Chin. Ann. Math.
36B(4), 2015, 561–578
DOI: 10.1007/s11401-015-0951-3

Chinese Annals of Mathematics, Series B

Randomly Weighted LAD-Estimation for Partially Linear Errors-in-Variables Models

Xiaohan YANG[1] Rong JIANG[2] Weimin QIAN[3]

Abstract The authors consider the partially linear model relating a response Y to predictors (x, T) with a mean function $x^{\mathrm{T}}\beta_0 + g(T)$ when the $x's$ are measured with an additive error. The estimators of parameter β_0 are derived by using the nearest neighbor-generalized randomly weighted least absolute deviation (LAD for short) method. The resulting estimator of the unknown vector β_0 is shown to be consistent and asymptotically normal. In addition, the results facilitate the construction of confidence regions and the hypothesis testing for the unknown parameters. Extensive simulations are reported, showing that the proposed method works well in practical settings. The proposed methods are also applied to a data set from the study of an AIDS clinical trial group.

Keywords Partially linear errors-in-variables, LAD-estimation, Randomly weighted method, Linear hypothesis, Randomly weighted LAD-test
2000 MR Subject Classification 62E17, 62F10

1 Introduction

Consider a partially linear errors-in-variables (EV for short) model as follows:

$$\begin{cases} Y_i = x_i^{\mathrm{T}}\beta_0 + g(T_i) + \varepsilon_i, \\ X_i = x_i + u_i, \end{cases} \qquad i = 1, 2, \cdots, n, \tag{1.1}$$

where $x_i \in R^p$ are unobservable explanatory variables, $X_i \in R^p$ are manifest variables, $\beta_0 \in R^p$ is an unknown parameter vector, T_i is a scalar co-variate, the function $g(\cdot)$ is unknown, $Y_i \in R$ are responses, and $(\varepsilon, u^{\mathrm{T}})^{\mathrm{T}} \in R^{p+1}$ are independent with a common error distribution that is spherically symmetric. Spherical symmetry implies that ε_i and each component of u_i have the same distribution, which ensures model identifiability and means that $(\varepsilon, u^{\mathrm{T}})^{\mathrm{T}} =^d RU_{p+1}$ (R is a nonnegative random variable, U_{p+1} is a uniform random vector on $\Omega_p = \{a : a \in R^{p+1}, \|a\| = 1\}$, R and U_{p+1} are independent), and $(\varepsilon, u^{\mathrm{T}})^{\mathrm{T}}$ and x are independent. A detailed coverage of linear errors-in-variables models can be found in [7]. More work on nonlinear models with measurement errors can be found in [2]. Recently, the model (1.1) has been studied by Cui and Li [5], Liang et al. [13], Zhu and Cui [24] and so on. Cui and Li [5] and Liang et al.

Manuscript received May 17, 2013. Revised January 16, 2015.
[1]Department of Mathematics, Tongji University, Shanghai 200092, China.
E-mail: xiaohyang@tongji.edu.cn
[2]Department of Mathematics, Donghua University, Shanghai 200051, China.
E-mail: jrtrying@dhu.edu.cn
[3]Corresponding author. Department of Mathematics, Tongji University, Shanghai 200092, China.
E-mail: wmqian2003@aliyun.com

[13] discussed the least square estimators for the parametric and nonparametric components by the nearest neighbor estimation and the general kernel smoothing for the nonparametric component, respectively. The quantile estimate of the slope parameter β_0 has been studied by He and Liang [8].

It is well known that the least square (LS for short) method is one of the oldest and most widely used statistical tools for linear models. But, the LS estimate can be sensitive to outliers and therefore, non-robust. Unlike the LS method, the least absolute deviation (LAD for short) method is not sensitive to outliers and produces robust estimates. Due to the developments in theoretical aspects and the availability of computing power, the LAD method has become increasingly popular. In particular, it has many applications in econometrics and biomedical studies (see [1, 10]), among many others.

However, the asymptotic distribution of the estimators by the LAD method is generally related to nuisance parameter that can not be conveniently estimated. The randomly weighted method can provide a way of assessing the distribution of the estimators without estimating the nuisance parameter. The random weighting method was first proposed by Zheng [23]. An advantage of the random weighting method is that no observation is repeatedly used within each replicate of the random weighting, though each observation may be weighted unequally. This method has been used in many applications as an alternative to the bootstrap method. For example, Rao and Zhao [16] used this method to derive the approximate distribution of the M-estimator in the linear regression model. Cui et al. [6] proposed a random weighting method for the proportional hazards model. Wang et al. [19] extended the method to the censored regression model. Jiang et al. [9] discussed randomly weighted least square estimators for the unknown parameters in semi-linear EV model. A statistical analysis of the LAD method used in the partially linear regression model (1.1) with additive measurement errors, however, still seems to be missing. The objective of the present paper is to fill this gap.

In this paper, our objective is to apply the randomly weighted LAD-estimation (RWLADE for short) to partially linear EV models, and establish the asymptotic normality of the RWLADE for the parameter. These results can be used to construct confidence intervals for β_0. Furthermore, we propose a LAD-test for partially linear EV models. The LAD-test has been used by Zhao and Chen [22] to test linear hypotheses in the linear model. But the critical values of the test statistic are related to estimators of nuisance parameters. Chen et al. [3] proposed an easy and convenient randomly weighting resampling method to determine the critical values for testing linear hypotheses in the least absolute deviation regression. Motivated by this idea, we also use the randomly weighted method to determine the critical values for testing hypotheses in partially linear EV models.

The outline of the paper is as follows. In Section 2, we define the weighting scheme to be used, hence the RWLADE for β_0, and then the test statistics of it. Section 3 is the statement of the main results for β_0, and the chi-square distributions of test statistics of the proposed estimators are also given in this section. In Section 4, simulations are carried out to assess the finite sample performance of the method and also an illustration of the method to a real example is given in this section. Some concluding remarks are given in Section 5. All the technical proofs are delayed in the appendix of Section 6.

2 Definition of the Estimators

For technical convenience we will assume that T_i are confined to the interval $[0,1]$. Throughout, we shall employ a constant C $(0 < C < \infty)$ to denote some constant not depending on n, which however may assume different values at each appearance. In our proofs and statement of the results, we will let the $x's$ be independent random variables.

