Inflation Forecasting in Pakistan

Advanced Statistics

Submission Date: 4th May 2025

Project Team

Dania Waseem	23I-2622
Tabia Tariq	23I-2618

Department of Data Science

National University of Computer and Emerging Sciences
Islamabad, Pakistan

,

Contents

List of Figures

List of Tables

Inflation forecasting in Pakistan

0.1 Introduction

0.1.1 What is inflation?

Prices throughout whole economies tend to increase persistently during periods of time which defines inflation. Money loses purchasing ability when inflation occurs therefore each unit of currency obtains lower quantities of products and services. The Consumer Price Index (CPI) and Wholesale Price Index (WPI) serve as common tools to monitor inflation rates in economies. The Pakistani economy heavily depends on inflation because it impacts economic stability while shaping interest rates together with investment patterns and public living conditions. The accurate prediction of inflation remains essential for public administrators as well as business operators and consumer populations who need to base their choices in well-informed ways. The research applies ARIMA alongside Ridge Regression, LASSO and Elastic Net machines to process Pakistani historical data for forecasting future inflation patterns.

0.1.2 How to Identify Inflation and When It Becomes Dangerous?

When the prices of basic items such as food and fuel together with clothing and transport and housing escalate continuously over time it indicates that a specific country experiences inflation. The rising cost of basic items requires citizens to allocate a bigger amount of their budget to maintain their standard of purchasing power. The increasing cost of a loaf of bread from 50 rupees to 70 rupees last year until today represents rising inflation. Economists make use of Consumer Price Index (CPI) and Wholesale Price Index (WPI) to track price trends over time because they present average price data. The continual increase in these index numbers indicates inflation is taking place. Small levels of inflation exist naturally in economies since they stimulate spending and new investments. Too much inflation develops serious issues for the economy. Inflation devalues money possessions and damages savings while increasing living expenses which creates difficulties for low and middle-class households to buy necessities. Economists label inflation as hazardous when it persists at 6% to 7% above natural levels. The emergence of hyperinflation occurs with inflation rates exceeding 10% thus creating a state of economic malfunction which threatens market stability. The monitoring of inflation is essential because proper steps must be implemented when inflation reaches dangerous heights.

0.1.3 What is inflation forecasting?

Forecasters predict changes in economic goods and service pricing through the process of inflation forecasting. The forecast indicates future inflation possibilities for economists as well as governments and businesses to make informed decisions. Experts apply mathematical and statistical models to past data consisting of price trends and interest rates and money supply and economic growth in order to generate their predictions. The precision of inflation predictions remains essential because it allows central banks to establish interest rates and directs government policies and enables businesses and individuals to take sound financial decisions. A central bank would implement interest rate increases when predicting substantial inflation to keep it under control. This project employs ARIMA and Ridge and LASSO and Elastic Net models to predict future inflation trends based on past data points so Pakistan can prepare for economic difficulties.

0.1.4 Objectives of this report

This report consists of core purposes that are designed to fulfill specific objectives related to inflation forecasting in Pakistan.

1. To utilize multiple forecasting tools such as ARIMA, Ridge Regression, LASSO, and Elastic Net to predict inflation levels in Pakistan. The selected models will process past economic data to forecast future inflation patterns, providing insights into upcoming price movement tendencies.

2. To compare the performance of the selected models in terms of accuracy, reliability, and effectiveness. Developed models will undergo analysis to identify the one that presents the most effective method for Pakistani inflation forecasting.

3. To establish which variables hold the greatest effect on inflation levels in Pakistan. The study evaluates distinct model features to determine which economic variables—such as interest rates, exchange rates, and money supply measurements—specifically affect inflation.

0.2 Literature Review

Paper 1: *Inflation Forecasting Using Machine Learning Methods*

The research paper titled *"Inflation Forecasting Using Machine Learning Methods"* by Baybuza I., published in the *Russian Journal of Money and Finance* in 2018, focuses on enhancing inflation prediction using modern machine learning techniques. This study was conducted on the Russian economy and utilized various macroeconomic variables including consumer price indices, exchange rates, industrial production, and interest rates. The forecasting models applied in this research include **Boosting** and **Random Forest**, which are ensemble learning methods capable of capturing complex nonlinear relationships. The base years used for training the models span from 2003 to 2017. The study concludes that machine learning methods, especially ensemble models, offer more accurate and robust forecasts compared to traditional models, particularly during periods of economic instability or structural shifts.

Paper 2: *Phillips Curve Inflation Forecasts*

The paper titled *"Phillips Curve Inflation Forecasts"* by James H. Stock and Mark W. Watson explores the use of the Phillips Curve model to forecast inflation in the **United States**. This research analyzes the relationship between inflation and economic activity, focusing on variables such as unemployment, labor costs, and inflation expectations. The study compares multiple variations of the Phillips Curve, including those with expectations terms and real activity indicators. Models were trained using data primarily from the **1960s to the early 2000s**, and the paper was published in **2008**. The results show that while the Phillips Curve still has forecasting value, its accuracy varies over time, particularly during periods of structural economic change.

Paper 3: *Short-term Inflation Forecasting Models for Turkey and a Forecast Combination Analysis*

The research paper *"Short-term Inflation Forecasting Models for Turkey and a Forecast Combination Analysis"* analyzes various inflation forecasting models applied to the economy of **Turkey**. Published by authors including Fethi Öğünç and others, this study uses models such as **ARIMA**, **Factor Models**, and **VAR** (Vector Autoregression). The models are evaluated over the base period from **2005 to 2014**. The variables used include **CPI, exchange rates, output gap, import prices**, and other macroeconomic indicators. The paper finds that combining forecasts from multiple models improves prediction accuracy, especially in high-volatility environments.

Paper 4: *Forecasting China's Economic Growth and Inflation*

The paper titled *"Forecasting China's Economic Growth and Inflation"* by Patrick Higgins, Tao Zha, and Wenna Zhong focuses on macroeconomic forecasting in **China**. Published in **2022**, the study applies models such as **Bayesian Vector Autoregression (BVAR)** and **Factor-Augmented Models**. The forecasting models are built using data from **1995 to 2019**. Key variables include **GDP, CPI, industrial production, producer prices**, and monetary indicators. The paper concludes that models incorporating international factors and financial conditions outperform traditional domestic-only models.

Paper 5: *Evaluating Performance of Inflation Forecasting Models of Pakistan*

This paper, titled *"Evaluating Performance of Inflation Forecasting Models of Pakistan"*, was authored by Muhammad Nadim Hanif and Muhammad Jahanzeb Malik and published in **2015** in the SBP Research Bulletin. The study focuses on **Pakistan** and compares the effectiveness of models such as **ARIMA**, **ARFIMA**, and **State Space Models**. It covers data from **2001 to 2014**, using variables like **headline CPI, interest rates, monetary aggregates**, and **exchange rates**. The paper finds that no single model performs best in all conditions and highlights the importance of model selection based on economic context.

Paper 6: *Growth and Inflation Forecasts for Germany: A Panel-Based Assessment of Accuracy and Efficiency*

The study *"Growth and Inflation Forecasts for Germany: A Panel-Based Assessment of Accuracy and Efficiency"*, published in **2006**, examines professional forecasts in **Germany** using panel data methods. Authors assess the efficiency and accuracy of forecasts from different institutions using historical data from **1991 to 2005**. Key variables analyzed include **GDP growth, CPI inflation, industrial output**, and **survey-based expectations**. The paper uses **panel regression models** and concludes that although forecasts are unbiased on average, their efficiency varies over time and across institutions.

Paper 7: *Forecasting UK Consumer Price Inflation Using Inflation Forecasts*

This paper, titled *"Forecasting UK Consumer Price Inflation Using Inflation Forecasts"* by Hossein Hassani and Emmanuel Sirimal Silva, focuses on inflation prediction in the **United Kingdom**. The study utilizes a combination of **ex-ante forecast-based models** and traditional time series methods, applied on datasets spanning from **1993 to 2012**. Variables considered include **CPI, inflation expectations, economic sentiment indices**, and **lagged inflation rates**. The paper concludes that incorporating expert forecasts as explanatory inputs enhances predictive accuracy compared to purely data-driven models.

Paper 8: *Inflation Rate Determinants in Saudi Arabia: A Non-Linear ARDL Approach*

The study *"Inflation Rate Determinants in Saudi Arabia: A Non-Linear ARDL Approach"* by Abdulrahman A. Albahouth investigates inflation in **Saudi Arabia** using advanced econometric techniques. Published in **2025** in the journal *Sustainability*, the paper uses the **Non-Linear ARDL** model to capture asymmetries in the relationships. It is based on data from **1990 to 2022** and includes variables like **oil prices, money supply (M2), government spending, exchange rate**, and **interest rates**. The findings show that inflation in Saudi Arabia is more responsive to positive shocks in oil prices and money supply than to negative shocks.

Paper 9: *Short-Term Forecasting of Inflation in Bangladesh with Seasonal ARIMA Processes*

The paper *"Short-Term Forecasting of Inflation in Bangladesh with Seasonal ARIMA Processes"* by Tahsina Akhter applies time series modeling to inflation data from **Bangladesh**. Published in **2013**, this study uses **Seasonal ARIMA (SARIMA)** models trained on data from **2000 to 2012**. Variables include **monthly CPI**, **seasonal dummy variables**, and **trend terms**. The study concludes that SARIMA effectively captures seasonal and trend components in Bangladesh's inflation, making it suitable for short-term policy planning.

Paper 10: *Forecasting Inflation and Economic Growth of Pakistan Using Two Time Series Methods*

This study titled *"Forecasting Inflation and Economic Growth of Pakistan Using Two Time Series Methods"* by Kishwer Sultana et al. focuses on the **Pakistani** economy. The paper, published in **2022**, applies **ARIMA** and **Exponential Smoothing** models to inflation and GDP data from **2000 to 2020**. It utilizes variables such as **CPI, GDP, interest rate**, and **money supply**. The results show that both models can reasonably predict inflation trends, but ARIMA provides more accurate forecasts, especially during periods of economic shocks.

