. The first one is penalized linear models or shrinkage models. Such methods date back at least to Tikhonov (1943). In Statistics and Econometrics, regularized estimators gained attention after the seminal papers by Willard James and Charles Stein who popularized the bias variance tradeoff in statistical estimation (Stein, 1956; James and Stein, 1961). We consider the Ridge Regression estimator put forward by Hoerl and Kennard (1970) and the Least Absolute Shrinkage and Selection (LASSO) estimator of Tibshirani (1996). We also include a combination of those two penalties models which is called Elastic Net model.

The second group of machine learning techniques focuses on nonlinear models such as ensemble methods (e.g., Random Forest) and ANN (Multiple- Layers Perceptron, Neural Network Autoregressive) models. However, this study did not consider advanced deep learning techniques like Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNNs) which are specifically designed to handle sequential data. While LSTMs and RNNs offer advantages, they are also more complex, challenging to train, and require high-dimensional data compared to ANN models (Kousta, 2020). Furthermore, the NARX model was discovered to be computationally more robust than the NAR, RNN, Extreme gradient boosting models (Ruiz *et al.*, 2016).

3.5.1 Penalized Linear Models (shrinkage)

In the world of linear regression models, Ridge and LASSO regression stand out as two fundamental techniques, both designed to enhance the prediction accuracy and interpretability of the models. This approach used regularization, which is a method to prevent overfitting by adding a penalty to the loss function.

Given a sample with T realizations of the random vector $(\mathrm{Y_t}, \mathrm{Z_t'})'$ the goal is to predict $\mathrm{T+h}$ for horizons $\mathrm{h} = \mathrm{H}$, which sample size testing data set equal to 20 where Y_t is HI_t and Z_t' random vector of covariates. For (usually predetermined) integers $p \geq 1$ and $r \geq 0$ define the *n*-dimensional vector of predictors $X_t = \{Y_{t-1}; \dots; Y_{t-p}; Z_t'; \dots Z_{t-r}'\}$ where, $n = p + d(r + 1)$, where p is number of lags of HI_t, d is numbers of covariate, r is number of lags of each covariate and consider the following direct forecasting model.

$$HI_{t+h} = f_h(X_t) + \varepsilon_{t+h}; \qquad h = 1, 2, \dots, 20; \ t = 1, \dots . \ T \qquad (3.10)$$

Where, $f_h: R^n \to R$ is an unknown (measurable) function and $\varepsilon_{t+h} = HI_{t+h} - f_h(X_t)$ is error term assumed to have zero mean and finite variance. The f_h could be the conditional expectation function,

$f_h(x) = E(Y_t + h|X_t = x)$, or simply the best linear projection of HI_{t+h} onto the space spanned by X_t. We consider the family of linear models where $f(x) = \beta_0' X$ in equation (3.13) for a vector on known parameters $\beta_0 \in R^n$. Notice that we drop the subscript h for clarity. However, the models as well as the parameter β_0 have to be understood for particular value of the forecasting horizon h, is 20. These models contemplate a series of well-known specifications in time series analysis (Hamilton, 1994).

In particular, equation (3.10) becomes

$$HI_{t+h} = \beta_0' X_t + \varepsilon_{t+h}; \qquad h = 1; \dots.; H; \ t = 1; : \dots.; T \tag{3.11}$$

where, under squared loss, β_0 is identified by the best linear projection of HI_{t+h} onto X_t which is well defined whenever $E({X_0}' X_t)$ is non-singular. In that case, ε_{t+h} is orthogonal to X_t by construction and this property is exploited to derive estimation procedures such as the Ordinary Least Squares (OLS).

In penalized regressions the estimator $\hat{\beta}$ for the unknown parameter vector β_0 minimizes the Lagrangian form or loss function

$$Q(\beta) = \sum_{t=1}^{T-h} (HI_{t+h} - \beta' X_t)^2 + \rho(\beta)$$

$$= \|HI - X\beta\|_2^2 + \rho(\beta) \tag{3.12}$$

Where $HI = (HI_{h+1}; \dots HI_T)', X := (X_1, \dots . X_{T-h})' and \ p(\beta) : p(\beta; \lambda; \gamma; Z) \geq 0$ is a penalty function that depends on a tuning parameter $\lambda \geq 0$, that controls the trade-off between the goodness of fit and the regularization term. Naturally, the estimator $\hat{\beta}$ also depends on the choice of λ and α.

In the presence of high variance, β_{ols} become unstable and have a tendency to over-fit on the training data resulting in poor out of sample predictions (Friedman, and Tibshirani, 2017). To reduce the variance of β_{ols}estimates albeit at the cost of having a positive bias and thereby improving the accuracy of the out of sample predictions, the machine learning literature proposes many penalized regression/shrinkage methods. Depending on the type of penalty used, there are different kinds of shrinkage methods. Here we consider three shrinkage methods.

3.5.1.1 Ridge Regression Model

The ridge regression was proposed by Hoerl and Kennard (1970) as a way to fight highly correlated regressors and stabilize the solution of the linear regression problem. The idea was to introduce a small bias but, in turn, reduce the variance of the estimator. The ridge regression is also known as a particular case of Tikhonov Regularization (Tikhonov, 1943, 1963; Tikhonov and Arsenin, 1977), in which the scale matrix is diagonal with identical entries.

Ridge regression is one of the most robust versions of linear regression in which a small amount of bias is introduced so that we can get better long term predictions. The amount of bias added to the model is known as Ridge Regression penalty. We can compute this penalty term by multiplying with the lambda to the squared weight of each individual features. It is also called as l_2 regularization.The equation for Ridge regression given by:

$$Q(\beta) = Min\left(\sum_{t=1}^{T-h}(HI_{t+h} - \beta' X_t)^2 + \lambda\sum_{i=1}^{n}(\beta_i)^2\right) \tag{3.14}$$

Where, λ is shrinkage parameter of Ridge regression and β_i is coefficient obtained from Ridge regression which not shrunk exactly zero. Due to this Ridge regression has the advantage to compute analytic solution, where the coefficients associated with the least relevant predictors are shrunk towards zero, but never reaching exactly zero. We note that despite the fact that Ridge Regression shrinks the parameter estimates to zero compared to the least squares estimator in the sense that $\left|\left|\hat{\beta}_{ridge}(\lambda)\right|\right|^2 < \left|\left|\hat{\beta}_{OLS}\right|\right|^2 for\ \lambda > 0,\ where\ \hat{\beta}_{ridge}(\lambda)$ is coefficients of covariates obtained from Ridge regression which not *exactly* shrink to zero and it is thus not directly useful for variable selection and λ is shrinkage parameters and $\hat{\beta}_{OLS}$ is coefficients of covariates obtained from Ordinary Least Square (OLS) estimation. Therefore, it cannot be used for selecting predictors, unless some truncation scheme is employed.

3.5.1.2 LASSO Regression Model

The LASSO was proposed by Tibshirani (1996) and Chen et al. (2001) as a method to regularize and perform variable selection at the same time. LASSO is one of the most popular regularization methods applied in both high and low data dimensional. It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights. Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0. It is also called as l_1 regularization. The equation for Lasso regression given by:

$$Q(\beta) = Min\left(\sum_{t=1}^{T-h}(HI_{t+h} - \beta' X_t)^2 + \lambda\sum_{i=1}^{n}|\beta_i|\right) \tag{3.15}$$

Where, λ is shrinkage parameter of LASSO and β_i is coefficient obtained from LASSO model which almost shrunk to zero. The solution of the LASSO is efficiently calculated by coordinate descent algorithms (Hastie et al., 2015), LASSO is not feasible to test combinations or models. Despite attractive properties, there are still limitations to the LASSO. A large number of alternative penalties have been proposed to keep its desired properties whilst overcoming its limitations. The important thing to note is that the LASSO performs estimation and variable selection in one step. This is in stark contrast to traditional procedures which would estimate β_0 by, say, least squares then decide which entries of β_0 are (non)-zero by means of hypothesis tests. However, the final model one arrives at by such a testing procedure depends heavily on the order in which such tests are carried out. The Lasso is only variable selection consistent under rather precise assumptions, which are rarely satisfied for time series, (Zou, 2006; Zhao and Yu, 2006)

3.5.1.3 Elastic Net Model

The Elastic Net model was proposed by Zou and Hastie (2005) as a way of combining strengths of LASSO and ridge regression. While the ℓ_1 part of the method performs variable selection, the ℓ_2 part stabilizes the solution. This conclusion is even more accentuated when correlations among predictors become high. As a consequence, there is a significant improvement in prediction accuracy over the LASSO (Zou and Zhang, 2009).

$$Q(\beta) = Min\left(\sum_{t=1}^{T-h}(HI_{t+h} - \beta' X_t)^2 + \lambda\left[\sum_{i=1}^{n}\alpha|\beta_i| + (1-\alpha)\sum_{i=1}^{n}(\beta_i)^2\right]\right) \tag{3.16}$$

where $\alpha \in [0,1], \beta_i, (\beta_i)^2$ is coefficient of Elastic Net obtained from LASSO and Ridge regression respectively and λ is shrinkage parameter. If $\alpha = 0$ then a ridge regression model is fitted, and if $\alpha = 1$ then a LASSO model is fitted. It also determines penalty parameter λ which used to shrink coefficients to zero. The Elastic Net has both the LASSO and ridge regression as special cases. Just like in the LASSO regression, the solution to the Elastic Net problem is efficiently calculated by coordinate descent algorithms.

Theoretical properties of penalized models

In this section we give an overview of the theoretical properties of penalized regression estimators previously discussed. Most results in high-dimensional time series estimation focus on model selection consistency, oracle property and oracle bounds, for both the finite dimension (*n* fixed, but possibly larger than *T*) and high-dimension (*n* increases with *T*, usually faster). More precisely, suppose there is a population, parameter vector β_0 that minimizes equation (3.10) over repeated samples. Suppose this parameter is sparse in a sense that only components indexed by $S_0 \subset \{1, 2, \dots, n\}$ are non-null. Let $\hat{S}_0 = \{j\colon \hat{\beta}_j \neq 0\}$. We say a method is ***model selection consistent*** if the index of non-zero estimated components converges to S_0 in probability.

$$P(\hat{S}_0 = S_0) \to 1, T \to \infty$$

Consistency can also be stated in terms of how close the estimator is to the true parameter for a given norm. We say that the estimation method is L^q-consistent if for every $\varepsilon > 0$

$$P\left(\|\hat{\beta}_0\|_q > \varepsilon\right) \to 0, T \to \infty$$

We say a penalized estimator has the oracle property if its asymptotic distribution is the same as the un-penalized one only considering the S_0 regressors. Finally, oracle risk bounds are finite sample bounds on the estimation error of $\hat{\beta}$ that hold with high probability. These bounds require relatively strong conditions on the curvature of objective function, which translates into a bound on the minimum restricted eigenvalue of the covariance matrix among predictors for linear models and a rate condition on λ that involves the number of non-zero parameters $|S_0|$.

3.5.1.4 Selection of the Penalty Parameters

The estimation of linear models by the penalized regression methods involves the choice of the penalty (tuning) parameter(α, λ). In general our feeling is that specifying a value of the penalty parameter(s) with theoretical performance guarantees (e.g., the oracle property or oracle inequalities) is still an open problem for most time series models. The penalty parameter λ is directly related to the number of variables in the model and thus is a key quantity for correct model selection.

Cross-Validation

One of the methods that is most used for model (variable) selection is cross validation (CV). In the context of penalized regressions, CV methods have been used to select the penalty parameters. The idea of CV methods is to split the sample into two disjoint subsets: the training set ("in-sample") and the validation set ("out-of-sample") or test set. The parameters of the model are estimated using solely the training set and the performance of the model using the estimated parameters is tested on the test set. The loss often used for cross-validation is the mean squared-error (MSE). The goal is to produce the so-called "cross validation curve", which is built by computing the MSE as a function of the tuning parameter chosen over a pre-selected grid.

Let Λ *and* A be the sets potential values of λ *and* α. Furthermore, let $V \subseteq \{1, \ldots, T\}$ be the indices of the observations in the validation set and $T \subseteq \{1, \ldots, T\}$ be the indices of the observations in the training set. Often, but not always, $T := V^c$. Let $\hat{\beta}_T(\lambda, \alpha)$ be the parameter estimate based on the training data T using the tuning parameters$(\lambda, \alpha) \in \Lambda \times A$. For each (λ, α) denotes the corresponding prediction error over the validation set V.

$$CV\ (\lambda, \alpha, V) = \sum_{t \epsilon V} \left(HI_t - X_t' \hat{\beta}_T(\lambda, \alpha) \right)^2 \tag{3.17}$$

Let $V = \{V_1, \ldots, V_B\}$ be a user specified collection of validation sets (with corresponding training sets $\{T_1, \ldots, T_B\}$), which we will talk more about shortly. The cross validation error for the parameter combination(λ, α) is then calculated as

$$CV\ (\lambda, \alpha) = \sum_{i=1}^{B} CV\ (\lambda, \alpha, Vi) \tag{3.18}$$

Where $\widehat{\beta}$ estimated coefficient of covariates and $(\hat{\lambda}, \hat{\alpha})$ selected tuning parameters at minimum MSE. The final parameter estimate is then found as $\widehat{\beta}\ (\hat{\lambda}, \hat{\alpha})$ based on all observations 1, . . . , T .

3.5.2 Nonlinear Machine Learning Models

Non-linear machine learning can handle non-linearity natures of inflation as we discussed in previous chapter. It works in a non-linear fashion, having a potential advantage in analysis of variables with complex correlations compared to linear models (Aronsson, L., *et al.*, 2021).

3.5.2.1 Random Forest Model

The theory for random forest (RF) models has been developed only for independent and identically distributed random variables. For instance, Scornet *et al.*, (2015) proves consistency of the RF approximation to the unknown function $f_h(X_t)$. Moreover, recently, Wager and Athey (2018) proved consistency and asymptotic normality of the RF estimator. RF consists of decision trees which are simple, fast, non-linear, and nonparametric machine learning approaches for both regression and classification purposes. As an ensemble method, Random forest was proposed by Breiman (2001) to alleviate the overfitting problem that regression trees frequently face and to lower the large variance of forecasts. Training data are randomly sampled with replacement to increase diversity between forecasts and a random subset of features is selected for each tree to make them more unrelated to each other. The final decision of the RF model is generally found by taking the average of the individual trees for a regression problem or by using majority voting for classification.

Given a dependent variable HI_{t+h} a set of predictors X_t and a number of terminal nodes K, the splits are determined to minimize the sum of squared errors of the following regression model,

$$\widehat{HI}_{t+h}|t = \sum_{k=l}^{K} C_k I_k(X_t, \theta_k) \quad (3.19)$$

C_k is a constant which estimated as the sample average of realizations of HI_{t+h} that "fall" within region R_k, θ_k the set of parameters define the kth region and $I_k(X_t, \theta_k)$ is indicators function which

$$I_k(X_t, \theta_k) = \begin{cases} 1 \; if \; x_t \in R_k(\theta_k) \\ 0, \quad otherwise \end{cases} \quad (3.20)$$

where $R_k(\theta_k)$ is the kth region. The randomly constructed regression trees are aggregated based on bootstrapping, where B = 500 bootstrap samples are used, for which a subset of original regressors is chosen at random. For each sample $b, b = 1; \dots; B$, a tree with K_b regions are estimated for a randomly selected subset of the original regressors. K_b is determined in order to leave a minimum number of observations in each region. For every bootstrap sample eventually a tree with K_b regions will exist. Finally, the forecasts are then given by

$$\widehat{HI}_{t+h}|t = \frac{1}{B}\sum_{b=1}^{B}\left\{\sum_{k=l}^{K}\hat{C}_{k,b} I_{k,b}\left(X_t;\ \hat{\theta}_{k,b}\right)\right\} \quad (3.21)$$

Where B is sample size of bootstrap sample, $\hat{C}_{k,b}$ and $\hat{\theta}_{k,b}$ is a constant and estimated parameter kth region of bootstrap sample respectively. The current study of inflation forecasting using ML methods, data were made normalized. In traditional econometric models, this is done to avoid so-called 'multicollinearity' problem. However, the results of many ML algorithms, including the RF model are not subject to this problem. It is important to note that, in all specifications, the RF models were trained and used to obtain inflation forecasts (standardized or normalized), but not forecasts of accumulated CPI values.

Regarding the application of Random Forests for time series forecasting (Kane et al., 2014) present an application of Random Forest for forecasting of avian influenza H5N1 outbreaks and a comparison to ARIMA. The authors founded that Random Forest provides enhanced predictive ability over the ARIMA model for the prediction of infectious disease outbreaks.

3.5.2.2 Artificial Neural Network (ANN) Model

Artificial Neural Network (ANN) technique is attempts to mimic the working nature of the human brain to extract the hidden pattern between a set of inputs and outputs (Haykin, 1999). ANN approach is an efficient forecasting tool (Aladag, C.H, 2011). These features are generating and exploring new knowledge by learning (Kourentzes, *et al* 2014). The essential elements that determine the ANN are architecture structure and learning algorithm (Kuleyin *et al.* 2022). The architecture is determined by deciding the number of layers and number of neuron nodes in each layer and there is no general rule for determining the best architecture (C.H. Aladag, *et.al.*, 2009). The three essential features of an artificial neural network (ANN) are the basic processing elements referred to as neurons or nodes; the network architecture describing the connections between nodes; and the training algorithm used to find values of the network parameters for performing a particular task. An ANN consists of elementary processing elements (neurons), organized in layers (See Figure 2.2). The layers between the input and the output layers are called "hidden". The number of input units is determined by the application.

Feed-forward back propagation network is one of the most neural networks architectures that widely used for forecasting due to its simple usage and success (Z.X. Guo, 2012). Back Propagation algorithm is one of the most used learning algorithms which update the weights based on the difference between the output value of the ANN and the desired real value. In the forecasting, the inputs are the past observations and the output is the predicted value.

A neuron is simply a linear combination of inputs, plus a constant term (called a *bias*), and transformed through a function (called an activation function), $i.e\ f\ (XW+\ \alpha),$ where $\boldsymbol{x}$ is an *n*-length vector of inputs, θ is a corresponding vector of weights, and α is a scalar bias term.

