STATISTICS
2018, VOL. 52, NO. 6, 1193–1211
https://doi.org/10.1080/02331888.2018.1506922

Statistical inference on asymptotic properties of two estimators for the partially linear single-index models

Jing Yang[a], Fang Lu[a] and Hu Yang[b]

[a]Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), College of Mathematics and Statistics, Hunan Normal University, Changsha, People's Republic of China;
[b]College of Mathematics and Statistics, Chongqing University, Chongqing, People's Republic of China

ABSTRACT

The outer product of gradients (OPG) estimation procedure based on least squares (LS) approach has been presented by Xia et al. [An adaptive estimation of dimension reduction space. J Roy Statist Soc Ser B. 2002;64:363–410] to estimate the single-index parameter in partially linear single-index models (PLSIM). However, its asymptotic property has not been established yet and the efficiency of LS-based method can be significantly affected by outliers and heavy-tailed distributions. In this paper, we firstly derive the asymptotic property of OPG estimator developed by Xia et al. [An adaptive estimation of dimension reduction space. J Roy Statist Soc Ser B. 2002;64:363–410] in theory, and a novel robust estimation procedure combining the ideas of OPG and local rank (LR) inference is further developed for PLSIM along with its theoretical property. Then, we theoretically derive the asymptotic relative efficiency (ARE) of the proposed LR-based procedure with respect to LS-based method, which is shown to possess an expression that is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. Moreover, we demonstrate that the new proposed estimator has a great efficiency gain across a wide spectrum of non-normal error distributions and almost not lose any efficiency for the normal error. Even in the worst case scenarios, the ARE owns a lower bound equalling to 0.864 for estimating the single-index parameter and a lower bound being 0.8896 for estimating the nonparametric function respectively, versus the LS-based estimators. Finally, some Monte Carlo simulations and a real data analysis are conducted to illustrate the finite sample performance of the estimators.

ARTICLE HISTORY
Received 27 August 2015
Accepted 9 July 2018

KEYWORDS
Partially linear single-index models; outer product of gradients; local rank regression; asymptotic property; asymptotic relative efficiency

AMS SUBJECT CLASSIFICATIONS
62G08; 62G05; 62G20

1. Introduction

Partially linear single-index models, which allow the response variable to depend linearly on some covariates and nonlinearly on a single linear combination of other variables called as single-index through a nonparametric link function, have attracted much attention in recent decades since its first introduction by Carroll et al. [1], because this class of models inherits both the explanatory power of parametric models and the flexibility of the single-index model. Specifically, let Y be the univariate response variable, X and Z are the associated p-dimensional and q-dimensional covariates, respectively. The PLSIM can be written as follows:

$$Y = g(X^T\beta_0) + Z^T\theta_0 + \varepsilon, \tag{1}$$

CONTACT Jing Yang yang2009jing@163.com Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), College of Mathematics and Statistics, Hunan Normal University, Changsha, 410081, China

where ε is the random error follows some probability density function $h(\cdot)$, $g(\cdot)$ is an unknown univariate smooth function, $\beta_0 \in R^p$ and $\theta_0 \in R^q$ are the corresponding parameter vectors. For the sake of model identification, it is often assumed that $\|\beta_0\|_2 = 1$ and the first component of β_0 is positive, where $\|\cdot\|$ denotes the Euclidean norm.

Various estimation procedures have been proposed for model (1) in the literature. Carroll et al. [1] developed a backfitting algorithm under a generalized version of PLSIM. But the resulting estimators may be unstable, Yu and Ruppert [2] proposed a penalized spline estimation procedure. Xia and Li [3] obtained the estimator of β_0 via the least squares method combined with kernel smoothing. Xia and Härdle [4] studied this model by combining the minimum average variance estimation (MAVE, [5]) method with the idea of dimension reduction. Zhu and Xue [6] presented the empirical likelihood-based inference for the parameters in PLSIM. Wang et al. [7] proposed a two-stage estimation procedure under the extra condition that a few indices based on X suffice to explain Z. Liang et al. [8] proposed a profile least squares approach for model (1). Nevertheless, all these existing procedures were built on least squares or likelihood based method, which may suffer in the presence of outliers or heavy-tail error distributions. Especially, when the error follows Cauchy distribution, the local LS estimate will break down. To overcome this drawback, considerable efforts have been devoted to constructing robust estimates and quantile regression proposed by Koenker and Bassett [9] has received much attention since it can provide richer information than the classic mean regression, related literature see [10–13] for a comprehensive review. Later on, the quantile regression method has also been extended to PLSIM but with limited results. Fan and Zhu [14] presented a quantile regression version of OPG estimation method to estimate the single-index parameter in general semi-parametric models. Yang and Yang [15] proposed a stepwise least absolute deviation procedure in PLSIM. Lv et al. [16] proposed minimizing average check loss estimation procedure to conduct quantile regression of PLSIM.

Although the quantile based estimator is robust, its efficiency is proportional to the density at the median. Accordingly, it may be inefficient for Gaussian distributions, worse still, the efficiency can be arbitrarily small if $h(0)$ is close to zero. Recently, Leng [17] proposed a shrinkage estimator for the classical linear model using the idea of rank regression. Wang et al. [18] proposed a LR estimation procedure based on lank regression and local polynomial smoothing for varying-coefficient models. Feng et al. [19] studied the single-index models by combining the ideas of OPG and rank-based regression inference. Sun and Lin [20] developed a robust estimation method for the varying-coefficient partially linear model via LR technique. All these proposed estimators based on LR method are shown to be highly efficient across a wide class of error distributions and own comparable efficiency in the worst case scenario. To the best of our knowledge, no such literature exists for the PLSIM up to now. Motivated by these observations, here we devoted to extending the LR regression method to PLSIM, and the main focus of this paper is on the robust estimation of the single-index parameter β_0 in model (1). Specifically, this paper mainly offer the following three contributions: (i) We theoretically establish the asymptotic property of the LS-based OPG estimate of β_0 proposed by Xia et al. [5]; (ii) We develop an efficient estimator for β_0 by integrating the ideas of LR regression and OPG approach, and its asymptotic property are also derived; (iii) Based on these asymptotic theories, the ARE of our proposed LR estimator with respect to the LS estimator obtained in Xia et al. [5] is derived, which is shown to has an expression that is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. Both the theoretical conclusions and numerical results reveal that the proposed LR procedure is highly efficient across a wide class of non-normal error distributions and almost not loss efficiency for normal error.

The rest of this paper is organized as follows. In Section 2, we first state the motivation of the LS-based OPG estimator (lsOPG) proposed by Xia et al. [5] and then present the idea of our proposed LR-based OPG estimator (lrOPG) for the single-index parameter. In Section 3, we establish the asymptotic properties of the two estimators as well as the LR estimate of nonparametric function, and the ARE of LrOPG with respect to LsOPG is also analyzed in this section. Two simulation examples and a real data analysis are performed to reflect the performance of the estimators in Section 4.

Some conclusion remarks and extension are followed in Section 5. All technical proofs are collected in Section 1.

2. The estimation procedures

2.1. Estimation via LS-based OPG method

Let $G(x,z) = E\{Y \mid X = x, Z = z\}$, it is easy to verify that $G(X,Z) = g(X^T\beta_0) + Z^T\theta_0$ and $\nabla_X G(X,Z) = \partial G(X,Z)/\partial X = g'(X^T\beta_0)\beta_0$. Therefore, the fact that all such partial derivatives are parallel to the single-index direction β_0 allows someone to estimate β_0 via estimating the partial derivative $\nabla_X G(X,Z)$ or $E(\nabla_X G(X,Z))$ instead, which is called the average derivative estimation (ADE) method. Because a high-order kernel function is used, the ADE approach still suffers the issue of 'curse of dimensionality'. Another disadvantage of this procedure is that when $E\{g'(X^T\beta_0)\} = 0$ or equivalently $E\{g'(X^T\beta_0)\}\beta_0 = 0$, it fails to estimate the single-index parameter β_0. Thus, one can use the OPG, i.e., $E\{\nabla_X G(X,Z)\nabla_X^T G(X,Z)\}$ to overcome this shortcoming. The estimate of β_0 is the eigenvector corresponding to the largest eigenvalue of $E\{\nabla_X G(X,Z)\nabla_X^T G(X,Z)\}$ due to the fact that $E\{\nabla_X G(X,Z)\nabla_X^T G(X,Z)\} = E\{g'(X^T\beta_0)^2\}\beta_0\beta_0^T$ has only one nonzero eigenvalue.

To implement the estimation, Xia et al. [5] considered estimating $\nabla_X G(X,Z)$ through local polynomial smoothing technique with a refined lower-dimensional kernel function. In detail, suppose that $\{(X_i, Y_i, Z_i),\ i = 1,2,\ldots,n\}$ is an independent identically distributed (i.i.d.) random sample from model (1), for X_i in the neighbourhood of X_j, the unspecified link function $g(\cdot)$ can be locally linearly approximated as

$$g(X_i^T\beta_0) \approx g(X_j^T\beta_0) + g'(X_j^T\beta_0)(X_i^T\beta_0 - X_j^T\beta_0) = a_j + b_j^T X_{ij},$$

where $a_j = g(X_j^T\beta_0)$, $b_j = g'(X_j^T\beta_0)\beta_0$ and $X_{ij} = X_i - X_j$. Assume that β is an initial estimate of β_0, they considered a series of the following minimization problems

$$(\hat{a}_j^{ls}, \hat{b}_j^{ls}, \hat{\theta}^{ls}) = \arg\min_{a_j, b_j, \theta} \sum_{i=1}^{n} \{Y_i - a_j - b_j^T X_{ij} - Z_i^T\theta\}^2 K_h(X_{ij}^T\beta), \quad j = 1,2,\ldots,n, \tag{2}$$

where $a_j = g(X_j^T\beta)$, $b_j = g'(X_j^T\beta)\beta$, $K_h(\cdot) = K(\cdot/h)/h$ and $K(\cdot)$ is the refined only one-dimensional kernel weight function. Therefore, the estimate of $E\{\nabla_X G(X,Z)\nabla_X^T G(X,Z)\}$ is given by

$$\hat{\Sigma}_{ls} = \frac{1}{n}\sum_{j=1}^{n} \hat{b}_j^{ls}\hat{b}_j^{ls^T},$$

and the first eigenvector of $\hat{\Sigma}_{ls}$, denoted as $\hat{\beta}_{\text{lsOPG}}$, is the estimate of β_0.