0.3 Description of Variables and Data

The dataset used in this study contains annual macroeconomic indicators for Pakistan from 1980 to 2025. The primary objective is to forecast the annual inflation rate (%) using various independent variables. Below is a detailed explanation of each variable included in the dataset.

- **Inflation (%)**: This is the dependent variable in our study. It represents the annual percentage change in the general price level of goods and services consumed in the economy.
- **Year**: Indicates the calendar year of the recorded data.
- **Population Growth (%)**: Measures the annual growth rate of the population.
- **Consumer Price Index (CPI)**: Reflects changes in the price level of a basket of consumer goods and services.
- **Trade (% of GDP)**: Represents the total trade (exports + imports) as a percentage of Gross Domestic Product.
- **Imports of Goods and Services (Annual % Growth)**: Annual growth rate of imported goods and services.
- **Gross National Expenditure (% of GDP)**: The total domestic and foreign investment and consumption expenditure as a percentage of GDP.
- **External Debt Stocks (% of GNI)**: Measures a country's external debt in relation to its Gross National Income.
- **Wholesale Price Index**: Reflects changes in the price level of goods at the wholesale level.

- **Food Production Index**: Index measuring changes in food production volume.
- **World Crude Oil Rates ($)**: Global average crude oil prices, a crucial input cost indicator.
- **Unemployment Rate (% of Total Labor Force)**: Indicates the percentage of the labor force that is without work but available and seeking employment.
- **Adolescent Fertility Rate**: The number of births per 1,000 women aged 15–19 years.
- **Age Dependency Ratio (% of Working-age Population)**: Ratio of dependents (younger than 15 or older than 64) to the working-age population.
- **Urban Population Growth (Annual %)**: Annual growth rate of the population living in urban areas.
- **Total Reserves (% of Total External Debt)**: Measures a country's reserves in relation to its external debt.
- **Exports of Goods and Services (Annual % Growth)**: Annual growth in exports.
- **Adjusted Savings: Gross Savings (% of GNI)**: Represents savings after accounting for education expenditures, depletion of natural resources, and pollution damages.

0.3.1 Training and Testing Data

For the purpose of this study, the dataset is divided into two parts:

- **Training Data (1980–2020):** This portion of the dataset is used to train machine learning models such as ARIMA, Ridge Regression, LASSO, and Elastic Net. The models learn patterns, trends, and relationships between the independent variables and the target variable, which is the annual inflation rate (%).
- **Testing Data (2021–2025):** This subset is used to evaluate the model's performance and accuracy in predicting inflation. It is important to ensure that all independent variables (e.g., CPI, oil prices, unemployment rate) are available for these years so that the models can generate valid forecasts. If any variable data is missing for the testing period, it must be estimated or forecasted using suitable methods before applying the trained model.

This separation ensures that the models are not overfitted and are capable of generalizing well to unseen data.

0.4 Model Estimation (Introduction and Equations)

In this section, we estimate different types of models to analyze the impact of macroeconomic variables on inflation. Each model serves a unique purpose depending on the nature of the data (time series, multicollinearity, overfitting, etc.). Below we describe each model, what it is used for, what problems it solves, and how it treats variables and model complexity.

0.4.1 1. ARIMA Model (Autoregressive Integrated Moving Average)

The ARIMA model is used to analyze and forecast univariate time series data. It captures autocorrelation patterns through autoregressive (AR) and moving average (MA) terms, while differencing (I) ensures stationarity.

ARIMA is a **time series model** used to forecast future values based on past trends, lags, and error terms. It is particularly effective when the dependent variable shows autocorrelation or trends over time. ARIMA helps in making the data stationary by differencing, which solves the problem of non-stationarity and autocorrelation. It models the relationship using:

The ARIMA model, which stands for **Autoregressive Integrated Moving Average**, is used for time series forecasting. It combines three elements:

- **AR (Autoregressive)**: Uses past values (lags) of the dependent variable to predict the current value.
- **I (Integrated)**: Applies differencing to make the series stationary (removes trends or seasonality).
- **MA (Moving Average)**: Uses past error terms to model the current value.

0.4.1.1 1. Autoregressive (AR) Component

The AR part regresses the current value of the time series on its past values. The AR(p) model is defined as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \tag{1}$$

Example: If we are forecasting inflation in year 2024 (Y_{2024}), then:

- β_0 is the base/intercept level of inflation.

- $\beta_1 Y_{2023}$ represents the influence of last year's inflation.
- $\beta_2 Y_{2022}$ captures inflation from two years ago.
- ε_t is the error term capturing unexplained shocks in 2024.

In simple terms, this means: *"Inflation this year depends on inflation in previous years, plus some random shock."*

0.4.1.2 2. Moving Average (MA) Component

The MA part models the current value as a function of previous error terms (shocks). The MA(q) model is:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{2}$$

Interpretation:

- μ is the mean of the series.
- ε_t is the current period's shock or error.
- $\theta_1 \varepsilon_{t-1}$ captures the impact of last year's forecasting error.

So for inflation in 2024, this equation says: *"The inflation in 2024 is based on this year's shock and previous years' forecast errors."*

0.4.1.3 3. Integrated (I) Component – Differencing

The "Integrated" part deals with making the series stationary by differencing. If the series is not stationary (i.e., has trends), we use first differencing:

$$\Delta Y_t = Y_t - Y_{t-1} \tag{3}$$

This is referred to as an **I(1)** process (Integrated of order 1).

Explanation: We do this to remove trends. Instead of modeling the actual inflation level, we model the *change* in inflation from one year to the next.

If differencing is applied once ($d = 1$), then we're modeling:

$$\Delta Y_{2024} = Y_{2024} - Y_{2023}$$

If the series is already stationary (no need for differencing), then $d = 0$ and this step is skipped.

0.4.1.4 4. Full ARIMA(p, d, q) Model

The complete ARIMA model combines all three components into one:

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t \tag{4}$$

Where:

- $\phi(B)$ is the autoregressive operator (lag terms),
- $(1-B)^d$ is the differencing operator (integration),
- $\theta(B)$ is the moving average operator (error lags),
- B is the backshift operator: $BY_t = Y_{t-1}$.

This model forecasts current inflation by combining:

- How past inflation values impact the present (AR),
- How past forecast errors affect the prediction (MA),
- And whether the data needs differencing to become stable (I).

0.4.2 2. Ridge Regression

Ridge Regression is a **regularized linear regression model** used when predictor variables are highly correlated (multicollinearity). In standard regression, multicollinearity can make coefficient estimates unstable and inflate their variances. Ridge adds an L_2 penalty to the loss function which reduces the size of all coefficients — giving **less weight to irrelevant variables and more to important ones**, but it does not eliminate any variables.

This model reduces overfitting, improves prediction accuracy, and is especially useful when we have more variables than observations or when variables are correlated.

0.4.2.1 Ridge Regression Equation

The Ridge objective function is:

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{5}$$

Where:

- y_i is the response variable (e.g., **Inflation** for year i).
- x_{ij} are the predictor variables (e.g., **CPI**, **Unemployment Rate**, **WPI**, etc.).
- β_0 is the intercept.
- β_j are the coefficients for the predictors.
- λ is the tuning (regularization) parameter that controls the amount of shrinkage.

0.4.2.2 Interpretation in This Project

- When predictors like **CPI**, **WPI**, and **Oil Prices** are highly correlated, traditional OLS estimates become unstable or overly sensitive to small changes in the data.
- Ridge Regression addresses this by shrinking the coefficients: large coefficients are penalized more heavily.
- It does **not eliminate** any variables (i.e., it keeps all predictors in the model), but shrinks irrelevant or less important variables' coefficients closer to zero.
- In this inflation forecasting context, Ridge allows us to include all potentially relevant economic indicators without worrying too much about multicollinearity.
- Unlike OLS, Ridge often gives more reliable and stable predictions, especially when the number of predictors is large or they are correlated.

0.4.3 3. Lasso Regression

Lasso Regression is another **regularization technique**, but unlike Ridge, it uses an L_1 penalty which forces some coefficients to become exactly zero. This means Lasso performs **variable selection by automatically removing less important variables**. This is useful when we suspect that many predictors have little or no effect on the dependent variable.

However, when predictors are highly correlated, Lasso tends to keep only one and discard the rest, sometimes randomly. It can lead to better interpretability but may lose useful information in correlated data.

0.4.3.1 Lasso Objective Function

The Lasso estimator solves the following optimization problem:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{6}$$

Where:

- y_i is the target variable (e.g., inflation rate in year i).
- x_{ij} are the predictor variables (e.g., CPI, WPI, Oil prices, etc.).
- β_j are the coefficients associated with each predictor.
- λ is the tuning (regularization) parameter.

0.4.3.2 Interpretation in This Project

- Lasso is beneficial when economic variables such as **CPI**, **Oil prices**, **Unemployment rate**, etc., are potentially highly correlated.
- Unlike Ridge Regression, Lasso can **shrink some coefficients exactly to zero**, thereby performing automatic **variable selection**.
- In the context of inflation modeling, this means Lasso helps identify the most relevant economic indicators, discarding the ones that do not contribute significantly.
- This reduces model complexity, increases interpretability, and often improves out-of-sample prediction.
- If $\lambda = 0$, the model behaves like ordinary least squares (OLS). If λ is very large, most coefficients are pushed to zero.
- This is helpful when we have many economic indicators, but not all are strong predictors of inflation.

0.4.3.3 Advantages

- Handles multicollinearity by shrinking correlated variable coefficients.
- Automatically removes irrelevant variables.
- Enhances prediction accuracy and interpretability.

0.4.4 4. Elastic Net Regression

Elastic Net Regression is a regularized regression method that linearly combines the penalties of Lasso (L1) and Ridge (L2) regression. It is especially useful in datasets where there are multiple correlated predictor variables and where feature selection is desired.

