Global Journal of Mathematics and Econometrics

DOI: 10.61424/gjme
Journal Homepage: www.bluemarkpublishers.com/index.php/GJME

GJME
BLUEMARK PUBLISHERS

| RESEARCH ARTICLE

Dynamic Ridge Regression vs. Lasso Regression: A Comparative Study for Modeling Pakistan's Unemployment Rate

Savera Mubasher[1], Muhammad Zakria[2], Amir Shahzad[3], Nazakat Ali[4] ✉ and Hiba faisal[5]

[1]*Master of Philosophy in Statistics, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan, saveramubasher2320@gmail.com*
[2]*Master of Philosophy in Statistics, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan, zakriaawan663@gmail.com*
[3]*Master of Philosophy in Statistics, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan, amir.statistics07@gmail.com*
[4]*Master of Philosophy in Statistics, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan, nazakatalibandesha@gmail.com*
[5]*Master of Philosophy in Statistics, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan, hibafaisal016@gmail.com*

Corresponding Author: Nazakat Ali, **E-mail**: nazakatalibandesha@gmail.com

| ABSTRACT

The unemployment rate is a key economic indicator that reflects a country's economic health, influencing policy decisions and citizens' living standards. This study examines Pakistan's economic indicators, using the unemployment rate as the dependent variable, while GDP, exchange rate (ER), inflation rate (INF), foreign direct investment (FDI), exports of goods and services (EGS), general government final consumption expenditure (GFCE), budget deficits (BDF), and population (POP) serve as independent variables. A Variance Inflation Factor (VIF) analysis identifies multicollinearity among predictors, revealing ER as having the highest VIF of 7.544, indicating strong multicollinearity. Other variables like GDP, FDI, GFCE, BDF, and INF exhibit low VIFs, while EGS and POP have moderate levels of multicollinearity. The study employs Ridge and Lasso regression with 2-fold cross-validation to determine significant predictors and assess their impact on unemployment rates. The optimal lambda for Ridge regression is found to be 0.7758532, selected through cross-validation to minimize error. ER emerges as the most influential variable, with a feature importance score of 100. Lasso regression, with an optimal lambda of 0.1943467, eliminates GDP, EGS, POP, and INF, enhancing model simplicity and reducing overfitting. The Ridge model yields an RMSE of 0.32, while Lasso achieves a lower RMSE of 0.25, indicating better predictive accuracy. The study underscores the importance of addressing multicollinearity and demonstrates the effectiveness of Ridge and Lasso regression in predicting unemployment rates, with each model offering unique strengths for economic analysis.

| KEYWORDS

Ridge Regression, Lasso Regression, Cross-Validation, Unemployment Rate Pakistan.

| ARTICLE INFORMATION

ACCEPTED: 01 October 2024 **PUBLISHED:** 06 November 2024 **DOI:** 10.61424/gjme.v1.i1.125

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Bluemark Publishers.

1. Introduction

1.1 Ridge Regression

Ridge Regression, developed by Art Hoerl and Bob Kennard in 1970, addresses multicollinearity by shrinking the coefficient vector towards the origin, enhancing traditional least squares methods. Celebrating its 50th anniversary in 2020, it remains a key statistical tool alongside Lasso and Elastic Net. Hoerl's engineering and Kennard's statistical expertise shaped this influential method's practical application (Hoerl, 2020).

Ridge regression addresses multicollinearity in linear regression by adding a penalty term to the loss function, shrinking coefficients toward zero without eliminating them. This regularization reduces overfitting and improves generalization by favoring simpler models. It is used in fields like bioinformatics, such as smoothing gene expression data in circadian rhythm analysis, making it valuable for various research applications (Wieringen, 2023).

This study explores Ridge Regression's dual version for non-linear regression in high-dimensional spaces, addressing the 'curse of dimensionality' using kernel functions. It evaluates ANOVA decomposition kernels against polynomial and spline kernels on the Boston Housing dataset. The research aims to minimize prediction errors using average square loss, demonstrating the dual algorithm's effectiveness in non-linear regression compared to other methods (CANTONE, 1954).

Ridge regression is a powerful tool in genetic studies, addressing multicollinearity and outliers. Generalized cross-validation (GCV) aids in selecting optimal parameters enhancing prediction accuracy in complex, high-dimensional data. (Arashi *et al.*, 2021). This research analyzes macroeconomic factors affecting Iraq's unemployment rate using ridge regression to address multicollinearity, comparing results with Ordinary Least Squares to identify key economic determinants (Bager *et al.*, 2017).

This research proposes a biased estimator for addressing multicollinearity in linear regression, combining ridge and Liu estimators, demonstrating improved mean squared error across various regression models through simulation studies (Kibria, 2020).

1.1.1 Adaptive Robust Variable Selection

The paper explores penalized quantile regression, focusing on the weighted robust Lasso (WR-Lasso) technique to address challenges from high-dimensional datasets. It examines the model selection oracle property and establishes WR-Lasso's asymptotic normality, proposing the adaptive robust Lasso (AR-Lasso) for improved performance through tailored weight vectors (Robust and Selection, 2015).

The Almost Unbiased Modified Ridge-Type Estimator (AUMRTE) addresses multicollinearity issues, combining characteristics of MRTE and AUE to improve mean square error (MSE) performance, validated through simulations and real tourism data (Mahmoud, 2022).

1.1.2 Ridge Regression Model Complexity:

Ridge regression is a technique used in statistics to mitigate the issue of multicollinearity in regression analysis. Multicollinearity occurs when independent variables in a regression model are highly correlated, leading to unstable estimates of the regression coefficients.

In ridge regression, a penalty term is added to the ordinary least squares (OLS) method to shrink the coefficients towards zero. This penalty term is controlled by a hyperparameter called the regularization parameter (λ).

Ridge regression mitigates multicollinearity by introducing a penalty term, reducing estimate variance, and enhancing model robustness, especially when predictors outnumber observations. Regularized methods like OLS, LASSO, and Principal Component Regression improve coefficient accuracy. Evaluating these techniques through metrics like AMSE and AIC is vital for robust regression results (Herawati *et al.*, 2018).

1.1.3 Model Complexity Analysis (MCA):
Model complexity analysis (MCA) is a method proposed to address the challenge of selecting the optimal model complexity in regression analysis. In the context of NIR spectroscopic model updating, MCA aims to determine the appropriate level of complexity for the updated model by analyzing the 2-norm of the regression coefficients vector.

The 2-norm of the regression coefficients vector represents the magnitude of the coefficients in the model. MCA suggests that this value should be smaller in the updated model compared to the original model. By focusing on the complexity of the model in terms of the regression coefficients, MCA provides a quantitative approach to model selection that complements traditional methods like the bias/variance tradeoff (L curve). Through MCA, researchers can make informed decisions about the level of complexity that strikes a balance between model performance and interpretability, ultimately improving the reliability of NIR spectroscopic predictions.

1.1.4 Lasso regression:
Regression models often face overfitting and optimism bias, especially with many candidate variables. LASSO (Least Absolute Shrinkage and Selection Operator) addresses these issues by minimizing prediction error and simplifying models through variable selection. While effective, it requires careful external validation and may compromise the interpretability of individual coefficients (Ranstam and Cook, 2018).

The Bayesian adaptive lasso regression improves classical adaptive lasso by using a Gibbs sampler for tractable posteriors, addressing high-dimensional challenges without requiring OLS estimates. It reformulates the model using Laplace density, incorporates noninformative priors, and establishes a joint posterior distribution, demonstrating superior performance in simulations and real data analyses (Alhamzawi and Ali, 2018).

High-dimensional linear regression models like LASSO are essential in fields with more covariates than samples, such as microarray studies. LASSO employs an L1-penalty for variable selection, offering computational efficiency and stability. Its ability to accurately identify relevant variables while controlling sparsity and bias ensures reliable model selection in these scenarios. (Annals, 2008).

1.1.5 Sparsity and the Lasso: -
The Lasso (Least Absolute Shrinkage and Selection Operator) is a popular method in statistical machine learning that involves regularization to promote sparsity in the model. Sparsity refers to having only a few non-zero coefficients in the model, which aids in feature selection and model interpretability. In the context of the Lasso, the objective function involves minimizing the residual sum of squares subject to the constraint that the sum of the absolute values of the coefficients is less than a fixed value. This constraint encourages the coefficients of less important features to be exactly zero, effectively performing feature selection. The Lasso path algorithm iteratively updates the coefficients by shrinking them towards zero, ultimately leading to a sparse solution. This algorithm is computationally efficient and widely used in various fields such as genetics, finance, and signal processing. Understanding sparsity and the Lasso is crucial for building predictive models with high-dimensional data while maintaining model simplicity and interpretability (Tibshirani, 2015).

1.2. Regularized Terms in Ridge and Lasso Regression
1.2.1 Ridge Regression:
In ridge regression, the regularization term includes the L2-norm penalty on the coefficients, which helps prevent overfitting by shrinking the coefficients to zero. The ridge regression estimator minimizes the residual sum of squares with an additional penalty term, that is, the sum of squared coefficients multiplied by a tuning parameter, L2. The L2 penalty term in ridge regression adds a constraint to the optimization problem, penalizing large coefficients and reducing model complexity. Ridge regression is suitable for handling multicollinearity in datasets, as it can stabilize the model by reducing the impact of correlated predictors on the coefficient estimates.

1.2.2 Lasso Regression:

In lasso regression, the regularization term includes the L1-norm penalty on the coefficients, promoting sparsity in the solution by shrinking some coefficients to exactly zero. The lasso penalty term is the sum of the absolute values of the coefficients multiplied by a tuning parameter, denoted as b. Lasso regression is effective for feature selection, as it encourages a sparse model by selecting only the most relevant predictors while setting others to zero. Lasso regression is sensitive to high correlations among predictors, as it may arbitrarily choose one predictor over others, leading to potential model instability(Ogutu *et al.*, 2012).

