Inflation Forecasting for Pakistan

Project Team

Noor Ul Huda	23I-2099
Malaika Zainab	23I-2605

Department of AI and DS

National University of Computer and Emerging Sciences
Islamabad, Pakistan

,

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Inflation represents a sustained increase in the general price level of goods and services in an economy over time. It decreases the purchasing power of money, meaning that as inflation rises, each unit of currency buys fewer goods and services. Inflation is typically measured on an annual basis using indices such as the Consumer Price Index (CPI), Producer Price Index (PPI), or the GDP deflator. Among these, CPI is the most widely used measure, reflecting the changes in prices experienced by households. By analyzing time series data from key variables such as exchange rates, money supply growth, oil rents, import-export ratios, and sectoral contributions to GDP, we attempt to uncover patterns and relationships that influence inflation dynamics.

1.1 Objectives

The main objectives of this project are:

1. To explore and preprocess a dataset containing inflation and various economic indicators from 1980 to 2025.
2. To build and evaluate four different models for forecasting inflation:
   - A time series model: ARIMA (AutoRegressive Integrated Moving Average)
   - Machine learning regression models:
     - Lasso Regression (Least Absolute Shrinkage and Selection Operator)
     - Ridge Regression
     - Elastic Net Regression
3. To compare the performance of the models in terms of forecasting accuracy and interpretability using appropriate evaluation metrics.

4. To identify significant determinants of inflation using Lasso Regression's feature selection capability.

1.2 Methodology Overview

1.2.1 Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) regression is a regularized linear regression technique that not only minimizes prediction error but also performs variable selection by adding an L1 penalty to prevent overfitting. It effectively shrinks some coefficients to zero, thus selecting a simpler model and identifying the most influential predictors.

1.2.2 Ridge Regression

Another regularisation technique that increases the loss function's L2 penalty is ridge regression. In contrast to Lasso, Ridge reduces model complexity and multicollinearity by shrinking coefficients towards zero rather than setting them exactly to zero. Although it may suffer from high correlation, it works especially well when all predictors are anticipated to have an impact on the target variable.

1.2.3 Elastic Net Regression

Another regularisation technique that increases the loss function's L2 penalty is ridge regression. In contrast to Lasso, Ridge reduces model complexity and multicollinearity by shrinking coefficients towards zero rather than setting them exactly to zero. Although it may suffer from high correlation, it works especially well when all predictors are anticipated to have an impact on the target variable.

1.2.4 ARIMA Model

A traditional statistical method for simulating time series data is the ARIMA (AutoRegressive Integrated Moving Average) model. By combining autoregressive terms, moving average components, and differencing (to make the series stationary), it is able to capture autocorrelations in the data. When the inflation time series is univariate and shows significant temporal relationships, ARIMA is very appropriate.

1.3 Model Comparison

The four models used in this study offer different strengths and are suited to different types of data patterns:

- **Lasso Regression** effectively selects features by reducing irrelevant coefficients to zero. It is perfect for determining the most significant economic indicators and manages multicollinearity effectively. It might, however, ignore time series data's temporal dependencies.
- **Ridge Regression** minimises the complexity of the model by reducing all of the coefficients, but it doesn't carry out variable selection. It works well when a large number of predictors have weak, correlated effects, but it is not interpretable and does not remove variables.
- **Elastic Net Regression** combines the benefits of Lasso and Ridge. It selects variables like Lasso while handling correlated features more effectively, making it a balanced approach when multicollinearity is present alongside the need for feature selection.
- **ARIMA Model**is designed for time series data, forecasting based on historical values and trends. Although it performs a good job of capturing temporal dynamics, it may not be able to handle complicated feature interactions and does not take into account many external factors unless it is extended to ARIMAX.

Chapter 2

Literature Review

In their article [7] "Trade-off between Inflation, Interest and Unemployment Rate of Pakistan: Revisited," Sumera Arshad and Amajd Ali study Pakistan's economy from 1974 to 2013. Using techniques like the Augmented Dickey-Fuller (ADF) test, ARDL model, and other diagnostic tests, they examine how population growth, exchange rate, external debt, and political instability impact unemployment, inflation, and interest rates. Their results show that broad money supply significantly increases inflation, while exchange rate and political instability help reduce it.

In the article [8]"Inflation and Economic Growth: Evidence from Pakistan," Shahzad Hussain investigates this relationship using co-integration and error correction models for both the short-run and long-run. His study shows that inflation and economic growth are positively related in Pakistan, meaning that a moderate level of inflation supports growth. This finding matches Malik and Chowdhury (2001) but differs from Mubarik (2005), who found a negative relationship. Hussain explains this positive link through the Tobin portfolio-shift effect, where higher inflation encourages more investment in physical assets. The Granger causality test further reveals that inflation influences economic growth after three lags, but growth does not significantly cause inflation.

In the article [5] "Relationship between Inflation and Interest Rate: Evidence from Pakistan," Ayub et al. (2010) investigate this connection by focusing on the validity of the Fisher Hypothesis in Pakistan. Using time series data from 1973 to 2010, the study first finds that both the nominal interest rate and inflation rate are non-stationary at their original levels, as confirmed through Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root tests. However, after first differencing, both series become stationary at 1 percent and 5 percent significance levels. Applying Engle-Granger and Johansen co-integration techniques, the authors conclude that there is a significant long-run equilibrium relationship between inflation and nominal interest rate in Pakistan, supporting the Fisher Hypothesis. Their findings suggest that changes in inflation are closely linked with movements in nominal interest rates over the long term.

