

Article

Not peer-reviewed version

Penalty Strategies in Semiparametric Regression Models

[Ayuba Jack Alhassan](#) , [S. Ejaz Ahmed](#) , [Dursun Aydin](#) ^{*} , [Ersin Yilmaz](#) ^{*}

Posted Date: 13 February 2025

doi: 10.20944/preprints202502.0947.v1

Keywords: Shrinkage estimators; penalty functions; kernel smoothing; semiparametric regression



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Penalty Strategies in Semiparametric Regression Models

Ayuba Jack Alhassan ¹, S. Ejaz Ahmed ², Dursun Aydin ¹ and Ersin Yilmaz ^{1,*}

¹ Department of Statistics, Mugla Sitki Kocman University, Mugla, Turkey

² Department of Mathematics and Statistics, Brock University, Ontario, St. Catharines, Canada

* Correspondence: yilmazersin13@hotmail.com

Definition: Penalty strategies in semiparametric regression models refer to regularization methods—such as Ridge, Lasso, SCAD, ElasticNet, MCP, and shrinkage estimators—that balance bias and variance in partially linear models. By penalizing or shrinking certain coefficients, these approaches handle challenges like multicollinearity and high-dimensional data, helping to improve estimation accuracy and interpretability for both the parametric and nonparametric components.

Abstract: This study includes a comprehensive evaluation of six penalty estimation strategies for partially linear models (PLRMs), focusing on their performance in the presence of multicollinearity and their ability to handle both parametric and nonparametric components. The methods under consideration include Ridge regression, Lasso, Adaptive Lasso (aLasso), smoothly clipped absolute deviation (SCAD), ElasticNet, and minimax concave penalty (MCP). In addition to these established methods, we also incorporate Stein-type shrinkage estimation techniques that are standard and positive shrinkage, and assess their effectiveness in this context. To estimate the PLRMs, we considered a kernel smoothing technique grounded in penalized least squares. Our investigation involves a theoretical analysis of the estimators' asymptotic properties and a detailed simulation study designed to compare their performance under a variety of conditions, including different sample sizes, numbers of predictors, and levels of multicollinearity. The simulation results reveal that aLasso and shrinkage estimators, particularly the positive shrinkage estimator, consistently outperform the other methods in terms of Mean Squared Error (MSE) relative efficiencies (RE), especially when the sample size is small, and multicollinearity is high. Furthermore, we present a real data analysis using the Hitters dataset to demonstrate the applicability of these methods in a practical setting. The results of the real data analysis align with the simulation findings, highlighting the superior predictive accuracy of aLasso and the shrinkage estimators in the presence of multicollinearity. The findings of this study offer valuable insights into the strengths and limitations of these penalty and shrinkage strategies, guiding their application in future research and practice involving semiparametric regression.

Keywords: Shrinkage estimators; penalty functions; kernel smoothing; semiparametric regression

1. Introduction

The Partially Linear Regression Model (PLRM), introduced by [16], serves as a versatile tool for utilizing the advantage of both parametric linear models and nonparametric regression models. It allows for modeling complex relationships by incorporating both linear and nonlinear components, thereby offering greater flexibility and interpretability across various scientific disciplines, including social sciences, biology, and economics. Formally, the PLRM is defined as:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

where y_i 's are the observations of the response variable, $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$ and t_i 's are the values of the explanatory variable, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ is an unknown k -dimensional parameter vector to be estimated, $f(\cdot)$ is an unknown smooth function, and ε_i 's are assumed to be uncorrelated random variables with mean zero and a common variance σ^2 and independent of the explanatory variables. Note also that the observations of y_i depend linearly on the entries of \mathbf{x}_i and nonlinearly on the values of a univariate variable t_i . The PLRM can be re-expressed in matrix and vector form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f} + \boldsymbol{\varepsilon} \quad (1.2)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ is an $(n \times k)$ design matrix with $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$ denoting the i .th k -dimensional row vector of the design matrix \mathbf{X} , and $\mathbf{f} = (f(t_1), \dots, f(t_n))'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ are random error vectors with $E(\boldsymbol{\varepsilon}) = 0$ and $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$. For more discussion on Model (1.1), see [39], [20], and [35], among others. Since it was initially presented by [16], this model has been very popular and is frequently used in the social, biological, and economic sciences. The PLRM generalizes both the parametric linear regression model and nonparametric regression model. When the $\mathbf{f} = 0$ model (1.1) reduces to the linear regression model $y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$, when the $\boldsymbol{\beta} = 0$ model (1.1) becomes a nonparametric regression model,

$$y_i = f(t_i) + \varepsilon_i \quad (1.3)$$

with a univariate covariate. In addition, because PLRMs contain both linear and nonlinear components, they are more flexible than linear models are. In this work, we are interested in estimating the parameter vector $\boldsymbol{\beta}$ and the function \mathbf{f} .

Estimating the parameters and the smooth function in PLRM has been a focal point of research, with various methodologies proposed to enhance estimation accuracy and model interpretability. Early contributions by [22] and [39] utilized kernel smoothing techniques for estimating the nonparametric component, laying the groundwork for subsequent advancements.

The advent of regularization methods introduced by [41] and [54] revolutionized parameter estimation by addressing issues such as multicollinearity and high-dimensionality through penalty functions. In particular, [4], [44] explored penalized least squares in semiparametric models, emphasizing the integration of penalty functions with kernel smoothing to enhance estimator performance. Further, [10,11,47] and [48].

Adaptive methods, such as the Adaptive Lasso (aLasso) introduced by [52], have been incorporated into PLRM frameworks to improve variable selection consistency. Moreover, SCAD proposed by [18] and MCP introduced by [51] penalties have been employed to mitigate the bias inherent in L_1 penalties, providing more accurate parameter estimates in complex models. Recent studies by [46] and [12] highlight the asymptotic properties of penalized estimators in semiparametric regression, emphasizing the significance of penalty strategies in PLRM. Challenges in balancing bias and variance persist, particularly in sparsity contexts with multicollinearity.

In regression problems, multicollinearity emerges when variables are highly correlated with each other and the dependent variable. This situation inflates confidence intervals, weakens parameter estimates, and leads to less reliable predictions. It also raises sampling variance, reducing accuracy for both inference and prediction. To counter these challenges in PLRMs, we integrate shrinkage estimation with penalty functions to bolster robustness and manage both multicollinearity and variable selection:

- We tackle variable selection and multicollinearity in Model (1.1), which features strongly correlated covariates, using six penalty functions: Ridge, Lasso, aLasso, SCAD, Elastic Net, and MCP.
- We then review and compare these methods, highlighting their strengths, limitations, and uses within PLRMs.
- To further enhance parameter estimation, we incorporate shrinkage estimation techniques—both standard and positive shrinkage estimators—into the penalty-based framework.
- Finally, kernel smoothing via penalized least squares is employed to estimate both the parametric (linear) and nonparametric (smooth) components of the model.

The remainder of this paper is structured as follows. Section 1.2 summarizes the background on penalty functions, partially linear models, and their integration. Section 2 describes the kernel smoothing method. Section 3 introduces the kernel-type ridge estimator. In Section 4, we discuss penalty functions and shrinkage estimators based on kernel smoothing. Section 5 presents the evaluation metrics for both parametric and nonparametric components. Section 6 outlines the estimators' asymptotic properties. Section 7 details the Monte Carlo simulations. Section 8 compares the performance of our proposed estimators on real data. Finally, Section 9 provides concluding remarks and includes supplementary technical materials.

Background

In this section, we present a relevant literature review of partially linear regression models (PLRMs) and show why penalty functions and shrinkage estimators are essential for handling challenges like multicollinearity, high-dimensional data, and complex nonlinearities. We also explain how kernel smoothing complements these approaches by capturing intricate patterns in the data based on the previous studies. Altogether, these methods offer a strong foundation for building reliable and flexible statistical models.

Partially Linear Regression Models (PLRMs) are a cornerstone of semiparametric statistics, offering a powerful way to model relationships that aren't fully captured by purely parametric or nonparametric approaches. They blend the interpretability of linear models with the flexibility of nonparametric methods, making them ideal for data with both linear and nonlinear components. Their ability to isolate certain covariate effects while accommodating nonlinearities has led to widespread use in fields such as economics, biostatistics, and environmental science [25]. Recent developments have expanded their scope to high-dimensional, longitudinal, and survival data ([38]; [37]).

Regularization methods, particularly those using penalty functions, have revolutionized regression analysis by tackling issues like multicollinearity, high dimensionality, and model selection. By constraining coefficient magnitudes, these methods shrink coefficients toward zero and improve model stability. Different penalty functions yield diverse estimators suited to various data characteristics and research goals.

Ridge regression, pioneered by [21], is well-established, yet its application in high-dimensional contexts and its ties to Bayesian methods continue to be studied [45]. [41] Lasso has also evolved substantially, with research on its consistency in variable selection, high-dimensional settings, and extensions to generalized linear models [14]. Ahmed and colleagues have contributed to Lasso's development in PLRMs using Stein-type shrinkage estimators, especially under multicollinearity [48].

Building on [53] adaptive Lasso (aLasso), researchers have explored data-driven weights for improving variable selection and estimation. For SCAD, introduced by [18], recent work has focused on high-dimensional data, quantile regression, and robust estimation [42]. The researchers applied SCAD along with the other penalty functions and with shrinkage estimators in PLRMs, showing its effectiveness in handling outliers [10]. MCP, initially studied by [51], has inspired efficient algorithms and further theoretical exploration [13]. Elastic Net, proposed by [53], has also gained traction in areas like genomics, imaging, and finance, where it handles correlated predictors well [54]. Its benefits are showed for balancing bias and variance in PLRMs [3].

Beyond these penalty-based approaches, shrinkage estimation techniques—particularly Stein-type estimators—offer another route. [5] investigate partially linear models by splitting the coefficient vector into main and nuisance effects and comparing shrinkage, pretest, and absolute penalty estimators. Their results show that a Stein-type shrinkage estimator uniformly outperforms the conventional least squares approach—especially when the nuisance parameter space is large—while the pretest estimator only offers benefits in a limited region. These methods shrink coefficients toward a chosen target, thereby reducing variance and mitigating noise ([49]; [23]; [28]). Further, Ahmed's work demonstrates that these techniques can improve estimation accuracy and interpretability in partially linear regression models ([4]; [30]; [2]). Meanwhile, kernel smoothing remains central to nonparametric estimation for capturing nonlinear relationships, with advances such as adaptive bandwidth selection [24] and its integration with penalty and shrinkage methods [17]. [50] showed via simulation that applying Stein-type shrinkage to the parametric components of [26] semi-nonparametric regression model enhances accuracy. [9] proposed a two-stage approach using LASSO/Elastic-Net for variable selection followed by post-selection shrinkage estimation to improve predictions in high-dimensional sparse linear regression models, while [1] introduced shrinkage-type estimators for reliability and stress-strength parameters from progressively censored data that outperform MLEs in simulations and industrial applications. Additional insights on shrinkage estimators are provided in [29,31,32], [7,33,36], and [6].

Combining penalty functions with kernel smoothing has become a potent strategy for complex data in PLRMs. This approach refines the parametric part through penalty-driven selection and shrinkage, while kernel smoothing captures nonlinearities. Recent studies highlight theoretical and practical benefits, including improved estimation accuracy and consistency in variable selection under multicollinearity or high dimensionality [8]. The advanced penalized kernel smoothing in PLRMs using shrinkage estimators are shown by [2], showing strong performance in simulations and real data. Their work illustrates the advantages of using penalties like aLasso and shrinkage estimation alongside kernel smoothing, especially with high-dimensional covariates and intricate nonlinearities.

2. Estimation Based on Kernel Smoothing

Let us first consider the nonparametric estimation of the unknown regression function $f(t)$ in Model (1.1). For simplicity, we assume that β is known from equation (1.1). The relationship between $y_i - \mathbf{X}_i\beta$ and t_i in this instance may be represented as

$$(y_i - \mathbf{X}_i\beta) = f(t_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

The nonparametric part of the semiparametric model is equivalent to equation (2.1). This results in the Nadaraya-Watson estimator, commonly known as the kernel estimator, which was proposed by [27] and [43], as mentioned in [39]:

$$\hat{f}(h, t) = \sum_{i=1}^n w_i(h, t_i)(y_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{W}_h(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.2)$$

where \mathbf{W}_h is a kernel smoothing matrix with $j - th$ entries w_{hi} and where h is a smoothing parameter (or bandwidth), as shown by:

$$w_i(h, t_i) = K\left(\frac{t-t_i}{h}\right) / \sum_{i=1}^n K\left(\frac{t-t_i}{h}\right) = K(u_i) / \sum_i K(u_i) \quad (2.3)$$

To estimate $f(t)$, kernel smoothing (regression) uses appropriate weights $w_i(h, t)$, as shown in (2.1). The kernel function $K(u)$ with a smoothing parameter h , which defines the size of the neighborhood around t , directs the weights assigned to the observations t_i ([40]). The kernel or weight function $K(u)$ in equation (2.3) has the properties of $\int K(u)du = 1$ and $K(u) = K(-u)$. The kernel function is chosen to give the closest observations the most weight and the farthest from t the least weight.