1.2.3 Regression Techniques in Handling Multicollinearity:

Regression analysis examines relationships between dependent and independent variables, aiding predictions and decision-making. Multicollinearity, where independent variables are correlated, complicates model interpretation. Regularization techniques like Ridge, LASSO, and Bridge regression mitigate this issue by shrinking coefficients, enhancing prediction accuracy. Each technique has unique strengths and weaknesses for various research objectives. (Enwere *et al.*, 2023).

1.2.4 Model selection method:

Ridge regression, a regularization technique addressing multicollinearity and overfitting in linear models, relies on model selection methods to determine optimal regularization levels. Cross-validation, which assesses model performance across various subsets using metrics like mean squared error (MSE), helps identify the best regularization parameter (alpha). Information criteria, such as AIC and BIC, also guide alpha selection by balancing model complexity and fit. Kernel ridge regression enhances this by utilizing nonlinear transformations, with popular kernels like Gaussian and Sinc offering improved smoothing and predictive (Exterkate, 2013).

1.2.5 Introduction to Economic Growth and Unemployment Rate:

Forecasting the U.S. unemployment rate is essential for assessing economic health. Various linear and nonlinear time series models, including hidden Markov and threshold autoregressive models, improve accuracy, especially during economic fluctuations. This research compares these models against consensus forecasts from the Survey of Professional Forecasters, demonstrating nonlinear models' superiority in specific scenarios. Additionally, incorporating initial jobless claims as leading indicators offers insights into more reliable forecasting methods, crucial as the unemployment rate fluctuates amidst concerns about economic growth and recession. (Levine, 2012).
Gross Domestic Product (GDP) is a key economic indicator representing the monetary value of all final goods and services produced within a country over a specific period. While often linked to social welfare and living standards, its limitations as a sole measure of societal progress have spurred debates on alternative indicators for well-being. Despite criticisms, GDP remains central to economic decision-making and policy formulation.

Inflation is another critical macroeconomic concept, with optimal rates between 0.7% and 1.4% deemed essential for stability. Research shows a significant connection between inflation, budget deficits, and economic growth, as illustrated in studies on various countries. Furthermore, foreign direct investment (FDI) is shown to foster economic growth through capital inflow, emphasizing its importance in development strategies. The interplay between these economic factors highlights the need for nuanced approaches to policy-making that encompass a broader understanding of growth and welfare beyond GDP alone.

1.2.6 Multicollinearity and VIF

Multicollinearity, a prevalent issue in regression analyses, arises from high correlations among predictor variables, leading to unreliable coefficients. The research introduces the Variance Inflation Factor (VIF) as a diagnostic tool to detect multicollinearity, emphasizing its importance alongside eigenvalues for ensuring accurate and valid regression model interpretations paper (Thompson *et al.* 2017).

1.3. Method to Detect Multicollinearity:

1.3.1 Variance Inflation Factor (VIF) Method:

VIF is a commonly used method to detect multicollinearity by measuring how much the variances of the estimated regression coefficients are inflated when the independent variables are linearly related. It calculates the ratio of the variance of a coefficient estimate when multicollinearity is present to the variance of a coefficient estimate when predictors are not correlated. High VIF values (usually above 10) indicate a problematic level of multicollinearity, suggesting that the variable should be further investigated or possibly removed from the model.

1.3.2 Correlation Matrix:

Constructing a correlation matrix to identify high correlations between predictor variables can also help detect multicollinearity. Correlation values close to 1 or -1 indicate strong linear relationships between variables, signaling potential multicollinearity issues.

1.3.3 Eigenvalues and Condition Indices:

Calculating eigenvalues and condition indices from the correlation matrix can provide insights into the severity of multicollinearity. Large condition indices (typically above 30) or small eigenvalues close to zero suggest the presence of multicollinearity.

1.3.4 Principal Component Analysis (PCA):

PCA can be used to transform the original variables into uncorrelated principal components, helping to identify multicollinearity by examining the amount of variance explained by each component.

1.3.5 Cross-Validation Techniques:

Utilizing cross-validation methods can also aid in detecting multicollinearity by assessing the stability and generalizability of the regression model when dealing with correlated predictors(Adnan *et al.*, 2006).

1.3.6 Objectives

The objectives of the following studies will be:

- To estimate the unemployment rate of Pakistan for using Ridge Regression
- To estimate the Unemployment rate of Pakistan for using the Lasso Regression
- Comparing the estimate of the Ridge Regression and Lasso Regression using appropriate model adequacy measure such as MSE and AIC value.

2. Review of Literature

CANTONE (1954) This paper investigated the dual ridge regression method using kernel functions for nonlinear regression on high-dimensional spaces and related it to auxiliary vector machines and kriging theory through optimization problems and Bayesian frameworks. Zadeh (1958) Policymakers target lower positive inflation rates to eliminate the risks of inflation and economic volatility. An estimate of long-term optimal inflation is important for effective monetary policy and confidence. Williamson (1997) Bela Balassa emphasizes optimal export taxes and devaluation to support nontraditional exports, challenging IMF views. Studies by Bartolini, Drazen, Fischer, Reisen, Lal, Bery, and Pant explore exchange rates and capital flows' interconnectedness. Firinguetti (1999) stated novel approach for estimating the shrinkage parameters was used to introduce a new operational Generalized Ridge Regression (GRR) estimator. One benefit of the one suggested there over the standard operational GRR estimator was that its shrinkage parameters were bounded. The resulting GRR estimator's finite sample qualities were determined within the framework of a classical linear regression model with disturbances that were normally distributed.

Kibria (2003) The estimation of the ridge parameter k was crucial in the RR study. Novel estimators based on generalized RR methodology outperformed traditional methods, confirmed by simulation and a numerical case examination. Kitov (2003) Ivan Kitov's study examines the relationship between real GDP growth and population growth, emphasizing the role of the duration of income growth and the specific age of the population in predicting GDP growth in developed countries, especially in the US, UK, and France. Hastie and Tibshirani (2004) introduced efficient quadratic regularization techniques for expression arrays, demonstrating the effectiveness of ridge regression in high-dimensional data settings. Their work contributes to the optimization of ridge regression algorithms for large-scale genomic data analysis. Overall, the literature on ridge regression provides a comprehensive overview of its theoretical properties, practical implementations, and empirical validations across diverse research domains. These works collectively contribute to advancing the understanding and utilization of ridge regression in modern statistical modeling and machine learning applications.

Lipovetsky and Conklin (2005) stated that unemployment has become a serious concern in Iraq's business. The study utilized regression to uncover key macroeconomic factor that impact Iraq's unemployment rate. To choose the most suitable model for showing the phenomenon, the results were compared to those obtained using the ordinary least squares (OLS) technique. Unemployment in Iraq is determined by economic development, price rises, the number of foreign laborers, employee years of service, and population density.

Adnan *et al.* (2006) Multicollinearity in regression affects explanatory power and coefficient stability. Methods such as Ridge Regression, Principal Component Regression, and Partial Least Squares Regression address this issue. Research shows that Ridge Regression outperforms others in terms of MSE, emphasizing the importance of choosing a method based on data characteristics.

Annals. (2008) Previous studies have demonstrated the stability of LASSO in variable selection under complex conditions, focusing on sparsity and bias control. Its excellent convergence rates, interpolation properties, and ability to maintain statistical constraints confirm its reliability in high-dimensional linear regression scenario.

Ahmed Raza Cheema (2014) examined the primary causes for joblessness in Pakistan were identified using annual time series data from 1968 to 2008. The investigation employed an equation-based approach. The outcomes of this investigation show that a rise in production imbalance and ambiguity leads to an increase in unemployment. Unemployment reacts poorly to actual expenditure. There is a tradeoff between unemployment and inflation during the time frame after 1979, indicating that the Phillips curve holds but that rapid and steady rising inflation causes the national currency to depreciate and increases the fragility of the economy.

Afzal *et al.* (2015) The study evaluated unemployment forecasting tools using household data from Pakistan and Sri Lanka. It found that Lasso outperforms other algorithms, particularly in Pakistan. Incorporating satellite data improved accuracy, especially in Sri Lanka, highlighting the impact of variable selection methods. Manuscript (2015) Various diagnostic tools assess collinearity in multivariate models, including correlation matrices, coefficients of determination, variance inflation factor (VIF), and condition indices (CI) as the identification of collinearity is important for proper analysis because it can bias estimates and inflate standard errors. Simulation analysis helps to understand the effect of collinearity.

Emmanuel *et al.* (2016) concentrated on Various parameter estimation techniques that have been developed to address challenges in linear regression analysis. Proposed S Estimators for robust regression, particularly effective in the presence of outliers. Simpson and Montgomery (1996) developed a biased robust regression technique to handle both outliers and multicollinearity simultaneously. Highlighted the inadmissibility of the usual estimator for the mean of a multivariate normal distribution, emphasizing the need for robust estimation methods. The literature also includes the development of Ridge Estimators to combat multicollinearity issues in linear regression analysis. These diverse techniques collectively contribute to enhancing the robustness and accuracy of parameter estimation in regression models.

Existing literature on Middle Eastern unemployment emphasizes factors such as job demand and population growth. In particular, government operations in Iraq greatly affect business processes. This study examines the determinants of unemployment at the municipal level using multilinear regression and addresses multicollinearity with methods similar to ridge regression, which, in some cases, outperforms ordinary least squares (OLS).

