



Inferences for extended partially linear single-index models

Zijuan Chen¹ · Suojin Wang¹

Received: 17 May 2022 / Accepted: 15 December 2022 / Published online: 21 February 2023
© The Author(s) under exclusive licence to Sociedad de Estadística e Investigación Operativa 2023

Abstract

In partially linear single-index models, there are two different covariate matrices in the model for the linear part and nonlinear part. All covariate information needs to be divided into two parts before the model can be fitted. In contrast, in the extended partially linear single-index models, all the covariate variables are included in one matrix, which is contained in both the linear part and nonlinear part of the model. We propose local smoothing estimators for the model parameters and unknown function with computationally efficient and accurate computation methodologies and obtain the asymptotic properties of the model parameter estimators. We also employ the LASSO penalty to obtain penalized estimators with consistency and oracle property in order to carry out estimation and variable selection simultaneously. Then we develop a linear hypothesis test for the model parameters. Furthermore, we extend the proposed methodology to the increasing dimensional settings under certain assumptions. Simulation studies are presented that support our analytic results. In addition, a real data analysis is provided for illustration.

Keywords LASSO · Local smoothing estimator · Partially linear single-index model · Variable selection

Mathematics Subject Classification 62G05 · 62F03 · 62J07

1 Introduction

Regression models are powerful tools for the analysis of the relationship between a response variable and some covariates. In many simple cases, a linear model $E(y|x) = x^T\beta$ can be used to fit the data, where x is the covariate vector and y is the response. However, the linear assumption may be violated in some cases and the relationship between the response and the covariates can be better explained via a link function g :

✉ Suojin Wang
sjwang@stat.tamu.edu

¹ Department of Statistics, Texas A&M University, College Station, TX 77843, USA

$E(y|x) = g(x^T \beta)$. This single-index model has been proposed and studied in detail (see Ichimura 1993; Horowitz and Härdle 1996). However, sometimes neither of these two models is able to sufficiently represent the relationship between the response and covariates. Carroll et al. (1997) introduced the partially linear single-index model (PLSIM), which is a combination of the linear model and single-index model, and is defined by

$$y = z^T \beta + g(x^T \theta) + e, \quad (1)$$

where y is the response, x and z are covariates with dimensions p and q , respectively, θ and β are p -dimensional and q -dimensional vectors of parameters, g is the unknown link function, and e is the random error. Different estimation and testing methodologies have been proposed for model selection and estimations of the parameters and the link function; see Yu and Ruppert (2002), Xia et al. (2002), Liang and Wang (2005), Xia and Härdle (2006), and Liang et al. (2010) for more details. The partially linear single-index models have also been generalized for analysis of data with more complicated correlation structure, such as longitudinal data, which are widely studied in many fields such as epidemiology and biology. See, for example, Wang et al. (2005), Li and Hsing (2010), Chen et al. (2015), Cai and Wang (2019a), Cai and Wang (2019b), and Cai et al. (2020). However, none of these works consider extended partially linear single-index model (EPLSIM) defined in Eq. (2). In particular, in Cai and Wang (2019b), methods for the generalized PLSIM (GPLSIM) were proposed, which are especially useful for a discrete (such as binary) response variable and are not applicable to EPLSIM.

There are mainly two challenges when applying the existing works on PLSIM to EPLSIM. Firstly, EPLSIM is more general compared with PLSIM and requires more regularity assumptions on the covariates. For example, when dealing with longitudinal data, the correlation structure is more complicated compared with i.i.d. cases. Therefore, many desired properties obtained on PLSIM might not be achievable on EPLSIM based on the existing approaches.

Another challenge is that EPLSIM has higher computational complexity. If they are applied to EPLSIM, the existing methodologies on PLSIM might lead to unreasonably high computational costs, or even become infeasible in some cases. Therefore, better methodologies need to be carried out specifically for EPLSIM.

In practice, the covariates in the PLSIM (1) are usually manually divided into two parts before the model is fitted to the data: One part becomes the covariates of the linear form z , and the other becomes the covariates of the single-index form x . However, this procedure may result in a misspecification of the model: some covariate variables of x should actually belong to z , while some covariate variables of z should actually belong to x . The model selection methods for model (1) are not able to detect the misspecification after x and z are specified. This problem can be solved by using EPLSIM, which was firstly introduced in Xia et al. (1999) and has the form

$$y = x^T \beta + g(x^T \theta) + e. \quad (2)$$

Here x denotes all the covariates in the model, and it appears in both the linear part and the single-index part of the model. The extended partially linear single-index model

(2) is an extension of the PLSIM (1) and is able to prevent the misspecification problem mentioned above. While model (2) may still be misspecified (for example, the true model is an additive model with more than three terms), it has a quite broad form and can be applied to more complex types of data.

Since x appears both in the linear part and the single-index part, it is natural to first consider the identifiability of the model parameters before deriving estimation methodologies. Xia et al. (1999) and Lin and Kulasekera (2007) have proposed and investigated the identifiability problem of extended partially linear single-index models and obtained the regularity conditions that ensure the identifiability. Xia et al. (1999) also proposed a simple kernel estimation for estimating the model parameters, which is similar to the estimation for the parameters of partially linear single-index models (1). Since the dimension of the parameters of extended partially linear single-index models is much larger than the dimension of the parameters of partially linear single-index models, similar estimation methods would be less computationally efficient for extended partially linear single-index models. Recently, Dong et al. (2016) introduced a new estimation method based on orthogonal series expansion. We propose the local linear smoothing estimators for the estimation of extended partially linear single-index models and introduce the profile estimating procedure of the estimation. We show that the solution to the optimization of the profile objective function is unique and can be expressed as linear forms, which leads to fast and accurate computations for the parameter estimation.

Although extended partially linear single-index models (2) eliminate the possibility of the model misspecification discussed above, the number of model parameters becomes larger since all covariate variables appear twice. Therefore, it is of importance to conduct variable selection to prevent model overfitting, which may lead to biased or inefficient estimators and predictions. We propose the penalized local linear smoothing estimation for extended partially linear single-index models with the use of penalty functions, such as the least absolute shrinkage and selection operator (LASSO). The estimators defined by the approach have several advantages [see Chapters 2 and 3 of Fan and Gijbels (1996)], and the variable selection procedure is automatically completed during the estimation procedure. As shown in the empirical study in Sect. 6, when the model parameters contain a substantial number of zeros (sparsity), the penalized estimators have better performance compared with the non-penalized estimators, even if the sample size is large. Therefore, in practice, estimating the parameters using the penalized estimators is always recommended when some covariates are known to be redundant.

After studying the estimation procedure of the model parameters, hypothesis testing for the linear constraints of the parameters is also considered. Based on the difference between the minimum values of the objective function in the null space and in the alternative space, a chi-squared type of test statistic is proposed for this type of linear hypotheses.

In increasing dimensional settings where p is not fixed but increases as the sample size n increases, some estimation methodologies are available in the existing literature for single-index models or partially linear single-index models. See, for example, Radchenko (2015) about LASSO estimators of increasing dimensional single-index models, Wang et al. (2012) about estimation in increasing dimensional single-index

models with non-convex penalties, and Ma and Zhu (2013) about estimation of heteroscedastic partially linear single-index models. These approaches provide efficient and robust estimation for single-index models or partially linear single-index models, but none of them consider the estimation problem for the EPLSIM in increasing dimensional settings. In this paper, we will investigate the properties of the proposed penalized estimators when the models have increasing dimensional covariates.