For any $t \in [0,1]$, we arrange $|T_1 - t|, |T_2 - t|, \cdots, |T_n - t|$ in an increasing order:

$$|T_{R(1,t)} - t| \le |T_{R(1,t)} - t| \le \cdots \le |T_{R(1,t)} - t| \tag{2.1}$$

(ties are broken by comparing indices). Obviously, $R(1,t), R(2,t), \cdots, R(n,t)$ is a permutation of $\{1, 2, \cdots, n\}$. Choose a group of fixed nonnegative numbers $\{d_{ni} : i = 1, 2, \cdots, n\}$ and let $k \equiv k_n$ be a natural number dependent solely on n. Suppose that $\{d_{ni} : i = 1, 2, \cdots, n\}$ and k satisfy

$$\frac{k}{\sqrt{n}(\log n)^2} \to \infty, \quad \frac{k}{n^{\frac{3}{4}}} \to 0, \quad n \to \infty, \tag{2.2}$$

$$\sum_{i=1}^{n} d_{ni} = 1, \quad \max d_{ni} = O\Big(\frac{1}{k}\Big), \quad \sum_{i>k} d_{ni} = o(n^{-\frac{1}{2}}). \tag{2.3}$$

Now we can define a probability weight vector $w_{ni}(t) = w_{ni}(t; T_1, T_2, \cdots, T_n), i = 1, 2, \cdots, n$ which satisfies $w_{nR(i,t)}(t) = d_{ni}, i = 1, 2, \cdots, n$. Obviously, $1 \le d_{ni} \le n, 1 \le w_{ni}(t) \le n$ for any $i = 1, 2, \cdots, n,\ t \in [0,1]$. These assumptions are commonly assumed when defining weight nonnegative functions. For example,

$$w_{ni}^1(t) = \frac{1}{h_n} \int_{S_{i-1}}^{S_i} K\Big(\frac{t-s}{h_n}\Big) \mathrm{d}s,$$

$$w_{ni}^2(t) = \frac{K\Big(\dfrac{t-T_i}{h_n}\Big)}{\sum\limits_{j=1}^{n} K\Big(\dfrac{t-T_i}{h_n}\Big)},$$

where $S_i = \frac{1}{2}(T_{(i)} + T_{(i-1)}), i = 1, \cdots, n, S_0 = 0, S_n = 1$ for any $i = 1, 2, \cdots, n,\ t \in [0,1]$.

In this paper, for any sequence of variables or functions $(S_1, \cdots, S_n)$, we always denote $\mathbf{S}^{\mathrm{T}} = (S_1, \cdots, S_n)$, $\widetilde{S}_i = S_i - \sum\limits_{j=1}^{n} W_{nj}(T_i) S_j$ and $\widetilde{\mathbf{S}}^{\mathrm{T}} = (\widetilde{S}_1, \cdots, \widetilde{S}_n)$. The conversion from $\mathbf{S}$ to $\widetilde{\mathbf{S}}$ will be applied to $Y_i, X_i, x_i, \varepsilon_i, u_i$ and $g(T_i)$. For example, $\widetilde{\mathbf{X}}^{\mathrm{T}} = (\widetilde{X}_1, \cdots, \widetilde{X}_n)$, $\widetilde{X}_i = X_i - \sum\limits_{j=1}^{n} W_{nj}(T_i) X_j$; $\widetilde{\mathbf{G}}^{\mathrm{T}} = (\widetilde{g}_1, \cdots, \widetilde{g}_n)$, $\widetilde{g}_i = g(T_i) - \sum\limits_{j=1}^{n} W_{nj}(T_i) g(T_j)$. The fact that $g(t) = E(Y_i - x_i^{\mathrm{T}}\beta \mid T_i = t)$ suggests

$$\widehat{g}_n(t) = \sum_{i=1}^{n} w_{ni}(t)(Y_i - x_i^{\mathrm{T}}\beta_0) = \sum_{i=1}^{n} w_{ni}(t) Y_i - \Big(\sum_{i=1}^{n} w_{ni}(t) x_i\Big)^{\mathrm{T}} \beta_0 = \widehat{g}_{1n}(t) - \widehat{g}_{2n}(t)^{\mathrm{T}} \beta_0 \tag{2.4}$$

as the nearest neighbor pseudo-estimator of $g(\cdot)$.

However, since β_0 is an unknown vector, we have to estimate β_0 first. Since $x_i's$ are unobservable, the least square method may be invalid. Instead of the generalized least square method used in [5], we can obtain $\widehat{\beta}_n$, the estimator of β_0, as follows:

$$\widehat{\beta}_n = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta}{\sqrt{1 + \|\beta\|^2}} \Big|. \tag{2.5}$$

But the asymptotic covariance matrix of $\widehat{\beta}_n$ involves the density of the errors and nuisance parameters and therefore is difficult to estimate reliably. To overcome this problem, we propose the following distributional approximation based on random weighting by exogenously generated i.i.d. random variables. The approach can be implemented with the simple LAD programming again.

Let $v_1, \cdots, v_n$ be a sequence of independent and identically distributed (i.i.d.) nonnegative random variables, with mean and variance both equal to 1. The standard exponential distribution has mean and variance equal to 1. Define

$$\beta_n^* = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} v_i \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta}{\sqrt{1 + \|\beta\|^2}} \Big|. \tag{2.6}$$

In this paper, we are also interested in testing the hypothesis

$$H_0 : H^{\mathrm{T}}(\beta - b_0) = 0 \leftrightarrow H_1 : H^{\mathrm{T}}(\beta - b_0) \neq 0, \tag{2.7}$$

where H is a known $p \times q$ matrix of rank q, and $b_0 \in R^p$ is a known vector $(0 < q \leq p)$.

To develop an analogue with the least absolute deviation, it is natural to consider the test statistic

$$M_n = \sum_{i=1}^{n} \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \widehat{\beta}_{nc}}{\sqrt{1 + \|\widehat{\beta}_{nc}\|^2}} \Big| - \sum_{i=1}^{n} \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \widehat{\beta}_n}{\sqrt{1 + \|\widehat{\beta}_n\|^2}} \Big|, \tag{2.8}$$

where $\widehat{\beta}_{nc} = \arg\min_{H^{\mathrm{T}}(\beta - b_0) = 0} \sum_{i=1}^{n} \big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta}{\sqrt{1 + \|\beta\|^2}} \big|$.

But the limiting distribution of M_n also involves the density function of the error terms. Chen et al. [3] proposed an easy and convenient randomly weighted resampling method to determine the critical values for testing nested linear hypotheses in the least absolute deviation regression. Motivated by this idea, we introduce a test statistic M_n^* on randomly weighted method and on the suitable centering adjustments. The approach can be implemented with the simple LAD programming again. Define

$$\begin{aligned} M_n^* = & \Big\{ \sum_{i=1}^{n} v_i \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta_{nc}^*}{\sqrt{1 + \|\beta_{nc}^*\|^2}} \Big| - \sum_{i=1}^{n} v_i \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta_n^*}{\sqrt{1 + \|\beta_n^*\|^2}} \Big| \Big\} \\ & - \Big\{ \sum_{i=1}^{n} v_i \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \widehat{\beta}_{nc}}{\sqrt{1 + \|\widehat{\beta}_{nc}\|^2}} \Big| - \sum_{i=1}^{n} v_i \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \widehat{\beta}_n}{\sqrt{1 + \|\widehat{\beta}_n\|^2}} \Big| \Big\}, \end{aligned}$$

where $\beta_{nc}^* = \arg\min_{H^{\mathrm{T}}(\beta - b_0) = 0} \sum_{i=1}^{n} v_i \big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta}{\sqrt{1 + \|\beta\|^2}} \big|$.

3 Main Results

Let the components be $x_i = (x_{ij})$. Denote $h_i = (h_{i1}, h_{i2}, \cdots, h_{ip})^{\mathrm{T}} = x_i - E(x_i \mid T_i), 1 \le i \le n$. We make the following assumptions.

Assumption 3.1 the random weights $v_1, \cdots, v_n$ are i.i.d. with $P(v_i \ge 0) = 1$, $E(v_i) = \mathrm{Var}(v_i) = 1$, and the sequence v_i and Y_i, X_i, x_i are independent.