The activation function transforms the weighted sum of the inputs into the final output. There are several commonly used activation functions, including the linear function, the sigmoid (or logistic) function, the hyperbolic tangent ***(tanh)*** function, and the Rectified Linear Unit (***ReLU***) function. The selection of activation functions is critical in the architecture of an ANN. It is worth noting that if all of the activation functions were linear, the ANN would be reduced to a linear regression model, whereas non-linear functions such as the ***sigmoid*** and ***tanh*** functions enable ANNs to discover complex functional relationships that may exist between targets and features (Sharma *et al.*, 2022)

Neurons are typically stacked into layers. Layers can have various forms, but the most simple is called a dense, or fully connected layer. For a layer with p neurons, let $B\ =\ (\beta_1 \ldots \beta_p)$ so that has β the dimensions *(n × p)*, and let $\alpha\ =\ (\alpha_1 \ldots \alpha_p)$. The matrix $\boldsymbol{B}$ supplies weights for each term in the input vector to each of the p neurons while $\boldsymbol{\alpha}$ supplies the bias for each neuron. Given an *n*-length input vector $\boldsymbol{X}$, we can write a dense layer with p neurons as,

$$g(x) = f(XB + \alpha)$$

$$= \begin{bmatrix} f\ (\boldsymbol{XB_1} + \boldsymbol{\alpha_1}) \\ f\ (\boldsymbol{XB_2} + \boldsymbol{\alpha_2}) \\ . \\ . \\ f\left(\boldsymbol{XB_p} + \boldsymbol{\alpha_p}\right) \end{bmatrix}^T \tag{3.22}$$

Where $\beta_1 \dots \beta_p$ *weight of the covariate in each neuron*, $\alpha_1 \dots \alpha_p$ *are bias which known as error*. The layer described in equation **(3.22)** can also accept higher-order input such as an $m \times n$ matrix of several observations, $X = \{x_1, x_2, \dots x_m)$ in which case,

$$g(x) = f\,(XB + \alpha)$$

$$= \begin{bmatrix} f\,(\boldsymbol{x_1 B_1} + \boldsymbol{\alpha_1}) & \dots & f\,(\boldsymbol{x_1 B_p} + \boldsymbol{\alpha_p}) \\ \vdots & & \vdots \\ f\,(\boldsymbol{x_m B_1} + \boldsymbol{\alpha_1}) & \dots & f\,(\boldsymbol{x_m B_p} + \boldsymbol{\alpha_p}) \end{bmatrix}$$

A fully connected, feed forward network with K layers is formed by connecting dense layers together so that the output of the preceding layer serves as the input for the current layer. Let k index a given layer, then the output of the k-th layer is $g_k(X) = f_k(g_{k-1} - (X)B_k + \alpha_k)$ where the parameters of the network are all elements B_k, α_k $for\ k \in (1 \dots . K)$. For simplicity, denote these parameters by $\boldsymbol{\theta}$, where $\theta_k = (B_k, \alpha_k)$. Further, for simplicity, denote the final output of the network $G(X;\ \theta) = gK(X)$.

For a single-hidden-layer architecture, the number of hidden units A indexes the different classes of ANN models (S_θ) since it is an unambiguous descriptor of the dimension k of the parameter vector $(T = (n+1)J_T + (J_T + 1)$. A class of neural models is specified by

$$S_\theta = \{g_\theta(x, \beta), x \in M^m\ \beta \in B\} \tag{3.23}$$

where $g_\theta(x, \beta)$,is a nonlinear function of x with β being its parameter vector, A is the number of hidden .

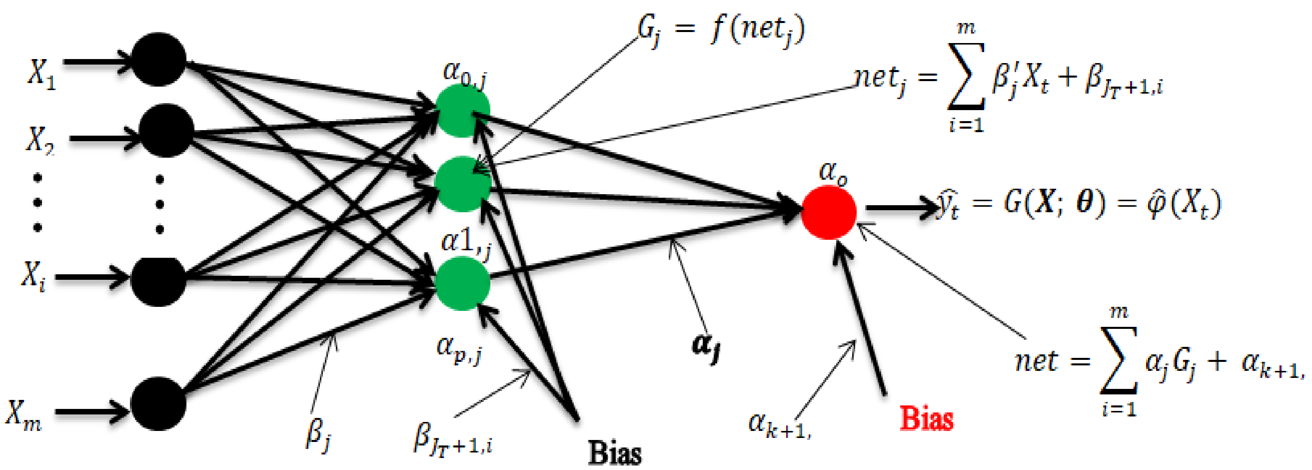


Figure 3.1: A multilayer feed forward ANN for approximating an unknown function $\hat{\varphi}\ (X_t)$ adopted from (H. Allende, et al., 2002)

3.5.2.2.1 Multi-Layer Perceptron (MLP*)*

Early exploration of neural networks, MLP as a first step because they are simpler to understand and implement compared to more complex architectures like LSTMs and it is the most popular type of NN (Moghaddamnia *et al.*, 1994) developed independently by Parker (1985) and Rumelhart et al. (1986). An MLP is a feed forward NN with multiple layers of fully connected neurons. A typical MLP network with one hidden layer and one output is shown in Figure 3.2. In this type of network, inputs are fed from the input layer and propagated forward to all neurons in the hidden layer (hence the term feed forward). Multi-layer ANN model where the approximating function $gK(X_t)$) is defined as

$$G(\boldsymbol{X};\ \boldsymbol{\theta}) =\ gK(\boldsymbol{X_t}) = \alpha_0 + \sum_{j=1}^{J_T} a_j f\left(\beta_j' X_t + \alpha_{0,j}\right) \tag{3.24}$$

$$=\ \alpha_0 + \sum_{j=1}^{J_T} a_j f\left(\beta_j' X_t\right)$$

$$= \mathrm{F}(XB)A$$

Where F activation function, B is matrix of parameters, A is vector of bias. In the above model, ${X_t}^T = \{1, X_t\}$, $f_j(.)$ is activation functions and the parameters are called weights and the parameter vector to be estimated is given by $\left(\alpha_0, \ldots.\alpha_k,\ \beta_1 \ldots.,\ \beta_{J_T},\ \beta_{0,1}, \ldots..,\beta_{0,J_T}\ \right)$. Each hidden layer unit receives the inputs, multiplies them by their corresponding weights, and adds them all together with a bias (α). That is, the hidden layer unit (j) calculates:

$$net_j = \sum_{j=1}^{J_T} \beta_j' X_i + \alpha_{0,j}, i = 1, 2, \ldots., n, and\ j = 1,2, \ldots.p \tag{3.25}$$

where (X_t) is the i^th^ independent variables at period t and B $= \left(\ \beta_1 \ldots.,\ \beta_{J_T},\ \beta_{0,1}, \ldots..,\beta_{0,J_T}\right)^T$ is a weight connecting the i^th^ independent variables to the hidden unit (j).

The output of the hidden layer unit (j), (G_j) is a transformation of the net, as follows:

$$G_j =\ f\left(net_j\right) \tag{3.26}$$

Where, $f(.)$is an activation function mostly the non-linear tanh function$i.e.\left(f(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}}\right)$. The output units receive the outputs of the hidden layer units (G_j) as their input. The process in the output layer units is the same as that of the hidden layer units. That is, the output unit calculates net, the sum of the product of the inputs and weights, and bias as follows:

$$net_h = \sum_{i=1}^{n} \alpha_j G_j + \alpha_{0,} \tag{3.27}$$

Where, "net_h" is the net value for the output unit (h), (α_j) is the weight connecting the hidden unit (j) to the output unit (h), and (G_j) is the output of the hidden unit (j), which is input for the output unit (h). The output unit then applies a transfer function, (*f*), to the (net_h) which is linear function connect with target variable headline inflation.

3.5.2.2.2 Neural network autoregression (NNAR)

NNAR is simple mathematical models of the brain form the basis of artificial neural networks (ANN), which is used in forecasting inflation. Complex linear and nonlinear relationships between the response and its predictors are possible with their help. Lagged values of the dependent variable HI_t are used as inputs to the feed-forward neural network, which also has a single hidden layer of k size nodes. The model is valid for a wide variety of fitted repetition networks, all of which have initial weights chosen at random. When making predictions, these are then averaged. The network is optimized for making predictions in a single step. In this research, we focus exclusively on feed-forward networks with a single hidden layer and we designate the number of lagged inputs and the number of hidden nodes in the network with the notation NNAR (p, k). Where p is number of lagged value of headline inflation used as input, k numbers of nodes in the hidden. More generally, with seasonal data, it is useful to also add the last observed values from the same season as inputs. $NNAR(p, P, k)m$ model has inputs

$$\left(HI_{t-1}, HI_{t-2}, \dots, HI_{t-p}, HI_{t-m}, HI_{t-2m}, HI_{t-Pm}\right)$$

and k neurons in the hidden layer. A NNAR(p,P,0)m model is equivalent to an SARIMA(p,0,0)(P,0,0)m model but without restrictions on the parameters that ensure stationary. The NNAR model is a feed forward neural network which involves a linear combination function and activation function. The formations of this function are defined as,

$$net_j = \sum_{i=1}^{n} w'_{ij} y_{ij} + \alpha_{0,j}, \qquad i = 1, 2, \dots, n,$$

Where, w_{ij} weight of the input variables, y_{ij} headline inflation $\alpha_{0,j}$ $bias\ or\ constant$ it is the same to equation (3.25)

Another model where NNARX, which stands for Nonlinear (Neural Network) Auto-Regressive model with eXogenous inputs, is a type of neural network model used for time series forecasting. The NNARX model is particularly powerful because it combines the ability to model complex, nonlinear relationships with the ability to incorporate both past values of the target variable and external (exogenous) variables that might influence future values (Khashei, M., and Bijari, M., 2011).

3.5.2.3 Training Artificial Neural Networks

Machine learning models require strong assumptions about the data generating process. For instances neural network models are quite sensitive to several properties of the data. When feeding a model with more than one feature, it is important that the features are at roughly similar scales (to within about an order of magnitude). In theory, a neural network should be able to adjust to inputs of differing scales. But in the initial iterations of training, larger-scaled inputs will dominate gradients and thus parameter adjustments. This can lead to premature saturation of the neurons or very slow model convergence. Pre-scaling the model inputs to have similar scales will alleviate this problem. Typical approaches include scaling inputs to normalization, and scaling inputs to the interval (0, 1) through the following affine transformation:

$$x^{*} = \frac{x - min(x)}{max(x) - min(x)}$$

Neural networks excel at generating predictions that generally lie within the boundaries of the training data. Out-of-bounds predictions are subject to more error. The training algorithm is to optimize the performance of the neural network by adjusting the weights to minimize the observed error. To enable the model to learn from its mistakes, a loss function is introduced. This loss function calculates the difference or error between the actual value and the predicted value at a given moment (Gurney, 2018).

3.5.3 Parameter Estimation

Many cases of machine learning, we are not directly calculating the optimal parameters by maximum likelihood estimation and OLS method. Instead, we are using some algorithm to approximate the optimal parameters, instead of calculating them directly. The algorithms we use here are called Optimizers. While using an algorithm to approximate the optimal parameters is simple, usually approximates well and applies to most machine learning algorithms that need to do parameter estimation. The common algorithm that we use to do the approximations to the optimal parameter is Stochastic Gradient Descent (SGD) for time series data.

Stochastic Gradient Descent (SGD)

For Stochastic Gradient Descent (SGD), we start form some imperfect parameters θ_k and update them by

$$\theta_k = \theta_k - \gamma \nabla_{\theta_k} L(y, \theta) = \theta_k - \frac{\gamma}{n} \sum_{i=1}^{n} \nabla L_i(\theta_k) \tag{3.28}$$

Where, γ controls the size of the update and is sometimes referred to as the learning rate. $L(\theta_k)$ is loss function , and $\nabla L(\theta_k)$ is the gradient between our current prediction and true label. n is the number of observations we consider to optimize the parameters in each iteration.

The computation of $\nabla_{\theta_k} L(y, \theta)$ is costly and increases with the size of ***X*** and ***HI***. To reduce this cost, the overall computation time needed a stochastic gradient descent (SGD) used. This is a modification of gradient descent in which updates to θ are calculated using only one observation at a time. We may infer that in each iteration, it is trying to approximate a little bit to the optimal value by shrinking the loss between current prediction and the true label, and after a number of iterations, we may get to a place where we are very close to the optimal parameter θ^*

The algorithm of SGD is basically running a for-loop with the above update operation until the loss between our prediction and the true label could be small enough. A later proposed version of SGD is SGD with momentum (SGDM), which the update becomes:

$$\Delta\theta = \alpha\Delta\theta - \gamma\nabla L(\theta)$$
$$\theta = \theta + \Delta\theta$$

where α is a forgetting factor between 0 and 1 or momentum, and γ is learning rate of algorithm. By doing this, the optimizer tends to remember the mainstream of the previous updating directions, and thus less affected by the random special cases in the data set.

Gradient Estimation

Each of the optimization routines described in the previous section rely upon the computation of $\nabla_{\theta_k} L(y;\ \theta)$ for all $\theta_k \in \theta$. This achieved through the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986), which is a generalization of the chain rule from calculus. Consider, for example, a network $G(X;\ \theta) = \mathrm{F}(XB)A$ with an accompanying loss

$$L(\boldsymbol{y};\ \boldsymbol{\theta}) = \frac{1}{2m} \|G(\mathrm{X};\ \theta) - \mathrm{y}\|_2^2 = \frac{1}{2m} \sum_{i=1}^{m} \|G(\mathrm{X};\ \theta)_i - \mathrm{y_i}\|_2^2 \tag{3.29}$$

$$\hat{\theta} = arg, min\|Y_i - \mathrm{F}(XB)A\|_2^2$$

The loss function gives us a measure of the accuracy with which an estimator $A\backslash$, fits the observed data but it does not account for the estimator's (model) complexity). Then

$$\nabla_{G(X;\,\theta)} L(y;\theta) \quad = \quad \frac{1}{m}(G(X;\,\theta) - y) \tag{3.30}$$

To derive $\nabla\theta_k\, L(y;\, \theta)$ we simply apply chain rule to the above equation, suppressing arguments to G and g for notational simplicity.

$$\nabla\theta_k\, L(y;\, \theta) = \frac{\partial G^T}{\partial \theta_k}\, \nabla_G L(y) \tag{3.31}$$

Where, $\frac{\partial G}{\partial \theta_k}$ is a generalized form of a Jacobian matrix, capable of representing higher order tensors which used to store and manipulate data. We can extend the application of backpropagation to calculate the gradient of the loss with respect to any of the set of parameters θ_k:

$$\nabla\theta_k\, L(y;\, \theta) = \frac{\partial L(y;\theta)}{\partial G_k}\, \frac{\partial g_K}{\partial g_{K-1}}\, \frac{\partial g_{K-1}}{\partial g_{K-2}}\, \ldots .\frac{\partial g_{k+1}}{\partial g_k}\, \frac{\partial g_k}{\partial \theta_k} \tag{3.32}$$

3.5.4 Model Selection and Diagnostic for Machine Learning Methods

In the hidden layer different numbers of neuron are used to choose the best architecture, the activation function used is the hyperbolic tangent function $f(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. One neuron is used in the output corresponding to the forecast, and it uses a linear activation function to obtain values in the real space. The forecasts were obtained using the data available (samples).

The weights (parameters) to be used in the NN model are estimated from the data by minimizing the mean squared error within- in sample and out sample forecasting, because some data may be lost differentiating. To train the network a backpropagation algorithm with momentum was used, which is an enhancement of the backpropagation algorithm. The network 'learns' by comparing the actual network output and the target; then it updates its weights by computing the first derivatives of the objective function, and uses momentum to avoid local minima. To choose the architecture of the model that best fits the data, one can use the mean squared error, but the larger the model is made (more neurons), the smaller becomes sum of residuals and residual standard deviation, and then model is selected. Finally, residual of the fitted model is diagnostically by LB, ACF residuals and histogram similar to traditional time series models.

3.6 Forecasting Strategy

There are two kinds of forecasts: in sample period forecasts (first quarter of 2000 to fourth quarter of

2023) and post-sample period forecasts (first quarter of 2019 to fourth quarter of 2023). The former are used to develop confidence in the model and the latter to generate genuine desired forecasts. In forecasting, the goal is to predict future values of a time series, HI_{t+h} , $h = 1, 2, \ldots$ based on the data collected to the present, $HI = \{HI_t, HI_{t-1}, \ldots, HI_1\}$. For ARIMA and SARIMA model, we assume HI_t is stationary and the model parameters are known.

3.7 Forecast Evaluation Methods

The forecast error is the difference between the actual inflation and the predicted inflation. There are several measures of forecast errors. The measure adopted in the study was used by Moshiri (1997) and Haider & Hanif (2007).

3.7.1 Root Mean Squared Error (RMSE)

RMSE is a measure of the size of the forecast error, that is, of a magnitude of typical mistakes made using a forecasting model (Stock & Watson, 2007). The RMSE is computed as the measure of the forecast accuracy for each model. The RMSE is calculated as follows

$$RMSE = \sqrt{\frac{1}{h}\sum_{i=1}^{h}\left(HI_t - \widehat{HI_{t+h}}\right)^2} \tag{3.33}$$

where HI_t is observed series of headline inflation (training or test data set) $\widehat{HI_{t+h}}$ predicted values of inflation (in-sample or out-sample). RMSE is commonly used for regression tasks, where the goal is to predict a continuous value. It is a popular choice because it is easy to interpret, as it is expressed in the same units as the predicted and true values (Ghasemi and Zahedi, 2010).

3.7.2 Mean Absolute Error (MAE)

MAE is another common performance measure for regression tasks, and is calculated as the average of the absolute differences between the predicted and true values. The MAE is defined as:

$$MAE = \frac{1}{h}\left|HI_t - \widehat{HI_{t+h}}\right| \tag{3.34}$$

MAE is also easy to interpret, as it is expressed in the same units as the predicted and true values. However, unlike RMSE, it is not sensitive to outliers in the data, as it only takes into account the magnitude of the error, rather than the squared error (Zhang and Yang, 2015)

In general, RMSE is a good choice when the goal is to minimize the prediction error, and the data is free of outliers. MAE is a good choice when the goal is to minimize the prediction error and the data contains outliers, or when the range of possible values is important to consider.

3.7.3 Mean Absolute Percentage Error (MAPE)

MAPE is the most common measure used to forecast error, probably because the variable's units are scaled to percentage units, which makes it easier to understand. According to Gartner (2018 Gartner Sales & Operations Planning Success Survey), the most popular evaluation metric for forecasts in Sales and Operations Planning is Mean Absolute Percentage Error (MAPE). It works best if there are no extremes to the data (and no zeros). It is often used as a loss function in time series forecasting and model evaluation. MAPE is defined as

$$MAPE = \frac{1}{h}\left|\frac{HI_t - \widehat{HI_{t+h}}}{HI_t}\right| \tag{3.35}$$

MAPE is expressed as a percentage, which is scale-independent and can be used for comparing forecasts on different scales (Kothari, J., and VB, A., 2024).

3.8 Software for Data Analysis

In our analysis of forecasting inflation, we utilized online Jupyter Notebook and RStudio 4.3.2. For time series forecasting, we utilized the forecast package, particularly the arima() function to model and forecast tseries() to set inflation as time series object. The glmnet package (Friedman, Hastie & Tibshirani, 2010) was used to implement the Lasso, Ridge and Elastic Net. As for the selection of $\lambda,$ glmnet package is implemented only by cross-validation. The function cv.glmnet() was used to perform cross-validation and also select the optimal tuning parameters, ensuring that our linear models were well-regularized to prevent over fitting. The glmnet() function was then used for modeling, providing a robust framework for implementing linear machine learning models. We also explored neural network models in R using the nnfor and neuralnet libraries. The mlp() function from the RSNNS package was used to construct Multi-Layer Perceptron (MLP) models, providing a flexible architecture for capturing complex nonlinear relationships in the data. Additionally, we used the nnetar() function from the forecast package and the neuralnet() function from the neuralnet package to construct neural network models tailored for time series data, facilitating the capture of intricate temporal patterns.

