IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 22, Issue 5, Ser. III (Sep. – Oct. 2026), PP 34-39
www.iosrjournals.org

Estimation of the Partial Linear Single-lndex Model Using Spline-Based and Robust Local Linear Huber Methods: A Comparative Study

NOOR ABDUL-KAREEM FAYADH
(Department of Statistics , College of Administration & Economics University Of Sumer ,lraq

Abstract:
Background: This document investigates the partially linear single-index semiparametric model, a statistical structure that incorporates both linear and nonparametric elements, allowing for the examination of both linear and nonlinear connections between explanatory variables and a response variable.
Materials and Methods The research outlines the theoretical basis of the model and approaches for estimating its parts, with an emphasis on two particular techniques: spline-based estimation, which uses spline functions to approximate the nonparametric aspect; and robust local linear estimation founded on the Huber loss function, which seeks to improve estimation accuracy and reduce the impact of outliers in the dataset.
Results: . Additionally, the research performs a statistical assessment of these two techniques by analyzing their precision and effectiveness in estimating the parameters of the model and its nonlinear aspect under a variety of statistical scenarios
Conclusion: , The study aiming to identify the appropriateness and performance traits of each technique for estimating the suggested model.

Key Word: *Partial Linear Single-lndex Model , spline-based estimation, robust local linear estimation, Huber function, nonparametric estimation.*

Date of Submission: 27-10-2026 Date of Acceptance: 06-11-2026

I. Introduction

The single-index partially linear model is a framework that combines linear and non-parametric components, allowing for the representation of both linear and non-linear relationships between explanatory variables and a response variable. This model relies on a single linear index to reduce the dimensionality of the non-parametric component, while preserving the flexibility of the non-parametric function in capturing complex relationships within the data. The significance of this model lies in its ability to strike a balance between the simplicity of linear models and the flexibility of non-parametric models, making it well-suited for analyzing data that contains relationships of varying natures.

The model consists of two fundamental components: the first is a linear component based on a set of explanatory variables and parametric coefficients that are to be estimated, while the second is a non-parametric component based on an unknown function. This function is expressed through a linear index, obtained by multiplying the vector of explanatory variables by a vector of index coefficients. This structure helps reduce the dimensionality used to represent the non-linear relationship, while maintaining the model's flexibility in describing various patterns within the data

II. Research problem

The research problem stems from the fact that the partially linear single-index semiparametric model combines both linear and non-linear components. This necessitates the use of estimation methods capable of accurately and efficiently capturing the nature of the relationships between the explanatory variables and the response variable. The challenge lies in selecting an appropriate estimation method that ensures high accuracy in estimating the parameters of the linear component and the non-parametric function particularly when faced with varying sample sizes, different levels of variance, and the presence of outliers in the data.

Consequently, there is a need to examine and compare two estimation methods for this model: spline-based estimation and robust local linear estimation (using the Huber loss function) .

III. Research objective

This study investigates the partially linear single-index semiparametric framework from both theoretical aspects and practical applications. It explores two techniques for estimating the elements of the model: estimation based on splines and robust local linear estimation utilizing the Huber loss function. This research focuses on the application of two estimation methods Spline- Based and Robust Local Linear Huber Method. It aims to compare the performance of these two methods through simulation experiments, while examining how the accuracy of the estimation is affected by varying sample sizes and different parameter levels. Furthermore, the research seeks to determine the efficiency of each method in estimating model parameters and the non-parametric function, and to identify which method achieves the best statistical performance based on established evaluation criteria.

Partial Linear Single-lndex Model: Regression models are used to study the relationship between a response variable and a set of explanatory variables. However, some relationships between these variables may be non-linear, rendering traditional linear models inadequate for representing them accurately. To address this issue, the partially linear single-index semi-parametric model offers a flexible approach by integrating both linear and non-parametric components within a single statistical framework; it allows for the estimation of linear effects for certain variables, while representing the non-linear effects of other variables through an unknown function [4] [2]

$$y_i = \beta_0 + X_i^T \beta + g(Z_i^T \alpha) + \in_i \qquad (1)$$

Here, y_i represents the dependent variable[14] , X_i^T is the vector of explanatory variables included in the linear component, and β is the vector of linear parameters. Z_i^T denotes the vector of variables included in the single-index component, α is the vector of single-index parameters, $g(Z_i^T \alpha)$ is an unknown non-parametric function, and $\in_i$ represents the random error term[3] [7]

