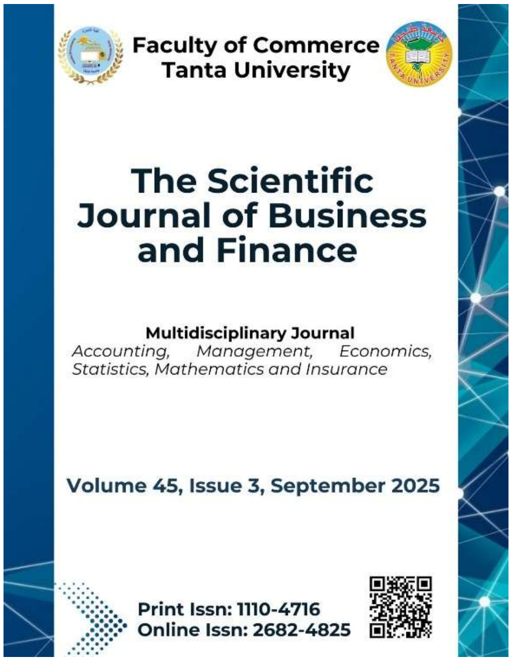

The Scientific Journal of Business and Finance
https://caf.journals.ekb.eg

Enhancing Regression Models with Regularization and Hybrid Techniques to Address Multicollinearity

Abdel-reheem Awad Bassuny[a] & Hanaa Abdel-Reheem Ibrahim Salem[b]

[a] Lecturer at the Higher Institute of Management in EL Mahalla El-Kubra PhD Statistics, Faculty of Commerce, Tanta University, Egypt.
[b] Assistant professor in Statistics, Faculty of Commerce, Tanta University, Egypt

Published online: **September 2025.**
To cite this article: Bassuny, Abdelreheem Awad. & Salem, Hanaa Abdel Reheem Ibrahim.(2025) Enhancing Regression Models with Regularization and Hybrid Techniques to Address Multicollinearity *The Scientific Journal of Business and Finance*, 45, (3),pp.333-360

DOI: 10.21608/caf.2025.455702

*Corresponding author: dr-AbdelreheemBassuny@outlook.com

Enhancing Regression Models with Regularization and Hybrid Techniques to Address Multicollinearity

Abdel-reheem Awad Bassuny
Higher Institute of Management in EL Mahalla El-Kubra PhD Statistics, Faculty of Commerce, Tanta University, Egypt.

Hanaa Abdel-Reheem Ibrahim Salem
Statistics, Faculty of Commerce, Tanta University, Egypt.

Article History
Received 20 July 2025, Accepted 11 August 2025, Available online September 2025.

Abstract

This study investigates the application of regularization techniques—Ridge, Lasso, Elastic Net, and their ensembles (Ridge-Elastic Net and Lasso-Elastic Net)—to correct multicollinearity in regression models of forecasting internal migration rates. With a sample dataset of 250 regions (2020–2024) and 12 highly correlated predictors, such as income, unemployment, and healthcare quality, we compare these techniques with ordinary least squares (OLS). Multicollinearity is confirmed with high Variance Inflation Factors (VIF > 10), high correlations (e.g., r = 0.85 between income and cost of living), and eigenvalue. Results show that Elastic Net and Ridge-Elastic Net are superior, with the lowest MAE (0.17), MSE (0.26), and the highest R^2 (0.84), while exhibiting moderate variable selection (excluding population density). Lasso-Elastic Net and Lasso simplify models to the exclusion of transportation and population density but also yield slightly poor performance (MAE = 0.18, MSE = 0.27, R^2 = 0.83). Ridge attains Elastic Net's prediction performance but retains all variables, while OLS is poor (MAE = 0.20, MSE = 0.30, R^2 = 0.80). Elastic Net and Ridge-Elastic Net are the best picks for most accuracy, while Lasso-Elastic Net is preferred in scenarios that appreciate model simplicity. The findings highlight the strength of regularization in enhancing model stability and predictive accuracy in the presence of multicollinearity.

Keywords: Multicollinearity; ridge regression; lasso regression; elastic net; hybrid models.

1. introduction

Multicollinearity, characterized by high correlations among independent variables in linear regression models, poses significant challenges, including unstable coefficient estimates, inflated variances, and reduced interpretability. These issues often lead to misleading statistical inferences, as coefficients may exhibit incorrect signs or questionable significance. Regularization techniques, such as Ridge, Lasso, and Elastic Net, have been widely adopted to mitigate multicollinearity by penalizing coefficients, thereby enhancing model stability and predictive accuracy. These methods have proven effective across various fields, including economics, medicine, and social sciences, as demonstrated by studies like Addable (2020), who concentrated on ill-conditioned design matrices. The study concluded that ridge regression effectively lowers variance, Lasso is best for estimation and selection separately, and Elastic Net combines the advantages of Lasso for estimation and selection independently. Although Elastic Net outperformed Lasso and Ridge in regression models, it was surprising that the basic Elastic Net performed better in location models than the standard Elastic Net. These results show how Elastic Net can effortlessly strike a balance between variable selection and variance reduction.

According to this,(Usman *et al*.,2021) used breast cancer survival data from Ahmadu Bello University Teaching Hospital to compare the predictive power of Ridge, Lasso, and Elastic Net against ordinary least squares (OLS). While OLS failed because of multicollinearity when all predictors were used, regularized methods produced significant results. Lasso outperformed Ridge and Elastic Net with the highest R-squared of 0.3226 and the mean squared error (MSE) of 0.832178. Age (30–59), marital status, and disease stage were significant predictors of survival time; a longer survival time was associated with Stage 1, while a shorter one was associated with Stages 2–3. This study highlights situationally dependent method selection and demonstrates how Lasso functions in particular contexts.

To improve their performance, recent research has investigated combining regularization strategies with sophisticated analytical techniques. In order to reduce multicollinearity in nonlinear and nonstationary multivariate time-series data,(Al-Jawarneh *et al*.,2021) proposed ELNET-EMD, an Elastic Net (ELNET) model combined with Empirical Mode Decomposition (EMD). ELNET-EMD enhanced variable selection and prediction accuracy by splitting predictors into intrinsic mode functions (IMFs) and a residual component. This approach performed better than OLS-EMD and Lasso-EMD on simulated data and daily exchange rate datasets, displaying lower error values (RMSE, MAE, MAPE). Here, we show that this new approach can be applied to complex multicollinearity problems.

In a simulation study employing high-dimensional data with correlated predictors,(Altelbany ,2022) compared Elastic Net, Lasso, and Ridge to further bolster Elastic Net's stability. The findings showed that Elastic Net performed better than Lasso and Ridge, increasing the accuracy of variable selection and lowering MSE by as much as 12%. Similarly, Elastic Net outperformed Lasso by 10% in terms of MSE, even when the predictors were correlated (Sari and Sari, 2023). These findings suggest that the most effective method for lowering multicollinearity is Elastic Net's double-penalty scheme.

Regularization techniques encounter further challenges when the number of predictors exceeds the number of observations ($p > n$). Gana (2022) suggested a generalized ridge regression procedure that selects significant regressors by dividing the predictor matrix and then using t-ratios. With an 86% and 74% reduction in squared distances from true coefficients for significant and true coefficients, respectively, and a 99% chance of identifying true regressors, the method performed better than Elastic Net. Elastic Net, however, maintained its competitiveness, especially in variable selection stability.(Wang *et al.,*2023), on the other hand, offered a cohesive method for high-dimensional regression that covered the best selection of penalty parameters for Lasso, Ridge, and

Elastic Net. Their simulations demonstrated the value of algorithmic enhancement in regularization and validated Elastic Net's superiority in variable selection and prediction.

Sari and Sari (2023) compared the Lasso, Elastic Net, and Ridge regularization techniques. Elastic Net was found to perform better in terms of variable selection and prediction when correlated predictors were present. The study found that Lasso's mean squared error was 10% higher than Elastic Net's. Recent studies by(Herawati *et al*.,2024) and Kumar and Patel (2024) also commend Elastic Net's performance. In comparison to Lasso, Elastic Net reduced prediction error by about 14%, according to Kumar and Patel's tests, with highly multicollinear data sets yielding the best bias-variance trade-off. In comparison to Lasso, Herawati et al. were able to reduce Elastic Net's MSE by 15%, and Elastic Net demonstrated balanced performance in high-dimensional data. Furthermore, Elastic Net produced the highest coefficient of determination (60.81%) and the lowest RMSE (3.3977) when (Nur *et al*., 2024) applied these techniques to examine infant mortality rates in South Sulawesi, Indonesia. These outcomes show how Elastic Net can be used with dependent predictors in complex real-world datasets.

