Inflation Forecasting in Pakistan: Using Advanced Statistical Models

Name	Roll Number
Shaista Shahid	23i-2519
Eman Ali	23i-2564

Section: BDS-4A
Course Name: Advance Statistics
Instructor Name: Sir Hazber Samson
Submission Date: May 4, 2025

Contents

1. Introduction

1.1 Overview of Inflation

Inflation is a fundamental concept in macroeconomics, reflecting the general rise in the prices of goods and services within an economy over a certain period of time. It is measured as the percentage increase in the Consumer Price Index (CPI) or other similar indices that track the cost of a fixed basket of goods. Inflation erodes the purchasing power of money, meaning that each unit of currency buys fewer goods and services. While inflation is an inevitable part of economic growth, its excessive or volatile levels can be harmful to an economy.

In Pakistan, inflation has been a persistent challenge, fluctuating significantly due to a variety of internal and external factors. Key contributors include changes in energy prices, food costs, exchange rates, and monetary policy decisions. Inflation in Pakistan has historically been high and unpredictable, with periods of hyperinflation, particularly during times of fiscal deficit, political instability, or global economic crises. Consequently, understanding inflationary trends and accurately forecasting future inflation rates has become an essential task for policymakers and financial analysts in the country.

1.2 Importance of Inflation Forecasting in Pakistan

Inflation forecasting is of paramount importance for countries like Pakistan, where the economy is subject to both domestic structural imbalances and external shocks. Forecasting inflation helps policymakers, particularly the State Bank of Pakistan (SBP), in adjusting monetary policies (such as interest rates and money supply) to stabilize prices and ensure economic stability. A predictable inflationary environment fosters confidence in investment, allows for better budget planning, and helps consumers and businesses make informed financial decisions.

In Pakistan, inflation forecasting is also crucial for long-term economic planning. Given the country's dependence on imports for key goods and its vulnerability to currency devaluation, accurately predicting inflation trends allows the government to implement preemptive actions to control price instability and mitigate the adverse effects of rising inflation on lower-income populations. Moreover, effective inflation forecasting is key to shaping wage negotiations, social security programs, and pension schemes, as well as for managing external debts.

1.3 Objectives of the Study

The primary objective of this study is to develop and compare inflation forecasting models for Pakistan, using both traditional statistical methods and advanced machine learning techniques. Specifically, the study aims to:

- Analyze the relationship between inflation and key macroeconomic variables such as interest rates, money supply (M2), unemployment rate, exchange rates, and GDP growth.
- Evaluate the performance of several forecasting models, including:

  - ARIMA (AutoRegressive Integrated Moving Average) model
  - Ridge Regression
  - LASSO (Least Absolute Shrinkage and Selection Operator)
  - Elastic Net

- Compare forecasted values with actual inflation data using evaluation metrics such as Mean Squared Error (MSE).
- Identify the most effective model for forecasting inflation in Pakistan.
- Contribute to the literature on inflation forecasting in developing countries, especially Pakistan.

2. Literature Review

2.1 Evaluating the Performance of Inflation Forecasting Models of Pakistan

The study titled *"Evaluating the Performance of Inflation Forecasting Models of Pakistan"* compares various inflation forecasting models by utilizing variables such as the Consumer Price Index (CPI), money supply, output gap, and global oil prices. The research tests a range of models including ARDL, VAR, BVAR, and several benchmark models like Random Walk (RW), ARIMA, and AR(1). The study finds that ARDL models perform best during periods of high inflation, whereas forecast averaging methods are more effective in moderate inflation environments. These results emphasize the state-dependent nature of forecasting accuracy and underscore the importance of selecting appropriate models based on the prevailing inflation regime.

2.2 Inflation Forecasting in Pakistan Using Artificial Neural Networks

Haider and Hanif's study titled *"Inflation Forecasting in Pakistan Using Artificial Neural Networks"* compares ANN, AR(1), and ARIMA models for predicting Pakistan's YoY CPI inflation. Using a univariate approach with lagged inflation data, their 12-layer ANN model outperformed traditional methods with lower RMSE. The ANN forecasts revealed an upward inflation trend for FY2008, demonstrating its effectiveness for economic forecasting in developing economies. The results suggest ANN's superiority over conventional time-series models for inflation prediction.

2.3 Inflation in Pakistan

The study *"Inflation in Pakistan"* by Mohsin S. Khan and Axel Schimmelpfennig analyzes inflation determinants using monetary variables (broad money, private sector credit), real

GDP, exchange rates, interest rates, and wheat support prices. The authors employ autoregressive distributed lag (ADL) models and vector error correction models (VECM) on monthly data (1998–2005). Results highlight monetary factors (12-month lag) as dominant inflation drivers, with private sector credit growth serving as a key leading indicator. The leading indicator model (LIM) forecasts inflation accurately, reinforcing monetary policy's role in stabilizing prices.

2.4 Three Attempts at Inflation Forecasting in Pakistan

The paper *"Three Attempts at Inflation Forecasting in Pakistan"* by Madhavi Bokil and Axel Schimmelpfennig compares three models: a leading indicators model (LIM), a univariate ARIMA model, and a VAR model with Phillips-curve dynamics. Key variables include CPI inflation, broad money growth, credit growth, output gap, and interest rates. The LIM, based on monetary indicators, outperforms the others in forecasting accuracy. The ARIMA model struggles with capturing turning points, while the VAR model underperforms despite incorporating structural relationships. The study focuses on post-1998 data to address structural breaks and uses 12-month moving averages to manage seasonality. Overall, monetary variables with a 12-month lag emerge as the strongest predictors.

2.5 Modeling and Forecasting Pakistan's Inflation by Using Time Series ARIMA Models

The research paper *"Modeling and Forecasting Pakistan's Inflation by Using Time Series ARIMA Models"* by Muhammad Abdus Salam, Shazia Salam, and Mete Feridun focuses on short-term inflation forecasting using ARIMA time series models. The study emphasizes CPI as the primary variable, along with other indices like WPI, SPI, and the GDP deflator. It outlines practical steps for applying ARIMA methodology, selecting the best model based on diagnostic and evaluation criteria. Both in-sample and out-of-sample forecasts show that ARIMA has strong predictive power. The paper concludes that ARIMA is a reliable tool for monthly inflation forecasting in Pakistan.

2.6 Threshold Inflation in Pakistan

In the research paper *"Threshold Inflation in Pakistan"* published in the SBP Research Bulletin Vol-13, No.1, 2017, the authors explore the relationship between CPI inflation and real GDP growth to identify the threshold level of inflation that begins to negatively affect economic growth. The study employs two econometric models: a quadratic regression model, where inflation and its squared term are included, and a regression kink model based on Hansen's (2017) methodology. The only additional explanatory variable used in the models is real investment growth, selected based on significance and the Akaike Information Criterion (AIC). The findings aim to inform monetary policy as Pakistan considers adopting a flexible inflation targeting regime

2.7 Does Globalization Matter for Inflation in pakistan?

The research paper titled *"Does Globalization Matter for Inflation in Pakistan?* Empirical Evidence from Global Slack Hypothesis" is authored by Mukhtiar Khan, Wajid Ali, and Asmat Khan. It explores how globalization affects inflation in Pakistan, focusing on the global slack hypothesis. The authors use time-series data to examine the impact of global output gaps on domestic inflation. Their findings show that global economic conditions significantly influence inflation in Pakistan. This suggests that policymakers should consider global factors when managing inflation.