Our main contributions in this paper are several folds. We propose local linear smoothing estimators for extended partially linear single-index models and provide very computationally efficient ways to compute the estimators. Moreover, we propose penalized local linear smoothing estimators to estimate the parameters with sparsity and to conduct variable selection simultaneously, which is more efficient when the model parameters contain sparsity. In addition, we introduce a chi-squared test statistic for general linear hypothesis testings for the parameters in this setting. Furthermore, we extend the proposed methodology to the increasing dimensional settings under certain assumptions and show the consistency of the penalized estimators.

In principle, all the results obtained in this manuscript should be applicable to PLSIM. One can estimate the model parameters under the constraints that some entries are equal to zero, and then the EPLSIM is essentially equivalent to PLSIM. However, we do not recommend doing that since EPLSIM has much higher dimensions and leads to more computational complexity. If one is able to identify and separate the covariates in the linear part and nonlinear part of the model, one should always use PLSIM instead of EPLSIM.

The rest of this paper consists of the following sections. In Sect. 2, we propose the local linear smoothing estimating methodology for estimating β , θ , and the link function $g(\cdot)$. Then we discuss the uniqueness of the solution of the optimization problem resulting from the estimation procedure, and derive the large sample theory for the estimators. In Sect. 3, we introduce the penalized local smoothing estimators and show the asymptotic properties of the proposed penalized estimators. In Sect. 4, we provide a test statistic for the linear hypothesis testings for the model parameters. In Sect. 5, we derive the asymptotic properties of the penalized estimators in the increasing dimensional settings with p being allowed to go to infinity with certain constraints. In Sect. 6, several numerical studies are conducted to assess the performance of the proposed methods. In Sect. 7, the EPLSIM is fitted to a publicly available data set about concrete slump test. Section 8 gives some additional remarks and concludes the paper. All the technical proofs of the main results are given in the supplementary material.

2 Local smoothing estimators

Formally, an extended partially linear single-index model with independent and identically distributed covariates and errors can be expressed as

$$y_i = x_i^T \beta_0 + g(x_i^T \theta_0) + e_i, \quad i = 1, 2, \dots, n, \quad (3)$$

where β_0 and θ_0 are the true model parameters, (x_i, e_i) 's are independent and identically distributed pairs of covariates and errors, and x_i and e_i are independent. For identifiability, we assume θ_0 has unit L_2 norm, namely $\|\theta_0\| = 1$, the first element of θ_0 is positive and θ_0 is orthogonal to β_0 , namely $\beta_0^T \theta_0 = 0$ [see Lin and Kulasekera (2007) for more details].

The nonparametric link function $g(u)$ and its derivative $g'(u)$ are estimated by local smoothing estimation as

$$(\hat{g}(u|\beta, \theta), \hat{g}'(u|\beta, \theta)) = \arg \min_{a,b} \sum_{i=1}^n \left\{ y_i - x_i^T \beta - a - b (x_i^T \theta - u) \right\}^2 K_h (x_i^T \theta - u),$$

where $K_h (x_i^T \theta - u) = K ((x_i^T \theta - u) / h) / h$ with a symmetric kernel function $K(\cdot)$ and bandwidth h . By basic calculations, $\hat{g}(u|\beta, \theta)$ can be expressed as [see Chen et al. (2015) for similar results for partially linear single-index models]

$$\hat{g}(u|\beta, \theta) = \sum_{i=1}^n s_i(u|\theta) (y_i - x_i^T \beta),$$

which may be abbreviated as $\hat{g}(u)$ when no confusion arises, where $s_i(u|\theta)$ depends on the kernel function and the observed data but is independent of β [see Eqs. (8) and (9)]. Explicitly,

$$s_i(u|\theta) = \frac{\left(\sum_{j=1}^n K_j t_j^2\right) K_i - \left(\sum_{j=1}^n K_j t_j\right) K_i t_i}{\left(\sum_{j=1}^n K_j\right) \left(\sum_{j=1}^n K_j t_j^2\right) - \left(\sum_{j=1}^n K_j t_j\right)^2}, \quad i = 1, 2, \dots, n,$$

where $t_j = x_j^T \theta - u$ and $K_j = K_h(t_j)$.

Then the local smoothing estimators (LSE) of β_0 and θ_0 can be computed through

$$\left(\hat{\beta}, \hat{\theta}_1\right) = \arg \min_{\beta, \theta} G(\beta, \theta) = \arg \min_{\beta, \theta} \sum_{i=1}^n \left\{ y_i - x_i^T \beta - \hat{g}(x_i^T \theta | \beta, \theta) \right\}^2, \quad (4)$$

with restriction $\hat{\beta}^T \hat{\theta}_1 = 0$. The estimate can be done in two steps with profiling. The first step is to fix θ and calculate

$$\hat{\beta}_\theta = \arg \min_{\beta^T \theta = 0} G(\beta, \theta). \quad (5)$$

Then the estimator $\hat{\theta}_1$ is

$$\hat{\theta}_1 = \arg \min_{\theta} \sum_{i=1}^n \left\{ y_i - x_i^T \hat{\beta}_\theta - \hat{g}(x_i^T \theta | \hat{\beta}_\theta, \theta) \right\}^2 = \arg \min_{\theta} G(\hat{\beta}_\theta, \theta). \quad (6)$$

Note that this is equivalent to optimizing (β, θ) in terms of β and θ simultaneously, but is easier to compute. Finally, we standardize $\hat{\theta}_1$ and get the estimator of θ by

$$\hat{\theta} = \hat{\theta}_1 / \|\hat{\theta}_1\|. \tag{7}$$

The asymptotic properties of the LSE are of great interest. We present the following regularity conditions to show the asymptotic normality and consistency of the LSE. We say that C is an inner compact subset of A if C is compact and there exists an open set B such that $C \subset B \subset A$.

Regularity Assumptions:

Assumption 1 The density of $x^T\theta$, $f_\theta(\cdot)$, is positive, bounded away from 0 and three times continuously differentiable in \mathcal{U} , an inner compact subset of $\{x^T\theta : \theta \in \Theta, x \in \mathcal{X}\}$, where Θ is the compact parameter space of θ and \mathcal{X} is the compact support of x .

Assumption 2 For any $\theta \in \Theta$, the second derivative of the function $\rho_x(u|\theta) = E(x|x^T\theta = u)$ with respect to u is bounded and continuous.

Assumption 3 The link function $g(\cdot)$ is three times continuously differentiable and $g''(\cdot) \neq 0$ on an open subinterval in \mathcal{U} .

Assumption 4 The third derivatives of $g(x^T\theta)$ and $f_\theta(x)$ with respect to x are uniformly Lipschitz continuous over $\Theta \subset \mathbb{R}^p$ for all $x \in \mathcal{X}$.

Assumption 5 The kernel function $K(\cdot)$ is a symmetric, bounded and continuously differentiable probability density function. Furthermore, $K(\cdot)$ is positive on the whole real line, \mathbb{R} , and $\int_{\mathbb{R}} |v|^i (K(v))^j dx < \infty$ ($i, j = 1, 2$).

Assumption 6 The variance of e , σ^2 , is positive, and $E(|e|^\gamma) < \infty$ for some $\gamma \geq 3$.

Assumption 7 The bandwidth h satisfies $nh^6 \rightarrow 0$ and $nh^{3+3/(\gamma-1)} / \log n \rightarrow \infty$ as $n \rightarrow \infty$.