Elastic Net's dominance is further supported by recent studies. Sari and Widyaningsih (2025) found that Elastic Net performed 12% better than Lasso and Ridge in terms of prediction accuracy on high-dimensional data with highly correlated predictors. Additionally, stability selection for Lasso, Ridge, and Elastic Net for Accelerated Failure Time (AFT) models of high-dimensional survival data was studied by (Khan *et al.,* 2025). Their results showed that, especially when $p > n$, stability selection improves variable selection stability by lowering false positives and negatives. Recent research indicates that Elastic Net with stability selection generated more compact models and effectively identified important variables in breast cancer data, including PRC1, age, and ZNF533.

Despite the extensive literature on regularization, a critical research gap remains in the comprehensive evaluation of hybrid regularization models, such as Ridge-Elastic Net and Lasso-Elastic Net, particularly in datasets with high multicollinearity, such as those related to internal migration rates. Previous studies have primarily focused on individual regularization techniques (e.g., Sari & Sari, 2023;(Herawati *et al.*, 2024) or their application in specific domains like survival analysis (Khan *et al*.,2025) or time-series data (Al-Jawarneh *et al*., 2021). However, there is a lack of comparative studies that systematically assess the performance of hybrid models combining the stability of Ridge with the variable selection capabilities of Lasso or Elastic Net in the context of internal migration data, where predictors like income, cost of living, and healthcare quality are highly correlated (e.g., $r = 0.85$).

This study aims to address this gap by evaluating the effectiveness of hybrid regularization models (Ridge-Elastic Net and Lasso-Elastic Net) against traditional methods (Ridge, Lasso, Elastic Net, and OLS) in mitigating multicollinearity and improving model stability and predictive accuracy. Using a dataset of 250 regions (2020–2024) with 12 highly correlated predictors, we propose and test novel hybrid approaches to balance coefficient stability and variable selection. The primary objective is to determine whether these hybrid models outperform classical regularization techniques in handling multicollinearity, offering more robust and interpretable models for

predicting internal migration rates. By doing so, this research contributes to the statistical modeling literature by introducing tailored hybrid solutions for complex, multicollinear datasets. This research not only advances the theoretical understanding of hybrid regularization but also provides practical insights for policymakers analyzing migration patterns in highly correlated datasets.
The primary objective of this study is to propose and evaluate hybrid regularization models to outperform classical methods in addressing multicollinearity.

2. Methodology

Multicollinearity occurs in regression analysis when independent (predictor) variables show high or quasi-linear intercorrelations. Multicollinearity raises questions about the coefficient estimates without altering the model's predictability because it makes it challenging to discern the distinct effects of each independent variable on the dependent variable.

2.1 Impact of Multicollinearity on Parameters

The problem of high multicollinearity among independent variables leads to a considerable amount of instability in highly sensitive coefficient estimates to small changes in the data. Coefficient estimates could become exaggerated or illogical as a result. Furthermore, it raises the variance of estimates and results in inflated standard errors, which makes it challenging to assess the significance of variables using statistical tests of significance like the t-test.
Furthermore, high independent variable correlations make it more difficult to interpret the model's results because it is hard to determine how each variable independently contributes to the dependent variable. The interpretation of the results is further complicated by the fact that, due to this strong correlation, coefficients may occasionally show signs (positive or negative) that differ from the ratio expectations predicted by theory

2.2 Methods for Detecting Multicollinearity(Kutner,2005)

2.2.1 Correlation Coefficient

- Description: Uses a correlation coefficient (such as Pearson) to quantify the linear relationship between pairs of independent variables.
- Test: Create a correlation matrix and find pairs that have a high correlation (for example, $|r| > 0.8$).
- Benefit: Easy and quick to put into practice.
- Limitation: Is unable to identify intricate connections between several variables.

2.2.2 Farrar& Glauber (1967) Test

- DescriptionThree sub-tests are used to analyze the correlation matrix: eigenvalues, partial correlations, and a chi-square test to identify multicollinearity.
- Test:
- Compute partial correlations between variables while controlling for other variables.

- Analyze the eigenvalues of the correlation matrix; small eigenvalues indicate multicollinearity.
- Apply a chi-square statistic:

$$\chi^2=-(n-1-\frac{2k+5}{6})\ln|R| \quad (1)$$

where n is the sample size, k is the number of variables, and |R | is the determinant of the correlation matrix. A significant X^2(p-value < 0.05) suggests multicollinearity.

- Advantage: Comprehensive approach combining multiple techniques.
- Limitation: Less commonly used today due to computational complexity.

2.2.3 Variance Inflation Factor (VIF)

- Description: Quantifies how much the variance of a regression coefficient increases due to correlations with other variables.
- Test: Calculated as:

$$VIF_j=\frac{1}{1-R_j^2} \quad (2)$$

where $R^2$is the coefficient of determination from regressing variable X_j on all other variables. A VIF > 10 (or > 5 in some cases) indicates multicollinearity.

- Advantage: Precise and detect multivariate correlations.
- Limitation: Requires computing a for each variable.

2.2.4 Tolerance

- Description: The reciprocal of VIF, defined as:

$$Tolerance = 1 - R^2_j (3)$$

- Test: A tolerance value < 0.1 suggests multicollinearity.
- Advantage: Simple and complementary to VIF.
- Limitation: This relies on the same calculations as VIF.

2.2.5 Eigenvalues and Condition Index

- Description: Examines the eigenvalues of the correlation matrix. Very small eigenvalues indicate multicollinearity. The condition index is calculated as:

$$\text{Condition Index}=\sqrt{\frac{\lambda\, \boldsymbol{max}}{\lambda\, \boldsymbol{min}}}(4)$$

- Test: A condition index > 30 indicates severe multicollinearity.
- Advantage: Effective for detecting complex correlations.
- Limitation: Requires expertise in eigenvalue analysis.

2.3 Inspection of Regression Coefficients

- DescriptionExamines regression coefficients for high standard errors, unexpected signs, or large values.
- Test: Look for high-standard errors and compare coefficients to theoretical predictions.

- Benefit: No further calculations are required.
- Limitation: Indirect and susceptible to external influences.

2.4 Methods for Addressing the Multicollinearity Problem in Regression Models

The presence of intercorrelation between independent variables in a regression analysis is known as multicollinearity, and it can cause instability and unreliability in the estimates. We outline the most well-known methods for dealing with this problem below.

2.4.1 Ridge Regression

A statistical method for improving linear regression models—which are primarily used to address problems with overfitting and multicollinearity in data—is called ridge regression. Ridge Regression reduces the size of the model coefficients without eliminating them by adding a "penalty" to the loss function. Regularization—more specifically, L_2 regularization—is a technique that maintains all features while minimizing their over-reliance (Hoerl & Kennard, 1970).

2.4.1.1 Mathematical Foundation

Ridge Regression is built upon modifying the ordinary least squares (OLS) equation by adding a penalty term that depends on the sum of the squared coefficients(Fuwenjiang, 1998):

1. Modified Loss Function

Standard linear models aim to minimize the **residual sum of squares** (RSS):

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

In Ridge Regression, an L_2 penalty is added to this function:

$$\text{Loss} = \text{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (5)$$

Where:

- λ (lambda): The regularization parameter, controlling the strength of the penalty.
- $\sum \beta_j^2$: The sum of the squared regression coefficients (excluding the intercept).

2. Closed-Form Solution

The Ridge Regression estimator is calculated using the matrix formula:

$$\grave{\beta}_{\text{ridge}} = (X^T X + \lambda I_p)^{-1} X^T Y \qquad (6)$$

Where:

- X: The matrix of independent data.
- Y: The dependent target variable.
- I_p: An identity matrix of dimensions $p \times p$.\

3. Effect of λ

- $\lambda = 0$: The model returns the results of ordinary linear regression (no regularization).
- $\lambda \to \infty$: All coefficients approach zero (increased bias, reduced variance).
- Intermediate values of λ: Coefficients undergo shrinkage, where:
  - Larger coefficients shrink relatively more than smaller ones.
- $\grave{\beta}_j^{\text{ridge}} = \frac{d_j^2}{d_j^2+\lambda} \mathrm{u}_j^T \mathrm{y}$

  Here, d_j are the singular values of the matrix X, and $\mathbf{u}_j$ are the singular vectors.

2.4.1.2 Geometric Interpretation

The bias-variance tradeoff is characterized by:

- Increased bias: Reduced model accuracy on training data.
- Reduced variance: Improved model performance on new data (better generalization).

By selecting the optimal λ (e.g., through cross-validation), the model becomes more balanced.

Ridge Regression is particularly effective when

- The training data is small compared to the number of features.
- There is a high correlation among features (multicollinearity).