0.4.4.1 Elastic Net Objective Function

The Elastic Net estimates the regression coefficients by minimizing the following objective function:

$$\hat{\beta}^{enet} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\} \tag{7}$$

Alternatively, it is often represented using a mixing parameter α:

$$\hat{\beta}^{enet} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right] \right\} \tag{8}$$

Where:

- y_i is the response variable (e.g., inflation).
- x_{ij} are the economic predictor variables (e.g., CPI, WPI, oil prices, etc.).
- β_j are the regression coefficients.
- λ is the regularization parameter controlling the overall strength of the penalty.
- α is the mixing parameter:
  - $\alpha = 1$ corresponds to Lasso.
  - $\alpha = 0$ corresponds to Ridge.
  - $0 < \alpha < 1$ gives a mix of both.

0.4.4.2 Why Use Elastic Net in This Project?

- Some variables in the dataset (e.g., CPI and WPI) are highly correlated, which makes Lasso unstable.
- Elastic Net performs both **variable selection (via L1)** and **coefficient shrinkage (via L2)**, giving a more balanced model.

- It prevents the model from selecting just one among a group of correlated predictors — instead, it can retain grouped variables.
- This is beneficial in inflation modeling, where many macroeconomic indicators might influence inflation jointly.

0.4.4.3 Advantages

- Combines strengths of Ridge and Lasso.
- Effective when number of predictors p exceeds the number of observations n.
- Handles multicollinearity and helps in feature selection.
- Reduces the risk of overfitting by controlling model complexity.

0.5 Summary Statistics

0.5.1 Mean, Median, Mode, Quartiles of each variable

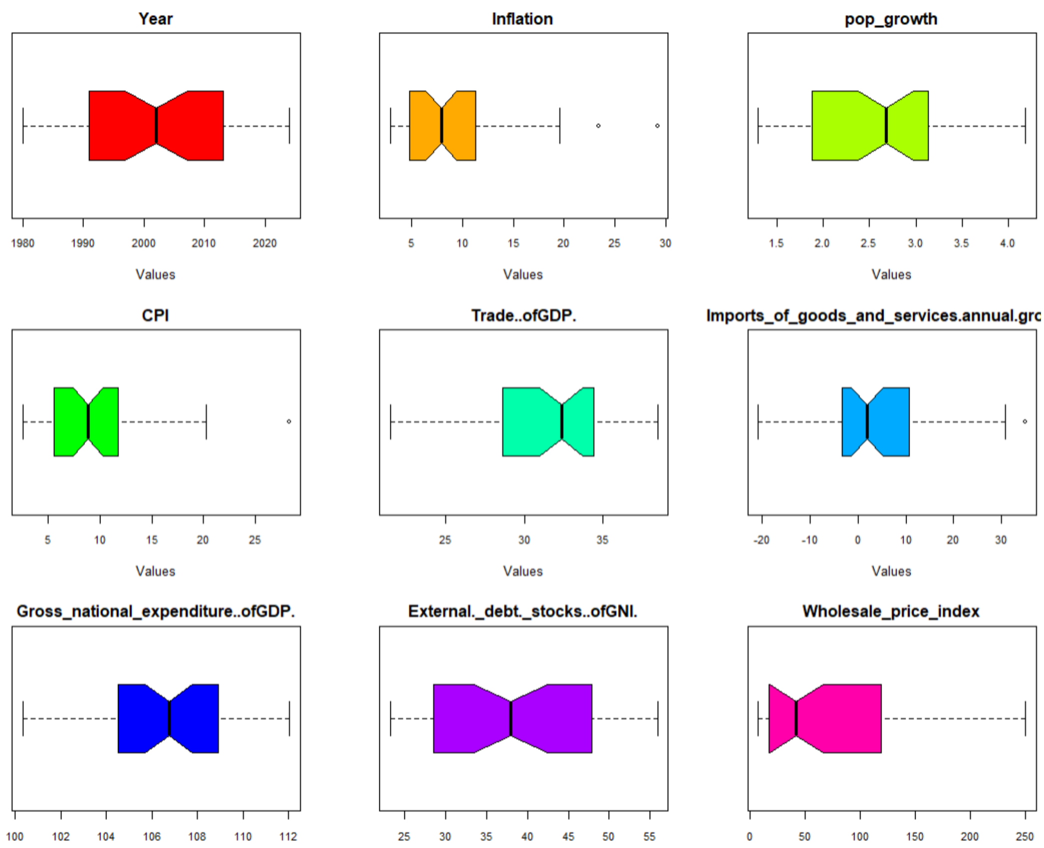
```
      Year        Inflation       pop_growth         CPI          Trade..ofGDP.
 Min.   :1980   Min.   : 2.90   Min.   :1.302   Min.   : 2.530   Min.   :21.46
 1st Qu.:1991   1st Qu.: 4.80   1st Qu.:1.887   1st Qu.: 5.610   1st Qu.:28.60
 Median :2002   Median : 7.90   Median :2.677   Median : 8.840   Median :32.36
 Mean   :2002   Mean   : 8.98   Mean   :2.578   Mean   : 8.993   Mean   :31.35
 3rd Qu.:2013   3rd Qu.:11.30   3rd Qu.:3.135   3rd Qu.:11.790   3rd Qu.:34.42
 Max.   :2024   Max.   :29.20   Max.   :4.178   Max.   :28.300   Max.   :38.50
 Imports_of_goods_and_services.annual.growth. Gross_national_expenditure..ofGDP. External._debt._stocks..ofGNI.
 Min.   :-20.892                              Min.   :100.3                      Min.   :23.28
 1st Qu.: -3.354                              1st Qu.:104.5                      1st Qu.:28.57
 Median :  1.992                              Median :106.7                      Median :38.01
 Mean   :  3.888                              Mean   :106.5                      Mean   :38.05
 3rd Qu.: 10.777                              3rd Qu.:108.9                      3rd Qu.:47.85
 Max.   : 35.051                              Max.   :112.0                      Max.   :55.90
 Wholesale_price_index Food_Production_index World_Crude_oil_rates... UnemploymentRate..oftotallaborforce.
 Min.   :  7.598       Min.   : 26.57        Min.   : 13.06           Min.   :3.10
 1st Qu.: 17.449       1st Qu.: 41.28        1st Qu.: 19.37           1st Qu.:4.70
 Median : 42.307       Median : 60.47        Median : 35.48           Median :5.90
 Mean   : 69.774       Mean   : 68.53        Mean   : 43.42           Mean   :5.51
 3rd Qu.:118.834       3rd Qu.: 98.04        3rd Qu.: 61.41           3rd Qu.:6.10
 Max.   :250.000       Max.   :125.88        Max.   :105.01           Max.   :8.30
 AdolescentFertilityRate AgeDependencyRatio..ofWorking.agePopulation. UrbanPopulationGrowth.annual..
 Min.   : 41.01          Min.   :69.40                                Min.   :1.878
 1st Qu.: 50.90          1st Qu.:74.62                                1st Qu.:2.624
 Median : 69.19          Median :85.54                                Median :3.328
 Mean   : 74.37          Mean   :82.34                                Mean   :3.323
 3rd Qu.:101.48          3rd Qu.:89.06                                3rd Qu.:3.970
 Max.   :110.80          Max.   :92.12                                Max.   :5.344
 TotalReserves..ofTotalExternalDebt. ExportsOfGoodsAndServices.annual.growth. AdjustedSavingsGrossSavings..ofGNI.
 Min.   : 4.380                      Min.   :-12.926                          Min.   : 6.30
 1st Qu.: 8.128                      1st Qu.: -1.354                          1st Qu.:14.90
 Median :13.754                      Median :  3.111                          Median :17.30
 Mean   :16.453                      Mean   :  6.466                          Mean   :17.15
 3rd Qu.:23.870                      3rd Qu.: 13.172                          3rd Qu.:19.28
 Max.   :37.081                      Max.   : 33.465                          Max.   :24.40
>
```