Assumptions 1, 2, and 4 are standard regularity assumptions for covariates x in PLSIM [see Liang et al. (2010) for more details]. Assumption 3 ensures the identifiability of the parameters in PLSIM and EPLSIM. Assumption 5 is for the consistency of the kernel estimators, and Assumptions 6 and 7 ensure the asymptotic properties of the estimators.

One key question is whether Eq. (5) has a solution and if so, whether the solution is unique. Theorem 1 shows that there exists a unique solution to Eq. (5), and provides an computationally efficient way to calculate the solution. The following definitions are necessary for the introduction of Theorem 1.

Denote the covariates matrix $X = (x_1, x_2, \dots, x_n)^T$ and the response vector $Y = (y_1, y_2, \dots, y_n)^T$, and let $\mathcal{C}(X)$ be the column space of X . Define $T = X\theta$ and $n \times n$ matrix D_θ with entries

$$(D_\theta)_{ij} = \frac{(\sum_{s=1}^n K_{is}t_{is}^2) K_{ij} - (\sum_{s=1}^n K_{is}t_{is}) K_{ij}t_{ij}}{(\sum_{s=1}^n K_{is}) (\sum_{s=1}^n K_{is}t_{is}^2) - (\sum_{s=1}^n K_{is}t_{is})^2}, \quad i, j = 1, 2, \dots, n, \tag{8}$$

where

$$t_{ij} = t_j - t_i = x_j^T \theta - x_i^T \theta, \quad K_{ij} = K_h(t_{ij}), \quad i, j = 1, 2, \dots, n.$$

When $T = X\theta \neq c1_n$ for any $c \in \mathbb{R}$, where $1_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n$, by the Cauchy–Schwarz inequality, we have

$$\left(\sum_{s=1}^n K_{is} \right) \left(\sum_{s=1}^n K_{is} t_{is}^2 \right) - \left(\sum_{s=1}^n K_{is} t_{is} \right)^2 > 0.$$

Then for any $Z = (z_1, z_2, \dots, z_n)^T \in \mathbb{R}^n$ and $i = 1, 2, \dots, n$, the optimization problem

$$(\hat{a}_i, \hat{b}_i) = \arg \min_{a_i, b_i \in \mathbb{R}} \sum_{j=1}^n (z_j - a_i - b_i t_{ij})^2 K_{ij},$$

has unique solution \hat{a}_i which can be expressed as $\hat{a}_i = \sum_{j=1}^n (D_\theta)_{ij} z_j$. When $Z = Y - X\beta$, we have

$$\hat{g}(x_i^T \theta) = \sum_{j=1}^n (D_\theta)_{ij} (y_j - x_j^T \beta). \tag{9}$$

Let $\tilde{X}_\theta = (I_n - D_\theta) X$ and $\tilde{Y}_\theta = (I_n - D_\theta) Y$, where I_n is the $n \times n$ identity matrix. Assume $\theta \neq 0$. Besides, let B^+ be the Moore–Penrose inverse of any matrix B . The following theorem gives an explicit simple expression of the estimators and provides an computationally efficient way to calculate them.

Theorem 1 *Suppose $n > p \geq 2$, $1_n \notin \mathcal{C}(X)$, $\text{rank}(X) = p$, and $K(\cdot) > 0$. Then the optimization problem in (5) has a unique solution expressed as*

$$\hat{\beta}_\theta = \left(\tilde{X}_\theta^T \tilde{X}_\theta \right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta = \left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta \theta^T \right)^{-1} \tilde{X}_\theta^T \tilde{Y}_\theta. \tag{10}$$

Remark: Equation (10) provides two different methods for calculating $\hat{\beta}_\theta$ when θ is fixed. The latter equation implies that the solution $\hat{\beta}_\theta$ can be obtained by solving a linear system, which is computationally efficient and accurate. However, to solve the linear system

$$\left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta \theta^T \right) \beta = \tilde{X}_\theta^T \tilde{Y}_\theta \tag{11}$$

in any software such as R the coefficient matrix should be nonsingular. This requires $K(\cdot)$ to be strictly positive on the whole real line \mathbb{R} . Many functions with good properties satisfy this condition, such as the standard normal density $\phi(\cdot)$, but they would

tend to 0 quickly. For instance, $\phi(v)$ would be close to 0 when $|v| > 3$. Consequently, when the sample size is not large enough and h is relative small, many entries of D_θ defined in (8) would be very close to 0. This would cause computational issues since it might make the coefficient matrix in (11) close to singular. Therefore, in these cases, though calculating the Moore–Penrose inverse numerically might be computationally inefficient and lead to larger computational errors, we need to use it in order to get rid of singularity issues.

The following theorem shows the asymptotic normality of the LSE. The proof of the theorem is given in the supplementary material.

Theorem 2 *Suppose that the regularity Assumptions 1–7 are satisfied. Then we have*

$$n^{\frac{1}{2}} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{d} N(0, \sigma^2 \Gamma^+)$$

as $n \rightarrow \infty$, where

$$\Gamma = E \left(\Lambda \Lambda^T \right), \quad \Lambda = \left(\left\{ x - \rho_x(x^T \theta_0) \right\}^T, \left[g'(x^T \theta_0) \left\{ x - \rho_x(x^T \theta_0) \right\} \right]^T \right)^T.$$

The LSE are well-defined and have an computationally efficient way to compute based on the discussion above. An estimating procedure based on kernel smoothing was introduced in Xia et al. (1999). This kernel smoothing estimator (KSE) has an objective function similar to Eq. (4), with \hat{g} based on kernel smoothing estimation. Therefore, by profiling one can also obtain the profile estimator $\tilde{\beta}_\theta$ for each fixed θ , as shown in Equation (3.2) of Xia et al. (1999). However, this profile estimator is obtained by optimizing the objective function without the constraint $\beta^T \theta = 0$, which implies that the profile KSE $\tilde{\beta}_\theta$ is not guaranteed to be orthogonal to θ . Therefore, due to identifiability issues, $\tilde{\beta}_\theta$ might not be close to the true value β_0 , even if θ is very close to θ_0 . This issue is indicated by the simulation results presented in Sect. 6. To resolve this issue, we add the condition $\beta^T \theta = 0$ when optimizing the objective function S_n proposed in Xia et al. (1999) and implement the method of the Lagrange multipliers to compute the estimators. The performance of this Lagrange kernel smoothing estimator (LKSE) is also assessed as shown in Sect. 6.

3 Penalized local smoothing estimators

In a real-world problem, the true model is usually unknown and either overfitting or underfitting of the model could happen, especially when the number of parameters is relatively large but not sufficient observations are available. Therefore, in these cases, we would like to estimate the parameters and conduct a variable selection simultaneously. This motivates us to use the penalized local smoothing estimators (PLSE) to perform the data analyses. In this section, we propose the PLSE with the implementation of the LASSO penalty to carry out variable selection as well as parameter estimation. The penalized estimators $(\hat{\beta}_{\lambda_1}, \hat{\theta}_{\lambda_2})$ are defined as

$$\begin{aligned} (\hat{\beta}_{\lambda_1}, \tilde{\theta}_{\lambda_2}) &= \arg \min_{\beta^T \theta = 0} G_p(\beta, \theta) \\ &= \arg \min_{\beta^T \theta = 0} \left\{ \frac{1}{2} G(\beta, \theta) + n\lambda_1 \|\beta\|_1 + n\lambda_2 \|\theta\|_1 \right\}, \\ \hat{\theta}_{\lambda_2} &= \tilde{\theta}_{\lambda_2} / \|\tilde{\theta}_{\lambda_2}\|, \end{aligned}$$

where $G(\beta, \theta)$ is defined in Eq. (4), λ_1 and λ_2 are the tuning parameters of β and θ , respectively, $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$, and $\|\theta\|_1 = \sum_{k=1}^p |\theta_k|$. Let S and T denote the sets of the subscripts of the nonzero elements of β_0 and θ_0 respectively. For any $l \in \mathbb{R}^p$ and $A = \{i_1, i_2, \dots, i_{|A|}\} \subset \{1, 2, \dots, p\}$, let $l_A = (l_{i_1}, l_{i_2}, \dots, l_{i_{|A|}})$ be the vector containing elements of l with subscripts in A and $A^c = \{1, 2, \dots, p\} \setminus A$. Similarly, let X_A be the matrix containing columns of X corresponding to the elements in A .

