STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **15**, April 2026, pp 2471–2487.
Published online in International Academic Press (www.IAPress.org)

# A Comparative Study of Ridge Robust and Reciprocal Lasso Estimators for Semiparametric Additive Partial Linear Models Under Multicollinearity

Hayder Raaid Talib [1,*], Noor Abdul-Kareem Fayadh [1], Sarah Sabah Akram [2]

[1] *College of Administration and Economics, Sumer University, Thi-Qar, Iraq*
[2] *College of Administration and Economics, Wasit University, Wasit, Iraq*

**Abstract** Estimating parameters in Semiparametric Additive Partial Linear Models (SAPLMs) accurately proves quite difficult under high-dimensional data and related explanatory variables. Multicollinearity among predictors not only increases the variance of parameter estimates but also makes statistical interpretation more difficult, especially when the number of variables exceeds the sample size. We contrast two strong estimating techniques (Ridge regression with R/W robust estimators and the Reciprocal Lasso method) to solve these problems. Our work assesses their efficacy in overcoming multicollinearity while concurrently choosing important variables. We evaluate the techniques by means of three criteria, namely: Average Absolute Deviation Error (AADE), Mean Squared Error (MSE), and coefficient of determination ($R^2$), using actual educational data on elements influencing the academic performance of special needs students. Results show that the Reciprocal Lasso approach offers more accurate predictions and improved variable selection capacity than both Ridge robust methods regarding the practical aspect, in terms of simulation methods, it was observed that the Lasso method is preferable when the sample size is less than the number of explanatory variables, the Ridge with W robust method is preferable when there is a moderate correlation between the explanatory variables, and the Ridge with R robust method is preferable when there is a strong relationship between the explanatory variables.

**Keywords** semiparametric regression; ridge robust estimator; Reciprocal Lasso; multicollinearity; variable selection; educational data.

**DOI:** 10.19139/soic-2310-5070-3299

## 1. Introduction

Integrating the benefits of both linear and nonlinear modeling, semiparametric additive partial linear models (SAPLMs) are essential statistical analysis tools [1]. They are especially useful in situations where relationships between independent and dependent variables cannot be strictly linear, since they provide flexibility in capturing complex patterns [2]. Comprising a linear component for parametric relationships and a nonlinear component for nonparametric effects, SAPLMs strike the best compromise between interpretability and computational efficiency [3]. In multiple regression models, an abundance of explanatory variables can complicate the analytical process. To address this challenge, it is essential to reduce the model's dimensionality, usually by implementing particular assumptions [3, 4]. The principle of scatter plot smoothing can be seamlessly extended to higher dimensions. In principle, regression smoothing for a predictor with $d$ dimensions adheres to the same tenets as in the one-dimensional scenario, where local averaging consistently yields asymptotically accurate approximations of the regression curve [5]. This method faces two major challenges. The regression function $m(x)$ manifests as a high-dimensional surface, and when $d > 2$, it becomes unvisualizable, thereby limiting its geometric interpretation in

*Correspondence to: Hayder Raaid Talib (Email: hayder.raaid@uos.edu.iq). College of Administration and Economics, Sumer University. P4F3+7H4, Rifai, Dhi Qar Governorate, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

elucidating the relationship between $X$ and $Y$. The efficacy of nonparametric smoothing, reliant on averaging within local neighborhoods, is compromised by data sparsity in high-dimensional spaces. Despite substantial sample sizes (e.g., $n \geq 1000$), the distribution of data points frequently remains too sparse to guarantee dependable smoothing [5, 6]. [7] conducted a study on the estimation of partial linear models using wavelet analysis and robust estimation techniques. The research aimed to enhance the accuracy of statistical model estimation by integrating linear and nonlinear relationships. Specifically, Huber's M-estimation was applied to estimate the linear component, while the Wavelet Thresholding method was used for the nonlinear component. The study concluded that combining these two approaches effectively addresses issues related to nonlinearity and outliers in statistical models. Researchers in 2014 conducted research on variable selection in Frailty Models using the H-likelihood method [8]. The research aimed at improving the manner in which we select models with random effects. The researchers compared three approaches to improving variable selection in semi-parametric models: Least Absolute Shrinkage and Selection Operator (LASSO), Smoothly Clipped Absolute Deviation (SCAD), and Penalized H-Likelihood (HL). Results showed that HL and SCAD performed better than LASSO in the selection of variables and error reduction of predictions. Simulated results were characterized by the sudden decline of prediction error rates through SCAD and HL compared to LASSO. The study concluded that HL and SCAD significantly enhance the study of semi-parametric models to a large extent and, thereby, are extremely helpful in the analysis of survival data and models with multiple random effects. Research carried out by [2] analyzed methods of dimension reduction of high-dimensional data along with variable selection within statistical modeling. The paper explained the development of the Minimum Average Variance Estimation (MAVE) method such that it was reducing the dimensions without defining a model initially. The paper introduced a new extension named the Minimum Average Variance Estimation Sparse (MAVE-S) method that involves adaptive techniques with the aim of enhancing variable selection. [1] on the other hand, investigated multicollinearity and the effects of outliers on the accuracy estimates of regression models. They noted that under such a scenario, ordinary least squares (OLS) gives inconsistent estimates. The authors suggested combining Ridge and Liu estimators and robust estimation techniques as a solution to this issue, ultimately developing new estimators known as Ridge and Liu[20, 21]. The study established that in cases where precise measurement matters, particularly during economic and industrial studies, the use of Ridge and Liu estimators in conjunction with robust regression enhances statistical accuracy. [9] contrasted two estimation methods for a semi-parametric regression model when autocorrelation exists. The research utilized a semi-parametric partial linear regression model with parametric and non-parametric components. Two estimation methods were used in the research: least squares estimation (LSEM) and semi-parametric generalized least squares estimation (SGLSE) [22]. The SGLSE method proved superior to LSEM with regard to the treatment of autocorrelation, as would be evident through the results from simulation investigations [10]. A comparison was undertaken on the grounds of the criterion of mean square error (MSE). [11] analyzed the relationship of money demand with key economic variables like gross domestic product, government consumer spending, and consumer price index during the period from 2000 to 2022. Authors discussed regression model coefficients when they encountered missing data using traditional estimation methods like maximum likelihood estimation (MLE), as well as more efficient methods like R-estimators, L-estimators, and both EM and W methods. We carried out comparative studies to identify the optimum procedure for various data loss situations. The finding indicated L-Estimators produced the most precise results when data loss was 3%. EM performed better at % data loss, and R-Estimators performed better than all procedures at 5% data loss. The study confirmed that in cases where losses are high, the use of more precise estimation methods is more effective compared to traditional methods when dealing with missing data. Recently, [12] examined how outliers and multicollinearity impacted the reliability of predictions that come from regression models. They joined the Least Trimmed Squares (LTS) technique with Ridge regression in an attempt to minimize the impacts of outliers as well as to enhance the precision of their estimations [23]. Their objective was to enhance the methods through which the estimates are made. The result, with an $R^2$ of 88%, indicated that the proposed model was very accurate. The research proved that applying the LTS method along with Ridge regression can significantly contribute to predictive models and efficiently deal with missing data. The research indicates that the authors recommend applying these methods to address the issues caused by extreme observations and multicollinearity in regression analysis.