Assumption 3.2 the distribution function F of $\varepsilon_1, \cdots, \varepsilon_n$ is absolutely continuous, with continuous density f uniformly bounded away from 0 and ∞ and $F(0) = \frac{1}{2}$.

Assumption 3.3 $ER^2 < +\infty$ and $P(R = 0) = 0$.

Assumption 3.4 The distribution of T_1 is absolutely continuous and its density $r(t)$ satisfies

$$0 < \inf_{0 \le t \le 1} r(t) \le \sup_{0 \le t \le 1} r(t) < \infty.$$

Assumption 3.5 $\Sigma = \mathrm{Cov}(x_1 - E(x_1|T_1))$ is a positive definite matrix.

Assumption 3.6 $E(|\varepsilon_1|^2 + \|x_1\|^2 + \|u_1\|^2) < \infty$; g and g_{2j} are continuous functions on the interval $[0, 1]$, where $g_{2j} = E(x_{1j} \mid T_1 = t)$ is the jth component of $g_2(t) = E(x_1 \mid T_1 = t)$ for $1 \le j \le p$.

Assumption 3.7 $E(|\varepsilon_1|^4 + \|x_1\|^4 + \|u_1\|^4) < \infty$; g and g_{2j} satisfy the Lipschitz condition and $g_{2j} = E(x_{1j} \mid T_1 = t)$ is a bounded function of t for $1 \le j \le p$.

Remark 3.1 Assumption 3.1 is commonly assumed in the random weighting method (see [19]). Assumptions 3.2–3.3 are often used in the LAD estimator (see [4, H1–H4]). Assumptions 3.4–3.6 are necessary for studying the optimal convergence rate of the nonparametric regression estimates and Assumption 3.7 guarantees the asymptotic normality of $\sqrt{n}(\beta_n^* - \beta)$, essentially the same as the conditions 1–4 of [5].

3.1 Random weighting LAD-estimation

Theorem 3.1 *Suppose that Assumptions* 3.1–3.7 *and* (2.2)–(2.3) *hold, and then*

$$\sqrt{n}(\beta_n^* - \beta_0) = \frac{\sqrt{1 + \|\beta_0\|^2}}{2f(0)} \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_i A_i + o_p(1) \xrightarrow{L} N(0, J_0^{-1} S J_0^{-1}). \tag{3.1}$$

Particularly, when $v_i \equiv 1$, *we have*

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = \frac{\sqrt{1 + \|\beta_0\|^2}}{2f(0)} \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i + o_p(1), \tag{3.2}$$

where

$$\begin{aligned} A_i &= \mathrm{sgn}(\varepsilon_i - u_i^{\mathrm{T}} \beta_0) \Big(h_i + u_i + \frac{(\varepsilon_i - u_i^{\mathrm{T}} \beta_0) \beta_0}{1 + \|\beta_0\|^2} \Big), \\ h_i &= x_i - E(x_i \mid T_i), \end{aligned}$$

$$J_0 = \frac{2f(0)}{\sqrt{1+\|\beta_0\|^2}}\Sigma,$$
$$S = \mathrm{Cov}\Big(\mathrm{sgn}(\varepsilon_1 - u_1^{\mathrm{T}}\beta_0)\Big(h_1 + u_1 + \frac{(\varepsilon_1 - u_1^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2}\Big)\Big).$$

Theorem 3.2 *Suppose that the conditions of Theorem* 3.1 *hold, and then*

$$\sqrt{n}(\beta_n^* - \widehat{\beta}_n) = \frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i - 1)A_i + o_p(1) \xrightarrow{L^*} N(0, J_0^{-1}SJ_0^{-1}). \tag{3.3}$$

Comparing (3.1) *with* (3.3)*, for the multivariate Kolmogorov-Smirnov distance between* $\sqrt{n}(\beta_n^* - \widehat{\beta}_n)$ *and* $\sqrt{n}(\widehat{\beta}_n - \beta_0)$*, we have*

$$\sup_{u\in R^p}|P^*(\sqrt{n}(\beta_n^* - \widehat{\beta}_n) \le u) - P(\sqrt{n}(\widehat{\beta}_n - \beta_0)) \le u)| \xrightarrow{L^*} 0 \quad \textit{in probability}, \tag{3.4}$$

where L^*, P^* *denote the corresponding distribution and probability conditionally on* $(X_1, Y_1, T_1), \cdots, (X_n, Y_n, T_n)$*. And the approximate to the distribution of* $\sqrt{n}(\widehat{\beta}_n - \beta_0)$ *by using random weights is valid in the weak sense.*

Remark 3.2 From Theorems 3.1–3.2, it is clear that β_n^* is a consistent estimator of β_0 and the conditionally limiting distribution of β_n^* for observations given is the same as that of $\widehat{\beta}$. Consequently, we can take the conditional distribution of β_n^* as an approximation to that of $\widehat{\beta}$ without estimating the asymptotic covariance matrix when making the confidence interval for parameters. In practical applications, this can be done by the Monte Carlo method. Specifically, one can generate random weights repeatedly for (2.6) and then obtain RWLADE of the regression parameters. Then the empirical distribution of the produced estimates is used as an approximation to the distribution of the LAD-estimator of β_0. For example, in deriving the $(1-\alpha)100\%$ confidence interval for β_0, one can implement the random weighting N times to obtain the estimates $\beta_n^{*(1)}, \beta_n^{*(2)}, \cdots, \beta_n^{*(N)}$ and hence use the lower and upper $\frac{\alpha}{2}$ quantiles of these quantities as the approximation of the lower and upper limits of the confidence interval.

3.2 LAD-test

Theorem 3.3 *Suppose that the conditions of Theorem* 3.1 *hold, and under the null hypothesis* (2.4)*, then*

$$M_n = \frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1), \tag{3.5}$$

where $H_n = \Sigma^{\frac{1}{2}}H(H^{\mathrm{T}}\Sigma H)^{-\frac{1}{2}}$, $A_i = \mathrm{sgn}(\varepsilon_i - u_i^{\mathrm{T}}\beta_0)\big(h_i + u_i + \frac{(\varepsilon_i - u_i^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2}\big)$ *and*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i \xrightarrow{L} N(0, H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}H_n), \tag{3.6}$$

where " $\xrightarrow{L}$ " *denotes approximation to the corresponding distribution,* $S = \Sigma + \frac{ER^2}{p+1}\big(I_p - \frac{\beta_0\beta_0^{\mathrm{T}}}{1+\|\beta_0\|^2}\big)$.

3.3 Random weighting LAD-test

Theorem 3.4 *Suppose that the conditions of Theorem* 3.1 *hold, and under the null hypothesis* (2.4)*, then*

$$M_n^* = \frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (v_i-1)H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1)$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (v_i-1)H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i \xrightarrow{L^*} N(0, H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}H_n). \tag{3.7}$$

Further by (3.6)*–*(3.7)*, we have*

$$L^*(M_n^*) \to L(Z) \leftarrow L(M_n),$$

where Z is the chi-squared variable with q degrees of freedom.