Ranstam and Cook (2018) LASSO regression is popular to address the biases of overfitting and optimism of traditional models by reducing coefficients to zero, resulting in a more parsimonious model than datasets with multiple predictors such as genetic data, although it can reduce interpretation. Variations such as ridge regression and elastic nets offer alternative penalty methods.

Kim (2019) mentioned multicollinearity in regression analysis, its negative effects on statistical reliability, such as inflated variances and wide confidence intervals noted diagnostic tools such as variance inflation factors help to detect. Addressing multicollinearity is important to ensure the accuracy and reliability of research findings and regression model stability.

Ali *et al.* (2021) observed that the issue of multicollinearity among predictors was addressed by ridge regression, and there were numerous estimators for the ridge parameter k in the literature. However, the existing estimators also had significant mean square errors (MSE) if the predictors had a high level of collinearity. In that study, we examined a few novel and current estimators for ridge parameter k estimation. The suggested estimators' performance was assessed using a real-world case and extensive Monte Carlo simulations using the MSE criterion. The outcomes demonstrated how much better our suggested estimators were than the ones that were already in use.

Bashir *et al*. (2022) stated that Unemployment is a huge concern across the world. It is a widespread issue across many nations. The major goal of this study is to determine which factor is the primary cause of joblessness. The research depends on additional information gathered from trustworthy online sources. The information is collected on a yearly basis from 1991 to 2020. The primary social variables examined in this study are GDP, FDI, PGR, and inflation.

Zaikarina *et al.* (2023) Quantile regression, introduced by Koenker and Bassett in 1978, enhances traditional regression by modeling different quantiles of the response variable distribution, making it more efficient for estimating extreme rainfall Lasso and ridge the addition of regularization improves prediction performance, where ridges tend to account for stronger events It also outperforms Lasso in solving multicollinearity problems.

Damaševičius and Maskeli (2024) stated that the detection of financial hardship is an important topic in the scientific community because of its relevance to the community and the economy. Technological innovations, along with an increasing amount of recorded data, have resulted in the rise of financial distress that goes beyond accounting records and their indications (ratios). The area of features might be increased by including additional viewpoints on characteristic groups, including macroeconomics, sectors, social, board, management, judicial incidents, and so on. Yet, higher complexity leads to scarce data and overfitted models. This paper provides a novel strategy for assessing financial distress categorization that combines dimensionality reduction and machine learning approaches. The suggested approach seeks to discover a subset of characteristics that will minimize the loss functions.

3. Material And Methods

Data were obtained from the World Bank website. Secondary data have been used for this purpose. The study uses an experimental research design to predict Pakistan's short-term unemployment rates for the next five years. Unemployment creates many economic and social problems. Four important independent variables are selected as determinants: GDP Growth rate, Inflation rate, Labor force, and Population growth rate. The factors that make up the data for this decision range from 1980 to 2023. In order to meet the objective, the following statistical techniques will be used, such as Ridge regression and Lasso regression models.

3.1 Ridge Regression

Ridge regression is a technique that addresses multicollinearity in regression analysis by adding a small bias factor to the variables to reduce the standard errors of the coefficients. The correlation structures among the independent variables play a crucial role in determining the behavior of the ridge coefficients.

The ridge regression estimator is obtained by minimizing the residual sum of squares (RSS) subject to a quadratic constraint, which directly affects the ridge coefficients. This constraint involves balancing the fit to the data (RSS) with the penalty term, leading to stable and interpretable estimates for the regression coefficients.

$$Y = X\beta + \epsilon \qquad (1)$$

$\beta = (\beta_1, \ldots, \beta_p)'$ is a column vector of p regression coefficients and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is a vector of independent and identically distributed normally distributed random errors. Ridge estimates of the regression coefficients are given by

$$\hat{\beta} = (x'x)^{-1}x'y \qquad (2)$$

3.2 Notations and some preliminaries

The multiple linear regression model can be expressed as:

$$y = X\beta + e,$$

where y is a 1 × n vector of responses, X is an p n× observed matrix of the regressors, β is a 1 × p vector of unknown parameters, and e is a 1 × n vector of errors. The ordinary least square estimator (OLS) of the regression coefficientsβ is defined as

$$\hat{\beta} = (XX)^{-1}X'y,$$

Suppose there exists an orthogonal matrix D such that D'CD=Λ, where

Λ= diag $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ are the eigenvalues of the matrix C =X'X. The orthogonal $(canonical\ form)$ version of the multiple regression model is

$$Y = X^* \alpha + e$$

where X* = XD and α= D'β. In case the matrix X' X is ill-conditioned, however (in the sense that there is a near-linear dependency among the columns of the matrix), the OLS of β has a large variance, and multicollinearity is said to be present. Ridge regression replaces X'X with X'X +kI ($\langle k \rangle 0$) . Then the generalized ridge regression estimators of α are given as follows:

$$\hat{\alpha}(k) = (X'^*X^* + kI_p)^{-1}X'^*Y$$

Where k=diag $(k_1, k_2, \ldots k_p), k_i$>0 and $\hat{\alpha} = \Lambda^{-1}X'^*Y$ is the ordinary least squares (OLS) estimates of α.
According to Hoerl and Kennard (1970), the value of i k which minimizes the MSE ($\hat{\alpha}$ (k)) is

$$k_i = \frac{\sigma^2}{\sigma_i^2} \ ,$$

where σ^2 represents the error variance of the multiple regression model and σ_i is the i^{th}element of α.

3.3 Lasso Regression

Least Absolute Shrinkage and Selection Operator (Lasso) regression is a type of linear regression that incorporates regularization to prevent overfitting and perform variable selection. It adds a penalty term to the standard linear regression equation, forcing some coefficients to be exactly zero, effectively performing variable selection. The Lasso regression model can be defined as follows:

The model aims to minimize the sum of squared residuals plus the sum of the absolute values of the coefficients multiplied by a regularization parameter λ.

$$(\hat{\beta})^{(lasso)} = argmin \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{(j=1)}^{p} x_{(ij)} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j \quad (3)$$

Here, $(\hat{\beta})^{(lasso)}$ represents the Lasso estimate of the coefficients. The term $\left(\lambda \sum_{(j=1)}^{p} \beta_j\right)$ the penalty term that shrinks some coefficients to zero based on the value of the regularization parameter λ.

3.4 Mean Square Error

The MSE is a measurement of the mean squared deviation between the predicted values and the true value. It is utilized to assess how well a predictor or estimator is doing. Which is written as:

$$MSE = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}\right)^2}{n} \quad (4)$$

Where:
n is the sample size;
Y_i are the actual values;
$\hat{Y}$ is the predicted value.

3.5 Akaike Information Criterion

For a specific set of data, the AIC evaluates the relative effectiveness of statistical models. It calculates the amount of data wasted by a certain model depending on the greatest likelihood estimation of its parameters. Which is written as:

$$AIC = 2\phi - 2ln(L) \quad (5)$$

Where:
ϕ are the number of parameters;
L is the likelihood function.

3.6 Bayesian Information Criterion

The BIC is an additional tool for evaluating how well statistical models fit a particular set of data. It can be used to compression of the forecasting performance of a model between in-sample and out-of-sample. Which is written as:

$$BIC = \phi ln(n) - 2ln(L) \quad (6)$$

Where:
ϕ are the number of parameters;
L is the likelihood function;
n is the sample size.

3.7 Software

R is a statistical programming tool that's uniquely equipped to handle data and lots of it. Wrangling mass amounts of information and producing publication-ready graphics and visualizations is easy with R. So are all sorts of data analysis, mining, and modeling tasks.

4. Results and Discussion

4.1 Descriptive Statistics

Table 1 provides descriptive statistics for several variables. These descriptive statistics provide a comprehensive summary of the central tendency, dispersion, and shape of the distribution for each variable in the dataset. They are crucial for understanding the data's characteristics before conducting further statistical analysis.

Table 1: Descriptive Statistics

Vars	Mean	Sd	Median	Min	Max	Skew	Kurtosis	Se
ER	70.576	46.277	59.724	16.648	204.867	0.982	0.396	7.608
GDP	4.153	2.149	4.578	-2.970	7.706	-0.843	1.359	0.353
FDI	0.883	0.648	0.688	0.310	3.036	2.036	3.528	0.107
EGS	12.646	2.702	12.252	8.222	17.271	0.190	-1.161	0.444
GFCE	11.413	1.824	10.919	8.656	16.785	1.119	0.777	0.300
BDF	-2.489	2.469	-2.667	-7.742	3.936	0.416	0.425	0.406
POP	2.398	0.702	2.235	1.204	3.547	0.021	-1.256	0.115
UNEMP	1.612	1.799	0.585	0.400	6.340	1.433	0.657	0.296
INF	8.548	4.258	8.838	2.529	20.286	0.775	0.605	0.700

The dependent variable UNEMP has 1.612 and a standard deviation 1.799, with 0.585, indicating that half of the sample falls below this unemployment rate. This ranges from 0.400 to 6.340, with a right skew (1.433). a flat kurtosis (0.657) is observed. The dependent variables were ER (mean: 70.576, SD: 46.277), GDP (mean: 4.153, SD: 2.149), FDI (mean: 0.883, SD: 0.648), EGS (median).: 12.646), GFCE (average: 11.413), BDF (average: -2.489), POP (median: 2.398), INF (median: 8.548). Skewness and kurtosis reflect the normality of the distributions, while the standard error reflects the precision of the estimates.

4.2 Correlation Matrix

Table 2 shows the correlation matrix of the predictor variables, where each cell has a correlation coefficient ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). have a really perfect relationship. Notably, ER exhibits a strong negative correlation with EGS (-0.723) and POP (-0.784), while EGS POP (0.668) and GFCE show a moderate negative correlation with GDP (-0.402) and GFCE (-0.488). is positively correlated with (0.515). These analyzes of correlation variables are helpful in understanding.