In Jupyter Notebook, integrated with the Python library scikit-learn (sklearn), algorithms for machine learning, such as Artificial Neural Networks (ANN) and Random Forest, are utilized in hyperparameter tuning and model selection. These algorithms aid in assessing the performance of models and selecting the best one for our dataset. Additionally, Jupyter Notebook supports parallel computing, enabling the acceleration of hyperparameter tuning by running multiple experiments simultaneously.

CHAPTER 4

4. RESULTS AND DISCUSSIONS

4.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) were involves visualizing the data through various statistical tools and graphical representations. Common techniques were descriptive statistics, correlation matrix and time series plots. The goal was to identify patterns, trends, seasonality, and potential outliers in the data.

Descriptive Statistics

The quarterly data of headline (year-on-year) inflation (in %) was obtained from the Central Statistical Service (CSS) covered period from 2000Q1 to 2023Q4. The descriptive statistics were used to describe the pattern and variation of the headline inflation series.

Table 4.1: Summary statistics for the quarter headline (year-on-year) inflation in percentage

Mean	Median	Maximum	Minimum	Std. dev	Skewness	Kurtosis	Jarque-Bera
14.966	11.750	61.833	-6.300	12.676	0.958	4.051	19.107(0.00)

Table 4.1 shows that the average value of the quarter headline inflation series was 14.966. The minimum and maximum quarter headline inflation series were (-6.300) and (61.833), respectively. There was also evidence of positive skewness. This implies that the series was positively skewed (non-symmetric). The kurtosis value (4.0512) was greater than three, which implies that the distribution of the headline inflation series was leptokurtic. This means series was not normally distributed. The empirical results were consistent with the Jarque-Bera (JB) test, which assessed whether the given series was normally distributed or not. The null hypothesis stated that the headline inflation series was normally distributed. The results of the JB test showed that the null hypothesis was rejected, which indicated that the observed series was not normally distributed.

Time series plot: Trend Analysis

A time series plot is a graph that shows information gathered over time from any process. The quarterly headline inflation series, from first quarter of 2000 to fourth quarter of 2023 are presented in Figure 4.1. The plot revealed that trend was decline from first quarter of 2000 to first quarter of 2002 below zero, which indicated that headline inflation during this period was low. After the series trend was rapidly increased up the first quarter of 2003 year, then after declined and turned to gradually increase up first quarter of 2008. Notably, the series reached a peak in 2008Q4, followed by a sharp decline in 2009. This coincided with the period of global economic risk, suggesting a potential structural break (nonlinearity) within the series. Consequently, traditional forecasting models might not have been reliable for this data due to the presence of structural break (sudden shift). The headline inflation series were gradually increased from 2015 to end of 2022 then turned to decline up end of study time (2023Q4). In generally, this trend was characterized by unconditional mean and variance over time.

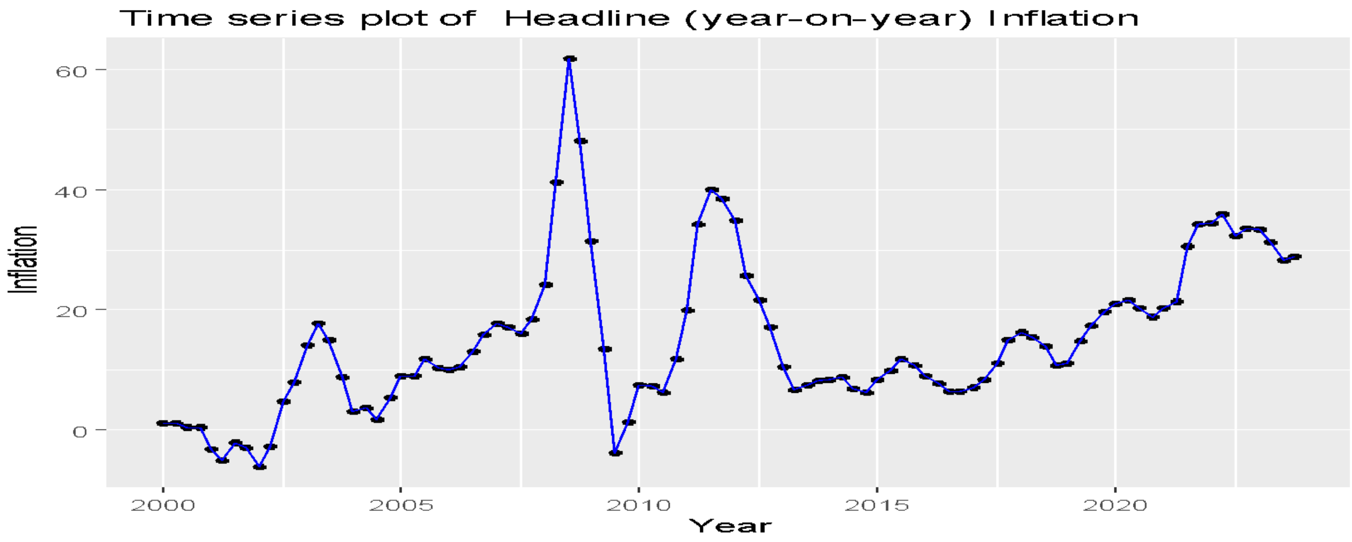


Figure 4.1: Trend plots of quarter data of headline inflation

The seasonally adjusted data allow for more meaningful comparisons of economic conditions from period to period. As we see from Figure 4.1 trend plot, the series was not stationary (the mean and the variance of series change over time) and had seasonal pattern. The quarterly headline inflation series has been seasonal adjusted to eliminate the effect of seasonal and time influences. A Figure 4.2 show there was no more difference between seasonal adjusted series and original series (Figure 4.1); this implied that there was no significance of seasonal effect.

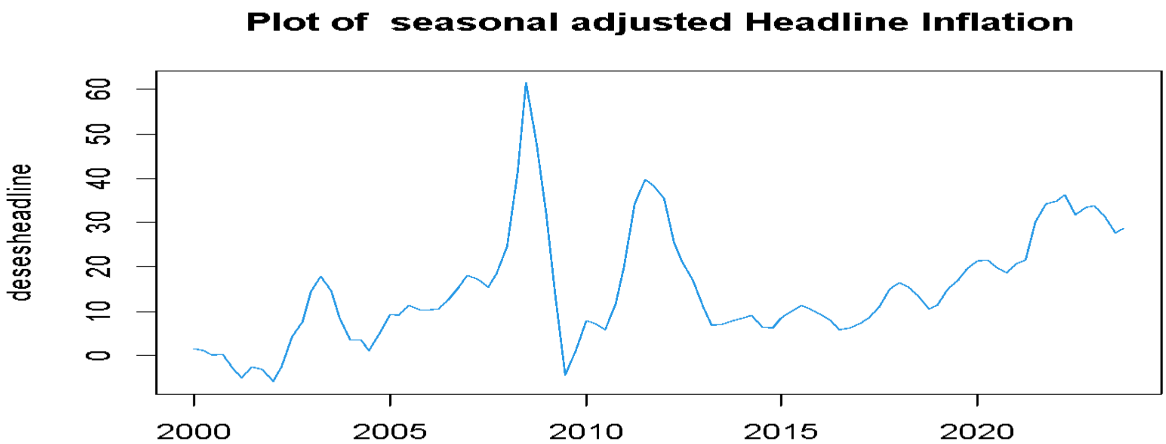


Figure 4.2: Trend plots of quarter data of seasonal adjusted headline inflation

Correlation Matrix

A correlation matrix was used to investigate covariance among predictor variables. Accordingly, many covariates such as broad money, gross national saving, lending interest rate, net demand deposit, agricultural production price, world oil price, expenditure, narrow money, and budget deficit were highly correlated with above 0.80 correlation coefficients; this indicated the existence of multicollinearity (See Appendix B).

4.2 Data Preprocess

Before developing the model, the data was preprocessed to make the series stationary (for time series analysis), normalized (for the machine learning model), and split into dataset.

4.2.1 Test for stationarity

The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are used to test the null hypothesis of a unit root against the alternative hypothesis of not unit root. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the null hypothesis that series is level or trend stationary. Other method stationary tests were ACF and PACF graphs. Figure 4.3 illustrate the ACF plot of original data had some seasonal pattern (sin wave) and exponentially decay that shows the series has a trend but the PACF shows the series is stationary accept at first and second lag.

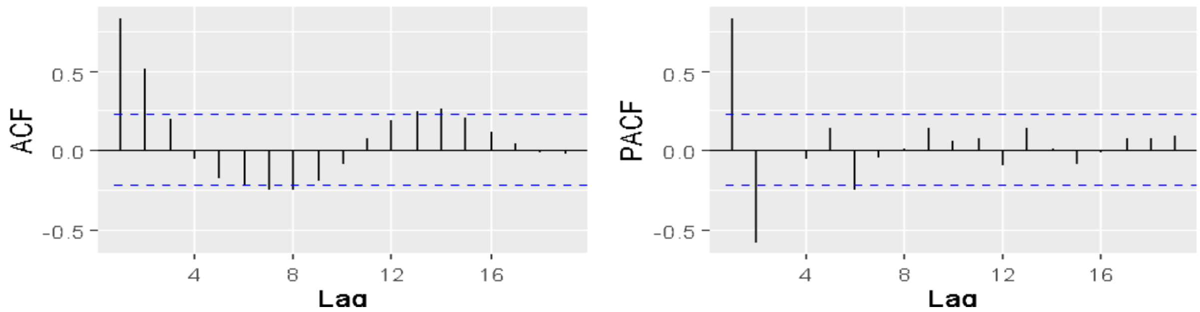


Figure 4.3: ACF (left hand) and PACF (right hand) plot of seasonal adjusted series

The formal test of stationarity where addressed in Table 4.2 instead of visualized graph which make ambiguity of whether the series is stationary or not. The KPSS test statistic (p-value=0.02) indicated that the series was not stationary at its original level, as the p-value less than 5% significance level. This result was confirmed by the ADF test (p-value=0.288), which also suggested the presence of a unit root (non-stationarity) at the level based on result in Table 4.2. To address this issue, the analysis proceeded by differencing the data (taking the first order difference). This removed the trend from the series.

Table 4.2: Unit Root test of headline Inflation test.

Test	At level			At first differenced		
	Statistics	Critical value at 5%	P-values	Statistics	Critical value at 5%	P-values
ADF	-0.697	-2.452	0.288	-19.398	-7.329	0.01
Phillips-Perron	-23.592	-22.281	0.050	-46.178	-10.932	0.01
KPSS	3.618	2.508	0.020	0.926	3.012	0.1

The subsequent KPSS test on the differenced data confirmed stationarity (p-value=0.1), since it was greater than the 5% significance level. Similarly, the ADF and PP tests on the differenced data yielded p-values of 0.01 and 0.01 respectively, both less than 5% significance level and supporting the conclusion of stationarity. However the unit root test shows first differenced data was stationary, the series had constant mean and conditional variance (See Figure 4.4).

Figure 4.4: Plot of first differenced of headline inflation

Figure 4.4 revealed that the differenced data exhibited volatility clustering. This means large changes were more likely to be followed by other large changes, and small changes by other small changes, regardless of positive or negative direction. This behavior suggested that volatility models like ARCH and GARCH might be suitable for this data. However, these models were not explored in this study due to their reported low accuracy in forecasting inflation.

4.2.2 Normalization of the Data

For machine learning models the data should be normally standardized (scaling data) to forecast inflation. Normalization is essential for ANN model because it helps to prevent any single feature from dominating the analysis. This can be done by scaling the data to a certain range, for example between -1 and 1.

4.2.3 Splitting of the Data

Before model analysis the transformed data divided into training and test sets in the ratios of 80:20 percent. The training data set used to train time series models and machine learning models covered the period from first quarter of 2000 and fourth quarter of 2018. The test (validation) data set used as our sample forecast covered the period from first quarter of 2019 and fourth quarter of 2023. Training data used to train and evaluate the model's performance during the development phase and to assess the model's accuracy in making predictions on new, unseen data. The training set is the portion of the data used to train the machine learning model. The model is trained on this data to learn the underlying patterns and relationships between the input features and output targets. After the model has been trained and evaluated on the training, it is then evaluated on the test set. The test set provides an unbiased evaluation of the model's performance on new data that it has not been trained on and assess the model's accuracy in making future values of headline inflation.

4.3 ARIMA Model (Benchmark Model)

Autoregressive integrated moving average (ARIMA) models is univariate time series and the most widely used in forecasting. Under time series analysis consideration after checked the stationarity of the series we built the model as follows.

4.3.1 Selection Order of ARIMA Models

An order of ARIMA (p, d, q) selected based on the ACF and PACF plot. As we see from Figure 4.5 we selected the lags order of ARIMA (p,d,q) based on plot of training data set. Thus ACF is spike out at lag 1 and lag 4 this implied that possible order of non-seasonal MA(q) is one (q=1) and MA(Q) is also one (Q=1). PACF plot is spike out lag 1, lag 2, lag 4, lag 7, lag 8 and lag 12 this resulted in possible order of non-seasonal AR(p) p=2 and seasonal AR(P) is P=3 but the seasonal component of AR(P) is exponentially decay which shows there was no seasonal component of AR (P). Moreover, the parsimonious ARIMA model didn't consider higher lags which not more spike out for simplicity of the models.

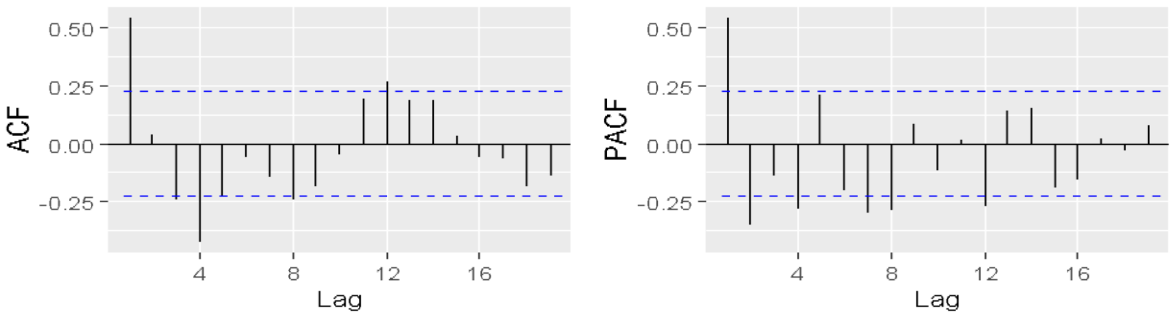


Figure 4.5: ACF and PACF plot of first differenced of training data set.

Therefore, based on the ACF and PACF graph the possible order of ARIMA (p, d, q)x(P,D,Q) were ARIMA(2,1,1)$(1,0,1)_4$, ARIMA(2,1,1)$(0,0,1)_4$, ARIMA(1,1,1) $(0,0,1)_4$ and ARIMA(1,1,1)$(1,0,1)_4$. Hence our data is quartet data the seasonal spike at 4 and it's multiply. This result revealed that existence of seasonal component which is Seasonal Autoregressive Integrated moving average (SARIMA) models were the best models to fit the series. An auto.arima() function in R selected the ARIMA(1,0,1)$(0,0,1)_4$.

4.3.2 SARIMA Models selection

Based on identified order of SARIMA models and others, we selected the best model by using model selection criteria .The first three model selection criteria listed in Table 4.3 were traditional model selection and used to select parsimonious model with minimum lags whereas, the latter two methods were used to select best model with minimum forecast error.

Table 4.3: The Candidate of SARIMA Models

Model	AIC	BIC	Log_Likelihood	RMSE	MAE
ARIMA(2,1,1)(1,0,1)	432.380	446.286	-210.190	3.869	47.584
ARIMA(2,1,1)(0,0,1)	430.415	442.003	-210.208	3.870	47.976
ARIMA(1,1,1)(0,0,1)	428.679	437.949	-210.339	3.877	47.499
ARIMA(1,1,1)(1,0,1)	430.644	442.231	-210.321	3.876	47.026
ARIMA(0,1,2)(0,0,1)	428.368	437.638	-210.184	3.868	47.614
ARIMA(1,0,1)(0,0,1)	437.163	448.816	-213.581	3.892	73.516

Table 4.3 shows ARIMA(0,1,2)(0,0,1) has the smallest values of AIC , BIC, log likelihood and RMSE, this implied that ARIMA(0,1,2)(0,0,1) models was the best model to fit training data, but auto.arima() function in R selected the ARIMA(1,0,1)(0,0,1) as the best model to fit training data. This implied why we said that selection of the order of the ARIMA or SARIMA model were based on experience of forecaster or subjective.

4.4 Machine Learning Methods of Analysis (Proposed Model)

4.4.1 Linear Model or Shrinkage Regressions Model

The shrinkage regressions models were used to fight multicollinearity among predictors, since EDA addressed multicollinearity among predictors. Shrinkage regression works by adding a penalty to the loss function, ultimately leading to the selection of important features that improved the forecast accuracy of inflation in this study.

4.4.1.1 Ridge Regression Model

To address the issue of collinearity, this study employed Ridge Regression. Ridge Regression tackles this problem by reducing the standard error of the estimates, although it introduces a slight bias into the results to achieve this benefit. We chose to implement the function over a grid of values ranging from $\lambda = 10^{10}$ to $\lambda = 10^2$ covering the full range of scenarios from the null model containing only the intercept to the least squares fit. The results showed that the coefficient estimates were much smaller in terms of the log(λ) norm, since λ=403 or log(λ) =6, grid values along with their log(λ) (See Appendix D, Figure D_1). It was similar to the result obtained by CV.

Figure 4.6 showed coefficient profile plot, where the *x*-axis is the logarithm of the penalty parameter λ and it provided the same information about shrink. Again each coefficient was represented by a line. The heavier penalty reduced the number of non-zero coefficient estimated; therefore, the number of lines in the plot decreased as the λ parameter increased. The numbers on top of the plot were the value of tuning parameter λ. In other words, as log lambda increases the all coefficients of variables shrank to zero, but not exactly zero like LASSO regression (from right to left). The top ten variables that significantly affect headline inflation at λ=0.14 which selected by 10-fold CV were food inflation, non-food inflation, political stability index, export price, reserve requirement, unemployment rate, numbers of vehicles, rainfall, world oil price and gross domestic fixed investment. But this did not show the orders of its importance in prediction headline inflation.