The following partial residuals may be obtained via the matrix and vector forms of model (1.2) in matrix form:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \hat{\mathbf{f}} = (\mathbf{I} - \mathbf{W}_h)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta} \quad (2.4)$$

where $\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{W}_h)\mathbf{X}$ and $\tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{W}_h)\mathbf{y}$. Hence, a transformed set of data based on kernel residuals is what we obtain as a result. The following weighted least squares (WLS) criterion is obtained by considering these partial residuals for the vector $\boldsymbol{\beta}$:

$$\mathbf{WLS}(\boldsymbol{\beta}; h) = ((\mathbf{I} - \mathbf{W}_h)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))^T (\mathbf{I} - \mathbf{W}_h)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \quad (2.5)$$

The result for the criterion $\mathbf{WLS}(\boldsymbol{\beta}; h)$ provided in equation (2.4) is easily demonstrated to be similar to the solution of ordinary least squares (OLS)

$$\hat{\boldsymbol{\beta}}_{KS} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}} \quad (2.6)$$

Additionally, revising the steps for $f(t)$ reduces equation (2.1) to the form:

$$\hat{f}(h, t) = \sum_{i=1}^n \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right)} (y_i - x_i \hat{\beta}) \quad (2.7)$$

Another way to express equation (2.7) in a matrix is as follows:

$$\hat{\mathbf{f}}_{KS} = \mathbf{W}_h(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{KS}). \quad (2.8)$$

Thus, our estimation for \mathbf{y}_{KS} becomes:

$$\hat{\mathbf{y}}_{KS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{KS} + \hat{\mathbf{f}}_{KS} = \mathbf{X}(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}} + \mathbf{W}_h(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{KS}) = \mathbf{H}_{KS}(h)\mathbf{y}$$

for the hat (projection) matrix

$$\mathbf{H}_{KS}(h) = \mathbf{W}_h + \tilde{\mathbf{X}}(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'(\mathbf{I} - \mathbf{W}_h) \quad (2.9)$$

3. Kernel Type Ridge Estimator in Semiparametric Model

For the sake of this paper, we limit our attention to kernel smoothing estimators of the vector parameter $\boldsymbol{\beta}$ and the unidentified smooth function $f(\cdot)$ in a semiparametric model. The corresponding estimators $\boldsymbol{\beta}$ and \mathbf{f} are based on model (1.2) for a given bandwidth parameter h , and by multiplying $(\mathbf{I}_n - \mathbf{W}_h)$ on both sides of model (1.2), we obtain the following model:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}} \quad (3.1)$$

where $\tilde{\mathbf{y}} = (\mathbf{I}_n - \mathbf{W}_h)\mathbf{y}$, $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{f}} + \boldsymbol{\varepsilon}^*$, $\tilde{\mathbf{f}} = (\mathbf{I}_n - \mathbf{W}_h)\mathbf{f}$ and $\boldsymbol{\varepsilon}^* = (\mathbf{I}_n - \mathbf{W}_h)\boldsymbol{\varepsilon}$.

These considerations show that Model (3.1) is the optimal solution for obtaining the vector $\boldsymbol{\beta}$ estimator corresponding to the parametric part of the semiparametric model (1.2). For the ridge regression problem, this model provides the following penalized least squares (PLS) criterion:

$$PLS_{RK} = \arg \min_{\boldsymbol{\beta}} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}'\boldsymbol{\beta} \quad (3.2)$$

where λ denotes a positive shrinkage parameter that regulates the severity of the penalty. The solution of the minimization problem (3.2) can be written as follows as ridge-type kernel smoothing estimator (see [46]) similar to [21] but modified based on partial residuals obtained using smoothing matrix \mathbf{W}_h :

$$\hat{\boldsymbol{\beta}}_{RK}(\lambda) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \lambda \mathbf{I}_k)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}} \quad (3.3a)$$

The ridge-type kernel smoothing estimate simplifies to an ordinary least squares estimation problem when $\lambda = 0$ on the basis of the local residuals specified in equations (2.6) and (2.8). Additionally, to estimate the unknown function \mathbf{f} , we imitate equation (2.8) and define it.

$$\hat{\mathbf{f}}_{RK} = \mathbf{W}_h(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{RK}(\lambda)) \quad (3.3b)$$

Thus, in the semiparametric model (1.2), the estimator (3.3b) is therefore defined as the ridge type kernel estimator of the unknown function \mathbf{f}

4. Penalty Functions and Shrinkage Estimators

Many penalty functions for linear and generalized regression models have been studied ([2,3]). Here, we consider the *MCP*, *Lasso*, *SCAD*, *aLasso*, and *ElasticNet* functions and shrinkage techniques additionally. Note that *ElasticNet* is a regularized regression technique that linearly combines the L_1 and L_2 penalties of the *Lasso* and *ridge* regression methods, respectively.

In this paper, we introduce the kernel smoothing estimators based on several penalties for the semiparametric regression model's components. For a given penalty function and tuning parameter λ , the general form of the penalty estimators is defined by the following penalized least squares (PLS) criterion:

$$PLS = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2 \right\} + P_{\lambda}(\boldsymbol{\beta}) = (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) + P_{\lambda}(\boldsymbol{\beta}) \quad (4.1)$$

where $P_{\lambda}(\boldsymbol{\beta})$ denotes the penalty function based on the shrinkage (or tuning) parameter λ . This parameter needs to be selected with any selection criteria.

Note that [19] bridge estimator is the vector $\hat{\boldsymbol{\beta}}$ that minimizes the *Lasso* and *Ridge* penalties in (4.1). However, *SCAD*, *ElasticNet*, *aLasso*, and *MCP* have different penalties, which are shortly introduced in the remainder of this study. Note that $|\beta_j|^q$ fulfills the L_q norm of the regression coefficients β_j because of the penalty functions used.

Hence, for various values of the shrinkage parameter λ , different penalty estimators that correspond to the parametric and nonparametric parts of the semiparametric model can be defined.

4.1. Estimation Procedure for the Parametric Component

According to (4.1), it is possible to obtain *ridge* estimates for the parametric component for $q = 2$ by minimizing the following penalized residual sum of squares.

$$\hat{\boldsymbol{\beta}}_{RK} = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^k (\beta_j)^2 \right\} \quad (4.2)$$

where $\tilde{\mathbf{x}}_i'$ is the i .th row of the matrix $\tilde{\mathbf{X}}$, and \tilde{y}_i is the i .th observation of $\tilde{\mathbf{y}}$. Note that the regularization estimate in solution (4.2) matches that in (3.3a). Alternative methods to ridge are described below.

Lasso: [41] proposed *Lasso*, a regularization method that uses the L_1 penalty for estimation and variable selection. The modified kernel smoothing estimators based on the *Lasso* penalty can be defined as

$$\hat{\boldsymbol{\beta}}_{LK} = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^k |\beta_j| \right\} \quad (4.3)$$

The absolute penalty term prevents *Lasso* from having an analytical solution, despite the apparent simplicity of equation (4.3).

aLasso: [52] proposed using adaptive weights on L_1 penalties on the regression coefficients to modify the *Lasso* penalty. The term "*aLasso*" refers to this weighted *Lasso* with oracle properties. The kernel smoothing estimator $\hat{\boldsymbol{\beta}}_{aLK}$ using *aLasso* penalty is defined as follows:

$$\hat{\boldsymbol{\beta}}_{aLK} = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^k \hat{w}_j |\beta_j| \right\} \quad (4.4)$$

where w is a weight function given by

$$\hat{w}_j = \frac{1}{|\hat{\beta}_j^*|^q}, q > 0$$

Notably, $\hat{\beta}^*$ is an appropriate estimator of β . In addition, an ordinary least squares (OLS) estimate can be used as a reference value. After the OLS estimate is obtained, $q > 0$ is selected, and the weights are computed to obtain the adaptive-Lasso estimates in (4.4).

SCAD: The *Lasso* method's penalty term increases linearly with the regression coefficient's size, leading to highly biased estimates for coefficients with large values. To address this issue, [18] proposed a *SCAD* penalty that can be obtained by replacing $|\beta_j|$ in (4.3) with $P_{\alpha,\lambda}|\beta_j|$. A modified kernel smoothing estimator $\hat{\beta}_{SK}$ based on the *SCAD* penalty can be expressed as

$$\hat{\beta}_{SK} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i' \beta)^2 + \sum_{j=1}^k P_{\alpha,\lambda} |\beta_j| \right\} \quad (4.5)$$

where $P_{\alpha,\lambda}(\cdot)$ is the *SCAD* penalty defined by

$$P_{\alpha,\lambda} = \lambda \left\{ \mathbf{I}(|\beta|) \leq \lambda + \frac{(\alpha\lambda - |\beta|)_+}{(\alpha - 1)\lambda} \mathbf{I}(|\beta| > \lambda) \right\}, \text{ for } \beta \geq 0 \quad (4.5a)$$

The penalty parameters in this case are $\lambda > 0$ and $\alpha > 2$, where $(t)_+ = \max(t, 0)$ and $\mathbf{I}(\cdot)$ is the indicator function. When $\alpha = \infty$, (4.5a) is equivalent to the L_1 penalty.

ElasticNet: *ElasticNet* is a penalized least squares regression method proposed by [53] that is widely used for regularization and automatic variable selection to choose groups of correlated variables. It combines the L_1 and the L_2 penalty term, where the L_1 penalty term ensures sparsity, whereas the L_2 penalty term selects correlated variable groups. Accordingly, the modified kernel-smoothing estimator $\hat{\beta}_{ENK}$ using an *ElasticNet* penalty is the solution to the following minimization problem:

$$\hat{\beta}_{ENK} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i' \beta)^2 + \lambda_1 \sum_{j=1}^k |\beta_j|^2 + \lambda_2 \sum_{j=1}^k |\beta_j| \right\} \quad (4.6)$$

In the expression above, the positive regularization parameters are λ_1 and λ_2 . Note that (4.6) ensures that the estimates correspond to the parametric part of the semiparametric regression model (1.2).

MCP: [51] introduced *MCP* as a different method to obtain less biased estimates of nonzero regression coefficients in a sparse model. For given regularization parameters $\lambda > 0$ and $\alpha > 0$, the kernel smoothing estimator $\hat{\beta}_{MCK}$ based on the *MCP* penalty can be defined as

$$\hat{\beta}_{MCK} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (\tilde{y}_i - \tilde{\mathbf{x}}_i' \beta)^2 + \sum_{j=1}^k P_{\alpha,\lambda}(|\beta_j|) \right\} \quad (4.7)$$

where $P_{\alpha,\lambda}(\cdot)$ is the *MCP* penalty expressed as

$$P_{\alpha,\lambda}(\beta) = \int_0^{|\beta|} \left(\frac{\lambda - x}{\alpha} \right)_+ dx = \left(\lambda |\beta| - \frac{\beta^2}{2\alpha} \right) \mathbf{I}(0 \leq |\beta| < \lambda\alpha) + \frac{\lambda^2 \gamma}{2} \mathbf{I}(|\beta| \geq \lambda\alpha)$$

4.2. Shrinkage Estimators

Shrinkage estimation offers a robust method for parameter estimation, especially in sparse environments where traditional methods may face high variance or multicollinearity. This section introduces standard and positive shrinkage estimators to improve accuracy and stability. Shrinkage estimators balance full model estimators ($\hat{\beta}^{FM}$), which use extensive predictors, with sub model estimators ($\hat{\beta}^{SM}$), which focus on fewer, significant predictors using any of introduced penalty functions (see [2]). To better understand the shrinkage process, it is necessary to partition the design matrix X into subsets that correspond to significant and less significant predictors. Let X represent the $n \times k$ design matrix, where n is the number of observations, and k is the total number of predictors. We partition X into two submatrices: $\tilde{X} = [\tilde{X}_1, \tilde{X}_2]$ where \tilde{X}_1 corresponds to the subset of significant nonzero predictors (p_1 predictors) and \tilde{X}_2 , contains the remaining predictors ($p - p_1$) where $p = p_1 + p_2$. This partitioning ensures that the shrinkage process selectively regularizes the coefficient associated with less significant predictors while preserving the interpretability of the model.