This study compares two robust estimation methods: Ridge regression with R/W robust estimators and Reciprocal Lasso. Our goal is to tackle multicollinearity and identify important variables. We look at three key metrics to assess performance: AADE, MSE, and $R^2$. This analysis looks at actual educational data to explore the academic success of students with special needs.

## 2. Methodology

### *2.1. Problem Statement*

Research problem lies in the fact that partially linear additive semiparametric models are among the most flexible and efficient statistical models for analyzing data that combine linear and nonlinear relationships, as they provide a balance between interpretability and accuracy in representing complex phenomena. however, the application of these models in practice—particularly when dealing with real, multivariate data—faces substantial statistical challenges, foremost among them the problem of multicollinearity arising from strong correlations among explanatory variables. this problem becomes more severe when the number of independent variables is large relative to the sample size, leading to instability in parameter estimators, inflated variances, and difficulty in disentangling the individual effect of each variable on the response variable. The impact of multicollinearity is not limited to computational issues; rather, it directly affects the interpretation of model results and the accuracy of statistical inference. In such cases, regression coefficients become unreliable and may exhibit illogical signs or exaggerated magnitudes, thereby weakening hypothesis testing results and reducing the predictive power of the model. Moreover, traditional estimation methods such as Ordinary Least Squares (OLS) are unable to efficiently address this problem, as they are highly sensitive to outliers and data deviations and lack effective mechanisms for selecting the most important variables within the model. Although several penalized and robust estimation methods—such as Ridge Robust estimators and Lasso-type approaches—have been developed to address multicollinearity, reduce estimator variance, and enhance model stability, the performance of these methods varies depending on the nature of the data and the correlation structure among explanatory variables. In cases of high correlation among predictors, some of these methods may fail to select all relevant variables or to achieve the desired balance between estimation accuracy and variable selection, particularly within the framework of partially linear additive semiparametric models that simultaneously incorporate parametric and nonparametric components. accordingly, the research problem arises from the pressing need to evaluate and compare the efficiency of Ridge Robust and Reciprocal Lasso estimators in addressing multicollinearity within partially linear additive semiparametric models, in terms of their ability to improve estimation accuracy, enhance parameter stability, and strengthen the selection of explanatory variables with true effects. this, in turn, ensures the development of more reliable and interpretable statistical models for practical applications, especially in the fields of educational data and social sciences.

### *2.2. Semi-parametric Additive Partially Linear Model*

The Semi-Parametric Additive Partially Linear Model is a blend of linear and nonlinear methods that helps to shed light on how independent and dependent variables relate to each other [13]. This model does a great job of looking at data sets that have complex patterns.

$$y_i = X_i^T \beta + \sum_{j=1}^{p} f_j(Z_{ij}) + e_i, \qquad i = 1,2,\dots,n,\ j = 1,2,\dots,p \tag{1}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^{p} f_j(Z_{1j}) \\ \vdots \\ \sum_{j=1}^{p} f_j(Z_{nj}) \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

Here, $y$ is a vector of the degree $(n \times 1)$ for the response variable observations, $x$ is a matrix of the degree $(n \times p)$ for the independent variable, $\beta$ is a vector from the degree $(p \times 1)$ that represents the linear part, $f_j(Z_{1j})$ is a vector from the degree $(n \times 1)$, which represents the non-linear part, and $e$ is a vector of the degree $(n \times 1)$ for a random error with mean zero and variance $\sigma^2$.

The above model is divided into the linear part $(X_i^T \beta)$, which deals with the clear linear effects between the variables, and the non-linear part $(g_j(Z_{ij}))$, which represents the non-linear relationships that cannot be easily explained using linear methods [8].

### *2.3. Multicollinearity and High Dimensionality*

**Multicollinearity problem:** Multicollinearity occurs when two or more independent (explanatory) variables in a linear regression model exhibit a strong linear correlation, making it difficult to isolate the individual effect of each variable on the dependent variable. This issue complicates the interpretation of regression coefficients and can lead to inaccurate estimates, ultimately compromising the reliability of statistical hypothesis testing [3].

****Causes of Multicollinearity**

1. The nature of the data itself: If the independent variables originate from the same source or are founded on the same phenomena, there may be a close link among them.
2. Introduction of redundant or duplicate variables.
3. Small samples or the paucity of data may make cross-correlations between variables seem to exist.
4. When variables are mathematically altered (e.g., squaring or taking the logarithm of a variable), a linear relationship may occur [5].

****Tests for detecting multicollinearity:**

1. Simple Correlation Coefficient: It is a measure of the linear relationship between two independent variables.
2. Variance Inflation Factor (VIF): It is a measure of how much the variance in a regression coefficient is being inflated due to collinearity.
3. Tolerance: It is the reverse of VIF, i.e., the share of variance explained by other predictors.
4. Eigenvalues and Condition Index: This method allows us to investigate multicollinearity by studying the eigenvalues of the correlation matrix and their condition numbers.
5. Determinant of the correlation matrix: If the determinant is small, then there is high multicollinearity.
6. Graphical Analysis: Residual plots and scatterplots are excellent ways of presenting the relationship between variables.
7. Extended Durbin–Watson Test: A modified version of the Durbin–Watson test for autocorrelation that also detects collinearity [3].

- **Effects of Multicollinearity**

1. When regression coefficients are more heterogeneous, this leads to less stable estimates that can reduce the reliability of our statistical inferences.
2. Ambiguous or confusing results: Regression coefficients may carry negative signs or unrealistic magnitudes, leading to interpretation problems.
3. Adjusted forecast: From a precision standpoint, multicollinearity can make the model less predictive by increasing standard errors and leading to overfitting [3].

### *2.4. Calculation Methods for the Semiparametric Additive Partially Linear Model:*

The Semiparametric Additive Partially Linear Model is a combination of a parametric regression function and an additive nonparametric regression function. For estimation, the parametric component can be in the form of a simple linear model, multiple linear models, or even a nonlinear function. Conversely, there exist various methods for estimating the nonparametric component. The model structure plays a significant role in the choice of estimation methods, merging parametric and nonparametric estimation to increase accuracy and flexibility.