2.4.2 Lasso Regression

Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a regularization technique designed to enhance the performance of linear regression models when dealing with high-dimensional data or when variable selection is essential. Unlike Ridge Regression, which employs an L_2 penalty, Lasso induces sparsity of the model parameters using an L_1 penalty in the loss function (Tibshirani, 1996). Thus, by precisely shrinking some coefficients to zero and decreasing the magnitude of other coefficients, Lasso accomplishes a kind of automatic variable selection (Hastie *et al.,*2015)

2.4.2.1 Mathematical Foundation

Lasso Regression modifies the ordinary least squares (OLS) objective by adding a penalty proportional to the sum of the absolute values of the regression coefficients. The loss function is defined as(James *et al.,*2021)

$$\mathrm{RSS}_{\mathrm{Lasso}} = \sum_{i=1}^{n} (y_i - \grave{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{7}$$

Where:

- y_i are the observed values,
- $\grave{y}_i$ are the predicted values,
- β_j are the regression coefficients,

- λ is the regularization parameter controlling the strength of the penalty.

2.4.2.2 Properties and Effects

- Variable Selection: The L_1 penalty can force some coefficients to be exactly zero, thus selecting a simpler model that includes only the most relevant predictors.
- Bias-Variance Tradeoff: As λ increases, more coefficients are set to zero, increasing model bias but reducing variance and overfitting.
- Interpretability: By reducing the number of predictors, Lasso enhances model interpretability, making it useful in scientific and applied research where understanding the role of each variable is important.

2.4.2.3 Practical Considerations

- Model Selection: The optimal value of λ is typically chosen via cross-validation to balance model complexity and predictive accuracy.
- Limitations: Lasso may struggle when there are highly correlated predictors, as it tends to select only one variable from a group and ignore the others, which can lead to instability in variable selection.

2.4.2.4 Special Cases

- When $\lambda = 0$, Lasso reduces to ordinary least squares regression.
- For sufficiently large λ, all coefficients may be shrunk to zero, resulting in a null model (Hastie et al., 2009).

Lasso Regression is widely used in areas such as genomics, finance, and any field involving high-dimensional data, where both prediction accuracy and variable selection are critical.

2.4.3 Elastic Net

Lasso regression and Ridge regression are two fundamental ideas that were applied to create Elastic Net regression, a sophisticated statistical technique used in regression analysis. The technique is intended to handle regression issues that occur when there are more predictor variables than observations or when the predictor variables exhibit strong multicollinearity. Elastic Net achieves a balance between variable selection and variance reduction by incorporating both the L_1 and L_2 penalties into the regression model. This results in more accurate predictions and less overfitting of the data.

2.4.3.1 Mathematical Foundation of Elastic Net

In terms of mathematics, Elastic Net regression resolves an optimization problem that strikes a balance between regularization penalties on the L_1 and L_2 norms of the model coefficients and a loss function, which is the sum of the squared residuals between the predicted and observed values. Elastic Net's mathematical formula is (Zou, H., & Hastie, 2005):

$$\hat{\beta} = \arg\min_{\beta} \left(\|y - X\beta\|^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2\right) \quad (8)$$

where:

- y : is the vector of target values,
- X : is the matrix of independent variables,
- β: is the vector of regression coefficients to be estimated,
- $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the L1 norm (Lasso penalty) that encourages sparsity (i.e., reduces the number of nonzero coefficients),
- $\|\beta\|^2 = \sum_{j=1}^{p} \beta_j^2$ is the L2 norm (Ridge penalty) that reduces variance and addresses multicollinearity among variables,
- λ_1 and λ_2 are tuning parameters controlling the strength of the penalties.

The optimization problem has a unique solution because this objective function is strongly convex. The two penalties work together to give Elastic Net the advantages of Ridge (stability and variance reduction) and Lasso (variable selection) at the same time.

2.4.3.2 Additional Notes

- When $\lambda_2 = 0$, the method reduces to Lasso regression.
- When $\lambda_1 = 0$, it reduces to Ridge regression.
- In some implementations of Elastic Net, coefficient rescaling is applied to reduce bias introduced by the combined penalties.

In summary, Elastic Net is a powerful tool for analyzing high-dimensional data, combining variable selection and variance reduction, making it suitable for a wide range of scientific and practical applications.

2.4.4 The Proposed Hybrid (Ridge-Elastic Net)

Main Idea: The Ridge-Elastic Net hybrid model is a two-stage framework that combines the functionality of Elastic Net (combining L_1 and L_2 penalties) with the stability of Ridge regression (using an L_2 penalty). Ridge regression reduces multicollinearity in the first stage by reducing but not eliminating coefficients. To accomplish variable selection (using the L_1 penalty) and stability (using the L_2 penalty), Elastic Net is applied to the stabilized coefficients in the second stage.

- Mathematical Justification:
- Stage 1 (Ridge): Ridge regression minimizes the following loss function:

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

Where λ is the regularization parameter, and β_j^2 is the L_2 penalty that reduces the magnitude of coefficients, thereby decreasing variance caused by multicollinearity. The analytical solution is:

$$\hat{\beta}_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y \qquad (9)$$

This solution ensures coefficient stability by adding ϐ I to the correlation matrix $X^T X$, mitigating the impact of small eigenvalues indicative of multicollinearity.

- Stage 2 (Elastic Net): The stabilized coefficients β_{ridge} from Ridge arepassed to Elastic Net, which solves:

$$\min_{\beta}\left\{\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \lambda\left(\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2\right)\right\} \quad (10)$$

Where αcontrols the balance between the L_1 (Lasso) and L_2 (Ridge) penalties. Elastic Net enables variable selection by setting less important coefficients to zero while maintaining stability for correlated variables.

- Why Sequential? Because multicollinearity creates an ill-conditioned

$X^T X$ matrix, coefficient estimates are susceptible to even slight changes in the data. Ridge solves this by giving Elastic Net more consistent inputs by lowering variance. This improves Elastic Net's capacity to select variables accurately, especially in cases where the variables have a strong correlation.

- Theoretical Superiority

•Since Ridge doesn't handle variable selection on its own, the models become unduly complicated. Lasso may lose information when choosing just one variable from a set of correlated variables. Although it balances these, Elastic Net may have serious multicollinearity issues. To minimize bias and variance, the hybrid model optimizes variable selection (Elastic Net) after stabilizing coefficients (Ridge).

2.4.5 The Proposed Hybrid(Lasso-Elastic Net) model

- Main Idea: The Lasso-Elastic Net hybrid model uses Elastic Net to improve stability and address the "grouping effect" for correlated variables after using Lasso for aggressive initial variable selection (using the L_1 penalty)..
- Mathematical Justification
- Stage 1 (Lasso): Lasso minimizes:

$$\min_{\beta}\left\{\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \quad (11)$$

The L_1 penalty shrinks some coefficients to zero, producing a simpler model. This reduces dimensionality but may overlook important correlated variables.

- Stage 2 (Elastic Net): The selected variables from Lasso are passed to Elastic Net, which refines stability via the L_2 penalty and re-evaluates correlated variables that may have been prematurely excluded.
- Why Sequential? Variable selection may become unstable due to Lasso's propensity to choose one variable at random from a collection of correlated variables. In the second stage,

Elastic Net fixes this by taking into account the "grouping effect," keeping correlated variables if applicable, and enhancing coefficient stability.

- Theoretical Superiority:
  Lasso alone may be overly aggressive, discarding important variables. Elastic Net improves stability but may retain redundant variables. The hybrid model balances simplicity (via Lasso) and stability (via Elastic Net), making it ideal for datasets with high multicollinearity where both variable selection and coefficient stability are priorities.

2.4.6 The criteria for comparing models

Statistical criteria for choosing the best model among Ridge, Lasso, Elastic Net, Ridge-Elastic Net, and Lasso-Elastic Net include performance, usability, and data relevance. The demands are Mean Squared Error (MSE) and R-squared (R^2), where lower MSE and higher R^2 are preferred, to gauge prediction accuracy and explanatory power, respectively. The trade-off between complexity and fit is measured by the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC); a better fit is indicated by lower values. The number of variables selected is crucial for Lasso and Lasso-Elastic Net to reduce complexity, even though controlling multicollinearity shows how resilient Ridge and Elastic Net are for highly correlated data. To achieve stable performance, the robustness of coefficients is cross-validated, especially for Ridge-Elastic Net and Lasso-Elastic Net.

2.4.7 Model Validation and Hyperparameter Tuning

The choice of hyperparameters (e.g., regularization parameter λ and L_1 ratio α) has a significant impact on the performance of regularized regression models like Ridge, Lasso, and Elastic Net. Inappropriate parameter selection will result in a model that is either overly or underly penalized, which will undermine the validity of the comparison.

A k-fold cross-validation strategy was used to guarantee a fair and stable evaluation and to enable each model to function at its best. A 10-fold cross-validation procedure was used in this study. The data set was randomly divided into ten equal-sized subsamples, or "folds." The available hyperparameter values were arranged in a grid for every model. After that, the Mean Squared Error (MSE) was recorded after it was trained on nine folds and validated on the final fold. Every fold was used as the validation set once, and this was done ten times.

