Estimating the Link Function in a Partial Linear Single Index Model for Longitudinal Data using LSTM Neural Networks

Hussein Jabbar Bayyoodh and Mohammed Sadeq Aldouri
Department of Statistics, College of Administration and Economics, University of Baghdad, 10001 Baghdad, Iraq
husseenjb@gmail.com, dr_aldouri@coadec.uobaghdad.edu.iq

Keywords: Partial Linear Single-Index Model, Longitudinal Data, Long Short-Term Memory, Local Polynomial, Semiparametric, Deep Neural Networks.

Abstract: This paper investigates and compares the performance of two estimation approaches - Long Short-Term Memory (LSTM) networks and the semi-parametric Minimum Average Variance Estimation (SMAVE) method - for the Partial Linear Single Index Model (PLSIM) in the context of longitudinal data. The PLSIM combines linear and nonlinear components, offering modeling flexibility for complex data structures often encountered in repeated measurements. We conduct extensive simulations with varying sample sizes (N = 50, 100, 150) to evaluate the prediction accuracy of both methods in estimating the unknown link function. Evaluation metrics such as Mean Squared Error (MSE), bias, and coefficient of determination (R^2) are used to assess estimation quality. Results show that LSTM significantly outperforms SMAVE in estimating both the linear parameters and the nonlinear link function. The LSTM method consistently achieves lower MSE and bias values, as well as higher R^2 scores for both the model and the nonlinear function, highlighting its superior ability to capture temporal dependencies and complex nonlinear relationships in longitudinal data. In contrast, SMAVE's performance is more sensitive to bandwidth selection and sample size. These findings suggest that deep learning models like LSTM offer a powerful alternative to traditional semi-parametric methods in longitudinal data analysis.

1 INTRODUCTION

Longitudinal data, also known as panel data, are a common statistical tool in many applied fields, such as economics, medicine, epidemiological studies, and clinical trials. These data are characterized by the fact that they involve repeated measurements of the same experimental units (such as individuals) across different time periods. This distinguishes them from cross-sectional data, which assumes that observations are independent and non-repeated. The correlation between measurements within each individual unit is an essential property of longitudinal data; these measurements are considered to be correlated over time, whereas measurements between different individuals are typically treated as independent.

Given this complex correlational structure, analyzing longitudinal data requires specialized statistical models that account for this temporal correlation. Statistical methods used to analyze regression models with longitudinal data typically fall into three main categories, differing in their assumptions about the relationship between observations within each unit and between different units. In recent years, semiparametric models have gained widespread attention due to their flexibility in incorporating both parametric and non-parametric components within the same model, enhancing its explanatory and predictive power. These models are particularly suitable for analyzing longitudinal data, despite the challenges they pose at the estimation level, particularly when dealing with the non-parametric component [1].

Among semiparametric models, the Partially Linear Single-Index Model (PLSIM) is a flexible framework that combines the advantages of linear models and non-linear structures. It allows for the representation of complex relationships between variables while maintaining relative mathematical simplicity. This model has gained increasing attention in many economic, medical, and social applications due to its ability to balance interpretability with estimation flexibility [2].

Although previous research has examined the model for independent data, generalizing this model to longitudinal data remains a fertile area of research, given the complexity of the temporal structure and the internal correlation between observations. Several studies have focused on developing efficient estimation methods for the parametric and nonparametric components of the model in this context. [3] addressed estimation and testing for independent variables, while [4] proposed efficient semiparametric least squares estimators using profile estimation. Some studies, such as, relied on simplified mathematical methods to calculate semiparametric information limits.

On the other hand, some studies, such as [5] and [6], addressed the generalization of the PLSIM model to longitudinal data, focusing on the challenges associated with estimating model parameters and the correlation function. [7] proposed a unified method using generalized estimating equations (GEE) to analyze imbalanced longitudinal data, whether dense or sparse. The results showed that the rates of convergence and asymptomatic variances of the estimators varied depending on the data structure. However, these estimators were not efficient in the semiparametric sense.

In this work, we present a new estimation method based on Long-Short-Term Memory (LSTM) deep neural networks to estimate the link function in a PLSIM model for longitudinal data. The proposed method was compared with the semi-parametric estimation method known as Minimum Average Variance Estimation (MAVE). Simulation results showed that the LSTM-based method outperforms the estimation accuracy, especially in representing complex nonlinear relationships between variables.