#### *2.4.1. The Difference method*

This method is employed to estimate nonlinear functions $f_j(Z_{ij})$ and is particularly useful for avoiding the complexities associated with direct derivative calculations [14]. We assume that $f_j(Z_{ij})$ is differentiable, and its first derivative at the point $Z_{ij}$ can be approximated using forward differences as follows:

$$f'_j(z_{ij}) \approx \frac{f_j(z_{ij}+h) - f_j(z_{ij})}{h} \tag{2}$$

Then, $f_j(Z_{ij})$ is replaced by the approximate derivative using the difference method, as shown in the following formula:

$$y_i = X_i^T \beta + \frac{f_j(z_{ij}+h) - f_j(z_{ij})}{h} + e_i \tag{3}$$

Rearranging Equation (3) allows for the estimation of the coefficients corresponding to the linear and nonlinear components, as expressed in the following formula:

$$y_i = X_i^T \beta + \sum_{j=1}^{p} \frac{f_j(z_{ij}+h)}{h} - \sum_{j=1}^{p} \frac{f_j(z_{ij})}{h} + e_i \tag{4}$$

Then, the $\beta$ coefficients are estimated using the Ridge Regression method:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T \left(y - F(z)\right) \tag{5}$$

$F(z)$:Estimated values of functions $f_j(Z_{ij})$

After estimating the linear component, the function $f_j(Z_{ij})$ is estimated as follows:

$$f_j(z_{ij})^{t+1} = h. \left( y_i - x_i^T \hat{\beta} - \sum_{k \neq j} f_k(z_{ik}) \right) \tag{6}$$

$t$: Number of iterations.

$\hat{\beta}$: Estimated parameters in the equation (6)

The estimation process alternates between the nonparametric and parametric components until stability is achieved, meaning the optimal solution is reached and the convergence condition is satisfied, as follows:

$$\left\| f_j^{(t)}(Z_{ij})^{t+1} - f_j(Z_{ij})^t \right\| < \varepsilon \tag{7}$$

#### *2.4.2. Ridge Robust Estimation*

The Ridge Robust Estimator successfully remedies such problems by reducing variance and enhancing model steadiness via parameter tuning. The technique enhances the precision of statistical models when there are large data drifts. The standard estimation process may lead to inaccuracy or unsafe results in situations where independent variables are highly collinear or shift over time. Ridge Robust Estimation offers safer and precise estimates even against unexpected or unusual data conditions.

According to the provided formula, Frank and Friedman (1993) developed this technique. Parametric estimation can be problematic, including multicollinearity and non-stationarity. In the Ridge technique, a penalty term is incorporated to reduce these problems and improve the model's stability. Initially, the parametric component is assumed to be unknown, with its initial value set as $X_i^T\beta = 0$. The non-parametric component is then estimated using the Smoothing Kernel method, as demonstrated in the following formula [2, 4].

$$\hat{f}_j^{(0)}(Z) = \frac{\sum_{i=1}^{n} K_h(Z - Z_{ij})\, y_i}{\sum_{i=1}^{n} K_h(Z - Z_{ij})} \tag{8}$$

where $\hat{\mathbf{f}}_{\mathbf{j}}^{(\mathbf{0})}(\mathbf{Z})$ represents the initial estimate of the non-parametric functions, and then the residual is calculated to remove the effect of the non-parametric part as shown in the following formula [7]:

$$r_i^{(0)} = y_i - \sum_{j=1}^{p} \hat{f}_j^{(0)}(Z_{ij}) \tag{9}$$

Then, the $\beta$ will be estimated using Ridge with the residual $r_i^{(0)}$. This is done by minimizing the modified objective function, as follows:

$$Q_{\text{Ridge}}(\beta) = \left\| r_i^{(0)} - X\beta \right\|^2 + \lambda \|\beta\|^2 \tag{10}$$

By applying partial differentiation to $Q_{Ridge}(\beta)$, with respect to $\beta$ and setting it equal to zero, we obtain:

$$\hat{\beta}_{\text{Ridge}}^{(1)} = (X^T X + \lambda I)^{-1} X^T r_i^{(0)} \tag{11}$$

Through equation (4) and using $\hat{\beta}_{\text{Ridge}}^{(1)}$, we calculate the new variable $(y_i')$ and then use $y_i'$ to calculate $\hat{f}_j^{(1)}(Z_{ij})$ using the difference method [9].

$$y_i' = y_i - X_i^T \hat{\beta}_{\text{Ridge}}^{(1)} \tag{12}$$

$$\hat{f}_j^{(1)}(Z_{ij}) = \frac{\sum_{i=1}^{n} K_h(Z - Z_{ij})\, y_i'}{\sum_{i=1}^{n} K_h(Z - Z_{ij})} \tag{13}$$

Iteration is done between estimating the parametric and non-parametric parts until stability occurs, i.e., reaching the optimal solution and achieving the convergence condition, as follows [6]:

$$\left\| \hat{\beta}^{(t)} - \hat{\beta}^{(t-1)} \right\| < \varepsilon \tag{14}$$

$$\left\| \hat{f}_j^{(t)}(Z) - \hat{f}_j^{(t-1)}(Z) \right\| < \varepsilon \tag{15}$$

By continuing the repetition process, we get the final estimate, as shown below [6]:

$$\hat{\beta} = \hat{\beta}^{(t)} \quad and \qquad \hat{f}_j = \hat{f}_j^{(t)}(Z) \tag{16}$$

- **Robust Estimations:**Due to the increasing emphasis on robust estimation and insights derived from prior statistical studies, various methodologies for robust estimation of location and scale have been established. The R method and the W method are approaches designed to improve estimation accuracy by reducing the impact of outliers.