The following theorem shows the asymptotic efficiency of the penalized estimators as well as their oracle property (the zero elements of the true parameters are correctly estimated as zero when the sample size is sufficiently large), which enables us to conduct the estimation and variable selection simultaneously.

Theorem 3 *Suppose that the regularity Assumptions 1–7 are satisfied and that $\lambda_i \rightarrow 0$, $n^{1/2}\lambda_i \rightarrow \infty$ for $i = 1, 2$. Then we have*

- (a) $P(\hat{\beta}_{\lambda_1 S^c} = 0 \text{ and } \hat{\theta}_{\lambda_2 T^c} = 0) \rightarrow 1$ as $n \rightarrow \infty$;
- (b)

$$n^{1/2} \begin{pmatrix} \hat{\beta}_{\lambda_1 S} - \beta_{0S} \\ \hat{\theta}_{\lambda_2 T} - \theta_{0T} \end{pmatrix} \xrightarrow{d} N(0, \sigma^2 \Gamma_r^+),$$

where

$$\Gamma_r = E(\Lambda_r \Lambda_r^T), \Lambda_r = \left(\left\{ x_S - \rho_{x_S}(x^T \theta_0) \right\}^T, \left[g'(x^T \theta_0) \left\{ x_T - \rho_{x_T}(x^T \theta_0) \right\} \right]^T \right)^T.$$

In practice, the tuning parameters, λ_1 and λ_2 , can be chosen via cross-validation. Other methods, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), can also be applied to determine λ_1 and λ_2 (see Liang et al. (2010) for more details). For the bandwidth h , one can use some adaptive methods such as the Lepski procedure to determine its value. However, in the simulation studies, we found that the value of h does not have a noticeable impact on the estimation efficiency as long as the values are in a reasonable range (see the table for MSEs in the supplementary material). We believe that choosing h via cross-validation is good enough in our case and it can help reduce the computational burden.

To select the values of h and λ_j via a K -fold cross-validation, the data set is randomly divided into K folds. For a specific value of $(h, \lambda_1, \lambda_2)$ and each fold k , the parameters and the nonparametric function are estimated based on the data with the k th fold removed, and predictions of the response values of the k th fold are performed based on the estimation. Then the mean of the K MSEs is obtained for each value of $(h, \lambda_1, \lambda_2)$ and the optimal value of the tuning parameters can be obtained by minimizing the mean MSE.

If the distribution of X is close to a degenerated distribution, the proposed method might have bad performance and could lead to numerical issues in practice. Theoretically, we assume that x_i are independent and identically distributed so that when $n \rightarrow \infty$, and the sign of the derivative of the objective function with respect to a single parameter is controlled by the tuning parameters. Therefore, as long as the model is identifiable (i.e., X has a full rank distribution), we have the oracle property when p is fixed and $n \rightarrow \infty$.

4 Hypothesis testing

Consider the general linear hypothesis

$$H_0 : W\xi = 0 \quad \text{versus} \quad H_1 : W\xi \neq 0, \tag{12}$$

where $\xi = (\beta^T, \theta^T)^T$ and W is a $m \times 2p$ full rank matrix. Let Ω_0 and Ω_1 be the parameter spaces of H_0 and H_1 , respectively. Define

$$G(H_0) = \inf_{\xi \in \Omega_0} G(\xi), \quad G(H_1) = \inf_{\xi \in \Omega_1} G(\xi),$$

and the test statistic

$$V = \frac{n \{G(H_0) - G(H_1)\}}{G(H_1)}. \tag{13}$$

Then we have the following theorem for testing the hypotheses in (12).

Theorem 4 *Suppose that the regularity Assumptions 1–7 are satisfied. We have:*

- (a) *under H_0 in (12), $V \rightarrow \chi_m^2$ in distribution;*
- (b) *under H_1 in (12), the test is consistent;*
- (c) *under the local alternative of $n^{1/2}W\xi \rightarrow d$ for some m dimensional $d \neq 0$, V converges in distribution to a noncentral chi-squared distribution with m degrees of freedom and noncentrality parameter*

$$\psi = \sigma^{-2}d^T (W\Gamma + W^T)^{-1}d,$$

where Γ is defined in Theorem 2.

5 Increasing dimensional settings

In this section, we consider the PLSE of EPLSIM with covariate dimension $p \rightarrow \infty$. Recall that x is the p -dimensional covariate vector and we assume $E(x) = 0$ for simplicity. let $\text{Cov}(x) = E(xx^T)$ be the covariance matrix of x . In addition to the regularity assumptions mentioned in Section 2, we introduce the following assumption to ensure the consistency of the PLSE as $p \rightarrow \infty$:

Assumption 8 The largest eigenvalue (spectra radius) of $\text{Cov}(x)$, denoted by $\gamma(\text{Cov}(x))$, is bounded for all p : $\sup_p \gamma(\text{Cov}(x)) < \infty$.

This is a standard and widely used regularity assumption for LASSO estimators in linear models. See, for example, Huang et al. (2008) and Zhang and Huang (2008) for more details. Let c_{jk} denote the (j, k) -th element of $\text{Cov}(x)$, which is just the covariance of the j^{th} and k^{th} element of x . Then by the Gershgorin Circle Theorem, a sufficient condition for Assumption 8 is:

Assumption 8'. The covariance function $\text{Cov}(x)$ satisfies $\sup_j \sum_{k=1}^{\infty} |c_{jk}| < \infty$.

If the elements of x are (weakly) stationary and $c_{jk} = c_{|j-k|}$, the regularity assumption above is equivalent to $\sum_{s=0}^{\infty} |c_s| < \infty$.

Note that when Assumption 8 or 8' is satisfied, we have $x^T \theta = O_p(1)$ since the parameter space is compact and

$$\sup_p E\left(\|x^T \theta\|_2^2\right) \leq \sup_p \gamma(\text{Cov}(x)) \cdot \sup_{\theta} \|\theta\|^2 < \infty.$$

Similarly, we have $x^T \beta = O_p(1)$. Since the penalized objective function $G_p(\beta, \theta)$ does not possess convexity, it is difficult to show the usual convergence rate, $O_p(\sqrt{\log p/n})$, of the penalized estimators, as presented in many existing methods of LASSO estimators for linear models (see, e.g., Chapter 11 of Hastie et al. (2015)). However, based on equation (5), we can still prove the consistency of the PLSE at a slower convergence rate as shown in the following theorem.