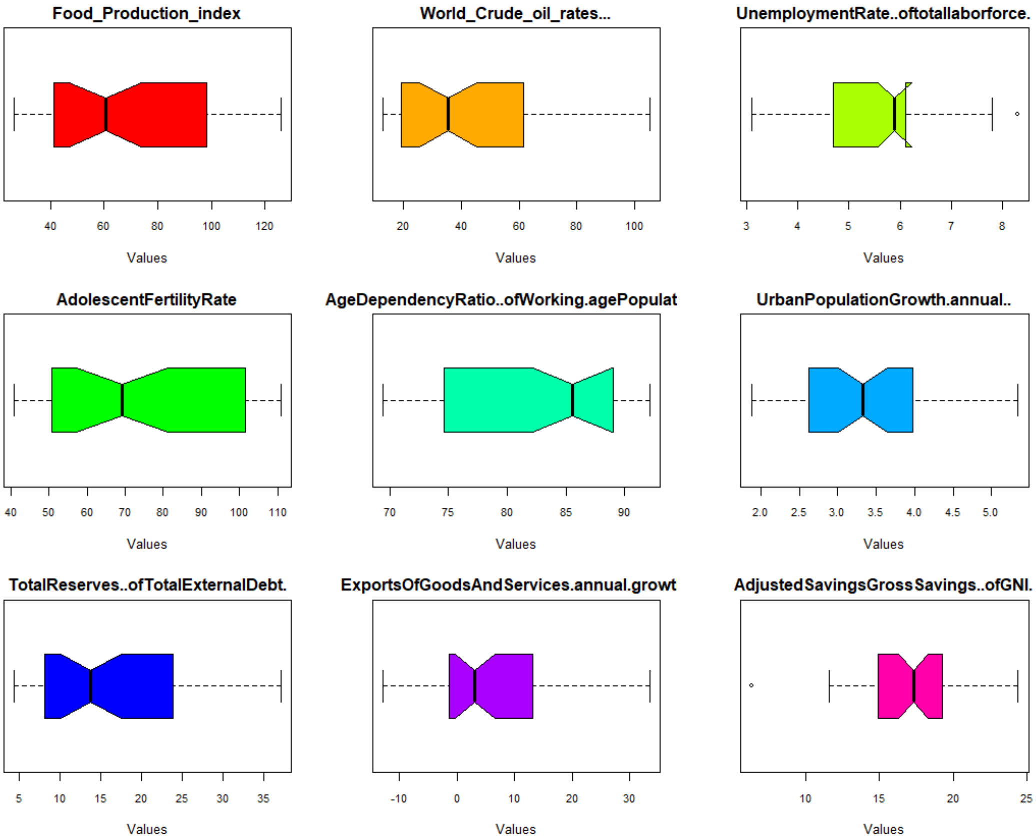
Figure 1: summary stats

Variable	Min	1st Qu.	Median	Mean	Max
Year	1980	1991	2002	2002	2024
Inflation	2.90	4.80	7.90	8.98	29.20
Population Growth	1.302	1.887	2.677	2.578	4.178
CPI	2.530	5.610	8.840	8.993	28.300
Trade (% of GDP)	21.46	28.60	32.36	31.35	38.50
Imports Growth	-20.89	-3.35	1.99	3.89	35.05
GNE (% of GDP)	100.3	104.5	106.7	106.5	112.0
External Debt (% of GNI)	23.28	28.57	38.01	38.05	55.90
WPI	7.60	17.45	42.31	69.77	250.00
Food Prod. Index	26.57	41.28	60.47	68.53	125.88
Oil Prices	13.06	19.37	35.48	43.42	105.01
Unemployment Rate	3.10	4.70	5.90	5.51	8.30
Adolescent Fertility Rate	41.01	50.90	69.19	74.37	110.80
Age Dependency Ratio	69.40	74.62	85.54	82.34	92.12
Urban Pop. Growth	1.878	2.624	3.328	3.323	5.344
Reserves (% of Debt)	4.380	8.128	13.754	16.453	37.081
Exports Growth	-12.93	-1.35	3.11	6.47	33.47
Gross Savings (% of GNI)	6.30	14.90	17.30	17.15	24.40

Table 1: Summary statistics for macroeconomic variables used in the analysis.

0.5.2 Box and whisker plots





- **Year:** Min 1985, Max 2024, Median 2002, Mean 2000, Outliers: None.

- **Inflation:** Min 3, Max 18, Median 7, Mean 17.5, Outliers: 23, 29, Skewness: Right-skewed.
  Interpretation: Most inflation values are moderate, but a few high years (including outliers like 23 and 29) drastically increase the mean.

- **Population Growth:** Min 1.3, Max 4.2, Median 2.7, Mean 2.7, Outliers: None, Skewness: Symmetrical.
  Interpretation: Population growth is stable and evenly distributed between 1.3% and 4.2%.

- **CPI:** Min 3, Max 21, Median 9, Mean 30, Outliers: 27, Skewness: Strongly Right-skewed.
  Interpretation: While most CPI values are moderate, one very high outlier (27) pulls the mean far above the median, indicating occasional price surges.

- **Trade (% of GDP):** Min 20, Max 40, Median 33, Mean 30, Outliers: None, Skewness: Left-skewed.
  Interpretation: Trade has mostly hovered around 33%, with a few lower years bringing the mean down.

- **Imports Growth (annual %):** Min -20, Max 30, Median 2, Mean 5, Outliers: 33, Skewness: Right-skewed.
  Interpretation: Though most years show moderate growth, some high spikes (like +33%) increase the mean.

- **Gross National Expenditure (% of GDP):** Min 100, Max 112, Median 106, Mean 106, Outliers: None, Skewness: Symmetrical.
  Interpretation: National expenditure is tightly clustered around 106%, showing a very consistent pattern.

- **External Debt Stocks (% of GNI):** Min 24, Max 56, Median 36, Mean 40, Outliers: None, Skewness: Right-skewed.
  Interpretation: Most values are moderate, but higher debt years pull the mean above the median.

- **Wholesale Price Index:** Min 5, Max 250, Median 40, Mean 125, Outliers: None, Skewness: Strongly Right-skewed.
  Interpretation: While most wholesale price index values are clustered at the lower end, extreme high values cause a major skew.

- **Food Production Index:** Min 35, Max 125, Median 60, Mean 80, Outliers: None, Skewness: Right-skewed.
  Interpretation: Most values are on the lower side (around 60), but some high values pull the mean up, suggesting occasional surges in food production.

- **World Crude Oil Rates:** Min 10, Max 105, Median 35, Mean 60, Outliers: None, Skewness: Right-skewed.
  Interpretation: While crude oil rates were generally moderate, some high price years cause the mean to exceed the median.

- **Unemployment Rate (% of labor force):** Min 3.1, Max 7.8, Median 5.9, Mean > 5.5, Outliers: 8.5, Skewness: Right-skewed.
  Interpretation: Unemployment was mostly stable around 5–6%, but one or more high outlier years increased the mean, indicating brief periods of labor market stress.

- **Adolescent Fertility Rate:** Min 40, Max 110, Median 69, Mean 71, Outliers: None, Skewness: Symmetrical.
  Interpretation: Fertility rates are evenly distributed, showing a stable trend over time with no extreme values.

- **Age Dependency Ratio:** Min 69, Max 100, Median 86, Mean 80, Outliers: None, Skewness: Left-skewed.

Interpretation: Most values are high, but a few lower values reduce the mean slightly, indicating a high dependent population.

- **Urban Population Growth (annual %):** Min 1.7, Max 5.5, Median 3.4, Mean 3.5, Outliers: None, Skewness: Symmetrical.
  Interpretation: Urban population growth has been consistently moderate with no extreme deviations.

- **Total Reserves (% of External Debt):** Min 5, Max 37, Median 13, Mean 20, Outliers: None, Skewness: Right-skewed.
  Interpretation: While reserves are generally low-to-moderate (10–25%), some years had significantly higher reserves.

- **Exports Growth (annual %):** Min -13, Max 35, Median 2, Mean 10, Outliers: None, Skewness: Right-skewed.
  Interpretation: Most years had modest export growth, but a few very high growth years increased the mean.

- **Adjusted Gross Savings (% of GNI):** Min 11, Max 24, Median 17, Mean 17.5, Outliers: 4, Skewness: Left-skewed.
  Interpretation: Savings rates are generally healthy, but one very low year drags the mean down, hinting at economic stress.

0.5.3 Scatterplots

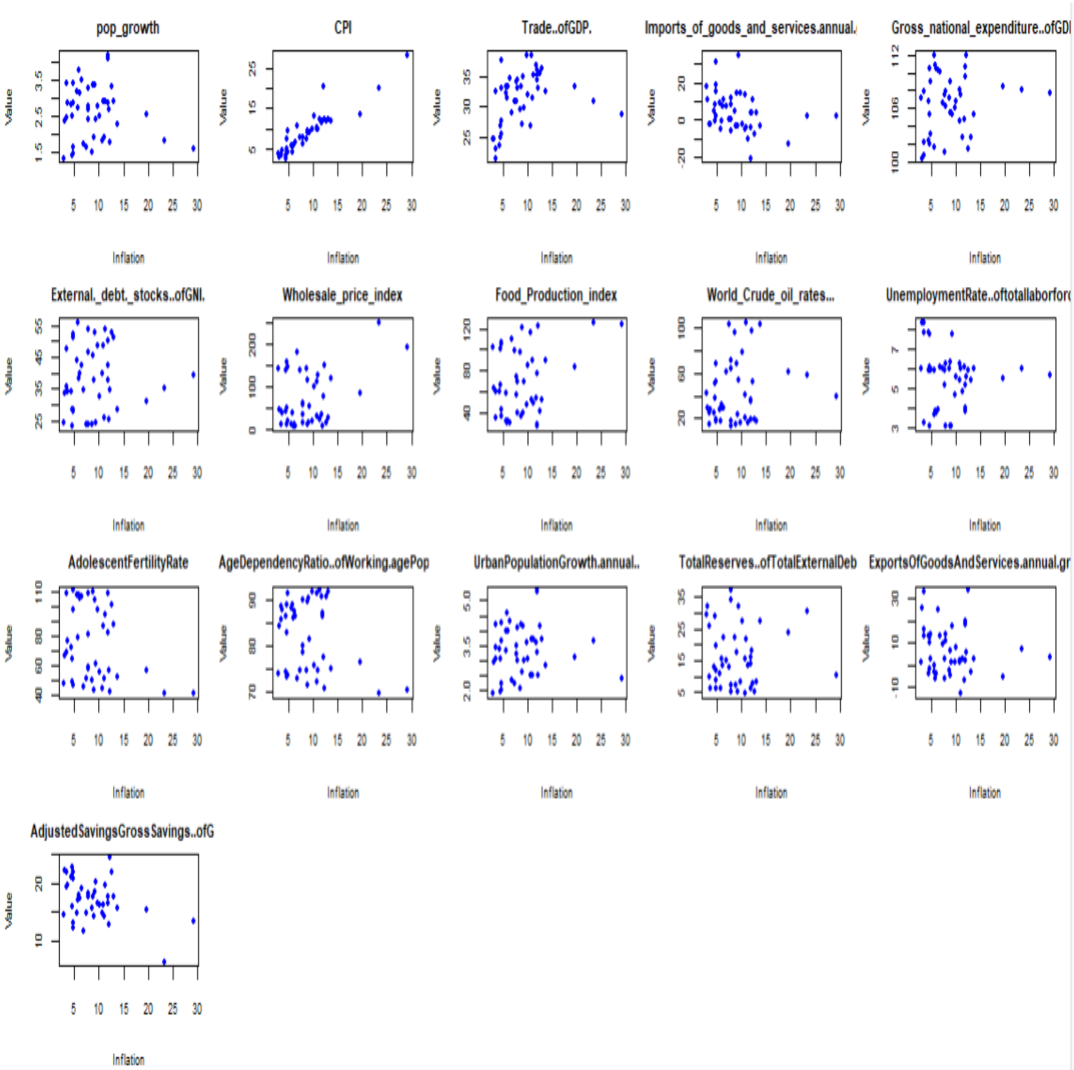


Table 2: *
Summary of Correlation Direction and Strength

Variable	Correlation with Inflation
Population Growth (pop_growth)	Weak positive
Consumer Price Index (CPI)	Strong Positive
Trade as % of GDP (Trade..ofGDP)	Positive
Imports of Goods and Services	weak negative
Gross National Expenditure (% of GDP)	weak Positive
External Debt Stocks (% of GNI)	weak Positive
Wholesale Price Index	Positive
Food Production Index	Weak Negative
World Crude Oil Rates	positive
Unemployment Rate	No Clear Correlation
Adolescent Fertility Rate	no clear correlation
Age Dependency Ratio	Weak Negative
Urban Population Growth	Weak Negative
Total Reserves (% of External Debt)	Weak Negative
Exports of Goods and Services	Weak Positive
Adjusted Gross Savings (% of GNI)	Weak Negative