Based on the above, this research aims to develop an advanced estimation framework for the partial linear single-index model (PLSIM) in the context of longitudinal data. This approach employs deep learning techniques, specifically LSTM neural networks, to improve the estimation of the nonlinear component (link function) and effectively address temporal correlation within individuals. This approach represents a paradigm shift in longitudinal data modeling, as it combines the theoretical foundations of semi-parametric models with the high predictive capabilities provided by machine learning algorithms. The research focuses on comparing the experimental performance of the proposed method with conventional methods in terms of estimation accuracy, parameter efficiency, and convergence speed, using extensive simulation experiments covering a variety of scenarios in terms of data density and the number of iterations per unit. This methodology is expected to enhance our understanding of the dynamic relationships between variables in longitudinal data and pave the way for more accurate and flexible applications in various applied fields

2 MODEL AND METHOD ESTIMATION

2.1 Model

The partial linear single-index model in longitudinal data is written in the form [8]:

$$Y_{it} = Z_{it}^T\theta + g(X_{it}^T\beta) + \varepsilon_{it} \ , \quad (1)$$
$$i = 1, \dots, N \ , t = 1, \dots, T.$$

Where Y_{it} represents the response variable of individual i at time t, X_{it} and Z_{it} are explanatory variables of the linear and non-linear parts, respectively, with dimensions R^q and R^p, (θ, β) are unknown the parameter vectors of the variables to be estimated with dimension $R^q$and R^p, $\parallel \beta \parallel = 1$ ($\|\cdot\|$ representing the Euclidean Norm), and the first element must be positive. $g(\cdot)$ is an unknown link function and ε_{it} are random error and $E(\varepsilon_{it} \setminus X_{it}, Z_{it}) = 0$

2.2 Methods Estimation

In this section, the estimation methodologies used to estimate the parameters of a partial single-indicator model in the context of longitudinal data will be presented and detailed. Two different estimation methods have been adopted, representing two methodological approaches that differ in their philosophy and mathematical foundation:

The first method relies on deep neural networks of the LSTM type, a deep learning algorithm specifically designed for processing time-series data. This method is particularly suitable for analyzing longitudinal data due to its ability to capture temporal dynamics and internal correlations between measurements across time.

The second method is MAVE method, a classical statistical technique that aims to estimate a nonlinear link function and trend indicator in a way that minimizes the mean variance within a semiparametric framework. It is a standard reference for this type of model.

The purpose of employing these two approaches is to conduct a systematic comparison between traditional methods, which rely on rigorous statistical foundations, and modern methods, which are based on machine learning techniques. The comparison focuses on estimation accuracy, parameter efficiency, and the ability to capture the complex structure of longitudinal data. In the following paragraphs, each method will be presented in detail, outlining the estimation steps, required conditions, and the theoretical basis on which it is built.

2.2.1 Long Short-Term Memory

The LSTM neural network is an advanced model in the field of recurrent neural networks. It differs from a traditional recurrent network (RNN) in that the hidden unit contains a memory cell instead of a traditional node. The memory cell provides a sophisticated mechanism for transferring information across time through a series of operations that allow the model to selectively retain or forget information, helping to overcome the problems of vanishing/exploding gradients.

The cell state (c_t) is a key component within an LSTM cell, acting as a conduit for information transfer across time. Gates are used to control the flow of information into or out of the cell, based on the resulting value of sigmoid activation functions. When values approach zero, information is blocked, while information is passed through if values approach one. In this way, the LSTM can determine which information to retain or forget at each time step, maintaining the model's long-term effectiveness.

Below is a breakdown of the basic components of an LSTM module along with their associated equations:

1) Input node (Candidate Cell State). This node, known as $\tilde{c}_t$, represents the new candidate cell state. It is calculated by summing the current input x_t and the previous hidden state h_{t-1} using the tanh activation function.

$$\tilde{c}_t = \emptyset(W_{\tilde{c}x}x_t + W_{\tilde{c}h}h_{t-1} + b_{\tilde{c}}). \quad (2)$$

Where:

- $\emptyset()$ stands for an element-wise application of the $\tan h$ function;
- $W_{\tilde{c}x}$ is the input weights matrix;
- $W_{\tilde{c}h}$ is the hidden weights matrix.