- **R-Estimators:**The term "Robust Estimation" is derived from its basis in rank tests. This idea was first introduced by Hodges and Lehmann, who used the following Walsh rates in their approach:

$$W_{ij} = x_i + x_j/2, \quad \text{where} \quad (i = 1, 2, \ldots, n) \quad \text{and} \quad (j = 1, 2, \ldots, n).$$

The Walsh rates are utilized to compute the R estimators for the location parameter via the subsequent formula:

$$T_R = \text{med}\left\{\frac{x_i + x_j}{2}\ ;\ i < j\right\}$$

This formula is the most efficient in terms of computational steps when compared to the formula for $(i < j)$ or the extended form for all value pairs $(i, j)$. Regarding the R estimators for the scale parameter, Rousseeuw and Croux introduced the following statistics:

$$S_R = C\,med_i\left\{med_j\,|x_i - x_j|\right\}$$

For each $i$, the median is computed as $\{|x_i - x_j|; j = 1, 2, ..., n\}$, resulting in n values that contribute to the final estimate of SR. The constant c is introduced to ensure that the estimate aligns with the presumed distribution, where $c = 1.6982, 0.7071, 1.1926$ for the normal, Cauchy, and exponential distributions, respectively. When applying the formula above, the median corresponds to the ordered statistic $h = [(n+1)/2]$, while the inner median, $med_j$, corresponds to the ordered statistic $h = [(n/2)+1]$. In the same study, the researchers also proposed an alternative R estimator for estimating the scale parameter, as follows [15, 16]:

$$Q_R = d\left\{|x_i - x_j|\ ;\ i < j\right\}_{(k)}$$

The constant d is employed to align the estimate with the assumed distribution, with values of (3.476, 1.207, 1.0483) allocated to the normal, Cauchy, and exponential distributions, respectively [1].

$$k = \binom{h}{2} \cong \binom{n}{2}/4; \qquad h = \lfloor n/2 \rfloor + 1$$

- **W-Estimation:**The W estimator provides an alternative to the M estimator, with $(T_n)$ being the M estimator, which is defined in the following way:

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - T_n}{CS_n}\right) = 0$$

By defining W based on the formula "$UW$"$(u) = \psi(u)$ and replacing $\psi(w)$ through equivalence, we derive the following:

$$\sum_{i=1}^{n}\left(\frac{x_i - T_n}{CS_n}\right) W\left(\frac{x_i - T_n}{CS_n}\right) = 0$$

or by rearranging:

$$T_n = \frac{\sum_{i=1}^{n} x_i\, w[(x_i - T_n)]/CS_n}{\sum_{i=1}^{n} w[(x_i - T_n)]/CS_n}$$

as $T_n$ is the $x_i$'s weighted average.

Due to the infrequency of obtaining an algebraic solution to the specified equation, akin to calculating the mean, a numerical method is utilized. Let $T^k$ represent the estimate at the $k^{th}$ iteration [17], then:

$$u_i^{(k)} = \frac{x_i - T_n^{(k)}}{CS_n}$$

Then, the iterative form is represented as:

$T_n^{(k+1)} = \frac{\sum_{i=1}^{n} x_i\, w[u_i^{(k)}]}{\sum_{i=1}^{n} w[u_i^{(k)}]}$, The equation is known as the Iteratively Reweighted Least Squares (IRLS) method, with weights defined as:

$$W_i^{(k)} = W[U_i^{(k)}]$$

.

The Ordinary Least Squares (OLS) technique is utilized to determine T, aiming to minimize the following sum of squared deviations: $\sum_{i=1}^{n}(x_i - T)^2$

And it is in the form: $T = \frac{\sum_{i=1}^{n} x_i}{n}$

The Weighted Least Squares (WLS) estimator, denoted as $T_w$, is determined based on fixed weights $W_i$.

The objective is to find $T_w$ that minimizes the following weighted sum of squared differences: $\sum_{i=1}^{n} w_i(x_i - t)^2$

which is in the following form:$T_w = \frac{\sum_{i=1}^{n} w_i(x_i - t)}{\sum_{i=1}^{n} w_i}$

The estimation process starts with a preliminary estimate of the W estimator of the spatial parameter. The Iteratively Reweighted Least Squares (IRLS) algorithm refines this estimate step by step in a sequence of iterations until it converges and gives a stable solution of good accuracy [12]. In the same fashion, while the M estimator is a generalization of the Maximum Likelihood Estimator (MLE), the W estimator is a general type of the Least Squares Estimator (LSE). While in some instances both of these two estimators will produce similar results; in most cases they will produce estimates that will be similar but distinct.

### *2.4.3. Reciprocal Lasso (RLasso)*

The Reciprocal Lasso (RLasso) methodology, formulated by Zou and Hastie (2005), aims to address specific limitations of Lasso regression.

1. When regression coefficients are more heterogeneous, this leads to less stable estimates that can reduce the reliability of our statistical inferences.
2. Ambiguous or confusing results: Regression coefficients may carry negative signs or unrealistic magnitudes, leading to interpretation problems.
3. Adjusted forecast: From a precision standpoint, multicollinearity can make the model less predictive by increasing standard errors and leading to overfitting [3].

The RLasso estimator is formulated using the subsequent expressions [14, 18]:

$$\hat{\beta}(\text{RLasso}) = \arg\min_{\beta} \|Y - X\beta\|_2^2 + \sum_{t=1}^{p} \frac{\lambda}{|\beta_t|} \tag{17}$$

$$\hat{\beta}(\text{RLasso}) = \arg\min_{\beta} \sum_{i=1}^{n} \left[ Y_i - \gamma^T B(Z_i) - X_i^T \beta \right]^2 + \sum_{t=1}^{p} \frac{\lambda}{|\beta_t|} \tag{18}$$

So, $\lambda$ represents the control parameter or penalty parameter for RLasso and is non-negative. That is, $\lambda \geq 0$

P: represents the variables number of the linear explanatory in the model and $(t = 1, 2, ..., p)$.

$B(Z_i)$:B-Spline basic function.

We used cubic B-spline functions with 4 internal knots placed at the 20th, 40th, 60th, and 80th percentiles of each predictor variable. This created 7 parameters for each nonlinear component. The analysis continued until the parameter changes became very small (less than 0.000001) or until reaching 1000 cycles. All computations were done using R software.

The Reciprocal Lasso method has the characteristics of Oracle Op's.

### 2.5. Selection of Penalty Parameter

The penalty parameter (also referred to as the regularization parameter or synthesis parameter) is denoted by $\lambda$ and plays a crucial role in model estimation. It directly influences the estimator's properties, such as bias and variance. The penalty parameter regulates the extent of parameter shrinkage and determines the selection of explanatory variables in the model. When the penalty parameter is set to zero $(\lambda = 0)$, the penalty estimators reduce to classical estimators. However, as $\lambda$ increases, it leads to excessive variable selection. In the extreme case of $\lambda \to \infty$ , all coefficients are forced to zero. Conversely, when $\lambda$ is too small, the penalty effect becomes minimal, resulting in negligible shrinkage of the coefficients.