The study [10] "Determinants of Inflation: A Case of Pakistan (1970–2007)" by Khan and Gill investigates the major factors influencing inflation in Pakistan over a long period. The authors examine data from 1971–72 to 2005–06 using four distinct price indicators: the GDP Deflator, Sensitive Price Indicator (SPI), Wholesale Price Index (WPI), and Consumer Price Index (CPI). They discovered that growing import prices and currency rate depreciation are the main causes of inflation. Inflation is also impacted by higher support prices for important crops, but budget deficits have little long-term effect on it.

The study [2]"Determinants of Recent Inflation in Pakistan" by Abdul Aleem Khan, Syed Kalim Hyder Bukhari, and Qazi Masood Ahmed investigates the key factors behind the rising inflation in Pakistan during the period 1972-73 to 2005-06. The study, which uses an economic framework with Ordinary Least Squares (OLS) estimation and is backed by a number of tests, concludes that growing import prices, private sector loan growth, and adaptive expectations were the main causes of inflation, particularly in 2005–06. Due to easy lending, expansionary monetary policy raised GDP growth but also consumer prices. While exchange rate stability assisted in containing imported inflation, fiscal policy and government borrowing had minimal effect on inflation. According to the report, price increases were supported by inflation expectations, raising concerns about Pakistan's capacity to control inflation in light of rising future expectations and a growing trade imbalance. To stabilise growth, effective monetary tightening and improved private sector credit management are required.

The study [6] "The Determinants of Inflation in Pakistan: An Econometric Analysis" by Nazima Ellahi explores the complex nature of inflation, emphasizing its dual role as both a driver of economic activity when mild and a damaging force when excessive. The study determines the main factors affecting inflation in Pakistan by using time series data from 1975 to 2015 using the Auto Regressive Distributed Lag (ARDL) methodology. The results indicate that while national spending and the import of goods and services have a positive effect on inflation, the money supply has a negative effect. It's interesting to note that inflation is negatively impacted by GDP growth, suggesting that improved economic performance can aid in price stabilisation. However, none of the factors had a substantial impact on inflation in the short term.

The study [4]"Inflation and Unemployment in Pakistan: An Empirical Analysis" by Dr. Ghulam Muhammad Mangnejo, Saqib Wahab Mahar, and Bakhtiar Ahmed investigates the relationship between inflation and unemployment in Pakistan, focusing on the existence of the Phillips Curve. The study, which uses EViews9 to analyse time series data from 1991 to 2015, discovers that periods of high inflation—especially in the 1990s and between 2005 and 2010—were linked to falling unemployment rates. On the other hand, unemployment increased during times of declining inflation, such as 2001 to 2005. The Phillips Curve theory is supported in the case of Pakistan by these trends, which imply an inverse relationship between unemployment and inflation.

The study [9]"Inter-relationship among Economic Growth, Savings and Inflation in Pakistan" by Muhammad Ilyas, Hazoor Muhammad Sabir, Anam Shehzadi, and Naeem Shoukat explores the complex linkages among economic growth, savings, and inflation in Pakistan using annual data from 1973 to 2010. Employing a simultaneous equation model and the 2SLS (Two-Stage Least Squares) technique, the research examines how these variables interact with each other. The findings revealed a bidirectional negative relationship between inflation and economic growth over the long run, with inflation and real interest rates harming economic growth, while the depreciation rate had a positive effect.

The study [3]"An Empirical Analysis of Fiscal Imbalances and Inflation in Pakistan" by Asif Idrees Agha and Muhammad Saleem Khan explores the long-run connection between fiscal indicators and inflation in Pakistan using annual data from 1973 to 2003.Using Johansen cointegration analysis, the study treats real GDP and exchange rates as exogenous variables and concludes that inflation is substantially correlated with fiscal deficits and the government's reliance on bank borrowing for budgetary support. Fiscal operations are a major factor in generating inflationary trends in the nation, as evidenced by the Vector Error Correction Model (VECM) results, which show that inflation reacts to fiscal imbalances through large error correction terms.

Agha and Khan (2006) [1] explored the long-run dynamics between fiscal indicators and inflation using annual data from FY 1973 to FY 2003. Employing Johansen's cointegration analysis, they found significant evidence that inflation in Pakistan is closely linked to fiscal imbalances, particularly government borrowing from the banking sector. Their results show that fiscal deficits and total bank borrowing had a considerable and long-term causal impact on inflation, even when actual GDP and exchange rates were considered exogenous. Inflation is predominantly a fiscal issue in Pakistan, as evidenced by the Vector Error Correction Model (VECM) results, which also showed that inflation reacts strongly to disequilibria brought on by borrowing and fiscal deficits.

Chapter 3

Variables

3.1 Variable Description and Relationship with Inflation

The data for each variable is taken from WDI (World development indicators).

Description of Variables

- **Time**: We took data from 1980 to 2025.
- **Inflation**: The increase in the price of goods and services over time, reducing purchasing power.
- **Official Exchange Rate (LCU per US$, period average)**: The average rate at which a country's currency is exchanged for US dollars during a specific period.
- **Broad Money Growth (Annual %)**: The annual percentage change in the total money supply in an economy, including cash and bank deposits.
- **Oil Rents (% of GDP)**: The share of a country's GDP that comes from oil revenues, such as exports and production.
- **Imports of Goods and Services (% of GDP)**: The percentage of a country's GDP spent on purchasing goods and services from other countries.
- **Exports of Goods and Services (% of GDP)**: The percentage of a country's GDP earned from selling goods and services to other countries.
- **External Debt Stocks (% of GNI)**: The portion of a country's Gross National Income that is owed to foreign lenders.