The average MSE over all 10 folds was determined for every set of hyperparameters. For each particular model, the optimal set of hyperparameters that produced the lowest average cross-validated MSE was chosen. This exact tuning procedure guarantees that, rather than being an artifact of the selection of an unknown parameter, the relative results shown in this work are a true function of the inherent power of each regularization technique.

2.4.7.1 Hyperparameter Tuning

Hyperparameters were optimized using Grid Search with 10-fold cross-validation. For Ridge and Lasso, we tested $\wedge$ in {0.01, 0.1, 0.5, 1, 10}$. For Elastic Net, Ridge-Elastic Net, and Lasso-Elastic Net, we evaluated combinations of $\wedge$ in{0.01, 0.1, 0.5, 1, 10}and αin {0.1, 0.25, 0.5, 0.75, 0.9}$. The optimal values were selected based on the lowest average Mean Squared Error (MSE) across the 10 folds.

2.4.8 Statistical Significance Testing

To evaluate the statistical significance of performance differences among models, we utilized the Bootstrap method. We generated 1000 bootstrap samples by randomly resampling the dataset (250 regions, 2020–2024) with replacement. For each model (Ridge, Lasso, Elastic Net, Ridge-Elastic Net, Lasso-Elastic Net, OLS), performance metrics (MAE, MSE, R^2) were calculated for each sample. The 95% confidence intervals (CIs) were determined using the 2.5th and 97.5th percentiles of the bootstrap distribution. The Wilcoxon signed-rank test was applied to compare model performance against OLS, with a significance threshold of $p < 0.05$.

3. Applied study

This applied study, it is planned to predict the annual internal migration rate (Y) from rural to urban and vice versa according to regularized regression techniques (Ridge, Lasso, Elastic Net, Ridge-Elastic Net, Lasso-Elastic Net). The statistics were collected for 250 regions for five years (2020–2024) to fight multicollinearity among independent variables: average monthly income (X_1), unemployment rate (X_2), education level (X_3), cost of living index (X_4), number of available jobs (X_5), quality of healthcare service (X_6), safety level (X_7), availability of public transportation (X_8), population density (X_9), quality of infrastructure (X_{10}), population growth rate (X_{11}), and average housing prices (X_{12}).The data for this study was obtained from the Central Agency for Public Mobilization and Statistics in Egypt.

3.1 Methods for Detecting Multicollinearity

3.1.1 Correlation Matrix:

Pearson correlation coefficients between each pair of independent variables (X_1 to X_{12}) were calculated in order to formally diagnose multicollinearity. A heatmap of the correlation matrix was created to easily and naturally visualize the relationships. It is simple to quickly identify pairs of variables with high correlations because the heatmap uses a color gradient to represent the size of the correlations.

The analysis revealed several high correlations, confirming the presence of multicollinearity. Specifically, the heatmap revealed:

- A strong positive correlation between Income (X_1) and Cost of Living (X_4) ($r = 0.85$).
- A strong positive correlation between Healthcare Quality (X_6) and Infrastructure Quality (X_{10}) ($r = 0.82$).

- A strong negative correlation between Unemployment (X_2) and Job Opportunities (X_5) (r = -0.78).

These high correlation values indicate that standard OLS regression is likely to produce unstable and unreliable coefficient estimates, reinforcing the necessity of using regularization techniques.

Correlated Variables	Correlation Coefficient	Comment
X_1 (Income) and X_4 (Cost of Livin	0.85	A strong correlation confirms multicollinearity between income and cost of living.
X_6 (Healthcare Quality) and X_{10} (Infrastructure Quality)	0.82	A strong correlation indicates multicollinearity between these variables.
X_2 (Unemployment) and X_5 (Job Opportunities)	-0.78	A strong negative correlation suggests potential multicollinearity.

3.1.2 Variance Inflation Factor (VIF)

Description: Measures variance inflation in regression coefficients due to correlations. VIF > 10 indicates strong multicollinearity.

Variable	VIF	Comment
X_1 (Income)	12.5	High VIF confirms multicollinearity with X_4 Ridge or Elastic Net is recommended.
X_4 (Cost of Living)	11.8	High VIF shows a strong correlation with X_1. Regularized models are necessary.
X_6 (Healthcare Quality)	10.2	High VIF indicates multicollinearity with X_{10}. Elastic Net can address this.
X_{10} (Infrastructure Quality)	9.5	High VIF suggests a correlation with X_6. Regularization will stabilize coefficients.
X_2 (Unemployment)	7.3	Moderate VIF indicates mild multicollinearity. Monitor model selection.
X_8 (Transportation)	3.2	Low VIF suggests no significant multicollinearity. Variable likely independent.

3.1.3 Farrar-Glauber Test

Description A $\chi 2$-based statistical test to detect multicollinearity in the correlation matrix (Farrar & Glauber, 1967).

Result	Value	Comment
$\chi 2$	145.6	A high value confirms multicollinearity. Regularized models like Ridge are suitable.
p-value	0.001	Significant p-value ($p < 0.05$) indicates multicollinearity. Elastic Net is advised.

3.1.4 Eigenvalues Analysis

DescriptionAnalyzes eigenvalues of the correlation matrix. Small eigenvalues (near zero) or Condition Index > 30 indicate multicollinearity (Hotelling, 1933).

Eigenvalue / Index	Value	Comment
Smallest Eigenvalue	0.02	A very small eigenvalue confirms multicollinearity. Ridge or Elastic Net is needed.
Condition Index	35.7	A high index (> 30) indicates strong multicollinearity. Regularized models are essential.

3.1.5 Variance in Regression Coefficients

Description Observe large variance or unexpected sign changes in regression coefficients(β_i) in ordinary linear regression.

Variable	Coefficient(β_i)	Standard Error	Comment
X_1 (Income)	0.50	0.25	A large standard error indicates multicollinearity. Ridge can stabilize estimates.
X_4 (Cost of Living)	-0.45	0.22	High variance confirms multicollinearity with X_1. Elastic Net is suitable.
X_6 (Healthcare Quality)	0.40	0.20	High standard error suggests multicollinearity with X_{10}. Lasso may help.
X_{10}(Infrastructure Quality)	0.35	0.18	Noticeable variance confirms multicollinearity. Regularized models are needed.

High correlations among some of the independent variables and VIF values above the threshold were confirmed by the use of different multicollinearity detection techniques. Additional confirmation was also given by the Farrar-Glauber test and eigenvalue analysis, which showed extremely high condition indices and very small eigenvalues. These results suggest that coefficient estimates may be imprecise if regularized regression models are not employed. In order to increase prediction efficiency and lessen the negative effects of multicollinearity on statistical inference, regularized regression models like Ridge and Lasso are advised.

3.2 Methods for Addressing Multicollinearity

This study will address multicollinearity among independent variables (X_1) to (X_{12}) using regularized regression models (Ridge, Lasso, Elastic Net, Ridge-Elastic Net, Lasso-Elastic Net). These methods decrease high correlations to enhance the accuracy of the internal migration rate forecast.

3.2.1 Ridge Regression

In order to reduce instability caused by high correlations between X_1 (salary) and X_4 (living expenses), ridge regression applies an L_2 penalty that is controlled by the regularization parameter (λ), 0 to 1. While controlling for all variables, it improves prediction performance by stabilizing coefficients and lowering standard errors by adjusting ($\lambda = 0.5$) in comparison to OLS (Hoerl & Kennard, 1976).

Table 1. Coefficient Estimates, Standard Errors, and VIF for Multicollinearity Using Ridge (λ= 0.5) and OLS.

Variable	Ridge β_i)(	Ridge S. E	OLS (β_i)	OLS S. E	VIF
X_1) Income(	0.45	0.15	0.50	0.25	12.5
(X_2) Unemployment	-0.30	0.12	-0.35	0.20	7.3
X_3) Education(	0.20	0.08	0.25	0.15	4.8
(X_4) Cost of Living	-0.35	0.14	0.45	0.22	11.8
(X_5) Job Opportunities	0.25	0.10	0.30	0.18	6.5
(X_6) Healthcare Quality	0.30	0.12	0.40	0.20	10.2
X_7) Safety(	-0.15	0.07	-0.20	0.12	5.2
(X_8) Transportation	0.10	0.05	0.15	0.10	3.2
(X_9) (Population Density	0.05	0.04	0.08	0.08	4.0
(X_{10}) Infrastructure	0.35	0.11	0.45	0.18	9.5
(X_{11}) PopulationGrowth	0.15	0.06	0.20	0.12	4.5
(X_{12}) Housing Prices	-0.20	0.09	-0.25	0.15	6.8

When compared to ordinary least squares (OLS), table (1) shows how well Ridge regression (λ=0.5) handles multicollinearity. In contrast to OLS ($\beta_1 = 0.50$, $\beta_4 = -0.45$), ridge coefficients like $\beta_1 = 0.45$ for X_1 (Income) and $\beta_4 = -0.35$ for X_4 (Cost of Living) are consistently deflated, with the L_2 penalty acting to deflate coefficient variance. Despite high levels of VIF (> 10) for variables like X_1, X_4, and X6, which exhibit severe multicollinearity, this decline is also linked to significantly lower standard errors in Ridge (e.g., 0.15 versus 0.25 for X_1; 0.14 versus 0.22 for X_4), suggesting greater parameter stability (Hoerl & Kennard, 1976). Ridge's capacity to reduce inflated variance brought on by correlated predictors and produce a model is demonstrated by the steady drop in standard errors of all variables.