The penalty parameter significantly influences the properties of estimated parameters, such as unbiasedness, stability, and Oracle properties. Therefore, selecting an appropriate penalty parameter with high precision is crucial. Several criteria exist for determining the optimal penalty parameter, including [13]:

#### 2.5.1. Robustified Cross Validation(RCV)

Scientists Robinson and Moyeed proposed the Smoothing Parameter Selection method using the RCV approach. This method is designed to determine the optimal bootstrap parameter in small samples, minimizing errors associated with limited sample sizes. The corresponding mathematical formula is as follows [12, 16]:

$$\mathrm{RCV} = n^{-1} \frac{1 + n^{-1} + \mathrm{tr}(R_\lambda)^2}{1 + n^{-1} + \mathrm{tr}(R_\lambda)^2} \left\| (I - R_\lambda) y \right\|^2 \tag{19}$$

Where $R_\lambda = (X^T X + \lambda I)^{-1}$

#### 2.5.2. Improved Akaike Information Criterion

This measure is considered more suitable than Akaike's Information Criterion (AIC) as it addresses the bias that arises in small sample sizes [8]. To correct this issue, a modified version, denoted as AICc, was proposed [1, 19].

$$\mathrm{AIC}_C = \log \left[ \frac{\sum \left\{ y_i - \hat{\beta}_\lambda(t) \right\}^2}{n} \right] + 1 + \frac{2\{\mathrm{tr}(R_\lambda) + 1\}}{n - \mathrm{tr}(R_\lambda) - 2} \tag{20}$$

$$\mathrm{AIC}_C = \log \left[ \frac{\left\| (R_\lambda - I) y \right\|^2}{n} \right] + 1 + \frac{2\{\mathrm{tr}(R_\lambda) + 1\}}{n - \mathrm{tr}(R_\lambda) - 2} \tag{21}$$

Optimization of Penalty Parameters The selection of the regularization parameter $\lambda$ was meticulously performed using criterion-specific optimization techniques: For Ridge Robust Estimators: The penalty parameter $\lambda$ was determined by minimizing the Robust Cross-Validation (RCV) criterion formulated in Equation (19). The optimization landscape was explored through a fine-grained grid search over the interval $\lambda \in [10^{-3}, 10^{3}]$ with logarithmic spacing, ensuring identification of the global minimum of the RCV function. For Reciprocal Lasso Estimator: The adaptive penalty parameter was optimized using the Improved Akaike Information Criterion (AICc) specified in Equation (21). The AICc formulation incorporates a finite-sample correction that provides unbiased model selection in high-dimensional settings with$p/n \approx 0.77$. The optimization procedure employed the Brent's algorithm for unimodal intervals and simulated annealing for multimodal regions, guaranteeing robust convergence to optimal $\lambda$ values that balance bias-variance tradeoff while maintaining Oracle properties.

## 3. Practical Part

This section presents the practical aspect of the study, utilizing real educational data from students with special needs. The key factors influencing students' academic performance (grade point average) were identified, and data was collected from the Iraqi Ministry of Education for wasit Governorate.

### 3.1. Description of the sample and study variables

This research examines a cohort of students with special needs from Wasit Governorate, Iraq, assessing the availability of critical data to facilitate the study's completion. Twenty-three research variables were examined, encompassing academic performance (final stage average), demographic factors (age, income, type of study, study hours, intelligence level, residence type), family background (parents' education levels, family size, parents' ages), and additional influential factors such as class size, teacher count, birth order, vision acuity, distance to school, pedagogical methods, and school attendance days, with a sample size of 30 students.

| | | | |
|---|---|---|---|
| $X_1$ | Number of study hours | $X_{12}$ | teacher count |
| $X_2$ | Student's age | $X_{13}$ | birth order |
| $X_3$ | income | $X_{14}$ | vision acuity |
| $X_4$ | type of study | $X_{15}$ | distance to school |
| $X_5$ | study hours | $X_{16}$ | pedagogical methods |
| $X_6$ | intelligence level | $X_{17}$ | and school attendance days |
| $X_7$ | residence type | $X_{18}$ | father's educational level |
| $X_8$ | parents' education levels | $X_{19}$ | mother's educational level |
| $X_9$ | family size | $X_{20}$ | student weight |
| $X_{10}$ | mother's age | $X_{21}$ | student gender |
| $X_{11}$ | class size | $X_{22}$ | father's age |
| y | final stage average | | |

Table 1. Calculations and variables selection of the coefficients utilizing the Semi-parametric Additive Partially Linear Model.

| j | Ridge with R robust | | Ridge with W robust | | Reciprocal Lasso | |
|---|---|---|---|---|---|---|
| | B | SE | B | SE | B | SE |
| $X_1$ | 3.6728 | 1.1728 | 3.8027 | 1.2819 | 3.7991 | 1.2109 |
| $X_2$ | 3.8917 | 1.1927 | 3.8928 | 1.0082 | 4.3892 | 1.6659 |
| $X_3$ | 0.2126 | 0.0076 | 0.2881 | 0.0105 | 0.2917 | 0.0121 |
| $X_4$ | 7.2910 | 1.5689 | 7.4489 | 1.7811 | 8.1192 | 1.8922 |
| $X_5$ | 6.1982 | 1.2007 | 5.9982 | 1.2901 | 4.8911 | 1.1185 |
| $X_6$ | 0.6291 | 0.1728 | 0.5983 | 0.1728 | 0.7103 | 0.1922 |
| $X_7$ | 1.2781 | 0.2781 | 1.3819 | 0.3183 | 2.1988 | 0.3004 |
| $X_8$ | 2.3897 | 0.4891 | 2.0283 | 0.3872 | 3.0064 | 0.4110 |
| $X_9$ | -0.6781 | 0.1887 | -0.5770 | 0.1779 | -0.6013 | 0.2005 |
| $X_{10}$ | -0.7099 | 0.3811 | -0.8018 | 0.2768 | -0.9004 | 0.3301 |
| $X_{11}$ | 7.9011 | 0.9027 | 8.8001 | 0.8992 | 6.8899 | 0.7099 |
| $X_{12}$ | -1.2776 | 0.6022 | -1.5583 | 0.4930 | Exclude | - |
| $X_{13}$ | 2.1276 | 0.5478 | 2.6657 | 0.4884 | 3.4376 | 0.7693 |
| $X_{14}$ | Exclude | - | Exclude | - | 0.9547 | 0.1264 |
| $X_{15}$ | Exclude | - | Exclude | - | Exclude | - |
| $X_{16}$ | 1.5722 | 0.3684 | 1.2115 | 0.4004 | 1.5873 | 0.3760 |
| $X_{17}$ | Exclude | - | Exclude | - | 0.6003 | 0.1638 |
| $X_{18}$ | 0.7693 | 0.1077 | 0.9812 | 0.2094 | 0.8662 | 0.1770 |
| $X_{19}$ | Exclude | - | 0.6844 | 0.1157 | Exclude | - |
| $X_{20}$ | Exclude | - | Exclude | - | Exclude | - |
| $X_{21}$ | Exclude | - | Exclude | - | 0.8944 | 0.2028 |
| $X_{22}$ | 4.3569 | 0.6581 | 5.2352 | 0.6833 | 6.1774 | 0.6988 |

### *3.2. Estimation*

The study employed Ridge with R robust, Ridge with W robust, and Reciprocal Lasso methods for variable estimation and selection. The analysis utilized an R-based computational method, producing the subsequent results as demonstrated in Table 1.