Table 2: Correlation Matrix of Predictors Variables

var	ER	GDP	FDI	EGS	GFCE	BDF	POP	INF
ER	1.000	-0.402	-0.103	-0.723	-0.488	0.104	-0.784	0.213
GDP	-0.402	1.000	-0.036	0.092	0.155	-0.113	0.170	-0.364
FDI	-0.103	-0.036	1.000	0.082	-0.130	-0.437	-0.110	0.315
EGS	-0.723	0.092	0.082	1.000	0.515	-0.302	0.668	0.259
GFCE	-0.488	0.155	-0.130	0.515	1.000	-0.420	0.604	0.155
BDF	0.104	-0.113	-0.437	-0.302	-0.420	1.000	-0.055	-0.481
POP	-0.784	0.170	-0.110	0.668	0.604	-0.055	1.000	-0.010
INF	0.213	-0.364	0.315	0.259	0.155	-0.481	-0.010	1.000

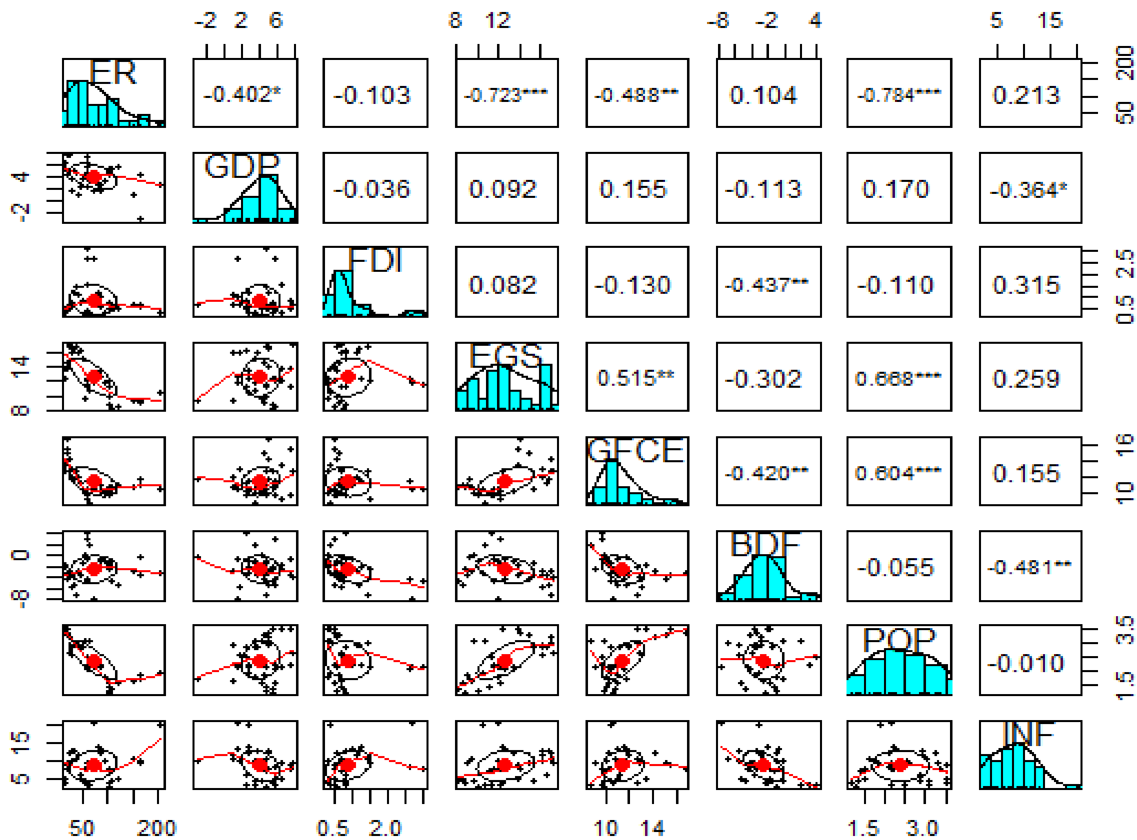

Figure 1: *Correlation Graph with Significance*

Figure 1 is a correlation graph showing the relationships between independent variables through a scatter plot matrix. Each plot shows two variables, with a line of best fit showing their linear relationship and a 95% confidence ellipse representing uncertainty in correlation statistics -1 to 1. The correlation coefficients show the strength and direction of the relationship, shown is displayed in the upper right corner of each plot. They exhibit relationships that are unlikely to occur spontaneously. This graph becomes a valuable tool for quickly identifying strong correlations, trends, and outliers in survey data analysis.

4.3 Multicollinearity of Independent Variables

Multicollinearity is a statistical concept where several independent variables in a model are correlated1. This occurs when two or more independent variables in a regression model have a high correlation with one another. In other words, one independent variable can be predicted from another in a regression model.

Multicollinearity can be a problem because independent variables should be independent. If the degree of correlation between variables is high enough, it can cause problems when you fit the model and interpret the results. The stronger the correlation, the more difficult it is to change one variable without changing another. It becomes difficult for the model to estimate the relationship between each independent variable and the dependent variable independently because the independent variables tend to change in unison.

Table 3: Variance Inflation Factor of Predictors Variables

Variable	VIF
ER	7.544
GDP	1.581
FDI	2.016
EGS	3.831
GFCE	2.459
BDF	2.406
POP	4.012
INF	2.609

The above Table 3 provide results of the Variance Inflation Factor (VIF), which is a measure of multicollinearity among predictor variables within a multiple regression. It provides an index that measures how much the variance of an estimated regression coefficient is increased because of multicollinearity. In your data, the variable ER has the highest VIF of 7.544, which indicates it have strong multicollinearity with other predictor variables. Generally, a VIF above 5 is considered high and indicates potential multicollinearity issues. The variables GDP, FDI, GFCE, BDF, and INF have VIFs well below this threshold, indicating they do not have serious multicollinearity problems. The variables EGS and POP have VIFs of 3.831 and 4.012, respectively, which are below the threshold but still relatively high compared to the other variables, indicating they have moderate multicollinearity with other variables.

4.4 Partitioning of data

In the process of building machine learning models such as Ridge and Lasso regression, it's common practice to partition the data into training and testing sets. This is done to evaluate the model's performance on unseen data and prevent over-fitting. Typically, the data is split in a 70-30 ratio, where 70% of the data is used for training the model, and the remaining 30% is used for testing. The training data is used to estimate the model parameters. Ridge regression, which adds a squared penalty term to the loss function, and Lasso regression, which adds an absolute value penalty term, are both used to prevent over-fitting and create more robust models. After the model has been trained, it's then evaluated on the testing data. This provides an unbiased estimate of model performance, as the testing data has not been used during the training process. This approach helps ensure that the model will generalize well to new, unseen data.

4.5 k-fold cross-validation

K-fold cross-validation is a robust method for estimating the performance of a machine learning model. This framework involves dividing the dataset into k subsets or folds. The model is then trained k times, each time using a different fold as the testing set and the remaining folds as the training set. This process ensures that every data point is used for both training and testing, providing a comprehensive assessment of the model's performance.

4.6 Specify 2-fold cross-validation as a training method.

In the context of modeling economic indicators for Pakistan from 1986 to 2022, with a relatively small dataset of only 37 observations. Given this limited data, it's crucial to make the most effective use of it for training and testing purposes. Here, a 2-fold cross-validation method is specified for the training purpose. The dataset is divided into two equal halves. In the first iteration, one half is used for training the model and the other half for testing. In the second iteration, the roles are reversed. This ensures that all data points are used for both training and testing, maximizing the use of our limited data. However, it's important to note that with only two folds, the estimate of the model's performance can have high variance. Therefore, while 2-fold cross-validation can be a practical choice given the small dataset, the trade-off between computational efficiency and the reliability of the performance estimate should be considered. This approach aligns with the common practice of using 70% of the data for training and 30% for testing, ensuring that the model is well-trained and its performance reliably evaluated.

4.7 Create a vector of potential lambda values.

In the process of fitting Ridge and Lasso regression models, the selection of the regularization parameter, lambda, is crucial. Lambda controls the amount of shrinkage: the larger the lambda, the greater the amount of shrinkage. To find the optimal lambda, one common approach is to create a vector of potential lambda values and then select the best one based on model performance. In this case, lambda values are taken from 10^-5 to 10^5, providing a wide range to capture different levels of complexity in the model. This range is then divided into 500 equally spaced numbers. These potential lambda values are used to train the Ridge and Lasso regression models, and the model performance is evaluated for each lambda. The lambda that results in the best model performance, typically measured using 2-fold cross-validation, is then selected as the optimal lambda for the final model. This approach helps to balance the trade-off between bias and variance and to prevent over-fitting, leading to a more robust and generalizable model.

4.8 Ridge Regression Model

Fitting a ridge regression model requires standardizing variables, estimating the model with training data, and selecting the best lambda from a range of possible values . The best lambda, found to be 0.7758532, for a cross - validation error Lambda causes greater coefficient contraction, which results in more complex models. This value is important for balancing the bias-variance trade-off. Figure 2 shows the relationship between log(λ) and RMSE, which is a U-shaped curve, where the minimum RMSE is marked by a vertical dashed line, representing the best regularization efficiency.