0.5.4 Scatterplots of Significant variables

```
Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                                 -45.34762   79.53263  -0.570  0.57386
pop_growth                                    6.98034    6.75113   1.034  0.31147
CPI                                           0.34538    0.15599   2.214  0.03657 *
Trade..ofGDP.                                 0.64469    0.19704   3.272  0.00323 **
Imports_of_goods_and_services.annual.growth. -0.10062    0.03869  -2.600  0.01569 *
Gross_national_expenditure..ofGDP.           -0.17859    0.33945  -0.526  0.60363
External._debt._stocks..ofGNI.               -0.14338    0.09747  -1.471  0.15427
Wholesale_price_index                        -0.03580    0.03499  -1.023  0.31643
Food_Production_index                         0.35239    0.16887   2.087  0.04770 *
World_Crude_oil_rates...                     -0.01073    0.03697  -0.290  0.77409
UnemploymentRate..oftotallaborforce.          1.94116    0.84619   2.294  0.03084 *
AdolescentFertilityRate                       0.34439    0.16830   2.046  0.05184 .
AgeDependencyRatio..ofWorking.agePopulation. -0.03819    0.38263  -0.100  0.92133
UrbanPopulationGrowth.annual..               -3.40237    5.17348  -0.658  0.51702
TotalReserves..ofTotalExternalDebt.           0.14354    0.09590   1.497  0.14749
ExportsOfGoodsAndServices.annual.growth.     -0.05265    0.03712  -1.418  0.16898
AdjustedSavingsGrossSavings..ofGNI.          -0.44993    0.37671  -1.194  0.24401
---
```

Figure 2: Significant Variables

As shown above, the scatter plots visualize the relationships between the selected variables. The variables marked with an asterisk (*) were found to be statistically significant, i.e., their p-values were less than 0.05. These variables demonstrate a notable linear relationship with other indicators in the dataset.

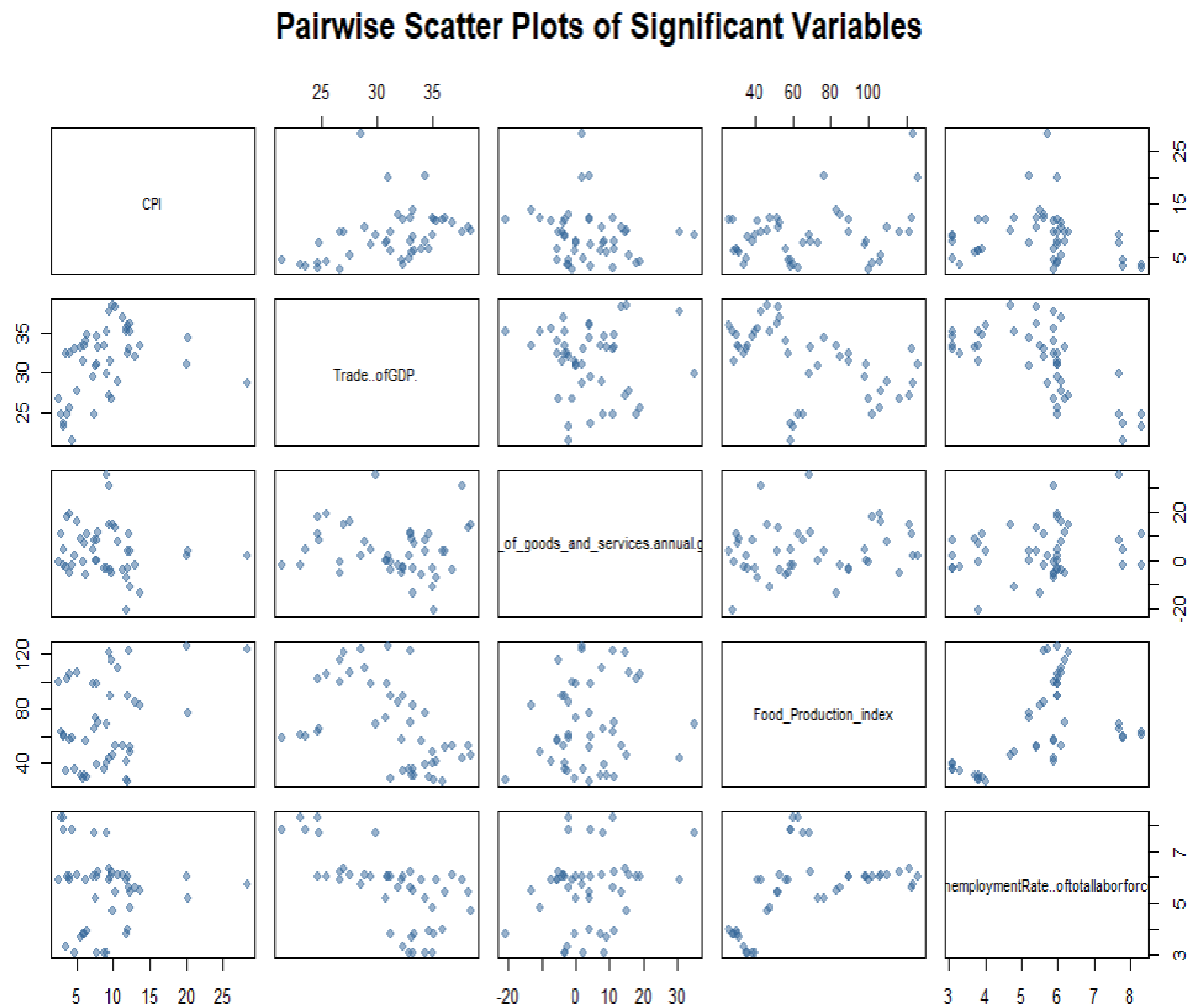


Figure 3: scatter plot

	CPI	Trade..ofGDP.	Imports_of_goods_and_services.annual.growth.	Food_Production_index	UnemploymentRate..oftotallaborforce.
CPI	1.00000000	0.35312398	-0.15702058	0.3012100	-0.09425034
Trade..ofGDP.	0.35312398	1.00000000	-0.09365753	-0.4834669	-0.64227467
Imports_of_goods_and_services.annual.growth.	-0.15702058	-0.09365753	1.00000000	0.1435286	0.22409575
Food_Production_index	0.30120998	-0.48346694	0.14352859	1.0000000	0.48826655
UnemploymentRate..oftotallaborforce.	-0.09425034	-0.64227467	0.22409575	0.4882665	1.00000000

>

Figure 4: Correlation Matrix

0.6 Findings/Results

0.6.1 ARIMA

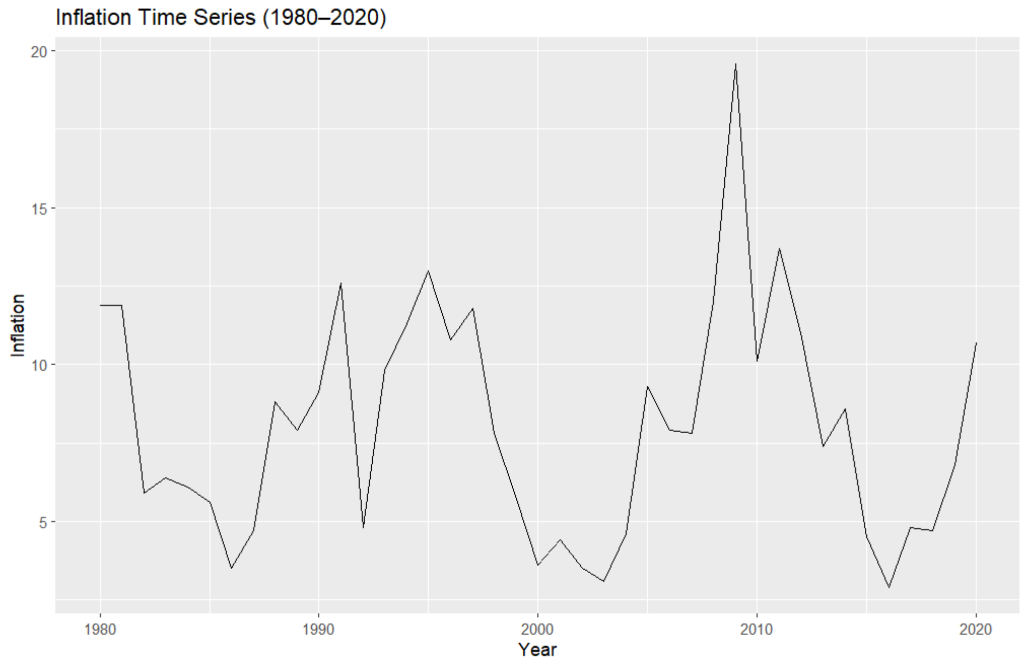


Figure 5: [1980-2020 Inflation Time Series (Training Data]

	Year	Actual	Predicted
1	2021	8.9	0.1530886
2	2022	12.2	2.2365798
3	2023	29.2	7.8824529
4	2024	23.4	4.6583828

>

Figure 6: [Actual Inflation Values and predicted using ARIMA]