Theorem 5 Suppose that the regularity Assumptions 1–8 are satisfied. In addition, assume that $p \log p/n \rightarrow 0$, $\lambda_i \rightarrow 0$, $n^{1/2} \lambda_i \rightarrow \infty$, and $\lambda_i = o_p(\sqrt{\log p/n})$ for $i = 1, 2$. Then we have:

(a)

$$\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} = O_p\left(\sqrt{\frac{p \log p}{n}}\right);$$

(b) For each fixed $k = 1, 2, \dots$, if $\beta_{0k} = 0$ or $\theta_{0k} = 0$, then $P(\hat{\beta}_k = 0) \rightarrow 1$ or $P(\hat{\theta}_k = 0) \rightarrow 1$ as $n \rightarrow \infty$.

Note that we only propose an effective computational procedure for non-penalized estimators since the solution has a linear expression. When the number of parameters is moderate, conducting the linear hypothesis testing with non-penalized estimation is more efficient compared with that with penalized estimation.

6 Simulation study

In this section, we evaluate our proposed methods empirically, and compare them with the method introduced in Xia et al. (1999) via simulation. We provide five examples here. The first example is from Xia et al. (1999), where the number of the parameters

is relatively small. In the second example, the number of the parameters is relatively large, where the penalized estimators are expected to have better performance. The third example is about hypothesis testings in extended partially linear single-index models. The fourth example is about penalized estimation in models with increasing dimensional covariates. In the fifth example, the covariates in the linear and nonparametric parts of the predictors are disjoint.

Note that Assumption 1 holds for all of the examples since we essentially let u vary in an inner compact subset of $\{x^T\theta : \theta \in \Theta, x \in \mathcal{X}\}$ when we carry out discrete computations. For instance, in R for computations of normally distributed covariates, $u \in [-10^{32}, 10^{32}]$, which is a compact subset of \mathbb{R} . Assumption 8 also holds for all of the examples since the elements of x in these examples are either moving average series or independent series.

Example 1 We firstly considered the example shown in Xia et al. (1999), which can be written as

$$y_i = 0.3x_i + 0.4x_{i-1} + \exp\left\{-2(0.8x_i - 0.6x_{i-1})^2\right\} + 0.1e_i,$$

where

$$x_i = 0.8x_{i-1} + \epsilon_i + 0.5\epsilon_{i-1}, \quad e_i, \epsilon_i \sim N(0, 1),$$

and all e_i, ϵ_i are independent of each other. The model above can also be expressed as

$$y_i = \beta_1 x_i + \beta_2 x_{i-1} + \exp\left\{-2(x_i \cos \alpha - x_{i-1} \sin \alpha)^2\right\} + 0.1e_i, \quad (14)$$

where $\beta_1 = 0.3, \beta_2 = 0.4$, and $\alpha = \arcsin(0.6) = 0.6435$.

Five different methods were used in this example: the KSE introduced in Xia et al. (1999), the LKSE and the LSE described in Sect. 2, the penalized kernel smoothing estimators (PKSE) and the PLSE proposed in Sect. 3. Although penalized estimation was not discussed in Xia et al. (1999), for comparison purposes, we could simply add the LASSO penalty to the objective function of their KSE to obtain the PKSE. We simulated 500 independent data sets with sample size $n = 50, n = 100$, and $n = 200$. The estimation procedure of the KSE is the same as shown in Xia et al. (1999). For the LSE, it is expected that h is determined by cross-validation for each simulated data set. However, due to a substantial computational burden, for each n , we firstly ran the simulation with a wide range of h (say h ranges from 0.01 to 1) but with a small number of replications (50 for each value of h). We observed that the MSE started increasing noticeably when h was less than 0.1 or greater than 0.3. We then fixed $h \in [0.1, 0.3]$, ran 500 replications, and calculated the mean squared error of the parameters. Then h_n was obtained by minimizing the mean squared error. After h was determined, the tuning parameters λ_1 and λ_2 were determined in a similar way. All optimizations were done in R via the `nloptr()` function from the `nloptr` package (Johnson (2020)).

Table 1 Simulation results for model (14)

n	Parameters	Methods	50		100		200	
			Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1		KSE	-0.0040	0.3045	0.0090	0.3094	0.0171	0.3056
		LKSE	0.0004	0.0068	-0.0001	0.0040	0.0000	0.0021
		LSE	-0.0008	0.0054	-0.0005	0.0033	-0.0002	0.0020
		PKSE	-0.0010	0.0154	-0.0009	0.0071	-0.0007	0.0035
		PLSE	-0.0024	0.0087	-0.0014	0.0048	-0.0009	0.0029
β_2		KSE	0.0030	0.2284	-0.0067	0.2322	-0.0130	0.2291
		LKSE	-0.0006	0.0072	0.0001	0.0042	-0.0001	0.0025
		LSE	0.0003	0.0059	0.0003	0.0037	0.0000	0.0023
		PKSE	-0.0010	0.0122	0.0002	0.0063	-0.0006	0.0033
		PLSE	-0.0003	0.0069	0.0005	0.0044	-0.0005	0.0027
α		KSE	0.0014	0.0121	-0.0005	0.0066	0.0002	0.0039
		LKSE	0.0013	0.0120	-0.0003	0.0065	0.0001	0.0039
		LSE	-0.0016	0.0087	-0.0011	0.0054	-0.0003	0.0034
		PKSE	-0.0005	0.0337	-0.0017	0.0161	-0.0005	0.0078
		PLSE	-0.0036	0.0171	-0.0028	0.0099	-0.0009	0.0059

For each value of the sample size n , 500 replications were simulated to compute the bias and RMSE

Table 1 shows the bias and the square root of mean squared error (RMSE) for the five methods obtained in our simulations. Although the KSE and the LKSE for α have similar performance, the LKSE for β is significantly better than the KSE of β in terms of bias and mean squared error. Furthermore, while the biases of the LSE and LKSE are similar with both of them being nearly negligible relative to the RMSE, the RMSE of the LSE is noticeably smaller than the RMSE of the LKSE, especially when the sample size n is relatively small. In addition, both penalized estimators have worse performance compared with the two non-penalized estimators, even for a relatively small sample size. The main reason is that the parameters are not sparse here. Moreover, for sample size $n = 200$, we compared the computation time of calculating the LSE using and without using Theorem 1. About 42% of the computation time had been reduced by applying the results of Theorem 1, which implies that the calculation methods provided by Theorem 1 have successfully accelerated the estimating procedure.

We assume i.i.d. in all the theorems and proofs for simplicity, although the results might still hold for more complicated correlation structures. Example 1 includes an $ARMA(1, 1)$ process which violates the i.i.d. assumption, but we still include it here since it was used in Xia et al. (1999) when EPLSIM was firstly introduced. In this example, we examined the robustness of the proposed approach in a model misspecified case. In addition, this allows us to compare our methods with the method introduced by Xia et al. (1999). The empirical results suggest that the asymptotic properties might still hold even if the correlation structure is more complicated.