Remark 3.3 Theorems 3.3–3.4 show that the limiting distribution of M_n^* under the null hypothesis (2.7) is the same as the null limiting distribution of M_n. Therefore, we can directly use the conditional distribution of M_n^* as an approximation to the null distribution of M_n and determine the critical values of the test statistic M_n without estimating the nuisance parameters. It is desired to determine a sequence $c_n(\alpha)$ such that $\lim\limits_{n\to\infty} P(M_n > c_n(\alpha)) = \alpha$ under H_0, for a given level $\alpha \in (0,1)$. As shown in the sequel, the $(1-\alpha)$ quantile $c_n^*(\alpha)$ of the conditional distribution of M_n^* for given $\{Y_i, X_i\}_{i=1}^n$ can be taken as an approximation to $c_n(\alpha)$, and this can be carried out by the following procedure. Take N large enough and generate N independent replicates of random weights to obtain N randomly weighting estimates $M_{nj}^*, j = 1, \cdots, N$, so then the p-value of testing the hypothesis is approximately equal to $\sharp\{j : M_{nj}^* > M_n, j = 1, \cdots, N\}/N$. A test at the nominal significance level α is to reject H_0 if M_n is larger than the sample $(1-\alpha)$ quantile of $M_{n1}^*, M_{n2}^*, \cdots, M_{nN}^*$ and to accept H_0 otherwise. It is easy to show that, for the given nominal significant level $\alpha \in (0,1)$, the test M_n with the critical value $c_n^*(\alpha)$ has the same asymptotic level and asymptotic power as the test with the critical value $c_n(\alpha)$ obtained by estimating nuisance parameters.

4 Simulation and Real Data Study

In this section, we conduct simulation studies to assess the finite sample performance of the proposed procedures and illustrate the proposed methodology on AIDS clinical trials.

Example 4.1 The data are generated from model (1.1), where the explanatory variable x is generated from uniform distribution on the interval $(3,5)$ and $\beta_0 = 2$. $\varepsilon \sim N(0,1)$, $u \sim N(0,1)$, $g(t) = \sin(2\pi t)$, $T \sim U(0,1)$. The randomly weighting variables v_i are taken to be exponential distribution and Poisson distribution with means 1 respectively (Exp(1) and P(1)). We use the Nadaraya-Watson kernel $K(u) = \frac{15}{16}(1-u^2)^2 I\ (|u| \le 1)$; and then

$$W_{ni}(t) = \frac{K\Big(\dfrac{t-t_i}{h}\Big)}{\sum\limits_{i=1}^n K\Big(\dfrac{t-t_i}{h}\Big)}$$

is the weight function with the bandwidth $h = n^{-\frac{1}{5}}$. Since the objective is to estimate β_0, our limited experience indicates that the choice of the bandwidth h here is not as critical as it is in direct nonparametric function estimation. Sample size n is taken to be 50, 100 and 200, respectively, and we do 500 repetitions for each sample size. The number of randomly weighting is $N = 500$.

We first study the performance of parameter estimators by using our proposed method (RWLADE for short). The mean values of parameter estimators and their standard errors are respectively reported in Table 1. Table 1 shows that the performance of β_n^* is very close to the true value in all terms. Moreover, β_n^* is much more accurate when sample sizes increase.

Table 1 Simulation results for β^*

		n		
ε	w	50	100	200
N(0,1)	Exp(1)	2.0108(0.1197)	2.0343(0.1101)	2.0309(0.0910)
	P(1)	2.0101(0.1198)	2.0354(0.1100)	2.0305(0.0909)
t(2)	Exp(1)	2.0986(0.2457)	2.0188(0.1803)	2.0343(0.1448)
	P(1)	2.0981(0.2472)	2.0190(0.1804)	2.0342 (0.1455)
t(3)	Exp(1)	2.0194(0.1955)	2.0089 (0.1460)	2.0351 (0.0938)
	P(1)	2.0190(0.1952)	2.0084 (0.1452)	2.0345 (0.0939)

We then investigate the length of confidence intervals and empirical coverage rates by the randomly weighted method at the nominal levels 90% and 95%. Simulation results are respectively reported in Tables 2–3. From Table 2, it can be seen that the empirical coverage rates are reasonably close to the true values in all cases, which indicates that the randomly weighted method is valid. As expected, the coverage levels based on the different cases are much closer to the nominal levels when sample sizes increase. Table 3 shows that the length of confidence intervals is small. Not unexpectedly, the length of confidence intervals decreases with sample sizes. Finally, Tables 1–3 show that the performances of Poisson weights are exactly similar to those of exponential weights.

Table 2 Simulation results for coverage probability of confidence intervals

		0.90			0.95		
ε	w	50	100	200	50	100	200
N(0,1)	Exp(1)	0.8902	0.8980	0.8970	0.9295	0.9387	0.9411
	P(1)	0.8881	0.8973	0.8966	0.9313	0.9399	0.9418
t(2)	Exp(1)	0.8957	0.8886	0.8920	0.9392	0.9389	0.9374
	P(1)	0.8962	0.8878	0.8942	0.9410	0.9367	0.9407
t(3)	Exp(1)	0.8913	0.8934	0.8994	0.9323	0.9391	0.9441
	P(1)	0.8884	0.8956	0.8956	0.9367	0.9376	0.9376

Next, the approximation of the null distribution of the LAD-test statistics M_n, by its randomly weighted version M_n^*, is evaluated under the null hypotheses. We also study the empirical significance level and the powers of the M-test with the critical values given by the random weighting method. Throughout our simulation study, the convex function is taken as $\rho(u) = |u|$.

The null hypothesis is $H_0 : \beta_0 = 0$. Here, the randomly weighted variables are only taken to be the exponential distribution with means 1.

Table 3 Simulation results for length of confidence intervals

		0.90			0.95		
ε	w	50	100	200	50	100	200
N(0,1)	Exp(1)	0.1499	0.1002	0.0830	0.1726	0.1200	0.0994
	P(1)	0.1449	0.0993	0.0834	0.1706	0.1238	0.0995
t(2)	Exp(1)	0.2434	0.1390	0.0873	0.2814	0.1640	0.1044
	P(1)	0.2443	0.1384	0.0849	0.2830	0.1641	0.1018
t(3)	Exp(1)	0.1785	0.1066	0.0799	0.2110	0.1079	0.0954
	P(1)	0.1806	0.1245	0.0796	0.2143	0.1290	0.0924

Table 4 lists the power functions at significance levels $\alpha = 0.10$ and 0.05 for various choices of error distributions ($N(0, 1)$, $t(2)$ and $t(3)$), different sample sizes $n = 100$ and 200, and different β values 0, 0.1, 0.2 and 0.5. Note that the empirical significant levels when the true $\beta = 0$ are close to the nominal levels, implying that the randomly weighted LAD-test is a valid test. As expected, the test has a bigger power for the larger sample sizes.