2.2. Estimation via LR-based OPG method

In this subsection, we aim to propose an efficient estimator of β_0 through a rank-based robust method together with the ideal of OPG approach. Motivated by the rank method studied in nonparametric regression analysis of Wang et al. [18] and Feng et al. [19], we propose the following LR-based objective function with refined kernel weights, which is defined as

$$L_n(b_k, \theta) = \frac{1}{n(n-1)} \sum_{i \neq j} |e_{ik} - e_{jk}| K_h(X_{ik}^T\beta) K_h(X_{jk}^T\beta), \tag{3}$$

where $e_{ik} = Y_i - a_k - b_k^T X_{ik} - Z_i^T\theta$, a_k, b_k, X_{ik} and $K_h(\cdot)$ are similarly defined as in (2). It is worth noting that the above rank-based loss function cannot generate the estimate of a_k because it is cancelled out in $e_{ik} - e_{jk}$, which is a unique feature of using this type of estimator in the present problem since our main interest is only on the gradient b_k. Moreover, the loss function $L_n(b_k, \theta)$ essentially belongs to a local version of Ginis mean difference, which is a classical measure of concentration or dispersion, see [21] for details.

Define the minimizer of (3) by $(\hat{b}_k^{lr}, \hat{\theta}^{lr})$ for $k = 1, 2, \ldots, n$, that is

$$
\begin{aligned}
(\hat{b}_k^{lr}, \hat{\theta}^{lr}) &= \arg\min_{b_k,\theta} L_n(b_k, \theta) \\
&= \arg\min_{b_k,\theta} \frac{1}{n(n-1)} \sum_{i\neq j} |Y_i - Y_j - b_k^T X_{ij} - Z_{ij}^T \theta| K_h(X_{ik}^T \beta) K_h(X_{jk}^T \beta).
\end{aligned} \tag{4}
$$

Therefore, following the idea of OPG technique, the estimate of $E\{\nabla_X G(X,Z) \nabla_X^T G(X,Z)\}$ based on LR regression can be calculated by

$$
\hat{\Sigma}_{lr} = \frac{1}{n} \sum_{k=1}^{n} \hat{b}_k^{lr} \hat{b}_k^{lr^T},
$$

and the eigenvector corresponding to the only nonzero eigenvalue of $\hat{\Sigma}_{lr}$ is the estimate of β_0, which is denoted by $\hat{\beta}_{\mathrm{lrOPG}}$.

Note that, as stated in [18], although the calculation of $\hat{b}_k^{lr}$ is more complicated compared with that of $\hat{b}_j^{ls}$ since (4) has no close-form solution, the proposed LR based estimator of $(b_k^T, \theta^T)^T$ can be solved by fitting a weighted L_1 regression on $n(n-1)/2$ pseudo observations $(\tilde{X}_i - \tilde{X}_j, Y_i - Y_j)$ with weights $w_{ij} = K_h(X_{ik}^T \beta) K_h(X_{jk}^T \beta)$, where $\tilde{X}_i = (X_i^T, Z_i^T)^T$ for $1 \leq i, j \leq n$. This algorithm is shown to be quite reliable and efficient in practice, and many statistical software packages can be selected to carry out weighted L_1 regression. In our simulation studies, we use the function *rq* in the R package *quantreg*.

Remark 2.1: Because both the weights in (2) and (4) involve the initial estimate β, following Xia and Härdle [4] and Yang and Yang [15], a repeated procedure can be considered to deal with this issue. That means we can update the weights with the latest estimator of β_0 until convergence. On the other hand, we assume in the following context that the initial estimate β is $\sqrt{n}$-consistent, which is feasible by using existing methods such as [4,14].

3. Large sample properties

In this part, we first establish the asymptotic properties of the two estimators discussed in Section 2 as well as the LR estimate of nonparametric function. Furthermore, based on the asymptotical distributions, the asymptotic relative efficiency is also analyzed.

3.1. Asymptotic properties

In order to establish the theoretical results of the resulting estimators, the following regularity conditions are required which may be weakened.

(C1$^\#$) ε is the random error with $E(\varepsilon \mid X, Z) = 0$.

(C1) The errors ε has a positive density function $h(x)$ satisfying $\int [h'(x)]^2/h(x)\,\mathrm{d}x < \infty$, which means that ε has finite Fisher information.

(C2) The density function $f_\beta(\cdot)$ of $X^T\beta$ is positive and satisfies Lipschitz condition of order one for β in a neighbourhood of β_0. Further $X^T\beta_0$ has a positive and bounded density function on its support. In addition, $E(|X|^4) < \infty$.

(C3) The kernel function $K(\cdot) \geq 0$ is a symmetric density function with bounded first-order derivative and compact support.
(C4) The link function $g(\cdot)$ has bounded and continuous second-order derivatives.
(C5) With probability 1, the variable X lies in a compact set $\mathcal{D}$. The conditional expectations $E(V \mid X^T\beta_0 = x^T\beta_0)$ and $E(VV^T \mid X^T\beta_0 = x^T\beta_0)$ have bounded derivatives with $V = (X^T, Z^T)^T$. Moreover, the matrix $\Omega = E(\tilde{Z}^{\otimes 2})$ is positive definite, where $\tilde{Z} = Z - E(Z \mid X^T\beta_0)$ and $M^{\otimes 2} = MM^T$ for any matrix or vector M.

These conditions above are commonly used and easily satisfied in many applications. (C1$^{\#}$) is only required in LS regression for establishing the consistency of $\hat{\beta}_{\text{lsOPG}}$, similar condition can also be seen in Liang et al. [8] and Wang et al. [7]. (C1) is a regular condition on the random errors which is the same as that used in rank regression such as [18–20,22]. (C2)–(C4) are standard conditions for nonparametric smoothing as that in Xia and Härdle [4] and Liang et al. [8]. In particular, (C2) guarantees the existence of any ratio terms with density appearing as part of the denominator. The moment requirement on X in (C2) is to make the proof simpler since the existence of finite moments is sufficient, see Härdle et al. [23] and Xia and Härdle [4]. Condition (C5) is a standard assumption used for deriving the asymptotic distribution of the resulted estimates, see Xia et al. [5] and Xia and Härdle [4].

Before presenting the main theoretical results, we first give some notations for convenience of expression. Let $u_\beta(x) = E(X \mid X^T\beta = x^T\beta)$, $v_\beta(x) = x - u_\beta(x)$, $\omega_\beta(x) = E(XX^T \mid X^T\beta = x^T\beta)$, $W_0(x) = v_\beta(x)v_\beta^T(x)$ and $W(x) = \omega_\beta(x) - u_\beta(x)u_\beta^T(x)$.

Theorem 3.1: *Suppose that conditions* (*C1$^{\#}$*) *and* (C2)–(C5) *hold. Let $\hat{\beta}_{\text{lsOPG}}$ be the estimator produced in Section* 2.1. *If $h \sim n^{-\lambda}$ with $1/6 < \lambda < 1/4$, then*

$$\sqrt{n}(\hat{\beta}_{\text{lsOPG}} - \beta_0) \xrightarrow{d} N\left(0, \Sigma_{\text{lsOPG}}\right), \tag{5}$$

where $\Sigma_{\text{lsOPG}} = E\{g'(X^T\beta_0)^2[W(X) - \Xi]^+[W_0(X) - \Xi][W(X) - \Xi]^+\varepsilon^2\}/E\{g'(X^T\beta_0)^2\}^2$, $\Xi = \Gamma\Omega^{-1}\Gamma^T$, $\Gamma = E\{v_\beta(Z)\tilde{Z}^T\}$, *$\Omega$ is defined in Condition* (C5), *and '+' denotes the Moore-Penrose inverse.*

Theorem 3.2: *Suppose that conditions* (C1)–(C5) *hold. Let $\hat{\beta}_{\text{lrOPG}}$ be the estimator generated in Section* 2.2. *If $h \sim n^{-\lambda}$ with $1/6 < \lambda < 1/4$, then*

$$\sqrt{n}(\hat{\beta}_{\text{lrOPG}} - \beta_0) \xrightarrow{d} N\left(0, \Sigma_{\text{lrOPG}}\right), \tag{6}$$

where $\Sigma_{\text{lrOPG}} = (1/4\tau^2)E\{g'(X^T\beta_0)^2[W(X) - \Xi]^+[W_0(X) - \Xi][W(X) - \Xi]^+[2H(\varepsilon) - 1]^2\}/E\{g'(X^T\beta_0)^2\}^2$, *$\Xi$ is same defined as in Theorem* 3.1, $\tau = \int h(x)^2\,\mathrm{d}x$ *and $H(\cdot)$ denotes the cumulative distribution function* (*CDF*) *of ε.*

As we can see from Theorems 3.1 and 3.2, both the estimators $\hat{\beta}_{\text{lsOPG}}$ and $\hat{\beta}_{\text{lrOPG}}$ eventually provide a $\sqrt{n}$-consistent estimate of the single-index parameter β_0 but with different asymptotic covariances. Therefore, there is a necessary to compare the efficiencies of the two different types of estimators, which is the main context in the next subsection.

Note that model (1) can be transformed into a partially linear model after the LR estimate of β_0 is derived. As pointed out previously that the rank-based loss function cannot generate the estimate of a_k because it is cancelled out in $e_{ik} - e_{jk}$. Therefore, according to Shang et al. [24], here we estimate θ_0 and $g(\cdot)$ by applying a three-stage estimation procedure based on local Walsh-average regression, which has been shown to own the same efficiency to the local rank regression. For the convenience of expression, we also denote the estimates of θ_0 and $g(\cdot)$ by $\hat{\theta}_{lr}$ and $\hat{g}_{lr}(\cdot)$, respectively. The asymptotic distributions of $\hat{\theta}_{lr}$ and $\hat{g}_{lr}(\cdot)$ are summarized in the following proposition.

Table 1. Asymptotic relative efficiency for various error distributions.