2) Input Gate. This gate is a unique element of the LSTM architecture, controlling the amount of new information that will be input into the cell state. It is calculated as follows:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i). \quad (3)$$

Where:

- $\sigma()$ represents a sigmoid function applied to each element. The resulting value ranges from 0 (unimportant information) to 1 (important information)
- $W_{ix}$is the input weights matrix
- $W_{ih}$is the hidden weights matrix.

3) Forget Gate. This gate controls whether information from the cell's previous state is retained or not. It is calculated using the equation:

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right). \quad (4)$$

It regulates the flow of information from the previous internal state c_{t-1}, W_{fx} is the input weights matrix, W_{fh} is the hidden weights matrix.

4) Internal Cell State. c_t represents the updated state of the cell, and is calculated by taking the Hadamard product between the forget gate and the previous state, then adding the product between the input gate and the new candidate state:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t. \quad (5)$$

Where $\odot$ represents the Hadamard product, i.e. the product between any two elements of the same position.

5) Output Gate. This gate determines whether the information stored in the cell will be used to update the current hidden state. The output gate is first calculated:

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_0),$$
$$h_t = \emptyset(c_t) \odot o_t. \quad (6)$$

Together, these components form the architecture of an LSTM cell, enabling the model to learn from long-term temporal dependencies, making it suitable for analyzing longitudinal or sequential data see [9] and [10] that involve complex, nonlinear temporal relationships.

2.2.2 Semiparametric Minimum Average Variance Estimation (SMAVE)

Xia et al. proposed [11] an estimation method known as MAVE, which provides a general framework for estimating partial single-index models. This method relies on the joint estimation of the coefficient vector β and the nonlinear function g(·) within a single-index model. It is considered a distinctive approach due to its ease of implementation compared to traditional methods and its ability to deliver efficient numerical solutions through a local linear approximation of the nonparametric function.

Subsequently, [5] and [12] extended this methodology to longitudinal data by explicitly accounting for temporal correlations within each observational unit.

The estimation objective is formulated by solving the following problem:

$$(\beta, \theta) = \underset{\beta,\theta}{\operatorname{argmin}}\, E[Y_{it} - g(X_{it}^T\theta) - Z_{it}^T\beta]^2. \quad (7)$$

Let $Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{NT})^T$,

$$Z = (Z_{11}, \dots, Z_{1T}, \dots, Z_{NT})^T, X = (X_{11}, \dots, X_{1T}, \dots, X_{NT})^T, \varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1T}, \dots, \varepsilon_{NT})^T$$

$$g(X, \theta) = (g(X_{11}^T\theta), \dots, g(X_{1T}^T\theta), \dots, g(X_{NT}^T\theta))$$

$$X_{it}(\theta) = ((X_{11} - X_{it})^T\theta, \dots, (X_{1T} - X_{it})^T\theta, \dots, ((X_{NT} - X_{it})^T\theta)^T.$$

The model (1) can be rewritten as

$$Y = g(X, \theta) + Z\beta + \varepsilon.$$

A local linear approximation is adopted using the derivative of η as follows:

$$g(X_{it}^T\theta) \approx g(X^T\theta) + \acute{g}(X^T\theta)(X_{it} - x)^T\theta.$$

It symbolizes: $a_{it} = g(X^T\theta)$, $b_{it} = \acute{g}(X^T\theta)$

The parameters are estimated by solving the following problem:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}[Y - Z\beta - (e_{NT}, X_{it}(\theta))(a_{it}, b_{it})^T]^T \times M_{it}[Y - Z\beta - (e_{NT}, X_{it}(\theta))(a_{it}, b_{it})'], \quad (8)$$

where: $M_{it} = diag(m_{11,it}, \dots, m_{11,it}, \dots, m_{11,it})$.