Table 2 Simulation results for model (15)

n	Parameters	Methods	50			100			200		
			RMSE	C	I	RMSE	C	I	RMSE	C	I
β		KSE	0.5492	0.91	0	0.5183	1.34	0	0.4314	1.94	0
		LKSE	0.1158	1.01	0	0.0630	1.67	0	0.0427	2.31	0
		LSE	0.1171	1.11	0	0.0676	1.77	0	0.0412	2.44	0
		PKSE	0.0436	2.03	0	0.0278	2.53	0	0.0170	2.82	0
		PLSE	0.0280	2.51	0	0.0199	2.75	0	0.0132	2.89	0
θ		KSE	0.0906	0.66	0	0.0602	0.87	0	0.0402	1.19	0
		LKSE	0.0951	0.60	0	0.0578	0.92	0	0.0391	1.20	0
		LSE	0.0984	0.52	0	0.0583	0.85	0	0.0354	1.22	0
		PKSE	0.0196	2.75	0	0.0137	2.77	0	0.0059	2.97	0
		PLSE	0.0057	2.99	0	0.0048	2.98	0	0.0029	3.00	0

C, the average number of the true zero parameters that were correctly set to zero (less than 0.01); I, the average number of the truly nonzero parameters that were incorrectly set to zero. For each value of the sample size n , 500 replications were simulated to compute the RMSE, C and I

Example 2 We now consider an extended partially linear single-index model with more parameters, which is model (3) with link function and parameters

$$g(u) = (1 + u^2)^{-1}, \quad \beta_0 = (2, -1, 0, 0, 0), \quad \theta_0 = (1, 2, 0, 0, 0)/\sqrt{5}, \quad (15)$$

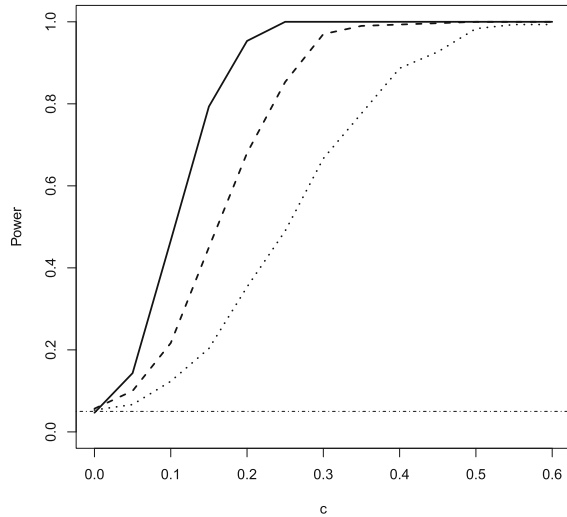
and the covariates and random errors are independent and identically distributed as

$$x_{ij} \sim N(0, 1), \quad e_i \sim N(0, 0.1^2).$$

We simulated 500 independent data sets with sample sizes $n = 50$, $n = 100$, and $n = 200$ from this model. Since the sample size is relatively small and the model parameters are sparse, the penalized estimators are expected to have better performance.

Table 2 shows the RMSE, the average number of the true zero parameters that were correctly set to zero and the average number of the truly nonzero parameters that were incorrectly set to zero for the five methods obtained in our simulations. Again, for β , the mean squared error of the LKSE is significantly smaller than the mean squared error of the KSE. This implies that the implementation of the method of Lagrange multipliers has led to huge improvement in performance of the KSE. Besides, the results in Table 2 also indicate that while the LKSE and the LSE have similar performance, the penalized estimators have much better performance compared with the estimators without penalty. Although the computation was heavy, we also tried to simulate a small number of replications for $n = 400$, and the results are similar to those for $n = 200$. Therefore, we can conclude that even if the sample size is relatively large, the penalized estimators are more preferable as long as the model contains sparsity. In addition, Table 2 also indicates that PLSE perform significantly better than the PKSE, especially when the sample size is relatively small ($n = 50$ and $n = 100$).

Fig. 1 The power function (or type I error when $c = 0$) versus c for sample size $n = 50$ (dotted), $n = 100$ (dashed), and $n = 200$ (solid). The nominal level is equal to 0.05 (horizontal dot-dash). For each value of the sample size n , 300 replications were simulated



Example 3 To investigate the performance of the test statistic V described in Sect. 4, we consider model (3) with link function and parameters

$$g(u) = 3u^2, \quad \beta_0 = (2, -3, c, c), \quad \theta_0 = (3, 2, 0, 0)/\sqrt{13},$$

where c ranges from 0 to 0.6 with increment 0.05. The covariates and random errors are independent and identically distributed as

$$x_{ij} \sim U(0, 1), \quad e_i \sim N(0, 0.1^2).$$

For each value of c , we simulated 300 independent data sets with sample sizes $n = 50$, $n = 100$, and $n = 200$ from the model, and considered the following null and alternative hypotheses:

$$H_0 : \beta_3 = \beta_4 = 0 \quad \text{versus} \quad H_1 : \beta_3 = \beta_4 = c > 0$$

with the nominal level equal to 0.05. The power function (or type I error when $c = 0$) versus c is plotted in Fig. 1.

Figure 1 shows that when $c = 0$, the type I error of the test is equal to 0.05, 0.06, and 0.05 for $n = 50$, $n = 100$, and $n = 200$, respectively, which is close to the nominal level apart from its standard errors. Also, Fig. 1 implies that the power function increases quite fast as c increases. Overall, V leads to a powerful test whose size is also well controlled.

Example 4 Consider model (15) but in the increasing dimensional setting with $p = \lfloor 0.85 * \sqrt{n} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. Thus, for $n = 50, 100, 200$, $p = 6, 8, 12$, respectively. We simulated 500 independent data sets for each value of (n, p) , and the results for PKSE and PLSE are shown in Table 3.

Table 3 Simulation results for model (15) with increasing dimensional settings

(n, p)	Methods	(50, 6)			(100, 8)			(200, 12)		
		RMSE	AC	AI	RMSE	AC	AI	RMSE	AC	AI
β	PKSE	0.0303	0.81	0	0.0205	0.95	0	0.0139	0.98	0
	PLSE	0.0256	0.88	0	0.0167	0.97	0	0.0126	0.99	0
θ	PKSE	0.0065	1.00	0	0.0039	1.00	0	0.0033	1.00	0
	PLSE	0.0059	1.00	0	0.0032	1.00	0	0.0031	1.00	0

AC, the average number of the true zero parameters that were correctly set to zero (less than 0.01) divided by the number of true zero parameters; AI, the average number of the truly nonzero parameters that were incorrectly set to zero divided by the number of true nonzero parameters. For each value of the sample size n , 500 replications were simulated to compute the RMSE, MC and MI

The results in Table 3 indicate that for both methods, the RMSE and the average number of the true zero parameters that were correctly set to zero divided by the number of true zero parameters (AC) increase as (n, p) increases. In this example, PLSE continues to perform better compared with PKSE, especially when the sample size n is relatively small.

The normal Q-Q plots of $\hat{\beta}_1$ and $\hat{\theta}_1$ for different values of n and p are shown in Figure 2. The results imply that the distributions of the estimators are not close to normal when the sample size is relatively small or even moderate. The distributions of the estimators are either converging to normal distributions slowly or even not converging to normal distributions. This is an interesting question for further research.

Example 5 We now consider model (15) with different values of the parameters listed as follows:

$$\beta_0 = (0, 0, 1), \quad \theta_0 = (3, 1, 0)/\sqrt{10}. \tag{16}$$

Note that in this scenario, the sets of indices of non-zero elements of β_0 and θ_0 are disjoint, and thus this is indeed a PLSIM. We simulated 200 independent data sets with sample sizes $n = 50$, $n = 100$, and $n = 200$ from this model. We treated the model as a PLSIM (3 parameters) and an EPLSIM (6 parameters), and estimated the parameters using KSE, LSE, PKSE, and PLSE. The results are shown in Table 4.