Symmetric errors				Asymmetric errors	
Error	ARE	Error	ARE	Error	ARE
Normal	0.955	Logistic	1.097	Lognormal	7.354
Laplace (DE)	1.500	Cauchy	∞	Exp(1)	3.000
$t(3)$	1.900	$T(0.1, 3)$	1.373	$\chi^2(3)$	1.901
$t(4)$	1.402	$T(0.05, 10)$	4.768	$\chi^2(4)$	1.500
$t(5)$	1.240	$T(0.1, 10)$	7.188	$F(4, 6)$	7.837

Proposition 3.3: *Suppose that same conditions given in Theorem* 3.2 *hold, then we have*

$$\sqrt{n}(\hat{\theta}_{lr} - \theta_0) \xrightarrow{d} N\left(0, \Omega^{-1}/12\tau^2\right),$$
$$\sqrt{nh}(\hat{g}_{lr}(u) - g(u) - \tfrac{1}{2}\mu_2 h^2 g''(u)) \xrightarrow{d} N\left(0, \nu_0 f_{\beta_0}^{-1}(u)/12\tau^2\right),$$

where $\mu_j = \int u^j K(u)\,\mathrm{d}u$ *and* $\nu_j = \int u^j K^2(u)\,\mathrm{d}u$ *for* $j = 0,1,2$.

3.2. Asymptotic relative efficiency

To measure the efficiency, we consider the asymptotic mean squared error (MSE) of the estimators $\hat{\beta}_{\mathrm{lsOPG}}$ and $\hat{\beta}_{\mathrm{lrOPG}}$ since they are all asymptotic unbiased. Recall that the asymptotic distributions of $\hat{\beta}_{\mathrm{lsOPG}}$ and $\hat{\beta}_{\mathrm{lrOPG}}$ presented in (5) and (6) by Theorems 3.1 and 3.2, we immediately have the following theorem holds.

Theorem 3.4: *Assume that* ε *is independent of* (X, Z), *then the ARE of LR based estimator* $\hat{\beta}_{\mathrm{lrOPG}}$ *to LS based estimator* $\hat{\beta}_{\mathrm{lsOPG}}$ *for the single-index parameter* β_0 *is*

$$\mathrm{ARE}(\hat{\beta}_{\mathrm{lrOPG}}, \hat{\beta}_{\mathrm{lsOPG}}) = \frac{\mathrm{MSE}(\hat{\beta}_{\mathrm{lsOPG}})}{\mathrm{MSE}(\hat{\beta}_{\mathrm{lrOPG}})} = \frac{\mathrm{Var}(\hat{\beta}_{\mathrm{lsOPG}})}{\mathrm{Var}(\hat{\beta}_{\mathrm{lrOPG}})} = 12\sigma^2\tau^2,$$

where $\sigma^2 = E(\varepsilon^2)$. *This ARE has a lower bound 0.864 for estimating the parameter component, which is attained at the random error density* $h(x) = \frac{3}{20\sqrt{5}}(5 - x^2)I(|x| \le 5)$.

It is worth noting that the above obtained ARE is essentially related to the ARE of the signed-rank Wilcoxon test with respect to the t-test. The AREs of the LR based estimator versus the LS based estimator are reported in Table 1 for some commonly used error distributions, including both the symmetric and asymmetric errors, in which '$T(\rho, \sigma)$' denotes the Tukey contaminated normal with CDF $F(x) = (1 - \rho)\phi(x) + \rho\phi(x/\sigma)$, $\phi(x)$ is the CDF of a standard normal distribution and $\rho \in [0, 1]$ stands for the contamination proportion. Moreover, From Table 1, we can clearly see that our proposed LR estimator is highly efficient than that of LS except for normal error, but the loss of efficiency in this case is slight.

Remark 3.1: According to the similar approach done in Shang et al. [24], it is not difficult to obtain that $\mathrm{ARE}(\hat{\theta}_{lr}, \hat{\theta}_{ls}) = 12\sigma^2\tau^2$ with a lower bound equalling to 0.864 and $\mathrm{ARE}(\hat{g}_{lr}(u), \hat{g}_{ls}(u)) = (12\sigma^2\tau^2)^{4/5}$ with a lower bound being 0.8896, respectively.

4. Simulation studies

4.1. Monte Carlo simulations

In this section, two numerical examples are conducted to compare the finite sample performance of our proposed estimator (lrOPG) with those obtained by Xia et al. [5] (lsOPG) and Fan and Zhu

[14] (qOPG), where the quantile is choosen as 0.5 in [14]. In our simulations, the Gaussian kernel is selected as the kernel function, and in order to evaluate the performance of the resulted estimators, we consider the following criteria: (1) average absolute deviation (AAD) of the estimated coefficients along with its corresponding standard deviation (SD). (2) Estimation error (EE) which is defined as $\mathrm{EE}(\hat{\beta}) = \sqrt{1 - |\hat{\beta}^T \beta_0|}$ in [14] for any estimator $\hat{\beta}$. The closer the value of $\mathrm{EE}(\hat{\beta})$ is to 0, the better the estimator $\hat{\beta}$ is. (3) Mean ratio of the absolute deviation (MRAD) that has the definition MRAD = $\mathrm{E}\{\mathrm{AD}(\hat{\beta}_{\mathrm{lsOPG}})/\mathrm{AD}(\hat{\beta}_{\mathrm{lrOPG}})\}$, where $\mathrm{AD}(\hat{\beta}) = \sum_{j=1}^{p} |\hat{\beta}_j - \beta_{0j}|$ for any estimator $\hat{\beta}$ of β_0.

To assess the performance of estimator $\hat{g}$ for the conditional link function g, we apply the square root of average square errors (RASE) of $\hat{g}$, which is defined as

$$\mathrm{RASE}(\hat{g}) = \left\{ \frac{1}{n_{\mathrm{grid}}} \sum_{i=1}^{n_{\mathrm{grid}}} \|\hat{g}(u_i) - g(u_i)\|^2 \right\}^{1/2},$$

where $\{u_i, i = 1, 2, \ldots, n_{\mathrm{grid}}\}$ are the grid points at which the function $g(\cdot)$ is evaluated.

Example 4.1: In this example, we consider the PLSIM with the following form

$$Y = g(X^T \beta_0) + Z^T \theta_0 + 0.2\varepsilon, \tag{7}$$

where $X = (X_1, X_2, X_3)$ are generated from standard normal distribution and the correlation between X_i and X_j to be $0.5^{|i-j|}$, $\beta_0 = (1, 2, 2)^T/3$. $Z = (Z_1, Z_2, Z_3, Z_4)^T$ are independent random vectors uniformly distributed on $[-1, 1]^4$, $\theta_0 = (3, 1.5, 0, 2)^T$ and $g(u) = u^2 - 0.5\exp(u - 1)$. In order to show the robustness of our proposed estimator lrOPG, the following five various symmetric error distributions are considered: standard normal $N(0, 1)$, $t(3)$, logistic, double exponential distribution with median 0 and scale parameter 1 (DE) and the mixed normal $T(0.1, 10)$.

In our simulation experiments, 200 repetitions are carried out with the sample size $n = 200$ for every error structure, and the corresponding results are reported in Table 2. As we can see from the observations, the proposed lrOPG estimator performs slightly worse than the lsOPG estimator when the error comes from standard normal distribution. Whereas, for the other symmetric non-normal errors, the performance of LS based estimator is seriously affected by the outliers or heavy-tailed data, and our LR-based estimator has a significantly improvement in terms of all above mentioned evaluation criterion as expected, i.e., the AAD, EE and RASE of lrOPG are obviously smaller than those of lsOPG. Similar phenomenon can also be reflected by the MRAD that its values are all great than 1 except for the normally distributed error. In addition, although both qOPG and lrOPG are robust to the error distributions, the proposed LR-based estimators gain a higher efficiency than that of quantile-based estimators. In a word, all these conclusions indicate that the variability of ε has little effect on the efficiency of the proposed lrOPG estimator, which corroborates our theoretical findings that LR-based approach is robust to the data subjected to outliers in the response or heavy-tailed error distributions, compared with LS-based method.

Example 4.2: In this example, we focus on the asymmetric error distributions to show the robustness of the proposed lrOPG estimator. The model set-up is same to (7) and the following four different asymmetric errors are generated: exponential distribution (Exp(1)), lognormal, chi-square distribution with its degree equalling to 3 ($\chi^2(3)$) and F distribution with its degree equalling to (4,6) ($F(4, 6)$). we also generate 200 random samples, each consisting of 200 observations, and the corresponding results are summarized in Table 3. Similar to the conclusions of Example 4.1, this table shows that our proposed LR-based method is highly efficient and the MEAD of lrOPG with respect to lsOPG are great than 1 for all the asymmetric errors under consideration. Therefore, based on the results of Examples 4.1 and 4.2, we can conclude that the new proposed estimator lrOPG does have a great

Table 2. Estimation accuracy comparisons in Example 4.1 and their standard errors in brackets.

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$				
Dist.	Method	AAD (SD)	AAD (SD)	AAD (SD)	EE($\hat{\beta}$) (SD)	MRAD (SD)	EE($\hat{\theta}$) (SD)	RASE (SD)
$N(0,1)$	lsOPG	0.0019 (0.0352)	0.0028 (0.0416)	0.0023 (0.0391)	0.0694 (0.0389)	–	0.0708 (0.0415)	0.1042 (0.0507)
	qOPG	0.0037 (0.0438)	0.0046 (0.0469)	0.0049 (0.0483)	0.0962 (0.0461)	0.7816 (0.0458)	0.0821 (0.0527)	0.1205 (0.0614)
	lrOPG	0.0023 (0.0396)	0.0030 (0.0437)	0.0024 (0.0409)	0.0719 (0.0403)	0.9671 (0.0414)	0.0764 (0.0458)	0.1129 (0.0538)
$t(3)$	lsOPG	0.0103 (0.1027)	0.0097 (0.1008)	0.0114 (0.1085)	0.2436 (0.1058)	–	0.1846 (0.1039)	0.2114 (0.1153)
	qOPG	0.0071 (0.0496)	0.0077 (0.0521)	0.0082 (0.0553)	0.1347 (0.0532)	1.3450 (0.0549)	0.1249 (0.0694)	0.1426 (0.0729)
	lrOPG	0.0034 (0.0419)	0.0041 (0.0470)	0.0039 (0.0458)	0.0961 (0.0485)	1.4068 (0.0472)	0.1089 (0.0582)	0.1275 (0.0634)
Logistic	lsOPG	0.0091 (0.0436)	0.0087 (0.0502)	0.0094 (0.0523)	0.1175 (0.0514)	–	0.1041 (0.0558)	0.1265 (0.0594)
	qOPG	0.0078 (0.0489)	0.0084 (0.0517)	0.0097 (0.0550)	0.1238 (0.0531)	0.9157 (0.0516)	0.1097 (0.0591)	0.1283 (0.0612)
	lrOPG	0.0038 (0.0392)	0.0031 (0.0406)	0.0035 (0.0412)	0.0874 (0.0423)	1.0315 (0.0413)	0.0853 (0.0497)	0.1183 (0.0575)
DE	lsOPG	0.0089 (0.0624)	0.0102 (0.0714)	0.0093 (0.0679)	0.1458 (0.0694)	–	0.1432 (0.0794)	0.1559 (0.0873)
	qOPG	0.0037 (0.0392)	0.0044 (0.0407)	0.0039 (0.0390)	0.0941 (0.0387)	1.3914 (0.0409)	0.1168 (0.0610)	0.1319 (0.0626)
	lrOPG	0.0034 (0.0419)	0.0042 (0.0438)	0.0037 (0.0405)	0.0934 (0.0428)	1.2143 (0.0435)	0.0895 (0.0522)	0.1217 (0.0591)
$T(0.1,10)$	lsOPG	0.0165 (0.1532)	0.0176 (0.1598)	0.0169 (0.1514)	0.2889 (0.1634)	–	0.2638 (0.1457)	0.3011 (0.1652)
	qOPG	0.0083 (0.0594)	0.0079 (0.0632)	0.0086 (0.0613)	0.1531 (0.0628)	1.8791 (0.0637)	0.1340 (0.0783)	0.1648 (0.0816)
	lrOPG	0.0049 (0.0487)	0.0047 (0.0503)	0.0058 (0.0539)	0.1004 (0.0531)	2.6318 (0.0524)	0.1125 (0.0629)	0.1372 (0.0693)