### *3.3. Comparison*

The estimation techniques and variable selection methods were assessed using average absolute deviation, mean squared error, and coefficient of determination. The results of this comparison are delineated in the table below.

Table 2. Performance comparison of regression methods.

| Methods | AADE | MSE | $R^2$ |
|---|---|---|---|
| Ridge with R robust | 3.0207 | 6.8794 | 73.59% |
| Ridge with W robust | 3.0067 | 7.0163 | 73.78% |
| Reciprocal Lasso | 2.6708 | 5.3849 | 77.04% |

Table 3. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.2), ($\sigma$ =0.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 4.883 | 65.455 | 3.688 | Reciprocal Lasso |
| | **Ridge with W robust** | 4.458 | 66.517 | 3.096 | |
| | **Reciprocal Lasso** | 3.728 | 69.116 | 2.684 | |
| 40 | **Ridge with R robust** | 4.327 | 66.835 | 3.255 | Reciprocal Lasso |
| | **Ridge with W robust** | 4.007 | 67.920 | 3.107 | |
| | **Reciprocal Lasso** | 3.226 | 70.044 | 2.938 | |
| 75 | **Ridge with R robust** | 3.727 | 71.637 | 3.005 | Ridge with W robust |
| | **Ridge with W robust** | 3.548 | 72.117 | 2.874 | |
| | **Reciprocal Lasso** | 3.894 | 71.895 | 2.965 | |
| 100 | **Ridge with R robust** | 3.562 | 72.907 | 2.785 | Ridge with W robust |
| | **Ridge with W robust** | 3.165 | 74.556 | 2.276 | |
| | **Reciprocal Lasso** | 3.650 | 72.176 | 2.800 | |

Based on the table's results and the comparison criteria applied, the R Lasso method demonstrates superior performance by achieving lower average absolute deviation and mean squared error compared to the other approaches. Furthermore, its coefficient of determination surpasses that of the other methods, indicating better model accuracy.

## 4. Simulation Part

In this section, the finite sample performance of the proposed procedure is investigated by Monte Carlo simulations. The methods used in Part 2, are compared, as they use the quadratic approximation of the nonparametric functions. using the model (1) where $k$ =2

$$y_i = X_i^T\beta \ + \ \sum_{j=1}^{p} f_j(Z_{ij}) \ + \ \varepsilon_i$$

$$f_1(Z_1) = 5\ \sin(1.5\pi Z_1)$$

$$f_2(Z_2) = 15\left(e^{-3.25Z_2} - 7e^{-6.4Z_2} + 4e^{-2.33Z_2}\right)$$

The experiments are replicated 1000 with different sample sizes (n= 15, 40, 75 and 100) with assuming there are 50 explanatory linear variables and two nonlinear variables. we assume $\beta$ =(3, 2.3, 0.4, -1.2, 2.4, 5, 4.6, -3.5, 3.4, -6.2, 0 , 0, ..., 0, 3.4, -0.8, 2.8, 2.9, -5.3) and ($\sigma$ =0.5, 1.5, 3). The distribution of $X$ and $\varepsilon$ are standard normal. $X$ and $\varepsilon$ are independent, the correlation between $X_i$ and $X_j[\rho^{|i-j|}]$ with $\rho$ =(0.2, 0.6 and 0.9). $Z_1$ and $Z_2$ are independent and uniformly distributed on [0,1] and B-splines are used to approximate the nonparametric functions and the number of knots in the approximation for each non-parametric component ranges from 2 to 8, the results were as follows:

Table 4. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.6), ($\sigma$ =0.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 5.844 | 62.657 | 4.488 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.034 | 62.288 | 4.187 | |
| | **Reciprocal Lasso** | 4.657 | 64.980 | 3.674 | |
| 40 | **Ridge with R robust** | 5.365 | 65.783 | 4.156 | Reciprocal Lasso |
| | **Ridge with W robust** | 4.859 | 63.932 | 4.001 | |
| | **Reciprocal Lasso** | 4.226 | 69.681 | 3.476 | |
| 75 | **Ridge with R robust** | 4.845 | 70.772 | 3.878 | Ridge with W robust |
| | **Ridge with W robust** | 4.477 | 71.366 | 3.277 | |
| | **Reciprocal Lasso** | 4.922 | 70.537 | 3.903 | |
| 100 | **Ridge with R robust** | 4.476 | 72.576 | 3.276 | Ridge with W robust |
| | **Ridge with W robust** | 4.154 | 74.699 | 3.004 | |
| | **Reciprocal Lasso** | 4.790 | 71.588 | 3.411 | |

Table 5. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.9), ($\sigma$ =0.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 6.844 | 61.768 | 5.166 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.632 | 61.005 | 5.001 | |
| | **Reciprocal Lasso** | 5.154 | 63.769 | 4.548 | |
| 40 | **Ridge with R robust** | 6.547 | 62.622 | 5.024 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.361 | 62.166 | 4.895 | |
| | **Reciprocal Lasso** | 4.576 | 63.677 | 4.439 | |
| 75 | **Ridge with R robust** | 5.260 | 65.355 | 4.374 | Ridge with R robust |
| | **Ridge with W robust** | 5.308 | 64.997 | 4.577 | |
| | **Reciprocal Lasso** | 5.899 | 64.123 | 4.824 | |
| 100 | **Ridge with R robust** | 3.657 | 67.032 | 3.582 | Ridge with R robust |
| | **Ridge with W robust** | 3.894 | 66.277 | 4.005 | |
| | **Reciprocal Lasso** | 4.265 | 65.003 | 4.376 | |