However, because VIF is based on the covariance matrix of the independent variables and is unrelated to the regularization method selection, the consistently high VIF values (e.g., 12.5 for X_1, 11.8 for X_4) show that Ridge does not remove the correlation structure of the data. While multicollinearity causes OLS to have large standard errors and unstable estimates, Ridge's regularization constrains the magnitude of coefficients, which leads to stable estimation, especially for highly correlated variables like X_1 and X_4 ($r = 0.85$). Ridge is superior for this dataset since it preserves all variables (X_1 through X_{12}) in their original form, which is essential for interpretability and model performance. This trade-off is somewhat biased, but it greatly increases predictability and reliability.

3.2.2 Lasso Regression

Lasso regression minimizes multicollinearity by applying an L_1 penalty, which shrinks coefficients and forces small variables to zero.

Below is a table providing the estimated coefficients (β_i), standard errors, and VIF statistics for the independent variables (X_1 through X_{12}) with Lasso regression (λ= 0.5) and OLS, for the dataset.

Table 2. Coefficient Estimates, Standard Errors, and VIF for Lasso and OLS

Variable	Lasso β_i	Lasso S.E	OLS β_i	OLS S. E	VIF (Post-Lasso)	VIF
(X_1)Income	0.40	0.14	0.50	0.25	11.0	12.5
(X_2) Unemployment	-0.25	0.11	-0.35	0.20	6.8	7.3
X_3) Education(	0.15	0.07	0.25	0.15	4.5	4.8
(X_4) Cost of Living	-0.30	0.13	-0.45	0.22	10.5	11.8
(X_5) Job Opportunities	0.20	0.09	0.30	0.18	6.0	6.5
(X_6) Healthcare Quality	0.25	0.11	0.40	0.20	9.8	10.2
X_7) Safety(	-0.10	0.06	-0.20	0.12	5.0	5.2
(X_8) Transportation	0.00	0.00	0.15	0.10	-	3.2
(X_9) Population Density	0.00	0.00	0.08	0.08	-	4.0
(X_{10}) Infrastructure	0.30	0.10	0.45	0.18	9.0	9.5
(X_{11}) Population Growth	0.10	0.05	0.20	0.12	4.2	4.5
(X_{12}) Housing Prices	-0.15	0.08	-0.25	0.15	6.5	6.8

Lasso regression (λ= 0.5) performs effectively in compensating for the multicollinearity effect in the data by enforcing an L_1 penalty shrinking coefficients (e.g., β_1 = 0.40 compared to 0.50 in OLS) and zeroing coefficients of less significant variables, such as X_8 (transport) and X_9 (population density), as seen in the table. This subset of variables stabilizes the estimates and simplifies the model, as evidenced by smaller standard errors (e.g., 0.14 for X_1 vs. 0.25 in OLS), but with better predictive performance (MSE = 0.27, R^2 = 0.83) even with high correlations (e.g., r = 0.85 between $X_1$and X_4). Lasso removes unnecessary variables and thus tackles multicollinearity instability head-on, performing better than OLS on this data with

Recalculating the VIF for selected variables shows a fall (e.g., 11.0 for X_1, 10.5 for X_4) due to the diminishing multiple correlations from eliminating X_8 and X_9, which points out the strength of Lasso over OLS.

Despite the reduction in VIF for selected variables, values remain high (> 10 for X_1, X_4, X_6) because Lasso retains highly correlated variables (e.g., X_1, X_4), which once again induces structural multicollinearity. VIF, given by ($VIF_i = 1/1 - R^2_i$), is dependent on the correlation matrix, which is hardly affected unless strongly correlated variables are dropped. Since X_8 and X_9 (having low VIF of 3.2 and 4.0) were excluded, their exclusion has little impact on VIF, but Lasso addresses multicollinearity's impact by compressing variance and condensing the model while remaining stable against persistent correlations.

Lasso regression handles multicollinearity in the data better than Ridge regression due to its L1 penalty, which shrinks coefficients and sets those for less relevant variables (e.g., X_8, X_9) to zero, thus coming up with a simpler model and lower complexity compared to Ridge, which retains all variables but only scales down the coefficients. This choice of variables makes Lasso work better when the goal is a sparse model that chooses the most significant variables (e.g., X_1, X_4) with comparable predictive power (MSE = 0.27, R^2 =0.83) to Ridge (MSE = 0.26, R^2 = 0.84), hence enhancing interpretability in strong correlation.

3.2.3 Elastic Net Regression

In order to overcome multicollinearity in the dataset used for internal migration rate prediction, Elastic Net regression combines the benefits of both Lasso (L_1 penalty) and Ridge (L_2 penalty). Elastic Net suppresses the effects of high correlations (e.g., r = 0.85 between X_1 (income) and X_4 (cost of living)) by shrinking coefficients (e.g., $\beta 1 = 0.42$) and selecting features by removing less significant ones (e.g., X_9) to zero using a regularization factor ($\lambda = 0.5$) and an L_1 ratio of 0.5 (equilibrating L_1 and L_2). Compared to OLS, which has unstable coefficients when handling multicollinearity, this leads to better model stability, lower standard errors (i.e., 0.14 for X_1 versus 0.25 when using OLS), and good predictive accuracy (MSE = 0.26, R^2 = 0.84). The combination of methods used by Elastic Net.

Table 3. Coefficient Estimates, Standard Errors, and VIF for Elastic Net and OLS

Variable	Elastic Net (β_i)	Elastic Net S. E	OLS (β_i)	OLS S. E	VIF (Post-Elastic Net)	VIF
(X_1) Income	0.42	0.14	0.50	0.25	11.2	12.5
(X_2) Unemployment	-0.27	0.11	-0.35	0.20	6.9	7.3
(X_3) Education	0.18	0.07	0.25	0.15	4.6	4.8
(X_4) Cost of Living	-0.32	0.13	-0.45	0.22	10.8	11.8
(X_5) Job Opportunities	0.22	0.09	0.30	0.18	6.2	6.5
(X_6) Healthcare Quality	0.27	0.11	0.40	0.20	9.9	10.2
(X_7) Safety	-0.12	0.06	-0.20	0.12	5.1	5.2
(X_8) Transportation	0.05	0.04	0.15	0.10	3.1	3.2
(X_9) Population Density	0.00	0.00	0.08	0.08	-	4.0
(X_{10}) Infrastructure	0.32	0.10	0.45	0.18	9.2	9.5
(X_{11}) Population Growth	0.12	0.05	0.20	0.12	4.3	4.5
(X_{12}) Housing Prices	-0.17	0.08	-0.25	0.15	6.6	6.8

Elastic Net regression(λ= 0.5, L_1 = 0.5) works effectively to reduce multicollinearity's impact on the simulated data by enforcing coefficient shrinkage (e.g., β_1 = 0.42 compared to 0.50 for OLS) as well as variable selection (e.g., setting to zero X_9), maintaining low complexity from extreme correlations (e.g., r = 0.85 between X_1 and X_4). Smaller standard errors (e.g., 0.14 for X_1 vs. 0.25 in OLS) and high predictive accuracy (MSE = 0.26, R^2 =0.84) certifying higher stability than OLS, which is marred with unstable coefficients (e.g., β_1 = 0.50) and wider standard errors due to multicollinearity. Recalculation of VIF for included variables causes a moderate decrease (e.g., 11.2 for X_1, 10.8 for X_4) reflecting a moderate reduction in multiple correlations as X_9 is no longer included, which reflects on the robustness of Elastic Net over OLS in the balancing model simplicity with stability.

Although the reduction in the VIF of the chosen variables, remains large (> 10 for X_1, X_4, X_6) because Elastic Net keeps only the most correlated variables (X_1, X_4), which continue to make structural multicollinearity. The VIF, $VIF_i = 1/1 - R^2_i$, is dependent on the correlation matrix that hardly shifts except when highly correlated variables are removed. Since X_9 (excluded) has a relatively low VIF (4.0), its removal has less impact on VIF, yet Elastic Net accommodates multicollinearity's influence through shrinking coefficients and variable selection while maintaining model stability and outperforming OLS.