2.8 Impact of Inflation on Economic Growth in Pakistan

Ijaz Uddin's (2021) research paper, *"Impact of Inflation on Economic Growth in Pakistan,"* explores the relationship between Gross Domestic Product (GDP) growth and inflation in Pakistan, using time series data from 1990 to 2015. The study employs the Augmented Dickey Fuller (ADF) test for stationarity and the Engel Granger Co-integration test to analyze both short-run and long-run associations. The findings indicate a positive and significant relationship between inflation and GDP growth, with a 1(percent) increase in inflation leading to a 0.27(percnt) increase in GDP.

2.9 Determinants of Inflation in Pakistan

The research paper titled *Determinants of Inflation in Pakistan* An Econometric Analysis Using Johansen Co-Integration Approach" by Furrukh Bashir, Shahbaz Nawaz, Kalsoom Yasin, Usman Khursheed, and Jahanzeb Khan, investigates the factors influencing inflation in Pakistan. The study focuses on both demand-side and supply-side determinants of inflation, considering variables such as money supply, government expenditures, exports, wages, and fiscal deficits. The authors use the Johansen Co-Integration approach to examine the long-term relationships between macroeconomic variables and inflation. This econometric model helps in identifying the causal effects of various factors on inflation in the Pakistani economy.

2.10 Inflation in Pakistan: Money or Wheat?

The research paper *"Inflation in Pakistan: Money or Wheat?"* by Mohsin S. Khan and Axel Schimmelpfennig examines the driving forces behind inflation in Pakistan. The study analyzes the role of monetary factors, including money supply, private sector credit, and the exchange rate, as well as the wheat support price as a supply-side factor. Using a stylized inflation model, the authors estimate the model with monthly data from January 1998 to June 2005. The results suggest that while monetary factors are the primary drivers of inflation with a lag of about one year, the wheat support price affects inflation in the short run and influences inflation over the medium term only if accommodated by monetary policy.

3. Description of Variables and Data Description

This research examines the dynamics of inflation in Pakistan and, in particular, how past macroeconomic trends have been used to predict future inflation patterns. The evidence is based on a lengthy dataset covering 136 years (1886–2021), which provides a deep historical framework within which to analyze the underlying structural causes of inflation in Pakistan. The data are sourced from historical records, World Bank, IMF, Pakistan Bureau of Statistics, and other pertinent historical economic documents.

3.1 Dependent Variable

Inflation (Consumer Prices, Annual %):
This is the variable of particular interest, expressed as the year-on-year change in the Consumer Price Index (CPI). It is a general indicator of the cost of living and indicates changes in the overall price level over time. Due to the long historical coverage, the earliest CPI estimates (before independence) are reconstructed using colonial and initial administrative records.

3.2 Independent Variables

To explain and predict inflation, the following macroeconomic variables are utilized as explanatory variables:

- **Exchange Rate (PKR/USD or historical equivalents):**
  Captures the relative value of the Pakistani Rupee versus the US Dollar. Historical exchange rate proxies are utilized for years prior to Pakistan's independence in 1947. Exchange rate volatility influences import prices and foreign inflation pressures.

- **GDP Growth (Annual %):**
  Measures the growth rate of Gross Domestic Product annually. High growth may result in higher demand and possible inflationary pressure. Pre-independence GDP is reconstructed on the basis of historical estimates of output and trade volumes.

- **Unemployment Rate (% of labor force):**
  Shows the percentage of the labor force that is unemployed. Unemployment affects wage dynamics and consumer expenditure.

- **Broad Money Supply (M2, Annual % Growth):**
  A major monetary aggregate that encompasses currency in circulation and demand deposits. Historical estimates are derived from central bank records and historical financial accounts.

- **Exports (as % of GDP):**
  The value of goods and services exported to other nations. Export performance influences foreign exchange earnings and indirectly affects inflation through trade balances.

- **Imports (as % of GDP):**
  Refers to the value of foreign goods and services bought. Dependence on imports, particularly of energy and food, makes the domestic inflation vulnerable to global price fluctuations.
- **Oil Rents (% of GDP):**
  Refers to oil production revenue. While Pakistan is not a country with natural oil reserves, the effect of world oil prices on domestic inflation is recorded here. Data in early periods corresponds to dependence on coal and kerosene as proxies for energy.
- **Remittances (Personal Transfers, % of GDP):**
  Remittances from foreign Pakistani workers. Such inflows increase income for households and domestic consumption, which may result in demand-pull inflation.

4. Data Preparation and Exploration

4.1 Data Import and Renaming

`The data includes annual observations from 1986 to 2021 for several economic variables. After loading, we renamed the columns for easier reference. The variables included:`

- Inflation (Consumer Prices - annual %)
- Exchange Rate
- GDP Growth
- Unemployment
- Broad Money (Money Supply)
- Exports (% of GDP)
- Imports (% of GDP)
- Oil Rents (% of GDP)
- Remittances (% of GDP)

Summary Statistics

We used the `summary()` function to get a basic understanding of each variable. This gave us the minimum, maximum, mean, median, and quartiles. For example:

- Inflation ranged from 2.5% to 20.3%, with a mean of around 8.24%.
- The Exchange Rate ranged from 16.65 to 162.91, showing a large increase over time.
- GDP Growth had a minimum of -1.3% and a maximum of 7.8%.

- Remittances varied widely, from about 1% to 31% of GDP.

These statistics help us understand the spread and distribution of our data and identify any major outliers or skewed variables.

```
> # Summary statistics
> summary(df$BroadMoney)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  224.8   788.3  2560.0  6766.2 10662.5 28620.0 
> summary(df$Inflation)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.500   4.625   8.350   8.244  10.450  20.300 
> summary(df$ExchangeRate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  16.65   31.37   59.62   66.85   95.33  162.91 
> summary(df$GDPGrowth)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 -1.300   2.925   4.400   4.347   5.650   7.800 
> summary(df$Unemployment)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.100   5.350   5.900   5.783   6.225   8.300 
> summary(df$Exports)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.920   9.297  14.385  17.469  27.902  31.550 
> summary(df$Imports)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   6.37   10.85   15.87   28.20   48.46   67.82 
> summary(df$OilRents)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.20    0.40    0.50    0.60    0.75    1.20 
> summary(df$Remittances)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.996   1.710   3.950   8.237  14.165  31.310 
> 
```