Table 6. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.2), ($\sigma$ =1.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 5.375 | 60.188 | 4.897 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.764 | 61.375 | 4.327 | |
| | **Reciprocal Lasso** | 4.899 | 64.045 | 3.866 | |
| 40 | **Ridge with R robust** | 5.005 | 62.478 | 4.438 | Reciprocal Lasso |
| | **Ridge with W robust** | 4.903 | 63.677 | 4.274 | |
| | **Reciprocal Lasso** | 4.438 | 65.327 | 3.922 | |
| 75 | **Ridge with R robust** | 4.565 | 65.834 | 3.855 | Ridge with R robust |
| | **Ridge with W robust** | 4.677 | 64.467 | 3.906 | |
| | **Reciprocal Lasso** | 4.896 | 64.677 | 4.325 | |
| 100 | **Ridge with R robust** | 4.244 | 66.903 | 3.587 | Ridge with R robust |
| | **Ridge with W robust** | 4.374 | 65.388 | 3.843 | |
| | **Reciprocal Lasso** | 4.558 | 65.117 | 4.002 | |

Table 7. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.6), ($\sigma$ =1.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 5.684 | 59.588 | 5.377 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.934 | 58.889 | 5.174 | |
| | **Reciprocal Lasso** | 5.496 | 61.277 | 4.892 | |
| 40 | **Ridge with R robust** | 5.436 | 59.165 | 5.026 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.564 | 59.002 | 5.000 | |
| | **Reciprocal Lasso** | 5.334 | 60.547 | 4.739 | |
| 75 | **Ridge with R robust** | 5.211 | 61.276 | 4.547 | Ridge with W robust |
| | **Ridge with W robust** | 5.165 | 61.875 | 4.176 | |
| | **Reciprocal Lasso** | 5.288 | 60.798 | 4.653 | |
| 100 | **Ridge with R robust** | 5.009 | 64.687 | 4.234 | Ridge with W robust |
| | **Ridge with W robust** | 4.576 | 67.123 | 3.767 | |
| | **Reciprocal Lasso** | 5.435 | 62.365 | 4.146 | |

Table 8. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.9), ($\sigma$ =1.5) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 6.143 | 62.266 | 5.546 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.788 | 61.588 | 5.174 | |
| | **Reciprocal Lasso** | 5.376 | 62.699 | 4.644 | |
| 40 | **Ridge with R robust** | 6.000 | 62.689 | 5.433 | Reciprocal Lasso |
| | **Ridge with W robust** | 5.547 | 62.975 | 5.003 | |
| | **Reciprocal Lasso** | 5.166 | 63.366 | 4.153 | |
| 75 | **Ridge with R robust** | 5.218 | 65.227 | 4.588 | Ridge with R robust |
| | **Ridge with W robust** | 5.377 | 64.733 | 4.615 | |
| | **Reciprocal Lasso** | 5.473 | 62.683 | 4.795 | |
| 100 | **Ridge with R robust** | 5.008 | 67.565 | 4.043 | Ridge with R robust |
| | **Ridge with W robust** | 5.124 | 65.377 | 4.435 | |
| | **Reciprocal Lasso** | 5.265 | 63.447 | 4.505 | |

Table 9. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.2), ($\sigma$ =3) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 6.747 | 58.908 | 5.277 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.580 | 59.226 | 5.199 | |
| | **Reciprocal Lasso** | 5.682 | 61.160 | 4.747 | |
| 40 | **Ridge with R robust** | 6.537 | 60.004 | 5.153 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.366 | 61.254 | 5.004 | |
| | **Reciprocal Lasso** | 5.740 | 63.476 | 4.358 | |
| 75 | **Ridge with R robust** | 6.411 | 65.577 | 4.484 | Ridge with R robust |
| | **Ridge with W robust** | 6.272 | 64.805 | 4.898 | |
| | **Reciprocal Lasso** | 6.682 | 63.785 | 5.178 | |
| 100 | **Ridge with R robust** | 6.015 | 67.179 | 4.300 | Ridge with R robust |
| | **Ridge with W robust** | 6.177 | 65.115 | 4.473 | |
| | **Reciprocal Lasso** | 6.377 | 64.480 | 5.026 | |

Table 10. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.9), ($\sigma$ =3) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 6.565 | 60.157 | 6.188 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.776 | 60.100 | 6.201 | |
| | **Reciprocal Lasso** | 5.376 | 63.179 | 5.547 | |
| 40 | **Ridge with R robust** | 6.373 | 61.006 | 5.578 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.436 | 62.254 | 5.576 | |
| | **Reciprocal Lasso** | 5.266 | 64.688 | 5.366 | |
| 75 | **Ridge with R robust** | 5.834 | 66.146 | 5.365 | Ridge with R robust |
| | **Ridge with W robust** | 5.699 | 67.784 | 5.169 | |
| | **Reciprocal Lasso** | 6.179 | 65.947 | 5.893 | |
| 100 | **Ridge with R robust** | 5.358 | 67.784 | 5.175 | Ridge with R robust |
| | **Ridge with W robust** | 5.022 | 69.479 | 5.040 | |
| | **Reciprocal Lasso** | 6.004 | 66.609 | 5.688 | |

Table 11. (AADE), ($R^2$) and (MSE) with ($n$ = 15, 40, 75, 100), ($\rho$ =0.9), ($\sigma$ =3) , ($P$ =50).

| n | Methods | AADE | $R^2$ | MSE | Best |
|---|---|---|---|---|---|
| 15 | **Ridge with R robust** | 6.702 | 66.387 | 5.955 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.765 | 66.894 | 6.100 | |
| | **Reciprocal Lasso** | 5.902 | 68.025 | 5.968 | |
| 40 | **Ridge with R robust** | 6.579 | 57.892 | 5.254 | Reciprocal Lasso |
| | **Ridge with W robust** | 6.412 | 67.115 | 5.177 | |
| | **Reciprocal Lasso** | 5.599 | 69.174 | 5.015 | |
| 75 | **Ridge with R robust** | 5.474 | 68.246 | 5.010 | Ridge with R robust |
| | **Ridge with W robust** | 5.797 | 67.899 | 5.265 | |
| | **Reciprocal Lasso** | 6.176 | 66.166 | 5.556 | |
| 100 | **Ridge with R robust** | 5.199 | 70.165 | 4.768 | Ridge with R robust |
| | **Ridge with W robust** | 5.540 | 69.062 | 5.100 | |
| | **Reciprocal Lasso** | 5.547 | 67.037 | 5.476 | |

## 5. Conclusion

- Offering more flexibility than many other models, the Semi-parametric Additive Partially Linear Model is a quick way to solve high-dimensional problems. It includes both linear and nonlinear additive components, therefore allowing the capture of complex interactions. Its ability to combine several linear explanatory variables and non-linear functions makes this model relevant in many industries, including health, education, social sciences, finance, and economics.