- **Agriculture, Forestry, and Fishing, Value Added (% of GDP)**: The percentage of a country's GDP generated from agriculture, forestry, and fishing activities.
- **Total Reserves (includes gold, current US$)**: The total value of a country's reserves held in foreign currencies, including gold, usually by the central bank.
- **Unemployment Rate**: The percentage of people in the workforce who are actively looking for a job but cannot find one.
- **Urban Population Growth**: The annual increase in the population of urban areas, often due to migration from rural areas.
- **Industrial Growth**: The percentage change in the output of industries such as manufacturing, mining, and energy.
- **GDP Growth**: The rate at which a country's total economic output (GDP) increases over a period of time.

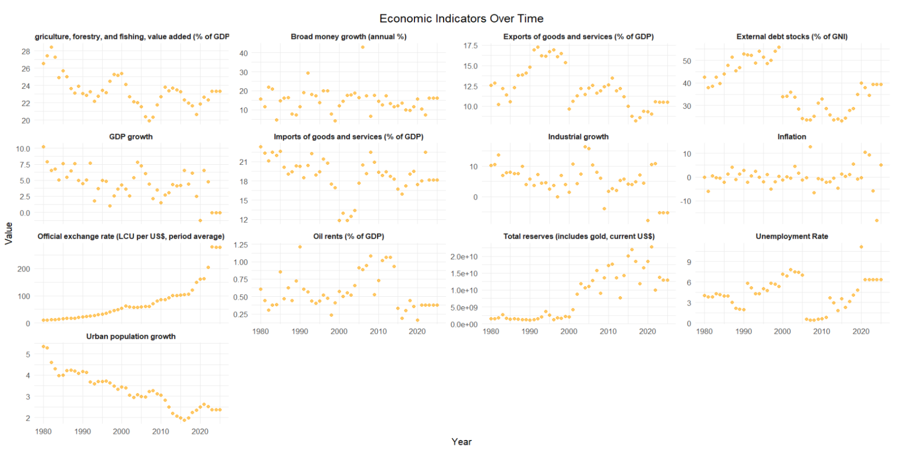

Figure 3.1: Scatter Plot

3.1.1 Scatter plot interpretaion

Most variables show consistent variability. However notable are exchange rate and Total reserves with positive linear relationship. population with negative linear relationship. Exports, GDP and imports donot have consistent trend.

Inflation VS Time: Analysis

The time series graph shows the trend of inflation from the year 1980 to around 2024.

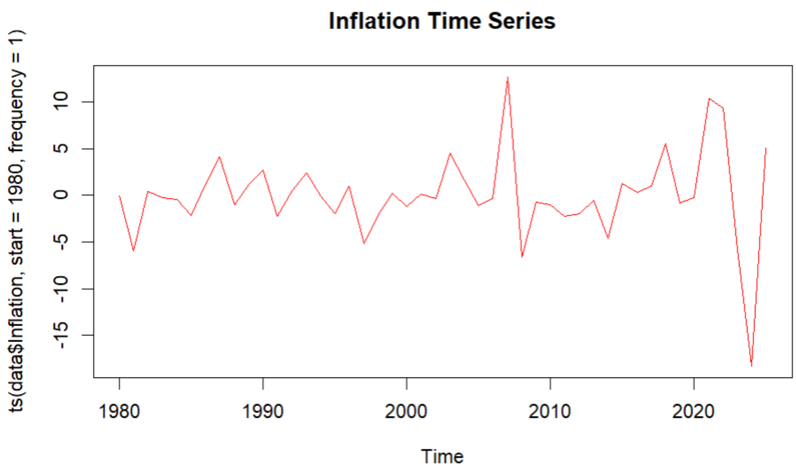

Figure 3.2: inflation vs Time

The time series does not show clear seasonality (repeating yearly pattern), but it does show occasional trends and sharp increases in certain years. The large spikes in 2008 and after 2020 may be due to major economic or political events.

Chapter 4

Models

4.1 Models Used in the Analysis

In this study, four widely used models were applied to analyze and forecast inflation: the **ARIMA (AutoRegressive Integrated Moving Average)** model, and three machine learning models—**LASSO**, **Ridge**, and **Elastic Net** regression.

4.1.1 ARIMA Model

The ARIMA model is a time-series forecasting method. It is especially useful for data that changes over time. The ARIMA model is defined by three main components:

- **AR (AutoRegressive)**: This component uses past values to predict current values.
- **I (Integrated)**: This makes the time series stationary by removing trends through differencing.
- **MA (Moving Average)**: This component uses past forecast errors to improve predictions.

An ARIMA model is usually written as ARIMA(p, d, q), where:

- p is the number of autoregressive terms,
- d is the number of times the data is differenced to remove trends,
- q is the number of lagged forecast error terms.

$$Y'_t = c + \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \cdots + \phi_p Y'_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Where:

- Y'_t is the value of the series at time t,
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients,
- ε_t is the error term at time t,
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients,
- c is the constant term.

Advantages: It is effective when analyzing and forecasting a single variable that changes over time, such as inflation rates in Pakistan.