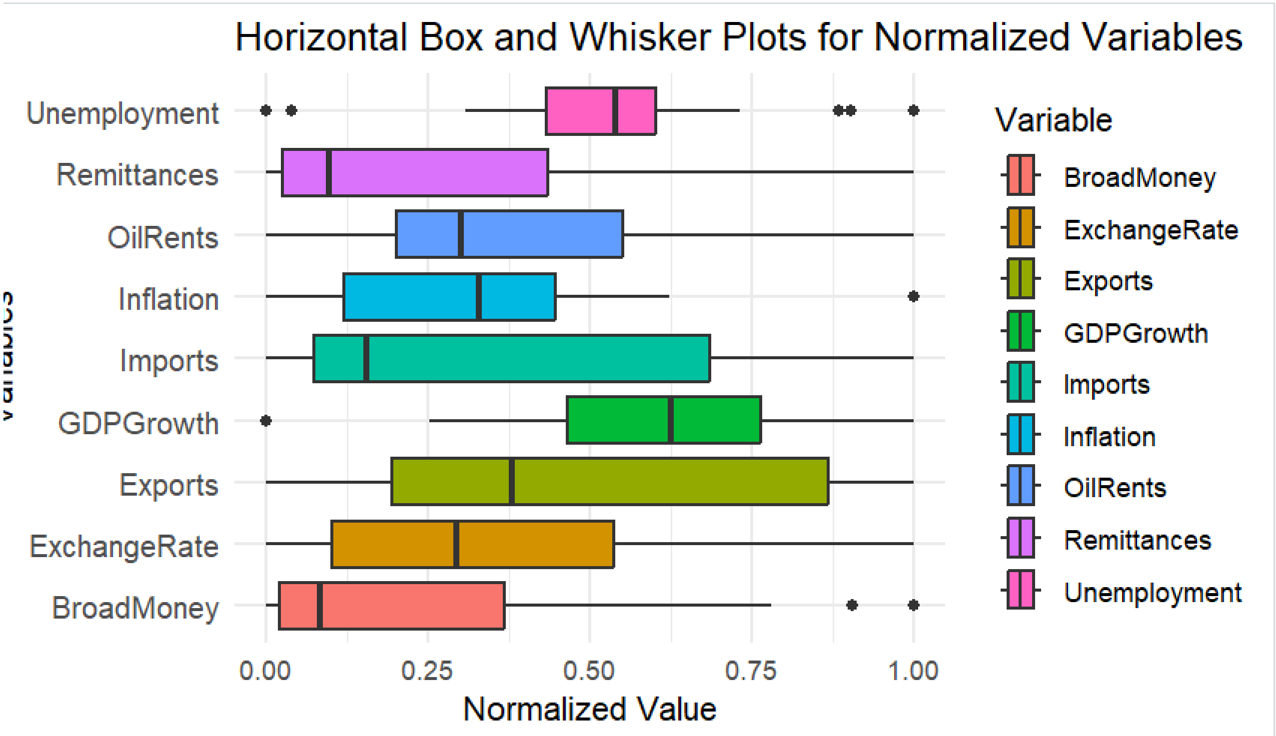
Normalization

To prepare the data for modeling, we normalized all variables (except the year). Normalization helps bring all variables into the same scale (0 to 1), which is important for models like Ridge, LASSO, and Elastic Net that are sensitive to the scale of data.

The formula used for normalization was:

$$\text{Normalized} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Boxplot Visualization



4.2 Data Normalization, Reshaping, and Visualization

After normalizing, we reshaped the data into a long format using `pivot_longer()` so we could easily plot it. Then, we created a horizontal boxplot using `ggplot2` to see the distribution of all normalized variables side-by-side. This plot helped visualize the central tendency, spread, and potential outliers of each variable. For instance, if any variable had a longer tail or unusual spikes, we could identify it easily through this boxplot.

Boxplot Interpretation

The horizontal box-and-whisker plot shows the distribution of all normalized variables used in our inflation forecasting study. The values have been scaled between 0 and 1 for uniform comparison.

Key Observations

- Broad Money and Exchange Rate have most of their values in the lower to mid-normalized range. This suggests that earlier years had lower values, which steadily increased over time.
- Exports and Imports are more spread out across the entire range, indicating significant variation across the years.

- GDP Growth shows a fairly symmetrical distribution, suggesting balanced ups and downs over the time period.
- Inflation itself appears mostly centered, with values not showing too many extreme spikes.
- Remittances have a wide spread, with many lower values and a long upper tail. This suggests remittances were low in earlier years and have increased significantly in more recent years.
- Unemployment is clustered in a narrow range, indicating it didn't vary as much over time.
- Oil Rents show a tight distribution, indicating relatively low variation, with no extreme highs.

Outliers

Outliers are shown as dots beyond the whiskers of each boxplot. These represent years where the values were unusually high or low compared to the rest.

- Unemployment has multiple outliers on both the lower and higher sides, indicating certain years where unemployment deviated significantly from the average trend.
- Broad Money, Remittances, and Inflation also show outliers, which may be due to sharp economic policy shifts, global events, or inflationary shocks.
- GDP Growth shows a lower outlier, which likely reflects a year of economic downturn or recession.
- Imports and Exports show long whiskers but fewer outliers, suggesting more continuous but variable trends.

4.3 Correlation and Scatter Plot Analysis

After normalizing the data, we performed a correlation analysis and created scatter plots to visually explore the relationships between the dependent variable (Inflation) and all the independent variables.

Scatter Plot Matrix

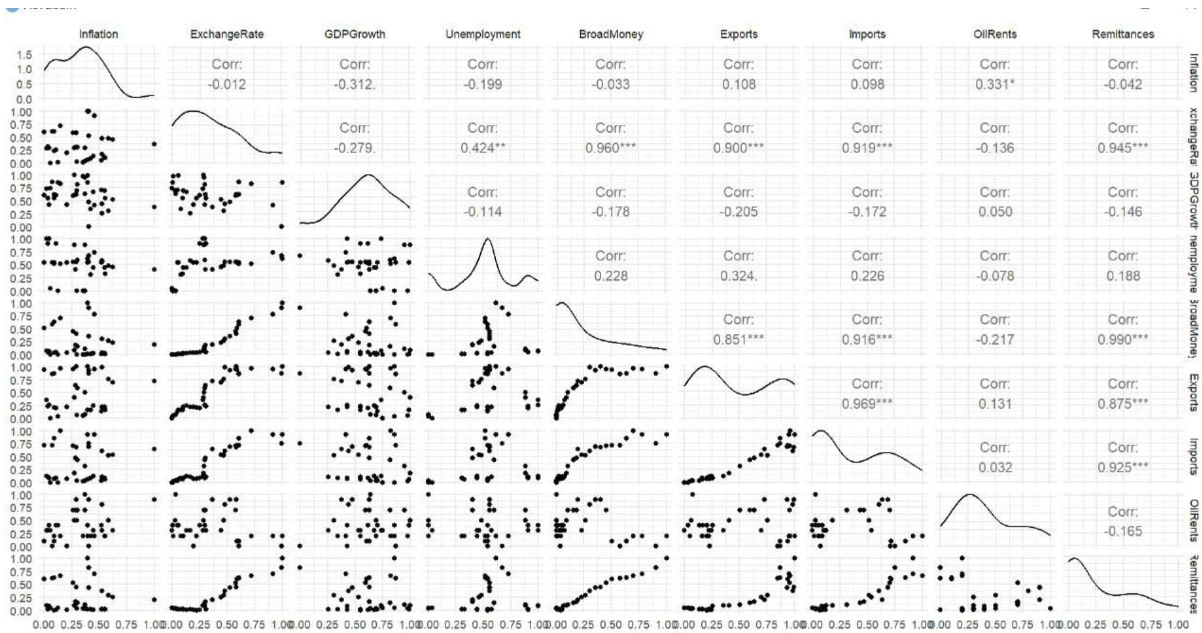
We started with a scatter plot matrix using `ggpairs()` from the `GGally` package. This matrix displays pairwise relationships among all the numeric variables in the dataset. It provides:

- A quick visual check for linear or non-linear relationships.
- Possible patterns or clusters in the data.