IV. Model estimation methods

Spline- Based Estimation: The estimation method using spline basis functions relies on approximating the non-parametric component of a semi-parametric regression model—specifically one involving a linear part and a single index by representing the unknown function in terms of a set of basis functions. This allows for the estimation of the non-linear relationship between the dependent variable and the single index without the need to pre-specify a functional form. The mathematical model is given by the following expression: [13]

$$y_i = \beta_0 + X_i^T \beta + g(Z_i^T \alpha) + \in_i \qquad (2)$$

To facilitate the estimation process, an individual indicator is defined as follows:

$$T_i = Z_i^T \alpha \qquad (3)$$

Thus, the model can be rewritten as follows:

$$y_i = \beta_0 + X_i^T \beta + g(T_i) + \in_i \qquad (4)$$

The non-parametric function $g(T_i)$is approximated using a set of spline basis functions, expressed as follows[8]

$$g(T_i) \approx \sum_{K=1}^{K} \gamma_K \; B(T_i)^T \gamma \qquad (5)$$

:

By substituting into the original model, we obtain

$$y_i = \beta_0 + X_i^T\beta + B(T_i)^T\gamma + \in_i \qquad (6)$$

To ensure stable and smooth estimates of a non-parametric function, a penalized least squares function is used, which takes the following form[13]

$$Q(\beta_0, \beta, \gamma, \alpha) = \sum_{i=1}^{n}[y_i - \beta_0 - X_i^T\beta - B(T_i)^T\gamma]^2 + \lambda\gamma^T P\gamma \quad (7)$$

where λ represents the smoothing parameter and P represents the penalty matrix; the parameter estimates are obtained by minimizing the aforementioned objective function, subject to a constraint that defines the single index $\alpha^T\alpha = 1$ The model can be written in matrix form as follows[

$$Y = D_\alpha\theta + \varepsilon \qquad (8)$$

Whereas

$D_\alpha = [1, X, B_\alpha]$

B_α represents a matrix of spline basis function values, while θ represents a vector of parameters[6]

$$\theta = \begin{pmatrix} \beta_0 \\ \beta \\ \gamma \end{pmatrix} \qquad (9)$$

By differentiating the objective function with respect to θ and setting the derivative equal to zero, we obtain the following estimation equations[5]

$$(D_\alpha^T D_\alpha + \lambda P^*)\hat{\theta} = D_\alpha^T Y \qquad (10)$$

Thus, the final form of the estimator is as follows:

$$\hat{\theta} = (D_\alpha^T D_\alpha + \lambda P^*)^{-1} D_\alpha^T Y \qquad (11)$$

P^* represents a penalty matrix, where the indicator vector is estimated iteratively by updating its value in a way that minimizes the objective function while satisfying a constraint $\alpha^T\alpha = 1$.

The updating process continues until the estimated values converge across successive iterations. Once the final estimates are obtained, the estimated value of the dependent variable is as follows [13]

$$\hat{y}_i = \hat{\beta}_0 + X_i^T\hat{\beta} + B(Z_i^T\hat{\alpha})^T\hat{\gamma} \qquad (12)$$

Robust Local Linear Huber Estimation: This method relies on estimating the non-linear component through local linear regression, where the non-linear function is approximated around each point of the independent variable using a local linear function. To enhance estimation efficiency in the presence of outliers, the Huber loss function is employed instead of the traditional ordinary least squares approach; thus, this method combines the flexibility of local estimation with robustness against the influence of outliers [9].

The non-linear part is approximated as follows:

$$g(T_i) \approx a_j + b_j(T_i - T_j) \qquad (13)$$

where $a_j$represents the value of the function at the point T_j, and b_j represents the slope of the function at the same point. By substituting the non-linear part into the model, we obtain [11]

$$Y_i \approx \beta_0 + X_i^T\beta + a_j + b_j(T_i - T_j) + \varepsilon_i \quad (14)$$

Whereas

$$e_{ij} = Y_i - \beta_0 + X_i^T\beta + a_j + b_j(T_i - T_j) \quad (15)$$

Used are To give greater importance to observations close to the point T_j, kernel [10]

$$w_{ij} = \frac{K\left((T_i - T_j)\backslash K\right)}{\sum_{i=1}^{n} K\left((T_r - T_j)\backslash h\right)} \quad (16)$$

K represents the kernel function, h represents the bandwidth parameter, and w_{ij} represents the weight assigned to observation i at point T_j.