4.1.2 LASSO Regression

LASSO stands for *Least Absolute Shrinkage and Selection Operator*. It is a regression technique used when we have many predictors (independent variables). LASSO not only fits a model but also selects the most important features by shrinking the less important ones to zero.

$$\min_{\beta_0, \beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - X_i \cdot \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Where:

- y_i is the target variable (e.g., inflation),
- X_i is a vector of predictor variables (e.g., exchange rate, imports, etc.),
- β are the coefficients,
- λ is the regularization parameter that controls the amount of shrinkage.

Advantages: It helps in identifying which economic factors have the most influence on the target variable by reducing the impact of irrelevant or weak predictors. Unlike regular linear regression, LASSO improves prediction accuracy and prevents overfitting by automatically selecting the most relevant features.

4.1.3 Ridge Regression

Ridge regression is a regularization method that addresses multicollinearity by adding a penalty to the size of the coefficients. Unlike LASSO, Ridge does not reduce coefficients to exactly zero but shrinks them closer to zero, keeping all predictors in the model.

$$\min_{\beta_0,\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - X_i \cdot \beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

Where the terms are the same as in LASSO, but the penalty term uses the square of coefficients instead of their absolute values.

Advantages: Ridge regression is suitable when many predictors are moderately related to the target. It provides stable predictions, especially when the number of predictors is high or when predictors are highly correlated.

4.1.4 Elastic Net Regression

Elastic Net regression is a hybrid approach that combines both LASSO and Ridge penalties. It is particularly useful when predictors are highly correlated or when the dataset contains more features than observations.

$$\min_{\beta_0,\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - X_i \cdot \beta)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right) \right\}$$

Where:

- $\alpha \in [0,1]$ controls the balance between LASSO (L1) and Ridge (L2) penalties,
- λ controls the overall regularization strength.

Advantages: Elastic Net combines the variable selection capability of LASSO with the stability of Ridge, making it a robust model when dealing with complex, high-dimensional data.

Chapter 5

Estimation

5.1 Summary Statistics for Key Economic Indicators

Description: df [13 × 8]

Variable <chr>	Mean <dbl>	Median <dbl>	Mode <dbl>	Q1 <dbl>	Q3 <dbl>	IQR <dbl>	Outliers <int>
Inflation	9.168686e+00	8.379511e+00	2.529328e+00	5.228710e+00	1.168733e+01	6.458616e+00	2
Official exchange rate (LCU per US$, period average)	7.659944e+01	5.888617e+01	9.900000e+00	2.412127e+01	1.014967e+02	7.737542e+01	3
Broad money growth (annual %)	1.497727e+01	1.518963e+01	1.625306e+01	1.156870e+01	1.756713e+01	5.998428e+00	2
Oil rents (% of GDP)	5.697267e-01	4.884851e-01	3.801974e-01	3.801974e-01	7.068504e-01	3.266530e-01	1
Imports of goods and services (% of GDP)	1.886762e+01	1.912525e+01	1.812037e+01	1.771469e+01	2.070293e+01	2.988233e+00	4
Exports of goods and services (% of GDP)	1.237139e+01	1.217107e+01	1.047994e+01	1.049510e+01	1.384647e+01	3.351375e+00	0
External debt stocks (% of GNI)	3.817445e+01	3.826030e+01	3.939438e+01	2.857930e+01	4.756158e+01	1.898228e+01	0
Agriculture, forestry, and fishing, value added (% of GDP)	2.339281e+01	2.328643e+01	2.332909e+01	2.223017e+01	2.408713e+01	1.856958e+00	3
Total reserves (includes gold, current US$)	8.377451e+09	8.223531e+09	1.297715e+10	1.614238e+09	1.366789e+10	1.205365e+10	0
Unemployment Rate	4.347011e+00	4.210000e+00	6.338000e+00	2.978000e+00	6.216000e+00	3.238000e+00	1
Urban population growth	3.273851e+00	3.249940e+00	2.364424e+00	2.505839e+00	3.905515e+00	1.399676e+00	0
Industrial growth	5.349911e+00	5.546263e+00	-5.261278e+00	3.735945e+00	8.018952e+00	4.283006e+00	7
GDP growth	4.455956e+00	4.451701e+00	-3.969252e-02	2.766984e+00	6.377183e+00	3.610199e+00	0

13 rows | 2-9 of 8 columns

Figure 5.1: summary statistics

5.2 Outliers (IQR Method)

The following table presents the outlier values identified in each numerical variable using the Interquartile Range (IQR) method. A value is considered an outlier if it lies below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$.

Table 5.1: Macroeconomic Indicators – Sample Data

Indicator	**Values**
Inflation	29.2, 23.4
Official exchange rate (LCU per US$, period average)	280.3561, 278.5807, 278.5807
Broad money growth (annual %)	29.30056, 42.90887
Oil rents (% of GDP)	1.214847
Imports of goods and services (% of GDP)	11.83034, 12.95104, 11.86606, 12.48552
Exports of goods and services (% of GDP)	—
External debt stocks (% of GNI)	—
Agriculture, forestry, and fishing, value added (% of GDP)	27.46229, 28.44777, 27.27320
Total reserves (includes gold, current US$)	—
Unemployment Rate	11.168
Urban population growth	—
Industrial growth	16.339861, 15.728007, -3.937173, -7.796860, -5.261278 (×3)
GDP growth	—

