Zeng and Liu *Journal of Inequalities and Applications* (2018) 2018:267
https://doi.org/10.1186/s13660-018-1857-x

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH **Open Access**

A difference-based approach in the partially linear model with dependent errors

Zhen Zeng[1] and Xiangdong Liu[1*]

*Correspondence:
tliuxd@jnu.edu.cn;
294596515@qq.com
[1]Department of Statistics, Jinan University, Guangzhou, P.R. China

Abstract

We study asymptotic properties of estimators of parameter and non-parameter in a partially linear model in which errors are dependent. Using a difference-based and ordinary least square (DOLS) method, the estimator of an unknown parametric component is given and the asymptotic normality of the DOLS estimator is obtained. Meanwhile, the estimator of a nonparametric component is derived by the wavelet method, and asymptotic normality and the weak convergence rate of the wavelet estimator are discussed. Finally, the performance of the proposed estimator is evaluated by a simulation study.

MSC: 62G05; 62G20

Keywords: NSD random variables; Partially linear model; Asymptotic normality; Finite difference; Least square

1 Introduction

Consider the partially linear model (PLM)

$$y_i = x_i^T \beta + f(t_i) + e_i, \quad 1 \le i \le n, \tag{1}$$

where the superscript T denotes the transpose, y_i are scalar response variables, $x_i = (x_{i1}, \ldots, x_{id})^T$ are explanatory variables, β is a d-dimensional column vector of the unknown parameter, $f(\cdot)$ is an unknown function, t_i are deterministic with $0 \le t_1 \le \cdots \le t_n \le 1$, and e_i are random errors.

PLM was first considered by Engle et al. [1], and now is one of the most widely used statistical models. It can be applied in almost every field, such as engineering, economics, medical sciences and ecology, etc. There are many authors (see [2–8]) concerned with various estimation methods to obtain estimators of the unknown parameters and nonparameters for partially linear model. Deep results such as asymptotic normality of estimators have been obtained.

In this paper, by a difference-based approach, we will use the ordinary least square and wavelet to investigate model (1). The differencing procedures provide a convenient means for introducing nonparametric techniques to practitioners in a way which parallels their knowledge of parametric techniques, and differencing procedures may easily be combined with other procedures. For example, Wang et al. [9] obtained a difference-based approach to the semiparametric partially linear model. Tabakan et al. [10] studied a difference-based ridge in partially linear model. Duran et al. [11] investigated the difference-based ridge and Liu type estimators in semiparametric regression models. Hu et al. [12] used a difference-based Huber Dutter estimator (DHD) to obtain the root variance σ and parametric β for partially linear model. Wu [13] constructed the restricted difference-based Liu estimator for the parametric component of partially linear model. However, in the majority of the previous work it is assumed that errors are independent. The asymptotic problem of difference-based estimators of partially linear model with dependent errors is in practice important. In this paper, we use a difference-based and ordinary least square method to study the partially linear model with dependent errors.

For the dependent errors e_i we confine ourselves to negatively superadditive dependent (NSD) random variables. There are many applications of NSD random variables in multivariate statistical analysis; see [14–23]. Hence, it is meaningful to study the properties of NSD random variables. The formal definition of NSD random variables is the following.

Definition 1 (Kemperman [24]) A function Φ: $\mathbf{R}^n \to \mathbf{R}$ is called superadditive if $\Phi(x \vee y) + \Phi(x \wedge y) \geq \Phi(x) + \Phi(y)$ for all $x, y \in \mathbf{R}^n$, where $\vee$ stands for componentwise maximum, and $\wedge$ for componentwise minimum.

Definition 2 (Hu [25]) A sequence $\{e_1, e_2, \dots, e_n\}$ is said to be NSD if

$$E\Phi(e_1, e_2, \dots, e_n) \leq E\Phi(Y_1, Y_2, \dots, Y_n), \tag{2}$$

where $Y_1, Y_2, \dots, Y_n$ are independent with $e_i \overset{d}{=} Y_i$ for each i, and Φ is a superadditive function such that the expectations in (2) exist. An infinite sequence $\{e_n, n \geq 1\}$ of random variables is said to be NSD if $\{e_1, e_2, \dots, e_n\}$ is NSD for all $n \geq 1$.