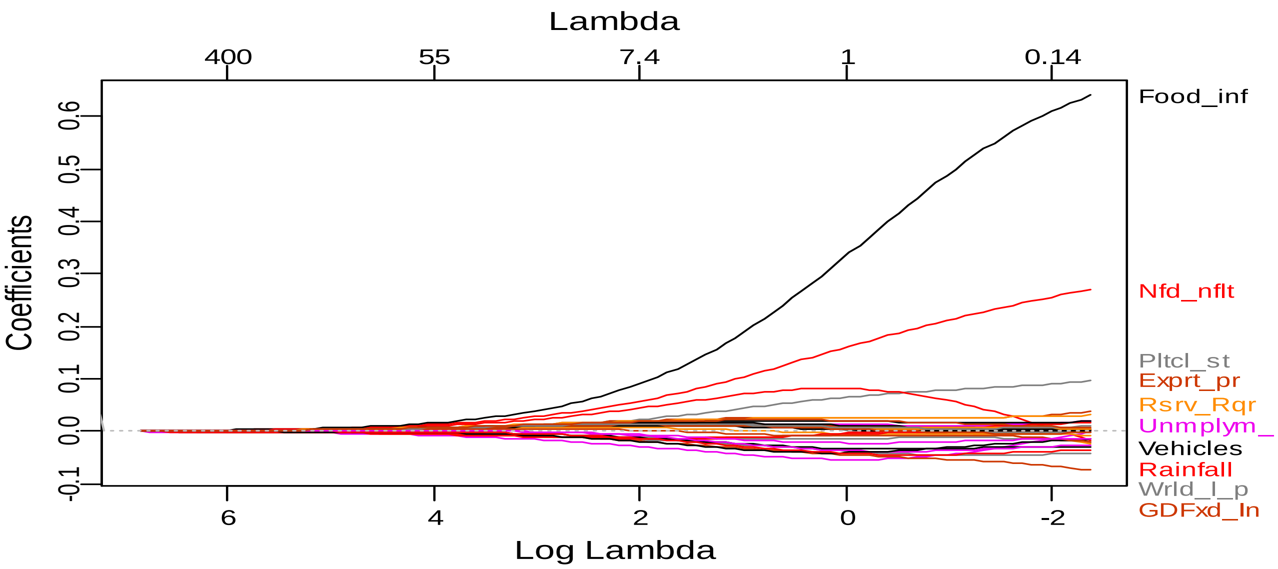


Figure 4.6: Coefficients of Ridge regression versus log lambda

Figure 4.7 illustrate, the headline number (34) shows numbers of predictors and the log(λ) shows magnitude of tuning parameter. As numbers of log(λ) increased the MSE of the model increased holdout all variables in the model.

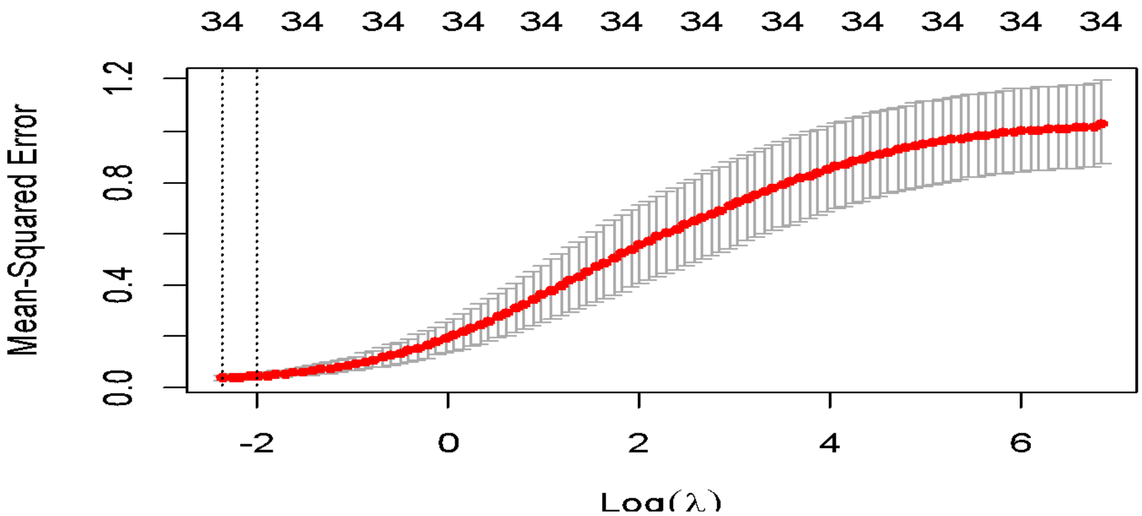


Figure 4.7: Results of 10-fold CV of Ridge regression model on training data

The sample means squared error (MSE) was plotted versus the logarithm of penalty parameter λ and the two vertical dashed lines locate the penalty corresponding to minimum MSE and a possible range of the penalty parameter. The left vertical dashed line locates the penalty for minimum MSE and the right vertical dashed line marks the one standard error limit of λ. From the plot, we chose of log(λ)=-2.45 seems reasonable. As a matter of the fact, the CV selected λ=0.09 which was close to exp(-2.45)= 0.09. The most important predictors variable in the model which minimize the MSE of the Ridge model were selected at **λ**=0.09. Accordingly the most important variables used to predict headline inflation were food inflation (100%), non-food inflation (42%), political stability index (14%), gross domestic fixed investment (11%), world oil price (7%), export price(5%),, rainfall, reserve requirement and numbers of vehicles imported others variable (See Appendix C, Figure C_1).

4.4.1.2 LASSO Model

The analysis also utilized LASSO regression to select important variables. Unlike Ridge Regression, LASSO works by shrinking the coefficients of less important predictors to zero. This effectively removes them from the model, leading to a more interpretable and simpler model with just the key variables influencing inflation forecasts. Figure 4.8 shows that there were 2 non-zero coefficient estimates of variables when log(λ) was -4. At $log(\lambda) = -8$ or λ=0.00034, only 7 variables had greater than 0.2 coefficient value. The food and non-food inflation took heavier penalty parameter, $\log(\lambda) > 0$; this implied that these variables were the most significant determinants of inflation. This implied that as the value of log(λ) decreased promotes feature selection, simplifies the model, and enhances its robustness to noise and outliers.

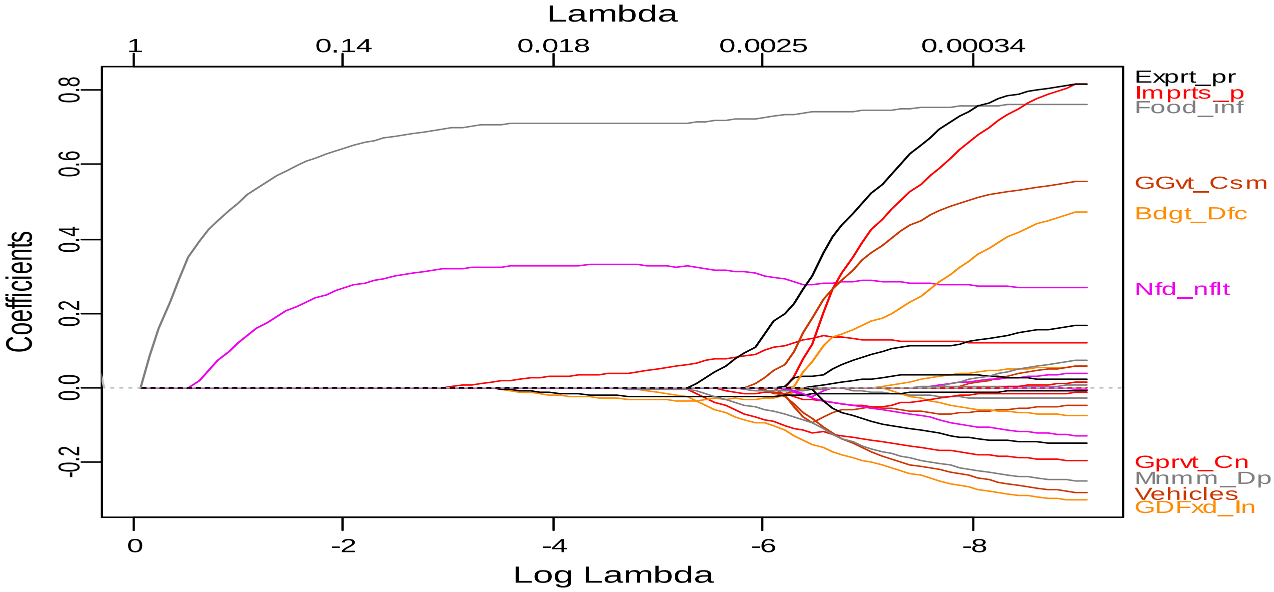


Figure 4.8: Coefficients of LASSO regression versus log(λ)

Figure 4.9 shows the result of 10-fold cross-validation as log(λ) increased and the numbers of predictors in the models decreased. The possible range of the penalty parameter corresponding to minimum MSE between log(λ) = -8.5 and log(λ) =-2.2 . From the plot, a minimum MSE at log(λ)=-8.01 seems reasonable. As a matter of the fact, the CV selects λ=0.0003 which was approach to exp(-8.1)= 0.0003. This was substantially lower than the test set MSE of the null model and of least squares, and very similar to the test MSE of Elastic Net regression with λ chosen by cross-validation. LASSO has a notable advantage over Ridge regression in that the resulting coefficient estimated were sparse.

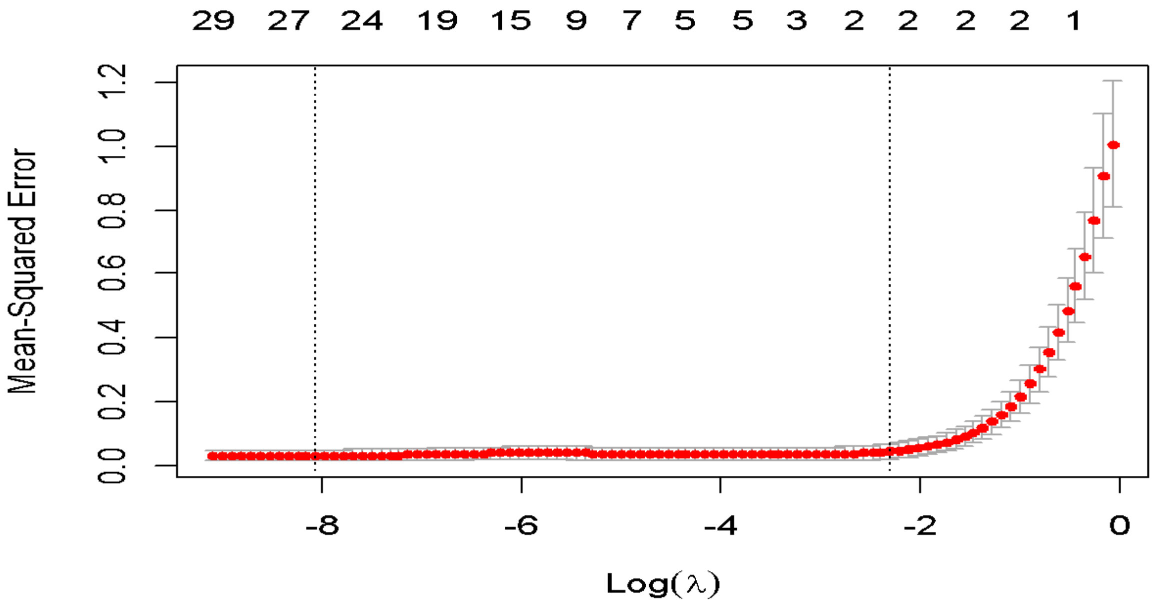


Figure 4.9: Result of 10-fold CV MSE versus log(λ)

LASSO model with λ = 0.0003 chosen by 10-fold CV selecting only 21 predictors with non-zero coefficients. Here we saw that 13 of the 34 coefficient estimated were exactly zero. The most important features (independent variable) that reduced forecasting error of headline inflation were chronological ordered from top to bottom as their importance. These were food inflation (100%), price of export and import good and service were 85% and 72% respectively, gross government consumption (53%), budget deficit (51%), non-food inflation (39%), gross domestic fixed investment (30%), minimum deposit (27%), numbers of imported roads motors and vehicles (23%), gross private consumption (22%,), political stability index (12%), imported food price (10%) were contribute more to increase the accuracy of LASSO model prediction (See Appendix C, Figure C_2).

From Table 4.4 the signs of coefficient of variables showed the direction of relationship of independent variables with the headline inflation and magnitudes showed significance of the variables, this implied that if coefficient of exogenous variables approach to one, it significantly affect response variables (headline inflation). Therefore, variables such as gross government consumption, food and non-food inflation, political stability index, world oil price, net foreign asset, budget deficit, foreign direct investment, unemployment rate, exchange rate, price of export and import goods and service are positively affect the headline inflation.

Table 4.4: Estimated non-zero coefficients of LASSO model

Variable	Coefficients	Variable	Coefficients
Minimum Deposit	-0.186**	Gross private consumption	-0.158**
GG Consumption	0.424**	Non-food inflation	0.327**
Food inflation	0.737***	Foreign direct investment	0.084
Political stability	0.104	Net Foreign Assets	0.179**
GD fixed Investment	-0.231**	Lending interest rate	-0.008
World oil price	0.012	Capital account	-0.068
Agricultural PP	-0.061	Vehicles	-0.223**
Export price	0.694***	T-bill sold	-0.017
Rainfall	-0.009	Budget deficit	0.117**
Import price	0.653***	Unemployment rate	0.011
Exchange rate	0.002	Price of food imported	-0.124**

**** sigf. coefficients more than absolute of 0.5, ** more than absolute of 0.1*

Whereas, variables such as minimum deposit, gross private consumptions, rainfall, capital account, government expenditure on facilities, gross fixed investment, T-bill sold, agricultural production supply, price of food imported and numbers of road motors and vehicles imported (transportation service) are negatively affect the headline inflations. The most significant variables affected headline inflation were food inflation, gross government consumption, non-food inflations, export and import price of goods and services, government fixed investment, gross private consumption, minimum deposit, price of food imported and numbers of vehicles and turned to improve accuracy of a model.

4.4.1.3 Elastic Net Model

We found that LASSO eliminated many variables, making the model too simplistic. To address this, we switched to the Elastic Net technique. This method offered a middle ground between Ridge Regression, which reduces error but can retain unimportant variables, and LASSO's variable selection. By using Elastic Net, we were able to capture more relevant variables while still keeping the model from becoming overly complex.

From Figure 4.10 as log lambda increased the all coefficients of variables shrank to zero. There were 3 non-zero coefficient estimates of variables when log(λ) is -4 which retained one variable compared to LASSO at same point. At $log(\lambda) = -8$ or λ=0.00034, only 8 variables had greater than 0.2 coefficient value. The food and non-food inflation took heavier penalty parameter greater than zero; this implied that these variables were the most significant determinants of inflation**.** The top ten variables that significantly affect headline inflation at $\alpha = 0.5$ and λ=0.00034 selected by 10-fold CV were import price of good and services, food inflation, gross government consumption, export price, non-food inflation, net foreign asset, imported food price, minimum deposit, gross domestic fixed investment and numbers of vehicles.

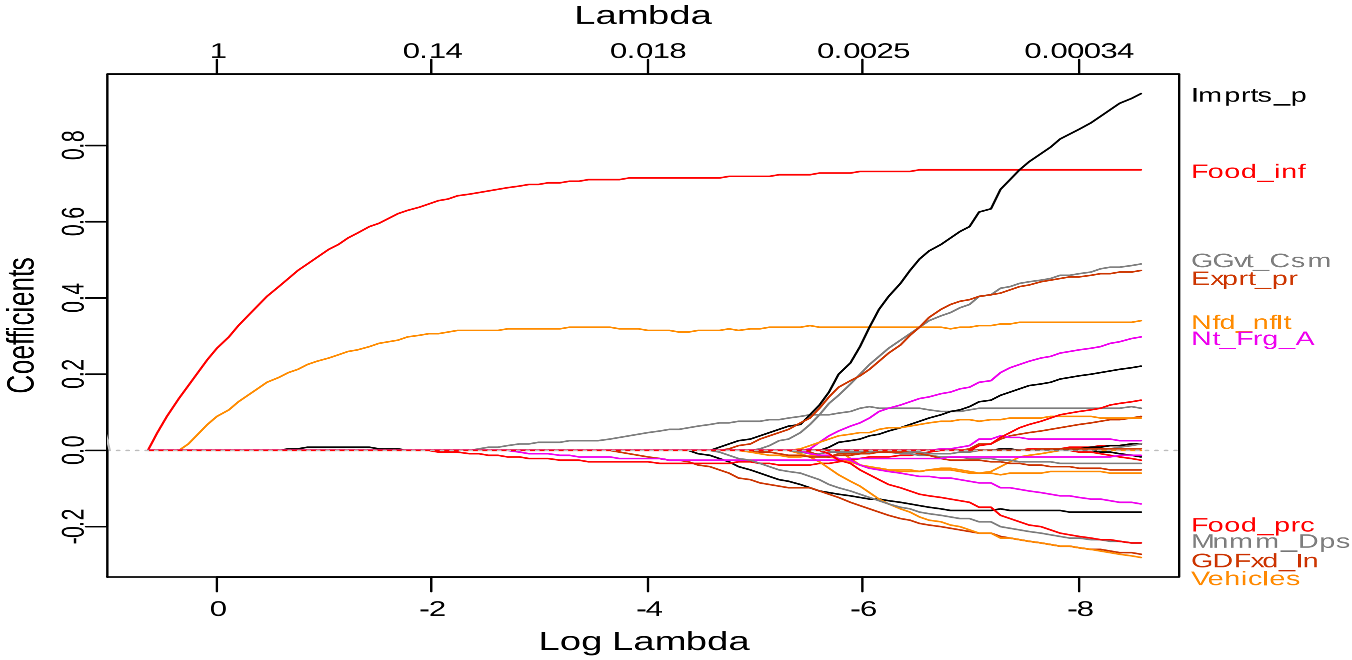


Figure 4.10: Coefficients of Elastic net regression versus log(λ)

The 10-fold CV with 3 time repetition selected the tuning parameter $\alpha = 0.5$ and λ= 0.0005 at which the MSE is minimum. When α approach to zero the elastic net model is similar to ridge regression and when approach to 1 it also similar to lasso regression. Figure 4.11 shows the minimum MSE at log(λ)= -7.669 which indicated by vertical dash line on the left side and when compared to for ridge model the region of tuning parameter λ is too large. Almost the value of tuning parameter λ for both LASSO and Elastic Net models is equal; this implied variables selected by both models almost similar.

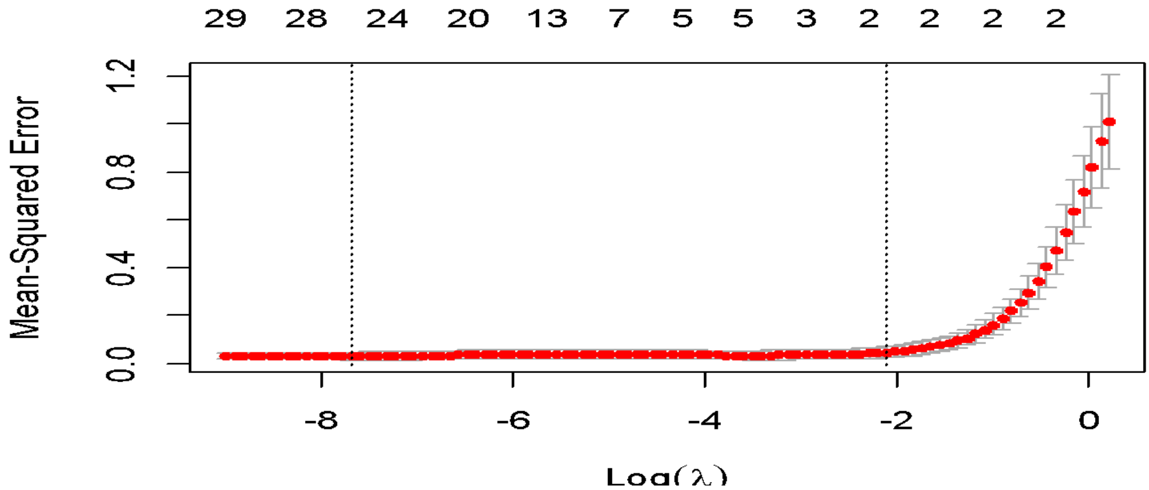


Figure 4.11: Result of 10-fold CV MSE versus log(λ) of Elastic Net model

The Elastic Net model selects the most important variables such as food inflation (100%) followed by price of import good and service (93%.), gross government consumption (57%.), price of export good and service (52%), non-food inflation(45%), numbers of imported roads motors and vehicles (30%), gross domestic fixed investment(30%), net foreign asset (27%), minimum deposit(27%), imported food price (20%) were majors contribute to increase the accuracy of Elastic Net model prediction (See Appendix C, Figure C_3).