- Initial signs of positive or negative correlations.

From the matrix, we can observe:

- Some variables show moderate to weak linear relationships with Inflation.
- For example, variables like Exchange Rate, Broad Money, and Imports appear to show stronger visual associations compared to others like Oil Rents or Unemployment.



Individual Scatter Plots

To explore each relationship in more detail, we created individual scatter plots between Inflation and each independent variable:

- Each plot includes:
  - Blue data points showing the relationship.
  - A red dashed regression line to highlight the trend using linear regression.

Key Observations:

- **Exchange Rate vs Inflation**: A clear upward trend suggests that as the exchange rate increases, inflation tends to rise, possibly due to import price effects.
- **Broad Money vs Inflation**: Also shows a positive relationship, which makes sense as increased money supply often contributes to inflation.
- **GDP Growth vs Inflation**: The relationship is weaker and may not be linear.
- **Imports and Exports**: Imports seem to have a more noticeable upward slope than exports, implying higher imports might contribute to inflationary pressure.

- **Remittances, Oil Rents, Unemployment**: These variables have weaker or scattered relationships with inflation, suggesting their effects might be indirect or lagged.

4.4 Interpretation of Scatter Plots of Inflation vs. Independent Variables

1. Inflation vs. Exchange Rate

- **Observation:** The scatter plot reveals an upward trend, indicating that as the exchange rate increases, inflation tends to rise.
- **Implication:** A weakening domestic currency (higher exchange rate) may lead to higher import prices, contributing to inflationary pressures.

2. Inflation vs. GDP Growth

- **Observation:** The relationship appears weak and somewhat scattered, with no clear linear trend.
- **Implication:** Changes in GDP growth do not show a consistent effect on inflation within this dataset, suggesting other factors may play more significant roles.

3. Inflation vs. Unemployment

- **Observation:** The scatter suggests a slight negative trend, but overall, the relationship is weak and dispersed.
- **Implication:** This weak correlation aligns with the classical or Phillips curve theories, where the trade-off between unemployment and inflation is not strongly observed here.

4. Inflation vs. Broad Money

- **Observation:** There is a clear upward trend, indicating that increases in broad money supply are associated with higher inflation.
- **Implication:** Expanding the monetary base can be inflationary, supporting monetary policy considerations.

5. Inflation vs. Exports

- **Observation:** The scatter shows a slight upward slope, but the relationship is not very strong.
- **Implication:** Higher exports may contribute to inflation, possibly through increased demand or currency appreciation, but the effect appears modest.

6. Inflation vs. Imports

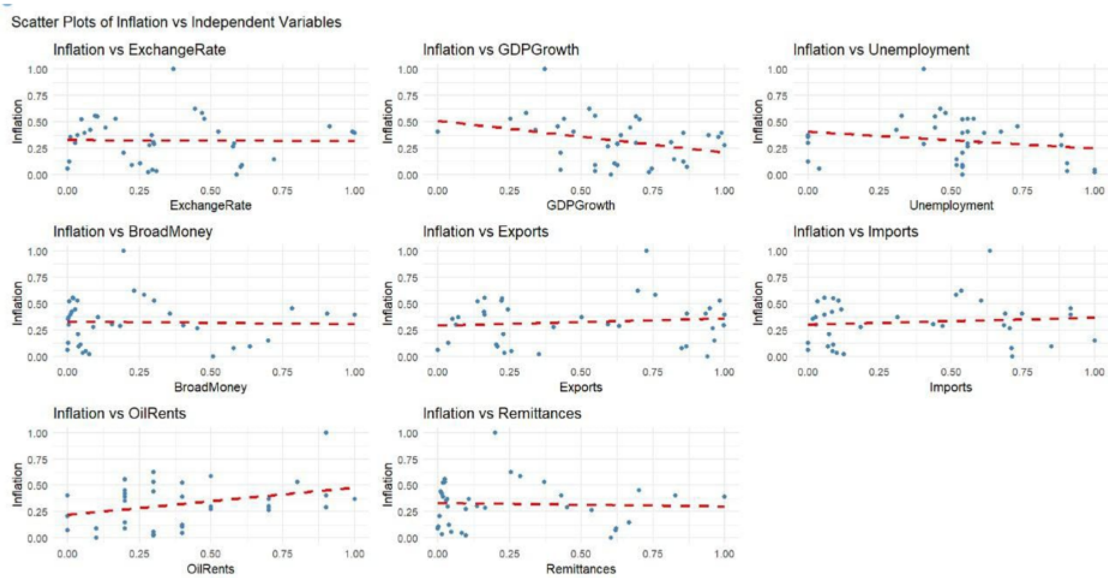
- **Observation:** The plot displays a more noticeable upward trend compared to exports.
- **Implication:** Increased imports might elevate inflation, potentially due to import price effects or trade deficits fueling domestic prices.

7. Inflation vs. Oil Rents

- **Observation:** The distribution of points is scattered with a slight upward trend.
- **Implication:** Oil rents could have a mild positive impact on inflation, possibly through increased revenue flows influencing aggregate demand or government spending.

8. Inflation vs. Remittances

- **Observation:** The data points are dispersed without a clear pattern, with slight fluctuations.
- **Implication:** Remittances might have an insignificant or complex relationship with inflation, potentially influencing inflation with a lag or through indirect pathways.



Summary

The scatter plots highlight that **Exchange Rate**, **Broad Money**, and **Imports** show the most consistent positive relationships with inflation. These variables could be significant drivers of inflationary trends in the observed context. Other variables like **GDP Growth**, **Unemployment**, **Oil Rents**, and **Remittances** display weaker and less consistent relationships, suggesting their effects are either indirect, lagged, or less impactful within this dataset.

5. Estimation of the Models

5.1 ARIMA Model for Inflation Forecasting

Model Overview

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful tool for time series forecasting. It combines three components:

- **AR (AutoRegressive):** This component uses the dependency between an observation and a number of lagged observations (previous time periods).
- **I (Integrated):** The integrated part makes the series stationary by differencing the series (subtracting a previous observation from the current observation).
- **MA (Moving Average):** This part models the relationship between an observation and a residual error from a moving average model applied to lagged observations.

In this project, ARIMA was used to forecast inflation in Pakistan over the next five years, based on historical data.

Model Fitting

To begin, the ARIMA model was applied to the inflation time series data (stored in `inflation_ts`). The ARIMA model was fitted using the `auto.arima()` function from the `forecast` package in R. This function automatically selects the best model by minimizing the Akaike Information Criterion (AIC), which is a common method for model selection in time series forecasting.