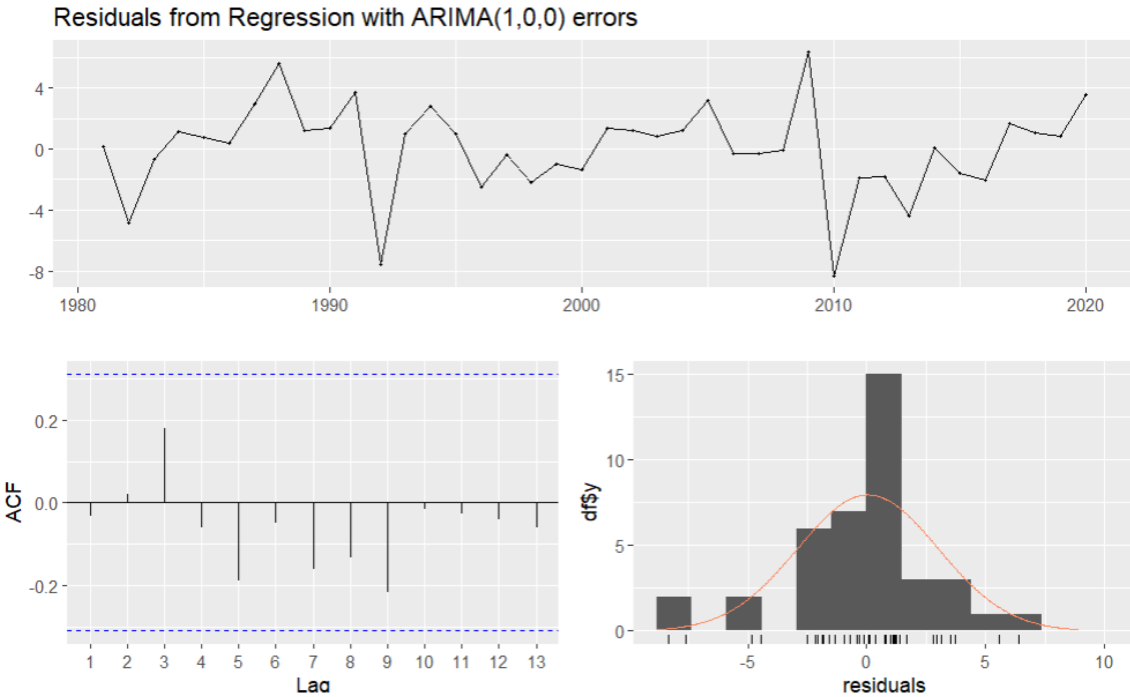


Figure 7: [ARIMA Residuals]

```
MSE: 245.3661
> cat("MAE:", round(mae, 4), "\n")
MAE: 14.6924
> cat("RMSE:", round(rmse, 4), "\n")
RMSE: 15.6642
```

Figure 8: [Evaluation Metrics ARIMA]

0.6.2 LASSO

```
Evaluation Metrics (2021-2024):
> cat("MSE:", round(mse, 3), "\n")
MSE: 30.416
> cat("RMSE:", round(rmse, 3), "\n")
RMSE: 5.515
> cat("MAE:", round(mae, 3), "\n\n")
MAE: 4.563
```

Figure 9: [Evaluation Metrics LASSO]

	Year	Actual	Predicted
1	2021	8.9	7.774508
2	2022	12.2	10.363025
3	2023	29.2	21.304407
4	2024	23.4	16.005350

Figure 10: [Actual Inflation Values and predicted using LASSO]

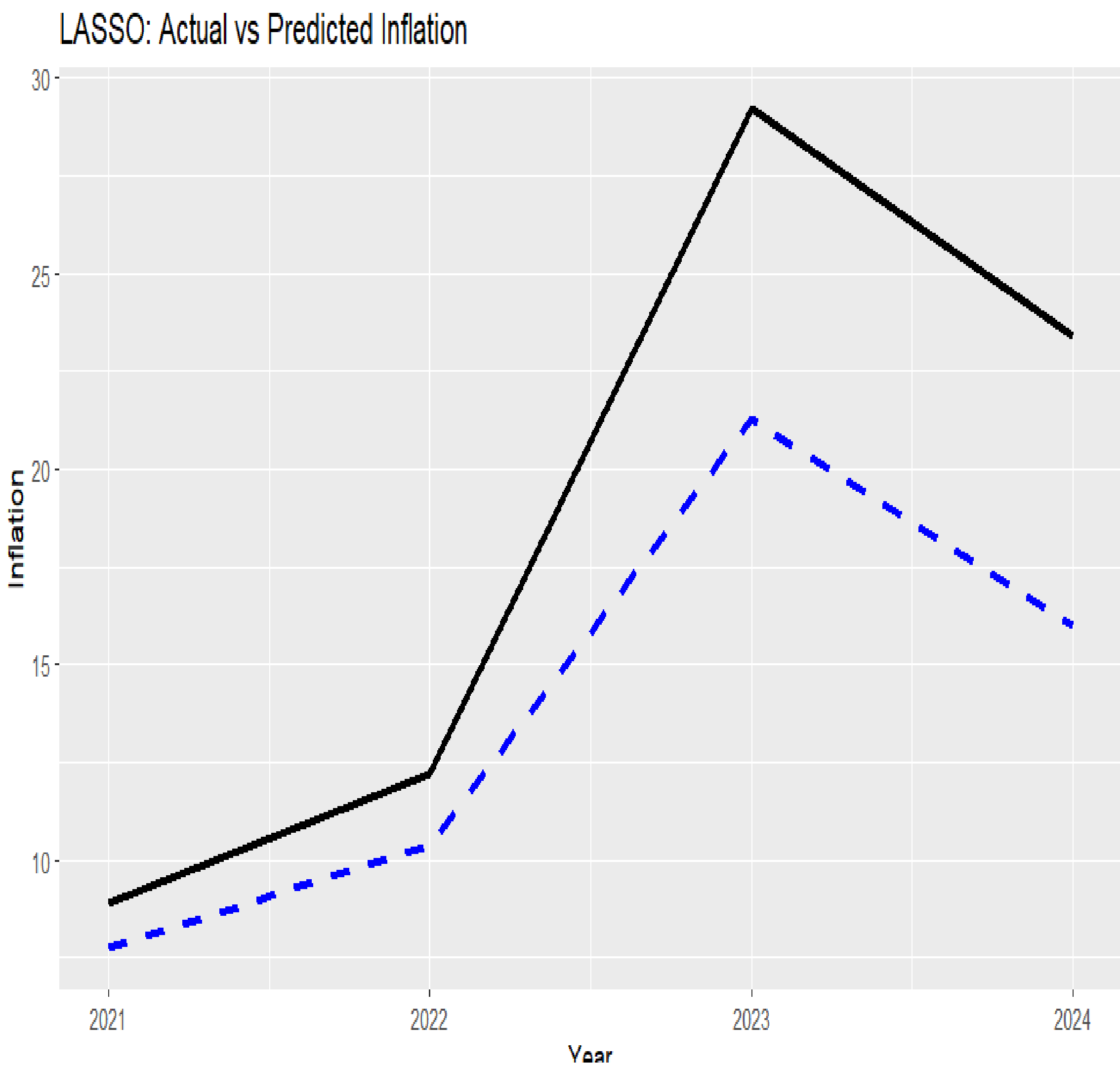


Figure 11: [Graph actual vs Predicted LASSO]

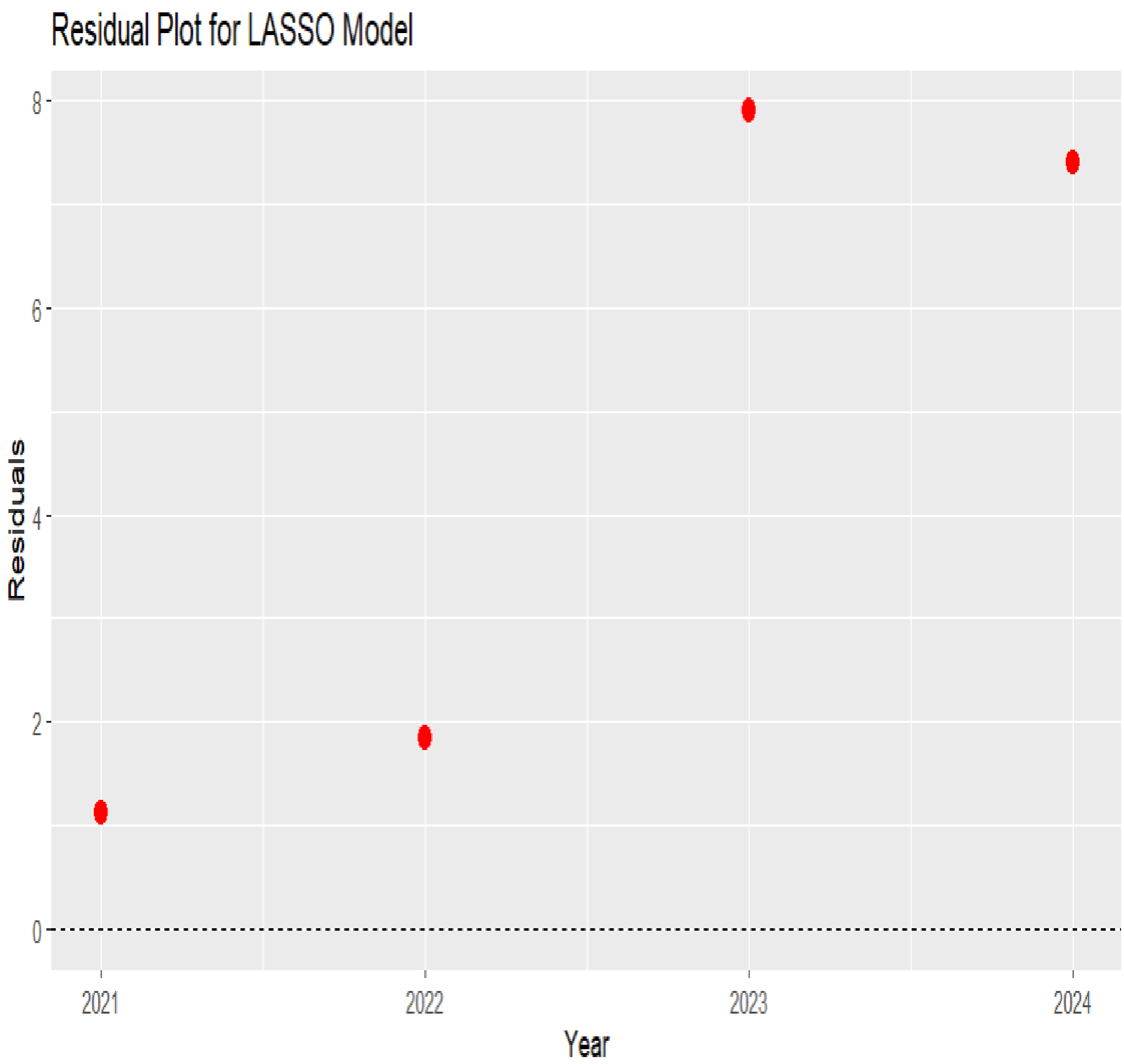


Figure 12: [LASSO Residuals]