4.4.2. Non-linear Model

4.4.2.1 Random Forest Model

Random Forest (RF) has powerful for capturing non-linear relationships and help to reduce tree correlation by injecting more randomness into the tree-growing process. More specifically, while growing a decision tree during the bagging process, random forests perform split-variable randomization, in k^{th} region where each time a split is to be performed, the search for the split variable is limited to a random subset of the original p feature. The split-variable randomization features of RF were controlled by *mtry* (hyperparameter) which helps to balance low tree correlation with reasonable predictive strength.

The analysis evaluated the impact of the number of trees in the Random Forest model and performance of method employed. The out-of-bag (OOB) estimate of the root mean squared error (RMSE) of bagging regression was 0.317 for 100 trees and decreased to 0.299 for 500 trees. This indicated that adding more trees resulted in averaging over decision trees with higher variance. The analysis also compared OOB error with 10-fold cross-validation (CV) using bagged CART (Classification And Regression Trees). The result shows RMSE of OOB error was 0.279 for 100 trees and 0.272 for 500 trees. However, calculating the OOB error of bagged CART with 500 trees took significantly longer (0.78 seconds) compared to bagged CART with 100 trees (0.20 seconds) on our machine. The 10-fold CV selected the best model with 18 randomly chosen features out of the original 34. This model achieved a minimum RMSE of 0.272 and likely included uncorrelated predictors with strong predictive power (See Table 4.5).

Table 4.5: Selection of a hyperparameter (mtry) from original 34 features

Techniques	mtry	RMSE of OOB
Bagging regression trees with 100 bootstrap replications	34	0.317
Bagging regression trees with 500 bootstrap replications	22	0.299
Bagged CART with 10-fold CV with 100 trees	13	0.279
Bagged CART with 10-fold CV with 500 trees	18	0.272

We observed a significant decrease in variance (and consequently, error) as the number of trees increased in the Random Forest model (See Figure 4.12). This suggested a rapid improvement in model accuracy. However, this trend eventually plateaued, indicating that adding more trees no longer offered substantial benefits. This plateau likely signified that a sufficient number of trees had been reached for optimal performance.

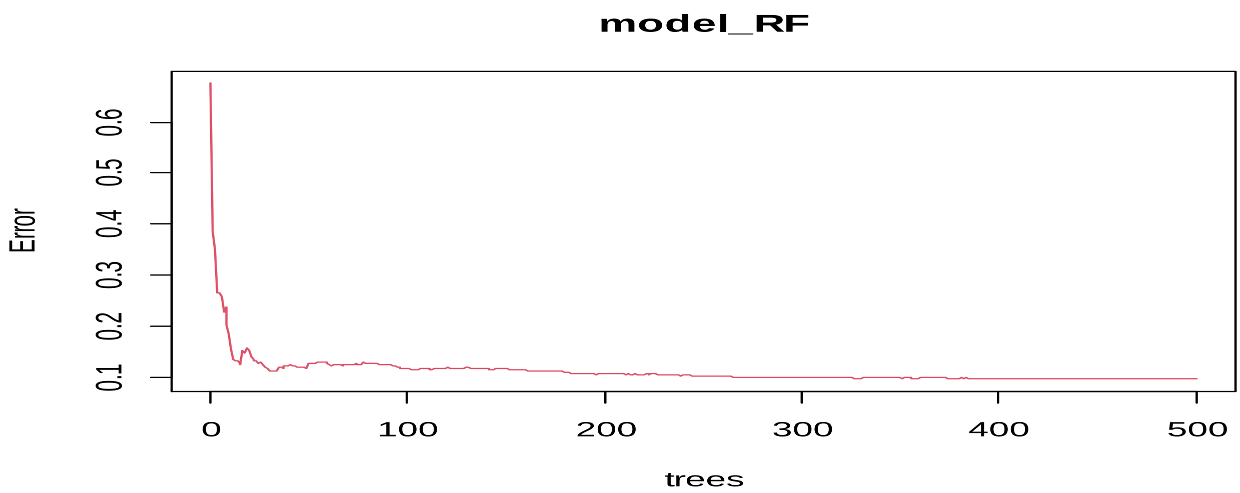


Figure 4.12: Result of RFs applied to the quarterly macroeconomic data with the growth inflation as the dependent variable. 500 trees and mtry = 18 are used.

Figure 4.12 shows that as the number of trees increased the error of the RF model decreased. This implies that high numbers of trees decrease the model forecast error, 500 trees are commonly used in much research to get numbers of nodes to classify this tree based on their purity of nodes. The result shows that 18 variables were used to split this tree. The RF models with 500 trees explained 94.9% of the variability of headline inflation. Random forest was also used to identify the most important features that explain the variability of response variables. The most important variables that significantly affected headline inflation were food inflation, non-food inflation, T-bill sales, agricultural production price, gross domestic fixed investment, and others (See Appendix C, Figure C_4).

4.4.2.2 Artificial Neural Network (ANN) Model

To reduce overfitting in the ANN models, a common issue with Neural Networks, the analysis used the 21 most important covariates identified by LASSO regression.

4.4.2.2.1 ANN Model Training

To optimize the ANN models, the analysis performed a hyperparameter tuning process. This involved evaluating various settings like learning rate, maximum iterations, activation functions, and network architecture using Jupyter Notebook. Five repetitions for each combination ensured consistency. The chosen settings aimed to minimize the Root Mean Squared Error (RMSE). This resulted in a maximum iteration range of 10,000-20,000 to ensure convergence, along with the selection of the hyperbolic tangent activation function and Stochastic Gradient Descent (SGD) optimizer for minimizing the loss function with momentum method. The learning rate, controlling the learning step size, was set $\gamma = constant\ with\ rate\ 0.1,\ \alpha = 0.9$ or momentum with initialization rates 0.0001. Finally, the training favored a single hidden layer with 3 possible node configurations (15, 10, or 5) to minimize the loss function and avoid overfitting by keeping the network architecture less complex.

Overall, the analysis adopted a meticulous approach to training the ANN models. By carefully tuning the hyperparameter and selecting the most effective activation function, optimizer, and hidden layer structure, the aim was to achieve optimal performance and minimize potential errors in the forecasting process. Accordingly, a single hidden layer ANN model was the best model used to forecast headline inflation. This model had a single hidden layer with different numbers of nodes, which was addressed as follows

4.4.2.2.2 Neural Network Autoregressive (NNAR) ANN Model

NNAR models specifically account for temporal dependencies by using past values of the output and inputs, making them suitable for headline inflation series modeling and forecasting. The NNAR modelling headline inflation into two ways; those were without exogenous and with exogenous variables.

NNAR without exogenous variables

In order to find the best performing model for the ANN architecture NNAR (univariate Neural Network Auto-Regressive), the analysis employed a grid search technique. This method involves evaluating the model's performance across a range of possible parameter settings. The grid search identified the NNAR (9,1,5) model as the optimal choice, based on its minimal mean squared error of 0.0024. These parameters: 9 were represented the number of lagged observations used as input features for the model, 1 was indicated that data was transformed by taking the firs seasonal difference, likely to address stationarity issues and 5 where signifies the number of neurons in the hidden layer. These units essentially help the model capture the underlying relationships within the data. By selecting the NNAR (9,1,5) model with the lowest mean squared error, the analysis aimed to achieve the most accurate inflation forecasts possible.

The analysis sought to improve the forecasting performance by averaging the results from 20 individual neural networks. Each network possessed a 9-5-1 architecture and network had 56 weights or parameters to adjust. These weights determined the strength and direction of connections between the nodes. The weights were calculated as follows: 9 input nodes multiplied by 5 (for the hidden layer) plus 5 bias terms for the hidden layer and 1 bias term for the output layer, totaling 56 weights or parameters. Similarly, the analysis employed 10-fold cross-validation (CV). The 20 networks were trained on 9 folds each time, with their performance evaluated on the remaining fold. Ultimately, the network architecture (9-5-1) with the lowest Mean Squared Error (MSE) of 0.036 across the 10 folds was chosen as the final model. This approach aimed to achieve a more robust and potentially more accurate inflation forecast compared to a single network by leveraging the collective intelligence of the ensemble and the validation power of CV.

NNAR with exogenous variables

The analysis also explored a more complex model called NNARX (p,P,k). This model extends the univariate NNAR by incorporating additional external variables (exogenous variables) to potentially improve inflation forecasting. A grid search technique was used to identify the optimal NNARX configuration based on minimizing the Mean Squared Error (MSE). Among various candidate models, NNAR (5,2,10) emerged as the best option with the lowest MSE of 0.00056.

The NNAR (5,2,10) model forecasted headline inflation using several key elements. It considered five lagged response variables, leveraging past observations to learn from historical trends. Additionally data was transformed by taking the second seasonal difference. The model featured a hidden layer with 10 neurons, which helped capture complex relationships and patterns in the data. This model resulted in a 27-10-1 network structure and network had 291 weights. The NNAR model used 27 input nodes: 5 for past inflation rates and 22 for external factors influencing inflation. The hidden layer had 10 neurons to process and learns complex relationships between these inputs. Finally, a single output unit generated the forecasted inflation.

The analysis acknowledged that visualizing the NNAR (5,2,10) model's behavior graphically proved difficult. This stemmed from the model's complexity, particularly its use of combined activation functions (hyperbolic tangent and linear). Despite this visualization challenge, the NNAR (5,2,10) model served a valuable purpose. By incorporating both historical inflation data and additional external factors, it aimed to generate more robust and potentially more accurate forecasts compared to simpler models. This approach leveraged a wider range of information, potentially leading to improved forecasting capabilities.

4.4.2.2.3 Multiple Perceptron Layer (MLP) Model

We employed MLP models to capture complex interactions between variables that couldn't be easily modeled by linear regression techniques and since hidden layer was not more than one. Moreover, unlike NNAR, MLP models do not inherently consider temporal dependencies unless explicitly structured. Similar to NNAR model we modelling headline inflation into ways, those were:

MLP without exogenous variables

The analysis also investigated MLP models for univariate inflation forecasting. These models resemble AR(p) models by using past inflation values (lags) as input features. The number of nodes in the hidden layer, a key factor in model complexity and 5-fold cross-validation identified the optimal configuration. Accordingly, an input layer has 3 lags (1, 2, and 4) of headline inflation, where 1and 2 non-seasonal components and 4 was seasonal that were independent component with 10 hidden nodes numbers which was prespecified. While the number of hidden layers can impact model complexity, the primary selection criterion here was minimizing the MSE combined with the number of hidden layers. This approach aimed to balance model performance with avoiding overfitting issues.

The tuning process within the MLP models revealed that the hyperbolic tangent activation function yielded superior performance compared to the logistic activation function. This is evident from the achieved Mean Squared Error (MSE) Hyperbolic Tangent (MSE = 0.014) and logistic (MSE = 0.017). This outcome suggests that the hyperparameter tuning effectively identified the most suitable activation function for minimizing forecasting errors in this specific case. See Table 4.6.

MLP with Exogenous Variables

The analysis also investigated MLP models that incorporated additional factors beyond past inflation data (exogenous variables) to forecast inflation. Similar to the previous analysis, 5-fold cross-validation played a crucial role in selecting the most effective model configuration. Including more lags and exogenous variables could potentially improve forecasting by capturing richer information, but it increased the risk of overfitting. The backpropagation algorithm iteratively adjusts the model's internal parameters based on calculated errors during training. Each adjustment was repeated 20 times with a predetermined learning rate. This iterative process aimed to find a global minimum, a state with a low level of error, for the model. In essence, the analysis prioritized finding a balance between incorporating relevant information and maintaining model interpretability. This approach aimed to achieve accurate inflation forecasts while still allowing for some understanding of how the model arrives at its predictions. The backpropagation algorithm with a fixed learning rate facilitated this training process.

Table 4.6: Architectures of MLP models without and with Exogenous Variables

Model	**Regressor**	**Nodes**	**Act.function**	**Lags**	**MSE**
1	No	10	Tangent-hyperbolic	3	0.014
2	No	10	Logistic	3	0.017
3	7	5	Logistic	0	0.071
4	7	5	Tangent-hyperbolic	0	0.063
5	21	5	Tangent-hyperbolic	1	0.001
6	21	5	Logistic	1	0.008

From Table 4.6, based on the Mean Squared Error (MSE), Model 5 stands out as the best performer with the lowest MSE of 0.001. This indicates it produces the most accurate forecasts compared to the other models. The analysis also compared models with different activation functions: hyperbolic tangent and logistic. Interestingly, Model 5 with the hyperbolic tangent function achieved a significantly lower MSE (0.001) compared to Model 4 with the logistic function MSE (0.008). This suggests that the hyperbolic tangent function is more effective in minimizing forecasting errors for this specific task. The non-linear nature of the hyperbolic tangent function allows it to capture more complex relationships within the data compared to the simpler logistic function. Model 5 has a network structure of 22-5-1, with 126 weights that the model needs to learn during training (See Appendix D, Figure D_3)

In generally, the result suggested that incorporating lags and using a tangent-hyperbolic activation function led to a substantial improvement in model performance (Model 5). The choice of activation function also appears to be important, with the logistic function potentially hindering performance in this specific case (Model 6 compared to Model 5).

4.5 Variables Importance

In machine learning analysis the selected features (variables) were the most important variables that improve forecast accuracy of the models. The first top ten selected variables were listed in below Table 4.7. Examining Table 4.7 offered valuable insights into the most important factors influencing inflation and improves forecasts accuracy of the models. All four models displayed a remarkable agreement when identifying the top ten most impactful features. This consistency suggests a strong understanding of the key drivers behind inflation and improves forecast accuracy of the models. Food inflation stood out as the most critical factor, consistently ranking first across all models. This emphasizes its significant influence on overall inflation and forecast accuracy of the models. Other features consistently selected in the top positions included non-food inflation, export price, and government spending (gross government consumption), gross domestic fixed investment and numbers of vehicles were the major factors determine the forecasting accuracy of the all models.

Table 4.7: Top 10 important variables selected by Machine Learning Models

Importance	**Machine Learning Models**			
	Ridge regression	LASSO Model	Elastic Net	Random Forest
1	Food inflation	Food inflation	Food inflation	Food inflation
2	Non-food inflation	Export price	Export price	Non-food inflation
3	**PSI**	GG consumption	GG consumption	GG consumption
4	GDF Investment	Budget deficit	Non-food inflation	**T-bills Sales**
5	**WOP**	Non-food inflation	GDF Investment	**Agricultural PP**
6	Export Price	GDF Investment	Budget deficit	GDF Investment
7	**Rainfall**	Minimum deposit	Vehicles	**T-bills Demand**
8	RR	Vehicles	Minimum deposit	Unemp. Rate
9	Vehicles	GP consumption	GP consumption	RR
10	Unemp. Rate	**Capital Account**	**Net foreign asset**	Exchange rate

Table 4.7 also revealed some model-specific features. These features appeared only in the "highlighted variables" section and were deemed important by a particular model but not necessarily by all. For instance, ridge regression models identified political stability index, world oil price, and rainfall as influential factors, suggesting their potential impact within this specific approach. Random forest, on the other hand, focused on different aspects, selecting T-bill sales, agricultural production price, T-bill demand, and exchange rate as some of its top ten features. Interestingly, the features chosen by LASSO and Elastic Net models were quite similar. This similarity likely stems from the way the Elastic Net's tuning parameters were set they closely resembled those of LASSO. Both approaches aim to minimize the influence of less important predictors, leading to similar feature selection in this case. However, there was a minor difference in the tenth position between LASSO and Elastic Net. The LASSO model identified *"capital account*" as the tenth most important variable, while the Elastic Net model favored *"net foreign asset."* In conclusion, analysis identified a core set of features consistently chosen by all models, alongside

some additional factors highlighted by specific modeling approaches. Understanding these influential factors is essential for developing accurate and robust inflation forecasting models.

4.6 Model Comparison

4.6.1 Univariate Analysis

The univariate analysis was done for both SARIMA and machine learning models (MLP, NNAR). The traditional model (time series model) used as bench mark model to compared with machine learning models without assumptions of stationarity.

Table 4.8: Model Comparison without exogenous variables

Models	Training data set (in-sample)			Test data set (out-sample)		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SARIMA(0,1,2)(0,0,1)	3.868	2.499	**47.614**	15.577	13.946	**51.058**
SARIMA(1,0,1)(0,0,1)	3.892	2.518	73.516	16.054	14.443	53.069
NNAR(9,1,5)[4]	**0.049**	**0.033**	53.78	1.343	1.229	373.89
MLP 5 hidden node	0.166	0.113	154.270	**0.563**	**0.463**	162.196

In-Sample Forecasting (Training Data): The Neural Network Autoregressive (NNAR(9,1,5)) model excelled in in-sample forecasting. It achieved the lowest Root Mean Squared Error (RMSE) of 0.049 and the lowest Mean Absolute Error (MAE) of 0.033, indicating it produced the most accurate forecasts on the training data compared to other models. This suggests that the NNAR model effectively learned the patterns within the training data.

In Out-of-Sample Forecasting (Testing Data): The picture changed when evaluating performance on unseen data (testing data). Here, the MLP (Multi-Layer Perceptron) model with 5 hidden nodes emerged as the leader. It achieved the lowest RMSE (0.563) and MAE (0.463) on the testing data, suggesting it generalized better to unseen data compared to other models. However**,** based on MAPE the SARIMA(0,1,2)(0,0,1) model, a type of univariate time series analysis model, outperformed all other models based on the Mean Absolute Percentage Error (MAPE) for both in-sample and out-of-sample forecasting. Overall, the NNAR model performed well on the training data, the MLP model seems to be a more robust choice in this case. It exhibits a better balance between training and testing performance, suggesting it might generalize better to unseen data**.**

4.6.2 Multivariable Analysis: Machine Learning Models

The machine learning analysis based on exogenous variables categorized as linear and non-linear models. This is used to identify which model can be handle the pattern of headline inflation and outperformed**.**

Table 4.9: Machine Learning Models Comparison

Models	Training data set (in-sample)			Test data set (out-sample)		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Ridge Model	0.166	0.099	35.865	0.140	0.101	29.324
LASSO Model	0.110	0.057	45.085	0.172	0.087	14.868
Elastic Net model	0.112	0.058	48.025	0.196	0.096	14.978
MLP 5 hidden nodes with exogenous	0.034	0.025	25.110	1.908	1.415	189.402
NNAR(5,2,10)[4] with exogenous	**0.000**	**0.000**	**0.137**	**0.128**	**0.067**	**10.576**
Random Forest	0.132	0.070	61.720	0.184	0.146	20.835

Table 4.9 offered valuable insights into how different models performed in forecasting inflation, particularly when focusing on out-of-sample forecasting (generalizing to unseen data). This analysis used Mean Absolute Percentage Error (MAPE) as the key metric. The analysis revealed that linear models, such as LASSO and Elastic Net, outperformed non-linear models like Multi-layer Perceptron (MLP) and Random Forest in out-of-sample forecasting based on MAPE. This suggests that linear models were more effective at capturing the proportional changes in inflation for unseen data points. Among the linear models, LASSO emerged as the leader. It achieved the lowest errors (RMSE = 0.172, and MAE= 0.087) on unseen data while maintaining a good MAPE. This indicates a well-balanced approach, capturing both accuracy and the proportional changes in inflation.