Table 5.1: Outliers

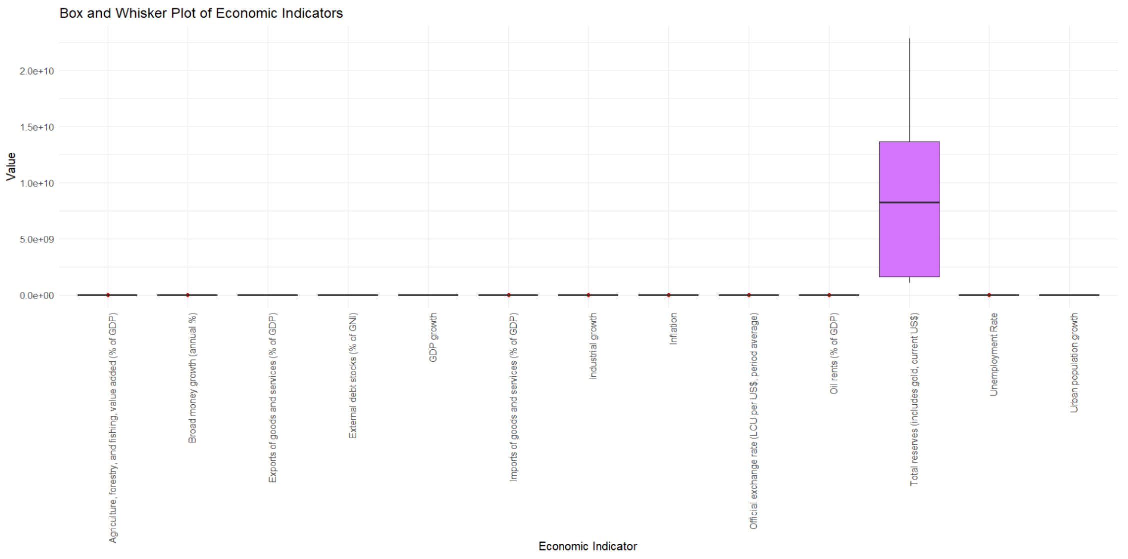

Figure 5.2: Box and Whisker Plot

5.2.1 Interpretation of Box and Whisker Plot

Interpretation of Box and Whisker Plot of Economic Indicators

The boxplot shows the distribution of several economic indicators. Most indicators, such as inflation, industrial growth, and unemployment rate, have relatively small values and appear compressed near the bottom of the plot. This is because one indicator, *Total reserves (includes gold, current US$)*, has extremely large values, in the range of billions or more. This large scale difference causes the boxplot of total reserves to dominate the vertical scale. As a result, the other indicators appear as flat lines, even if they have variability. The box for total reserves is relatively wide, suggesting a large spread (or variability) in reserve amounts over time. Red dots indicate outliers in some indicators like inflation, industrial growth, and unemployment rate. These are values that are unusually high or low compared to the rest of the data for those variables.

Chapter 6

Model estimation

6.0.1 Interpretation of ARIMA Model

- The model attempts to explain **Inflation** using several economic indicators as regressors. Following are notable.
  - **Broad Money Growth** has a negative coefficient (-0.0174), suggesting that higher money growth may slightly reduce inflation, though the effect is small and not statistically strong (high standard error: 0.0986).
  - **Oil Rents (% of GDP)** has a relatively large positive coefficient (0.2132), indicating that an increase in oil rents could be associated with higher inflation, although this estimate has a very high standard error (3.0556), making it statistically insignificant.
  - **Exports of Goods and Services** has a strong negative coefficient (-0.6104), possibly implying that more exports are associated with lower inflation.
  - **Agricultural Value Added (% of GDP)** has a moderately negative coefficient (-0.8722), indicating a potential dampening effect on inflation.
  - **Industrial Growth** and **GDP Growth** have positive coefficients (0.3387 and 0.2636, respectively), suggesting they may contribute to rising inflation.

Conclusion: While the model provides insights into how inflation relates to different macroeconomic indicators, the large standard errors and high MAPE suggest it may not be reliable for accurate forecasting without further refinement.

6.0.2 Interpretaion of The LASSO model

The LASSO regression model was employed to identify the most influential predictors of inflation while applying regularization to reduce overfitting and improve model interpretability. The final selected model retained three predictors with non-zero coefficients, indicating their relative importance in explaining inflation variability.

The intercept term is approximately 6.87, representing the baseline level of inflation when all predictors are at zero. The coefficient for **Agriculture, forestry, and fishing (as % of GDP)** is -0.35, suggesting that a higher share of agriculture in the economy is associated with lower inflation. This negative relationship may reflect the stabilizing influence of agricultural output on prices, especially in developing economies.

Industrial growth has a positive coefficient of 0.20, indicating that increased industrial output is associated with rising inflation, potentially due to higher production costs, energy demand, or wage pressures. Similarly, **GDP growth** has a smaller positive coefficient of 0.03, suggesting a mild inflationary effect as the overall economy expands.

These results highlight that inflation is modestly influenced by the sectoral composition of GDP and overall economic activity. The LASSO method has effectively reduced the model to only the most significant predictors, supporting a more interpretable and parsimonious regression structure.

6.0.3 Interpretation of Ridge Regression Model

- The Ridge regression model is used to predict **Inflation** using a subset of economic indicators selected through regularization, where insignificant variables are shrunk to zero.