In addition, using the wavelet method (see [26–29]), the weak convergence rate and asymptotic normality of the estimator of $f(\cdot)$ are obtained.

Throughout the paper we fix the following notations. β_0 is the true value of the unknown parameter β. $\mathbf{Z}$ is the set of integers, $\mathbf{N}$ is the set of natural numbers, $\mathbf{R}$ is the set of real numbers. Denote $x^+ = \max(x, 0)$, and $x^- = (-x)^+$. Let C_1, C_2, C_3, C_4 are positive constants. For a sequence of random variables η_n and a positive sequence d_n, write $\eta_n = o(d_n)$ if η_n/d_n converges to 0 and $\eta_n = O(d_n)$ if η_n/d_n is bounded. We can similarly define the notations of o_P and O_P for stochastic convergence and stochastic bounded. Weak convergence of a distribution is denoted by $H_n \xrightarrow{D} H$, and for random variables by $Y_n \xrightarrow{D} Y$. $\|x\|$ is the Euclidean norm of x, and $\lfloor x \rfloor = \max\{k \in \mathbf{Z} : k \leq x\}$.

2 Estimation method

Define the $(n - m) \times n$ differencing matrix D as

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & d_0 & d_1 & d_2 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & d_0 & d_1 & d_2 & \cdots & d_m \end{pmatrix},$$

where the positive integer number m is the order of differencing and $d_0, d_1, \ldots, d_m$ are differencing weights satisfying

$$\sum_{q=0}^{m} d_q = 0, \qquad \sum_{q=0}^{m} d_q^2 = 1. \tag{3}$$

This differencing matrix is given by Yatchew [30]. Using the differencing matrix to model (1), we have

$$DY = DX\beta + Df + De. \tag{4}$$

From Yatchew [30], the application of differencing matrix D in model (1) can remove the nonparametric effect in large samples, so we will ignore the presence of Df. Thus, we can rewrite (4) as

$$\tilde{Y} = \tilde{X}\beta + \tilde{e}, \tag{5}$$

where $\tilde{Y} = (\tilde{y}_1, \ldots, \tilde{y}_{n-m})^T$, $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_{n-m})^T$ and $\Sigma_n = \tilde{X}^T\tilde{X}$ is nonsingular for large n, $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_{n-m})^T$, $\tilde{y}_i = \sum_{q=0}^{m} d_q y_{i+q}$, $\tilde{x}_i = \sum_{q=0}^{m} d_q x_{i+q}$, $\tilde{e}_i = \sum_{q=0}^{m} d_q e_{i+q}$, $i = 1, \ldots, n-m$.

As a usual regression model, the ordinary least square estimator $\hat{\beta}_n$ of the unknown parameter β is given as

$$\hat{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{n-m} \left(\tilde{y}_i - \tilde{x}_i^T \beta\right)^2. \tag{6}$$

Then the estimator satisfies

$$-2\sum_{i=1}^{n-m} \tilde{x}_i\left(\tilde{y}_i - \tilde{x}_i^T \hat{\beta}_n\right) = 0,$$

and hence

$$\hat{\beta}_n = \Sigma_n^{-1}\tilde{X}^T\tilde{Y}. \tag{7}$$

In the following, we use wavelet techniques to estimate $f(\cdot)$ if $\hat{\beta}_n$ is known.