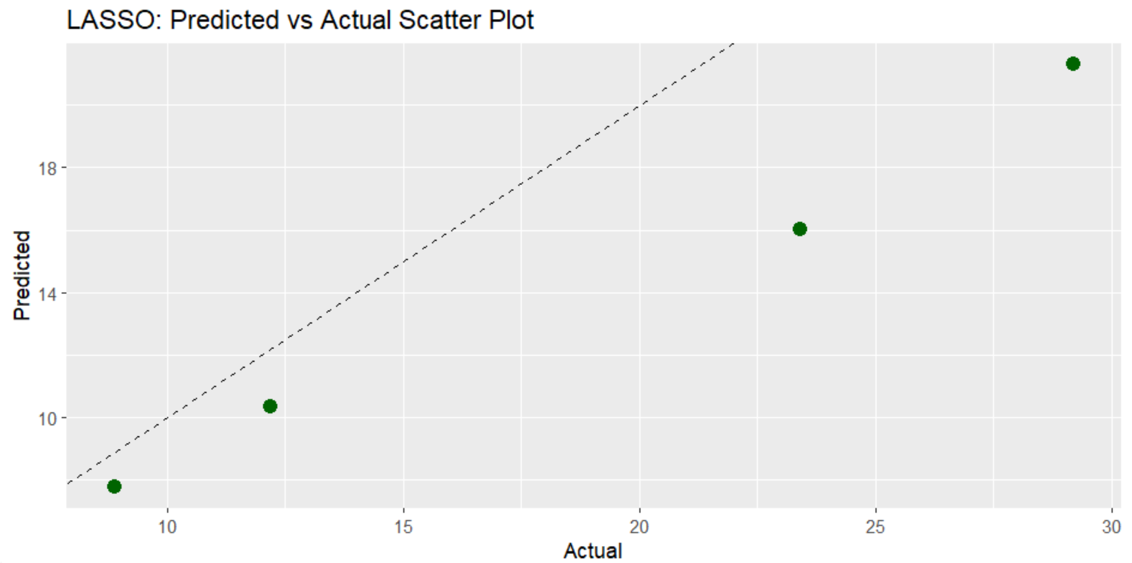
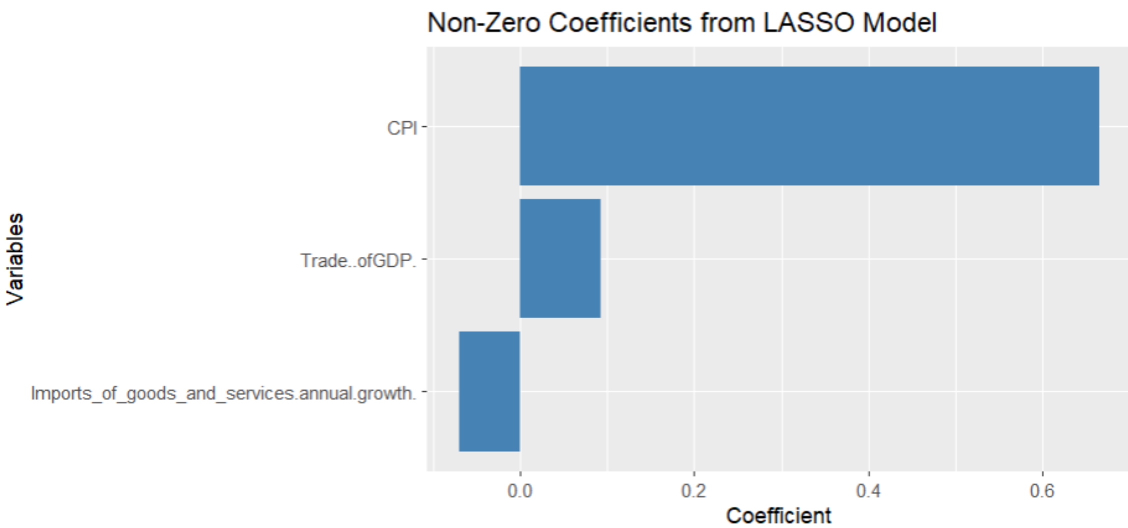


Figure 13: [Scatter plot with line]



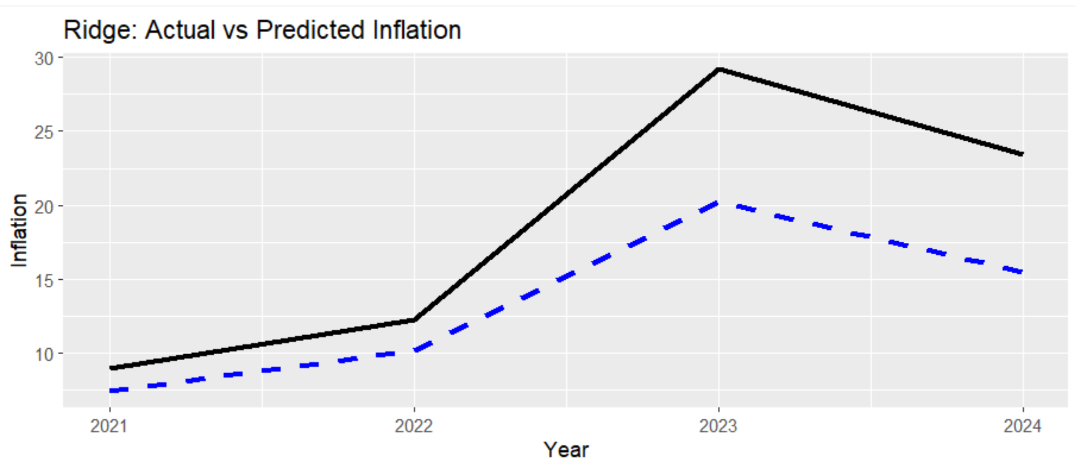
0.6.3 Ridge

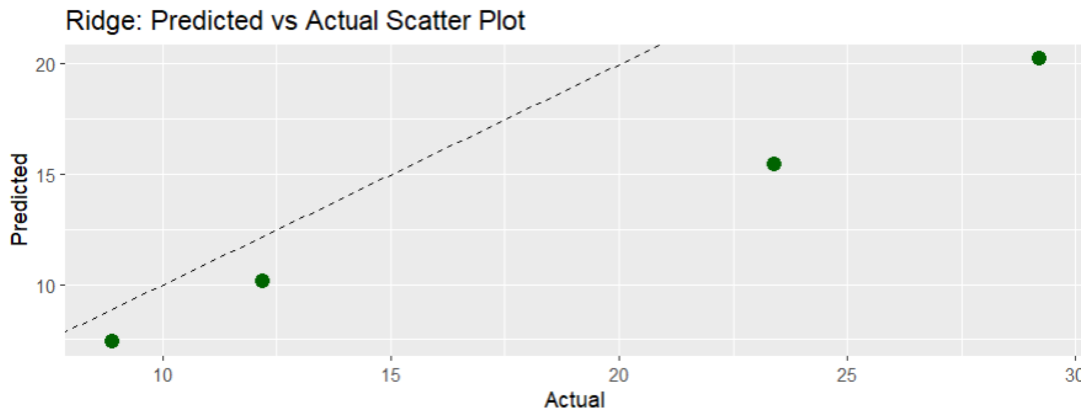
```
Ridge Model Evaluation Metrics (2021-2024):
> cat("MSE:", round(mse_ridge, 3), "\n")
MSE: 37.582
> cat("RMSE:", round(rmse_ridge, 3), "\n")
RMSE: 6.13
> cat("MAE:", round(mae_ridge, 3), "\n\n")
MAE: 5.116
```

Figure 14: [Evaluation Metrics Ridge]

```
  Year Actual Predicted
1 2021    8.9  7.406171
2 2022   12.2 10.168726
3 2023   29.2 20.227562
4 2024   23.4 15.433524
>
```

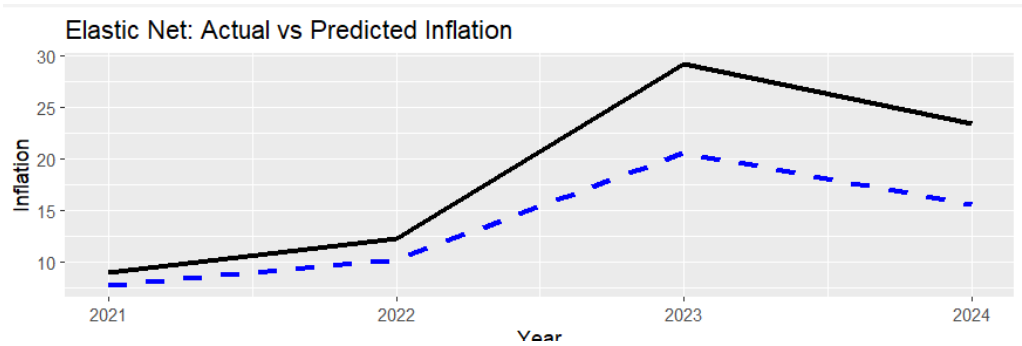
Figure 15: [Actual Inflation Values and predicted using Ridge]