After optimizing $\hat{\beta}^{FM}$ and $\hat{\beta}^{SM}$ we can define the shrinkage estimator $\hat{\beta}_S$ and positive part of the shrinkage estimator to avoid the over-shrinking of the regression coefficients $\hat{\beta}_{PS}$ as follows:

$$\begin{aligned} \hat{\beta}_S &= \hat{\beta}^{FM} - \frac{(p_2 - 2)}{\hat{T}(\hat{\beta}^{FM} - \hat{\beta}^{SM})} \\ \hat{\beta}_{PS} &= \hat{\beta}^{FM} - \left[\frac{(p_2 - 2)}{\hat{T}} \right]_1 (\hat{\beta}^{FM} - \hat{\beta}^{SM}) \end{aligned} \quad (4.8)$$

where $(\eta)_1 = \max[\eta, 1]$, p_2 represents the number of sparse parameters that are detected by the penalty functions to estimate $\hat{\beta}^{SM}$, $\hat{\mathbf{T}}$ is the distance measure which can be defined as $\hat{\mathbf{T}} = \sigma^{-2} n \hat{\beta}'_2 (\tilde{\mathbf{X}}'_2 \tilde{\mathbf{U}} \tilde{\mathbf{X}}_2) \hat{\beta}_2$ with sparse subset of regression coefficients denoted by $\hat{\beta}_2$, model variance σ^2 , $\tilde{\mathbf{X}}_2$ matrix of partial residuals of associated predictors with $\hat{\beta}_2$ and the projection matrix $\tilde{\mathbf{U}} = \mathbf{I}_n - [\tilde{\mathbf{X}}_1 (\tilde{\mathbf{X}}'_1 \tilde{\mathbf{X}}_1 + \lambda_s \mathbf{I}_{p_1})^{-1} \tilde{\mathbf{X}}'_1]$ where $\tilde{\mathbf{X}}_1$ is the partial residuals associated with the p_1 significant coefficients. Finally, λ_s is again shrinkage parameter. Notice that some of the advantages of using shrinkage estimators can be mentioned as follows: introducing the bias through the shrinkage, these estimators reduce variance and leading the smaller Mean Squared Error (MSE) values which means better prediction accuracy as well. In addition, shrinkage techniques shrink the coefficients in a controlled manner which preserved the interpretability.

Building upon the Shrinkage Estimation framework introduced above, we integrate both standard and Positive Shrinkage Estimators into the penalty estimation strategies. These estimators provide an additional layer of regularization, enhancing the estimation accuracy for sparse models. The estimation procedure involves:

- Parametric Component Estimation: We utilize the respective penalty functions (Ridge, Lasso, aLasso, SCAD, ElasticNet, MCP) to estimate the parameter vector β .
- Shrinkage Application: We apply the shrinkage techniques to obtain ordinary stein-type shrinkage $\hat{\beta}_s$ or positive shrinkage estimator $\hat{\beta}_{PS}$ to refine the parameter estimates.
- Nonparametric Component Estimation: We estimate the smooth function $f(t)$ using the refined parameter estimates via kernel smoothing as described in equation (4.8).

This comprehensive approach leverages the strengths of both penalty functions and shrinkage estimators, offering a robust estimation framework for PLRM. The incorporation of shrinkage estimators offers several advantages. By introducing a controlled bias, these estimators reduce the variance of parameter estimates, leading to more stable and reliable predictions. Additionally, they enhance model interpretability by focusing on significant predictors while mitigating the effects of multicollinearity. Moreover, the positive shrinkage estimator provides a practical safeguard against excessive regularization, striking a balance between bias and variance.

4.3. Estimation Procedure for the Nonparametric Component

For the parametric part of the semiparametric model in (1.2), modified kernel smoothing estimates based on various penalties are provided by equations (4.2– 4.7). Analogous to that of (3.3b), the estimation of the nonparametric part of the same model can be constructed via the vector of estimated parametric coefficients $\hat{\beta}_{LK}$ provided in (4.3). On the basis of the *Lasso* penalty of the unknown function, we obtain the transformed local estimates as follows:

$$\hat{\mathbf{f}}_{LK} = \mathbf{W}_h(\mathbf{y} - \mathbf{X}\hat{\beta}_{LK}) \quad (4.8)$$

as described in the previous section. Importantly, local estimates of the nonparametric component based on the adaptive-Lasso penalty are derived and are symbolically given as $\hat{\mathbf{f}}_{aLK}$ in equation (4.8), where $\hat{\beta}_{aLK}$ defined in (4.4) is written in place of $\hat{\beta}_{LK}$. Similarly, for the nonparametric part of the semiparametric model (1.2), the modified local kernel estimators $\hat{\mathbf{f}}_{SK}$, $\hat{\mathbf{f}}_{ENK}$, and $\hat{\mathbf{f}}_{MCK}$ based on the *SCAD*, *ElasticNet*, and *MCP* penalties, respectively, are obtained by replacing $\hat{\beta}_{LK}$ in (4.8) with $\hat{\beta}_{SK}$, $\hat{\beta}_{ENK}$, and $\hat{\beta}_{MCK}$.

To better understand the adaptation of the penalty functions to the estimation process of the partially linear model (1.1), a generic algorithm is introduced. As mentioned previously, to obtain a specific estimator on the basis of a determined penalty function, the corresponding penalty term is used in Algorithm 1. Therefore, the output estimators are denoted as $(\hat{\beta}, \hat{\mathbf{f}})$ to generalize the algorithm. For example, if we obtain the *Lasso* estimator of the parametric components in model (1.1), first, $\hat{\beta}_{LK}$ is found to minimize (4.2), then $\hat{\mathbf{f}}_{LK}$ is obtained by using the vector of estimated parametric coefficients $\hat{\beta}_{LK}$. A similar path is followed for the remaining estimators. The algorithm describing their findings is as follows:

Table 1. Computation of penalty estimators.

Input: Data matrix of parametric component $\mathbf{X} \in \mathbb{R}^{n \times k}$, data vector $t_i \in \mathbb{R}^{n \times 1}$, and response vector $\mathbf{y} \in \mathbb{R}^{n \times 1}$.
Output: Pair of estimates $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})$ based on a certain penalty function
1: Select an appropriate bandwidth h using a predetermined criterion and compute the smoother matrix \mathbf{W}_h , as defined in (3.9).
2: Compute the partial residuals $\tilde{\mathbf{X}} = (\mathbf{I}_n - \mathbf{W}_h)\mathbf{X}$ and $\tilde{\mathbf{y}} = (\mathbf{I}_n - \mathbf{W}_h)\mathbf{y}$.
3: To minimize $PLS(\boldsymbol{\beta}, \mathbf{f})$, determine the shrinkage parameter λ by a predetermined criterion.
4: Partition the partial residuals of \mathbf{X} in form $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)$, as defined in Section 4.2.
5: Apply shrinkage estimators $\hat{\boldsymbol{\beta}}^{PS}$, and $\hat{\boldsymbol{\beta}}^S$ based on used penalty functions
6: Find the estimate of parametric component $\hat{\boldsymbol{\beta}}$ associated with the $\tilde{\mathbf{X}}$ contains and $\tilde{\mathbf{y}}$.
7: Estimate the nonparametric smooth function \mathbf{f} as follows: $\hat{\mathbf{f}} = \mathbf{W}_h(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$
8: Return $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})$.

5. Measuring the Quality of Estimators

This section describes several performance metrics that can be used to assess the performance of the modified semiparametric kernel smoothing estimators based on penalty functions and shrinkage estimators defined in Section 4. Note that the estimators are indicated by the acronyms in parentheses. Each estimator's performance for the parametric component, nonparametric component, and overall estimated model is examined. Consequently, the performance metrics are explained in the following section.

5.1. Evaluation of the Parametric Component

The performance of a method is related to its ability to estimate the data. Note that the bias and variance of a penalty estimator $\hat{\boldsymbol{\beta}}_S$ are measured simultaneously via the mean squared error (*MSE*) matrix. Some metrics are defined in the following section.

MSE: This value is calculated by adding the variance and squared bias, which is provided by

$$MSE(\hat{\boldsymbol{\beta}}_S, \boldsymbol{\beta}) = E [(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_S)'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_S)] = Var(\hat{\boldsymbol{\beta}}_S) + (E(\hat{\boldsymbol{\beta}}_S) - \boldsymbol{\beta})^2 \quad (5.1)$$

where $\hat{\boldsymbol{\beta}}_S$ is the vector of the estimated regression coefficients determined by either of the methods. The key information regarding the quality standards for estimation is provided by equation (4.3). To evaluate the risk more accurately via the *MSE*, a scalar variant of the *MSE* (*SMSE*) is utilized, which yields an estimator's average quality score:

$$SMSE(\hat{\boldsymbol{\beta}}_S, \boldsymbol{\beta}) = E [(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_S)'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_S)] = tr [MSE(\hat{\boldsymbol{\beta}}_S, \boldsymbol{\beta})] \quad (5.2)$$

Additionally, the relative efficiencies of the methods for estimating $\boldsymbol{\beta}$ can be easily obtained via (5.2). To do this, the relative efficiency (*RE*) is defined in Definition 5.1.

Definition 5.1: Let S_1 and S_2 denote the two distinct estimation methods for $\boldsymbol{\beta}$, and the estimators are shown as $\hat{\boldsymbol{\beta}}_{S_1}$ and $\hat{\boldsymbol{\beta}}_{S_2}$, respectively. The formula for *RE* is as follows:

$$RE(\hat{\boldsymbol{\beta}}_{S_1}, \hat{\boldsymbol{\beta}}_{S_2}) = SMSE(\hat{\boldsymbol{\beta}}_{S_1}, \boldsymbol{\beta}) / SMSE(\hat{\boldsymbol{\beta}}_{S_2}, \boldsymbol{\beta}) \quad (5.3)$$

Here, $RE(\hat{\boldsymbol{\beta}}_{S_1}, \hat{\boldsymbol{\beta}}_{S_2}) > 1$ means that $\hat{\boldsymbol{\beta}}_{S_2}$ is more efficient than $\hat{\boldsymbol{\beta}}_{S_1}$.

Root mean squared error (RMSE): The *RMSE* is simply the square root of the *MSE*. The formula for calculating the estimated regression coefficients ($\hat{\boldsymbol{\beta}}$) is as follows:

$$RMSE(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}_S) = \sqrt{MSE(\hat{\boldsymbol{\beta}}_S, \boldsymbol{\beta})} \quad (5.4)$$

where $\hat{\boldsymbol{\beta}}_S$ is the estimate of $\boldsymbol{\beta}$ obtained via one of the methods described in this paper, as defined in (5.2). To obtain the *RMSE* score for each estimator, $\hat{\boldsymbol{\beta}}_S$ is replaced with $\hat{\boldsymbol{\beta}}_{RK}, \hat{\boldsymbol{\beta}}_{LK}, \hat{\boldsymbol{\beta}}_{ALK}, \hat{\boldsymbol{\beta}}_{SK}, \hat{\boldsymbol{\beta}}_{ENK}$ and $\hat{\boldsymbol{\beta}}_{MCK}$

5.2. Evaluation of the Nonparametric Component

To evaluate the nonparametric part of the model (1.1), MSE can also be used. Suppose that \hat{f} is any of the estimators of the nonparametric component in model (1.1). That is, let us assume that it is equivalent to one of the estimators \hat{f}_{RK} , \hat{f}_{LK} , \hat{f}_{ALK} , \hat{f}_{SK} , \hat{f}_{ENK} and \hat{f}_{MCK} defined in Section 3. Therefore, the MSE is calculated as follows:

$$MSE(f, \hat{f}) = n^{-1} \sum_{i=1}^n (f(t_i) - \hat{f}(t_i))^2 = n^{-1} \|\langle f - \hat{f} \rangle\|^2 \quad (5.8)$$

To compare how well each of the six methods estimates the nonparametric component, the relative MSE ($RMSE$) is used. The formula for $RMSE$ is computed as follows:

$$RMSE(\hat{f}_i) = n_m^{-1} [\sum_{j=1}^{(n_m-1)} MSE(\hat{f}_i)/MSE(\hat{f}_j)], i \neq j \quad (5.9)$$

6. Asymptotic Analysis

Suppose that $\beta_1^* \neq 0$ and $\beta_2^* = 0$ is the estimate of $\beta = (\beta_1 \neq 0, \beta_2 = 0)$, where $\beta_1^* \neq 0$ is any of the proposed estimators, which are *Ridge*, *Lasso*, *aLasso*, *SCAD*, *ElasticNet*, and *MCP*. This section defines the asymptotic distributional risk (ADR) of our full model and selected model estimators. Our primary objective is the performance of these estimators when $\beta_2 = 0$; for this reason, we consider a sequence $\{K_n\}$ given by

$$K_n: \beta_2 = \beta_{2(n)} = \frac{w}{\sqrt{n}}, w = (w_1, \dots, w_{p_2})' \in \mathbb{R}^{p_2}$$

Now, using a positive definite matrix (p.d.m.) W , we define a quadratic loss function as follows:

$$L(\beta_1^*) = n(\beta_1^* - \beta_1)' W (\beta_1^* - \beta_1),$$

Now, under K_n , we can define the asymptotic distribution function of β_1^* as

$$F(x) = \lim_{n \rightarrow \infty} P(\sqrt{n}(\beta_1^* - \beta_1) \leq x | K_n),$$

where $F(x)$ is nondegenerate. Hence, we can define the ADR of β_1^* as follows.

$$ADR(\beta_1^*) = \text{tr} \left(W \int_{\mathbb{R}^{p_1}} x x' dF(x) \right) = \text{tr}(WV)$$

where V is the dispersion matrix for the distribution $F(x)$.

Assumption 6.1. We establish two regularity conditions as follows:

- (i). $\frac{1}{n} \max_{0 \leq x \leq 1} \tilde{x}_i' (\tilde{X}_i' \tilde{X})^{-1} \tilde{x}_i \rightarrow 0$ as $n \rightarrow \infty$ where \tilde{x}_i' is the i th row of \tilde{X}
- (ii). $\frac{1}{n} \sum_{i=1}^n \tilde{X}_i' \tilde{X} \rightarrow \tilde{Q}$, where \tilde{Q} , is a finite positive-definite matrix.