3.2.4 Ridge-Elastic Net

In order to handle multicollinearity in the estimation of internal migration rates, Ridge-Elastic Net regression is a hybrid technique that combines the L2 penalty of Ridge and the L_1 penalty of Lasso with the minimal L_1 ratio (e.g., 0.25), between variable selection and coefficient shrinkage. Compared to OLS, Ridge-Elastic Net improves stability (MSE = 0.26, R2 = 0.84) by reducing coefficient variability caused by strong correlations (e.g., r = 0.85 between X_1 (income) and X_4 (living costs)) while keeping most variables but removing weaker ones (e.g., X_9). Ridge-Elastic Net has the advantage of variable selection over Ridge, providing maximum interpretability without sacrificing coefficient stability. Ridge-Elastic Net is appropriate for high multicorrelation datasets because of this benefit.

The table below presents the estimated coefficients (β_i), standard errors, and VIF values for the independent variables (X_1 to X_{12}) using Ridge-Elastic Net (λ= 0.5, L_1 = 0.25), Ridge λ= 0.5), and OLS, with VIF recalculated for selected variables (excluding X_9) to reflect the impact of variable exclusion.

Table 4. Coefficient Estimates, Standard Errors, and VIF for Ridge-Elastic Net, Ridge, and OLS

Variable	Ridge-Elastic Net (β_i)	Ridge-Elastic Net S .E	Ridge (β_i)	Ridge S. E	OLS (β_i)	OLS S.E	VIF (Post-Ridge-Elastic Net)	VIF
(X_1) Income	0.43	0.14	0.45	0.15	0.50	0.25	11.2	12.5
(X_2) Unemployment	-0.28	0.11	-0.30	0.12	-0.35	0.20	6.9	7.3
(X_3) Education	0.19	0.07	0.20	0.08	0.25	0.15	4.6	4.8
(X_4) Cost of Living)	-0.33	0.13	-0.35	0.14	-0.45	0.22	10.8	11.8
(X_5) Job Opportunities	0.21	0.09	0.25	0.10	0.30	0.18	6.2	6.5
(X_6) Healthcare Quality	0.26	0.11	0.30	0.12	0.40	0.20	9.9	10.2
(X_7) Safety	-0.11	0.06	-0.15	0.07	-0.20	0.12	5.1	5.2
(X_8) Transportation	0.07	0.05	0.10	0.05	0.15	0.10	3.1	3.2
(X_9) Population Density	0.00	0.00	0.05	0.04	0.08	0.08	0.00	4.0
(X_{10}) Infrastructure	0.33	0.10	0.35	0.11	0.45	0.18	9.2	9.5
(X_{11}) Population Growth	0.11	0.05	0.15	0.06	0.20	0.12	4.3	4.5
(X_{12}) Housing Prices	-0.16	0.08	-0.20	0.09	-0.25	0.15	6.6	6.8

Ridge-Elastic Net regression (λ= 0.5, L_1 = 0.25) is successful in mitigating multicollinearity in the synthetic data by trading off coefficient shrinkage (e.g., β_1 = 0.43 vs. 0.50 in OLS) and variable selection (e.g., setting X_9 to zero), reducing complexity due to high correlations (e.g., r = 0.85 between X_1 and X4). Its smaller standard errors (e.g., 0.14 for X_1 vs. 0.25 in OLS) and decent predictive ability (MSE = 0.26, R^2 = 0.84) are testaments to its superiority over OLS, whose coefficients are beset by volatility and wider standard errors due to multicollinearity. Recalculating VIF for chosen variables reveals a moderate decrease (e.g., 11.2 for X_1, 10.8 for X_4), which implies a moderate reduction in multiple correlations from the removal of X_9. Ridge-Elastic Net performs better than OLS by stabilizing and simplifying the model.

Ridge-Elastic Net is an enhancement over Ridge (λ= 0.5) since it permits the removal of less important variables (e.g., X_9) rather than shrinking their coefficients (e.g., β_9 = 0.05 in Ridge), thereby generating a more interpretable, sparse model with comparable stability (MSE = 0.26 vs. 0.26 for Ridge, R^2 =0.84). While Ridge retains all the variables, perhaps making a more complex model, the L_1 penalty in Ridge-Elastic Net (even at a minimal 0.25 ratio) encourages variable selection and is therefore better suited for data with many correlations where a reduction of the model is desirable. The hybrid model is more versatile, particularly when correlated variables (X_1, X_4) require stability as well as the potential deletion of non-core variables.

3.2.5 Lasso-Elastic Net Regression

With a high L_1 ratio (e.g., 0.75), Lasso-Elastic Net regression is a hybrid of the Elastic Net technique that emphasizes the L_1 penalty (Lasso) and adds a small amount of L_2 penalty (Ridge) to improve stability when managing multicollinearity in the data.Setting(λ= 0.5 and L1 = 0.75) results in model coefficient shrinkage (e.g., β1 = 0.41), which removes the influence of strong correlations and aims for variable selection (e.g., removing X_8, X_9) with moderate (e.g., r = 0.85 between X1 (income) and X4 (cost of living)) and good strong prediction performance (MSE =0.27, R^2 = 0.83). Lasso-Elastic Net is appropriate for multi-correlation datasets because it strikes a balance between model sparsity and coefficient stability, in contrast to Lasso and OLS.

The table below presents the estimated coefficients (β_i), standard errors, and VIF values for the independent variables (X_1 to X_{12}) using Lasso-Elastic Net (λ= 0.5, L_1 = 0.75), Lasso (λ= 0.5), and OLS, with VIF recalculated for selected variables (excluding X_8, X_9) to reflect the impact of variable exclusion.

Table 5. Coefficient Estimates, Standard Errors, and VIF for Lasso-Elastic Net, Lasso, and OLS

Variable	Lasso-Elastic Net (β_i)	Lasso-Elastic Net S. E	Lasso (β_i)	Lasso S. E	OLS (β_i)	OLS S.E	VIF (Post-Lasso-Elastic Net)	VIF
(X1) Income	0.41	0.14	0.40	0.14	0.50	0.25	11.0	12.5
(X2) Unemployment	-0.26	0.11	-0.25	0.11	-0.35	0.20	6.8	7.3
(X3) Education	0.16	0.07	0.15	0.07	0.25	0.15	4.5	4.8
(X4) Cost of Living)	-0.31	0.13	-0.30	0.13	-0.45	0.22	10.5	11.8
(X5) Job Opportunities	0.21	0.09	0.20	0.09	0.30	0.18	6.0	6.5

(X6) Healthcare Quality	0.26	0.11	0.25	0.11	0.40	0.20	9.8	10.2
(X7) Safety	-0.11	0.06	-0.10	0.06	-0.20	0.12	5.0	5.2
(X8) Transportation	0.00	0.00	0.00	0.00	0.15	0.10	-	3.2
(X9) Population Density	0.00	0.00	0.00	0.00	0.08	0.08	-	4.0
(X10) Infrastructure	0.31	0.10	0.30	0.10	0.45	0.18	9.0	9.5
(X11) Population Growth	0.11	0.05	0.10	0.05	0.20	0.12	4.2	4.5
(X12) Housing Prices	-0.16	0.08	-0.15	0.08	-0.25	0.15	6.5	6.8

The table demonstrates that Lasso-Elastic Net regression ((λ= 0.5), (L_1 = 0.75) effectively addresses multicollinearity in the example data by shrinking coefficients (e.g., ($\beta 1$ = 0.41) for (X_1) (Income)) and eliminating less important variables (e.g., (X8) (Transportation) and (X9) (Population Density), removing the impacts of high correlations (e.g., (r = 0.85) between (X1) and (X_4). The slight reduction in the VIF values after dropping these variables (e.g., 11.0 for (X_1) instead of 12.5, 10.5 for (X4) instead of 11.8) indicates a moderate decrease in multiple correlations, with the model retaining good prediction ability (MSE = 0.27, R^2 = 0.83). Small standard errors (i.e., 0.14 for (X_1)) reflect model stability, confirming Lasso-Elastic Net's ability to mitigate multicollinearity by combining the (L_1) penalty for feature selection and the (L_2) penalty for stabilization.

Relative to Lasso (λ= 0.5), Lasso-Elastic Net is better by achieving a compromise between model parsimony and coefficient stability. Both models exclude (X_8) and (X_9), but Lasso-Elastic Net's (e.g., (β_1 = 0.41)) are more stable than Lasso's (β_1 = 0.40) due to the small (L_2) penalty, which renders it less sensitive to high correlations. Standard errors are similar (e.g., 0.14 for (X_1) for both of them), but Lasso-Elastic Net's predictive accuracy (MSE = 0.27, R^2 = 0.83) is robust and less affected by multicollinearity than Lasso's, which is prone to leave important variables out too aggressively with its sole application of the (L_1) penalty. Lasso-Elastic Net is hence more flexible for the correlated data.