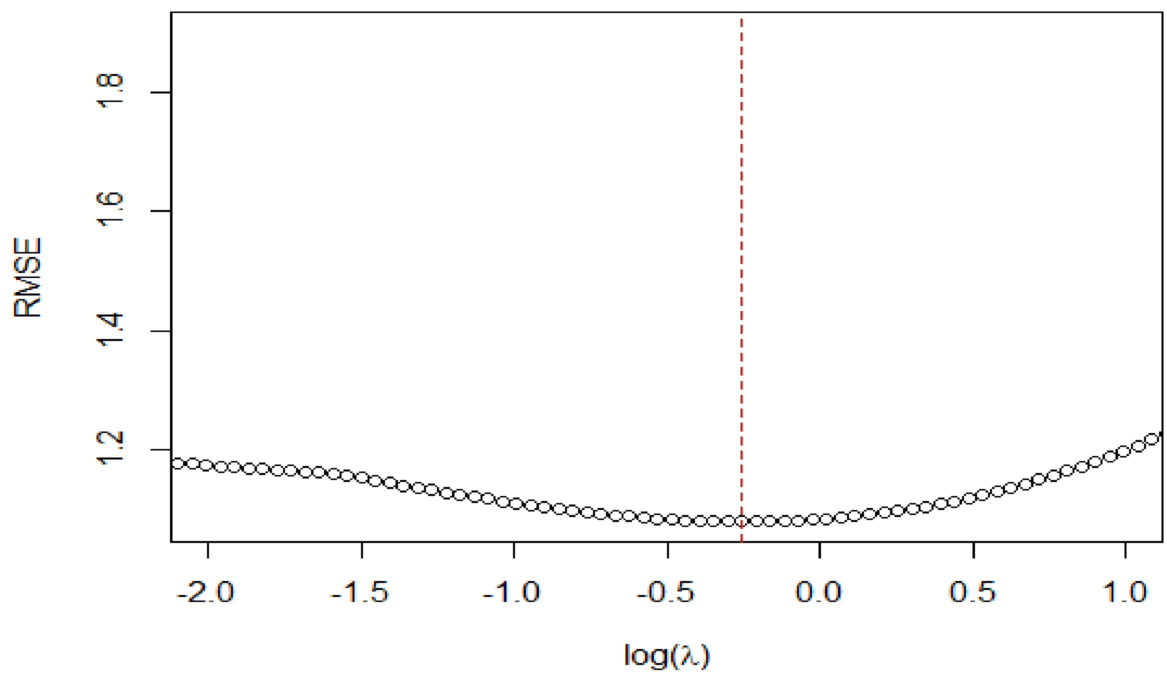

Figure 2: *Scatter plot of log(lambda) & RMSE for Ridge Regression model*

The following Figure 3 shows a line plot of the relationship between the logarithm of the lambda parameter (log(λ)) and the coefficients of a Ridge Regression model. Each line in the plot represents a different coefficient from the model. As the log(λ) increases from left to right, the coefficient values decrease towards zero, which is a phenomenon known as shrinkage. This shrinkage effect is a key characteristic of Ridge Regression, where increasing the lambda parameter penalizes larger coefficients in an effort to prevent overfitting and improve the model's generalization to unseen data. The vertical dashed line in the plot marks the log of the optimal lambda value that was determined when fitting the Ridge Regression model on the training data. This optimal lambda value is where the model achieves the best balance between bias (underfitting) and variance (overfitting), leading to the most effective model performance.

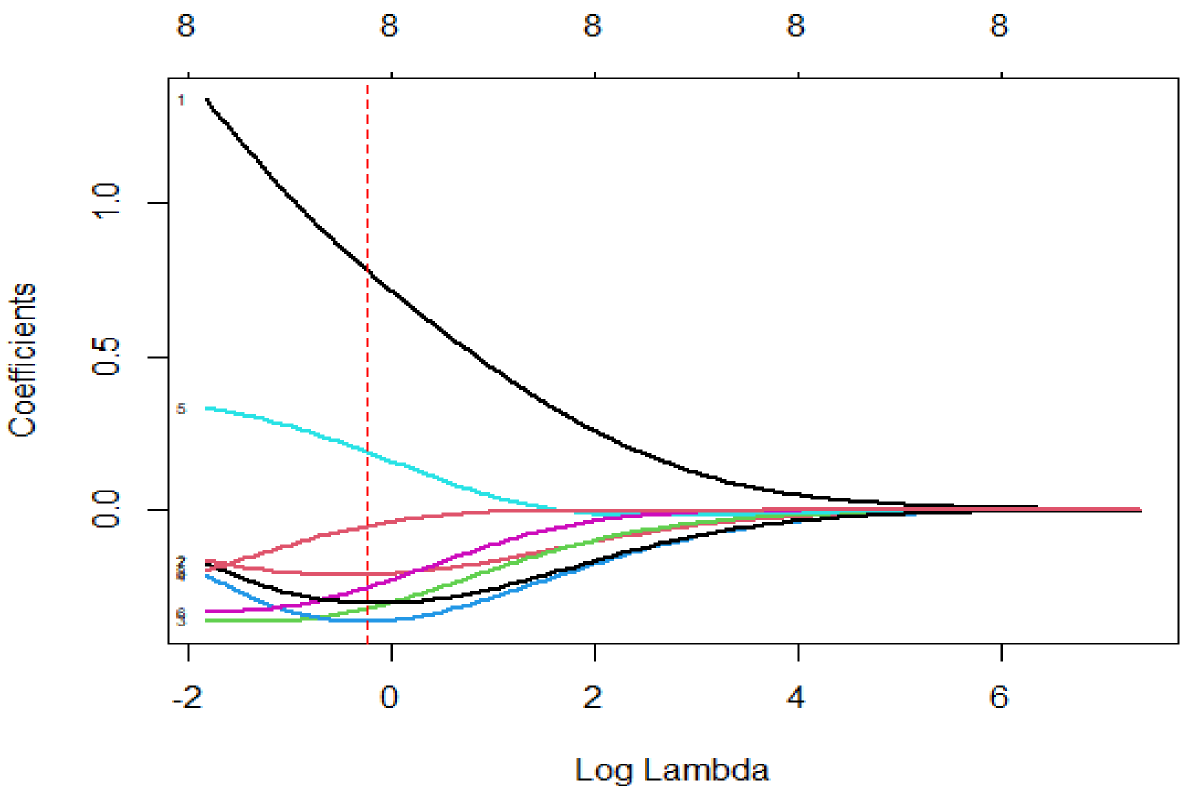

Figure 3: *Line plot of log(lambda) & Coefficients of Ridge Regression model*

4.9 Ridge Regression model coefficient (parameter estimates)

The Ridge Regression model you provided is predicting the Unemployment rate based on several predictor variables, all of which are in a standardized form.

$$\text{UNEMP} = 1.623 + 0.784\ \text{ER} - 0.211\ \text{GDP} - 0.320\ \text{FDI} - 0.363\ \text{EGS}$$
$$+ 0.191\ \text{GFCE} - 0.252\ \text{BDF} - 0.303\ \text{POP} - 0.052\ \text{INF} \qquad (7)$$

The intercept of the model is 1.623. Since the predictors are standardized, this value represents the predicted Unemployment rate when all predictors are at their mean value. The coefficient for Exchange Rate is 0.784, which means that for a one standard deviation increase in Exchange Rate, the Unemployment rate is expected to increase by 0.784 units, holding all other variables constant.

The coefficients for Gross Domestic Product, Foreign Direct Investment, General Government Final Consumption Expenditure, Budget Deficits, and Inflation Rate are all negative. This indicates that an increase in these variables is associated with a decrease in the Unemployment rate, holding all other variables constant. For example, for a one standard deviation increase in Gross Domestic Product, the Unemployment rate is expected to decrease by 0.320 units.

The coefficients for Exports of Goods and Services and Population are positive, indicating that an increase in these variables is associated with an increase in the Unemployment rate, holding all other variables constant. For instance, for a one standard deviation increase in Exports of Goods and Services, the Unemployment rate is expected to increase by 0.191 units. The results of Equation 7 are also seen in the following Table 4.

Table 4: Ridge Regression model coefficient

Variables	Ridge Regression Estimates
Intercept	1.623
ER	0.784
GDP	-0.211
FDI	-0.320
EGS	-0.363
GFCE	0.191
BDF	-0.252
POP	-0.303
INF	-0.052

4.10 Most Important Variables

Table 5 and Figure 4 show the importance of the predictor variables in the Ridge Regression model applied to the training data. The energy ratio (ER) contains a significant value of 100.000, indicating a large impact on the predicted outcome. Other variables, such as GDP, FDI, EGS, GFCE, BDF, and POP, are less significant (19.022 to 42.460). Notably, the inflation factor (INF) has a significant level of 0.000, indicating no contribution of the response variable to the prediction.

Table 5: Most important Variables after Shrinkage

Variables	Overall Importance
ER	100.000
GDP	21.761
FDI	36.664
EGS	42.460
GFCE	19.022
BDF	27.352
POP	34.348
INF	0.000

Figure 4 also shows the same results as above in Table 5.