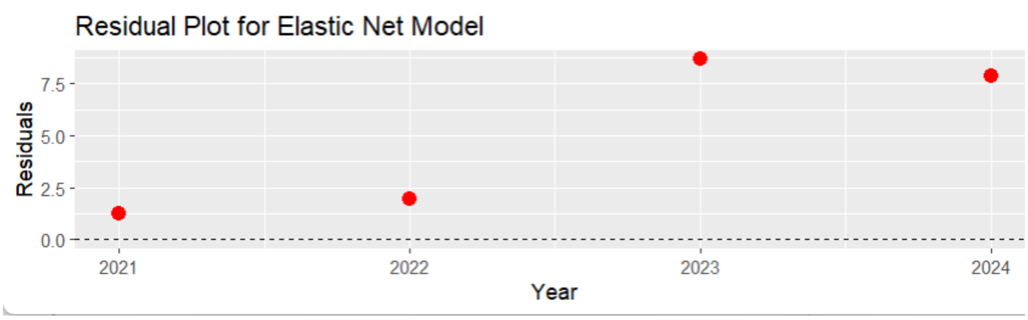


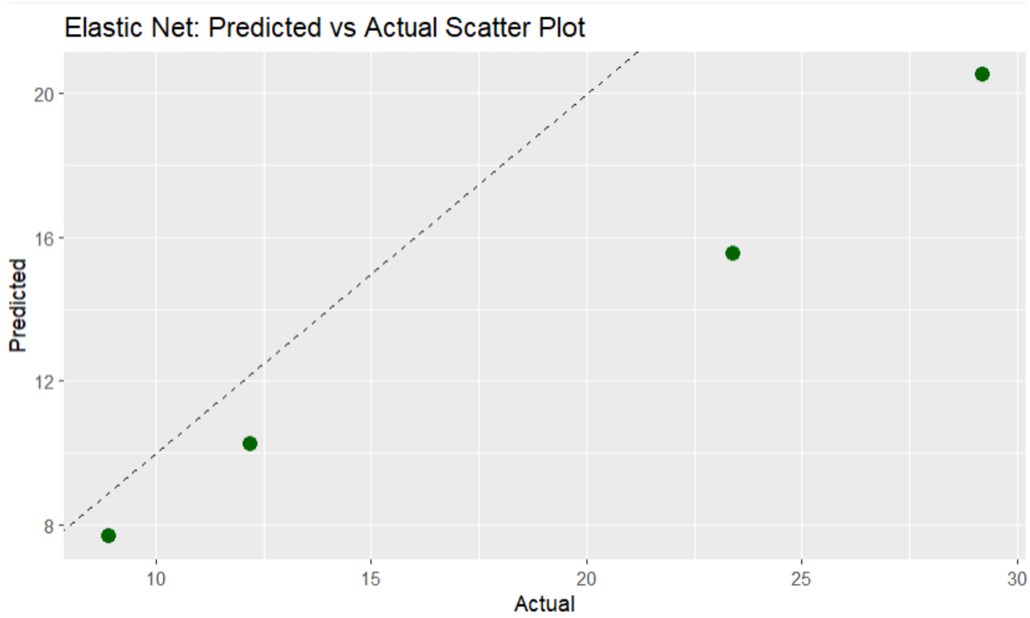


0.6.4 Elastic Net

```
Elastic Net Model Evaluation Metrics (2021-2024):
> cat("MSE:", round(mse_enet, 3), "\n")
MSE: 35.52
> cat("RMSE:", round(rmse_enet, 3), "\n")
RMSE: 5.96
> cat("MAE:", round(mae_enet, 3), "\n\n")
MAE: 4.924
```







Conclusion

In this study, we employed four different models to forecast inflation for the years 2021–2024 and compared their performances using the Mean Squared Error (MSE) metric:

- **ARIMA Model**: MSE = 245.37
- **LASSO Regression**: MSE = 30.42
- **Ridge Regression**: MSE = 37.58
- **Elastic Net**: MSE = 35.52

Among the models, the **LASSO Regression** achieved the lowest MSE of 30.42, indicating the best predictive performance on the test data. This suggests that LASSO not only fits the data well but also effectively avoids overfitting by performing automatic variable selection and shrinking the coefficients of less relevant predictors to zero. As a result, it simplifies the model and enhances interpretability.

In contrast, the ARIMA model showed significantly higher error (MSE = 245.37), primarily due to the presence of non-stationarity and its inability to capture the complex multivariate relationships among predictors. While ARIMA accounted for autocorrelation using time series components, its failure to fully address non-stationarity, as indicated by the Augmented Dickey-Fuller test (p-value = 0.15), led to poor forecasting performance.

Both Ridge and Elastic Net models performed reasonably well but were slightly inferior to LASSO in terms of MSE. The Ridge model includes all predictors by applying an L_2

penalty, which helps in handling multicollinearity but doesn't perform variable selection. The Elastic Net combines both L_1 and L_2 penalties, offering a compromise between Ridge and LASSO, but still couldn't outperform the sparsity-inducing nature of LASSO in this case.

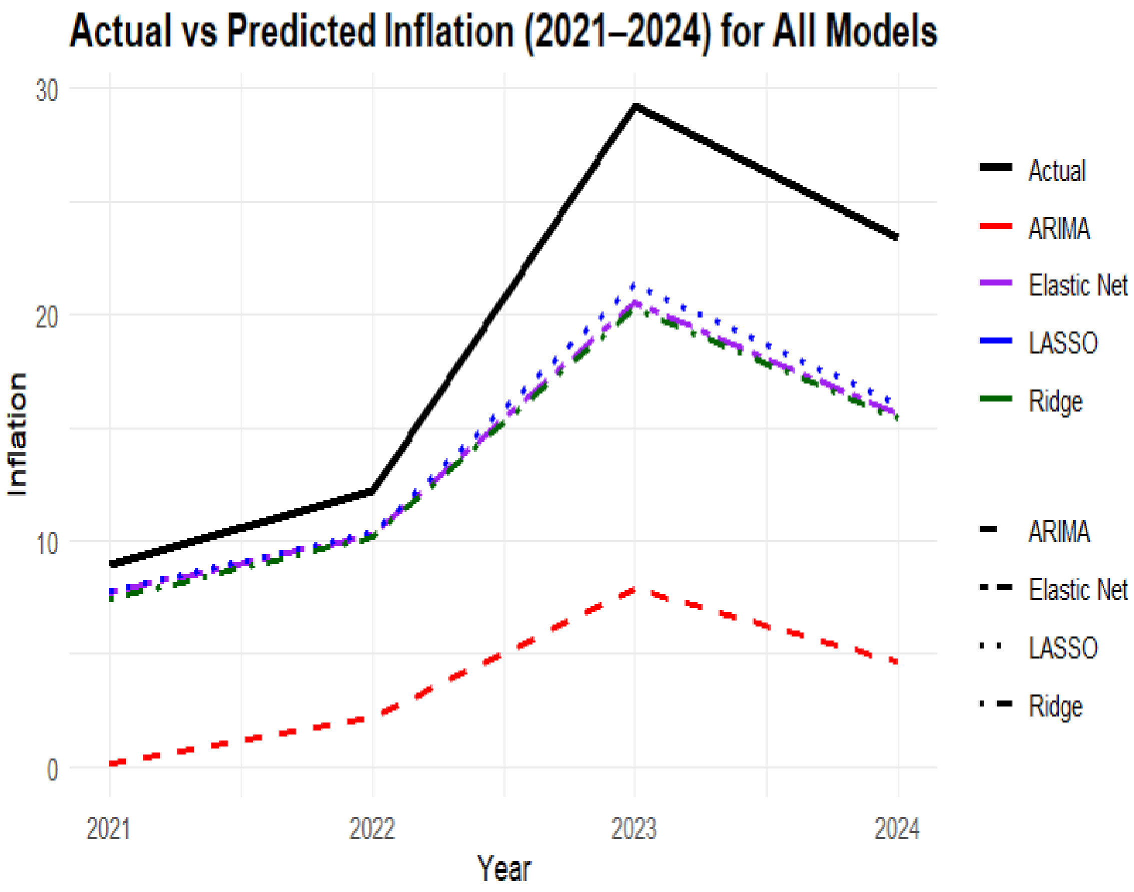


Figure 16: Actual vs Predicted Inflation (2021–2024) for All Models

In conclusion, **LASSO is the best-performing model** for forecasting inflation in this study due to its low prediction error, ability to handle irrelevant predictors via regularization, and simplicity in model interpretation. Future work can consider combining time series dynamics and regularized regression to further improve forecasting performance.

0.7 References

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

- Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts. In *NBER Working Paper Series*. National Bureau of Economic Research. https://www.nber.org/papers/w14322

- Öğünç, F., Akdoğan, K., Başer, S., Chadwick, M. G., Ertuğ, D., Hülagü, T., Kösem, S., Özmen, M. U., & Tekatlı, N. (2017). Short-term inflation forecasting models for Turkey and a forecast combination analysis. *Central Bank of the Republic of Turkey Working Paper*.

- Hanif, M. N., & Malik, M. J. (2015). Evaluating performance of inflation forecasting models of Pakistan. *SBP Research Bulletin*, 11(1).

- Hanif, M. N., & Malik, M. J. (2015). Evaluating performance of inflation forecasting models of Pakistan. *SBP Research Bulletin*, 11(1).

- Hassani, H., & Silva, E. S. (2014). Forecasting UK consumer price inflation using inflation forecasts. *Applied Economics*, Routledge.

- Albahouth, A. A. (2025). Inflation rate determinants in Saudi Arabia: A non-linear ARDL approach. *Sustainability*, 17(3), 1036. https://doi.org/10.3390/su17031036

- Akhter, T. (2013). Short-term forecasting of inflation in Bangladesh with seasonal ARIMA processes. *Munich Personal RePEc Archive (MPRA)*. https://mpra.ub.uni-muenchen.de/id/eprint/48599

- Sultana, K., Rahim, A., Moin, N., Aman, S., & Ghauri, S. P. (2022). Forecasting inflation and economic growth of Pakistan using two time series methods. *Journal of Economic Studies*.

- GeeksforGeeks, What is Ridge Regression?,
Available at: https://www.geeksforgeeks.org/what-is-ridge-regression/

- Lasso Regressions and Forecasting Models in Applied Stress Testing, Chan-Lau, Jorge A., IMF Working Paper No. 17/244, International Monetary Fund, 2017.

- International Monetary Fund. (2025). Inflation rate, average consumer prices – Pakistan. World Economic Outlook (April 2025).
Retrieved from
https://www.imf.org/external/datamapper/PCPIPCH@WEO/WEOWORLD/PAK.