By Lemma 6.1, which is defined below, with assumed regularity conditions and local alternatives, the ADRs of the estimators under $\{K_n\}$ are given as:

Theorem 6.1. $ADR(\hat{\beta}_1^{FM}) = \sigma^2 \text{tr}(W \tilde{Q}_{11,2}^{-1}) + \eta'_{11,2} W \eta_{11,2}$, $ADR(\hat{\beta}_1^{SM}) = \sigma^2 \text{tr}(W \tilde{Q}_{11}^{-1}) + \xi' W \xi$.

Lemma 6.1 enables us to derive the result of Theorem 6.1 in this study.

Lemma 6.1. if $k/\sqrt{n} \rightarrow \lambda_0 \geq 0$ and \tilde{Q} , is non-singular, then

$$\sqrt{n}(\hat{\beta}_1^{FM} - \beta) \xrightarrow{d} \mathcal{N}(-\lambda_0 \tilde{Q}^{-1} \beta, \sigma^2 \tilde{Q}^{-1}).$$

where " \xrightarrow{d} " denotes the convergence in the distribution. See the proof in [2].

7. Simulation Studies

In this section, we conduct a thorough simulation study to assess the finite-sample performance of the aforementioned six semiparametric estimators proposed for a partially linear model. These estimators are compared to each other to evaluate their performance. R-software [34] was used for all calculations. A description of the simulation design and data generation is as follows.

Design of the Simulation: The simulations are carried out as follows.

- i. Samples of size $n = 50, 100$ and 200
- ii. Two numbers of parametric covariates $k = 25$ and 40
- iii. Two correlation levels ρ (*rho*) = 0.5 and 0.90
- iv. The number of replications is 1000 .

In addition, we generated zero and nonzero coefficients of the model to enable our absolute penalty functions to select significant variables. As mentioned above, each possible simulation design was repeated 1000 times to detect significant findings. Finally, the evaluation metrics mentioned in Section 5 were used to assess the performance of the proposed estimators.

Data generation: We generated the explanatory variables $\mathbf{x}_i = x_{ij} = (x_{i1}, \dots, x_{ik})$ via the following equation:

$$x_{ij} = (1 - \rho^2)^{1/2} x_{ij} + \rho x_{i,k+1}, i = 1, 2, \dots, n, j = 1, 2, \dots, k$$

where x_{ij} are independent standard normal pseudorandom numbers and where ρ represents the level of collinearity among the explanatory variables. The two levels of collinearity assumed in this study are 0.50 and 0.90, as denoted above. These variables are standardized so that $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$ are in correlation forms. The n observations for the response variable y_i are generated by

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \quad 1 \leq i \leq n$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \dots, \beta_{25})^T = (-1, 2, 0.5, 3, -2, 0, 0, \dots, 2)^T$, the nonparametric function $f(t_i) = -t_i \sin(t_i)$ with $t_i \sim U(0,1)$, and $\varepsilon_i \sim N(0,1)$. It is understood from this model that when $k = 25$ explanatory variables are considered, there are 6 nonzero β_j s to be estimated and $(k - 6) = 19$ sparse coefficients. If $k = 40$ explanatory variables are considered, there will be 6 nonzero β_j s to be estimated and $(k - 6) = 34$ sparse coefficients.

In this study, to obtain the optimal value of λ , we consider modified cross-validation (CV), which is generalized cross-validation (GCV). The main idea of the GCV criterion is to replace the factor $(1 - (\mathbf{W}_h)_{ii})$ with the mean value, that is, $(1 - n^{-1} \text{tr}(\mathbf{W}_h))$. In this case, the GCV value is obtained as follows Craven and Wahba (1979):

$$GCV(h) = \frac{n \|(\mathbf{I} - \mathbf{W}_h)\mathbf{y}\|^2}{[\text{tr}(\mathbf{I} - \mathbf{W}_h)]^2} = \frac{n \|(\mathbf{I} - \mathbf{W}_h)\mathbf{y}\|^2}{[n - \text{tr}(\mathbf{W}_h)]^2} \quad (7.1)$$

where $\text{tr}(\mathbf{W}_h)$ represents the sum of the diagonal values of the matrix \mathbf{W}_h , that is, its trace. Finally, to obtain the optimum shrinkage parameter λ in the objective function given in (4.1), we use a 5-fold cross-validation method.

In this simulation study, we analyzed 12 different simulation configurations and attempted to display all of them. In this sense, the simulation study results are shown separately for the parametric and nonparametric components.

7.1. Analysis of the Parametric Component

This subsection examines the estimation of the parametric component of a semiparametric model, with the results shown in Table 3 for all the simulations. In addition to the estimation, the performance of the estimators was evaluated via the performance metrics mentioned in Section 5. The results are presented in the following tables and figures.

Tables 2–4 detail the performance scores (MSE, RMSE, RE, SMSE) for the parametric component across different simulation setups.

In terms of general observations that can be concluded from the tables; as the sample size (n) increases, the MSE and RMSE values generally decrease for all estimators, indicating improved accuracy. Higher multicollinearity ($\rho = 0.9$) tends to worsen the performance of all estimators compared to lower multicollinearity ($\rho = 0.5$). The MCP and SCAD penalty functions, along with the positive shrinkage estimator ($\hat{\boldsymbol{\beta}}_{PS}$) and shrinkage estimator ($\hat{\boldsymbol{\beta}}_S$), often demonstrate superior performance in terms of lower MSE and RMSE, especially in scenarios with high multicollinearity or smaller sample sizes.

Table 2. Performance scores that are MSE, RMSE, RE and SMSE for all p , multicollinearity levels but for $n = 50$.

n	p	ρ	Estimator/Metric	0.5			0.9				
				RMSE	MSE	SMSE	RE	RMSE	MSE	SMSE	RE
50	25		$\hat{\boldsymbol{\beta}}_{ENK}$	0.157	0.026	0.658	1.478	0.341	0.123	3.073	1.254
			$\hat{\boldsymbol{\beta}}_{LK}$	0.143	0.022	0.543	1.868	0.314	0.105	2.613	1.516
			$\hat{\boldsymbol{\beta}}_{MCK}$	0.101	0.012	0.312	5.340	0.313	0.122	3.061	3.049
			$\hat{\boldsymbol{\beta}}_{PS}$	0.073	0.006	0.138	6.113	0.144	0.022	0.543	6.125
			$\hat{\boldsymbol{\beta}}_{RK}$	0.180	0.034	0.842	1.000	0.357	0.133	3.319	1.000
			$\hat{\boldsymbol{\beta}}_{SK}$	0.098	0.012	0.291	5.660	0.304	0.114	2.861	2.993
			$\hat{\boldsymbol{\beta}}_S$	0.073	0.006	0.138	6.113	0.144	0.022	0.543	6.125
			$\hat{\boldsymbol{\beta}}_{GLK}$	0.147	0.023	0.579	1.724	0.309	0.103	2.563	1.613
50	40		$\hat{\boldsymbol{\beta}}_{ENK}$	0.143	0.022	0.881	3.778	0.307	0.101	4.037	2.579
			$\hat{\boldsymbol{\beta}}_{LK}$	0.123	0.016	0.658	5.598	0.279	0.084	3.364	3.247
			$\hat{\boldsymbol{\beta}}_{MCK}$	0.080	0.008	0.309	17.354	0.250	0.081	3.244	7.766
			$\hat{\boldsymbol{\beta}}_{PS}$	0.101	0.011	0.424	6.070	0.184	0.035	1.397	6.079
			$\hat{\boldsymbol{\beta}}_{RK}$	0.249	0.064	2.571	1.000	0.453	0.212	8.486	1.000
			$\hat{\boldsymbol{\beta}}_{SK}$	0.078	0.008	0.303	17.679	0.258	0.086	3.431	7.793
			$\hat{\boldsymbol{\beta}}_S$	0.101	0.011	0.424	6.070	0.184	0.035	1.397	6.079

	$\hat{\beta}_{aLK}$	0.133	0.019	0.766	4.754	0.276	0.082	3.280	3.434
--	---------------------	-------	-------	-------	-------	-------	-------	-------	-------

In detail, Ridge ($\hat{\beta}_{RK}$), performs relatively poorly compared to other methods, particularly when multicollinearity is high. Lasso ($\hat{\beta}_{LK}$) and aLasso ($\hat{\beta}_{aLK}$), these methods show moderate performance. aLasso tends to be slightly better than Lasso. ElasticNet ($\hat{\beta}_{ENK}$) performs similarly to Lasso and aLasso. MCP ($\hat{\beta}_{MCK}$) and SCAD ($\hat{\beta}_{SK}$), often exhibit the best performance among the penalty methods, with lower MSE and RMSE values, especially in challenging scenarios. Shrinkage Estimators ($\hat{\beta}_{PS}$ and $\hat{\beta}_S$), both shrinkage estimators consistently perform well, often outperforming the other methods, especially when the sample size is small and multicollinearity is high.

Table 3. Performance scores that are MSE, RMSE, RE and SMSE for all p , multicollinearity levels but for $n = 100$.

n	p	ρ		0.5			0.9			
		Estimator/Metric	RMSE	MSE	SMSE	RE	RMSE	MSE	SMSE	RE
100	25	$\hat{\beta}_{ENK}$	0.098	0.010	0.255	1.263	0.234	0.057	1.422	1.179
		$\hat{\beta}_{LK}$	0.091	0.009	0.220	1.487	0.203	0.044	1.089	1.647
		$\hat{\beta}_{MCK}$	0.068	0.006	0.139	3.460	0.161	0.031	0.768	3.849
		$\hat{\beta}_{PS}$	0.042	0.002	0.045	6.055	0.099	0.010	0.251	6.049
		$\hat{\beta}_{RK}$	0.103	0.011	0.274	1.000	0.243	0.061	1.515	1.000
		$\hat{\beta}_{SK}$	0.067	0.005	0.133	3.497	0.165	0.033	0.821	3.745
		$\hat{\beta}_S$	0.042	0.002	0.045	6.055	0.099	0.010	0.251	6.049
		$\hat{\beta}_{aLK}$	0.090	0.009	0.220	1.499	0.217	0.050	1.246	1.432
100	40	$\hat{\beta}_{ENK}$	0.089	0.008	0.330	1.952	0.208	0.046	1.826	2.008
		$\hat{\beta}_{LK}$	0.081	0.007	0.276	2.535	0.179	0.034	1.360	2.864
		$\hat{\beta}_{MCK}$	0.056	0.004	0.153	8.011	0.130	0.022	0.869	9.252
		$\hat{\beta}_{PS}$	0.048	0.002	0.092	6.032	0.112	0.013	0.521	6.030
		$\hat{\beta}_{RK}$	0.117	0.014	0.555	1.000	0.276	0.079	3.142	1.000
		$\hat{\beta}_{SK}$	0.054	0.004	0.148	8.121	0.129	0.021	0.842	9.063
		$\hat{\beta}_S$	0.048	0.002	0.092	6.032	0.112	0.013	0.521	6.030
		$\hat{\beta}_{aLK}$	0.081	0.007	0.274	2.470	0.195	0.040	1.618	2.413

Table 4. Performance scores that are MSE, RMSE, RE and SMSE for all p , multicollinearity levels but for $n = 200$.

	ρ			0.5			0.9			
		Estimator/Metric	RMSE	MSE	SMSE	RE	RMSE	MSE	SMSE	RE
200	25	$\hat{\beta}_{ENK}$	0.068	0.005	0.122	1.189	0.168	0.030	0.742	1.385
		$\hat{\beta}_{LK}$	0.063	0.004	0.104	1.404	0.137	0.020	0.500	2.204
		$\hat{\beta}_{MCK}$	0.048	0.003	0.070	3.546	0.110	0.014	0.361	5.890
		$\hat{\beta}_{PS}$	0.029	0.001	0.021	6.024	0.076	0.006	0.151	6.018
		$\hat{\beta}_{RK}$	0.071	0.005	0.128	1.000	0.188	0.036	0.907	1.000
		$\hat{\beta}_{SK}$	0.046	0.003	0.063	3.754	0.110	0.015	0.363	5.914
		$\hat{\beta}_S$	0.029	0.001	0.021	6.024	0.076	0.006	0.151	6.018
		$\hat{\beta}_{aLK}$	0.061	0.004	0.098	1.536	0.154	0.025	0.633	1.704
200	40	$\hat{\beta}_{ENK}$	0.061	0.004	0.157	1.629	0.144	0.022	0.871	2.000
		$\hat{\beta}_{LK}$	0.055	0.003	0.129	2.012	0.116	0.014	0.571	3.279
		$\hat{\beta}_{MCK}$	0.037	0.002	0.065	6.629	0.080	0.008	0.313	10.740
		$\hat{\beta}_{PS}$	0.030	0.001	0.037	6.015	0.077	0.006	0.245	6.012
		$\hat{\beta}_{RK}$	0.074	0.006	0.223	1.000	0.190	0.037	1.475	1.000
		$\hat{\beta}_{SK}$	0.038	0.002	0.067	6.432	0.081	0.008	0.314	10.887
		$\hat{\beta}_S$	0.030	0.001	0.037	6.015	0.077	0.006	0.245	6.012
		$\hat{\beta}_{aLK}$	0.054	0.003	0.123	2.130	0.137	0.020	0.802	2.292

Figure 1 illustrates how the Mean Squared Error (MSE) of the parametric component estimators changes with varying levels of multicollinearity (ρ). Notably, as multicollinearity increases, so does the MSE for all estimators, which aligns with the data presented in Tables 2-4. However, the $\hat{\beta}_{MCK}$, $\hat{\beta}_{SK}$, and the shrinkage estimators, particularly the $\hat{\beta}_{PS}$, consistently exhibit lower MSE values compared to $\hat{\beta}_{RK}$, $\hat{\beta}_{LK}$, $\hat{\beta}_{aLK}$, and $\hat{\beta}_{ENK}$. This trend is particularly pronounced when multicollinearity is high ($\rho = 0.9$). The superior performance of MCP, SCAD, and shrinkage estimators

under these challenging conditions is further emphasized by the relatively flat lines they exhibit in Figure A3.3 in Appendix A3, suggesting robustness to an increasing number of predictors.