As shown in Table 4, the estimation is more accurate when the model is treated as a PLSIM. The RMSEs of KSE(NE) and LSE(NE) are smaller than the RMSEs of LKSE(ET), LSE(ET), PKSE(ET), and PLSE(ET) since the dimension of the parameter space is twice as large when the model is treated as an EPLSIM. As expected, in this scenario, PKSE and PLSE are naturally no longer the best methods in this non-extended PLSIM since there are very few number of parameters. However, they still perform reasonably well when the sample size is sufficiently large. More generally, this example illustrates a price to be paid for employing an unnecessarily complex model when a simpler model is known to be valid.

Finally, note that we did not include the method proposed by Dong et al. (2016). Their method is based on orthogonal series expansion in $L^2(\mathbb{R})$, whose performance

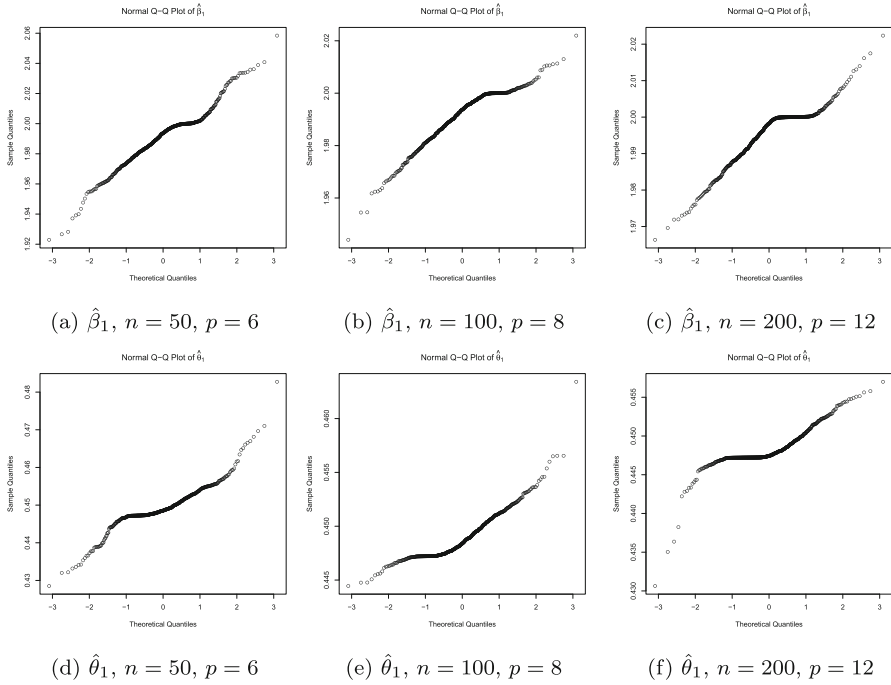


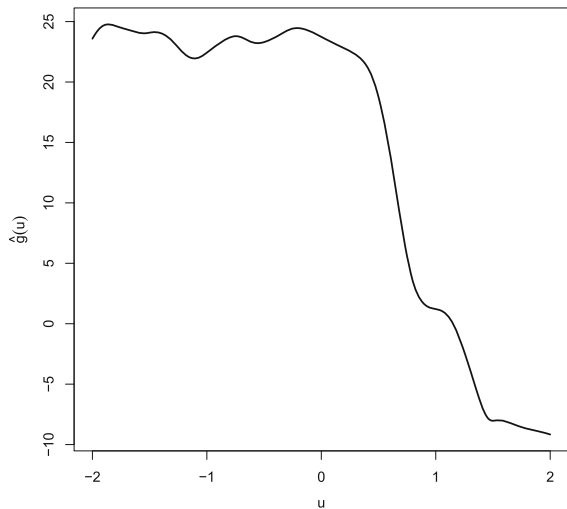
Fig. 2 Normal Q-Q plots of $\hat{\beta}_1$ and $\hat{\theta}_1$ for different values of n and p

Table 4 Simulation results for model (15) with parameters shown in (16)

n	Parameters	Methods	50		100		200	
			Bias	RMSE	Bias	RMSE	Bias	RMSE
β		KSE(NE)	0.0008	0.0196	0.0008	0.0116	0.0011	0.0077
		LSE(NE)	0.0010	0.0165	0.0016	0.0107	0.0001	0.0074
		LKSE(ET)	-0.0055	0.0621	-0.0011	0.0349	-0.0004	0.0243
		LSE(ET)	0.0003	0.0562	0.0017	0.0306	0.0007	0.0216
		PKSE(ET)	-0.0245	0.0551	-0.0070	0.0262	-0.0022	0.0135
		PLSE(ET)	-0.0138	0.0222	-0.0059	0.0141	-0.0023	0.0115
θ		KSE(NE)	-0.0008	0.0485	-0.0022	0.0280	-0.0005	0.0190
		LSE(NE)	-0.0004	0.0388	-0.0002	0.0246	0.0006	0.0168
		LKSE(ET)	-0.0003	0.0665	0.0009	0.0403	0.0008	0.0292
		LSE(ET)	-0.0057	0.0660	-0.0026	0.0371	-0.0004	0.0256
		PKSE(ET)	0.0243	0.0796	0.0097	0.0414	0.0061	0.0222
		PLSE(ET)	0.0105	0.0487	0.0047	0.0396	0.0017	0.0200

KSE(NE), LSE(NE) with the model treated as PLSIM (non-extended); LKSE(ET), LSE(ET), PKSE(ET), and PLSE(ET) with the model treated as EPLSIM (extended). For each value of the sample size n , 200 replications were simulated to compute the bias and RMSE

Fig. 3 The value of $\hat{g}(u)$ versus u obtained from the concrete slump test data set



highly depends on the similarity between the nonparametric function and the orthogonal basis they choose. We are able to show some simulations results where our method easily outperforms their method (say we choose the No. 100 function of the orthogonal basis) and vice versa.

7 Real data application

We applied the proposed methods to analyze a publicly available data set of concrete slump test data, which was firstly introduced and analyzed in Yeh (2007) [see Yeh (2006) and Yeh (2007) for more related information]. The high-performance concrete is highly complex, and thus it is very difficult to model its behavior using available information. In this data set, there are 7 input covariates: cement (kg/m^3), blast furnace slag (kg/m^3), fly ash (kg/m^3), water (kg/m^3), superplasticizer (kg/m^3), coarse aggregate (kg/m^3), and fine aggregate (kg/m^3), and 3 output variables: concrete slump (cm), concrete flow (cm) and 28-day compressive strength (mpa). We focused on modeling the concrete slump using all the 7 available input covariates. The data set contains 103 observations and a multiple linear regression model yields an R^2 value of 0.32. Some further exploratory analysis indicates strong nonlinear relationships between the concrete slump and the covariate variables, which leads to the use of nonlinear models for predictions and simulations of concrete slump (Yeh 2008, 2009).

The local smoothing estimation, Lagrange kernel smoothing estimation and their penalized versions were applied to the analysis of the data set after the covariates were standardized. For the penalized local smoothing estimation, we firstly performed a 10-fold cross-validation to the data set to select the bandwidth and the tuning parameters. Then we computed the estimates of the parameters with the tuning parameters and obtained

$$\begin{aligned}\hat{\theta} &= (0.152, 0.558, 0.031, -0.722, -0.166, -0.302, -0.158)^T, \\ \hat{\beta} &= (10.817, 12.927, 12.570, 6.056, 0.032, 11.407, 8.975)^T.\end{aligned}$$

The results indicate that the third element of $\hat{\theta}$ and the fifth element of $\hat{\beta}$ are effectively zero. We also performed a hypothesis test to test for $\theta_j = 0$ or $\beta_j = 0$ based on the proposed method. The obtained p -values associated with β_5 , θ_3 , and θ_7 are less than 0.05 (see the supplementary material for the exact p -values). Therefore, the nonparametric part of the model might not depend on the third covariate (fly ash), while the linear part of the model might not depend on the fifth covariate (superplasticizer).