- As for the simulation part, it was observed that the (Lasso) method was superior for small sample sizes (15 and 40) and for all cases with respect to the values of correlation and standard deviation. for sample sizes (75 and 100), the Ridge with W robust method was preferable for moderate correlations (r = 0.6) and for all standard deviation values (0.5, 1.5, and 3). However, for high correlations between explanatory variables (r = 0.9), the Ridge with R robust method was preferred.

- The Reciprocal Lasso approach is best for estimating and selecting variables in the semi-parametric additive partially linear model since it shows better efficiency than other methods regarding the practical part.
- These qualities increase its adaptability for several uses. Particularly in the education sector, the study found several important factors that significantly affect results and need focus.

## 6. Recommendations for future research

Based on the methodological insights gained from this study, several avenues for future research are recommended to validate and extend our findings

1. Large-Sample Validation: The primary recommendation is to replicate this comparative analysis using a substantially larger dataset. Future studies should prioritize a significantly higher observation-to-variable ratio $(n >> p)$ to mitigate the risk of overfitting, obtain more stable standard errors, and provide a more robust assessment of the true variable selection capabilities of the Reciprocal Lasso and Ridge Robust estimators in a high-dimensional setting.
2. Advanced Validation Techniques: Employing more rigorous resampling techniques is advised. Future work should utilize repeated k-fold cross-validation or bootstrap methods specifically designed for high-dimensional data to obtain unbiased estimates of prediction error and better evaluate the generalizability of the models.
3. Application to Other Domains: Applying this comparative framework to other fields with inherently large datasets (e.g., genomics, finance, or psychometrics) would be highly valuable. This would test the robustness and transportability of our conclusions across different domains and data-generating processes.

Addressing these points in future research will be instrumental in confirming the generalizability of our results and in solidifying the practical utility of the Reciprocal Lasso estimator for semiparametric additive partial linear models

## Acknowledgement

The authors would like to thank their affiliations for partially supporting this work.

## Conflict of Interest

There is no known conflict for this work.

REFERENCES

1. Adegoke, Ajiboye S and Adewuyi, Emmanuel and Ayinde, Kayode and Lukman, Adewale F, *A comparative study of some robust ridge and liu estimators*, Sci World J. 2016;11(4):16-20.
2. Alkenani, Ali and Yu, Keming, *Sparse MAVE with oracle penalties*, Adv Appl Stat. 2013;34(2):85.
3. Aydin, Dursun and Memmedli, Memmedaga and Omay, Rabia Ece, *Smoothing parameter selection for nonparametric regression using smoothing spline*, Eur J Pure Appl Math. 2013;6(2):222-38.
4. Boente Boente, Graciela Lina and Martinez, Alejandra Mercedes, *A robust spline approach in partially linear additive models*, Comput Stat Data Anal. 2023;178:107611.
5. Chan, Jireh Yi-Le and Leow, Steven Mun Hong and Bea, Khean Thye and Cheng, Wai Khuen and Phoong, Seuk Wai and Hong, Zeng-Wei and Chen, Yen-Lin, *Mitigating the multicollinearity problem and its machine learning approach: a review*, Mathematics. 2022;10(8):1283.
6. Fan, Jianqing, *Local polynomial modelling and its applications*, Monographs on Statistics and Applied Probability 66. London: Chapman & Hall; 2018.

7. Gannaz I, *Robust estimation and wavelet thresholding in partially linear models*, Stat Comput. 2007;17 (4):293-310.
8. Ha, Il Do and Pan, Jianxin and Oh, Seungyoung and Lee, Youngjo, *Variable selection in general frailty models using penalized h-likelihood*, J Comput Graph Stat. 2014;23(4):1044-1060.
9. Razaqa, Hassan H and Talibb, Hayder R and Kamilc, Reem T, *A comparison between two methods of estimating a semi-parametric regression model in the presence of the autocorrelation problem*, Adv Theory Nonlinear Anal Appl. 2024;7(4):54-9.
10. Hastie, Trevor J, *Computer Age Statistical Inference*, Algorithms, Evidence, and Data Science. Cambridge: Cambridge University Press; 2017.
11. Hayder Raaid Talib, Hassan Hopoop Razaq, Sarah Adel Madhioom, *Comparison between classical and robust estimation methods for regression model parameters in the case of incomplete data*, Adv Theory Nonlinear Anal Appl. 2024;7(4):17-34.
12. Waibusi H, *Estimasi Parameter Regresi Ridge Robust pada Data Profil Kesehatan Sulawesi Selatan*, Makassar:Universitas Hasanuddin; 2023.
13. Hurvich, Clifford M and Simonoff, Jeffrey S and Tsai, Chih-Ling, *Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion*, J R Stat Soc Series B Stat Methodol. 1998;60(2):271-293.
14. Liu, Xiang and Wang, Li and Liang, Hua, *Estimation and variable selection for semiparametric additive partial linear models*, Stat Sin. 2011;21(3):1225.
15. Moczo, Peter and Kristek, Jozef and Halada, Ladislav, *The finite-difference method for seismologists*, An Introduction. 2004;161.
16. Nguelifack, Brice M and Kemajou-Brown, Isabelle, *Robust signed-rank estimation and variable selection for semi-parametric additive partial linear models*, J Appl Stat. 2020;47(10):1794-819.
17. Sawa, Takamitsu, *Information criteria for discriminating among alternative regression models*, Econometrica. 1978;50(6):1273-1291.
18. Talib, Hayder Raaid and Madhloom, Sarah Adel and Aazer, Kareem Khalaf, *Comparison between classical and robust methods for linear model parameters in the presence of outlier's values*, Edelweiss Appl Sci Technol. 2024;8(6):8506-13.
19. Wei, Chuan-hua and Liu, Chunling, *Statistical inference on semi-parametric partial linear additive models*, J Nonparametr Stat. 2022;24(4):809-23.
20. Abed, Ahmed Razzaq and Nayef Al-Qazaz, Qutaiba N, *Comparison of two approaches robust-ridge estimator in restricted additive partially regression model*, AIP Conference Proceedings. 2025;3264(1):050023.
21. Alharthi, Muteb Faraj and Akhtar, Nadeem, *Newly Improved Two-Parameter Ridge Estimators: A Better Approach for Mitigating Multicollinearity in Regression Analysis*, Axioms. 2025;14(3):186.
22. Kitano, Takahiro and Noma, Hisashi, *Ridge, lasso, and elastic-net estimations of the modified Poisson and least-squares regressions for binary outcome data*, arXiv preprint arXiv:2408.13474,2024.
23. Shabbir, Maha and Chand, Sohail and Dar, Irum Sajjad, *Bagging-based heteroscedasticity-adjusted ridge estimators in the linear regression model*, Kuwait Journal of Science. 2025;52(3):100412.