  - **Oil Rents (% of GDP)** has a positive coefficient (0.0105), suggesting that an increase in oil rents is associated with higher inflation. This aligns with the economic intuition that rising oil revenues may fuel domestic price pressures.
  - **Imports of Goods and Services (% of GDP)** has a small positive coefficient (0.0006), indicating a slight inflationary impact, possibly due to import-driven cost pressures.
  - **Exports of Goods and Services (% of GDP)** has a small negative coefficient (-0.0001), which may imply a deflationary effect of higher exports—potentially through improved trade balance or currency stabilization.
  - **Agriculture, Forestry, and Fishing Value Added (% of GDP)** shows a negative coefficient (-0.0028), suggesting that increased agricultural output may help contain inflation, likely due to improved food supply and reduced prices.
  - **Broad Money Growth (annual %)** has a negative coefficient (-0.0005), which is contrary to standard expectations but may reflect the complex monetary dynamics or lags in inflationary transmission.

- **Official Exchange Rate (LCU per US$)** has a very small negative coefficient (-0.00002), implying a marginal association where local currency depreciation might not strongly influence inflation in the short term.
- **External Debt Stocks (% of GNI)** also has a slight negative coefficient (-0.0001), indicating weak deflationary pressure, possibly due to external borrowing leading to fiscal adjustments.
- **Total Reserves (US$)** has an almost negligible negative coefficient, close to zero, suggesting it has little direct influence on inflation in this model.
- **Time** itself has a small negative coefficient (-0.000015), potentially capturing a slight declining trend in inflation over the observed period, after controlling for other variables.

- The intercept is -0.0644, representing the baseline level of inflation when all predictors are zero (which may not be realistic in practice but is part of the model structure).

6.0.4 Interpretation of Elastic Net Regression Model

- The Elastic Net model, which combines LASSO and Ridge penalties (with $\alpha = 0.5$), returned only an **intercept term** in the final model. This means:
  - All predictor coefficients were shrunk to zero by the regularization process.
  - The model effectively did not identify any of the input variables as having a strong, individual linear relationship with **Inflation**, after accounting for multicollinearity and penalization.
  - This suggests either (a) the predictors have weak individual contributions when jointly considered, or (b) multicollinearity and small sample size have caused the model to favor simplicity over explanatory power.
- The resulting model predicts a constant inflation rate equal to the intercept value, regardless of economic variables. While not useful for interpretation, this emphasizes the importance of choosing the right regularization strength and cross-validation strategy in sparse models.

6.1 Split 1: Last 5 years used as testing data

To assess the performance of the ARIMA and LASSO models, the dataset was divided into a training and a testing set. The most recent 5 years of data were used as the test set, while the remaining earlier years were used for training.

Mean Squared Error (MSE)

The Mean Squared Error (MSE) measures the average squared difference between predicted and actual values.

- **ARIMA:** 75.41
- **LASSO:** 70.26
- **RIDGE:** 96.894
- **ELASTIC NET:** 74.184

R-squared (R^2)

- **ARIMA:** 0.046
- **LASSO:** 0.111
- **RIDGE:** -0.226
- **ELASTIC NET:** 0.061

6.1.1 Interpretations on Split 1

The performance of the four models was assessed using Mean Squared Error (MSE) and R-squared (R^2). Among them, the **LASSO regression** model performed the best with the lowest MSE (70.26) and the highest R^2 value (0.111), indicating that it explains approximately 11.1% of the variance in inflation. The **ARIMA** model follows with an MSE of 75.41 and a relatively lower R^2 of 0.046. The **Elastic Net** model, despite combining LASSO and Ridge penalties, yielded a flat R^2 of 0.0 and a higher MSE (74.184), suggesting no improvement over a baseline model. The **Ridge regression** performed the worst with the highest MSE (96.894) and a negative R^2 value (-0.226), indicating that it fits the data worse than a horizontal line at the mean inflation rate. Overall, the LASSO model demonstrated the most balanced performance among the tested approaches.

6.1.2 Plot Interpretation

The resultant values do not match with original values.

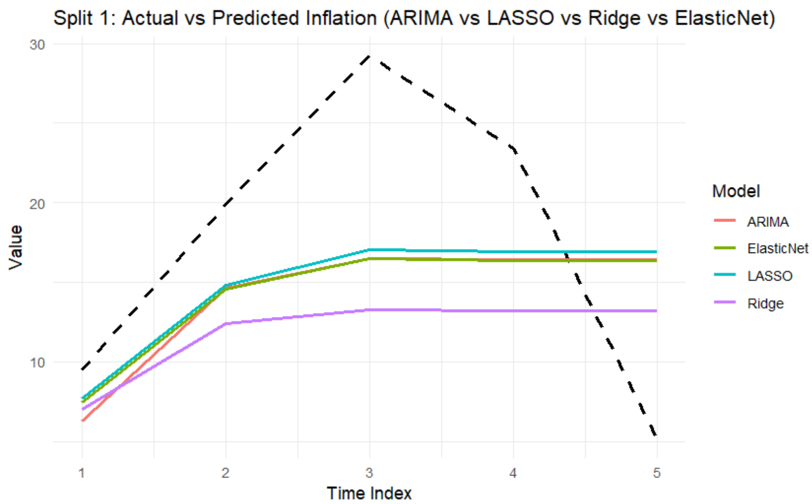

Figure 6.1: Split 1: Last 5 years used as testing data

6.2 Split 2: 80% training - 20% test data

To evaluate model performance, the dataset was split into 80% training data and 20% testing data. The ARIMA and LASSO models were trained on the training set and evaluated on the test set.