The NNAR (non-linear) model was a standout performer. It achieved the lowest values for RMSE, MAE, and MAPE in both in-sample and out-of-sample forecasting. However, its in-sample performance might be misleading because of potential overfitting, where the model learned the training data too well. Despite this concern, the NNAR model still achieved a relatively low MAPE on unseen data. This suggests it could capture some proportional changes in inflation. This was further supported by its high forecast accuracy (based on 1-MAPE). The accuracy was 99.86% for in-sample forecasts and 89.42% for out-of-sample forecasts. Considering its overall performance and ability to balance accuracy with capturing proportional changes, it appeared to be a more robust choice for both in-sample and out-of-sample headline inflation forecasting based on this result.

The RMSE, MAE, and MAPE values increased when going from in-sample forecasting to out-sample forecasting for non-linear machine learning models. This implied that non-linear machine learning models were unable to capture the structures of headline inflation at long forecast horizons; unlike they're in-sample forecasting. However, the forecast performance of linear machine learning models was worse than that of non-linear machine learning models; the RMSE, MAE, and MAPE decreased when going from an in-sample forecast to an out-sample forecast. These implied that linear machine learning models were able to capture the structures of headline inflation at long forecast horizons, unlike it's in sample forecasting. Finally, in the model performance comparison for univariate analysis (time series and machine learning analysis) and multivariable-based machine learning analysis, the machine learning analysis outperformed in both. The machine learning models based on multivariable data minimized the forecast error, indicating their ability to handle multidimensional data and improve forecast accuracy. For instance, from Tables 4.9 and 4.10, the RMSE of NNAR with exogenous variables and without exogenous variables was 0.000 and 0.050 for in-sample forecasting and 1.079 and 1.229 for out-of-sample forecasting, respectively. The NNAR (5, 2, 10) with exogenous variables was selected based on its performance. This model was checked for diagnostics before forecasting the future values of headline inflation.

4.7 Model Diagnostic for NNARX Model and Forecasting

4.7.1 NNAR (5, 2, 10) Diagnostic

Models diagnostic of NNAR models based on in sample and out sample forecasting of the models. Based on histogram of residuals of NNAR (5, 2, 10) in-sample and out-sample forecasts, the residuals were normally distributed with mean zero and constant variance. ACF graphs also shows that the there is no serial correlation between the lags (See Appendix D, Figure D_5 and D_6).

In-sample	Out-sample
Ljung-Box test	Ljung-Box test
data: Residuals from NNAR(5,2,10)[4]	data: Residuals from NNAR(5,2,10)[4]
Q* = 5.659, df = 8, p-value = 0.6854	Q* = 6.4831, df = 4, p-value = 0.1659
Model df: 0. Total lags used: 8.	Model df: 0. Total lags used: 4

In both in-sample and out-sample forecasting, the p-values were 0.6854 and 0.1659, respectively, both of which are greater than 0.05. These results indicate that there was no significant autocorrelation in the residuals up to lag 10 and lag 4 respectively. Therefore, the model appears to have adequately captured the autocorrelation structure in the data.

4.7.2 ANN Forecasting

The selected NNAR (5, 2, 10) model was used to forecast Ethiopia headline inflation by using test data set which covered from 2019Q1 to 2023Q4 for next five year or 20 horizon. From Figure 4.19 we observed that the forecasted values of inflation has similar pattern to the test data. This indicated that the performance of machine learning in the recognition pattern and incorporate the trend and seasonal information into its algorithm, giving you the best out-come (Sorjamaa, *et al.*, 2006)

The graphs show that the headline inflation will be low for next two years (2024 and 2025) compared to current year. It gradually increase from 2026Q1 reached peak point in 2027Q2 with maximum 36% and then gradually decline up to the 2028Q4 (See Appendix D, Table D_1).

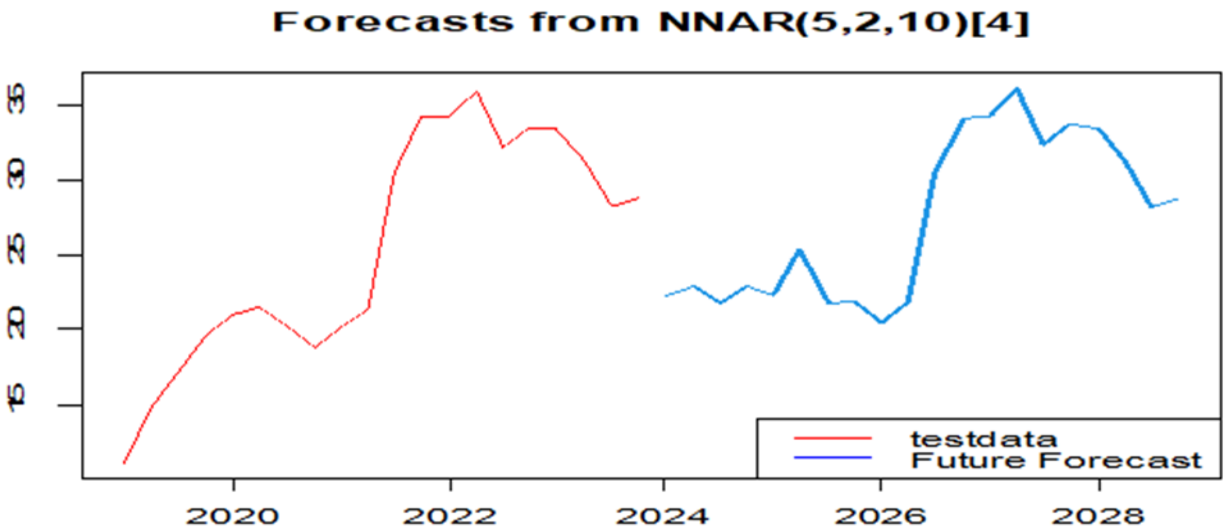


Figure 4.13: ANN Model future forecast

4.8 DISCUSSIONS

This study aimed to model and forecast Ethiopia's headline inflation and identifies the most important predictors that improve the forecast accuracy of machine learning models using data from 2000Q1 to 2023Q4. The SARIMA model was a commonly used time series analysis method for modeling and forecasting headline inflation series. In contrast, machine learning methods such as Ridge regression, LASSO, Elastic net, Multi-Layer Perceptron (MLP), Neural Network Autoregressive model (NNAR), and Random Forest were employed. These machine learning models aimed to identify crucial features predicting headline inflation and enhance the models' forecast accuracy.

Ridge regression and Random Forest were utilized to address collinearity among predictors and improve forecast accuracy. The LASSO model was used to select the important feature variables for predicting headline inflation, which then served as inputs for the ANN models. The NNAR models were employed to handle the seasonality of quarterly headline inflation, similar to time series models.

The multivariable analysis of the machine learning models identified the most important features selected by LASSO, Elastic net, ridge regression, and random forest. But the LASSO model is ability to perform feature selection, which significantly contributed to the improved forecasting accuracy. The important predictors identified such as food inflation, non-food inflation, gross private consumptions, rainfall, capital account, world oil price, government expenditure, budget deficit, gross fixed investment, unemployment rate, were consistent with previous studies (Mohamed F., 2023; Nakorji and Aminu, 2022; Tekeber N., et al., 2019; Bedada, et al., 2020; Mulugeta, 2020). These variables were crucial in explaining the variations in headline inflation, affirming their relevance in economic forecasting. The finding also was based incorporating variables like political stability indexes, as suggested by Mwangi (2016), could enhance the predictive capabilities of these models, offering a more comprehensive tool for economic. The number of vehicles (includes road motors) uniquely identified by this study which negatively affect the headline inflation in Ethiopia. But there literature consisted to this study.

The study's findings indicated that the LASSO model outperformed the benchmark model (SARIMA (0,1,2)(0,0,1)), Ridge regression, and Elastic net model in both in-sample and out-sample forecasting, with a minimum RMSE of 0.11 and 0.17, respectively. This result aligns with the findings of Özgür and Akkoç (2021), who concluded that LASSO performed best compared to all other baseline (ARIMA) and shrinkage models, such as ridge regression and Elastic Net, achieving a root mean squared error of 0.834. Furthermore, the LASSO model outperformed the Ridge model, random walk, and AR models, although the Elastic Net model produced results nearly identical to those of the LASSO model (Baybuza, 2018; Popoola, Y. 2023).

Similarly, the findings demonstrated that the ANN model outperformed traditional time series bench mark models (SARIMA) and linear machine learning models such as, Ridge regression, LASSO, Elastic Net in forecasting Ethiopia's headline inflation. Specifically, NNAR (5, 2, 10) achieved the lowest RMSE both in-sample and out-sample (0.000 and 0.128, respectively), highlighting its robustness and accuracy. This result is consistent with the findings of (Rita 2019; Yuniar I., *et al.*, 2020; Michael V., *et al.*, 2020; Nakorji and Aminu, 2022; Mahajan and Srinivasan, 2019; Nunoo, E.,and Kings, I. 2013), who also reported superior performance of the ANN model over other baseline models.

Finally this study demonstrated that the machine learning models with exogenous variables have improved the forecast accuracy of the models compared to without predictors. This finding related with (Rodrıguez-Vargas, 2020; Gustavo S. 2023), who suggested that excluding irrelevant data and add related variables improved forecasting accuracy of the economic and financial time series data.

CHAPTER 5

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

This study aimed to model and forecast Ethiopia's headline inflation, as well as to identify predictors significantly contributing to headline inflation using a machine learning approach. The trend of quarterly headline inflation in Ethiopian shows that there were slight increases from 2000 to 2007Q2 then sudden shift in 2008Q4 and rapidly increase from 2015Q1 to 2023Q4 which showed existence of both conditional mean and variance.

Gross government consumption, food and non-food inflation, world oil price, net foreign asset, budget deficit, unemployment rate, exchange rate, price of export and import goods and service, minimum deposit, gross private consumptions, rainfall, capital account, government expenditure on facilities, gross fixed investment, T-bill sold, agricultural production price, price of food imported were predictors significantly determining headline inflation in Ethiopia. Among these predictors' food and non-food inflation, price of export and import goods and service, gross fixed investment, numbers of vehicles were the most factors contributors of forecast accuracy of the models. Moreover, political stability index and numbers of vehicles (includes road motor) also uniquely identified variables that significantly affect headline inflation and improves the forecast accuracy of the models.

The NNAR model showed remarkable accuracy and outperformed all the other models used in the study. The forecasted values of headline inflation for next five years indicated that the headline inflation has fluctuated over the forecast periods. The empirical results revealed that a uncertainty in headline inflation follow the same pattern to recent five years.

5.2 RECOMMENDATIONS

Based on the findings of this study, the following recommendations are made:

- Since there is uncertainty in the headline inflation; the government bodies, policy makers, investor and concerned bodies should give attention to headline inflation.
- Policymakers and stakeholders such as National Bank of Ethiopia (NBE) pay attention to the key determinants of Ethiopian headline inflation.
- Government and concerned body should encourage the improvement in numbers of transportation facility.

- The government, political parties, and concerned bodies should maintain political stability in order to tackle inflation.

For future research, this study recommends to improve the forecast accuracy of machine learning methods in predicting Ethiopia's inflation by employing advanced deep learning models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and ensemble models. Since these modeling approaches have a significant promise in handling complex and sequential data, which can enhance the robustness and precision of inflation forecasts.

Limitation of the Study

This study is limited by the data employed. It utilized headline inflation data with 34 related covariates from January 2000 to December 2023. The analysis focused on in-sample and out-of-sample forecasting. Due to this restricted limitation, the findings may not be generalizable to all machine learning models. Machine learning typically deals with multidimensional data and often utilizes rolling window forecasting methods. Future research could address this limitation by incorporating additional covariates that might influence inflation. Additionally, employing rolling window techniques throughout the forecasting process could improve the overall accuracy of machine learning models across various forecasting horizons.

REFERENCES

Abd El-Aal, M. F. (2023). Analysis factors affecting Egyptian inflation based on machine learning algorithms. Data Science in Finance and Economics, 3(3), 285-304.

Abebe, A., Temesgen, A., & Kebede, B. (2023). Modeling inflation rate factors on present consumption price index in Ethiopia: threshold autoregressive models approach. Future Business Journal, 9(1), 72.

Abebe, T. H. (2021). Using Models of the GARCH Family to Estimate the Level of Food and Non-Food Inflation in Ethiopia. Journal of Applied Economic Research. 2021. Vol. 20.№ 4, 20(4), 726-749.

Abledu, G. K., & Agbodah, K. (2012). Stochastic forecasting and modeling of volatility of oil prices in Ghana using ARIMA time series model. European Journal of Business and Management, 4(16), 122-131.

Aladag, C. H., Egrioglu, E., & Kadilar, C. (2009). Forecasting nonlinear time series with a hybrid methodology. *Applied Mathematics Letters*, *22*(9), 1467-1470.

Araujo, G. S., & Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. Latin American Journal of Central Banking, 4(2), 100087.

Aronsson, L., Andersson, R., & Ansari, D. (2021). Artificial neural networks versus LASSO regression for the prediction of long-term survival after surgery for invasive IPMN of the pancreas. *Plos one*, *16*(3), e0249206.

Arsovski, L., & Gramatikov, R. D. (2023). Inflation Predictions in 2024: An Economic Analysis: The Inflation Process And Expectations In Republic Of North Macedonia: An Economic Analysis. Vizione, (41).

Balami, G. (2006). Innovation and Growth in Developing Countries. Journal of Development Economics, July; Growth in Developing Countries, Journal of Economic Literature, September: Growth in the Global Economy. Cambridge, Mass: MIT Press, 33(3).

Baybuza, I. (2018). Inflation forecasting using machine learning methods. Russian Journal of Money and Finance, 77(4), 42-59.

Bedada, T. D., Demissie, W. M., & Wolde, E. T. (2020). Determinants of inflationary experience in Ethiopia. Journal of Economics and Financial Analysis, 4(1), 15-54.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

Brockwell, P. J., Davis, R. A., Brockwell, P. J., & Davis, R. A. (2016). ARMA models. *Introduction to Time Series and Forecasting*, 73-96.

Bruckmeier, D. (2022). Forecasting Volatility in Financial Time Series (Doctoral dissertation).

Casdagli, M. (1992, December). Nonlinear forecasting, chaos and statistics In Modeling Complex Phenomena: Proceedings of the Third Woodward Conference, San Jose State University, April 12–13, 1991 (pp. 131-152). New York, NY: Springer New York.

Cassel, G. (1937). Keynes General Theory. Int'l Lab. Rev., 36, 437.

Cerqueira, V., Torgo, L., Pinto, F., & Soares, C. (2017). Arbitrated ensemble for time series forecasting. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part II 10* (pp. 478-494). Springer International Publishing.

Cerulli, Drago, (2021) G. An introduction to machine learning and text mining for economists using Stata, Python and R.

Chakraborty, C., & Joseph, A. (2017). Machine learning at central banks.

Cui, P., & Athey, S. (2022). Stable learning establishes some common ground between causal inference and machine learning. Nature Machine Intelligence, 4(2), 110-115.

Cukierman, A., & Gerlach, S. (2003). The inflation bias revisited: theory and some international evidence. The Manchester School, 71(5), 541-565.Economic Modelling, 22(5), 848-867

Emeru, G. M. (2020). The determinants of inflation in Ethiopia: A multivariate time series analysis. Journal of Economics and Sustainable Development, 11(21), 53-62.

Feldman, M., & Lowe, N. (2017). Evidence-based economic development policy. Innovations: Technology, Governance, Globalization, 11(3-4), 34-49.

Girijasankar, M., & Anis, C. (2001). Inflation and Economic Growth: Evidence from South Asian Countries, university of western Sydney. Hagos, M. K. (2014). Perception of inflation and effect of inflation on the salariat in ethiopia. Research Journal of Finance and Accounting.

Gurney, K. (2018). *An introduction to neural networks*. CRC press.

Hakizimana, C. (2022). Forecasting inflation of Rwanda using macroeconomics variables (2006Q1-2019Q4) (Doctoral dissertation, University of Rwanda).

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, *64*(1-2), 307-333.

Hansen, C., & Liao, Y. (2019). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. Econometric Theory, 35(3), 465-509.

Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint arXiv:1707.08692*.

HELE, B. K., Temesgen, A., & Abebe, A. (2022). Statistical Analysis Of Inflation Rate In Ethiopia: Threshold Auto Regressive Approach (Doctoral dissertation, Haramaya University, Haramay).

Jain, A., Sukhdeve, T., Gadia, H., Sahu, S. P., & Verma, S. (2021, March). Covid19 prediction using time series analysis. In 2021 International conference on artificial intelligence and smart systems (ICAIS) (pp. 1599-1606). IEEE.

Jhingan, M. L.(2009.) 13edition. Advanced Economic Theory (VRINDA PUBLICATIONS (P) LTD). 921-933

Jinghan, M. (1997). Monetary economics (4th ed.). Virinda India: Virinda Publication Press.

Jose, J., Gires, A., Tchiguirinskaia, I., Roustan, Y., & Schertzer, D. (2022). Scale invariant relationship between rainfall kinetic energy and intensity in Paris region: An evaluation using universal multifractal framework. *Journal of Hydrology*, *609*, 127715.

Joseph, V. R., & Vakayil, A. (2021). SPlit: An Optimal Method for Data Splitting. Technometrics, 64(2), 166–176. https://doi.org/10.1080/00401706.2021.1921037.

Kagy, G., & Negatu, D. (2023). The Coronavirus Pandemic: Ethiopia. The Coronavirus Pan- demic and Inequality: A Global Perspective, 131-140.

Kahssay, T. (2017). Determinants of inflation in Ethiopia: A time-series analysis. Journal of Economics and Sustainable Development, 8(19), 1-6.

Kebede, G. M. (2020). The Dynamics of Inflation in Ethiopia: Time Series Approach.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied soft computing*, *11*(2), 2664-2675.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied soft computing*, *11*(2), 2664-2675.

Kothari, J., & VB, A. (2024). Treasury Bond Price Prediction using Time Series and Sentiment Analysis. *Grenze International Journal of Engineering & Technology (GIJET)*, *10*.

Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, *41*(9), 4235-4244.

Kousta, T., & Bellet, C. S. Local Interpretable Model-agnostic Explanations for Long Short Term Memory Network used for Classification of Amazon Customer Reviews.

Kuleyin, A., Gök, A., Eroğlu, H. A., Özkaraova, E. B., Akbal, F., Jada, A., & Duply, J. (2022). Combining Electro-Fenton and adsorption processes for reclamation of textile industry wastewater and modeling by Artificial neural Networks. Journal of Electroanalytical Chemistry, 921, 116652.

Latin American Journal of Central Banking, 1(1-4), 100012.

Liu, M., Zeng, A., Xu, Z., Lai, Q., & Xu, Q. (2021). Time series is a special sequence: Forecasting with sample convolution and interaction. arXiv preprint arXiv:2106.09305, 1(9)

Lowe, A. (2017). On economic knowledge: toward a science of political economics.

Mao, S., & Xiao, F. (2019). Time series forecasting based on complex network analysis. IEEE Access, 7, 40220-40229.

McAdam, P., & McNelis, P. (2005). Forecasting inflation with thick models and neural networks.

Melaku, M. (2020). Determinants of inflation in Ethiopia. Academic Journal of Research and Scientific Publishing, 2(18).