After fitting the model, inflation was forecasted for a 5-year horizon using the `forecast()` function, where the forecast horizon was set to 5 years (`h = 5`). This forecast represents the predicted inflation values for the next five years.

Error Metrics - R^2

To assess the performance of the ARIMA model, the R^2 value was calculated. This value indicates how well the model explains the variance in the data. An R^2 value closer to 1 means a better fit, while a lower R^2 indicates a poorer fit.

- **Sum of Squared Errors (SSE):**

$$\text{SSE} = \sum (\text{residuals})^2$$

- **Total Sum of Squares (SST):**

$$\text{SST} = \sum (\text{inflation data} - \text{mean of inflation data})^2$$

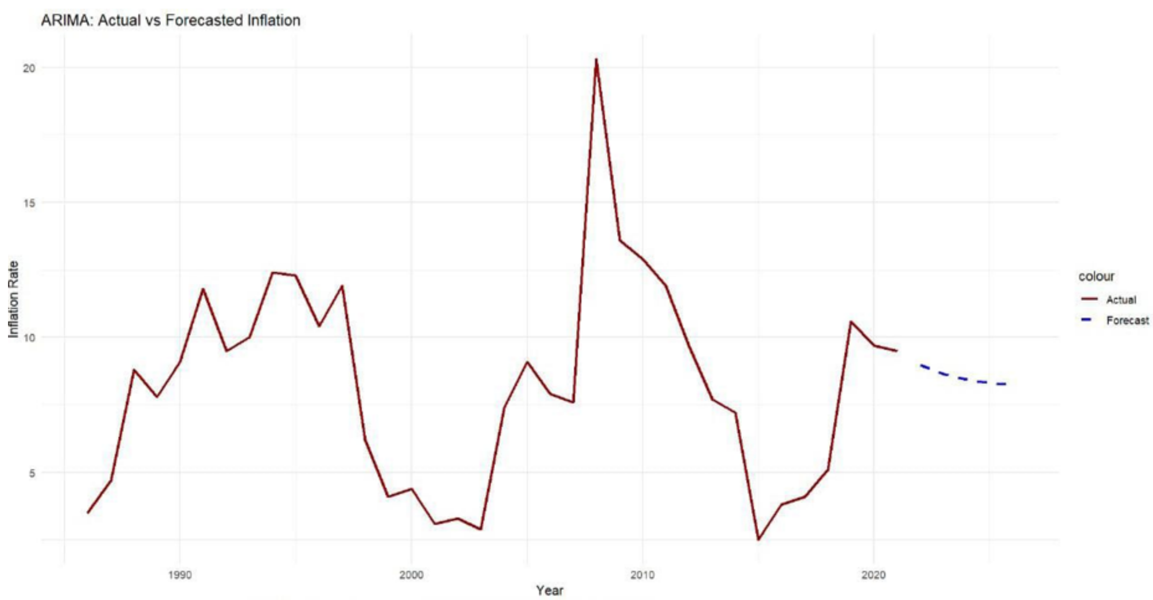
- **R-squared (R^2):**

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

In this case, the R^2 value was calculated to be **0.3941**, meaning that about 39.41% of the variance in inflation is explained by the ARIMA model. While this is not a very high value, it suggests that the model captures a reasonable portion of the inflation trend, though there may be other factors at play (such as external economic variables) that the model is not accounting for.

Forecasting the Inflation

The forecasted inflation values for the next 5 years were visualized using the `plot()` function, showing the actual inflation data alongside the forecasted values. However, an error occurred during plotting stating that the `figure margins were too large`. This usually happens when the plotting area in R is too small to accommodate the graph, which can be resolved by adjusting the plotting parameters or increasing the plotting window size.



5.1.1 ModelOverview

- The actual inflation data is shown as a line or points.
- The forecasted inflation is displayed as a continuation of the series.
- Confidence intervals are included to indicate the uncertainty in the predictions.

5.1.2Conclusion

The ARIMA model provided a structured approach to forecasting inflation in Pakistan using historical data. With a R^2 value of 0.3941, the model explained approximately 39.41% of the variability in inflation, indicating a moderate fit.

While this is not exceptionally high, the result suggests that the ARIMA model effectively captured some underlying trends in the data. However, its limitations in fully explaining inflation variability highlight the possible influence of external economic factors not included in this univariate time series model.
Despite the plotting issue encountered, the model's forecast offers a valuable baseline for short- to medium-term inflation projections. Future enhancements could include multivariate models or integration with exogenous variables to improve accuracy.

5.2 Model-Fitting

We standardized the predictor variables and performed 10-fold cross-validation using the `cv.glmnet()` function to determine the optimal penalty parameter (λ). The best λ value was then used to fit the Ridge regression model using the `glmnet()` function.

5.3 Model-Coefficients

The coefficients from the Ridge regression model are as follows:

- Intercept: 8.244
- ExchangeRate: -0.049
- GDPGrowth: -0.588
- Unemployment: -0.390
- BroadMoney: -0.081
- Exports: 0.187
- Imports: 0.194
- OilRents: 0.563
- Remittances: -0.140

These values indicate the impact of each predictor on inflation.

5.4 Model-Evaluation-R^2

The R^2 value was calculated as 0.1976, meaning the model explains only 19.76% of the variance in inflation. This suggests the model captures some trends but misses much of the variability.

5.5 Ridge-Regression-for-Inflation-Forecasting

Ridge regression is a regularized linear regression technique that adds a penalty to the model's coefficients to prevent overfitting, especially when predictors are highly correlated. In this analysis, we used Ridge regression to forecast inflation in Pakistan, utilizing predictors such as exchange rates, GDP growth, unemployment, and more.

5.5.1 Model-Fitting

We standardized the predictor variables and performed 10-fold cross-validation using the `cv.glmnet()` function to determine the optimal penalty parameter (λ). The best λ value was then used to fit the Ridge regression model using the `glmnet()` function.

5.5.2 Model-Coefficients

The coefficients from the Ridge regression model are as follows:

- Intercept: 8.244
- ExchangeRate: -0.049
- GDPGrowth: -0.588
- Unemployment: -0.390
- BroadMoney: -0.081
- Exports: 0.187
- Imports: 0.194
- OilRents: 0.563
- Remittances: -0.140