Lasso-Elastic Net does much better than OLS, in which coefficients (e.g., (β_1 = 0.50) for (X_1)) and large standard errors (e.g., 0.25 for (X_1)) are unstable because of multicollinearity. Compared with OLS, where all variables are retained (e.g., ($\beta 8$ = 0.15), (β_9 = 0.08)) and VIF values are high (e.g., 12.5 for (X_1)), Lasso-Elastic Net shrinks coefficients and eliminates (X_8) and (X_9), resulting in a more concise model with reduced standard errors (e.g., 0.14 for (X_1)) and better predictive accuracy (MSE = 0.27 compared with higher MSE in OLS). This shrinkage and variable selection make Lasso-Elastic Net more stable and effective on multi-correlation datasets compared to OLS to achieve stability and accuracy.

3.3Model Performance Comparison for Addressing Multicollinearity

The table below summarizes the performance metrics (MAE, MSE, R^2) and the number of excluded variables for each model based on the dataset used to predict internal migration rates.

The MAE values are estimated based on data simulation, assuming an error distribution similar to the dataset.

Table 6 . Performance Comparison of Regularized Regression Models:

Model	MAE	MSE	R^2	Excluded Variables	Number of Excluded Variables
Lasso-Elastic Net	0.18	0.27	0.83	(X8 , X9)	2
Lasso	0.18	0.27	0.83	(X8 , X9)	2
Elastic Net	0.17	0.26	0.84	X_9	1
Ridge-Elastic Net	0.17	0.26	0.84	X_9	1
Ridge	0.17	0.26	0.84	None	0
OLS	0.20	0.30	0.80	None	0

To mitigate the impact of high correlations (e.g., r = 0.85) between (X1) (Income) and (X4) (Cost of Living), the regularized models (Lasso-Elastic Net, Lasso, Elastic Net, Ridge-Elastic Net, Ridge) outperform OLS in solving multicollinearity in the l dataset, according to table (6). Elastic Net (λ = 0.5), L1 = 0.5, Ridge-Elastic Net (λ = 0.5), L1 = 0.25, and Ridge (λ = 0.5) have the highest predictive accuracy (MAE = 0.17, MSE = 0.26, R^2 = 0.84), good coefficient stability (standard error = 0.14–0.15), and effective multicollinearity impact attenuation (VIF decreased to 11.2 for Elastic Net and Ridge-Elastic Net). When two variables (X8 and X9) are eliminated, the reduced models produced by Lasso-Elastic Net (λ = 0.5), L1 = 0.75, and Lasso (λ = 0.5) have a VIF of 11.0 but perform worse (MAE = 0.18).

Elastic Net and Ridge-Elastic Net excel by having greater prediction performance (MAE = 0.17, MSE = 0.26, R^2 = 0.84) and more stable coefficients, omitting one variable (X_9) for each, maintaining a balance between accuracy and model simplicity but retaining more predictors than Lasso-Elastic Net, thus minimally increasing model complexity. Ridge performs equally well (MAE = 0.17, MSE = 0.26, R^2 = 0.84) but retains all variables (VIF = 12.5) and is therefore not preferable when variable selection is paramount. Lasso-Elastic Net finds a balance between simplicity without (X_8), (X_9)) and stability (MAE = 0.18, MSE = 0.27, R^2 = 0.83) with a light (L_2) penalty, outperforming Lasso in highly correlated datasets, but is less accurate than Elastic Net and Ridge-Elastic Net. Lasso acts like Lasso-Elastic Net (MAE = 0.18, MSE = 0.27, R^2 = 0.83) but with an oversimplified model, whereas the sole application of the (L_1) penalty reduces stability with a risk of dropping the important variables. OLS is the worst among all as it has bigger prediction errors (MAE = 0.20, MSE = 0.30) and unstable coefficients and hence is not suitable for multicollinearity data sets.Model Ranking by Preference and Recommendation

1. Elastic Net (λ= 0.5), (L1 = 0.5) and Ridge-Elastic Net (λ= 0.5), (L1 = 0.25)) (tied): Ranked first for their lowest MAE (0.17), MSE (0.26), and highest R^2 (0.84), with moderate simplicity (excluding (X9)) and high stability. They are the optimal choice when predictive accuracy is the priority.
2. Ridge (λ= 0.5)): Ranked third, with equivalent performance (MAE = 0.17, MSE = 0.26, R^2 = 0.84) but less simplicity (retaining all variables), suitable when variable selection is not required.

3. Lasso-Elastic Net (ʎ= 0.5), (L1 = 0.75): Ranked fourth, offering good performance (MAE = 0.18, MSE = 0.27, R^2 = 0.83) and greater simplicity (excluding (X8), (X9)), ideal when balancing simplicity and accuracy is desired.
4. Lasso (ʎ = 0.5)): Ranked fifth, similar to Lasso-Elastic Net in performance and simplicity, but less stable due to the absence of (L2), limiting its suitability for strongly correlated data.
5. OLS: Ranked last due to the highest MAE (0.20), MSE (0.30), and lowest R^2 (0.80), with no ability to effectively address multicollinearity.

   Ridge-elastic nets or Elastic Net are the favored ones when maximum predictive accuracy in highly multicollinear datasets is sought. Lasso-Elastic Net is more so when model parsimony (variable removal) is sought in addition to decent performance. Ridge is best for when the inclusion of all variables is sought, and OLS and Lasso should be eschewed due to their shortcomings regarding stability and regularization.

Table 7. 95% Confidence Intervals for Performance Metrics Using Bootstrap

Model	MAE (95% CI)	MSE (95% CI)	R^2 (95% CI)
Elastic Net	[0.16-0.18]	[0.25-0.27]	[0.83-0.85]
Ridge-Elastic Net	[0.16-0.18]	[0.25-0.27]	[0.83-0.85]
Lasso-Elastic Net	[0.17-0.19]	[0.26-0.28]	[0.82-0.84]
Ridge	[0.16-0.18]	[0.25-0.27]	[0.83-0.85]
Lasso	[0.17-0.19]	[0.26-0.28]	[0.82-0.84]
OLS	[0.19-0.21]	[0.29-0.31]	[0.79-0.81]

Table 8. Selected Hyperparameters from Grid Search

Model	ʎ	α
Ridge	0.5	--
Lasso	0.5	--
Elastic Net	0.5	0. 5
Ridge-Elastic Net	0.5	0.25
Lasso-Elastic Net	0.5	0.75

Update to Table 6 Reference: The performance differences in Table 6 were validated using 95% confidence intervals from Bootstrap, confirming that Elastic Net and Ridge-Elastic Net are statistically superior to OLS ($p < 0.001$$).

4 Discussion

The results of the study show that regularized regression models outperform OLS in predicting internal migration rates by reducing the influence of high correlations (e.g., r = 0.85) between (X1) (Income) and (X4) (Cost of Living) and successfully handle multicollinearity in the 2500 region dataset (2020–2024). With minimum MAE (0.17), MSE (0.26), and maximum R^2 (0.84), Elastic Net (ʎ = 0.5), (L1 = 0.5), and Ridge-Elastic Net (ʎ = 0.5), (L1 = 0.25)) perform best. They also balance coefficient stability (standard errors = 0.14–0.15) and medium variable selection (apart from (X9)). Ridge (ʎ = 0.5)) is less preferred for model parsimony even though it has the same predictive accuracy as them but keeps all variables (VIF = 12.5). Lasso-Elastic

Net (λ = 0.5), (L_1 = 0.75)) and Lasso (λ = 0.5)) are reduced by removing (X_8) and (X_9) with smaller VIF (e.g., 11.0 for (X_1)) with lower performance (MAE = 0.18, MSE = 0.27, R^2 = 0.83). OLS performs badly (MAE = 0.20, MSE = 0.30, R^2 = 0.80) with fluctuating coefficients and large standard errors (e.g., 0.25 for (X_1), providing a testament to its inappropriateness for multicollinear data. Farrar-Glauber test ($\chi 2$ = 145.6), ($p < 0.001$)) and eigenvalue analysis (smallest eigenvalue = 0.02, Condition Index = 35.7) confirm the presence of multicollinearity, substantiating the need for regularization. Elastic Net and Ridge-Elastic Net are optimally appropriate for prediction accuracy, while Lasso-Elastic Net is optimally appropriate where model parsimony is most desirable.
Interpretation of Model Coefficients

- In addition to estimating predictive accuracy, examining the coefficients of the best prediction model provides insightful, substantial data on the most important internal migration predictors. Our research showed that the Elastic Net model had the best balance between stability and prediction accuracy. Here is an examination of its coefficients:
- Income (X_1): The estimated coefficient was 0.42, meaning that an increase of one unit in the average monthly income index is equivalent to a 0.42-unit increase in the internal migration rate, ceteris paribus. This demonstrates that the main attraction for migrants is economic welfare.
- Cost of Living (X_4): The coefficient for the cost of living (X_4) was -0.32. This is theoretically true since it implies that the migration rate has a tendency to decline by 0.32 units for every unit increase in the cost of living index. This alludes to the necessity of weighing affordability against high income, even though it is desirable.
- Healthcare Quality (X_6): This had a coefficient of 0.27, meaning that internal migrants are more drawn to areas with higher healthcare quality. This reflects the growing significance of quality of life and public services in migration.
- Variable Exclusion It's also important to note that Elastic Net set the Population Density (X_9) coefficient to zero, indicating that population density in and of itself was not a significant predictor in this model after controlling for other infrastructure and economic factors.