The Huber function is used to reduce the impact of large residuals, and it is written as follows :

$$p_c(u) = \begin{cases} u^2\backslash 2, & |u| < c \\ c|u| - c^2\backslash 2, & |u| > c \end{cases} \quad (17)$$

where u represents the standardized residual, and c represents the tuning constant that determines the transition from quadratic loss to linear loss. The standardized residual is calculated using the following relationship:

$$\mathrm{u_{ij}} = \frac{e_{ij}}{s} \quad (18)$$

where s represents a robust measure of dispersion for the residuals. Accordingly, the robust objective function is defined as follows:

$$Q = \sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}\, p_c\left(\frac{\mathrm{e_{ij}}}{s}\right) \quad (19)$$

The parameters are estimated by minimizing the aforementioned objective function, while taking the constraint into account $\hat{\alpha}^T\hat{\alpha} = 1$

The estimated value of the response variable is then calculated through[12]

$$\widehat{Y_i} = \widehat{\beta_0} + X_i^T\hat{\beta} + \hat{g}(Z_i^T\hat{\alpha}) \quad (20)$$

V. Eexperimental Aspect

The simulation experiments were executed relying on the three sizes of samples (50,80,150) to generate the random certain data in the Partial Linear Single-lndex Model To present the various functions, many mathematical functions were used:

- the Linear function and its formula [1] $g(T_i)$=0.25+2t
- The linear function with invariable limit existence[1] $g(T_i)$= -0.5 + $10(t - 0.5)^2$
- Non-Linear Function [15] $g(T_i)$=exp(2t-1)
- Linear Function [15] $g(T_i)$=8t(1-t)

The implementation of the experiment requires the following:

1.Generate the variable x with zero mean and σ^2 varianc.e

2.Generate the variable T_i with zero mean and σ^2 variance.

3.Generate the variable e with zero mean and σ^2 variance , whereas follows the normally distribution and three levels of Standard deviation (0.5,1,1.5) substitut the variables that are generated in the Partial Linear Single-lndex Model to get the response variable y, for each level of criterion deviation and each Partial Linear Single-lndex Model.

The following non-parametric methods have already been compared :

1. Spline- Based Estimation (SBE)
2. Robust Local Linear Huber Estimation (RLLHE)

After getting all the experiments' stimulation results, relying on four functions that represent the functions of Partial Linear Single-lndex Model which were selected to stimulate the Partial Linear Single-lndex Models, if method (RLLHE) occurred, they got the best value of MSE for all standard deviation levels and samples. Table 1 shows the criterion values and average MSE for all standard deviation levels and all sample sizes for the used non-parametric methods to estimate the first Partial Linear Single-lndex Models.

Table no 1: shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the first Partial Linear Single-lndex Models.

		Standard deviation levels for the boundary of error		
N	nonparametric methods	0.5	1	1.5
50	SBE	1.142	2.709478	4.360352
	RLLHE	1.043208	2.328272	4.126534
80	SBE	1.113219	2.464955	4.357315
	RLLHE	1.016019	2.320047	3.985702
150	SBE	1.063232	2.441591	4.282632
	RLLHE	1.007542	2.191039	3.738212

Table no 2 shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the second Partial Linear Single-lndex Models .

		Standard deviation levels for the boundary of error		
N	nonparametric methods	0.5	1	1.5
50	SBE	1.141705103	2.709533087	4.361579881
	RLLHE	1.04155548	2.325208976	4.127916026
80	SBE	1.113413985	2.464553338	4.3605395
	RLLHE	1.019751839	2.320796556	3.999511667
150	SBE	1.063236856	2.44136103	4.281660291
	RLLHE	1.007503491	2.195520973	3.740296155

Table no 3: **shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the third Partial Linear Single-lndex Models**

		Standard deviation levels for the boundary of error		
N	nonparametric methods	0.5	1	1.5