Table 4 Empirical significant levels and power values

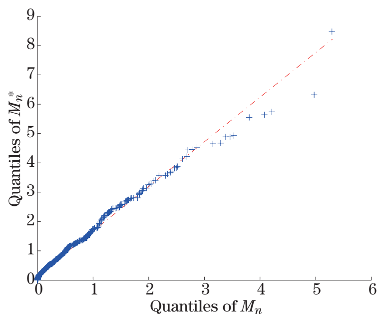
		N(0,1)		t(2)		t(3)	
n	β_2	0.10	0.05	0.10	0.05	0.10	0.05
100	0	0.0880	0.0540	0.0660	0.0340	0.0900	0.0410
	0.1	0.4780	0.1500	0.4100	0.2900	0.3300	0.2260
	0.2	0.9800	0.4800	0.7300	0.6100	0.7680	0.6700
	0.5	1.0000	0.9900	0.9600	0.9500	0.9980	0.9940
200	0	0.0980	0.0520	0.0880	0.0380	0.0943	0.0482
	0.1	0.9000	0.3300	0.6900	0.6100	0.5100	0.4160
	0.2	1.0000	0.8100	1.0000	0.9700	0.9460	0.9060
	0.5	1.0000	1.0000	1.0000	0.9700	1.0000	1.0000

Figure 1 shows quantile-quantile plots of M_n with respect to M_n^* for various choices of error distributions ($N(0, 1)$, $t(2)$ and $t(3)$), and different sample sizes $n = 100$ and 200, in which the straight lice indicates that M_n^* approximates well to the distribution of M_n. It shows that, when the sample size is increased from 100 to 200, the distribution approximation for the larger size is much more accurate than that for the small one.

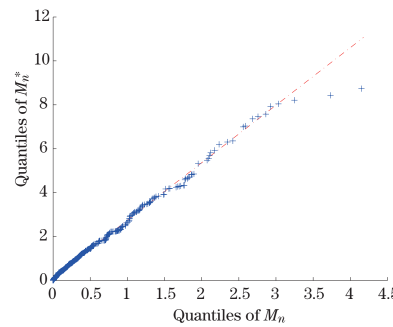
Example 4.2 In this section, we model the relationship between viral load and CD4+ cell counts in HIV-infected individuals during potent antiviral treatments based on the data from ACTG 315 study. In general, it is believed that the virologic response RNA (measured by viral load) and immunologic response (measured by CD4+ cell counts) are negatively correlated during antiviral treatment (see [12, 21]). And also the discordance between virologic and immunologic responses has been observed from several recent clinical studies (see [14–15, 17, 20]) which model the relationship between viral load and CD4+ cell counts by the mixed-effect varying-coefficient model based on these data. In their studies, exact tests and confidence intervals for parameters are not available. Instead, we present these analysis results by model (1.1). Here, we also focus on the data for the first 24 weeks of treatment, since virological or immunologic responses during this period are popular endpoints for many AIDS clinical trials. So both viral load and CD4+ cell counts were scheduled to be measured on days $t = 0, 2, 7, 10, 14, 28, 56, 84, 168$ after initiation of an antiviral therapy. We obtained 441 complete paries of viral load and CD4+ cell count observations from 48 evaluable patients. Let Y_i be the viral load and let x_i be the CD4+ cell count for subject i. To reduce the marked skewness of CD4+ cell counts and to make treatment times equal space, we take log-transformations of both variables (this is commonly used in AIDS clinical trials (see [14])). The x_i are measured with error. The model we used is

$$Y = \beta_0 + x\beta_1 + g(T) + \varepsilon, \quad X = x + u,$$

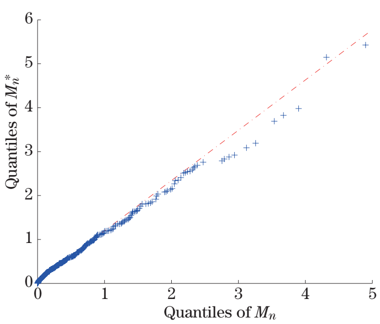
where X is the observed CD4 cell counts and T is time.



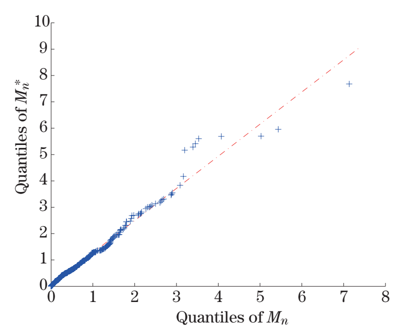
(a) $\varepsilon \sim N(0,1), n = 100$



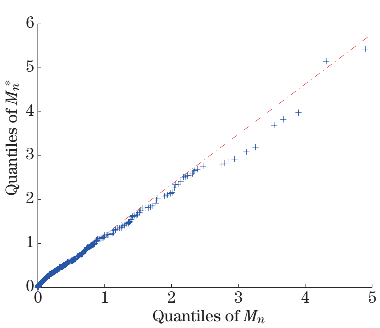
(b) $\varepsilon \sim N(0,1), n = 200$



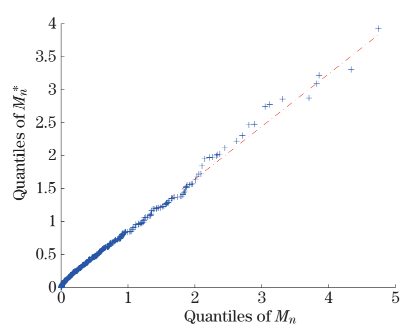
(c) $\varepsilon \sim t(2), n = 100$



(d) $\varepsilon \sim t(2), n = 200$



(e) $\varepsilon \sim t(3), n = 100$



(f) $\varepsilon \sim t(3), n = 200$

Figure 1 Q-Q plot of M_n^* v.s. M_n

The parameter estimators, by using our proposed methods, are $(\beta_0, \beta_1) = (2.7234, -0.1301)$. The 95% confidence interval of β_0 is $(2.6496, 2.7924)$ and that of β_1 is $(-0.1498, -0.1085)$. It can be seen that the length of confidence intervals is small. Furthermore, we test the linear hypothesis $H_0 : \beta_1 = 0$. The resulting p-value is 0, suggesting that β_1 is significant.

5 Discussion

The primary goal of this paper is to provide a convenient inference and a linear hypothesis testing for the partially linear EV model based on the LAD-estimate. The proposed inference procedure via resampling avoids the difficulty of density estimation and is convenient to implement with the availability of the standard linear programming and computing power. All simulation studies confirm that the performance of the random weighting method works well. We believe that the proposed statistical method is methodologically valuable. Some of the conditions assumed for the main results may be dropped or relaxed and, in particular, the samples usually may not be independent in many applications. In addition, it allows that the LAD can be extended to the M method, and the random weighting method can be used in other nonparametric regression models, such as the mixed-effect varying-coefficient model for AIDS data; the censored model or longitudinal data, which are common in survival analysis, and they are valuable subjects for future research.

6 Appendix

To prove the theorem, we first introduce the following three lemmas.

Lemma 6.1 (1) Suppose that Assumption 3.6 and (2.2)–(2.3) hold, and then

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)\varepsilon_s\Big| = o(1) \quad \text{a.s.}, \tag{6.1}$$

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)h_{sj}\Big| = o(1) \quad \text{a.s.}, \quad 1\le j\le p, \tag{6.2}$$

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)u_{sj}\Big| = o(1) \quad \text{a.s.}, \quad 1\le j\le p. \tag{6.3}$$

(2) Suppose that (2.2)–(2.3) hold, $E(|\varepsilon_1|^l + \|x_1\|^l + \|u_1\|^l) < \infty$, and g and g_{2j} satisfy the Lipschitz condition. Then

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)\varepsilon_s\Big| = o(n^{\frac{1}{l}-\frac{1}{2}}) \quad \text{a.s.}, \tag{6.4}$$

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)h_{sj}\Big| = o(n^{\frac{1}{l}-\frac{1}{2}}) \quad \text{a.s.}, \quad 1\le j\le p, \tag{6.5}$$

$$\max_{1\le i\le n}\Big|\sum_{s=1}^{n} W_{ns}(T_i)u_{sj}\Big| = o(n^{\frac{1}{l}-\frac{1}{2}}) \quad \text{a.s.}, \quad 1\le j\le p \tag{6.6}$$

for $l = 3$ or 4.