Table 3. Estimation accuracy comparisons in Example 4.2 and their standard errors in brackets.

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$				
Dist.	Method	AAD (SD)	AAD (SD)	AAD (SD)	EE($\hat{\beta}$) (SD)	MRAD (SD)	EE($\hat{\theta}$) (SD)	RASE (SD)
Exp(1)	lsOPG	0.0092 (0.0630)	0.0107 (0.0675)	0.0114 (0.0718)	0.1534 (0.0683)	–	0.1705 (0.0849)	0.2163 (0.1058)
	qOPG	0.0057 (0.0462)	0.0061 (0.0494)	0.0064 (0.0501)	0.1078 (0.0491)	1.2408 (0.0513)	0.1292 (0.0684)	0.1376 (0.0743)
	lrOPG	0.0046 (0.0441)	0.0043 (0.0428)	0.0049 (0.0475)	0.0983 (0.0468)	1.5372 (0.0474)	0.1134 (0.0527)	0.1198 (0.0639)
Lognormal	lsOPG	0.0176 (0.1638)	0.0185 (0.1701)	0.0181 (0.1716)	0.3019 (0.1832)	–	0.2751 (0.1860)	0.3142 (0.1994)
	qOPG	0.0092 (0.0674)	0.0106 (0.0725)	0.0094 (0.0693)	0.1347 (0.0718)	1.6893 (0.0724)	0.1428 (0.0759)	0.1537 (0.0873)
	lrOPG	0.0058 (0.0562)	0.0054 (0.0543)	0.0063 (0.0581)	0.1136 (0.0574)	2.3715 (0.0578)	0.1214 (0.0653)	0.1283 (0.0716)
$\chi^2(3)$	lsOPG	0.0143 (0.0716)	0.0157 (0.0735)	0.0149 (0.0720)	0.1592 (0.0743)	–	0.1738 (0.0864)	0.2075 (0.0983)
	qOPG	0.0059 (0.0534)	0.0056 (0.0539)	0.0065 (0.0587)	0.1194 (0.0558)	1.1350 (0.0562)	0.1270 (0.0649)	0.1318 (0.0724)
	lrOPG	0.0052 (0.0487)	0.0061 (0.0548)	0.0057 (0.0512)	0.1027 (0.0531)	1.2934 (0.0529)	0.1153 (0.0552)	0.1204 (0.0681)
$F(4,6)$	lsOPG	0.0189 (0.1762)	0.0173 (0.1691)	0.0206 (0.1758)	0.3145 (0.1867)	–	0.2916 (0.1937)	0.3279 (0.2018)
	qOPG	0.0098 (0.0714)	0.0104 (0.0756)	0.0095 (0.0703)	0.1538 (0.0744)	1.9813 (0.0756)	0.1605 (0.0824)	0.1782 (0.0917)
	lrOPG	0.0063 (0.0592)	0.0059 (0.0587)	0.0068 (0.0624)	0.1306 (0.0602)	2.5036 (0.0615)	0.1278 (0.0681)	0.1427 (0.0734)

Table 4. Estimations of the coefficients in air pollution data.

Method	Wind	SD	Temp	SD	Solar.R	SD
lsOPG	−0.8405	0.1099	0.5390	0.1304	0.0274	0.0051
qOPG	−0.7916	0.0652	0.6154	0.0627	0.0192	0.0036
lrOPG	−0.8103	0.0586	0.5892	0.0512	0.0209	0.0027

efficiency gain across a wide spectrum of non-normal error distributions and almost not lose any efficiency for the normal error.

4.2. A real data application

We consider an application of the proposed LR regression method to the air pollution data from an environmental study which had also been analyzed by Yu and Ruppert [2] for PLSIM. The data are daily measurements of three predictor variables and one response variable from $n = 111$ days from May to September, 1973, in the New York metropolitan area, of which the variable of interest is the mean ozone Y , and the three covariate variables are solar radiation (Solar.R), average wind speed (Wind) and maximum daily temperature (Temp), respectively. The purpose is to research how the concentration of the air pollution ozone depends on the three meteorological elements. Beforehand, both the response variable Y and the covariates are standardized so that they have mean zero and unit variance. Hence, as recommended in [2], we consider the average wind speed and maximum daily temperature as the single-index variables and the solar radiation as the linear variable. Then we apply the LS-based approach of [5], quantile-based method of [14] and our proposed LR-based procedure to analyze the data set by a partially linear single-index regression model stated as (1).

The average estimated coefficients of the single-index parameter with their corresponding standard deviations are summarized in Table 4. Although there is a relatively small discrepancies in terms of estimated coefficients, the LS-based method has a remarkably larger standard deviations than that of LR approach. Moreover, from the boxplots of the estimated coefficients displayed in Figure 1(a–c), we can also clearly see that the estimator lrOPG tends to generate significantly fewer extreme values and smaller variance than lsOPG, that is to say, LR-based approach has a much smaller probability of variability to the data compared with LS based method. In addition, the estimations for the nonparametric link function $g(\cdot)$ are presented in Figure 1(d–f) along with its 95% pointwise confidence intervals. As expected, the LR regression method provides narrower confidence intervals than those of LS and quantile methods. All these observations indicates that our proposed rank-based procedure is the most effective among the considered methodologies. Consequently, in practice, as we do not known the data structure in advance, LR regression method should be a reasonable alternative to existing works by taking its efficiency and robustness into account.

5. Conclusions and extension

In this paper, we focus our attention on the estimation for the single-index parameter in PLSIM. A new robust estimation procedure combining the ideas of OPG and local rank-based inference is developed along with its theoretical property, both the theoretical and numerical results demonstrate that our proposed estimator has a great efficiency gain across a wide spectrum of non-normal error distributions and almost not lose any efficiency for the normal error. Moreover, we have established the asymptotic property of the estimator based on the combination of OPG and LS method, which was proposed by Xia et al. [5] without giving its theoretical property. Based on the asymptotic results, we show that the ARE of the proposed LR-based procedure with respect to LS-based method has an expression which is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test, and the estimator lrOPG has a lower bound equal to 0.864 for estimating the single-index parameter with respect to the estimator lsOPG even in the worst case scenarios. Note that, here we

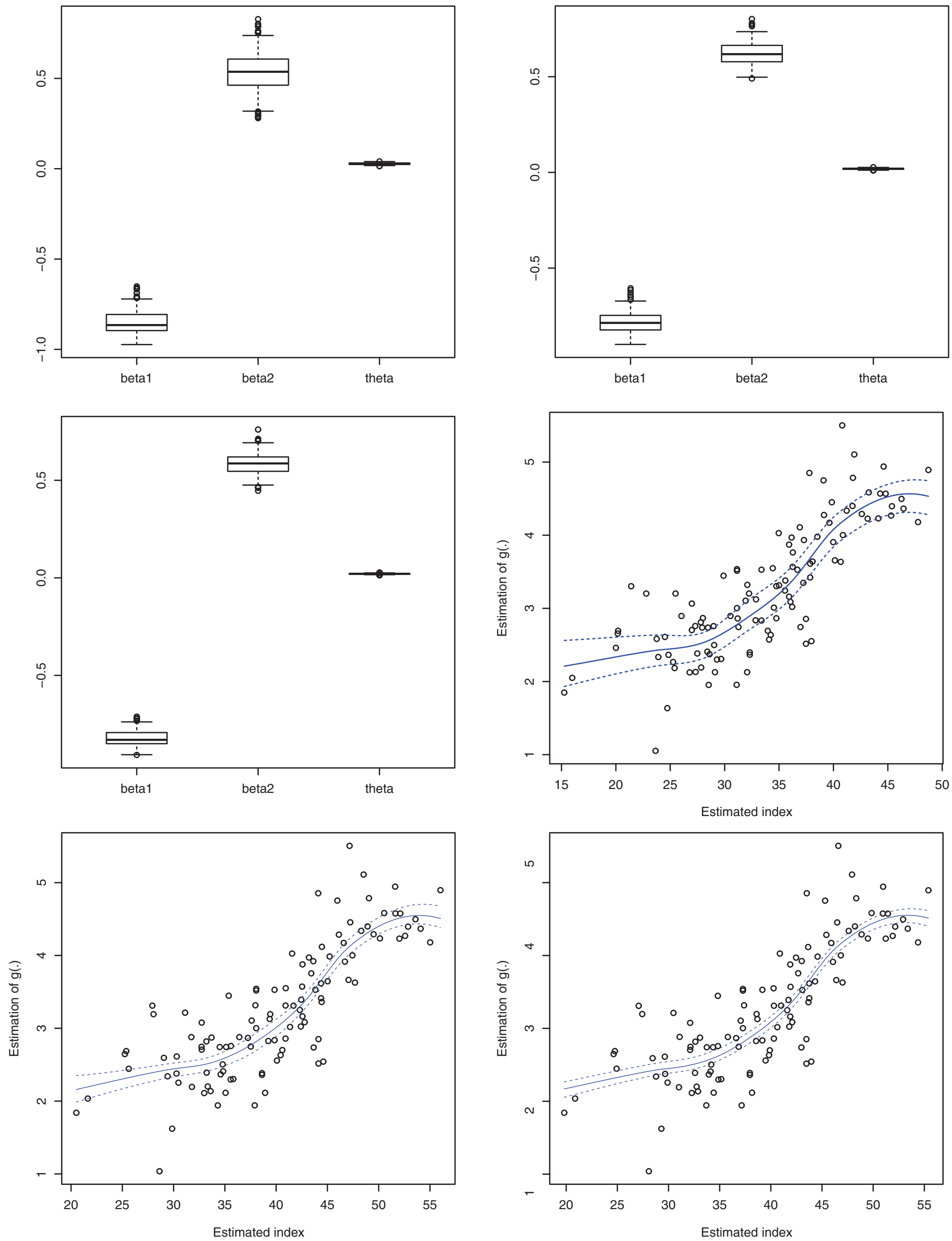

Figure 1. (a)–(c) are the boxplots of the estimated coefficients for air pollution data. (d)–(f) are the estimated coefficient functions (solid curves) against the estimated index, along with the associated 95% pointwise confidence intervals (dotted curves). (a) lsOPG, (b) qOPG, (c) lrOPG, (d) lsOPG, (e) qOPG and (f) lrOPG.