50	SBE	1.141968992	2.709474818	4.360479408
	RLLHE	1.042987621	2.327947271	4.126578833
80	SBE	1.113237857	2.464903317	4.357647932
	RLLHE	1.016189045	2.320074208	3.986666902
150	SBE	1.063229714	2.441565238	4.282520367
	RLLHE	1.00746742	2.191272995	3.738214105

Table no 4: **shows the standard value (MSE) for all Standard deviaion letvels, and all samples' sizes in the used non-Parametric levels to estimate the fourth Partial Linear Single-lndex Models**

		Standard deviation levels for the boundary of error		
N	nonparametric methods	0.5	1	1.5
50	SBE	1.142186379	2.709426318	4.359787569
	RLLHE	1.043959424	2.329290489	4.126275265
80	SBE	1.113157284	2.465120889	4.355679101
	RLLHE	1.015675252	2.319584705	3.982894958
150	SBE	1.063264162	2.441732079	4.283087829
	RLLHE	1.007733235	1.007733235	3.738639805

VI. Conclusions:

1. Comparing the estimation methods used for Partial Linear Single-Index Models reveals that the Robust Local Linear Huber Estimation (RLLHE) method is the best, as it yielded the lowest Mean Squared Error (MSE) across all sample sizes and various levels of standard deviation.

2. It was found that, for a constant standard deviation, the Mean Squared Error (MSE) decreases as the sample size increases.

3. It was found that, for a constant sample size, the Mean Squared Error (MSE) increases as the standard deviation increases.

References

[1]. Ahkim, M., Gijbels, I., & Verhasselt, A. (2017). Shape testing in varying coefficient models. Test, 26(2), 429-450.
[2]. Chen, L. P., Wu, J. C., & Yi, G. Y. (2026). Semiparametric estimation for error-prone partially linear single-index models. Journal of Computational and Graphical Statistics, 35(2), 669-684.
[3]. Cui, S., Guo, X., & Zhang, Z. (2025). Estimation and inference in ultrahigh-dimensional partially linear single-index models. Science China Mathematics, 68(8), 1807-1840.
[4]. Do, H., Wang, Y., Liu, M., & Lee, M. (2025). Neural Network-based Partial-Linear Single-Index Models for Environmental Mixtures Analysis. arXiv preprint arXiv:2512.11593.
[5]. Fitriyah, A. T., Chamidah, N., & Saifudin, T. (2025). Prediction of paddy production in Indonesia using semiparametric time series regression least square spline estimator. Data and Metadata, 4, 527-527.
[6]. Goepp, V., Bouaziz, O., & Nuel, G. (2025). Spline regression with automatic knot selection. Computational Statistics & Data Analysis, 202, 108043.
[7]. Huang, K., Chen, X., Guan, Y., & Li, Y. (2025). Partially linear single-index models and functional principal component analysis of spatially and temporally indexed point processes. Journal of Business & Economic Statistics, 43(4), 1077-1091.
[8]. Kirkby, J. L., Leitao, Á., & Nguyen, D. (2023). Spline local basis methods for nonparametric density estimation. Statistic Surveys, 17, 75-118.
[9]. Li, Z., Yan, P., & Hsu, L. T. (2026). Logistic-aided Huber M-estimator for robust GNSS positioning. arXiv preprint arXiv:2603.19640.
[10]. Liu, T., Xu, X., Lee, L. F., & Mei, Y. (2026). Robust estimation for the spatial autoregressive model. Econometric Theory, 1-51.
[11]. Loh, P. L. (2025). A theoretical review of modern robust statistics. Annual Review of Statistics and Its Application, 12(1), 477-496.
[12]. Wang, W., Peng, X., & Tong, T. (2025). Robust and efficient mediation analysis via huber loss. Psychometrika, 90(2), 734-756.
[13]. Xu, Z., Xiu, Y., Zhan, X., Chen, B., & Shimada, K. (2022). Vision-aided UAV navigation and dynamic obstacle avoidance using gradient-based B-spline trajectory optimization. arXiv preprint arXiv:2209.07003.
[14]. Zu, T., & Yu, Y. (2023). gplsim: An R Package for Generalized Partially Linear Single-index Models. R J., 15(1), 55-64.
[15]. Cai, Z., Fan, J., & Li, R. (1999). Generalized varying-coefficient models.