Proof This result is due to Lemma 1 of [5].

Lemma 6.2 (1) Assume that Assumption 3.2 holds and that f is a continuous function on interval $[0,1]$, and $\frac{k}{\log n} \to \infty$, $\frac{k}{n} \to 0 (n \to \infty)$. Then

$$\sup_{0 \leq t \leq 1} |T_{R(k,t)} - t| = o(1) \quad \text{a.s.}, \tag{6.7}$$

$$\max_{1 \leq i \leq n} \Big| f(T_i) - \sum_{s=1}^{n} W_{ns}(T_i) f(T_S) \Big| = o(1) \quad \text{a.s.} \tag{6.8}$$

(2) Assume that Assumption 2.2 holds and that f satisfies the Lipschitz condition and $\frac{k}{\log n} \to \infty$, $\frac{k}{n^{\frac{3}{4}}} \to 0$ $(n \to \infty)$. Then

$$\sup_{0 \leq t \leq 1} |T_{R(k,t)} - t| = o(n^{-\frac{1}{4}}) \quad \text{a.s.}, \tag{6.9}$$

$$\max_{1 \leq i \leq n} \Big| f(T_i) - \sum_{s=1}^{n} W_{ns}(T_i) f(T_S) \Big| = o(n^{-\frac{1}{4}}) \quad \text{a.s.} \tag{6.10}$$

Proof This result is due to Lemma 2 of [5].

Lemma 6.3 Under the condition of Theorem 3.1, we have

$$\frac{1}{n} \widetilde{x}^{\mathrm{T}} \widetilde{x} \to \Sigma \quad \text{a.s.} \tag{6.11}$$

Proof Observe that $x_i = h_i + g_2(T_i), 1 \leq i \leq n$, and we have

$$\begin{aligned}
&\frac{1}{n} (\widetilde{x}^{\mathrm{T}} \widetilde{x})_{s,m} \\
&= \Big(\frac{1}{n} \sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^{\mathrm{T}} \Big)_{s,m} \\
&= \frac{1}{n} \sum_{i=1}^{n} \Big[\Big(h_{is} - \sum_{j=1}^{n} W_{nj}(T_i) h_{js} \Big) + \Big(g_{2s}(T_i) - \sum_{j=1}^{n} W_{nj}(T_i) g_{2s}(T_j) \Big) \Big] \\
&\quad \times \Big[\Big(h_{im} - \sum_{j=1}^{n} W_{nj}(T_i) h_{jm} \Big) + \Big(g_{2m}(T_i) - \sum_{j=1}^{n} W_{nj}(T_i) g_{2m}(T_j) \Big) \Big] \\
&= \frac{1}{n} \sum_{i=1}^{n} h_{is} h_{im} + R_{1n}(s,m).
\end{aligned}$$

By virtue of Lemmas 6.1–6.2 and the strong law of large numbers, we have

$$\frac{1}{n} \sum_{i=1}^{n} h_{is} h_{im} \to E h_{1s} h_{1m} \quad \text{a.s.}, \quad R_{1n}(s,m) \to 0 \quad \text{a.s.}$$

and therefore,

$$\frac{1}{n} (\widetilde{x}^{\mathrm{T}} \widetilde{x})_{s,m} \to E h_{1s} h_{1m} \quad \text{a.s.},$$

so

$$\frac{1}{n} \widetilde{x}^{\mathrm{T}} \widetilde{x} \to E h_1 h_1^{\mathrm{T}} = \Sigma \quad \text{a.s.}$$

Next we proceed to prove the theorems.

Proof of Theorem 3.1 In this section, for simplicity in notation, let $\theta = \sqrt{n}(\beta - \beta_0)$.

$$\begin{aligned} Q_n(\theta) &= \sum_{i=1}^n v_i \Big(\Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta}{\sqrt{1+\|\beta\|^2}} \Big| - \Big| \frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} \Big| \Big) \\ &= \sum_{i=1}^n v_i \Big(\Big| \frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta}{\sqrt{1+\|\beta\|^2}} - \frac{\widetilde{x}_i^{\mathrm{T}}(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}} \Big| - \Big| \frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} \Big| \Big). \end{aligned}$$

Write

$$\begin{aligned} -S_i = B_i &= \frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta}{\sqrt{1+\|\beta\|^2}} - \frac{\widetilde{x}_i^{\mathrm{T}}(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}} - \frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} \\ &= -\frac{1}{\sqrt{1+\|\beta_0\|^2}} \Big(\widetilde{x}_i + \widetilde{u}_i + \frac{(\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2} \Big)^{\mathrm{T}} (\beta-\beta_0) + o(1). \end{aligned}$$

By virtue of Lemmas 6.1–6.2 and the strong law of large numbers, we have

$$-S_i = B_i = -\frac{1}{\sqrt{1+\|\beta_0\|^2}} \Big(h_i + u_i + \frac{(\varepsilon_i - u_i^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2} \Big)^{\mathrm{T}} (\beta-\beta_0) + o(1).$$

By applying the identity in Knight [11],

$$|r-s| - |r| = -s(I(r>0) - I(r<0)) + 2\int_0^s \{I(r \le t) - I(r \le 0)\}\mathrm{d}t.$$

We have

$$Q_n(\theta) = Q_{n1}(\theta) + Q_{n2}(\theta),$$

where

$$\begin{aligned} Q_{n1}(\theta) &= -\frac{1}{\sqrt{1+\|\beta_0\|^2}} \sum_{i=1}^n v_i \Big(h_i + u_i + \frac{(\varepsilon_i - u_i^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2} \Big)^{\mathrm{T}} (\beta-\beta_0)\mathrm{sgn}(\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0) + o(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{v_i A_i^{\mathrm{T}}\theta}{\sqrt{1+\|\beta_0\|^2}} + o_p(1), \\ Q_{n2}(\theta) &= 2\sum_{i=1}^n v_i \int_0^{-B_i} \Big[I\Big(\frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} \le t \Big) - I\Big(\frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} \le 0 \Big) \Big] \mathrm{d}t. \end{aligned}$$

Since

$$\frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta}{\sqrt{1+\|\beta\|^2}} =^d \frac{\varepsilon_i - \widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} =^d \frac{\varepsilon_i - u_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}} =^d \varepsilon_i,$$

where $=^d$ stands for obeying the same distribution, we have

$$\begin{aligned} EQ_{n2}(\theta) &= EQ_n(\theta) - EQ_{n1}(\theta) \\ &= E\sum_{i=1}^n \Big(\Big| \varepsilon_i - \frac{\widetilde{x}_i^{\mathrm{T}}(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}} \Big| - |\varepsilon_i| \Big) + o(1) \\ &= \frac{f(0)}{1+\|\beta\|^2} \theta^{\mathrm{T}} E\Big(\frac{1}{n}\sum_{i=1}^n \widetilde{x}_i\widetilde{x}_i^{\mathrm{T}} \Big)\theta + o(1) \\ &= \frac{f(0)}{1+\|\beta_0\|^2} \theta^{\mathrm{T}} E\Big(\frac{1}{n}\sum_{i=1}^n \widetilde{x}_i\widetilde{x}_i^{\mathrm{T}} \Big)\theta + o(1) \\ &\to \frac{f(0)}{1+\|\beta_0\|^2} \theta^{\mathrm{T}}\Sigma\theta. \end{aligned}$$