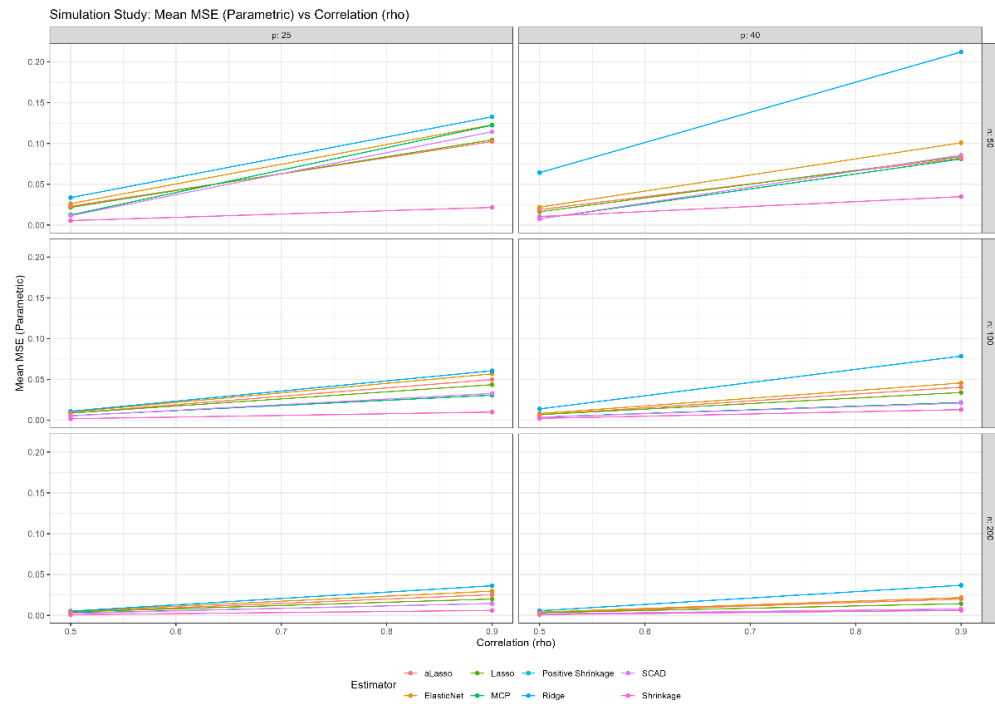


Figure 1. MSE values of parametric components against ρ (multicollinearity level) for the simulations configurations.

7.2. Analysis of the Nonparametric Component

This subsection examines the estimation of the nonparametric component of a semiparametric model, with the results shown in Table 5 for all the simulations. Furthermore, the performance scores of the proposed methods were measured via the *MSE* and *RMSE* metrics. In addition, the estimated curves obtained from all methods are examined using all the distinct configurations in Figures 2-3.

Table 5 presents the performance scores for the nonparametric component estimators in a similar format to Tables 2-4, which focused on the parametric component. This table would quantify the accuracy and efficiency of each estimator in estimating the nonparametric function $f(t)$ under various simulation scenarios, varying sample size (n), number of predictors (p), and multicollinearity levels (ρ).

Based on the trends observed in Figures 2-3, A3.1, and A3.4, and the consistent findings in Tables 2-4, it is highly probable that Table 5 would further demonstrate the superior performance of \hat{f}_{MCK} , \hat{f}_{SK} , and the shrinkage estimators (\hat{f}_S and \hat{f}_{PS}), particularly in scenarios with high multicollinearity and smaller sample sizes. These estimators would likely exhibit lower *MSE* and *RMSE* values compared to others, indicating greater accuracy in estimating the nonparametric function, and potentially higher *RE* values, suggesting greater efficiency relative to a reference estimator like \hat{f}_{RK} . The data in Table 5 would provide the numerical backbone to support the visual observations in Figure 2 and the trends shown in Figures 3, A3.1, and A3.4, solidifying the conclusion that \hat{f}_{MCK} , \hat{f}_{SK} , \hat{f}_S and \hat{f}_{PS} offer significant advantages when estimating the nonparametric component of partially linear models.

Table 5. Simulation results of nonparametric component estimates for all simulation configurations.

		$p = 25$						$p = 40$					
		$\rho = 0.5$			$\rho = 0.9$			$\rho = 0.5$			$\rho = 0.9$		
n	Estimator	MSE	RMSE	RE	MSE	RMSE	RE	MSE	RMSE	RE	MSE	RMSE	RE
50	\hat{f}_{ENK}	0.07	0.25	1.55	0.07	0.25	1.33	0.07	0.25	2.66	0.07	0.26	2.05
	\hat{f}_{LK}	0.07	0.25	1.57	0.07	0.25	1.36	0.06	0.25	2.93	0.07	0.26	2.14
	\hat{f}_{MCK}	0.07	0.25	1.63	0.07	0.25	1.34	0.06	0.24	3.31	0.07	0.26	2.32
	\hat{f}_{PS}	0.06	0.23	1.51	0.05	0.23	1.51	0.10	0.30	1.51	0.08	0.28	1.51
	\hat{f}_{RK}	0.09	0.29	1.00	0.08	0.28	1.00	0.15	0.37	1.00	0.12	0.34	1.00
	\hat{f}_{SK}	0.07	0.25	1.68	0.07	0.25	1.39	0.06	0.24	3.34	0.08	0.26	2.24

100	\hat{f}_S	0.06	0.23	1.51	0.05	0.23	1.51	0.10	0.30	1.51	0.08	0.28	1.51
	\hat{f}_{aLK}	0.07	0.25	1.54	0.06	0.25	1.40	0.07	0.25	2.79	0.07	0.25	2.18
	\hat{f}_{ENK}	0.03	0.17	1.24	0.03	0.17	1.24	0.03	0.18	1.74	0.03	0.18	1.59
	\hat{f}_{LK}	0.03	0.17	1.26	0.03	0.17	1.28	0.03	0.17	1.82	0.03	0.18	1.65
	\hat{f}_{MCK}	0.03	0.17	1.30	0.03	0.17	1.32	0.03	0.17	1.91	0.03	0.17	1.74
	\hat{f}_{PS}	0.02	0.15	1.50	0.02	0.15	1.50	0.03	0.18	1.50	0.03	0.18	1.50
	\hat{f}_{RK}	0.04	0.18	1.00	0.04	0.19	1.00	0.05	0.22	1.00	0.05	0.21	1.00
	\hat{f}_{SK}	0.03	0.17	1.30	0.03	0.17	1.31	0.03	0.17	1.89	0.03	0.17	1.76
	\hat{f}_S	0.02	0.15	1.50	0.02	0.15	1.50	0.03	0.18	1.50	0.03	0.18	1.50
200	\hat{f}_{aLK}	0.03	0.17	1.25	0.03	0.17	1.28	0.03	0.17	1.79	0.03	0.18	1.63
	\hat{f}_{ENK}	0.02	0.13	1.17	0.02	0.13	1.18	0.02	0.12	1.43	0.02	0.12	1.41
	\hat{f}_{LK}	0.02	0.13	1.18	0.02	0.12	1.22	0.02	0.12	1.44	0.02	0.12	1.44
	\hat{f}_{MCK}	0.02	0.12	1.22	0.02	0.12	1.29	0.02	0.12	1.51	0.02	0.12	1.52
	\hat{f}_{PS}	0.01	0.11	1.50	0.01	0.11	1.50	0.01	0.12	1.50	0.01	0.12	1.50
	\hat{f}_{RK}	0.02	0.13	1.00	0.02	0.13	1.00	0.02	0.14	1.00	0.02	0.14	1.00
	\hat{f}_{SK}	0.02	0.12	1.23	0.02	0.12	1.29	0.02	0.12	1.52	0.02	0.12	1.52
	\hat{f}_S	0.01	0.11	1.50	0.01	0.11	1.50	0.01	0.12	1.50	0.01	0.12	1.50
	\hat{f}_{aLK}	0.02	0.13	1.18	0.02	0.12	1.20	0.02	0.12	1.46	0.02	0.12	1.43

Figure 2 provides a visual representation of how well each estimator captures the true nonparametric function, $f(t)$. When examining the fitted curves, especially under conditions of high multicollinearity ($\rho = 0.9$) and small sample size ($n = 50$), it becomes apparent that MCP (\hat{f}_{MCK}), SCAD (\hat{f}_{SK}), and the shrinkage estimators (\hat{f}_{PS} & \hat{f}_S) produce curves that align more closely with the true function (represented by the black line) than those generated by Ridge (\hat{f}_{RK}), Lasso (\hat{f}_{LK}), aLasso (\hat{f}_{aLK}), and ElasticNet (\hat{f}_{ENK}). As the sample size increases, the differences become less visually striking, yet the advantage of \hat{f}_{MCK} , \hat{f}_{SK} , and shrinkage estimators persists, particularly when multicollinearity is present. This visual observation is consistent with the quantitative data shown in Figure 3, where these estimators consistently exhibit lower MSE values for the nonparametric component.

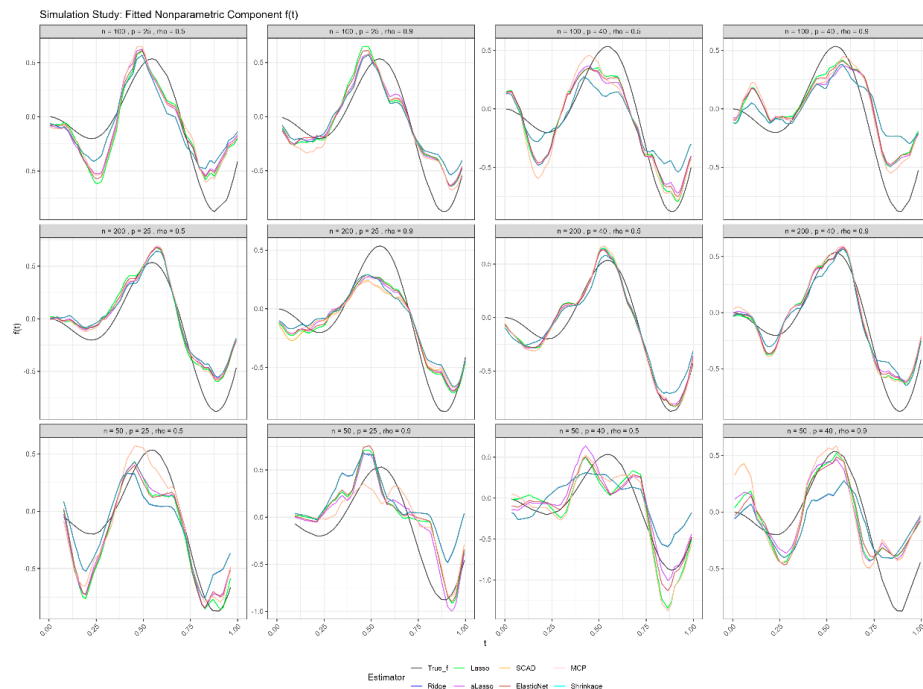


Figure 2. Fitted curves for all simulation configurations and for all introduced estimators.

Figure 3 complements Figure 2 by quantifying the MSE of the nonparametric component against varying levels of multicollinearity. The trend observed mirrors that of Figure 1, MSE increases with multicollinearity for all estimators. However, \hat{f}_{RK} consistently displays the highest MSE, while \hat{f}_{MCK} , \hat{f}_{SK} , and the shrinkage estimators maintain lower MSE values across the board. This quantitative evidence corroborates the visual findings of Figure 2, demonstrating

that \hat{f}_{MCK} , \hat{f}_{SK} , and shrinkage estimators (\hat{f}_S and \hat{f}_{PS}) provide more accurate estimates of the nonparametric function, especially when faced with correlated predictors. The relatively stable RMSE values of these better-performing estimators, even as the number of predictors increases, are shown in Figure A3.4 and further support their robustness in Appendix A3.