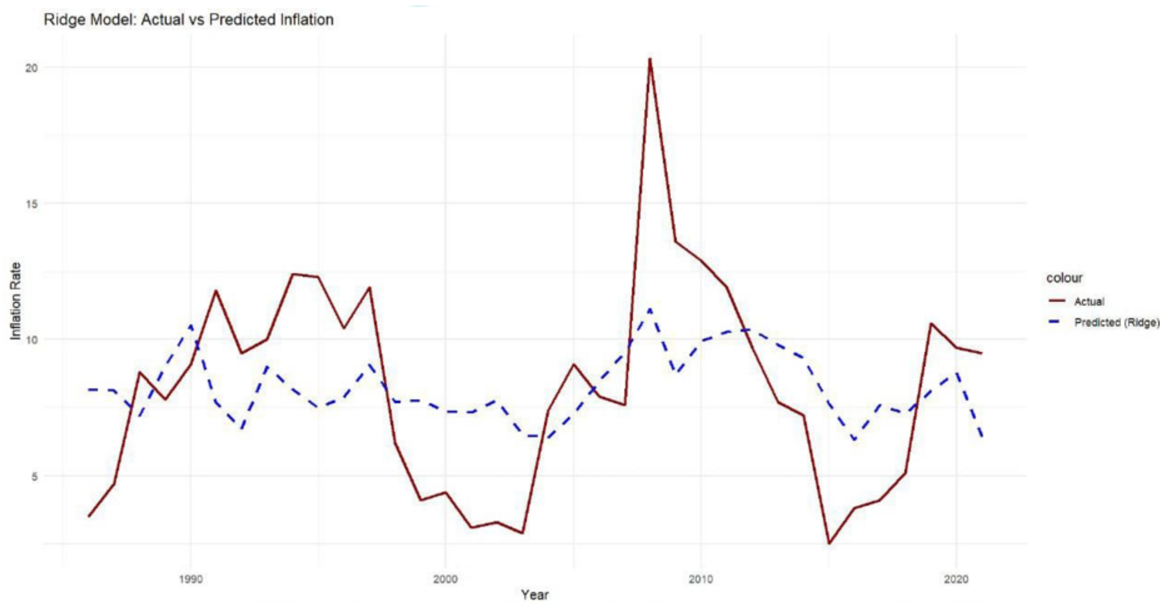
These values indicate the impact of each predictor on inflation.

5.5.3 Model-Evaluation-–R^2

The R^2 value was calculated as 0.1976, meaning the model explains only 19.76% of the variance in inflation. This suggests the model captures some trends but misses much of the variability.

5.6 Cross-Validation-Plot

The plot of cross-validation error versus λ could not be displayed due to a `figure margins too large` error. Once resolved, this plot would show the impact of different λ values on model performance.



5.7 Conclusion

Ridge regression was used to model inflation using several key economic indicators. Despite the low R^2 value of 0.1976, which indicates that the model explains a small portion of the variance in inflation, the Ridge regression model can still provide useful insights into the relationships between inflation and the selected predictors. However, given the low explanatory power, further refinement of the model (by including additional predictors or considering more advanced models like LASSO or Elastic Net) might be necessary for better forecasting.

5.8 Lasso-Regression-for-Inflation-Forecasting

5.9 Overview

Lasso (Least Absolute Shrinkage and Selection Operator) regression is a regularized linear model that performs both variable selection and regularization. By adding an $L1$ penalty term, Lasso encourages sparsity in the coefficients, effectively driving some coefficients to zero. This feature selection property makes Lasso useful when we suspect that only a subset of predictors significantly affects the target variable—in this case, inflation.

5.10 Model-Fitting

We fitted the Lasso regression model using the `glmnet()` function with $\alpha = 1$, indicating that we are using Lasso (as opposed to Ridge regression, which would use $\alpha = 0$). The model was trained on the scaled predictor variables (`x_scaled`) and the target variable (`y`, which represents inflation).

Next, we performed cross-validation with the `cv.glmnet()` function to determine the optimal penalty parameter (λ). The best λ value is selected based on the minimum mean cross-validation error.

5.11 Model-Coefficients

The Lasso regression model's coefficients are as follows:

- Intercept: 8.244
- GDPGrowth: -0.875
- Unemployment: -0.384
- OilRents: 0.876

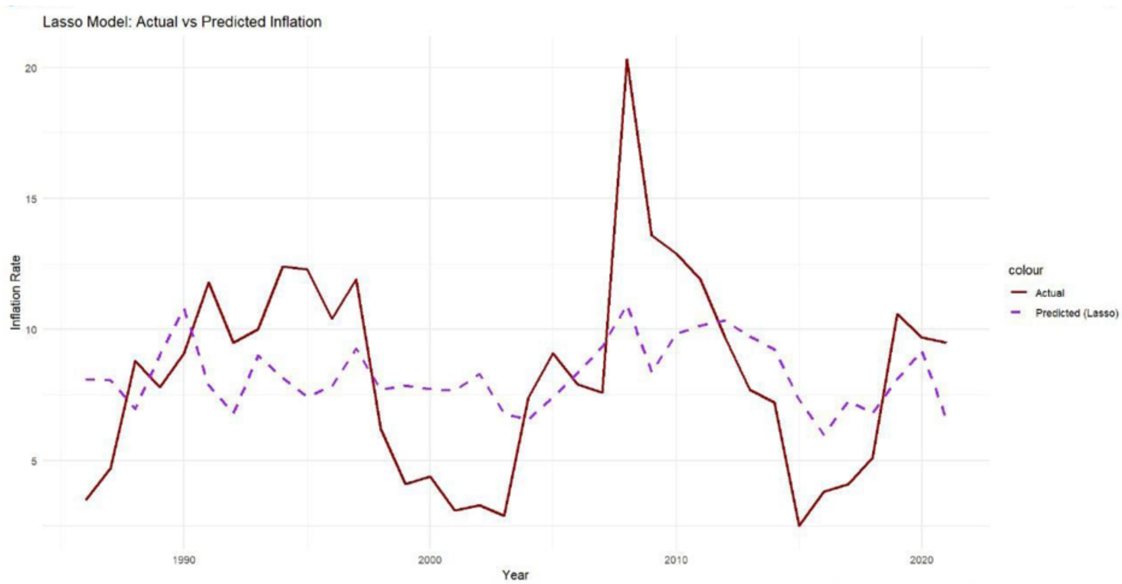
The coefficients for several predictors (ExchangeRate, BroadMoney, Exports, Imports, and Remittances) are driven to zero, indicating that Lasso selected only the most important variables for predicting inflation. For example, GDPGrowth and OilRents have significant coefficients, while other predictors do not contribute to the model.

5.12 Model-Evaluation-R^2

The R^2 value for the Lasso regression model is 0.2248, meaning the model explains 22.48% of the variance in inflation. Although this is slightly better than the Ridge regression model's R^2 (0.1976), it still indicates that a substantial portion of the variability in inflation is not explained by the predictors included in the model.

5.13 Cross-Validation-Plot

The cross-validation plot, generated by `plot(cv_lasso)`, was not displayed due to the "figure margins too large" error. This can be fixed by adjusting the plotting parameters. The plot would show how the mean cross-validation error changes with different λ values, helping us choose the optimal regularization strength.



5.14 Conclusion

Lasso regression provided a more sparse model by selecting a subset of predictors that significantly impact inflation. With an R^2 value of 0.2248, the model explains a modest amount of the variation in inflation. While Lasso is better than Ridge in terms of variable selection, further refinement or more complex models might be necessary to improve predictive accuracy.

6. Elastic-Net-Regression-for-Inflation-Forecasting

6.1 Overview

Elastic Net is a regularized regression technique that combines the strengths of both Lasso and Ridge regression. The key feature of Elastic Net is the ability to adjust the mixing parameter, α, to balance between Lasso ($L1$ penalty) and Ridge ($L2$ penalty). When $\alpha = 0.5$, Elastic Net is a mix of both Lasso and Ridge, leveraging the benefits of both techniques: the sparsity of Lasso and the regularization of Ridge.