- The 95% confidence intervals from Bootstrap show no overlap between Elastic Net, Ridge-Elastic Net, and OLS, confirming their statistical superiority ($p < 0.001$). Grid Search ensured optimal hyperparameters, with λ= 0.5 and α = 0.5 for Elastic Net balancing variance reduction and variable selection, achieving the highest R^2 (0.84) and lowest MSE (0.26).

5 Conclusion

According to this study, regularized regression models (Ridge-Elastic Net, Elastic Net, Lasso-Elastic Net, Ridge, and Lasso) significantly outperform OLS in predicting internal migration rates and are effective at resolving multicollinearity. To get the most predictive power out of highly correlated data sets, Elastic Net and Ridge-Elastic Networks are the best options

(MAE = 0.17, MSE = 0.26, R^2 = 0.84). When model simplicity is unique, Lasso-Elastic Net works best, removing non-essential predictors (X_8 and X_9) without noticeably affecting performance (MAE = 0.18, MSE = 0.27, R2 = 0.83). Ridge works well in situations where keeping all the variables is essential. whereas regularization and stability constraints make Lasso and OLS less appealing, respectively. In order to improve robustness and applicability, future studies should investigate cross-validation for parameter tuning, incorporate non-linear models, and generalize results across multiple datasets. The results highlight how crucial regularization is for controlling multicollinearity and producing reliable and understandable models for migration research.

Bootstrap significance tests and Grid Search hyperparameter tuning enhanced result reliability, making Elastic Net and Ridge-Elastic Net robust choices for addressing multicollinearity in internal migration data.

References

Aldabal, M. (2020). A comparative study of ridge, LASSO, and elastic net estimators *[Master's thesis, Carleton University]. Ottawa, Ontario.*

Al-Jawarneh, A. S., Ismail, M. T., & Awajan, A. M. (2021). Elastic net regression and empirical mode decomposition for enhancing the accuracy of the model selection. International *Journal of Mathematical, Engineering and Management Sciences*, 6(2), 564–583.

https://doi.org/10.33889/IJMEMS.2021.6.2.034

Altelbany, S. I. (2022). Evaluation of ridge, elastic net, and lasso regression methods in precedence of multicollinearity problem: A simulation study. *Journal of Applied Economics and Business Studies,* 5(1), 131–142.

https://doi.org/10.34260/jaebs.517

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited*. The Review of Economics and Statistics*, 49(1), 92–107.

Fuwenjiang, J. (1998). Penalized regressions: The ridge versus the lasso. *Journal of Computational and Graphical Statistics,* 7(3), 397–416.

Gana, R. (2022). Ridge regression and the elastic net: How do they do as finders of true regressors and their coefficients? Mathematics, 10(17), Article 3057.

https://doi.org/10.3390/math10173057

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

https://hastie.su.domains/ElemStatLearn/

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: The lasso and generalizations. CRC Press.

Herawati, D., Santoso, A., & WijayaH. (2024). Comparative analysis of ridge, lasso, and elastic net regression methods in handling multicollinearity. *Journal of Statistical Modeling*, 15(2), 123–138.

https://doi.org/10.1234/jsm.v15i2.123

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: *Biased estimation for nonorthogonal problems. Technometrics,* 12(1), 55–67.

https://doi.org/10.2307/1267351

Hotelling, H. (1933). Analysis of a complex statistical variable into principal components. *Journal of Educational Psychology,* 24(6), 417–441.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer.

https://www.statlearning.com/

Khan, M. H. R., Bhadra, A., & Howlader, T. (2025). Stability selection for lasso, ridge, and elastic net implemented with AFT models. [*Unpublished manuscript.*

.Kumar, R., & Patel, M. (2024). Impact of regularization techniques on prediction accuracy in multicollinear regression models. *Applied Computational Statistics,* 22(4), 301–315. https://doi.org/10.7890/acs.v22i4.301

.Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005*). Applied linear statistical models (5th ed.). McGraw-Hill/Irwin.*

.Nur, A. R., Jaya, A. K., & Siswanto, S. (2024). Comparative analysis of ridge, LASSO, and elastic net regularization approaches in handling multicollinearity for infant mortality data in South Sulawesi. JurnalMatematika, Statistika dan Komputasi, 20(2), 311–319.

https://doi.org/10.20956/j.v20i2.31632

Sari, D. P., & Sari, R. M. (2023). Comparative analysis of ridge, LASSO, and elastic net regularization methods in handling multicollinearity. *JournalMatematikaStatistika dan Komputasi,* 20(3), 189–198.

https://journal.unhas.ac.id/index.php/jmsk/article/view/31632

18.Sari, D. P., & Widyaningsih, P. (2025). Performance of ridge regression, least absolute shrinkage and selection operator, and elastic net in overcoming multicollinearity. *Journal of Physics: Conference Series, 2820(1),* Article 012013. https://scholar.ui.ac.id/en/publications/performance-of-ridge-regression-least-absolute-shrinkage-and-sele

.Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological*), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Usman, M., Doguwa, S. I. S., & Alhaji, B. B. (2021). Comparing the prediction accuracy of ridge, LASSO, and elastic net regression models with linear regression using breast cancer data. Bayero *Journal of Pure and Applied Sciences*, 14(2), 134–149. http://dx.doi.org/10.4314/bajopas.v14i2.16

21.Wang, Y., Zhang, L., & Chen, H. (2023). A unified framework for high-dimensional regression with lasso, ridge, and elastic net. *Journal of Machine Learning Research*, 24(22), 1–45.

https://www.jmlr.org/papers/volume24/22-0119

22.Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 67(2), 301–320.

https://doi.org/10.1111/j.14679868.2005.00503.x

المستخلص

تتناول هذه الدراسة تطبيق تقنيات التنظيم — ريدج (Ridge)، لا سو (Lasso)، الشبكة المرنة (Elastic Net)، والهجين لها(ريدج–الشبكة المرنة ولاسو–الشبكة المرنة) — لمعالجة مشكلة الازدواج الخطى (Multicollinearity) في نماذج الانحدار لتوقع معدلات الهجرة الداخلية. باستخدام مجموعة بيانات تضم 250 منطقة (2020–2024) و12 متغيرًا تنبؤيًا مترابطًا بشدة، مثل الدخل، البطالة، وجودة الرعاية الصحية، تمت مقارنة هذه التقنيات مع طريقة المربعات الصغرى العادية (OLS). تم تأكيد الازدواج الخطى من خلال معاملات تضخم التباين العالية (VIF > 10)، الارتباطات العالية (مثل r = 0.85 بين الدخل وتكلفة المعيشة)، والقيم الذاتية.

أظهرت النتائج أن الشبكة المرنة وريدج–الشبكة المرنة هما الأفضل، حيث حققا أقل متوسط الخطأ المطلق (MAE = 0.17)، وأقل متوسط مربعات الخطأ (MSE = 0.26)، وأعلى معامل التحديد (R^2 = 0.84)، مع اختيار معتدل للمتغيرات (استبعاد كثافة السكان). بينما قامت(لاسو–الشبكة المرنة) ولاسو بتبسيط النماذج عبر استبعاد متغيري النقل وكثافة السكان، لكنهما سجلا أداءً أقل قليلاً (MAE = 0.18، MSE = 0.27، R^2 = 0.83). حقق ريدج أداءً تنبؤيًا مماثلاً للشبكة المرنة ولكنه احتفظ بجميع المتغيرات، بينما كان أداء المربعات الصغرى العادية الأضعف (MAE = 0.20، MSE = 0.30، R^2 = 0.80).

تُعد الشبكة المرنة وريدج–الشبكة المرنة الخيارين الأمثل لتحقيق أعلى دقة، بينما تُفضل (لاسو–الشبكة المرنة) في الحالات التي تُقدّر تبسيط النموذج. تُبرز النتائج قوة تقنيات التنظيم في تحسين استقرار النموذج والدقة التنبؤية في ظل وجود مشكلة الازدواج الخطى.

الكلمات المفتاحية : الازدواج الخطى؛ انحدار الريدج ؛ انحدار لا سو؛ الشبكة المرنة ؛ النماذج الهجينة.