mainly consider the issue of estimation in the PLSIM, some irrelevant variables are often met in practice, then variable selection based on LR regression should be an interesting problem to investigate in the future. Furthermore, we may perform some hypothesis tests for the parametric or nonparametric components for purpose of dimension reduction and exploratory data.

Table 5. Average $m^2(\hat{\beta}_k, B_0)$ for model (8) based on LS and LR methods.

Dist.	Method	$m^2(\hat{\beta}_1, B_0)$	$m^2(\hat{\beta}_2, B_0)$	$m^2(\hat{\beta}_3, B_0)$
$N(0,1)$	lsOPG	0.0164	0.1026	0.1375
	lrOPG	0.0183	0.1139	0.1468
$t(3)$	lsOPG	0.0872	0.3451	0.4326
	lrOPG	0.0239	0.1274	0.1670
MN	lsOPG	0.1247	0.3712	0.4639
	lrOPG	0.0225	0.1348	0.1574
Exp(1)	lsOPG	0.1039	0.3647	0.4395
	lrOPG	0.0244	0.1292	0.1226

For a further study of the applicability of the proposed method, we extend the LR-based OPG approach from one to several directions. Consider the following multiple-index model

$$Y_i = (X_i^T \beta_1)^2 - (0.5 + X_i^T \beta_2)^2 + 15\cos(X_i^T \beta_3) + 0.5\varepsilon, \tag{8}$$

where $\beta_1 = (1,2,3,0,0)^T/\sqrt{14}$, $\beta_2 = (-2,1,0,1,0)^T/\sqrt{6}$ and $\beta_3 = (0,0,1,1,1)^T/\sqrt{3}$, the independent variable X_i has a standard normal distribution in R^5. Obviously, the effective dimension reduction space given by $B_0 = (\beta_1, \beta_2, \beta_3)$ has thus dimension $D = 3$. Note that, the squared distance function $m^2(\hat{B}, B_0) = \|(I - B_0 B_0^T)\hat{B}\|^2$ for $d < D = 3$ and $m^2(\hat{B}, B_0) = \|(I - \hat{B}\hat{B}^T)B_0\|^2$ for $d \geq D = 3$, which was defined in Xia et al. [5] with $D = 3$ is the true dimension of the reduced space used in our simulations and d denotes the dimension used for estimation, is applied to measure the estimation accuracy. Further to examine whether the proposed lrOPG estimators own robustness or not, we compare the results with that of lsOPG estimators proposed in Xia et al. [5] under N(0,1), t(3), the mixed normal T(0.1,10) and Exp(1) errors. The corresponding results with 200 repetitions and 200 samples are collected in Table 5. The observations reveal that the proposed lrOPG is likely to be a robust and efficient estimate versus lsOPG for vary errors, whereas it is hard to specify the asymptotic distribution, especially for the cases of $D > 3$ as stated in Wang et al. [25]. Furthermore, we may have no information about d beforehand, some suitable criteria are also needed to estimate it. Such studies require much more research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Jing Yang's research was supported by the National Natural Science Foundation of China (Grant 11801168) and the Natural Science Foundation of Hunan Province (Grant 2018JJ3322). Fang Lu's research was supported by the National Natural Science Foundation of China (Grants 11801169, 11571055). Hu Yang's research was supported by the National Natural Science Foundation of China (Grant 11671059).

References

[1] Carroll RJ, Fan J, Gijbels I, Wand MP. Generalized partially linear single-index models . J Amer Statist Assoc. 1997;92:477–489.
[2] Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. J Amer Statist Assoc. 2002;97:1042–1054.
[3] Xia Y, Li WK. On single-index coefficient regression models. J Amer Statist Assoc. 1999;94:1275–1285.
[4] Xia Y, Härdle W. Semi-parametric estimation of partially linear single-index models. J Multivariate Anal. 2006;97:1162–1184.
[5] Xia Y, Tong H, Li WK, Zhu L. An adaptive estimation of dimension reduction space. J Roy Statist Soc Ser B. 2002;64:363–410.
[6] Zhu L, Xue L. Empirical likelihood confidence regions in a partially linear single-index model. J Roy Statist Soc Ser B. 2006;68:549–570.

[7] Wang J, Xue L, Zhu L, Chong Y. Estimation for a partial-linear single-index model. Ann Statist. 2010;38:246–274.
[8] Liang H, Liu X, Li R, Tsai CL. Estimation and testing for partially linear single-index models. Ann Statist. 2010;38:3811–3836.
[9] Koenker R, Bassett G. Regression quantiles. Econometrica. 1978;46:33–50.
[10] Honda T. Quantile regression in varying coefficient models. J Stat Plan Inference. 2004;121:113–125.
[11] Wu TZ, Yu K, Yan Y. Single-index quantile regression. J Multivariate Anal. 2010;101:1607–1621.
[12] Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann Statist. 2011;39:305–332.
[13] Koenker R. Quantile regression. Cambridge: Cambridge University Press; 2005.
[14] Fan Y, Zhu L. Estimation of general semi-parametric quantile regression. J Stat Plan Inference. 2012;143:896–910.
[15] Yang H, Yang J. The adaptive L1-penalized LAD regression for partially linear single-index models. J Stat Plan Inference. 2014;151:73–89.
[16] Lv Y, Zhang R, Zhao W, Liu J. Quantile regression and variable selection of partial linear single-index model. Ann Inst Stat Math. 2015;67:375–409.
[17] Leng C. Variable selection and coefficient estimation via regularized rank regression. Stat Sinica. 2010;20:167–181.
[18] Wang L, Kai B, Li R. Local rank inference for varying coefficient models. J Amer Statist Assoc. 2009;488:1631–1645.
[19] Feng L, Zou C, Wang Z. Rank-based inference for the single-index model. Statist Probab Lett. 2012;82:535–541.
[20] Sun J, Lin L. Local rank estimation and related test for varying-coefficient partially linear models. J Nonparametr Statist. 2014;26:187–206.
[21] David HA. Early sample measures of variability. Stat Sci. 1998;13:368–377.
[22] Hettmansperger TP, McKean JW. Robust nonparametric statistical methods. 2nd ed.. New York: Chapman-Hall; 2011.
[23] Härdle W, Hall P, Ichimura H. Optimal smoothing in single-index models. Ann Statist. 1993;21:157–178.
[24] Shang S, Zou C, Wang Z. Local Walsh-average regression for semiparametric varying-coefficient models. Statist Probab Lett. 2012;82:1815–1822.
[25] Wang T, Xu P, Zhu L. Penalized minimum average variance estimation. Stat Sinica. 2013;23:543–569.
[26] Xia Y. Asymptotic distributions for two estimators of the single-index model. Econ Theory. 2006;22:1112–1137.
[27] Pollard D. Asymptotics for least absolute deviation regression estimators. Econ Theory. 1991;7:186–199.
[28] Bai ZD, Miao BQ, Rao CR. Estimation of directions of arrival of signals: asymptotic results. Adv Spectrum Anal Array Process. 1991;1:327–347.
[29] Hodges JL, Lehmann EL. The efficiency of some nonparametric competitors of the t-test. Ann Math Statist. 1956;27:324–335.

Appendix. Proofs of main results

In order to prove the theoretical results, we first give the following notations for convenience of expression. Define

$$\delta_n = \log^{1/2}(n)/\sqrt{nh}, \quad \delta_\beta = \beta_0 - \beta, \quad \gamma_n = (nh)^{-1/2},$$
$$X_{ix} = X_i - x, \quad K_{h,ij} = K_h(X_{ij}^T\beta), \quad K_{ij} = K(X_{ij}^T\beta/h), \quad d_i = (Z_i^T, X_i^T, \varepsilon_i)^T,$$
$$c_i = \gamma_n^{-1}\begin{pmatrix} \theta_0 - \theta \\ h(g'(X_i^T\beta_0)\beta_0 - b_i) \end{pmatrix}, \quad W_{i,l} = \begin{pmatrix} Z_i \\ X_{il}/h \end{pmatrix},$$
$$\Delta_{ij} = g(X_i^T\beta_0) - g(X_j^T\beta_0) - g'(X_i^T\beta_0)X_{ij}^T\beta_0, \quad W_{ij} = W_{i,l} - W_{j,l},$$
$$\xi_\beta(x) = E(XZ^T \mid X^T\beta = x^T\beta) - xE(Z \mid X^T\beta = x^T\beta)^T,$$
$$\omega_\beta(x) = E(XX^T \mid X^T\beta = x^T\beta) \quad \text{and} \quad \tilde{\omega}_\beta(x) = \omega_\beta(x) - xu_\beta(x)^T - u_\beta(x)^T x + xx^T.$$

The following two lemmas are frequently used in the sequel.