It is a diagonal matrix containing the kernel weights, and it achieves: $\sum_{j=1}^{N}\sum_{s=1}^{T} m_{js,it} = 1$ and e_{NT} is a NT-dimensional vector with all elements being 1 The estimation is performed using an iterative algorithm consisting of the following steps:

1) Local estimation of non-parametric coefficients
2) $(a_{it}, b_{it})^T = \left(\bar{X}_{it,*}^T(\theta) M_{it} \bar{X}_{it,*}^T(\theta)\right)^{-1} \bar{X}_{it,*}^T(\theta) M_{it}(Y_{it} - Z_{it}\beta)$.
   Where $\bar{X}_{it,*}^T(\theta) = (e_{NT}, X_{it}(\theta))$.
3) Update β and θ:
   $$(\beta^T, \theta^T)^T = \begin{pmatrix} Z_*^T M Z_* & Z_*^T M X_* \\ X_*^T M Z_* & X_*^T M Z_* \end{pmatrix}^{-1} \begin{pmatrix} Z_*^T \\ X_*^T \end{pmatrix} M(Y_* - A_*).$$
   Where: $Y_* = (Y_{11,*}^T, \dots, Y_{1T,*}^T, \dots, Y_{NT,*}^T)^T$,
   $$X_* = (b_{11}X_{11,*}^T, \dots, b_{1T}X_{1T,*}^T, \dots, b_{NT}X_{NT,*}^T)^T,$$
   $$X_{it,*} = (X_{11} - X_{it}, \dots, X_{1T} - X_{it}, \dots, X_{NT} - X_{it})^T,$$
   $$Z_* = (Z_{11,*}^T, \dots, Z_{1T,*}^T, \dots, Z_{NT,*}^T)^T,$$
   $$M = diag(M_{11}, \dots, M_{1T}, \dots, M_{NT}),$$
   $$A = (a_{11}e_{NT}^T, \dots, a_{1T}e_{NT}^T, \dots, a_{NT}e_{NT}^T)^T.$$
4) Iteration until convergence Steps 1 and 2 are repeated using the updated values of β and θ until numerical convergence is achieved.

Two sets of weights are used in the iterative procedure (multivariate weights):

$$w_{js,it} = \frac{H((X_{js} - X_{it})/h_1)}{\sum_{j=1}^{N}\sum_{s=1}^{T} H((X_{js} - X_{it})/h_1)}.$$

Where $H(\cdot)$ is a p-variate symmetric kernel function and h_1 is a bandwidth, single-index weights:

$$w_{js,it}^{\theta} = \frac{K((X_{js} - X_{it})^T\theta/h_2)}{\sum_{j=1}^{N}\sum_{s=1}^{T} K((X_{js} - X_{it})^T\theta/h_2)}.$$

Where $K(\cdot)$ a univariate symmetric kernel is function and h_2 is a bandwidth. Then, we can obtain the final estimators $\hat{\beta}$ and $\hat{\theta}$ by using $\tilde{\beta}$ and $\tilde{\theta}$ and the above iterations with the single-index weights. By substituting β, θ and $X_{it}^T\theta_0$ with $\hat{\beta}, \hat{\theta}$ and u, we can get the estimator of $g(u)$, which is denoted by $\hat{g}(u)$.

3 EVALUATION METRICS

Comparing estimation methods is a fundamental procedure in statistical research to identify the most appropriate techniques. Various evaluation metrics exist - some designed to compare parameter estimates, while others assess the overall model performance. In general, the choice of appropriate metrics depends on the nature of the data and the objectives of the statistical analysis. In this paper, the following evaluation metrics will be employed:

$$MSE = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it})^2}{N * T}.$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it})^2}.$$

$$MSE(g(u)) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(g(X_{it}^T\beta) - \hat{g}(X_{it}^T\beta))^2}{N * T}.$$

$$R^2(g(u)) = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(g(X_{it}^T\beta) - \hat{g}(X_{it}^T\beta))^2}{N * T}.$$

4 SIMULATION

In this simulation paper, we generate synthetic longitudinal data to evaluate estimation methods for the Partial Linear Single Index Model (PLSIM). The data consists of N = 50,100,150 subjects, each measured at T = 5 time points, resulting in NT total observations. The covariate vectors X (predictors for the nonparametric component) and Z (predictors for the linear component) are simulated from standard normal distributions.