By the Schwarz's inequality and the control limited theorem, it is easy to see that

$$\begin{aligned}&\mathrm{Var}(Q_{n2}(\theta))\\&\le 2\sum_{i=1}^{n}E(v_i^2)E\Big(\int_0^{S_i}\Big[I\Big(\frac{\varepsilon_i-\widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}}\le t\Big)-I\Big(\frac{\varepsilon_i-\widetilde{u}_i^{\mathrm{T}}\beta_0}{\sqrt{1+\|\beta_0\|^2}}\le 0\Big)\Big]\mathrm{d}t\Big)^2\\&\le\frac{2}{\sqrt{n}}E\Big(\max_{1\le i\le n}\Big|\frac{\widetilde{x}_i^{\mathrm{T}}\theta}{\sqrt{1+\|\beta\|^2}}\Big|\Big)E|Q_{n2}(\theta)|\end{aligned}$$

and Assumption 3.2 then implies that

$$Q_n(\theta)\to Q_0(\theta)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{v_iA_i^{\mathrm{T}}\theta}{\sqrt{1+\|\beta_0\|^2}}+\frac{f(0)}{1+\|\beta_0\|^2}\theta^{\mathrm{T}}\Sigma\theta.$$

The convextiy of the limiting objective function $Q_0(\theta)$ assures the uniqueness of the minimizer and consequently

$$\sqrt{n}(\beta_n^*-\beta_0)=\frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_iA_i+o_p(1).$$

Furthermore,

$$E(\sqrt{n}(\beta_n^*-\beta_0))=0,\quad \mathrm{Var}(\sqrt{n}(\beta_n^*-\beta_0))=J_0^{-1}SJ_0^{-1},$$

where

$$S=\mathrm{Cov}\Big(\mathrm{sgn}(\varepsilon_1-u_1^{\mathrm{T}}\beta_0)\Big(h_1+u_1+\frac{(\varepsilon_1-u_1^{\mathrm{T}}\beta_0)\beta_0}{1+\|\beta_0\|^2}\Big)\Big);$$

particularly, when $v_1\equiv 1$, we have

$$\sqrt{n}(\widehat{\beta}_n-\beta_0)=\frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_i+o_p(1).$$

By the central limited theorem, we have

$$\sqrt{n}(\beta_n^*-\beta_0)\to^{L}N(0,J_0^{-1}SJ_0^{-1}).$$

Proof of Theorem 3.2 By the result of Theorem 3.1, we have

$$\sqrt{n}(\beta_n^*-\widehat{\beta}_n)=\frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)A_i+o_p(1). \tag{6.12}$$

From Lemma 2.9.5 in [20], it follows that conditionally on $\{Y_i,X_i,T_i\}_{i=1}^n$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)A_i\xrightarrow{L^*}N(0,S) \tag{6.13}$$

for almost every sequence $\{Y_i,X_i,T_i\}_{i=1}^n$. Thus, by (6.12)–(6.13), it is easy to show that (3.3) holds true,

$$\sqrt{n}(\beta_n^*-\widehat{\beta}_n)\xrightarrow{L^*}N(0,J_0^{-1}SJ_0^{-1}).$$

By using the similar argument as in [16], (3.4) can be shown to hold true.

Proof of Theorem 3.3 Define K as a known $p\times(p-q)$ matrix of rank $p-q$ $(0<q<p)$ which satisfies $H^{\mathrm T}K=0$. Write

$$K_n=\Sigma^{\frac12}K(K^{\mathrm T}\Sigma K)^{-\frac12},\quad H_n=\Sigma^{-\frac12}H(H^{\mathrm T}\Sigma^{-1}H)^{-\frac12}$$

and then

$$K_n^{\mathrm T}K_n=I_{p-q},\quad H_n^{\mathrm T}H_n=I_q$$

and

$$H_n^{\mathrm T}K_n=0,\quad H_nH_n^{\mathrm T}+K_nK_n^{\mathrm T}=I_p$$

Without loss of generality, $H_0:H^{\mathrm T}(\beta-b_0)=0$ can be written as

$$\beta-\beta_0=K\gamma$$

for some $\gamma\in R^{p-q}$, so

$$\sqrt n(\widehat\beta_{nc}-\beta_0)=K\widetilde\gamma.$$

Let $v_i\equiv1$. By Theorem 3.1, we have

$$Q_n(\widetilde\gamma)=-\frac{1}{\sqrt n}\sum_{i=1}^n\frac{A_i^{\mathrm T}K\widetilde\gamma}{\sqrt{1+\|\beta_0\|^2}}+\frac{f(0)}{1+\|\beta_0\|^2}\widetilde\gamma^{\mathrm T}K^{\mathrm T}\Sigma K\widetilde\gamma+o_p(1).\tag{6.14}$$

It thus follows that

$$\widetilde\gamma=\frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}(K^{\mathrm T}\Sigma K)^{-1}\frac{1}{\sqrt n}\sum_{i=1}^nK^{\mathrm T}A_i+o_p(1).\tag{6.15}$$

Replacing (6.14) into (6.15), we get

$$Q_n(\widetilde\gamma)=-\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt n}\sum_{i=1}^nK_n^{\mathrm T}\Sigma^{-\frac12}A_i\Big\|^2+o_p(1).$$

Similarly

$$\begin{aligned}Q_n(\widehat\theta)&=-\frac{1}{\sqrt n}\sum_{i=1}^n\frac{A_i^{\mathrm T}\widehat\theta}{\sqrt{1+\|\beta_0\|^2}}+\frac{f(0)}{1+\|\beta_0\|^2}\widehat\theta^{\mathrm T}\Sigma\widehat\theta+o_p(1)\\&=-\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt n}\sum_{i=1}^n\Sigma^{-\frac12}A_i\Big\|^2+o_p(1),\end{aligned}$$

where $\widehat\theta=\sqrt n(\widehat\beta_n-\beta_0)$. When H_0 is true,

$$\begin{aligned}M_n&=\sum_{i=1}^n\Big|\frac{\widetilde Y_i-\widetilde X_i^{\mathrm T}\widehat\beta_{nc}}{\sqrt{1+\|\widehat\beta_{nc}\|^2}}\Big|-\sum_{i=1}^n\Big|\frac{\widetilde Y_i-\widetilde X_i^{\mathrm T}\widehat\beta_n}{\sqrt{1+\|\widehat\beta_n\|^2}}\Big|\\&=Q_n(\widetilde\gamma)-Q_n(\widehat\theta)\\&=-\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt n}\sum_{i=1}^nK_n^{\mathrm T}\Sigma^{-\frac12}A_i\Big\|^2+\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt n}\sum_{i=1}^n\Sigma^{-\frac12}A_i\Big\|^2+o_p(1)\\&=\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt n}\sum_{i=1}^nH_n^{\mathrm T}\Sigma^{-\frac12}A_i\Big\|^2+o_p(1).\end{aligned}$$