Mihretu, M. E. (2023). Inflation on Forecasting Models and Forecasting Combination Analysis: TheCase Of Ethiopia

Mirza, N., Rizvi, S. K. A., Naqvi, B., & Umar, M. (2024). Inflation prediction in emerging economies: Machine learning and FX reserves integration for enhanced forecasting. International Review of Financial Analysis, 94, 103238.

Moghaddamnia, A., Remesan, R., Kashani, M. H., Mohammadi, M., Han, D., & Piri, J. (2009). Comparison of LLR, MLP, Elman, NNARX and ANFIS Models—with a case study in solar radiation estimation. *Journal of Atmospheric and Solar-Terrestrial Physics, 71*(8-9), 975-982.

Mwangi, C. N. W. (2016). Artificial neural network model for inflation forecasting in Kenya(Thesis).

Nakorji, M., & Aminu, U. (2022). Forecasting inflation using machine learning techniques. The Review of Finance and Banking, 14(1), 45-55.

Nakorji, M., & Aminu, U. (2022). Forecasting inflation using machine learning techniques. The Review of Finance and Banking, 14(1), 45-55.

Ngai, E. W., & Wu, Y. (2022). Machine learning in marketing: A literature review, conceptual framework, and research agenda. *Journal of Business Research, 145*, 35-48.

Nigusse, T., Tadesse, T., & Melaku, T. (2019). Supply and demand side determinants of inflation in Ethiopia: Auto-regressive distributed lag model (ARDL). International Journal of Commerce and Finance, 5(2), 8-21.

Nunoo, E., & Kings, I. (2013). Inflation Forecasting in Ghana-Artificial Neural Network Model Approach (Doctoral dissertation).

Nyoni, T., & Mutongi, C. (2019). Modeling the long-run relationship between inflation and eco- nomic growth in Zimbabwe: a bi-variate cointegration (Engle-Granger Two-Step) approach.

Oyewale, A. M., Kasali, A. O., Phazamile, K., Abiodun, M. V., & Adeyinka, A. I. (2019). Fore- casting Inflation Rates Using Artificial Neural Networks. International Journal of Statistics and Applications, 9(6), 201-207

Popoola, Y. (2023). Machine Learning Approaches to Macroeconomic Forecasting; An Application toSwedish Inflation.

Prakash, K., & Maiti, D. (2016). Does devaluation improve trade balance in small island economies? The case of Fiji. Economic Modelling, 55, 382-393.

Purbasari, I. Y., Anggraeny, F. T., & Ardiningrum, N. A. (2020). Time-series Modeling for Consumer Price Index Forecasting using Comparison Analysis of Autoregressive Integrated Moving Average and Artificial Neural Network. In Proceedings of the International Conference on

Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies (CESIT 2020) (pp. 599-604).

Rich, R. W., & Steindel, C. (2005). A review of core inflation and an evaluation of its measures. *FRB of New York Staff Report*, (236).

Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with machine learning methods.

Routledge. Mahajan, K., & Srinivasan, A. (2020). Inflation Forecasting in Emerging Markets: A Machine Learning Approach.

Ruiz, L. G. B., Cuéllar, M. P., Calvo-Flores, M. D., & Jiménez, M. D. C. P. (2016). An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies*, *9*(9), 684.

Samoah, G., Ofori, J. & Ayikpah, A.A.Y. (2021). Forecasting Inflation Rates in Ghana using Regression, Artificial Neural Networks and Support Vector Machines. Journal of Technology, Engineering, Management &Computational Sciences, Vol. 1, No.1, pp.56-128.

Samuel, A. L. (1959). Machine learning. The Technology Review, 62(1), 42-45.

Santos, A. R. G. N. V. (2019). Forecasting Inflation in Portugal by Using Machine Learning Techniques (Doctoral dissertation, Universidade de Lisboa (Portugal)).

Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests.

Sharma, D. K., Hota, H. S., Brown, K., & Handa, R. (2022). Integration of genetic algorithm with artificial neural network for stock market forecasting. *International Journal of System Assurance Engineering and Management*, *13*(Suppl 2), 828-841

Sharma, R., Shih, R., Patel, C., & Sontag, J. (2007, January). Application of exploratory data analysis (eda) techniques to temperature data in a conventional data center. In *International Electronic Packaging Technical Conference and Exhibition* (Vol. 42770, pp. 851-857).

Siven, C. H. (2021). Bent Hansen's theory of inflation 19511. The European Journal of the History of Economic Thought, 28(2), 249-272

Sorjamaa, A., & Lendasse, A. (2006, April). Time series prediction using direct strategy. In Esann (Vol. 6, pp. 143-148)

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast? Journal of Money, Credit and banking, 39, 3-33.

Tafere, Kibrom. (2008). The sources of the recent inflationary experience in Ethiopia. Addis Ababa University.

Teräsvirta, T., Van Dijk, D., & Medeiros, M. C. (2005). Linear models, smooth transition autoregression, and neural networks for forecasting macroeconomic time series: A re-examination. International Journal of Forecasting, 21(4), 755-774.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical

Society Series B: Statistical Methodology, 58(1), 267-288.

Tikhonov, A. N. (1963). On the solution of ill-posed problems and the method of regularization. In *Doklady akademii nauk* (Vol. 151, No. 3, pp. 501-504). Russian Academy of Sciences.

Tong, H. (1990). Non-linear time series: a dynamical system approach. Oxford University Press.

Totonchi, J. (2015). Macroeconomic theory of inflation. In International Conference on Economics and Finance Research IPEDR (Vol. 4).

Yellen, J. L. (2017). Inflation, uncertainty, and monetary policy. Business Economics, 52, 194-207.

Zhang, C., Li, H., Wang, X., & Yang, X. (2015). Cross-scene crowd counting via deep convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 833-841).

Zhou, K., Wang, W. Y., Hu, T., & Wu, C. H. (2020, September). Comparison of time series forecasting based on statistical ARIMA model and LSTM with attention mechanism. In Journal of physics: conference series (Vol. 1631, No. 1, p. 012141). IOP Publishing

APPENDENCES

Appendix A: Variables Description

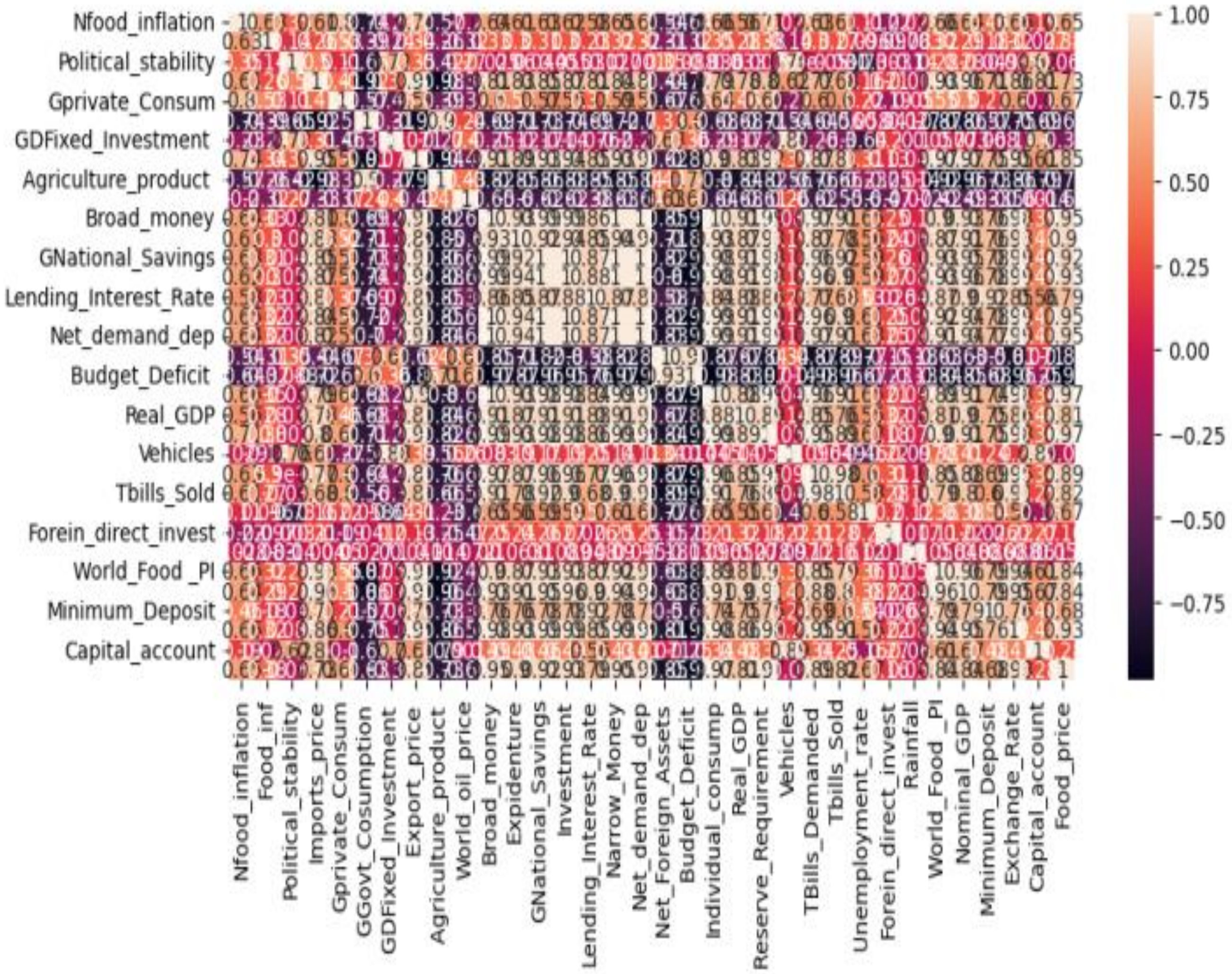
Table Appendix A: Dependent and Independent Variables Description

No	Variable Name	Measurement	Source
1	Headline Inflation (dependent variable)	In %	CSS
2	Political stability index	Number	World bank
3	Food inflation	In %	CSS
4	Non-food Inflation	in %	CSS
5	Price export of good and service	In million birr	NBE
6	Price import of good and service	In million birr	NBE
7	Gross government Consumption	In thousand birr	NBE
8	Gross private Consumption	In thousand birr	NBE
9	Minimum deposit	In thousand birr	NBE
10	Gross Domestic Fixed Investment	In thousand birr	NBE
11	Agricultural production Price	In Million birr	CSS
12	Broad Money Supply	In thousand birr	NBE
13	Expenditure	In Million birr	NBE
14	Narrow money	In thousand birr	NBE
15	Individual consumption	In thousand birr	NBE
16	Real GDP Per capital	In thousand birr	NBE
17	Reserve Requirement	In thousand birr	NBE
18	Numbers of Vehicles	In million (number)	NBE
19	T-Bill demand	In thousand birr	NBE
20	T-Bill sales	In thousand birr	NBE
21	Unemployment rate	In %	NBE
22	Foreign direct investment	In million birr	NBE
23	Rainfall	In milliliter	Ethiopia MA
24	World oil price	In thousand birr	World bank
25	Nominal GDP	In thousand birr	NBE
26	Exchange rate	In %	NBE
27	Capital account	In thousand birr	NBE

28	Imported food price	In million birr	NBE
29	Net demand deposit	In thousand birr	NBE
30	Gross national saving	In thousand birr	NBE
31	Investment	In thousand birr	NBE
31	Lending interest rate	In %	NBE
33	Net Foreign asset	In thousand birr	NBE
34	World food price index	In Number	World bank
35	Private Consumption	In thousand birr	NBE