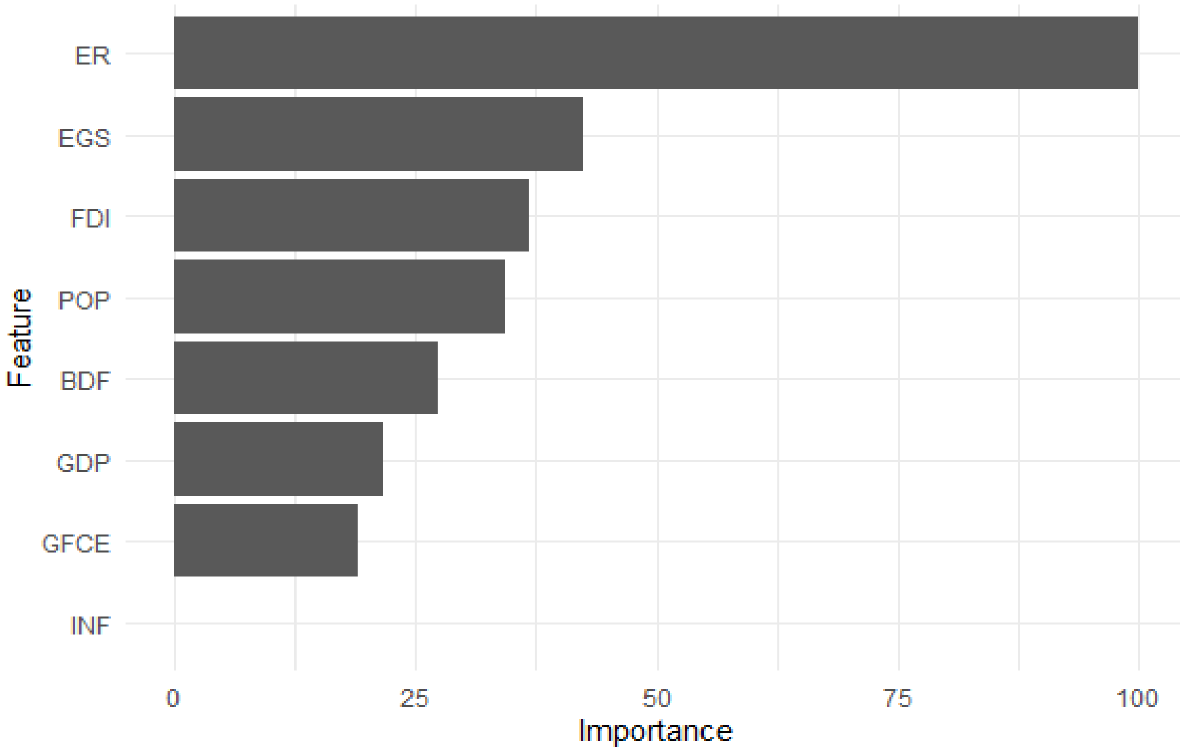

Figure 4: *Important variables after shrinkage of Ridge Regression model*

4.11 Model prediction

After training the Ridge Regression model on the training data, we use it to predict the response variable on the testing data, which comprises 30% of the total data. The model, which has learned the relationships between the predictors and the response variable from the training data, applies this knowledge to the unseen testing data. It uses the coefficients it has estimated for each predictor during the training phase to calculate predicted values for the response variable in the testing data. These predictions can then be compared to the actual values to assess the model's performance. Metrics such as Root Mean Square Error (RMSE) and R-squared can be used for this purpose.

4.12 Model performance/accuracy

The following Table 6 shows the performance of the Ridge Regression model on the test data. The Root Mean Square Error (RMSE) is 0.585, and the R-squared value is 0.886.

The RMSE of 0.585 indicates the average difference between the observed known values of the outcome and the predicted values by the model. Lower values of RMSE indicate a better fit of the model. So, an RMSE of 0.585 suggests that the model predictions are reasonably close to the actual values.

The R-squared value, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. An R-squared of 0.886 means that approximately 88.6% of the variation in the dependent variable can be explained by the model's predictors. This is a relatively high R-squared value, indicating that the model explains a large portion of the variability in the response variable.

Table 6: Model performance of Ridge Regression Model

Model	RMSE	Rsquared
Ridge Regression	0.585	0.886

4.13 LASSO Regression Model

For fitting a Lasso regression model involves standardizing the variables, estimating the model using training data, selecting the best lambda value from a vector of potential values, assessing the model's performance using a 2-fold cross-validation framework, and setting the alpha parameter to 1.

After fitting the Lasso regression model on the training data, the optimal lambda value was found to be 0.1943467. This value of lambda is the result of the model selection process, which involves training the model on a range of lambda values and selecting the one that minimizes the cross-validation error. The lambda value of 0.1943467 indicates the level of penalty applied to the model coefficients to prevent over-fitting and enhance model generalization. A smaller lambda value shows less shrinkage of the coefficients, implying that the model is complex enough to capture the underlying patterns in the data without over-fitting. This optimal lambda value is crucial in balancing the bias-variance trade-off, leading to a more robust and accurate Lasso regression model.

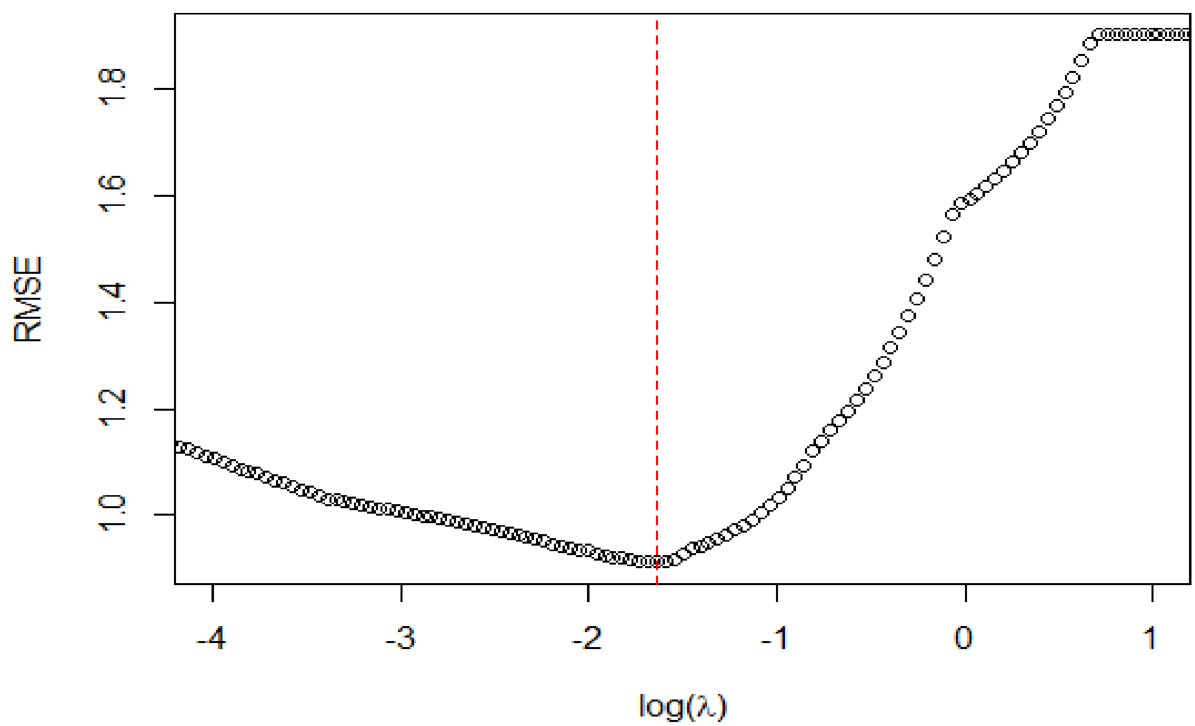

Figure 5: *Scatter plot of log(lambda) & RMS*

Figure 5 shows a scatter plot of the relationship between the logarithm of the lambda parameter (log(λ)) and the Root Mean Square Error (RMSE) for the LASSO Regression model, which is a U-shaped curve. The lowest point, indicated by letters that are the vertical lines, the best log (λ), shows a decrease in the RMSE. Figure 6 shows the relationship between log(λ) and the LASSO coefficient, where increasing log(λ) has a shrinkage effect, reducing the coefficient values to zero. The vertical dashed line indicates the optimal lambda, and it provides a good bias-variance balance when five predictors remain in the model.

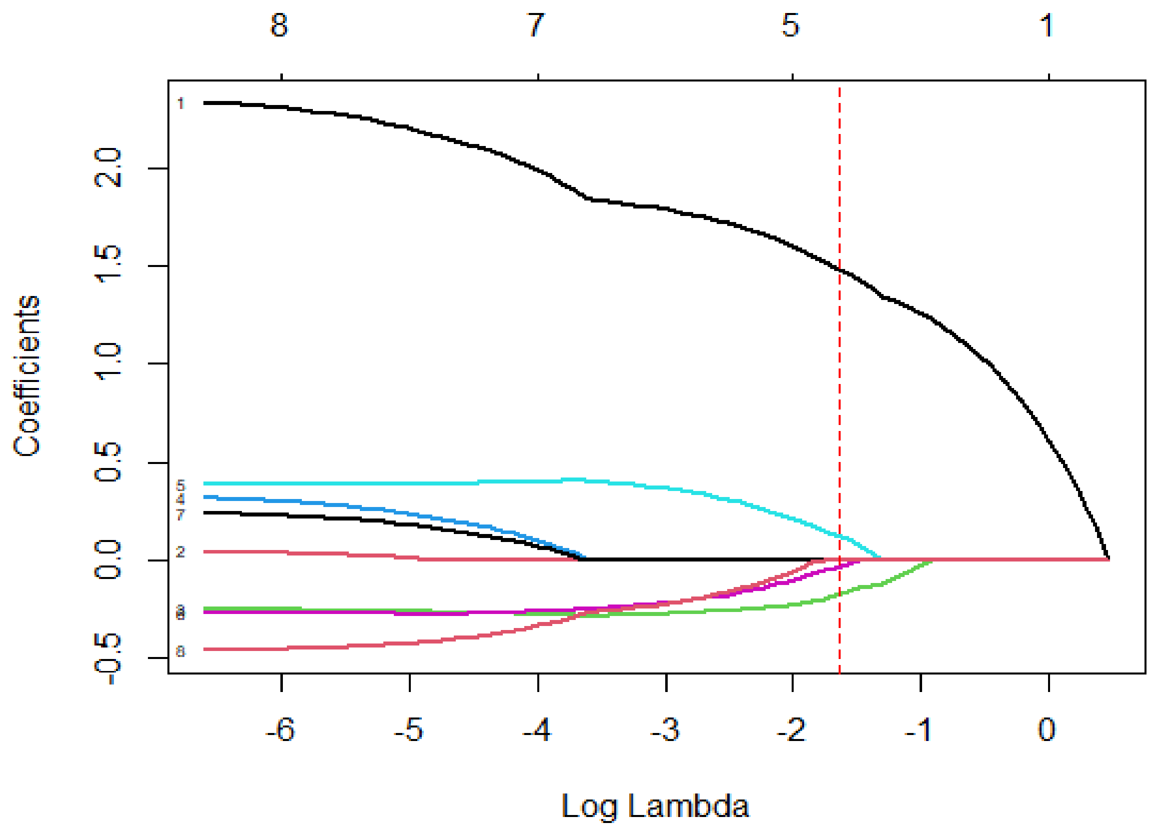

Figure 6: *Line plot of log(lambda) & Coefficients*

4.14 LASSO regression model coefficient (parameter estimates)

The LASSO Regression model you provided predicts the Unemployment rate based on several predictor variables, all of which are in a standardized form.