The true parameter vectors are defined as: $\beta = \frac{1}{\sqrt{3}}[1, -1, 1]^{T}$ normalized to unit norm, representing the direction for the single index $X_{it}^{T}\beta$ $\theta = [0.5, -0.8]^{T}$, corresponding to the coefficients in the linear part $Z_{it}^{T}\theta$ The nonlinear link function is defined as $g(u) = \sin(u)$ which operates on the single index $u = X_{it}^{T}\beta$ The response variable Y_{it} is generated

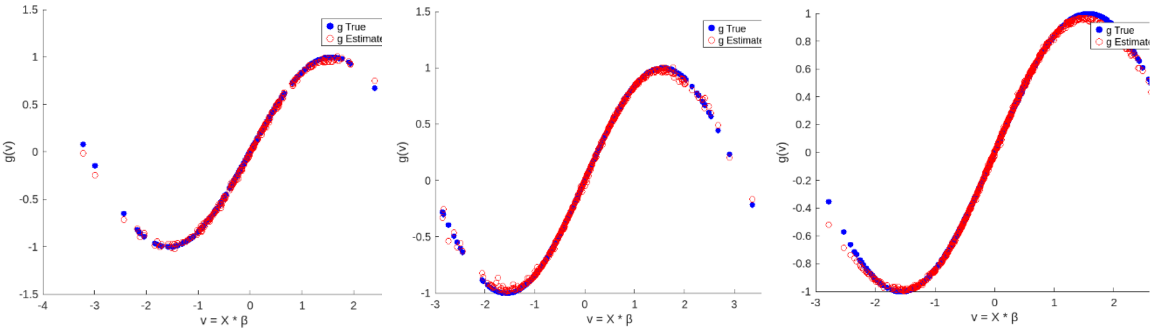
$$Y_{it} = Z_{it}^{T}\theta + g(X_{it}^{T}\beta) + \varepsilon_{it}\,.$$

The error term ε_{it} is generated to follow an AR(1) structure within each subject to induce temporal correlation Specifically:

- $\varepsilon_{i1} \sim N\,(0, \sigma^2)$, and for $t > 1$: $\varepsilon_{i}t = \rho * \varepsilon_{i}(t-1) + \eta_{i}t$, where $\eta_{i}t \sim N(0, \sigma^2)$.
- The parameters are set to $\sigma = 0.3$ *and* $\rho = 0.3$.

Table 1: Means and MSEs of the estimates of the parameters and link function using the LSTM estimation method.

Parameter	N=50		N=100		N=150	
	bias	MSE	bias	MSE	bias	MSE
θ_1	0.0053	0.0000	0.0155	0.0002	0.0048	0.0000
θ_2	0.0074	0.0001	0.0138	0.0002	0.0117	0.0001
β_1	0.0162	0.0003	0.0654	0.0043	0.0583	0.0034
β_2	0.0604	0.0036	0.0489	0.0024	0.1118	0.0125
β_3	0.0041	0.0000	0.0303	0.0009	0.1136	0.0129
$g(\cdot)$	0.0115	0.0007	0.0024	0.0004	0.0064	0.0004
MSE MODEL	0.1239		0.1027		0.1009	
R^2 MODEL	0.9096		0.9277		0.9260	
R^2 $g(\cdot)$	0.9984		0.9992		0.9990	

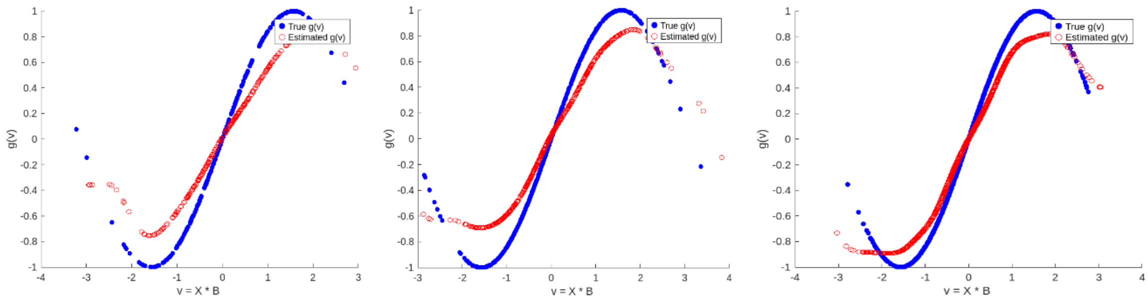


N=50 N=100 N=150

Figure 1: Scatter diagram of true link function and estimated link function using LSTM estimation method.

Table 2: Means and MSEs of the estimates of the parameters and link function using the SMAVE estimation method.