6.2 Model-Fitting

We fitted the Elastic Net regression model using the `cv.glmnet()` function with $\alpha = 0.5$. This setting ensures an equal mix of Lasso and Ridge regularization. We then performed cross-validation to determine the best λ (the penalty parameter), by minimizing the cross-validation error.

After selecting the optimal λ, the final model was fitted using `glmnet()` with the best λ value.

6.3 Model-Coefficients

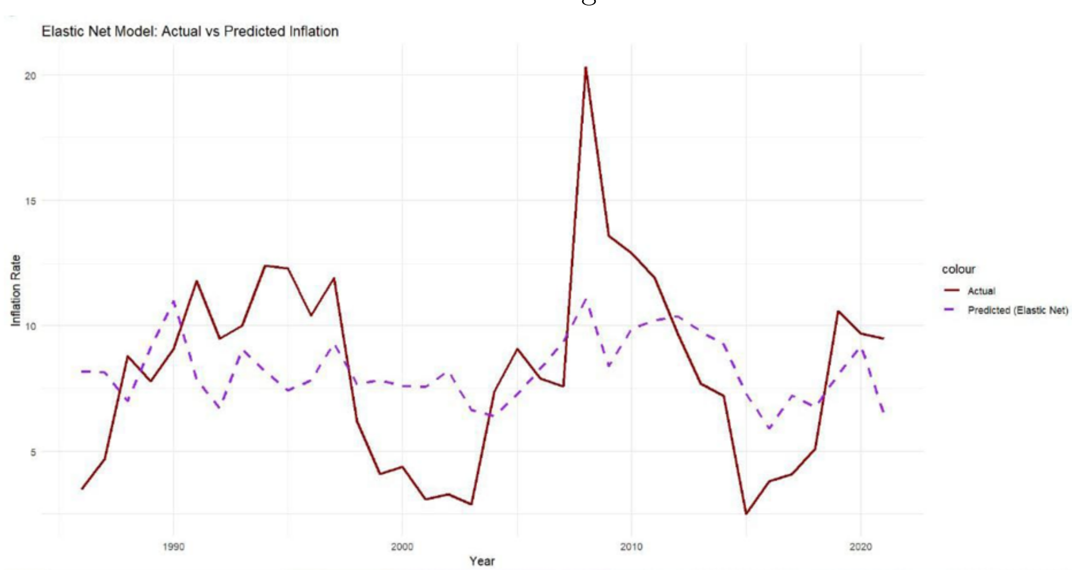
The coefficients from the Elastic Net regression model are:

- Intercept: 8.244
- GDPGrowth: -0.911
- Unemployment: -0.451
- OilRents: 0.901

Similar to the Lasso model, several predictors (ExchangeRate, BroadMoney, Exports, Imports, and Remittances) have been shrunk to zero, indicating that only a subset of predictors are important for forecasting inflation. For instance, GDPGrowth and OilRents have non-zero coefficients, indicating their significant relationship with inflation.

6.4 Model-Evaluation - R^2

The R^2 value for the Elastic Net regression model is 0.2317, meaning the model explains 23.17% of the variance in inflation. This is slightly better than both the Ridge (0.1976) and Lasso (0.2248) models, suggesting that Elastic Net offers a better balance between variable selection and regularization.



7. Conclusion

Elastic Net regression offers a promising approach for inflation forecasting by balancing the strengths of both Lasso and Ridge. The model explains 23.17% of the variance in inflation, which is an improvement over the Ridge and Lasso models. However, the relatively low R^2 indicates that

8. Results-and-Conclusion

1.Visual-Comparison-of-Model-Performance

The combined plots display actual inflation values (in dark red) against model predictions or forecasts (in purple or blue dashed lines) over time.

- **ARIMA** shows a reasonable ability to follow the inflation trend up to the end of historical data, then produces forecasts for the next five years. However, the forecast curve is relatively flat and conservative, reflecting ARIMA's nature of focusing on past trends without external predictors.
  - $R^2 = 0.394$, which is moderate, meaning the model explains about 39.4% of the variation.
- **Ridge Regression** tends to over-smooth the inflation curve. It follows the general trend but misses sharp fluctuations. This behavior is common in Ridge due to its L2 penalty, which shrinks coefficients but doesn't perform variable selection.
  - $R^2 = 0.198$, indicating poor fit.
- **Lasso Regression**, while slightly better than Ridge, still shows underfitting tendencies. However, it appears to perform better at certain turning points due to its feature selection capability via the L1 penalty.
  - $R^2 = 0.225$, still modest.
- **Elastic Net Regression** combines strengths of both Ridge and Lasso. It appears to capture trend direction slightly better than the others, but still not ideal.
  - $R^2 = 0.232$, the best among regularized regression models, but still under 25%.

2.Model-Equations

Below are the equations for each model used in the analysis.

ARIMA-Model

The ARIMA model is defined as:

$$\text{Inflation}_t = 8.0965 + 0.6250 \cdot \text{Inflation}_{t-1} + \epsilon_t$$

where: - Inflation_t is the inflation rate at time t, - Inflation_{t-1} is the inflation rate at the previous time point, - ϵ_t is the white noise error term at time t.

- **Inflation at time** t is predicted based on the inflation at the previous time point $t-1$, with a coefficient of 0.6250.
- The model also includes a constant term of 8.0965.

- ϵ_t represents random errors or noise that cannot be explained by past values of inflation.

In simple terms, this equation says that inflation at the current time depends on inflation in the previous period, with some randomness added in.

Ridge-Regression-Model

The Ridge regression model is given by:

$$\text{Inflation}_t = 8.2444 - 0.0487 \cdot \text{ExchangeRate}_t - 0.5882 \cdot \text{GDPGrowth}_t - 0.3904 \cdot \text{Unemployment}_t - 0.0806 \cdot \text{BroadM}$$

where: - Each variable represents the respective economic indicator at time t, - ϵ_t is the error term.

- **Inflation at time** t is influenced by multiple factors:
  - **ExchangeRate**: A negative relationship, meaning that as the exchange rate increases, inflation tends to decrease.
  - **GDPGrowth**: A negative relationship, implying that higher GDP growth is associated with lower inflation.
  - **Unemployment**: Also negatively related to inflation.
  - **BroadMoney**: A small negative effect on inflation.
  - **Exports and Imports**: Both have a positive effect on inflation, meaning as they increase, inflation also tends to rise.
  - **OilRents**: Positive effect, meaning higher oil rents (revenues from oil) are associated with higher inflation.
  - **Remittances**: A negative relationship, suggesting that more remittances reduce inflation.
- The constant term is 8.2444, and ϵ_t is the random error.

In simple terms, this model tells us that inflation is influenced by various economic factors such as exchange rates, GDP growth, unemployment, money supply, trade balances, oil revenues, and remittances.

Lasso-Regression-Model

The Lasso regression model equation is:

$$\text{Inflation}_t = 8.2444 - 0.8745 \cdot \text{GDPGrowth}_t - 0.3836 \cdot \text{Unemployment}_t + 0.8763 \cdot \text{OilRents}_t + \epsilon_t$$

where: - The non-zero coefficients are for GDP Growth, Unemployment, and Oil Rents.