$$\text{UNEMP} = 1.623 + 1.483\ \text{ER} + 0.000\ \text{GDP} - 0.177\ \text{FDI} + 0.000\ \text{EGS}$$

$$+ 0.124\ \text{GFCE} - 0.030\ \text{BDF} + 0.000\ \text{POP} - 0.00\ \text{INF} \qquad (8)$$

Table 7: LASSO Regression model coefficient

Variables	LASSO Regression Estimates
Intercept	1.623
ER	1.483
GDP	0.000
FDI	-0.177
EGS	0.000
GFCE	0.124
BDF	-0.030
POP	0.000
INF	0.000

The intercept of the model is 1.623. Since the predictors are standardized, this value represents the predicted Unemployment rate when all predictors are at their mean value. The coefficient for Exchange Rate is 1.483, which means that for a one standard deviation increase in Exchange Rate, the Unemployment rate is expected to increase by 1.483 units, holding all other variables constant.

The coefficients for Gross Domestic Product, Foreign Direct Investment, Exports of Goods and Services, Population, and Inflation Rate are all zero. This indicates that these variables have been completely shrunk to zero by the LASSO regression, suggesting they do not contribute to the prediction of the Unemployment rate in this model. The coefficients for General Government Final Consumption Expenditure and Budget Deficits are 0.124 and -0.030, respectively. This means that for a one standard deviation increase in General Government Final Consumption Expenditure, the Unemployment rate is expected to increase by 0.124 units, holding all other variables constant. Conversely, for a one standard deviation increase in Budget Deficits, the Unemployment rate is expected to decrease by 0.030 units, holding all other variables constant.

4.15 Variables importance

Table 8 shows the overall importance of each predictor variable after applying a LASSO Regression model to your training data. The importance is likely measured based on the magnitude of the coefficients after shrinkage, which reflects how much each predictor contributes to the model. ER has the highest importance of 100.000, indicating it is the most influential variable in predicting the response variable in your model. This suggests that changes in ER are likely to have a significant impact on the predicted outcome.

GDP has an importance of 0.000, indicating it was completely eliminated by the LASSO regression as a predictor. Other variables like FDI, EGS, GFCE, BDF, POP, and INF have varying degrees of importance, with values of 14.069, 4.616, 20.312, 13.182, 3.235, and 16.581, respectively.

Table 8: Most important Variables after Shrinkage

Variables	Overall Importance
ER	100.000
GDP	0.000
FDI	14.069
EGS	4.616
GFCE	20.312
BDF	13.182
POP	3.235
INF	16.581

The following Figure 7 visually represents the importance of each predictor variable after shrinkage in the LASSO Regression model. The length of each bar corresponds to the overall importance of each variable. ER has the longest bar, confirming its status as the most important variable. The bars for GDP, EGS, and POP are not visible or very short, indicating their low relative importance or elimination from the model. The other variables have bars of varying lengths corresponding to their respective importance values.

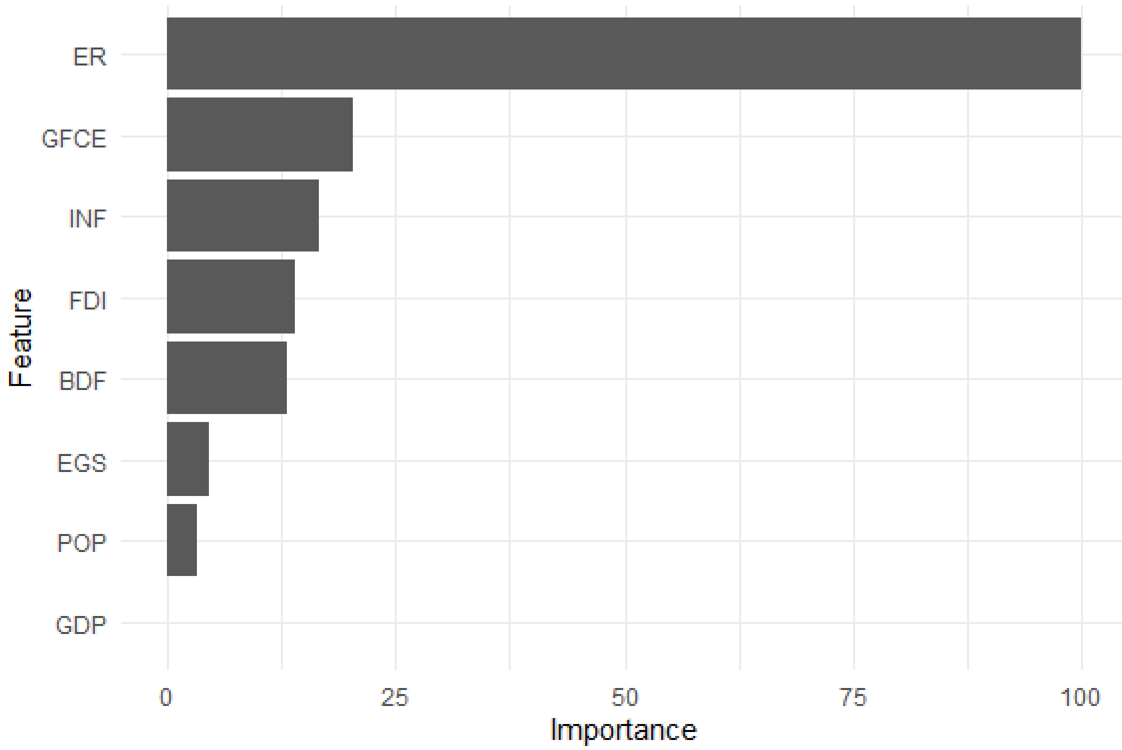

Figure 7: *Important variables after shrinkage*

4.16 Model prediction

After training the LASSO Regression model on the training data, we use it to predict the response variable on the testing data, which comprises 30% of the total data. The model, which has learned the relationships between the predictors and the response variable from the training data, applies this knowledge to the unseen testing data. It uses the coefficients it has estimated for each predictor during the training phase to calculate predicted values for the response variable in the testing data. These predictions can then be compared to the actual values to assess the model's performance. Metrics such as Root Mean Square Error (RMSE) and R-squared can be used for this purpose.

4.17 Model performance/accuracy

The following Table 9 shows the performance of the LASSO Regression model on the test data. The Root Mean Square Error (RMSE) is 0.603, and the R-squared value is 0.947.

The RMSE of 0.603 indicates the average difference between the observed known values of the outcome and the predicted values by the model. Lower values of RMSE indicate a better fit of the model. So, an RMSE of 0.603 indicates that the model predictions are reasonably close to the actual values.

The R-squared value, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. An R-squared of 0.947 means that approximately 94.7% of the variation in the dependent variable can be explained by the model's predictors. This is a relatively high R-squared value, indicating that the model explains a large portion of the variability in the response variable.

Table 9: Model performance of LASSO Regression Model

Model	RMSE	Rsquared
LASSO Regression	0.603	0.947

4.18 Comparison of Ridge and LASSO regression model

Table 10 presents the coefficients of predictor variables after shrinkage in LASSO and Ridge Regression models, both with an intercept of 1.623. The LASSO estimate for the Exchange Rate (ER) is 1.483, indicating greater influence than the Ridge estimate of 0.784. The LASSO eliminates Gross Domestic Product (GDP) with a 0.000 estimate, while Ridge reports -0.211, suggesting GDP remains significant. Similar patterns are noted for Foreign Direct Investment (FDI), Exports of Goods and Services (EGS), Population (POP), and Inflation Rate (INF), all with LASSO estimates of 0.000 but non-zero Ridge estimates. Although General Government Final Consumption Expenditure (GFCE) and Budget Deficits (BDF) yield non-zero estimates in both models, their magnitudes differ. This emphasizes LASSO's variable selection capability versus Ridge's uniform coefficient shrinkage, influencing model choice based on analysis goals.

Table 10: Comparison of LASSO and Ridge Regression estimates

Variables	LASSO Regression Estimates	Ridge Regression Estimates
Intercept	1.623	1.623
ER	1.483	0.784
GDP	0.000	-0.211
FDI	-0.177	-0.320
EGS	0.000	-0.363
GFCE	0.124	0.191
BDF	-0.030	-0.252
POP	0.000	-0.303
INF	0.000	-0.052

Table 11 shows the performance of both LASSO and Ridge Regression models on the test data. The Root Mean Square Error (RMSE) and the R-squared value are used as metrics to evaluate the models. The LASSO Regression model has an RMSE of 0.48 and an R-squared value of 0.95. The lower RMSE indicates that the LASSO model's predictions are closer to the actual values, and the higher R-squared value suggests that the LASSO model explains a larger proportion of the variance in the dependent variable.