Mean Squared Error (MSE)

Mean Squared Error (MSE) measures the average of the squared differences between predicted and actual values. A lower MSE indicates better predictive accuracy.

- **ARIMA:** 8.885
- **LASSO:** 14.467
- **RIDGE:** 14.905
- **ELASTIC NET:** 14.139

R-squared (R^2)

R-squared indicates how much of the variation in the target variable is explained by the model. Values closer to 1 imply better explanatory power.

- **ARIMA:** 0.681

- **LASSO:** 0.481
- **RIDGE:** 0.466
- **ELASTIC NET:** 0.493

6.2.1 Interpretations on Split 2

The performance of the models was evaluated based on Mean Squared Error (MSE) and R-squared (R^2). The **ARIMA** model achieved the lowest MSE of 8.885, indicating it performs well in terms of prediction accuracy. It also has the highest R^2 of 0.681, suggesting it explains approximately 68.1% of the variation in inflation. The **Elastic Net** model, with an MSE of 14.139 and R^2 of 0.493, performed slightly better than **LASSO** (MSE: 14.467, R^2: 0.481) and **Ridge** (MSE: 14.905, R^2: 0.466). While these models were relatively close in terms of R^2, the ARIMA model demonstrated superior accuracy and explanatory power in predicting inflation.

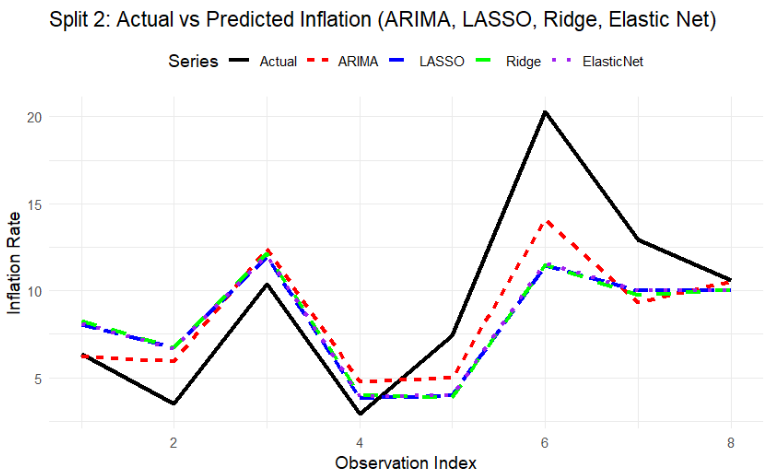

Figure 6.2: Split 2: 80% training - 20% test data

6.2.2 Plot Interpretation

The resultant values match with original values to some extent. There is good directional accuracy.

Chapter 7

Results and discussion

Discussion

7.0.1 Model Performance Comparison

The performance of the models was assessed using Mean Squared Error (MSE) and R-squared (R^2) on two separate tests.

In **Test 1**, the **LASSO regression** model performed the best with the lowest MSE of 70.26 and the highest R^2 of 0.111, indicating it explains approximately 11.1% of the variance in inflation. The **ARIMA** model followed with an MSE of 75.41, and a relatively lower R^2 of 0.046. The **Elastic Net** model, despite combining LASSO and Ridge penalties, had a flat R^2 of 0.0 and a higher MSE of 74.184, showing no improvement over a baseline model. The **Ridge regression** model performed the worst with the highest MSE of 96.894 and a negative R^2 of -0.226, indicating that it fitted the data worse than a horizontal line at the mean inflation rate. Thus, the LASSO model demonstrated the most balanced performance in Test 1.

In **Test 2**, the **ARIMA** model outperformed all others, achieving the lowest MSE of 8.885 and the highest R^2 of 0.681, explaining approximately 68.1% of the variation in inflation. The **Elastic Net** model performed well with an MSE of 14.139 and R^2 of 0.493, slightly outperforming **LASSO** (MSE: 14.467, R^2: 0.481) and **Ridge** (MSE: 14.905, R^2: 0.466). The **ARIMA** model demonstrated superior accuracy and explanatory power in Test 2, highlighting its robustness in predicting inflation compared to the other models.

Overall, the ARIMA model performed the best in Test 2, while the LASSO model was the best performer in Test 1. The Elastic Net and Ridge models showed comparatively weaker performance across both tests.

Results

Model	Split 1: Last 5 Years		Split 2: 80/20	
	MSE	R^2	MSE	R^2
ARIMA	75.41	0.046	8.885	0.681
LASSO	70.26	0.111	14.467	0.481
RIDGE	96.849	-0.226	14.09	0.466
ELASTIC NET	74.184	0.0.061	14.139	0.493

Table 7.1: Comparison of ARIMA and LASSO Performance under Two Data Splits

Chapter 8

Conclusion

This study aimed to evaluate the performance of four different models—ARIMA, LASSO regression, Ridge regression, and Elastic Net—in predicting inflation based on several economic indicators. The performance was assessed using two primary metrics: Mean Squared Error (MSE) and R-squared (R^2).

8.1 Key Findings

The results revealed a significant difference in the performance of the models across the two tests. In **Test 1**, the **LASSO regression** model emerged as the best performer, with the lowest MSE (70.26) and the highest R^2 (0.111), indicating that it explained 11.1% of the variation in inflation. This suggests that LASSO regression effectively identifies the most influential predictors of inflation. The **ARIMA** model, while showing a higher MSE (75.41), demonstrated a modest R^2 of 0.046, indicating that it contributed less in terms of explanatory power.