Appendix B: Correlation

Table Appendix B: Correlation of Exogenous Variables



Appendix C: Features Selected by Machine Learning Models

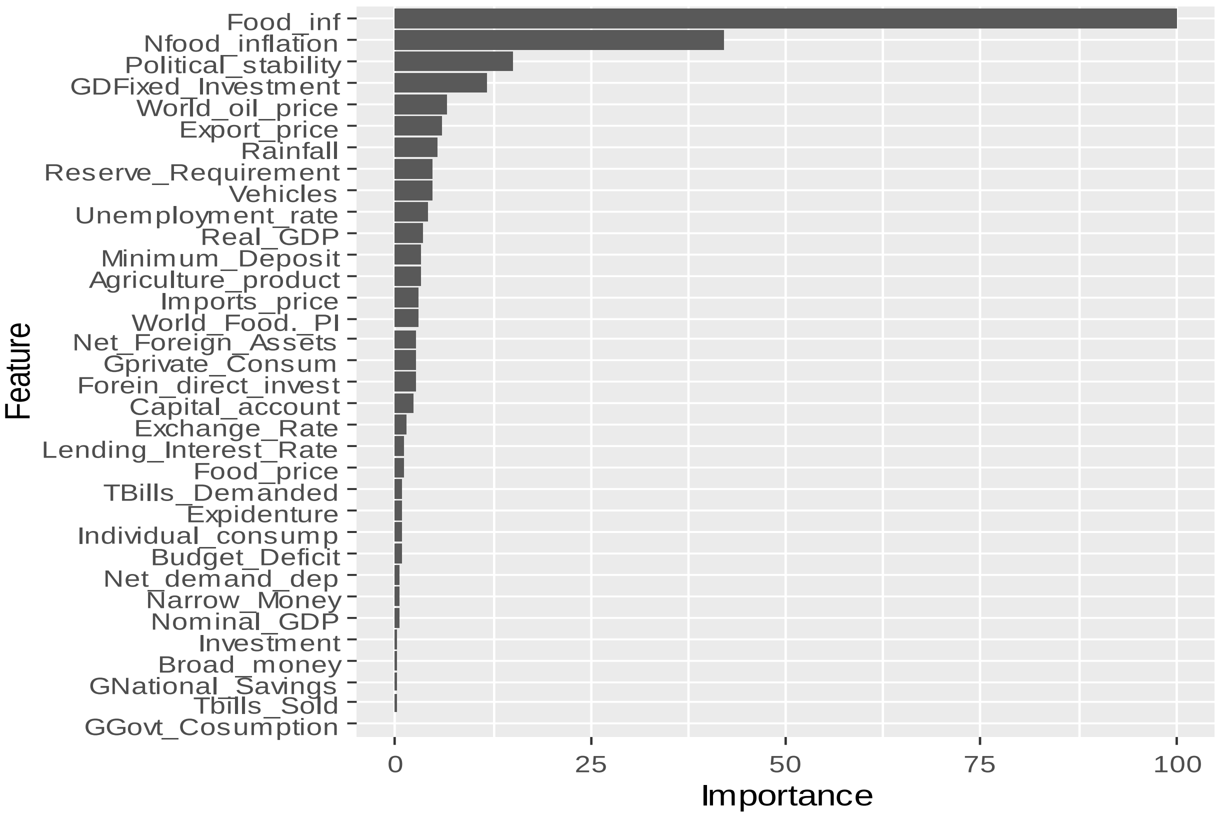


Figure Appendix C_1: Diagram of Features selected by Ridge Regression

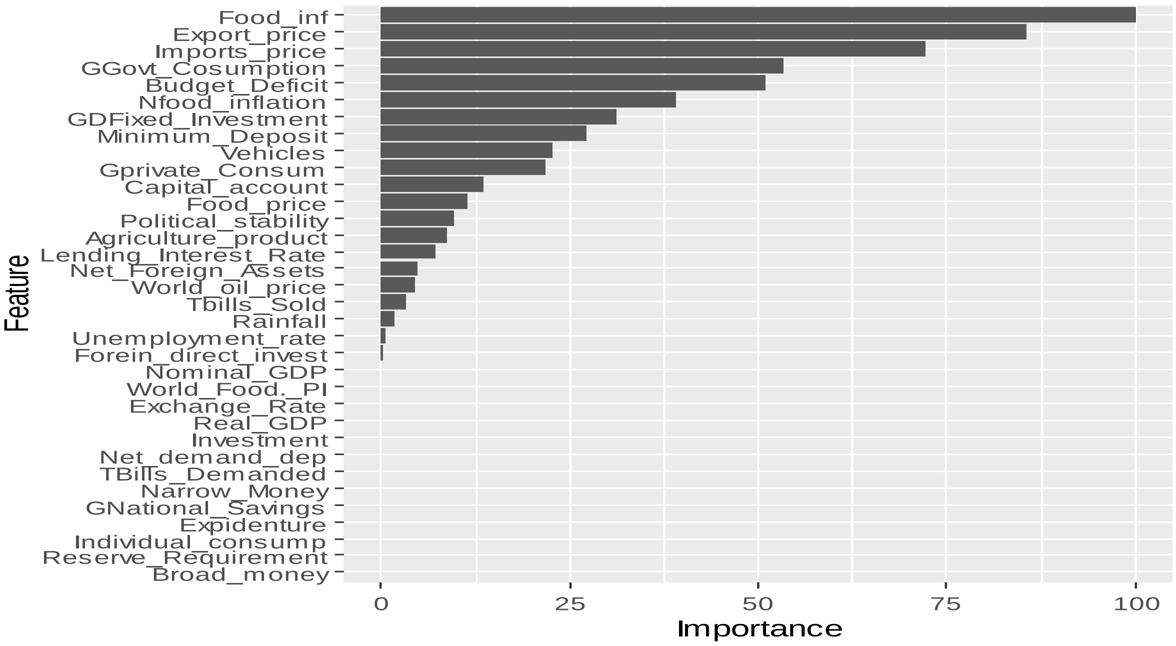


Figure Appendix C_2: Diagram of Features selected by LASSO model

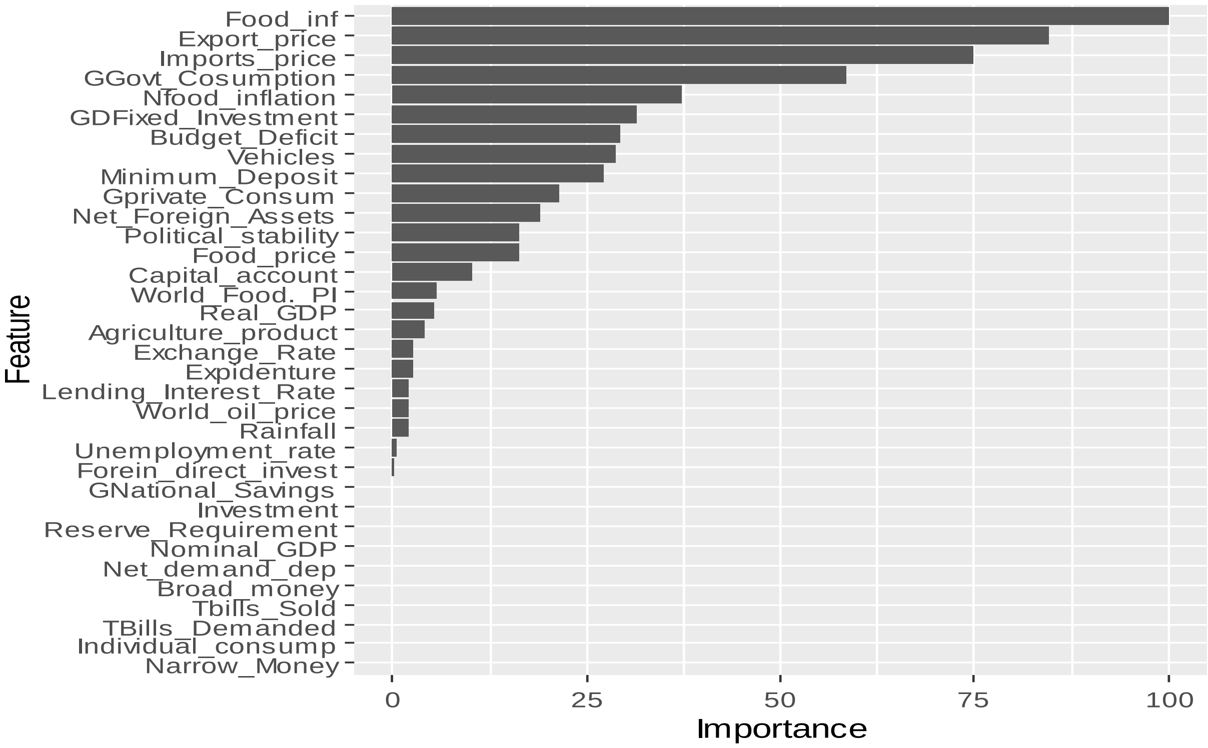


Figure Appendix C_3**:** Diagram of Features selected by Elastic Net Model

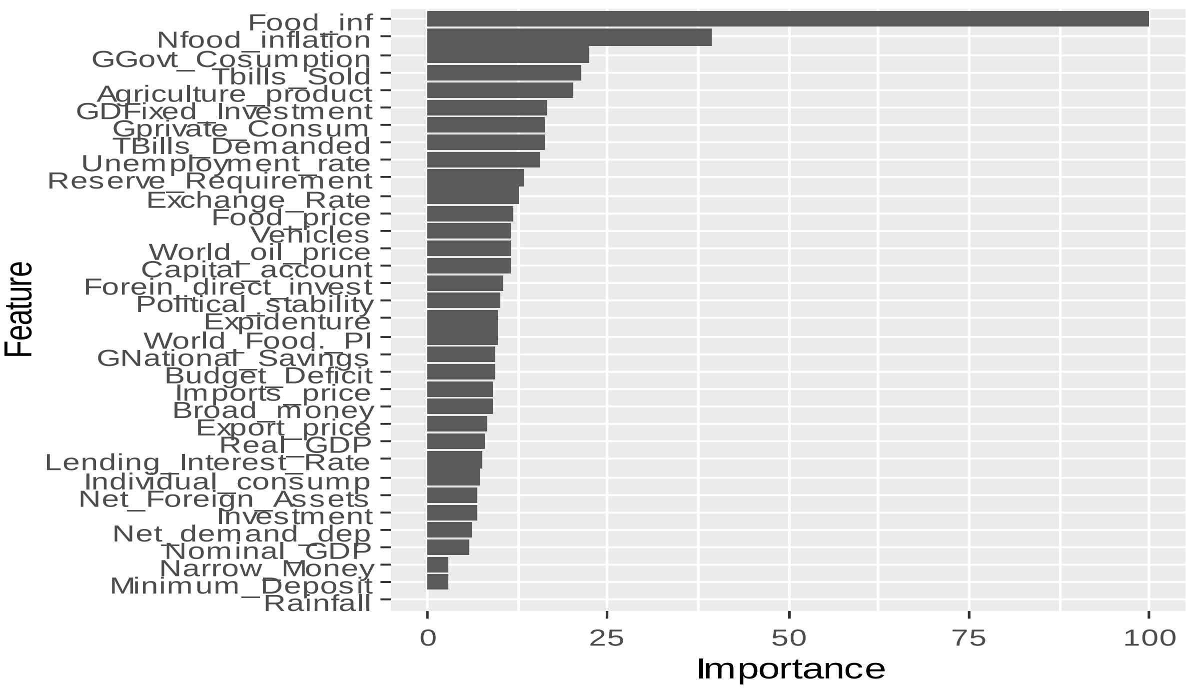


Figure Appendix C_4: Diagram of Features selected by Random Forest

Appendix D: Non-Linear Models of Machine Learning Models

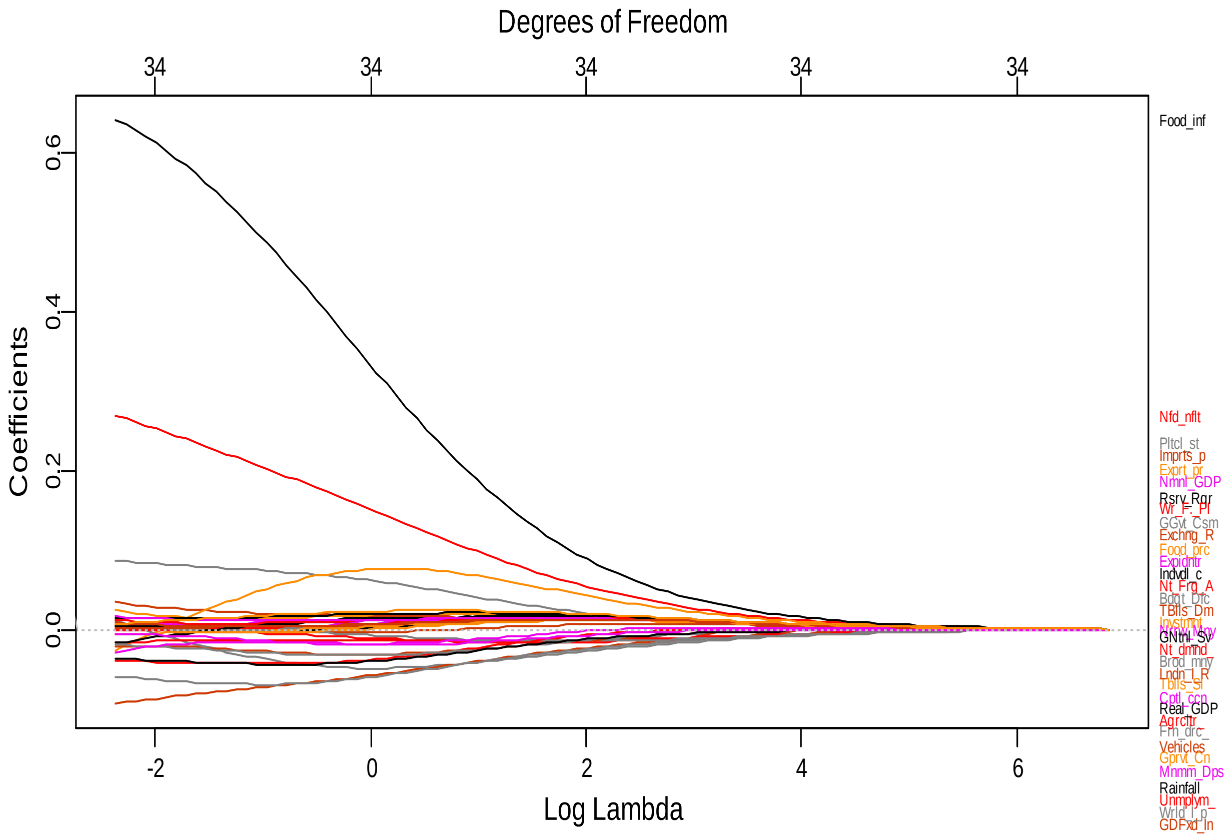


Figure Appendix D_1: Coefficients versus log(λ) Ridge regression

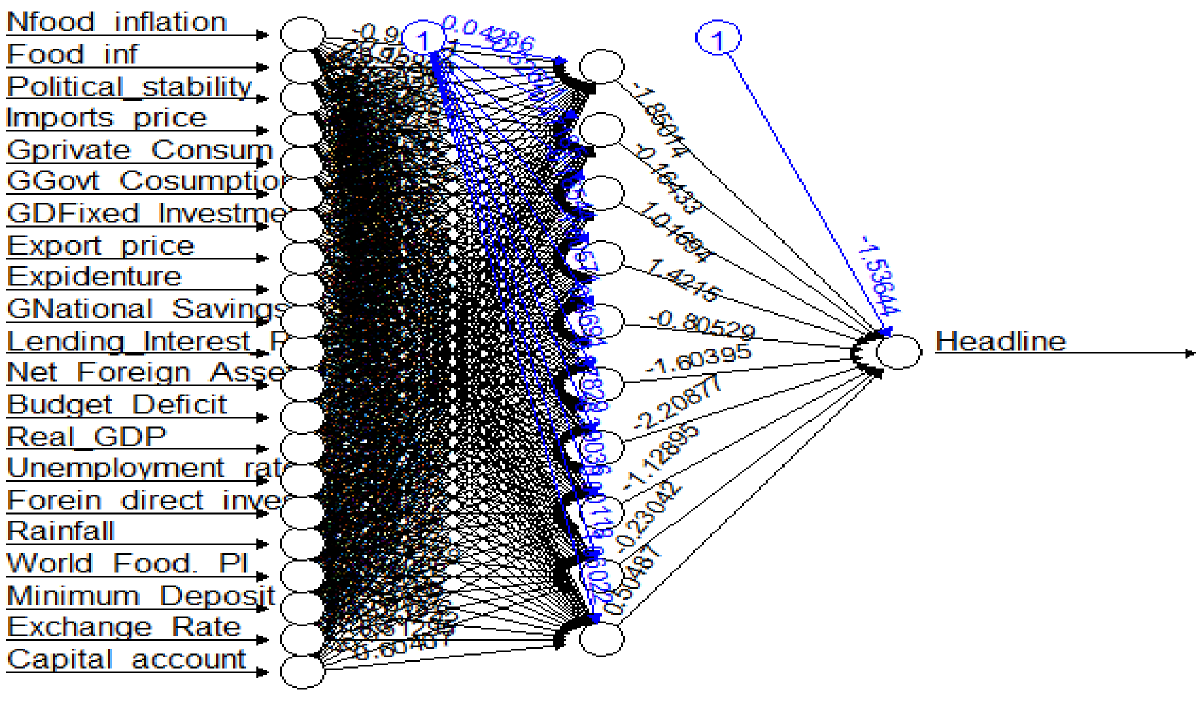


Figure Appendix D_2: Multi-Layer Perceptron (MLP) 22-10-1 network structure

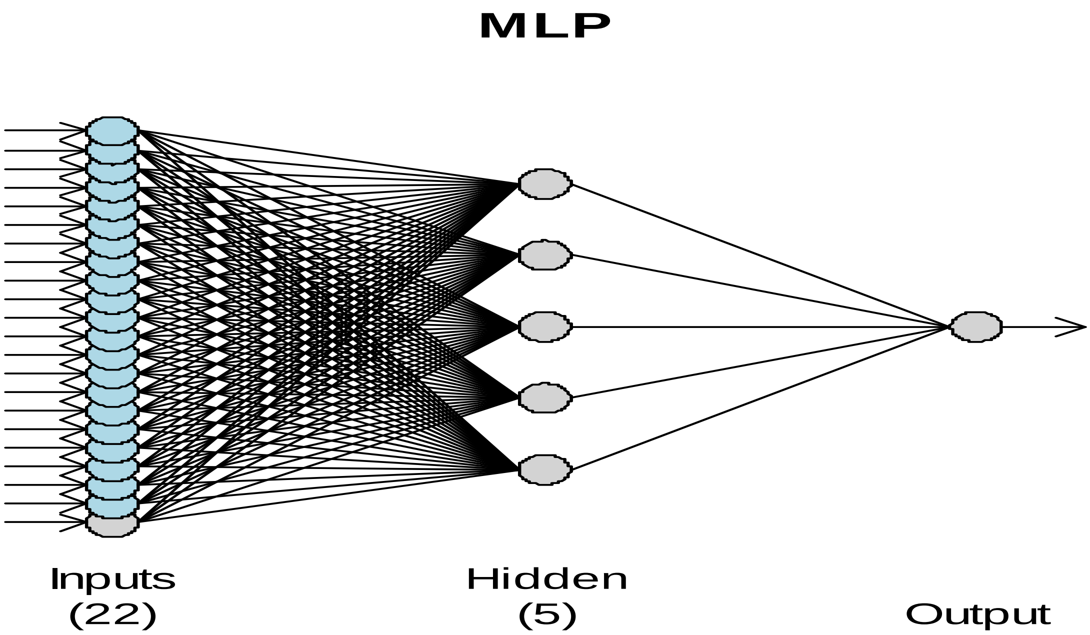


Figure Appendix $D_{3:}$ Structure Of best fitted ANN MLP model

The sky bright ones are deterministic inputs (shows seasonality in this case) While the grey input nodes are auto regressions or lags of headline inflation. The figure indicates that the resulting network has 5 hidden nodes, it was trained 20 times and the different forecasts were combining using median operator.

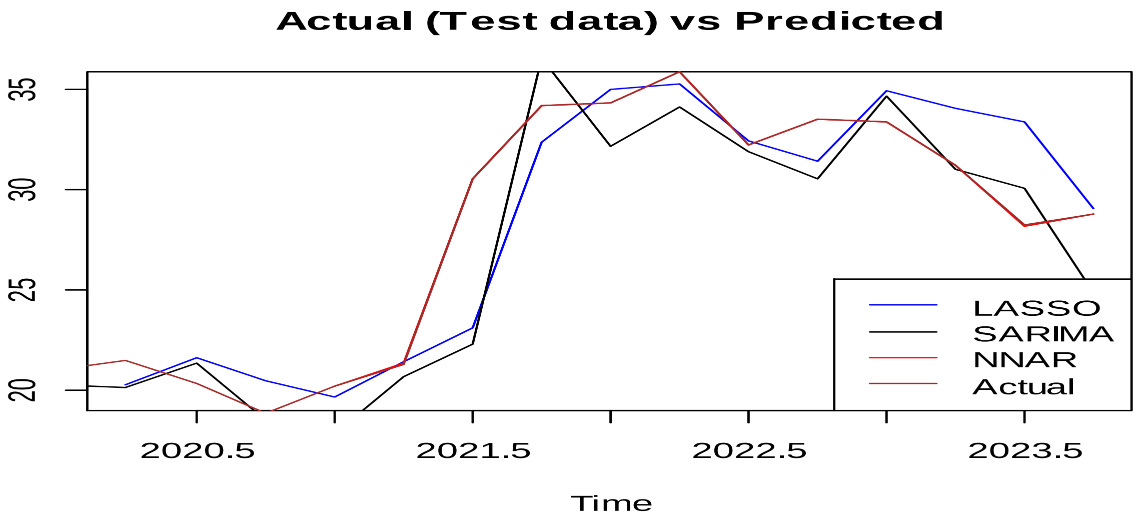


Figure Appendix D_4: Actual versus prediction of some selected Models

The out-sample prediction of NNAR model fitted line was overlap actual values of test data. Compared to SARIMA model LASSO model out-sample prediction was approach to the true values of the test data set.

Model Diagnostic of ANN Model

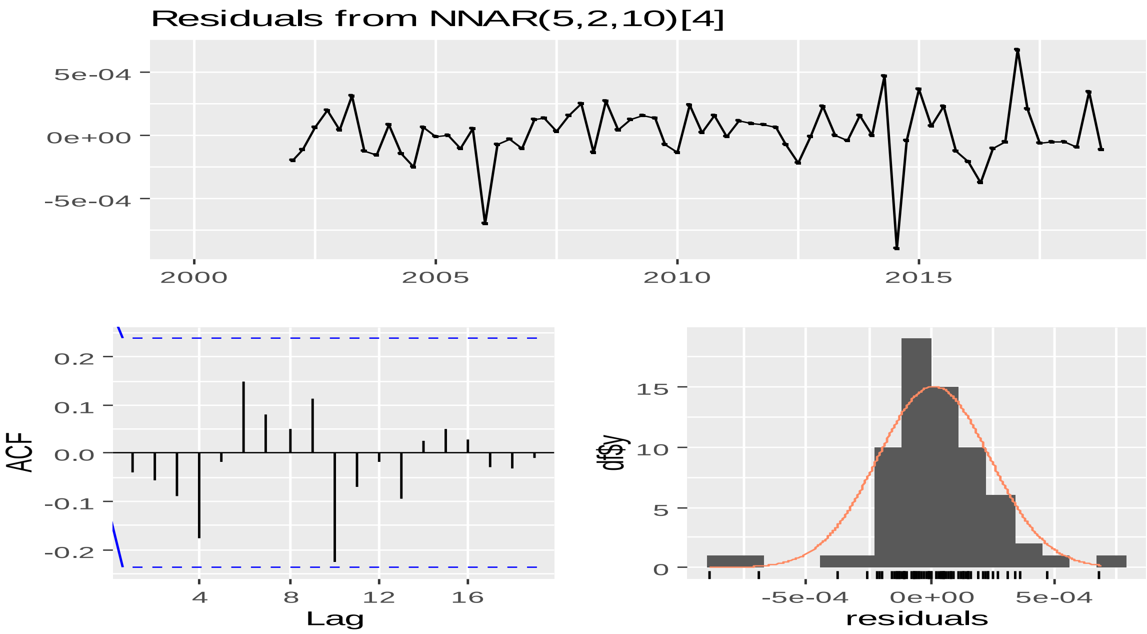


Figure Appendix D5: Residual diagnostic of in-sample forecasting of NNAR (5, 2, 10) with exogenous variable

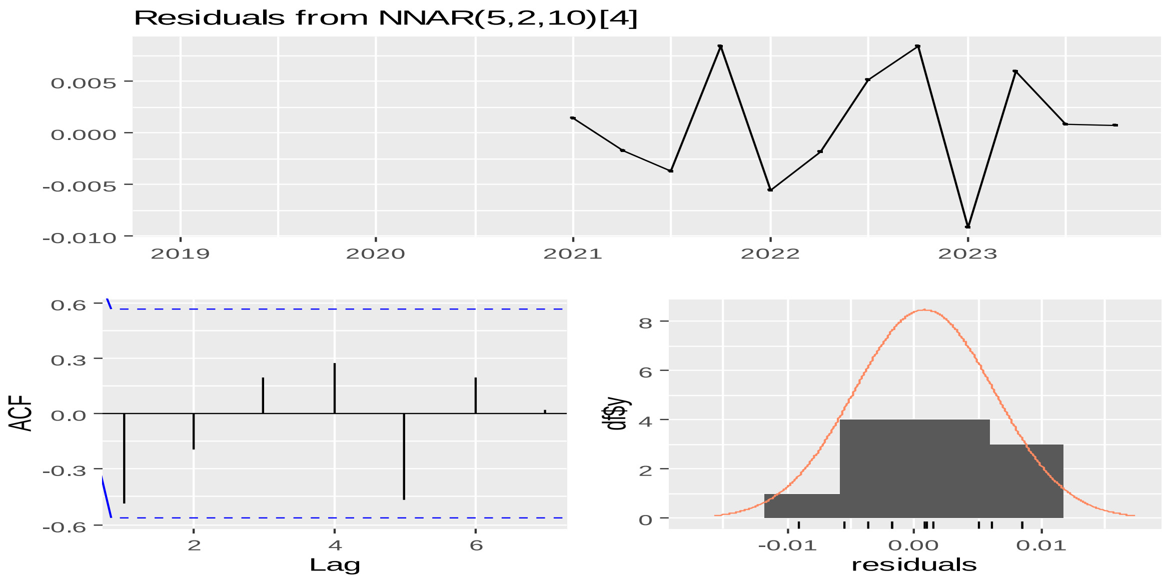


Figure Appendix D6: Residual diagnostic of out-sample forecasting of NNAR (5, 2, 10) with exogenous Variable

Forecast Values of ANN model

Table Appendix D1: Forecast values of NNARX (5,2,10) model

	Qtr1	Qtr2	Qtr3	Qtr4
2024	22.244	22.870	21.794	22.87531
2025	22.266	25.324	21.763	21.893
2026	20.442	21.826	30.464	34.086
2027	34.251	36.136	32.360	33.775
2028	33.439	31.218	28.179	28.776