Lemma A.1: *Suppose that $E(Z \mid X^T\beta = x^T\beta) = m(x^T\beta)$ and its derivatives up to second order are bounded supported and $E(|Z|^r)$ exists for some $r > 3$. Let (X_i, Z_i) be an i.i.d. sample from (X, Z), under the Conditions (C2)–(C3), we have*

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_{ix}^T\beta)(X_{ix}^T\beta/h)^r Z_i = f(x^T\beta)m(x^T\beta)\mu_r + \{f(x^T\beta)m(x^T\beta)\}'\mu_{r+1}h + O_p(h^2 + \delta_n),$$

where $\mu_r = \int K(u)u^r\,\mathrm{d}u$.

Lemma A.2: *Suppose that $m_n(\chi, Z)$, $n = 1, 2, \ldots$, are measurable functions of Z with index $\chi \in R^d$, where d is any integer number, such that*

(i) $|m_n(\chi, Z)| \le a_n M(Z)$ *with* $E\{M(Z)^\kappa\} < \infty$ *for some* $\kappa > 2$ *and* a_n *increases with* n *such that an* $a_n < c_0 n^{1-2/\kappa}$;
(ii) $E\{m_n(\chi, Z)^2\} \le a_n m_0(\chi)^2$ *with* $|m_0(\chi) - m_0(\chi')| \le c|\chi - \chi'|^{\alpha_1}$, *where* $\alpha_1 > 0$ *and* $c > 0$ *are two constants (without loss of generality, we assume* $m_0(\chi') \ge 1$);
(iii) $|m_n(\chi, Z) - m_n(\chi', Z)| \le |\chi - \chi'|^{\alpha_1} n^{\alpha_2} G(Z)$ *with some* $\alpha_2 > 0$ *and* $E\{G(Z)^2\}$ *exists.*

Suppose that $\{Z_i\}_{i=1}^n$ *is a random sample from* Z. *Then, for any* $\alpha_0 > 0$, *we have*

$$\sup_{|\chi| \le n^{\alpha_0}} |\{n m_0(\chi)\}^{-1} \sum_{i=1}^n m_n(\chi, Z_i) - E\{m_n(\chi, Z_i)\}| = O_p(\sqrt{a_n \log n / n}).$$

Note that, the detailed proofs of Lemmas .1 and .2 can be referred to Xia and Härdle [4].

***Proof of Theorem* 3.1.:** Assume β be the initial estimator of β_0. Denote by

$$S_n^\beta(x) = \frac{1}{n} \sum_{i=1}^n K_h(X_{ix}^T \beta) \begin{pmatrix} 1 \\ Z_i \\ X_{ix} \end{pmatrix} \begin{pmatrix} 1 \\ Z_i \\ X_{ix} \end{pmatrix}^T$$

and

$$\begin{pmatrix} a_x^\beta \\ \theta_x^\beta \\ b_x^\beta \end{pmatrix} = \{n S_n^\beta(x)\}^{-1} \sum_{i=1}^n K_h(X_{ix}^T \beta) \begin{pmatrix} 1 \\ Z_i \\ X_{ix} \end{pmatrix} Y_i. \tag{A1}$$

From Lemma A.1, we have

$$\begin{aligned} S_n^\beta(x) &= f_\beta(x^T\beta) \begin{pmatrix} 1 & E(Z \mid X^T\beta = x^T\beta)^T & v_\beta(x)^T \\ E(Z \mid X^T\beta = x^T\beta) & E(ZZ^T \mid X^T\beta = x^T\beta) & \xi_\beta(x)^T \\ v_\beta(x) & \xi_\beta(x) & \tilde{\omega}_\beta(x) \end{pmatrix} \\ &\quad + O_p(\delta_n + h^2)(1 + |x|^2) \\ &\triangleq f_\beta(x^T\beta) \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + O_p(\delta_n + h^2)(1 + |x|^2), \end{aligned}$$

where

$$A_{11} = \begin{pmatrix} 1 & E(Z \mid X^T\beta = x^T\beta)^T \\ E(Z \mid X^T\beta = x^T\beta) & E(ZZ^T \mid X^T\beta = x^T\beta) \end{pmatrix} \in R^{(q+1)\times(q+1)}$$

$$A_{12} = (v_\beta(x), \xi_\beta(x))^T \in R^{(q+1)\times p}, \quad A_{21} = A_{12}^T \quad \text{and} \quad A_{22} = \tilde{\omega}_\beta(x) \in R^{p\times p}.$$

Then, based on the inverse operation of block matrix, it follows that

$$\{S_n^\beta(x)\}^{-1} = f_\beta(x^T\beta)^{-1} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} + O_p(\delta_n + h^2)(1 + |x|^2), \tag{A2}$$

where

$$\begin{aligned} &B_{11} = A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22.1}^{-1} A_{21} A_{11}^{-1}, \quad B_{12} = -A_{11}^{-1} A_{12} A_{22.1}^{-1}, \\ &B_{21} = B_{12}^T, \quad B_{22} = A_{22.1}^{-1} \quad \text{and} \quad A_{22.1} = A_{22} - A_{21} A_{11}^{-1} A_{12}. \end{aligned}$$

On the other hand, by applying the Taylor expansion to $g(X_i^T\beta_0)$ at $x^T\beta_0$ yields

$$\begin{aligned} Y_i &= g(x^T\beta_0) + g'(x^T\beta_0) X_{ix}^T \beta_0 + \tfrac{1}{2} g''(x^T\beta_0)(X_{ix}^T\beta_0)^2 + Z_i^T\theta_0 + \varepsilon + O(|X_{ix}^T\beta_0|^3) \\ &= g(x^T\beta_0) + g'(x^T\beta_0) X_{ix}^T \beta_0 + \frac{1}{2} g''(x^T\beta_0)(X_{ix}^T\beta)^2 \\ &\quad + Z_i^T\theta_0 + \varepsilon + \Delta_n(X_i, x, \beta), \end{aligned} \tag{A3}$$

where the second equality holds by rewriting β_0 as $\beta_0 = \beta + \delta_\beta$, and

$$\Delta_n(X_i, x, \beta) = O(|X_{ix}^T\beta_0|^3 + |X_{ix}^T\beta||X_{ix}\delta_\beta| + |X_{ix}\delta_\beta|^2).$$

In addition, it is not difficult to verify that

$$\{n S_n^\beta(x)\}^{-1} \sum_{i=1}^n K_h(X_{ix}^T\beta) \begin{pmatrix} 1 \\ Z_i \\ X_{ix} \end{pmatrix} \{g(x^T\beta_0) + g'(x^T\beta_0) X_{ix}^T\beta_0 + Z_i^T\theta_0\} = \begin{pmatrix} g(x^T\beta_0) \\ \theta_0 \\ g'(x^T\beta_0)\beta_0 \end{pmatrix}. \tag{A4}$$

Further, by virtue of Lemmas A.1 and A.2, with some calculations, we have the following two expressions hold. That is

$$\{nS_n^{\beta}(x)\}^{-1}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}1\\ Z_i\\ X_{ix}\end{pmatrix}g''(x^T\beta_0)(X_{ix}^T\beta_0)^2 = O(h^2(\delta_n+h^2))(1+|x|^2), \tag{A5}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}1\\ Z_i\\ X_{ix}\end{pmatrix}|X_{ix}^T\beta|^l|X_{ix}|^s = O(h^l)(1+|x|^{s+1}) \tag{A6}$$

for $l,s = 0,1,2,3$. Accordingly, in terms of (A6), we have

$$\frac{1}{n}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}1\\ Z_i\\ X_{ix}\end{pmatrix}\Delta_n(X_i,x,\beta) = O(h^3+h|\delta_\beta|+|\delta_\beta|^2)(1+|x|^4). \tag{A7}$$

As to the noise term, let $Z_i^* = (1, Z_i^T)^T$, from Equation (A2), it follows that

$$\begin{aligned}
&\{nS_n^{\beta}(x)\}^{-1}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}1\\ Z_i\\ X_{ix}\end{pmatrix}\varepsilon_i\\
&= \{nf_\beta(x^T\beta)\}^{-1}\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}Z_i^*\\ X_{ix}\end{pmatrix}\varepsilon_i + O((\delta_n+h^2)\delta_n)(1+|x|^2)\\
&= \{nf_\beta(x^T\beta)\}^{-1}\sum_{i=1}^{n}K_h(X_{ix}^T\beta)\begin{pmatrix}B_{11}Z_i^* + B_{12}X_{ix}\\ B_{21}Z_i^* + B_{22}X_{ix}\end{pmatrix}\varepsilon_i\\
&\quad + O((\delta_n+h^2)\delta_n)(1+|x|^2).
\end{aligned} \tag{A8}$$

Taking into account of the definitions of A_{ij}, $i,j = 1,2$, and by applying the inverse operation of block matrix leads to

$$A_{11}^{-1} = \begin{pmatrix}1+E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}E(Z\mid X^T\beta = x^T\beta) & -E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}\\ -\Omega^{-1}E(Z\mid X^T\beta = x^T\beta) & \Omega^{-1}\end{pmatrix},$$

and

$$\begin{aligned}
A_{22.1} &= A_{22} - A_{21}A_{11}^{-1}A_{12}\\
&= \tilde\omega_\beta(x) - v_\beta(x)v_\beta(x)^T - v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}E(Z\mid X^T\beta = x^T\beta)v_\beta(x)^T\\
&\quad + \xi_\beta(x)\Omega^{-1}E(Z\mid X^T\beta = x^T\beta)v_\beta(x)^T + v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}\xi_\beta(x)^T\\
&\quad - \xi_\beta(x)\Omega^{-1}\xi_\beta(x)\\
&= \tilde\omega_\beta(x) - v_\beta(x)v_\beta(x)^T + v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}\big[\xi_\beta(x)^T\\
&\quad - E(Z\mid X^T\beta = x^T\beta)v_\beta(x)^T\big] + \xi_\beta(x)\Omega^{-1}\big[E(Z\mid X^T\beta = x^T\beta)v_\beta(x)^T - \xi_\beta(x)^T\big]\\
&= \tilde\omega_\beta(x) - v_\beta(x)v_\beta(x)^T + v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T\Omega^{-1}E(\tilde Z\tilde X^T) - \xi_\beta(x)\Omega^{-1}E(\tilde Z\tilde X^T)\\
&= \tilde\omega_\beta(x) - v_\beta(x)v_\beta(x)^T + \big[v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T - \xi_\beta(x)\big]\Omega^{-1}E(\tilde Z\tilde X^T)\\
&= \tilde\omega_\beta(x) - v_\beta(x)v_\beta(x)^T - E(\tilde X\tilde Z^T)\Omega^{-1}E(\tilde Z\tilde X^T)\\
&= E(\tilde X^{\otimes 2}) - E(\tilde X\tilde Z^T)\Omega^{-1}E(\tilde Z\tilde X^T) \triangleq \Psi_\beta(x),
\end{aligned} \tag{A9}$$

where last equality holds by expanding the expressions of $\tilde\omega_\beta(x)$ and $v_\beta(x)$. Therefore, with some simple calculations, we have

$$\begin{aligned}
B_{21}Z_i^* + B_{22}X_{ix} &= -A_{22.1}^{-1}A_{21}A_{11}^{-1}Z_i^* + A_{22.1}^{-1}X_{ix}\\
&= \Psi_\beta^-(x)(-v_\beta(x) + X_{ix})\\
&\quad + \Psi_\beta^-(x)\left\{v_\beta(x)E(Z\mid X^T\beta = x^T\beta)^T - \xi_\beta(x)\right\}\Omega^{-1}\left\{Z_i - E(Z\mid X^T\beta = x^T\beta)\right\}\\
&= \Psi_\beta^-(x)\left\{[X_i - u_\beta(x)] - E(\tilde X\tilde Z^T)\Omega^{-1}[Z_i - E(Z\mid X^T\beta = x^T\beta)]\right\}.
\end{aligned} \tag{A10}$$