Suppose that there exists a scaling function $\phi(\cdot)$ in the Schwartz space S_l and a multiresolution analysis $\{V_{\tilde{m}}\}$ in the concomitant Hilbert space $L^2(\mathbf{R})$, with the reproducing kernel $E_{\tilde{m}}(t,s)$ given by

$$E_{\tilde{m}}(t,s) = 2^{\tilde{m}} E_0\left(2^{\tilde{m}}t, 2^{\tilde{m}}s\right) = 2^{\tilde{m}} \sum_{k\in\mathbf{Z}} \phi\left(2^{\tilde{m}}t - k\right)\phi\left(2^{\tilde{m}}s - k\right).$$

Let $A_i = [s_{i-1}, s_i]$ denote intervals that partition $[0,1]$ with $t_i \in A_i$ for $1 \leq i \leq n$. Then the estimator of the nonparameter $f(t)$ is given by

$$\hat{f}_n(t) = \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta}_n\right) \int_{A_i} E_{\tilde{m}}(t,s)\, ds. \tag{8}$$

3 Preliminary conditions and lemmas

In this section, we give the following conditions and lemmas which will be used to obtain the main results.

(C1) $\max_{1\le i\le n}\|x_i\| = C_1 < \infty$.

(C2) $f(\cdot)\in H^\alpha$ (Sobolev space), for some $\alpha > 1/2$.

(C3) $f(\cdot)$ is Lipschitz function of order $\gamma > 0$.

(C4) $\phi(\cdot)$ belongs to S_l, which is a Schwartz space for $l\ge\alpha$. $\phi(\cdot)$ is a Lipschitz function of order 1 and has compact support, in addition to $|\hat{\phi}(\xi) - 1| = O(\xi)$ as $\xi\to 0$, where $\hat{\phi}$ denotes Fourier transform of ϕ.

(C5) s_i, $1\le i\le n$, satisfy $\max_{1\le i\le n}(s_i - s_{i-1}) = O(n^{-1})$, and $2^{\tilde{m}} = O(n^{1/3})$.

Remark 3.1 Condition (C1) is standard and often imposed in the estimator of partial linear models, once can refer to Zhao et al. [31]. Conditions (C2)–(C5) are used by Hu et al. [29]. Therefore, our conditions are very mild and can easily be satisfied.

Lemma 3.1 (Hu [25]) *Suppose that* $\{e_1, e_2, \dots, e_n\}$ *is NSD.*

(i) *If* $g_1, g_2, \dots, g_n$ *are nondecreasing functions, then* $\{g_1(e_1), g_2(e_2), \dots, g_n(e_n)\}$ *is NSD.*

(ii) *For any* $2\le m\le n$ *and* $1\le i_1 < i_2 < \cdots < i_m$, $\{e_{i_1}, e_{i_2}, \dots, e_{i_m}\}$ *is NSD.*

Lemma 3.2 (Wang et al. [17]) *Let* $p > 1$. *Let* $\{e_n, n\ge 1\}$ *be a sequence of NSD random variables with* $Ee_n = 0$ *and* $E|e_n|^p < \infty$ *for each* $n\ge 1$. *Then for all* $n\ge 1$,

$$E\left(\max_{1\le k\le n}\left|\sum_{i=1}^k e_i\right|^p\right) \le 2^{3-p}\sum_{i=1}^n E|e_i|^p \quad \text{for } 1 < p\le 2 \tag{9}$$

and

$$E\left(\max_{1\le k\le n}\left|\sum_{i=1}^k e_i\right|^p\right) \le 2\left(\frac{15p}{\ln p}\right)^p\left[\sum_{i=1}^n E|e_i|^p + \left(\sum_{i=1}^n Ee_i^2\right)^{p/2}\right] \quad \text{for } p > 2. \tag{10}$$

Lemma 3.3 *Let* $p > 1$. *Let* $\{e_n, n\ge 1\}$ *be a sequence of NSD random variables with* $Ee_n = 0$ *and* $E|e_n|^p < \infty$ *for all* $n\ge 1$, *and* $\{c_q, 0\le q\le m\}$ *be a sequence of real constants. Then for all* $n\ge 1$,

$$E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^k\sum_{q=0}^m c_q e_{i+q}\right|^p\right) \le 4m^{p-1}\sum_{i=1}^n\sum_{q=0}^m E|c_q e_{i+q}|^p \quad \text{for } 1 < p\le 2 \tag{11}$$

and, for $p > 2$,

$$\begin{aligned} &E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^k\sum_{q=0}^m c_q e_{i+q}\right|^p\right) \\ &\quad\le 2^{p+1}m^{p-1}\left(\frac{15p}{\ln p}\right)^p\left[\sum_{i=1}^n\sum_{q=0}^m E|c_q e_{i+q}|^p + \left(\sum_{i=1}^n\sum_{q=0}^m E(c_q e_{i+q})^2\right)^{p/2}\right]. \end{aligned} \tag{12}$$