- **Inflation at time** t depends on:
  - **GDPGrowth**: A negative relationship, meaning higher GDP growth is associated with lower inflation.

  - **Unemployment**: Negative relationship, higher unemployment tends to reduce inflation.
  - **OilRents**: A positive relationship, suggesting that higher oil revenues lead to higher inflation.

- The constant term remains 8.2444, and ϵ_t is the random error.

In simple terms, this equation highlights that inflation is influenced by GDP growth, unemployment, and oil revenues, with GDP growth and unemployment reducing inflation, while oil revenues increase inflation.

Elastic-Net-Model

The Elastic Net regression model equation is:

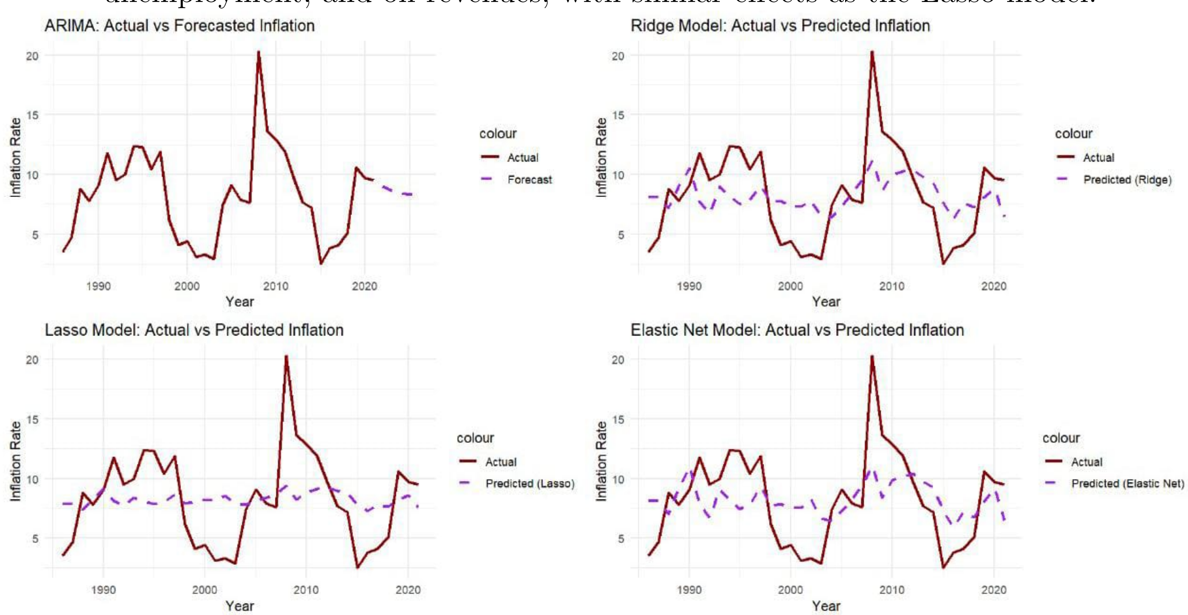
$$\text{Inflation}_t = 8.2444 - 0.9110 \cdot \text{GDPGrowth}_t - 0.4510 \cdot \text{Unemployment}_t + 0.9012 \cdot \text{OilRents}_t + \epsilon_t$$

where: - The significant variables are GDP Growth, Unemployment, and Oil Rents.

- **Inflation at time** t is explained by:
  - **GDPGrowth**: A negative relationship, with higher GDP growth leading to lower inflation.
  - **Unemployment**: A negative relationship, implying higher unemployment results in lower inflation.
  - **OilRents**: A positive relationship, where more oil revenues contribute to higher inflation.

- The constant term is 8.2444, and ϵ_t represents random errors.

In simple terms, this model also indicates that inflation is affected by GDP growth, unemployment, and oil revenues, with similar effects as the Lasso model.



2. Model Comparison Table (Summary)

Model	R^2 Value	MSE	Strengths	Weaknesses
ARIMA	0.394	15.06795	Forecasting capability, trend tracking	Lacks external predictors, flat forecasts
Ridge	0.198	14.09511	Regularization helps avoid overfitting	Underfits, misses fluctuations
Lasso	0.225	14.62025	Feature selection, slightly better fit	Still underfits, not very responsive
Elastic Net	0.232	14.54714	Balanced penalty, best among regressions	Still low R^2, slight trend deviation

Table 1: Summary of Model Performance for Inflation Forecasting

3.Conclusion

Among all models tested:

- **ARIMA** provided the best explanatory power (highest R^2) and is better suited for time series forecasting.
- **Elastic Net** was the best-performing among machine learning regression models, but all linear models had limited predictive power for inflation.
- The relatively low R^2 values across all models suggest that inflation in Pakistan is influenced by multiple complex factors not captured in the current model setup (e.g., global oil prices, monetary policy, geopolitical events).
- Future improvements could involve:
  - Adding exogenous variables (ARIMAX, VAR, etc.)
  - Trying non-linear models (Random Forest, XGBoost, Neural Networks)
  - Including macroeconomic indicators

9. References

References

[1] State Bank of Pakistan, *Evaluating the Performance of Inflation Forecasting Models of Pakistan*, 2015, `https://www.sbp.org.pk/research/bulletin/2015/Vol-11-1/Evaluating%20the%20Performance%20of%20Inflation%20Forecasting%20Models%20of%20Pakistan.pdf`

[2] Haider, Z. and Hanif, M. *Inflation Forecasting in Pakistan*, https://pu.edu.pk/images/journal/pesr/currentissues/7%20HAIDER%20and%20HANIF%20Inflation%20Forecasting%20in%20Pakistan.pdf

[3] Akhtar, M. *Inflation in Pakistan*, 2009, https://www.researchgate.net/publication/24046506_Inflation_in_Pakistan

[4] Pakistan Institute of Development Economics (PIDE), *Inflation and Monetary Policy in Pakistan*, 2015, https://file.pide.org.pk/pdf/PDR/206/Volume3/341-368.pdf

[5] Econstor, *Monetary Policy and Inflation in Pakistan*, 2010, https://www.econstor.eu/bitstream/10419/43393/1/550986480.pdf

[6] State Bank of Pakistan, *Inflation and Monetary Policy in Pakistan*, 2017, https://www.sbp.org.pk/research/bulletin/2017/Vol-13.1/Paper-01-2017.pdf

[7] Institute of Policy Research, *Inflation in Pakistan: Monetary or Fiscal Phenomena*, 2024, https://ipripak.org/wp-content/uploads/2024/08/inflation-in-Pakistan-monetary-or-fiscal-phenonmena-Final-File-21-08-2024.pdf

[8] State Counsellor, *Inflation and Monetary Policy*, 2021, https://statecounsellor.wordpress.com/wp-content/uploads/2021/06/210204.pdf

[9] Asian Journal of Business and Management Review, *Inflation and its Impact in Pakistan*, 2014, https://www.ajbmr.com/articlepdf/ajbmrv01n0509.pdf