Finally, combining Equations (A1), (A3)–(A5) and (A7)–(A10), it follows that

$$b_x^\beta = g'(x^T\beta_0)\beta_0 + \{nf_\beta(x^T\beta)\Psi_\beta(x)\}^- \sum_{i=1}^n K_h(X_{ix}^T\beta)$$
$$\cdot \left\{[X_i - u_\beta(x)] - E(\tilde{X}\tilde{Z}^T)\Omega^{-1}[Z_i - E(Z \mid X^T\beta = x^T\beta)]\right\}\varepsilon_i$$
$$+ O(h(\delta_n + h^2) + h|\delta_\beta| + |\delta_\beta|^2)(1 + |x|^4).$$

Denote by $\theta_\beta(x) = [X_i - u_\beta(x)] - E(\tilde{X}\tilde{Z}^T)\Omega^{-1}[Z_i - E(Z \mid X^T\beta = x^T\beta)]$, then b_x^β can be rewritten as

$$b_x^\beta = g'(x^T\beta_0)\beta_0 + \{nf_\beta(x^T\beta)\Psi_\beta(x)\}^- \sum_{i=1}^n K_h(X_{ix}^T\beta)\theta_\beta(x)\varepsilon_i$$
$$+ O(h(\delta_n + h^2) + h|\delta_\beta| + |\delta_\beta|^2)(1 + |x|^4).$$

It is worth noting that the above obtained b_x^β is parallel to that derived in Lemma 6.2 of Xia [26], which plays a key role in the related theoretical proofs for the OPG based estimation procedure. As a consequence, directly using a similar approach to the proof of Theorem 4.1 in [26], the asymptotic normality of $\hat{\beta}_{\text{lsOPG}}$ is followed by

$$\sqrt{n}(\hat{\beta}_{\text{lsOPG}} - \beta_0) \xrightarrow{d} N\left(0, \Sigma_{\text{lsOPG}}\right).$$

This completes the proof of Theorem 3.1. ■

In the next, we devoted ourselves to proving Theorem 3.2. Back to the previously defined notations, $L_n(b_k, \theta)$ can be equivalently transformed as follows

$$L_n(b_k, \theta) \Leftrightarrow L_n^*(c_k) = \frac{1}{n(n-1)} \sum_{i \neq j} |\varepsilon_i - \varepsilon_j + \gamma_n c_k^T W_{ij} + \Delta_{ik} - \Delta_{jk}| K_{h,ik} K_{h,jk}.$$

Furthermore, let $Q_n(c_k)$ be the gradient function of $L_n^*(c_k)$, that is

$$Q_k(c_k) = \nabla L_n^*(c_k) = \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} \text{sgn}\{\varepsilon_i - \varepsilon_j + \gamma_n c_k^T W_{ij} + \Delta_{ik} - \Delta_{jk}\} W_{ij} K_{h,ik} K_{h,jk},$$

where sgn$(\cdot)$ denotes the sign function.

Lemma A.3: *Suppose that same conditions in Theorem* 3.2 *hold. Then,*

$$\gamma_n{}^{-1}(Q_k(c_k) - Q_k(0)) = 2\gamma_n \tau f_\beta(X_k^T\beta)^2 \Phi_\beta(X_k) c_k + o_p(1),$$

where

$$\Phi_\beta(x) = \begin{pmatrix} E(\tilde{Z}^{\otimes 2}) & E(\tilde{Z}\tilde{X}^T)/h \\ E(\tilde{X}\tilde{Z}^T)/h & E(\tilde{X}^{\otimes 2})/h^2 \end{pmatrix}.$$

Proof: Let

$$U_k = \gamma_n{}^{-1}(Q_k(c_k) - Q_k(0)) = \sum_{i \neq j} M_k(d_i, d_j),$$

where

$$M_k(d_i, d_j) = (\text{sgn}\{\varepsilon_i - \varepsilon_j + \gamma_n c_k^T W_{ij} + \Delta_{ik} - \Delta_{jk}\} - \text{sgn}\{\varepsilon_i - \varepsilon_j + \Delta_{ik} - \Delta_{jk}\}) W_{ij} K_{h,ik} K_{h,jk}.$$

By virtue of the bandwidth assumption that $h \sim n^{-\lambda}$ with $1/6 < \lambda < 1/4$, we have

$$E\{\|M_k(d_i, d_j)\|^2\} \leq 4h^{-4} E\{W_{ij} W_{ij}^T K_{ik}^2 K_{jk}^2\} = O_p(h^{-2}) = o_p(n).$$

Therefore, with some calculations based on Lemma A.1 in [18], it follows that

$$U_k = E\{M_k(d_i, d_j)\} + o_p(1) = E\{E[M_k(d_i, d_j) \mid W_{ij}]\} + o_p(1)$$
$$= 2h^{-2}\gamma_n \tau E\{W_{ij} W_{ij}^T K_{ik} K_{jk}\} c_k (1 + o_p(1)) = 2\gamma_n \tau f_\beta(X_k^T\beta)^2 \Phi_\beta(X_k) c_k + o_p(1),$$

where the last equality holds by virtue of Lemma .1. This completes the proof of Lemma A.3. ■

Lemma A.4: *Let* $T_n(c_k) = {\gamma_n}^{-1} c_k^T Q_k(0) + \gamma_n \tau f_\beta (X_k^T \beta)^2 c_k^T \Phi_\beta(X_k) c_k + {\gamma_n}^{-1} L_n^*(0)$. *Suppose that same conditions in Theorem* 3.2 *hold. Then,*

$$P\left\{ \sup_{\|c_k\| \le c} |{\gamma_n}^{-1} L_n^*(c_k) - T_n(c_k)| \ge \epsilon \right\} \to 0.$$

Proof: Note that, from the conclusion of Lemma lemma:3, we have

$$\begin{aligned}\nabla[{\gamma_n}^{-1} L_n^*(c_k) - T_n(c_k)] &= {\gamma_n}^{-1} Q_k(c_k) - {\gamma_n}^{-1} Q_k(0) - 2\gamma_n \tau f_\beta (X_k^T \beta)^2 \Phi_\beta(X_k) c_k \\ &= {\gamma_n}^{-1}[Q_k(c_k) - Q_k(0)] - 2\gamma_n \tau f_\beta (X_k^T \beta)^2 \Phi_\beta(X_k) c_k + o_p(1) = o_p(1),\end{aligned}$$

which completes the proof. ■

***Proof of Theorem* 3.2:** Assume β be the initial estimator of β_0. In order to prove this Theorem, we first demonstrate that the following equality holds, i.e.,

$$\hat{b}_k^{lr} = g'(X_k^T \beta_0)\beta_0 + \frac{n^{1/2} h {\gamma_n}^{-1}}{2\tau f_\beta (X_k^T \beta)^2} \Psi_\beta(X_k)^- \Pi + o_p(1), \tag{A11}$$

where $\Psi_\beta(\cdot)$ is defined as in (A9) and

$$\Pi = \frac{\gamma_n}{n(n-1)} \sum_{i \ne j} \text{sgn}\{\varepsilon_i - \varepsilon_j\} \left\{ X_{ij} - E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1} Z_{ij} \right\} / h K_{h,ik} K_{h,jk}.$$

Let $\tilde{c}_k$ and $\hat{c}_k$ be the minimizer of $T_n(c_k)$ and $L_n^*(c_k)$, respectively. From Lemma A.4, we can obtain that

$${\gamma_n}^{-1}(L_n^*(c_k) - L_n^*(0)) \xrightarrow{p} {\gamma_n}^{-1} c_k^T Q_k(0) + \gamma_n \tau f_\beta (X_k^T \beta)^2 c_k^T \Phi_\beta(X_k) c_k.$$

Then, it follows from the convexity lemma in Pollard [27] that

$$\hat{c}_k = \tilde{c}_k + o_p(1) = \frac{-{\gamma_n}^{-2}}{2\tau f_\beta (X_k^T \beta)^2} \Phi_\beta(X_k)^{-1} Q_k(0) + o_p(1).$$

Denote as $c_{k2} = {\gamma_n}^{-1} h(g'(X_k^T \beta_0)\beta_0 - b_k)$. By the definition of $\Phi_\beta(x)$, it follows that

$$\Phi_\beta^{-1}(x) = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

with $C_{21} = -C_{22} E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1}/h$ and $C_{22} = \Psi_\beta(X_k)^- h^2$. Hence, by the expressions of $\Phi_\beta^{-1}(x)$ and $Q_k(0)$, we have

$$\hat{c}_{k2} = \frac{-{\gamma_n}^{-2}}{2\tau f_\beta (X_k^T \beta)^2} \{C_{21} Q_{k1}(0) + C_{22} Q_{k2}(0)\} + o_p(1),$$

which is equivalent to

$${\gamma_n}^{-1} h(g'(X_k^T \beta_0)\beta_0 - \hat{b}_k^{lr}) = \frac{-{\gamma_n}^{-2}}{2\tau f_\beta (X_k^T \beta)^2} \{C_{21} Q_{k1}(0) + C_{22} Q_{k2}(0)\} + o_p(1), \tag{A12}$$

where

$$\begin{aligned} Q_{k1}(0) &= \frac{\gamma_n}{n(n-1)} \sum_{i \ne j} \text{sgn}\{\varepsilon_i - \varepsilon_j + \Delta_{ik} - \Delta_{jk}\} Z_{ij} K_{h,ik} K_{h,jk}, \\ Q_{k2}(0) &= \frac{\gamma_n}{n(n-1)} \sum_{i \ne j} \text{sgn}\{\varepsilon_i - \varepsilon_j + \Delta_{ik} - \Delta_{jk}\} X_{ij}/h K_{h,ik} K_{h,jk}. \end{aligned}$$