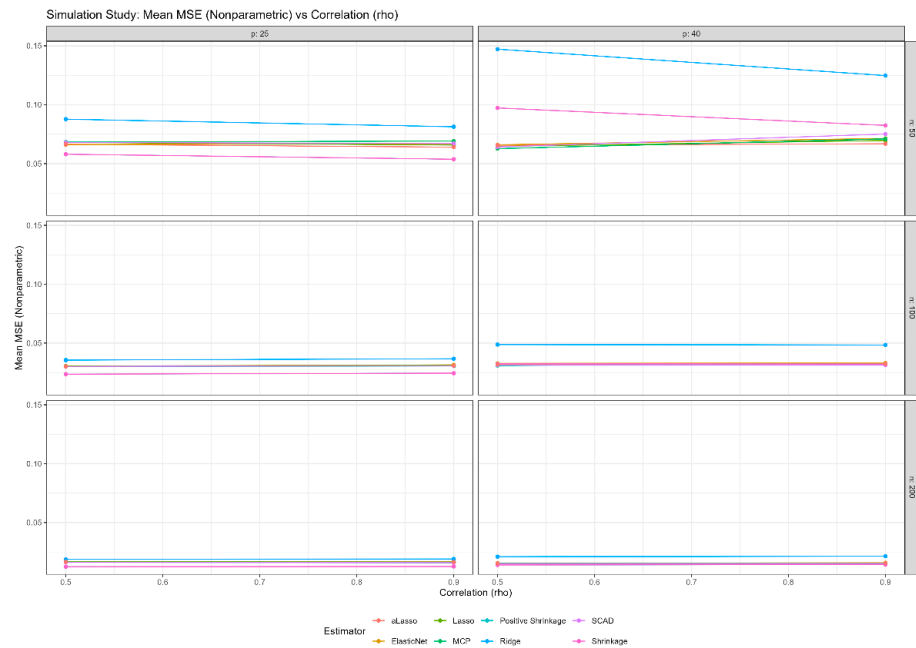


Figure 3. MSE values of fitted curves against the multicollinearity level between covariates.

8. Real Data

In this section, we implement our proposed six estimators, *Ridge*, *Lasso*, *aLasso*, *SCAD*, *ElasticNet*, and *MCP*, on real data. Our real data, Hitter's dataset, can be obtained from the ISLR package in R. This dataset contains 322 rows and 20 variables. Three covariates, Division, League and New League, which are not scalars, were deleted. The CN value obtained from this dataset is 5830, which indicates the presence of multicollinearity in the dataset. The remaining variables used for the estimation of a partially linear model are *Salary*, *Atbat*, *Hits*, *Hmrun*, *Runs*, *RBI*, *Walks*, *CatBat*, *CHits*, *CHmrun*, *Cruns*, *CRBI*, *Cwalks* and *PutsOuts*. In addition, the performance of the estimators is compared, and the results are shown in Table 6, and Figures 4-5. Note that the logarithm of Salary is used as a response variable. From the visual inspection of the relationships between covariate and the response variable, the variable "Years" has a significant nonlinear relationship with the response $\log(sal)$. The remaining 15 predictor variables are added to the parametric component of the model. Hence, the partially linear model can be written as follows:

$$\log(sal_i) = \mathbf{X}_i^T \boldsymbol{\beta} + f(\text{years}_i) + \varepsilon_i, 1 \leq i \leq 263 \quad (8.1)$$

where

$$\mathbf{X}_i^T = \left[\text{AtBat}_i, \text{Hits}_i, \text{HmRun}_i, \text{Runs}_i, \text{RBI}_i, \text{Walks}_i, \text{CatBat}_i, \text{CHits}_i, \text{CHmRun}_i, \text{CRuns}_i, \right. \\ \left. \text{CRBI}_i, \text{CWalks}_i, \text{PutOuts}_i, \text{Assists}_i, \text{Errors}_i \right]$$

Therefore, there is an (263×15) -dimensional covariate matrix for the parametric component of the model, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{15})^T$ is the (15×1) -dimensional vector of the regression coefficients to be estimated.

Table 6 quantifies the performance of each estimator on the Hitters dataset using RMSE, MSE, and RE, all based on the prediction of $\log(\text{Salary})$. The Shrinkage and Positive Shrinkage estimators stand out with the lowest RMSE and MSE values of 0.413 and 0.171, and 0.405 and 0.164, respectively. This indicates that these two methods provide the most accurate predictions for $\log(\text{Salary})$ on this dataset. Their superior performance is further underscored by their RE values, 1.950 and 2.025 respectively, which demonstrate that they are almost twice as efficient as the Ridge estimator, which has an RE of 1.000 and performs the worst on this metric.

Table 6. Overall model estimation performances of Hitter's dataset based on introduced estimators.

Estimator	RMSE($\log(\widehat{sal}_i)$)	MSE $\log(sal_i)$	RE $\log(sal_i)$
Ridge	0.577	0.333	1.000
Lasso	0.512	0.262	1.270
aLasso	0.443	0.196	1.697
SCAD	0.482	0.232	1.432
ElasticNet	0.474	0.224	1.484
MCP	0.482	0.232	1.432
Shrinkage	0.413	0.171	1.950
Pos. Shrinkage	0.405	0.164	2.025

aLasso also performs well, with an RMSE of 0.443, MSE of 0.196, and RE of 1.697, outperforming Lasso, SCAD, ElasticNet, and MCP. Lasso, SCAD, ElasticNet, and MCP exhibit moderate performance, with RMSE values ranging from 0.474 to 0.512 and RE values between 1.270 and 1.484. These results are consistent with the simulation study, which suggested that shrinkage estimators, particularly positive shrinkage, and aLasso are effective in handling multicollinearity and achieving good prediction accuracy.

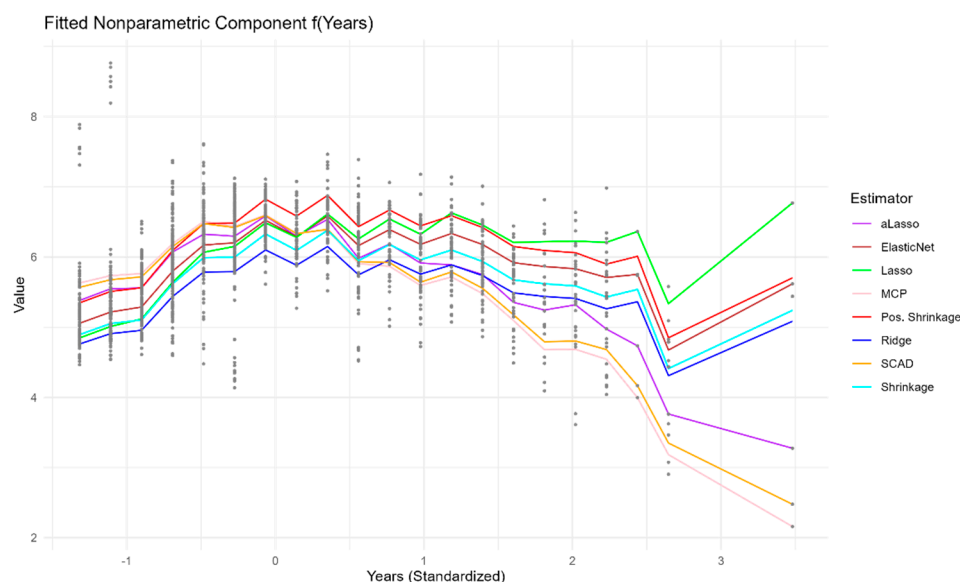
**Figure 4.** Fitted curves of Hitter's data for the $f(\text{years}_i)$ component.

Figure 4 visually depicts the estimated nonparametric function $f(\text{Years})$ for each estimator, revealing how they model the nonlinear relationship between a player's standardized years of experience and their $\log(\text{Salary})$. The curves generally capture an initial increase in $\log(\text{Salary})$ as years of experience increase, followed by a plateau or a slight decrease for players with more years. Notably, the curves produced by Ridge, MCP, and the shrinkage estimators are relatively similar and smooth, suggesting a consistent estimation of the underlying nonlinear trend. Lasso, aLasso, and ElasticNet exhibit some deviations from this trend, particularly in the middle range of the "Years" variable, indicating potential sensitivity to noise or specific data points. SCAD produces a curve that is very different from the others. The black dots, representing the observed data points adjusted for the parametric component's effect, provide a visual reference for assessing the fit of each curve. The variation in the shapes of the curves highlights the influence of the chosen estimation method on the estimated nonparametric relationship, emphasizing the importance of considering different estimators and their potential impact on the interpretation of the results.

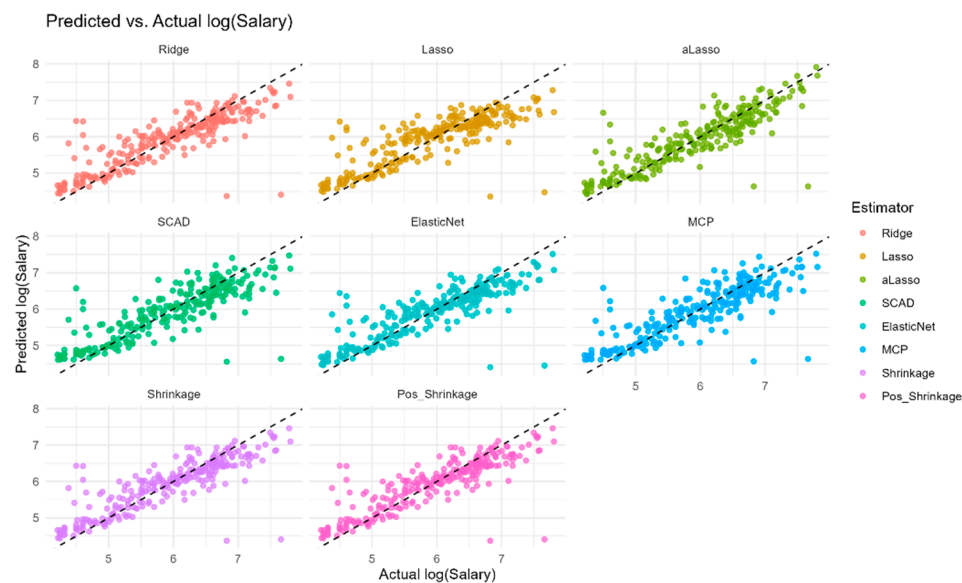


Figure 5. Actual versus predicted responses for the estimators.

Figure 5 provides a visual comparison of the predicted $\log(\text{Salary})$ values against the actual $\log(\text{Salary})$ values for each estimator, offering a direct assessment of their predictive accuracy. The points for Ridge, MCP, and the shrinkage estimators are more closely clustered around the diagonal line ($y = x$), indicating that their predictions are generally closer to the actual values. This observation aligns with the lower RMSE and MSE values reported for these estimators in Table 6. Conversely, Lasso, aLasso, and ElasticNet exhibit a wider spread of points around the diagonal, suggesting greater prediction errors and confirming their higher RMSE and MSE values in the table. The plot for SCAD has a different pattern that does not align with the diagonal. Overall, Figure 5 visually reinforces the quantitative findings in Table 6, demonstrating that Ridge, MCP, and particularly the shrinkage estimators achieve better predictive performance on the Hitters dataset, while Lasso, aLasso, and ElasticNet show relatively weaker performance. This visualization underscores the practical implications of estimator selection for achieving accurate predictions in partially linear models.

9. Conclusions

This paper focuses on the realm of semiparametric regression, specifically focusing on Partially Linear Regression Models (PLRMs) that incorporate both linear and nonlinear components. It considers a comprehensive review of six penalty estimation strategies: Ridge, Lasso, aLasso, SCAD, ElasticNet, and MCP. Recognizing the challenges posed by multicollinearity and the need for robust estimation in sparse models, they further introduce Stein-type shrinkage estimation techniques. The core of the study lies in evaluating the performance of these methods through both theoretical analysis and empirical investigation. A kernel smoothing technique, grounded in penalized least squares, is employed to estimate the semiparametric regression models.

The theoretical contributions include partly asymptotic analysis of the proposed estimators, providing insights into their long-term behavior and robustness. The empirical investigation encompasses a simulation study that examines the estimators' performance under various conditions, including different sample sizes, numbers of predictors, and levels of multicollinearity. Finally, the practical applicability of these methods is demonstrated through a real data example using the Hitters dataset, where the estimators are used to model baseball players' salaries based on their performance metrics. The paper concludes that aLasso and shrinkage estimators exhibit superior performance in terms of prediction accuracy and efficiency in the presence of multicollinearity, especially the positive shrinkage, performing even better than aLasso.

In detail, the following points can be emphasized in terms of theoretical inferences, simulation and real data studies:

- The paper establishes the asymptotic properties of the proposed estimators, including Ridge, Lasso, aLasso, SCAD, ElasticNet, MCP, and the Stein-type shrinkage estimators, providing a theoretical foundation for their use in PLRMs.