Parameter	N=50		N=100		N=150	
	bias	MSE	bias	MSE	bias	MSE
θ_1	0.008941	0.00007	0.01404	0.000197	0.02544	0.000647
θ_2	0.0179	0.0003205	0.01953	0.0003813	0.01832	0.0003355
β_1	0.4257	0.1812	0.2372	0.05627	0.1487	0.02212
β_2	0.2013	0.04054	0.5022	0.2522	0.08696	0.007562
β_3	0.03144	0.0009883	0.002175	4.73e-06	0.3998	0.1599
$g(\cdot)$	0.0014	0.1250	0.0042	0.1347	0.0176	0.0930
MSE MODEL	0.2328		0.2371		0.2020	
R^2 MODEL	0.8301		0.8329		0.8519	
R^2 $g(\cdot)$	0.6980		0.6990		0.7872	



N=50 N=100 N=150

Figure 2: Scatter diagram of true link function and estimated link function using SMAVE estimation method.

To evaluate the model's performance more accurately and to ensure measuring its generalization ability rather than relying solely on the training data performance, the total sample was divided into two subsets: a training set and a testing set. Specifically, 70% of the individuals were allocated for training, while the remaining 30% were reserved for testing. The LSTM model was trained exclusively on the training set, whereas the testing set was used to evaluate the model and calculate the performance metrics.

The estimates for the linear parameters θ_1 and θ_2. The mean squared error (MSE) for the parameters θ is very low across all sizes, and in some cases reaches values very close to zero, indicating a good ability of the model to recover the linear parameters. The bias is also small (<0.02), which enhances the reliability of the estimate (Table 1).

Estimates of the indicator vector β at N=150 show a significant improvement in the estimate, with the MSE and bias being relatively smaller compared to smaller values, especially for the parameters β_2 and β_3. In contrast, at N=50, the bias is relatively larger for some components (e.g., β_2), an expected behavior with a small sample size.

The mean square error of the estimate of the nonlinear function $g(\cdot)$ is very low (<0.001), reflecting a high accuracy in representing the nonlinear relationship. The relationship between the true and estimated link functions obtained using the LSTM method is illustrated in Figure 1.

The determined coefficient R^2 for the estimate of $g(\cdot)$ is close to 1 (approximately 0.999), demonstrating excellent estimation quality in all cases. Goodness of fit of the overall model the mean squared error of the full model decreases with increasing sample size (from 0.123 at N=100 to 0.1009 at N=150). R^2 for the full model increases slightly with increasing sample size, reaching about 0.927 at N=150, indicating a high fit.

The three graphs show a comparison between the true values of the nonlinear function $g(u)$ (blue dots) and the estimated values (red circles) across different sample sizes N=50, 100, and 150 (Fig. 1). The following observations can be drawn: At all three sample sizes, the estimate appears to agree very closely with the true shape of the function, with clear convergence between the blue dots and red circles. As the sample size increases (from 50 to 150), the dispersion around the true curve decreases, and the estimated values become smoother and more stable. This reflects the improved accuracy with increasing sample size, which is theoretically expected in semiparametric estimation.

The results indicate that the estimation of the linear coefficients (θ_1, θ_2) has a high degree of accuracy, as the mean deviation (Bias) and mean squared error (MSE) values remained relatively low across all sample sizes (Table 2). In contrast, the single-index coefficients $(\beta_1, \beta_2, \beta_3)$ recorded high bias, especially for the smaller sample (N=50). The bias for β_1, β_2 reached 0.4257 and 0.2013, respectively, with MSEs of 0.1812 and 0.04054, reflecting poor estimation accuracy in the nonlinear part of the model. The estimation results of the function $g(\cdot)$ also showed inconsistent performance, with the highest MSE recorded at N=100, at 0.1347, despite a low bias (0.0042), which may indicate fluctuations in the estimate resulting from the nature of the function or the sample size. As for the quality of the model as a whole, the performance indicators showed that the model's coefficient of determination (R^2 MODEL) gradually decreased from 0.8519 at N=150 to 0.8301 at N=50. Meanwhile, the accuracy of the representation of the function $g(\cdot)$ was significantly lower, with its R^2 reaching 0.6980 at N=50, demonstrating the superiority of the deep neural network method used in estimating the link function.