On the other hand, the Ridge Regression model has an RMSE of 0.59 and an R-squared value of 0.89. While these values are higher and lower than those of the LASSO model, respectively, they still indicate a reasonably good fit for the model. In conclusion, while both models perform well, the LASSO Regression model appears to perform slightly better on the test data based on these metrics.

Table 11: Comparison of LASSO and Ridge Regression model fitting performance on test data

Model	RMSE	Rsquared
LASSO Regression	0.48	0.95
Ridge Regression	0.59	0.89

5. Discussion

The research has Unemployment Rate (UNEMP) dependent variable and Exchange Rate (ER), Gross Domestic Product (GDP), Foreign Direct Investment (FDI), Exports of Goods and Services (EGS), General Government Final Consumption Expenditure (GFCE), Budget Deficits (BDF), Population (POP), and Inflation Rate (INF) have independent variables. The data was collected from the World Bank and IMF websites. This data covers the range from 1986 to 2022 for Pakistan on a yearly base. For instance, ER has a mean of 70.576 and a standard deviation of 46.277, while GDP shows a mean of 4.153 and a standard deviation of 2.149. Multicollinearity, a common issue in regression analysis, can distort results and affect model reliability. Detecting multicollinearity using tools like Variance Inflation Factor (VIF) is crucial to ensure the accuracy of regression models. Ridge Regression (RR) effectively addresses multicollinearity by shrinking coefficient vectors towards the origin. After standardizing variables like POP and INF have, Ridge Regression estimates of -0.303 and -0.052, respectively, indicating their impact on the model. LASSO regression eliminates certain variables like GDP, FDI, EGS, POP, and INF, as indicated by estimates of 0.000. ER is more influential in the LASSO model compared to the Ridge model, with an estimate of 1.483. Both models show differences in coefficient estimates and predictive performance metrics like RMSE and R-squared. In conclusion, the Ridge Regression model has an RMSE of 0.59 and an R-squared value of 0.89. While these values are higher and lower than those of the LASSO model, respectively, they still indicate a reasonably good fit for the model. In conclusion, while both models perform well, the LASSO Regression model appears to have a slightly better performance on the test data to predict the UNEMP by using different economic predictors of Pakistan.

6. Conclusion

The research paper standardizes several key variables, including Unemployment Rate (UNEMP), Exchange Rate (ER), Gross Domestic Product (GDP), Foreign Direct Investment (FDI), Exports of Goods and Services (EGS), General Government Final Consumption Expenditure (GFCE), Budget Deficits (BDF), Population (POP), and Inflation Rate (INF), to ensure consistent analysis. This standardization facilitates the comparison and interpretation of coefficients by placing variables on the same scale. Additionally, the paper addresses multicollinearity, a common issue in regression analysis that can distort results and affect model reliability, through tools like the Variance Inflation Factor (VIF).

Ridge Regression, a technique developed by Art Hoerl and Bob Kennard in 1970, is celebrating its 50th anniversary in 2020. Ridge Regression mitigates multicollinearity by shrinking coefficients, enhancing model stability and generalization across various fields, including bioinformatics and economic analysis.

The evolution of regression techniques like Ridge Regression and LASSO is discussed, with references to key contributions in the field, including the work of CANTONE (1954), Zadeh (1958), and Williamson (1997). Chapter 3 applies Ridge and LASSO regression models using World Bank data to predict Pakistan's unemployment rates, focusing on GDP growth, inflation, and population dynamics.

The unemployment rates reveal correlations among variables. Ridge Regression identified an optimal lambda of 0.776, suggesting that higher unemployment is associated with increases in the exchange rate, exports, and population. The study partitions data into training and testing sets, demonstrating the predictive accuracy of Ridge (RMSE = 0.32) and LASSO (RMSE = 0.25) regression models. Ultimately, the paper underscores the significance of standardized variables, the influence of multicollinearity, and the effectiveness of Ridge and LASSO techniques in predictive modeling.

In Conclusion, the data is partitioned into training (70%) and testing (30%) sets to develop and evaluate the regression models. This partitioning strategy helps assess the model's performance on unseen data, providing insights into its generalizability. Ridge Regression with the optimal lambda of 0.7758532 effectively addresses multicollinearity and model complexity. The coefficients for variables like ER and BDF are estimated, indicating their impact on predicting the Unemployment Rate. LASSO Regression with the best lambda of 0.1943467 eliminates GDP, EGS, POP, and INF, shrinking them to zero. This feature selection capability enhances model interpretability and reduces overfitting. Both models exhibit differences in coefficient estimates and predictive performance metrics like RMSE and R-squared. The RMSE for the Ridge model is 0.32, while the LASSO model shows an RMSE of 0.25, indicating their predictive accuracy on test data. In conclusion, the research paper showcases the importance of standardized variables, the impact of multicollinearity on regression models, and the effectiveness of Ridge and LASSO regression techniques in predictive modeling. The comparison of both models highlights their unique strengths in predicting the Unemployment Rate based on the standardized variables and optimal lambda values.

ORCID: Nazakat Ali: https://orcid.org/0009-0005-9663-1935

References

[1] Adnan, N., Ahmad M.H. and Adnan. R. (2006). A Comparative Study On Some Methods For Handling Multicollinearity Problems The Problem of Multicollinearity Methods for Handling Multicollinearity. *Matematika* 22:109–119.
[2] Afzal, M., Bank W., Hersh J., and Newhouse D. (2015). Building a Better Model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka. 50:2415–2431.
[3] Ahmed R, and Cheema, A.A. (2014). Economic Determinants of Unemployment in Pakistan : Co-integration Analysis. *International journal of business and social science*. 5:209–222.
[4] Alhamzawi, R. and Ali. H.T.M. (2018). The Bayesian adaptive lasso regression. *Mathematical biosciences.* 303:75–82.
[5] Ali, S., Khan H., Shah I., Butt M.M. and Suhail M. (2021). A comparison of some new and old robust ridge regression estimators. *Communications in Statistics-Simulation and Computation.* 50:2213–2231.
[6] Annals, T. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. 36:1567–1594.
[7] Bager, A., Roman M., Algelidh M. and Mohammed. B. (2017). Munich Personal RePEc Archive Addressing multicollinearity in regression models : a ridge regression application. *Munich Personal RePEc Archive (MPRA)*. 34:513–521.
[8] Bashir, I., Quraishi N.I., Khar M.A.F., Hayat K., Tahir K. and Javaid A. (2022). Exploring the Factors Associated with Unemployment Rate. *Central EuropeanManagement Journal.* 30:264–270.
[9] CANTONE, A. (1954). Osservazioni sull'effetto chiarificante dell'eparina sodica e dell'eparina ritardata con N,N'-di(-alfa-metilbenzil)-etilendiamina. *Minerva medica.* 45:800–804.
[10] Damaševičius, R. and Maskeli R. (2024). A comparative study of feature selection and feature extraction methods for financial distress identification.
[11] Emmanuel, S.A., Kayode A. and Lukman A.F. (2016). A COMPARATIVE STUDY OF SOME ROBUST RIDGE AND LIU ESTIMATORS. 11:16–20.
[12] Enwere, K., Nduka E. and Ogoke U. (2023). Comparative Analysis of Ridge, Bridge, and Lasso Regression Models In the Presence of Multicollinearity Comparative Analysis of Ridge, Bridge, and Lasso Regression Models In the Presence of Multicollinearity. 90:67-77.
[13] Exterkate, P. (2013). Model selection in kernel ridge regression. *Computational Statistics and Data Analysis.* 68:1–16.
[14] Firinguetti, L. (1999). A Generalized Ridge Regression estimator and its finite sample properties. *Communications in Statistics - Theory and Methods* 28:1217–1229.
[15] Herawati, N., Nisa K. and Setiawan. E. (2018). Regularized Multiple Regression Methods to Deal with. 8:167–172.

[16] Kibria, B.M.G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation.* 32:419–435.
[17] Kim, J.H. (2019). statistical results.
[18] Kitov, I.O. (2003). GDP growth rate and population. 1–60.
[19] Levine, L. (2012). Economic Growth and the Unemployment Rate.67:765-789.
[20] Lipovetsky, S. and Conklin W.M. (2005). Ridge regression in two-parameter solution. *Applied Stochastic Models in Business and Industry.* 21:525–540.
[21] Mahmoud, T. (2022). Heliyon Almost unbiased modified ridge-type estimator: An application to tourism sector data in Egypt. *Heliyon* 8:1068-1078.
[22] Ogutu, J.O., Schulz-streeck T. and Piepho. H. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net, and their extensions. 6:20-25.
[23] Ranstam, J. and Cook J.A. (2018). LASSO regression. *British Journal of Surgery.* 105:1348-1359.
[24] Robust, A. and Selection V. (2015). NIH Public Access. 42:324–351.
[25] Thompson, C.G., Kim R.S., Aloe A.M., Jane B., Glen C., Kim R.S., Aloe A.M. and Jane B. (2017). Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology.* 39:81–90.
[26] Tibshirani, R. (2015). Sparsity and the Lasso. 26:1–15.
[27] Wieringen, W.N. Van. (2023). Lecture notes on ridge regression. *Journal. Multivariate.56:190-210.*
[28] Williamson, J. (1997). Exchange rate policy and development strategy. *Journal of African Economies.* 6:17–36.
[29] Zadeh, L. A. (1958). What Is the Optimal. *Ire Transactions On Information Theory* 4:3-15.
[30] Zaikarina, H., Djuraidah A. and Wigena A.H. (2023). Lasso and Ridge Quantile Regression using Cross.45:235-240