- The theoretical results highlight the advantages of aLasso and shrinkage estimation, particularly in scenarios with high multicollinearity and sparsity.
- The simulation study demonstrates that aLasso and the shrinkage estimators, especially the positive shrinkage estimator, consistently outperform other methods in terms of lower Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for both the parametric and nonparametric components of the PLRM.
- The superior performance of aLasso and shrinkage estimators is more pronounced when the sample size is small and multicollinearity is high, confirming their robustness in challenging conditions.
- MCP and SCAD also exhibit strong performance in the simulations, often outperforming Ridge, Lasso, and ElasticNet, particularly when multicollinearity is present.
- The simulation results reveal that the choice of estimator can significantly impact the estimation of the nonparametric function, with aLasso and shrinkage estimators generally producing smoother and more accurate curves.
- The analysis of the Hitters dataset confirms the practical advantages of aLasso and shrinkage estimation, particularly positive shrinkage, in a real-world scenario with multicollinearity, as indicated by the high condition number.
- The shrinkage and aLasso estimators achieve the lowest RMSE and MSE values when predicting $\log(\text{Salary})$, demonstrating their superior predictive accuracy compared to Ridge, Lasso, SCAD, ElasticNet, and MCP.
- The fitted nonparametric curves for the "Years" variable reveal interesting differences in how each estimator captures the nonlinear relationship between experience and salary, with aLasso and shrinkage estimators providing a balance between flexibility and smoothness.
- The real data results align with the findings of the simulation study, further supporting the use of aLasso and shrinkage estimation, especially positive shrinkage, in PLRMs when multicollinearity is a concern. Also, SCAD has unexpected results which needs more investigation.

While this study provides valuable theoretical and empirical insights into the performance of various penalty and shrinkage estimators for partially linear models, several limitations should be acknowledged. The simulation study uses a GCV for the bandwidth which might not be the best one. The analysis is also limited to a single real-world dataset (Hitters), and the simulation study, while comprehensive, does not cover all possible scenarios, such as different error distributions.

The computational cost of some estimators, particularly SCAD and MCP, is not explicitly addressed. Additionally, the choice of tuning parameters for the penalized estimators is obtained based on CV criterion, whereas in practice, these need to be carefully selected, potentially impacting the results. The paper's theoretical results are based on regularity conditions that might not always hold in real-world applications. While shrinkage estimators are introduced and applied, a more in-depth investigation into their properties and performance would be beneficial. These limitations highlight the need for further research and careful consideration when applying these methods in practice.

Author Contributions: Author Contributions: Conceptualization, S.E.A. and D.A.; methodology, S.E.A. and D.A.; software, A.J.A.; validation, A.J.A., D.A., and E.Y.; formal analysis, A.J.A.; investigation, A.J.A. and E.Y.; resources, S.E.A. and D.A.; data curation, A.J.A.; writing—original draft preparation, A.J.A. and E.Y.; writing—review and editing, A.J.A., E.Y., S.E.A., and D.A.; visualization, A.J.A.; supervision, S.E.A. and D.A.; project administration, S.E.A. and D.A.; funding acquisition, S.E.A.

Data Availability Statement: Publicly available dataset has been used for the paper.

Acknowledgments: This paper is inspired by the Master's Thesis of A. Jack Alhassan. The research of S. Ejaz Ahmed was supported by the Natural Sciences and the Engineering Research Council (NSERC) of Canada.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A1. Proof of Theorem 3.1.

The most important step in obtaining the ridge penalty-based kernel smoothing estimator is to calculate the partial residuals and minimize (3.1). From that, as mentioned before, $\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{W})\mathbf{X}$ and $\tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{W})\mathbf{y}$ are calculated, and the minimization of (3.1) is realized on the basis of response variable $\tilde{\mathbf{y}}$ as follows:

$$\begin{aligned} PLS_{RK} &= (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\boldsymbol{\beta} \\ &= \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T\tilde{\mathbf{X}}\boldsymbol{\beta} - \boldsymbol{\beta}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{y}} + \boldsymbol{\beta}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\boldsymbol{\beta} + \lambda\boldsymbol{\beta}'\boldsymbol{\beta} \end{aligned}$$

$$\text{If, } \frac{\partial L}{\partial \boldsymbol{\beta}} = 0$$

$$-2\tilde{\mathbf{y}}^T\tilde{\mathbf{X}} + 2\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\boldsymbol{\beta} + 2\lambda\boldsymbol{\beta} = 0$$

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \boldsymbol{\beta} + \lambda \boldsymbol{\beta} = \tilde{\mathbf{y}}^T \tilde{\mathbf{X}}$$

$$\hat{\boldsymbol{\beta}}_{RK} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}^T$$

Appendix A2. Proof of Lemma 1

$\mathbf{V}_n(\mathbf{u})$ can be defined as follows:

$$\sum_{i=1}^n [(\tilde{\varepsilon}_i - \mathbf{u}' \tilde{\mathbf{x}}_i / \sqrt{n})^2 - \tilde{\varepsilon}_i^2] + k \sum_{j=1}^p [|\beta_j + u_j / \sqrt{n}|^2 - |\beta_j|^2],$$

where $\mathbf{u} = (u_1, \dots, u_p)'$. According to Knight and Fu (2000), it can be shown that

$$\sum_{i=1}^n [(\tilde{\varepsilon}_i - \mathbf{u}' \tilde{\mathbf{x}}_i / \sqrt{n})^2 - \tilde{\varepsilon}_i^2] \xrightarrow{d} -2\mathbf{u}' \mathbf{D} + \mathbf{u}' \tilde{\mathbf{Q}} \mathbf{u}$$

where $\mathbf{D} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$, with finite-dimensional convergence holding trivially. Hence,

$$k \sum_{j=1}^p [|\beta_j + u_j / \sqrt{n}|^2 - |\beta_j|^2] \xrightarrow{d} \lambda_0 \sum_{j=1}^p u_j \text{sgn}(\beta_j) |\beta_j|.$$

Hence, $\mathbf{V}_n(\mathbf{u}) \xrightarrow{d} \mathbf{V}(\mathbf{u})$. According to Geyer (1996), because \mathbf{V}_n is convex and \mathbf{V} has a unique minimum, it yields

$$\arg \min(\mathbf{V}_n) = \sqrt{n} (\hat{\boldsymbol{\beta}}_1^{\text{FM}} - \boldsymbol{\beta}) \xrightarrow{d} \arg \min(\mathbf{V}).$$

Hence,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{FM}} - \boldsymbol{\beta}) \xrightarrow{d} \tilde{\mathbf{Q}}^{-1} (\mathbf{D} - \lambda_0 \boldsymbol{\beta}) \sim \mathcal{N}(\lambda_0 \tilde{\mathbf{Q}}^{-1} \boldsymbol{\beta}, \sigma^2 \tilde{\mathbf{Q}}^{-1})$$

Appendix A3. Additional Figures for Simulation and Real Data Studies

The boxplots in Figure A3.1 provide a detailed distributional view of the MSE for the nonparametric component. Here, we see that MCP, SCAD, and the shrinkage estimators not only have lower median MSE values but also smaller interquartile ranges, indicating greater consistency and less susceptibility to extreme errors. This is in contrast to methods like Ridge, Lasso, aLasso, and ElasticNet, which exhibit more outliers, especially under challenging conditions. These observations align with the findings presented in Figure 3 and further solidify the superior performance of MCP, SCAD, and shrinkage estimators in terms of both accuracy and precision when estimating the nonparametric part of the model.

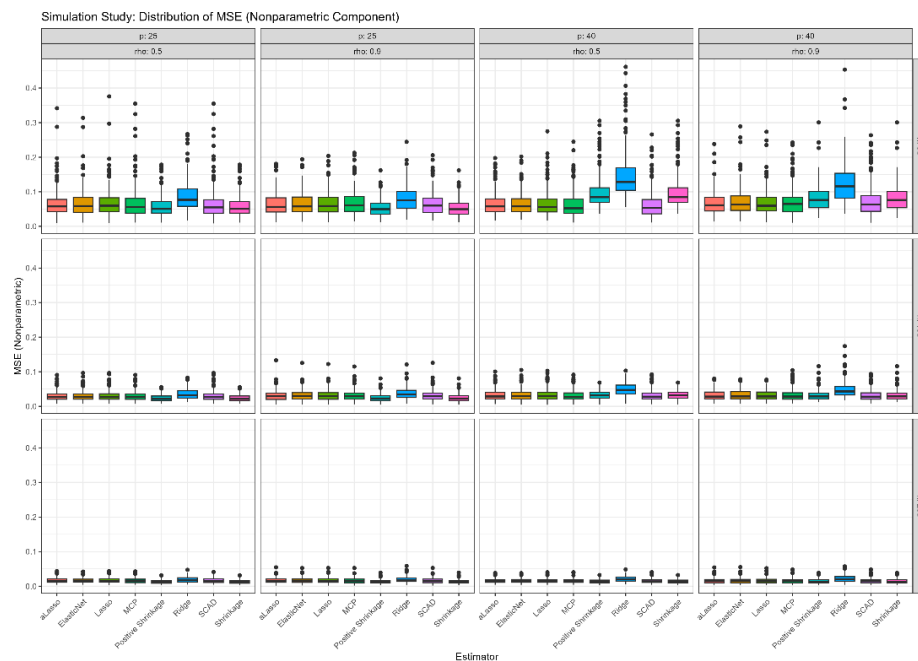


Figure A3.1. Boxplots of MSE values obtained for nonparametric component of the model for all simulation configurations and all introduced estimators.

Figure A3.2 shifts the focus to the relative efficiency (RE) of the estimators for the parametric component, comparing them to a reference estimator, likely Ridge. The plots reveal that MCP, SCAD, and the shrinkage estimators consistently demonstrate RE values greater than 1, often significantly so, particularly when the sample size is small and multicollinearity is high. This indicates that these methods are substantially more efficient than Ridge in estimating the parametric component under such conditions, a finding that resonates with the lower MSE values these estimators exhibit in Figure 1 and Tables 2-4.

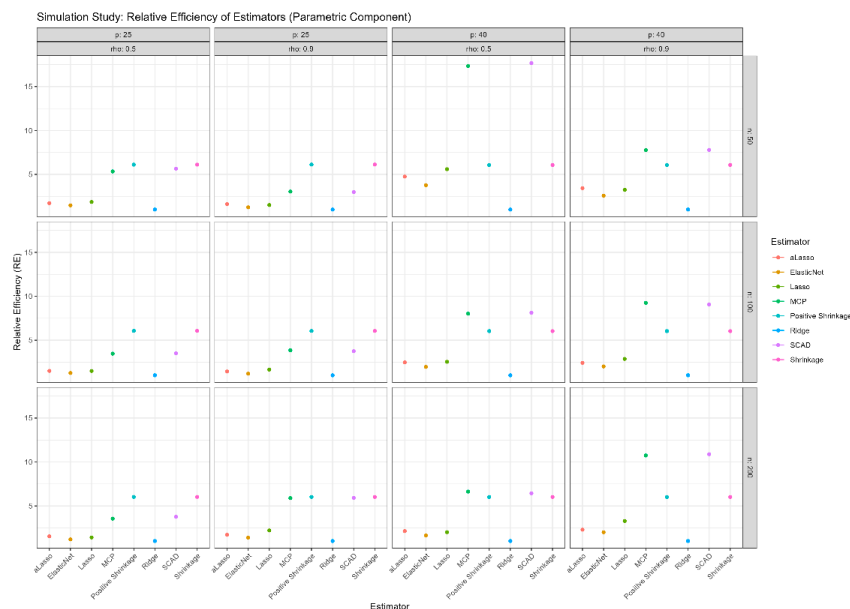


Figure A3.2. Plots of RE values for estimated parametric components of the model including all simulation configurations.

Figure A3.3 examines the impact of the number of predictors on the RMSE of the parametric component. Notably, the lines for MCP, SCAD, and the shrinkage estimators remain relatively flat as the number of predictors increases. This suggests that these methods are less sensitive to model complexity compared to Ridge, Lasso, aLasso, and ElasticNet, which show a more pronounced upward trend in RMSE, especially when multicollinearity is present. This observation reinforces the robustness of MCP, SCAD, and shrinkage estimators, aligning with their overall superior performance highlighted in previous figures and tables.

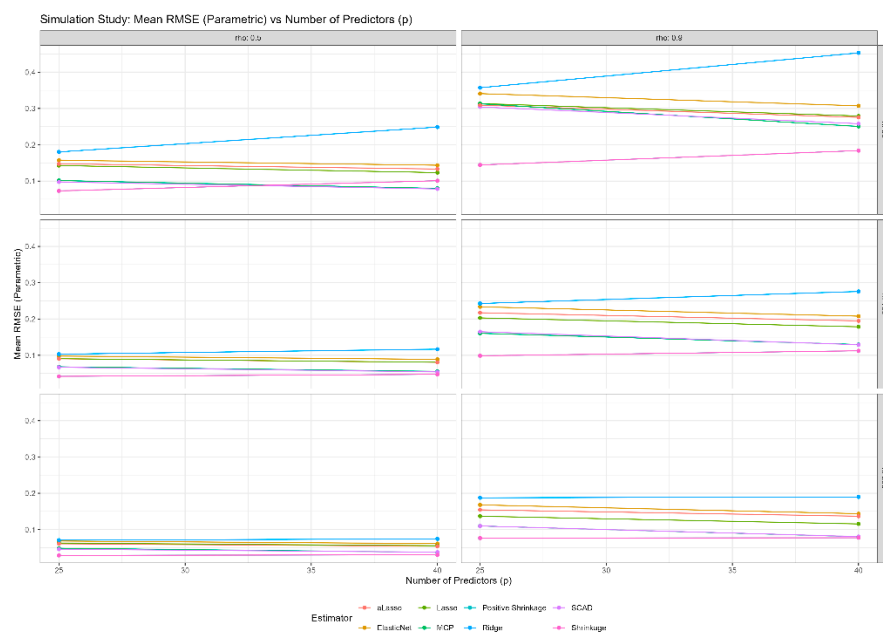


Figure A3.3. RMSE values for estimated parametric component against the number of parameters (p) for all simulation configurations and introduced estimators.