Note that,

$$
\begin{aligned}
Q_{k1}(0) &= \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} \operatorname{sgn}\{\varepsilon_i - \varepsilon_j + \Delta_{ik} - \Delta_{jk}\} Z_{ij} K_{h,ik} K_{h,jk} \\
&= \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} \operatorname{sgn}\{\varepsilon_i - \varepsilon_j\} Z_{ij} K_{h,ik} K_{h,jk} \\
&\quad + \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} [\operatorname{sgn}\{\varepsilon_i - \varepsilon_j + \Delta_{ik} - \Delta_{jk}\} - \operatorname{sgn}\{\varepsilon_i - \varepsilon_j\}] Z_{ij} K_{h,ik} K_{h,jk} \\
&\triangleq Q_{k11}(0) + Q_{k12}(0).
\end{aligned}
$$

By a similar proof of Lemma A.3, it is not difficult to verity that ${\gamma_n}^{-2} Q_{k12}(0) = O_p(h^2) = o_p(1)$. Therefore, we have

$$
Q_{k1}(0) = \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} \operatorname{sgn}\{\varepsilon_i - \varepsilon_j\} Z_{ij} K_{h,ik} K_{h,jk} + o_p(1). \tag{A13}
$$

In addition, via applying a similar arguments to $Q_{k2}(0)$ yields

$$
Q_{k2}(0) = \frac{\gamma_n}{n(n-1)} \sum_{i \neq j} \operatorname{sgn}\{\varepsilon_i - \varepsilon_j\} X_{ij}/h K_{h,ik} K_{h,jk} + o_p(1). \tag{A14}
$$

Then, substituting (A13) and (A14) into Equation (A12), combined with the expressions of C_{21} and C_{22} leads to

$$
\hat{b}_k^{lr} = g'(X_k^T \beta_0)\beta_0 + \frac{n^{1/2} h {\gamma_n}^{-1}}{2\tau f_\beta (X_k^T \beta)^2} \Psi_\beta (X_k)^- \Pi + o_p(1).
$$

Consequently, we have Equation (A11) holds.

Next, we put our main attention on the proof of Theorem 3.2 by using (A11). Let

$$
R_{\beta_0} = n^{-1/2} \sum_{k=1}^{n} \frac{h {\gamma_n}^{-1}}{2\tau f_\beta (X_k^T \beta)^2} \Psi_{\beta_0}(X_k)^- \Pi.
$$

By the expression of Π and the assumption that β is $\sqrt{n}$-consistent, it follows from Lemma A.2 that

$$
\begin{aligned}
R_{\beta_0} &= \frac{1}{n^2(n-1)} \sum_{k=1}^{n} \sum_{i \neq j} \frac{g'(X_k^T \beta_0)}{2\tau f_\beta (X_k^T \beta)^2} \Psi_{\beta_0}(X_k)^- \operatorname{sgn}\{\varepsilon_i - \varepsilon_j\} \left\{ X_{ij} - E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1} Z_{ij} \right\} K_{h,ik} K_{h,jk} \\
&= \frac{1}{n^2} \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{g'(X_k^T \beta_0)}{2\tau f_\beta (X_k^T \beta)} \Psi_{\beta_0}(X_k)^- [2H(\varepsilon_i) - 1] \\
&\quad \cdot \left\{ [X_i - u_{\beta_0}(X_k)] - E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1} [Z_i - E(Z \mid X^T \beta_0 = X_k^T \beta_0)] \right\} K_{h,ik} + o(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{g'(X_i^T \beta_0)}{2\tau} \Psi_{\beta_0}(X_i)^- [2H(\varepsilon_i) - 1] \left\{ v_{\beta_0}(X_i) - E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1} \tilde{Z}_i \right\} + o(n^{-1/2}),
\end{aligned}
$$

where $\tilde{Z}_i = Z_i - E(Z \mid X^T \beta_0 = X_i^T \beta_0)$. Taking into consideration of Equation (A11), then we have

$$
\frac{1}{n} \sum_{k=1}^{n} \hat{b}_k^{lr} \hat{b}_k^{lr^T} = \frac{1}{n} \sum_{k=1}^{n} (1 + O_p(h^2 + \delta_n + |\delta_\beta|)) g'(X_k^T \beta_0)^2 \|\beta_0 + \tilde{R}_{\beta_0}\|^2 \tilde{\beta}\tilde{\beta}^T + o_p(n^{-1/2}),
$$

where

$$
\tilde{\beta} = (\beta_0 + \tilde{R}_{\beta_0}) / \|\beta_0 + \tilde{R}_{\beta_0}\|, \tag{A15}
$$

$$
\tilde{R}_{\beta_0} = [E\{g'(X_k^T \beta_0)^2\}]^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{g'(X_i^T \beta_0)}{2\tau} \Psi_{\beta_0}(X_i)^- [2H(\varepsilon_i) - 1] \left\{ v_{\beta_0}(X_i) - E(\tilde{X}\tilde{Z}^T) E(\tilde{Z}^{\otimes 2})^{-1} \tilde{Z}_i \right\}
$$

.

On the other hand, as $\beta_0^T\tilde{X} = \beta_0^T(X - E(X \mid \beta_0^T X)) = \beta_0^T X - \beta_0^T X = 0$ and $\beta_0^T E(\tilde{X}\tilde{X}^T) = E(\beta_0^T \tilde{X}\tilde{X}^T) = 0$, it follows that $\beta_0^T \Xi = 0$. Accordingly, we have

$$\beta_0^T \Xi^{-1} = 0 \quad \text{and} \quad \beta_0^T \tilde{R}_{\beta_0} = 0,$$

this together with the restriction $\|\beta_0\|_2 = 1$ leads to

$$\|\beta_0 + \tilde{R}_{\beta_0}\|^2 = 1 + \|\tilde{R}_{\beta_0}\|^2 = 1 + o(n^{-1}). \tag{A16}$$

Hence, based on Equations (A15) and (A16), we obtain that

$$\tilde{\beta} = \beta_0 + \tilde{R}_{\beta_0} + o(n^{-1/2}).$$

Denote the nonzero eigenvector of $\sum_{k=1}^n \hat{b}_k^{lr}\hat{b}_k^{lr^T}/n$ in the next iteration by $\beta^{(k+1)}$, from Lemma 3.1 presented in Bai et al. [28], we have $\beta^{(k+1)} - \tilde{\beta} = o(n^{-1/2})$, which indicates that

$$\beta^{(k+1)} = \beta_0 + \tilde{R}_{\beta_0} + o_p(n^{-1/2}).$$

As a conclusion, by a similar approach as done in the proof of Theorem 1 in Feng et al. [19], and applying the central limit theorem to $\tilde{R}_{\beta_0}$, we complete the proof of Theorem 3.2. ■

***Proof of Proposition* 3.3:** Theorem 3.2 indicates that $\hat{\beta}_{\text{lrOPG}}$ is $\sqrt{n}$-consistent. Let $U = X^T\hat{\beta}_{\text{lrOPG}}$, taking into account of Condition (C5) that X lies in a compact set $\mathcal{D}$ with probability 1, then model (1) can be transformed into a partially linear model, which is a special case of varying-coefficient partially linear model studied in [24]. Therefore, Proposition 3.3 can be immediately proved by the similar proofs of Theorems 3.2 and 3.3 in Shang et al. [24]. ■

***Proof of Theorem* 3.4:** Based on the asymptotic results of Theorems 3.1 and 3.2, we have that the asymptotic covariances of the estimators $\hat{\beta}_{\text{lsOPG}}$ and $\hat{\beta}_{\text{lrOPG}}$ are

$$\sqrt{n}(\hat{\beta}_{\text{lsOPG}} - \beta_0) \xrightarrow{d} N\left(0, (\Sigma - \Gamma\Omega^{-1}\Gamma^T)^- E(\varepsilon^2)\right)$$

and

$$\sqrt{n}(\hat{\beta}_{\text{lrOPG}} - \beta_0) \xrightarrow{d} N\left(0, (\Sigma - \Gamma\Omega^{-1}\Gamma^T)^- E\{(2H(\varepsilon) - 1)^2\}/4\tau^2\right),$$

respectively. Taking into consideration of the assumptions that ε is independent of (X, Z) and $E(\varepsilon^2) = \sigma^2$, it follows that

$$\text{ARE}(\hat{\beta}_{\text{lrOPG}}, \hat{\beta}_{\text{lsOPG}}) = \frac{\text{Var}(\hat{\beta}_{\text{lsOPG}})}{\text{Var}(\hat{\beta}_{\text{lrOPG}})} = \frac{4\tau^2\sigma^2}{E\{(2H(\varepsilon) - 1)^2\}}. \tag{A17}$$

Moreover, it is easy to show that

$$\begin{aligned} E\{(2H(\varepsilon) - 1)^2\} &= \int (2H(\varepsilon) - 1)^2 h(\varepsilon)\,\mathrm{d}\varepsilon \\ &= \int 4H(\varepsilon)^2 h(\varepsilon)\,\mathrm{d}\varepsilon - 4\int H(\varepsilon)h(\varepsilon)\,\mathrm{d}\varepsilon + \int h(\varepsilon)\,\mathrm{d}\varepsilon \\ &= \int 4H(\varepsilon)^2\,\mathrm{d}H(\varepsilon) - 4\int H(\varepsilon)\,\mathrm{d}H(\varepsilon) + 1 = 1/3. \end{aligned} \tag{A18}$$

Therefore, substituting (A18) into (A17), we obtain that $\text{ARE}(\hat{\beta}_{\text{lrOPG}}, \hat{\beta}_{\text{lsOPG}}) = 12\tau^2\sigma^2$. In addition, a result of Hodges and Lehmann [29] indicates that the ARE has a lower bound 0.864, with this lower bound being obtained at the density $h(x) = \frac{3}{20\sqrt{5}}(5 - x^2)I(|x| \le 5)$. This completes the proof. ■