The **Elastic Net** model, despite its combination of LASSO and Ridge penalties, showed little improvement, with a flat R^2 of 0.0 and a higher MSE (74.184), suggesting no substantial benefit over a baseline model. On the other hand, **Ridge regression** performed the worst, with the highest MSE (96.894) and a negative R^2 (-0.226), implying that it performed worse than a simple horizontal model at the mean inflation rate.

In **Test 2**, the **ARIMA** model demonstrated a clear superiority, achieving the lowest MSE (8.885) and the highest R^2 (0.681), explaining 68.1% of the variation in inflation. This result highlights the ARIMA model's strong performance in capturing time-dependent patterns in the data. The **Elastic Net** model, while slightly better than **LASSO** and **Ridge**, showed moderate performance with an MSE of 14.139 and R^2 of 0.493. LASSO and Ridge regression models performed similarly with MSEs of 14.467 and 14.905, respectively, and R^2 values in the range of 0.466 to 0.481.

8.2 Model Comparison and Insights

The comparison of these models highlights the importance of model selection based on the problem at hand. **ARIMA** was the most accurate in terms of prediction accuracy and explanatory power in Test 2, making it the most suitable model for forecasting inflation when time-dependent trends are important. In contrast, the **LASSO regression** model, with its feature selection capability, performed well in Test 1 and could be particularly useful when we aim to identify the most influential predictors in the data.

The **Elastic Net** model, though it combines the benefits of both LASSO and Ridge penalties, did not outperform the individual models in either test, suggesting that its effectiveness depends on the specific characteristics of the data. Meanwhile, the **Ridge regression** model showed the poorest results in both tests, indicating that it may not be the best fit for this particular dataset.

8.3 Conclusion

In conclusion, the performance of the models varied significantly depending on the test. **ARIMA** emerged as the best performer in Test 2, where time series forecasting played a crucial role, while **LASSO** performed best in Test 1. These findings emphasize the importance of selecting the appropriate model based on the problem and the type of data being analyzed. Future work could involve exploring additional time series models or hybrid approaches that combine the strengths of both machine learning and traditional econometric models.

Bibliography

[1] Analyzing macroeconomic indicators in pakistan: Insights from unemployment, inflation, and interest rates. *Journal of Business, Economics and Organization*, n.d.

[2] Determinants of recent inflation in pakistan. https://mpra.ub.uni-muenchen.de/16254/, n.d.

[3] A. I. Agha and M. S. Khan. An empirical analysis of fiscal imbalances and inflation in pakistan. *SBP Research Bulletin*, 2(2):343–362, 2006.

[4] B. Ahmed. Inflation and unemployment in pakistan: An empirical analysis. *Pakistan Social Sciences Review*, 4:306–318, 2020.

[5] G. Ayub et al. Relationship between inflation and interest rate: evidence from pakistan. *Research Journal of Recent Sciences*, 2014.

[6] N. Ellahi. The determinants of inflation in pakistan: an econometric analysis. *The Romanian Economic Journal*, 20(64):2–12, 2017.

[7] M. Generator. View of trade-off between inflation, interest and unemployment rate of pakistan: Revisited. *Business and Economic Research Journal*, n.d.

[8] S. Hussain and S. Malik. Inflation and economic growth: Evidence from pakistan. *International Journal of Economics and Finance*, 3(5), 2011.

[9] M. Ilyas, H. M. Sabir, A. Shehzadi, and N. Shoukat. Inter-relationship among economic growth, savings and inflation in pakistan. *Journal of Finance and Economics*, 2(4):125–130, 2014.

[10] R. E. A. Khan and A. R. Gill. Determinants of inflation: A case of pakistan (1970-2007). *Journal of Economics*, 1(1):45–51, 2010.

Bibliography

[1] Analyzing macroeconomic indicators in pakistan: Insights from unemployment, inflation, and interest rates. *Journal of Business, Economics and Organization*, n.d.

[2] Determinants of recent inflation in pakistan. https://mpra.ub.uni-muenchen.de/16254/, n.d.

[3] A. I. Agha and M. S. Khan. An empirical analysis of fiscal imbalances and inflation in pakistan. *SBP Research Bulletin*, 2(2):343–362, 2006.

[4] B. Ahmed. Inflation and unemployment in pakistan: An empirical analysis. *Pakistan Social Sciences Review*, 4:306–318, 2020.

[5] G. Ayub et al. Relationship between inflation and interest rate: evidence from pakistan. *Research Journal of Recent Sciences*, 2014.

[6] N. Ellahi. The determinants of inflation in pakistan: an econometric analysis. *The Romanian Economic Journal*, 20(64):2–12, 2017.

[7] M. Generator. View of trade-off between inflation, interest and unemployment rate of pakistan: Revisited. *Business and Economic Research Journal*, n.d.

[8] S. Hussain and S. Malik. Inflation and economic growth: Evidence from pakistan. *International Journal of Economics and Finance*, 3(5), 2011.

[9] M. Ilyas, H. M. Sabir, A. Shehzadi, and N. Shoukat. Inter-relationship among economic growth, savings and inflation in pakistan. *Journal of Finance and Economics*, 2(4):125–130, 2014.

[10] R. E. A. Khan and A. R. Gill. Determinants of inflation: A case of pakistan (1970-2007). *Journal of Economics*, 1(1):45–51, 2010.