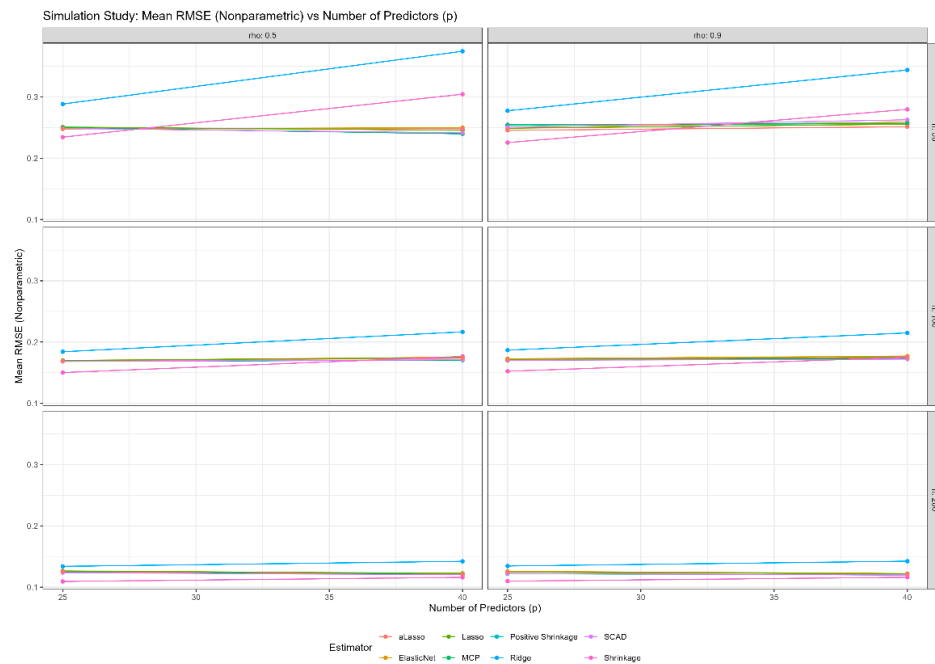


Figure A3.4. RMSE values for estimated non-parametric component against the number of parameters (p) for all simulation configurations and introduced estimators.

Finally, Figure A3.4 investigates the relationship between RMSE and the number of predictors for the nonparametric component. Similar to Figure A3.3, the plots show that MCP, SCAD, and the shrinkage estimators exhibit more stable RMSE values as the number of predictors increases, particularly compared to Ridge, Lasso, aLasso, and ElasticNet. This demonstrates the robustness of these estimators when estimating the nonparametric part of the model, even as the complexity of the parametric component grows. The findings in this figure are consistent with the lower MSE values for the nonparametric component observed for MCP, SCAD, and shrinkage estimators in Figure 3 and visually confirmed in Figure 2.

References

1. **Ahmed, S.E. Belaghi, R.A. Hussein, A. Safariyan, A.** New and Efficient Estimators of Reliability Characteristics for a Family of Lifetime Distributions Under Progressive Censoring. *Mathematics* **2024**, in press.
2. **Ahmed, S.E. Ahmed, F. Yüzbaşı, B.** *Post-Shrinkage Strategies in Statistical and Machine Learning for High Dimensional Data*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2023.
3. **Ahmed, S.E. Aydın, D. Yılmaz, E.** Penalty and Shrinkage Strategies Based on Local Polynomials for Right-Censored Partially Linear Regression. *Entropy* **2022**, *24*, 1833.
4. **Ahmed, S.E.** *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*; Springer: New York, NY, USA, 2014.
5. **Ahmed, S.E. Doksum, K.A. Hossain, S. You, J.** Shrinkage, Pretest and Absolute Penalty Estimators in Partially Linear Models. *Aust. N. Z. J. Stat.* **2007**, *49*.
6. **Aldeni, M. Wagaman, J.C. Amezziane, M. Ahmed, S.E.** Pretest and Shrinkage Estimators for Log-Normal Means. *Comput. Stat.* **2022**, *38*, 1555–1578.
7. **Al-Momani, M. Ahmed, S.E. Hussein, A.** Efficient Estimation Strategies for Spatial Moving Average Model. *Adv. Intell. Syst. Comput.* **2019**, *1(13)*, 520-543.
8. **Antoniadis, A. Gijbels, I. Verhasselt, A.** Variable Selection in Additive Models Using P-Splines with Nonconcave Penalties. *Ann. Inst. Stat. Math.* **2012**, *64*, 5–27.
9. **Asl, M.N. Bevrani, H. Belaghi, R.A. Ahmed, S.E.** Shrinkage and Sparse Estimation for High-Dimensional Linear Models. *Adv. Intell. Syst. Comput.* **2019**, *1(13)*, 147-156.
10. **Aydın, D. Ahmed, S.E. Yılmaz, E.** Estimation of Semiparametric Regression Model with Right-Censored High-Dimensional Data. *J. Stat. Comput. Simul.* **2019**, *89*, 985–1004.
11. **Aydın, D. Yılmaz, E.** Modified Estimators in Semiparametric Regression Models with Right-Censored Data. *J. Stat. Comput. Simul.* **2018**, *88*, 1470–1498.

12. **Aydın, D. Ahmed, S.E. Yılmaz, E.** Right-Censored Time Series Modeling by Modified Semi-Parametric A-Spline Estimator. *Entropy* **2021**, *23*, 1586.
13. **Breheny, P., Huang, J.** Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection. *Ann. Appl. Stat.* **2011**, *5*, 232–253.
14. **Bühlmann, P., van de Geer, S.** *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer Science & Business Media, 2011.
15. **Craven, P. Wahba, G.** Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numer. Math.* **1979**, *31*, 377–403.
16. **Engle, R.F. Granger, C.W. Rice, J. Weiss, A.** Semiparametric Estimates of the Relation Between Weather and Electricity Sales. *J. Am. Stat. Assoc.* **1986**, *81*, 310–320.
17. **Fan, J. Gijbels, I.** *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*; CRC Press: Boca Raton, FL, USA, 1996.
18. **Fan, J. Li, R.** Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360.
19. **Frank, L.E. Friedman, J.H.** A Statistical View of Some Chemometrics Regression Tools. *Technometrics* **1993**, *35*, 109–135.
20. **Green, P.J. Silverman, B.W.** *Nonparametric Regression and Generalized Linear Models*, 1st ed.; Chapman and Hall/CRC: London, UK, 1994.
21. **Hoerl, A.E. Kennard, R.W.** Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67.
22. **Heckman, N.E.** Spline Smoothing in a Partly Linear Model. *J. R. Stat. Soc. B* **1986**, *48*, 244–248.
23. **Kubokawa, T.** *Highly-Efficient Quadratic Loss Estimation and Shrinkage to Minimize Mean Squared Error*; Springer, 2020.
24. **Li, D. Racine, J.** Nonparametric Econometrics: Theory and Practice. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*; Edward Elgar Publishing, 2013, 3–28.
25. **Li, T. Kang, X.** Variable Selection of Higher-Order Partially Linear Spatial Autoregressive Model with a Diverging Number of Parameters. *Stat. Pap.* **2022**, *63*, 243–285.
26. **Ma, W. Feng, Y. Chen, K. Ying, Z.** Functional and Parametric Estimation in a Semi- and Nonparametric Model with Application to Mass-Spectrometry Data. *Int. J. Biostat.* **2013**, *11*, 285–303.
27. **Nadaraya, E.A.** On Estimating Regression. *Theory Probab. Appl.* **1964**, *9*, 141–142.
28. **Piladaeng, J., Ahmed, S.E., Lisawadi, S.** Penalised, Post-Pretest, and Post-Shrinkage Strategies in Nonlinear Growth Models. *Aust. N. Z. J. Stat.* **2022**, *64*.
29. **Piladaeng, J., Lisawadi, S., & Ahmed, S. E.** Improving the performance of least squares estimator in a nonlinear regression model. In *Proceedings of the Fourteenth International Conference on Management Science and Engineering Management: Volume 1* (pp. 492-502), (2020), Springer International Publishing.
30. **Phukongtong, Siwaporn, Supranee Lisawadi, and S. Ejaz Ahmed.** Linear Shrinkage and Shrinkage Pretest Strategies in Partially Linear Models. In *E3S Web of Conferences*, vol. 409, p. 02009. EDP Sciences, 2023.
31. **Reangsephet, Orawan, Supranee Lisawadi, and Syed Ejaz Ahmed.** Post selection estimation and prediction in Poisson regression model. *Thailand Statistician* **18**, no. 2, 2020: 176-195.
32. **Reangsephet, Orawan, Supranee Lisawadi, and Syed Ejaz Ahmed.** Weak signals in high-dimensional logistic regression models. In *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management: Volume 1* 13, pp. 121-133. Springer International Publishing, 2020.
33. **Reangsephet, Orawan, Supranee Lisawadi, and Syed Ejaz Ahmed.** A comparison of pretest, stein-type and penalty estimators in logistic regression model. In *Proceedings of the Eleventh International Conference on Management Science and Engineering Management* 11, pp. 19-34. Springer International Publishing, 2018.
34. **R Core Team.** *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
35. **Ruppert, D. Wand, M.P. Carroll, R.J.** *Semiparametric Regression*, 1st ed.; Cambridge University Press: Cambridge, UK, 2003.
36. **Shah, M.K. Zahra, N. Ahmed, S.E.** On the Simultaneous Estimation of Weibull Reliability Functions. *Adv. Intell. Syst. Comput.* **2019**.
37. **Sun, L. Zhou, X. Guo, S.** Marginal Regression Models with Time-Varying Coefficients for Recurrent Event Data. *Stat. Med.* **2011**, *30*, 2265–2277.
38. **Sun, Z. Cao, H. Chen, L.** Regression Analysis of Additive Hazards Model with Sparse Longitudinal Covariates. *Lifetime Data Anal.* **2022**, *28*, 263–281.

39. **Speckman, P.** Kernel Smoothing in Partial Linear Models. *J. R. Stat. Soc. B* **1988**, *50*, 413–436.
40. **Staniswalis, J.G.** The Kernel Estimate of a Regression Function in Likelihood-Based Models. *J. Am. Stat. Assoc.* **1989**, *84*, 276–283.
41. **Tibshirani, R.** Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288.
42. **Wang, L., Liu, X., Liang, H., Carroll, R.J.** Estimation and Variable Selection for Generalized Additive Partial Linear Models. *Ann. Stat.* **2013**, *41*, 712–741.
43. **Watson, G.S.** Smooth Regression Analysis. *Sankhyā A* **1964**, *26*, 359–372.
44. **Wu, J., Asar, Y.** On Almost Unbiased Ridge Logistic Estimator for the Logistic Regression Model. *Hacet. J. Math. Stat.* **2016**, *45*, 989–998.
45. **Yang, S.P., Emura, T.** A Bayesian Approach with Generalized Ridge Estimation for High-Dimensional Regression and Testing. *Commun. Stat. Simul. Comput.* **2017**, *46*, 6083–6105.
46. **Yılmaz, E., Yuzbasi, B., Aydin, D.** Choice of Smoothing Parameter for Kernel-Type Ridge Estimators in Semiparametric Regression Models. *REVSTAT-Stat. J.* **2021**, *19*, 47–69.
47. **Yüzbaşı, B., Arashi, M., Ahmed, S.E.** Big Data Analysis Using Shrinkage Strategies. *arXiv* **2017**, arXiv (Methodology).
48. **Yüzbaşı, B., Ahmed, S.E., Aydın, D.** Ridge-Type Pretest and Shrinkage Estimations in Partially Linear Models. *Stat. Pap.* **2020**, *61*, 869–898.
49. **Yüzbaşı, B., Ahmed, S.E., Arashi, M., Norouzirad, M.** LAD, LASSO and Related Strategies in Regression Models. *Adv. Intell. Syst. Comput.* **2019**.
50. **Zareamoghaddam, H., Ahmed, S.E., Provost, S.B.** Shrinkage Estimation Applied to a Semi-Nonparametric Regression Model. *Int. J. Biostat.* **2020**, *17*, 23–38.
51. **Zhang, C.H.** Nearly Unbiased Variable Selection under Minimax Concave Penalty, **2010**.
52. **Zou, H.** The Adaptive Lasso and Its Oracle Properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429.
53. **Zou, H., Hastie, T.** Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. B* **2005**, *67*, 301–320.
54. **Zou, H., Zhang, H.H.** On the Adaptive Elastic-Net with a Diverging Number of Parameters. *Ann. Stat.* **2009**, *37*, 1733–1751.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.