For the three plots in all three sizes, the estimate appears to be inconsistent with the true form of the function, indicating the difficulty of smoothing methods in interpreting the relationship for a nonlinear function (Fig. 2).

5 CONCLUSIONS

Comparison results between LSTM and SMAVE estimation methods on a Partial Linear Single Index Model (PLSIM) for longitudinal data showed a clear superiority of LSTM in terms of the accuracy of estimating linear and nonlinear parameters and the stability of results across different sample sizes. LSTM recorded very low MSE and deviation values in estimating the parameters θ and β, and demonstrated excellent performance in estimating the nonlinear function $g(\cdot)$ with R^2 values exceeding 0.99, reflecting its ability to capture complex relationships in longitudinal data. In contrast, SMAVE showed less accurate results, particularly in estimating the parameters β and the function $g(\cdot)$, where the MSE and Bias values were relatively high, especially at small sample sizes, indicating that it is affected by the choice of bootstrap parameters and the difficulty of estimation with limited samples. LSTM also outperformed in overall model quality, achieving the highest R^2 for the model, demonstrating a better fit to the data and superior prediction, while SMAVE was less stable and performed less well. Thus, these results demonstrate that LSTM represents a more efficient and accurate option for analyzing and estimating PLSIM models with longitudinal data compared to traditional semiparametric methods.

REFERENCES

[1] L. Su and A. Ullah, "Profile likelihood estimation of partially linear panel data models with fixed effects," Economics Letters, vol. 92, no. 1, pp. 75–81, Jul. 2006, doi: 10.1016/j.econlet.2006.01.019.

[2] R. J. Carroll, J. Fan, I. Gijbels, and M. P. Wand, "Generalized partially linear single-index models," Journal of the American Statistical Association, vol. 92, no. 438, p. 477, Jun. 1997, doi: 10.2307/2965697.

[3] H. Liang, X. Liu, R. Li, and C. L. Tsai, "Estimation and testing for partially linear single-index models," The Annals of Statistics, vol. 38, no. 6, pp. 3811–3836, 2010, doi: 10.1214/10-AOS835.

[4] T. Chen and T. Parker, "Semiparametric efficiency for partially linear single-index regression models," Journal of Multivariate Analysis, vol. 130, pp. 376–386, 2014, doi: 10.1016/j.jmva.2014.06.006.

[5] J. Chen, J. Gao, and D. Li, "Estimation in partially linear single-index panel data models with fixed effects," Journal of Business and Economic Statistics, vol. 31, no. 3, pp. 315–330, Jul. 2013, doi: 10.1080/07350015.2013.775093.

[6] P. Lai, G. Li, and H. Lian, "Quadratic inference functions for partially linear single-index models with longitudinal data," Journal of Multivariate Analysis, vol. 118, pp. 115–127, Jul. 2013, doi: 10.1016/j.jmva.2013.03.019.

[7] J. Chen, D. Li, H. Liang, and S. Wang, "Semiparametric GEE analysis in partially linear single-index models for longitudinal data," The Annals of Statistics, vol. 43, no. 4, pp. 1682–1715, Aug. 2015, doi: 10.1214/15-AOS1320.

[8] S. Ma, H. Liang, and C. L. Tsai, "Partially linear single index models for repeated measurements," Journal of Multivariate Analysis, vol. 130, pp. 354–375, 2014, doi: 10.1016/j.jmva.2014.06.011.

[9] "Deep learning-based prediction frameworks for longitudinal data analysis," Florida State University Libraries, 2022.

[10] N. Olawale Adeboye and N. Olawale Alabi, "Deep-learning modelling of dynamic panel data for African economic growth," Journal of Econometrics and Statistics, vol. 2, no. 1, 2022, doi: https://doi.org/10.46791/jes.2022.v02i01.03.

[11] Y. Xia and W. Härdle, "Semi-parametric estimation of partially linear single-index models," Journal of Multivariate Analysis, vol. 97, no. 5, pp. 1162–1184, May 2006, doi: 10.1016/j.jmva.2005.11.005.

[12] J. Q. Zhao, Y. Y. Zhao, and Z. X. Miao, "Analysis of panel data partially linear single-index models with serially correlated errors," Journal of Computational and Applied Mathematics, vol. 368, Apr. 2020, doi: 10.1016/j.cam.2019.112532.