Figure 3 shows the estimate of the link function $\hat{g}(u)$ with $-2 \leq u \leq 2$ obtained from the penalized local smoothing estimation. The function drops rapidly when $u > 0.5$. These estimates of parameters and the link function yields an R^2 value equal to 0.82, while the R^2 values obtained by using the local smoothing estimation, Lagrange kernel smoothing estimation and penalized kernel smoothing estimation are 0.57, 0.44, and 0.47, respectively. The R^2 value of the penalized kernel smoothing estimation is much smaller than the R^2 value of the penalized local smoothing estimation. This, together with the results of the simulations shown in Sect. 6, implies that the penalized local smoothing estimation has better performance and is more robust, especially for real world problems when no prior information of the model parameters is available. Overall, the local smoothing estimation method has the best performance among all the estimation methods, while the other three methods also lead to substantial improvements compared with the simple linear model approach.

8 Discussion

In this paper, we considered the EPLSIM (3), which are more flexible compared with the PLSIM (1). However, extended partially linear single-index models often have more parameters, making it more difficult to estimate the parameters. We proposed the LSE in Sect. 2 for parameter estimation, and introduced the chi-squared test statistic in Sect. 4 for testing general linear hypotheses. Furthermore, for data sets with too many covariates (which lead to sparse parameters), we proposed the PLSE in Sect. 3 for conducting parameter estimation and variable selection simultaneously, and studied its properties in the increasing dimensional setting with certain constraints. The uniqueness and linear expressions of the solution to the optimization of the profile objective function are shown in Sect. 2, resulting in fast and accurate computations for the solution. In addition, the performance of the KSE introduced in Xia et al. (1999) can be improved by implementing the method of Lagrange multipliers to calculate the profile estimator. Besides, asymptotic properties of the proposed estimators and test statistic were also introduced and discussed in detail.

Simulation studies were presented in Sect. 6 to assess the performance of the proposed estimators and test statistic. We compared the five estimation methods for a model containing a small number of parameters introduced in Xia et al. (1999), and for a model containing more parameters. The simulation results indicate that the LKSE have much better performance compared with the KSE, especially for β . Besides, the

results of the first example implies the LSE perform better than the LKSE, and the results of the second example implies the PLSE perform better than the PKSE. The results also indicate that the penalized estimators would generally outperform the non-penalized estimators when the model contains sparsity. For the test statistic V defined in (13), the simulation results show that it is powerful with good size control.

An interesting real-world data set of concrete slump test data was analyzed in Sect. 7. We fitted the EPLSIM to the data and used introduced methods to estimate the parameters and the link function. The estimated link function $\hat{g}(u)$ shown in Fig. 3 has a special pattern, which might result from some characteristic of the data. The fitted R^2 was more than doubled to 0.82 from 0.32 by fitting the EPLSIM with PLSE instead of a multiple linear regression model.

As a future research problem, it would be interesting to study the EPLSIM with more complicated correlation structure. For instance, the covariates could be time series with auto correlation, or the measurements are taken from different subjects over time as in longitudinal data.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11749-022-00845-8>.

Acknowledgements We would like to thank the Associate Editor and three referees for their constructive and helpful comments and suggestions that substantially improved an earlier version of this paper. This research was supported in part by the Simons Foundation Mathematics and Physical Sciences–Collaboration Grants for Mathematicians Program Award #499650.

References

- Cai L, Jin L, Wang S (2020) Oracally efficient estimation and simultaneous inference in partially linear single-index models for longitudinal data. *Electron J Stat* 14:2395–2438
- Cai Q, Wang S (2019a) Efficient estimation in partially linear single-index models for longitudinal data. *Scand J Stat* 46:116–141
- Cai Q, Wang S (2019b) Inferences with generalized partially linear single-index models for longitudinal data. *J Stat Plan Inference* 200:146–160
- Carroll RJ, Fan J, Gijbels I, Wand MP (1997) Generalized partially linear single-index models. *J Am Stat Assoc* 92:477–489
- Chen J, Li D, Liang H, Wang S (2015) Semiparametric gee analysis in partially linear single-index models for longitudinal data. *Ann Statist* 43:1682–1715
- Dong C, Gao J, Tjøstheim D (2016) Estimation for single-index and partially linear single-index integrated models. *Ann Statist* 44:425–453
- Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman & Hall, London, UK
- Hastie T, Tibshirani R, Wainwright M (2015) Statistical learning with sparsity: the lasso and generalizations. Chapman & Hall, Florida, USA
- Horowitz JL, Härdle W (1996) Direct semiparametric estimation of single-index models with discrete covariates. *J Am Stat Assoc* 91:1632–1640
- Huang J, Ma S, Zhang C (2008) Adaptive lasso for sparse high-dimensional regression models. *Stat Sin* 18:1603–1618
- Ichimura H (1993) Semiparametric least squares (sls) and weighted sls estimation of single-index models. *J Econom* 58:71–120
- Johnson SG (2020) The nlopt nonlinear-optimization package. <https://cran.r-project.org/web/packages/nloptr/nloptr.pdf>
- Li Y, Hsing T (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann Statist* 38:3321–3351

- Liang H, Liu X, Li R, Tsai C (2010) Estimation and testing for partially linear single-index models. *Ann Statist* 38:3811–3836
- Liang H, Wang N (2005) Partially linear single-index measurement error models. *Stat Sin* 15:99–116
- Lin W, Kulasekera KB (2007) Identifiability of single-index models and additive-index models. *Biometrika* 94:496–501
- Ma Y, Zhu L (2013) Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *J R Statist Soc B* 75:305–322
- Radchenko P (2015) High dimensional single index models. *J Multivar Anal* 139:266–282
- Wang N, Carroll RJ, Lin X (2005) Efficient semiparametric marginal estimation for longitudinal/clustered data. *J Am Stat Assoc* 100:147–157
- Wang T, Xu P, Zhu L (2012) Non-convex penalized estimation in high-dimensional models with single-index structure. *J Multivar Anal* 109:221–235
- Xia Y, Härdle W (2006) Semi-parametric estimation of partially linear single-index models. *J Multivar Anal* 97:1162–1184
- Xia Y, Tong H, Li WK (1999) On extended partially linear single-index models. *Biometrika* 86:831–842
- Xia Y, Tong H, Li WK, Zhu L (2002) An adaptive estimation of dimension reduction space. *J R Statist Soc B* 64:363–410
- Yeh I (2006) Exploring concrete slump model using artificial neural networks. *J Comput Civil Eng* 20:217–221
- Yeh I (2007) Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement Concrete Comp* 29:474–480
- Yeh I (2008) Prediction of workability of concrete using design of experiments for mixtures. *Comput Concrete* 5:1–20
- Yeh I (2009) Simulation of concrete slump using neural networks. *Construct Mater* 162:11–18
- Yu Y, Ruppert D (2002) Penalized spline estimation for partially linear single-index models. *J Am Stat Assoc* 97:1042–1054
- Zhang C, Huang J (2008) The sparsity and bias of the lasso selection in high-dimensional linear regression. *Ann Statist* 36:1567–1594

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.