Under the condition of Theorem 3.3. This means that the Lindeberg's condition holds. Moreover, note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i \xrightarrow{L} N(0, H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}H_n).$$

Proof of Theorem 3.4 Similar to the proof of Theorem 3.3, define

$$\begin{aligned}
Q_n(\theta^*) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{v_i A_i^{\mathrm{T}}\theta^*}{\sqrt{1+\|\beta_0\|^2}} + \frac{f(0)}{1+\|\beta_0\|^2}\theta^{*T}\Sigma\theta^* + o_p(1),\\
\theta^* &= \frac{\sqrt{1+\|\beta_0\|^2}}{2f(0)}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_i A_i + o_p(1),
\end{aligned}$$

and replacing into $Q_n(\theta^*)$, we have

$$Q_n(\theta^*) = -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_i\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1).$$

Similarly, it is easy to show that

$$\begin{aligned}
Q_n(\widehat{\theta}) &= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1),\\
Q_n(\widetilde{\gamma}^*) &= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_i K_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1),\\
Q_n(\widetilde{\gamma}) &= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} K_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1).
\end{aligned}$$

So

$$\begin{aligned}
Q_n(\theta^*) - Q_n(\widehat{\theta}) &= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1),\\
Q_n(\widetilde{\gamma}^*) - Q_n(\widetilde{\gamma}) &= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)K_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2 + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
&\theta^* = \sqrt{n}(\beta_n^* - \beta_0),\\
&\widehat{\theta} = \sqrt{n}(\widehat{\beta}_n - \beta_0),\\
&K\widetilde{\gamma}^* = \sqrt{n}(\beta_{nc}^* - \beta_0),\\
&K\widetilde{\gamma} = \sqrt{n}(\widehat{\beta}_{nc} - \beta_0).
\end{aligned}$$

Therefore

$$\begin{aligned}
M_n^* &= [Q_n(\widetilde{\gamma}^*) - Q_n(\theta^*)] - [Q_n(\widetilde{\gamma}) - Q_n(\widehat{\theta})]\\
&= [Q_n(\widetilde{\gamma}^*) - Q_n(\widetilde{\gamma})] - [Q_n(\theta^*) - Q_n(\widehat{\theta})]\\
&= -\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)K_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2
\end{aligned}$$

$$+\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)\Sigma^{-\frac{1}{2}}A_i\Big\|^2+o_p(1)$$
$$=\frac{1}{4f(0)}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i\Big\|^2+o_p(1).$$

From Lemma 2.9.5 in [18], it follows that conditionally on $\{Y_i, X_i, T_i\}_{i=1}^n$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i-1)H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}A_i \xrightarrow{L^*} N(0, H_n^{\mathrm{T}}\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}H_n).$$

Acknowledgement The authors are extremely grateful to the referees for their valuable comments and suggestions.

References

[1] Buchinsky, M., Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources*, **33**, 1998, 88–126.

[2] Carroll, R. J., Ruppert, D. and Stefanski, L. A., Nonlinear Measurement Error Models, Chapman and Hall, New York, 1995.

[3] Chen, K., Ying, Z. and Zhang, H., Analysis of least absolute deviation, *Biometrika*, **95**, 2008, 107–122.

[4] Cui, H. J., Asymptotic properties of generalized minimum L_1-norm estimates in EV model, *Science in China, Series A,* **27**, 1997, 119–131 (in Chinese).

[5] Cui, H. J. and Li, R. C., On parameter estimation for semi-linear errors-in-variables models, *Journal of Multivariate Analysis*, **64**, 1998, 1–24.

[6] Cui, W. Q., Li, K. and Yang, Y. N., Random weighting method for Cox's proportional hazards model, *Science in China, Series A*, **51**, 2008, 1843–1854.

[7] Fuller, W. A., Measurement Error Models, John Wiley and Sons, New York, 1987.

[8] Hardle, W., Liang, H. and Gao, J. T., Partially Linear Models, Physica-Verlag, Beilin, 2000.

[9] Jiang, R., Qian, W. M. and Zhou, Z. G., Randomly weighted estimators for parametric component in semi-linear errors-in-variables models, *Journal of Tongji University, Natural Science*, **39**, 2011, 768–772 (in Chinese).

[10] Jin, Z., Ying, Z. and Wei, L. J., A simple resampling method by perturbing the minimand, *Biometrika*, **88**, 2001, 381–390.

[11] Knight, K., Limiting distributions for L_1 regression estimators under general conditions, *Ann. Stat.,* **26**, 1998, 755–770.

[12] Lederman, M. M., Connick, E. and Landay, A., Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine and ritonavir: Results of AIDS clinical trials group protocol 315, *The Journal of Infectious Diseases*, **178**, 1998, 70–79.

[13] Liang, H., Hardle, W. and Carroll, R. J., Estimation in a semiparametric partially linear errors-in-variables model, *Annual Statistics*, **27**, 1999, 1519–1535.

[14] Liang, H., Wu, H. L. and Carroll, R. J., The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effect varying-coefficient semiparametric models with measurement error, *Biostatistics*, **4**, 2003, 297–312.

[15] Mallolas, J., Li, W. and Del, R. A., Clinical Outcome, CD4+ Cell Count, and HIV-1 Reverse Transcriptase and Protease Sequences in Patients Remaining Viremic during HAART, 7th Conference on Retroviruses and Opportunistic Infections, Abstract♯334, Jan 30-Feb 2, San Francisco, CA, 2000.

[16] Rao, C. R. and Zhao, L. C., Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap, *Sankhyā A*, **54**, 1992, 323–331.

[17] Sabin, C., Staszewski, S. and Phillips, A., Discordant Immunological and Virological Responses to HAART, 7th Conference on Retroviruses and Opportunistic Infections, Abstract♯333, Jan 30-Feb 2, San Francisco, CA, 2000.

[18] Van der Vaart, A. W. and Wellner, J. A., Weak Convergence and Empirical Processes, Springer-Verlag, New York, 1996.

[19] Wang, Z., Wu, Y. and Zhao, L. C., Approximation by randomly weighting method in censored regression model, *Science in China Series A*, **52**, 2009, 561–576.

[20] Wu, H., Connick, E. and Kuritzkes, D. R., Cell Kinetic Patterns and Their Relationships with Virologic Responses in HIV-1-Infected Patients Treated with HAART, 7th Conference on Retroviruses and Opportunistic Infections, Abstract♯340, Jan 30-Feb 2, San Francisco, CA, 2000.

[21] Wu, H. and Ding, A., Population HIV-1 dynamics in vivo: Applicable models and inferential tools for virological data from AIDS clinical trials, *Biometrics*, **55**, 1999, 410–418.

[22] Zhao, L. C. and Chen, X. R., Asymptotic behavior of M-test statistics in linear models, *Journal of Combine Information System Science*, **16**, 1991, 234–248.

[23] Zheng, Z. G., Random weighting method, *Acta Mathematicae Applilcate Sinica*, **10**, 1987, 247–253 (in Chinese).

[24] Zhu, L. and Cui, H. J., A semiparametric regression model with errors in variables, *Scan. Journal Statistics*, **30**, 2003, 429–442.