Proof Let $z_{1i} = \sum_{q=0}^m c_q^+ e_{i+q}$, $z_{2i} = \sum_{q=0}^m c_q^- e_{i+q}$, then $\sum_{q=0}^m c_q e_{i+q} = z_{1i} - z_{2i}$, and $\{c_q^+ e_{i+q}, i\ge 1\}$ and $\{c_q^- e_{i+q}, i\ge 1\}$ are both NSD random variables for all $0\le q\le m$ by Lemma 3.1. By the

C_r-inequality,

$$\begin{aligned}
&E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}\sum_{q=0}^{m}c_q e_{i+q}\right|^p\right)\\
&\quad= E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}(z_{1i}-z_{2i})\right|^p\right)\\
&\quad\le 2^{p-1}\left\{E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}z_{1i}\right|^p\right)+E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}z_{2i}\right|^p\right)\right\}\\
&\quad\le 2^{p-1}m^{p-1}\sum_{q=0}^{m}\left\{E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}c_q^+ e_{i+q}\right|^p\right)+E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}c_q^- e_{i+q}\right|^p\right)\right\}.
\end{aligned}$$

In the case $1 < p \le 2$, it follows from Lemma 3.2 that

$$\begin{aligned}
&E\left(\max_{1\le k\le n-m}\left|\sum_{k=1}^{k}\sum_{q=0}^{m}c_q e_{i+q}\right|^p\right)\\
&\quad\le 2^{p-1}m^{p-1}\sum_{q=0}^{m}\left\{E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}c_q^+ e_{i+q}\right|^p\right)+E\left(\max_{1\le k\le n-m}\left|\sum_{i=1}^{k}c_q^- e_{i+q}\right|^p\right)\right\}\\
&\quad\le 4m^{p-1}\left(\sum_{i=1}^{n-m}\sum_{q=0}^{m}E\left|c_q^+ e_{i+q}\right|^p+\sum_{i=1}^{n-m}\sum_{q=0}^{m}E\left|c_q^- e_{i+q}\right|^p\right). \qquad (13)
\end{aligned}$$

Note that $|c_q|^p = |c_q^+|^p + |c_q^-|^p$, the desired result (11) follows from (13) immediately. In the same way, we also have (12). The proof is completed. □

Remark 3.2 From Lemma 3.3 and Lemma 3.1, we have, for $1 < p \le 2$,

$$E\left(\left|\sum_{i\in S}\sum_{q=0}^{m}c_q e_{i+q}\right|^p\right)\le 4m^{p-1}\left\{\sum_{i\in S}\sum_{q=0}^{m}E|c_q e_{i+q}|^p\right\} \qquad (14)$$

and, for $p > 2$,

$$\begin{aligned}
&E\left(\left|\sum_{i\in S}\sum_{q=0}^{m}c_q e_{i+q}\right|^p\right)\\
&\quad\le 2^{p+1}m^{p-1}\left(\frac{15p}{\ln p}\right)^p\left[\sum_{i\in S}\sum_{q=0}^{m}E|c_q e_{i+q}|^p+\left(\sum_{i\in S}\sum_{q=0}^{m}E(c_q e_{i+q})^2\right)^{p/2}\right], \qquad (15)
\end{aligned}$$

where $S \subset \{1,2,\ldots,n\}$.

Lemma 3.4 *Let A and B be disjoint subsets of $\mathbf{N}$, and $\{X_j, j\in A\cup B\}$ be a sequence of NSD random variables. Let $f: \mathbf{R}\to\mathbf{R}$ and $g: \mathbf{R}\to\mathbf{R}$ be differentiable with bounded derivatives, and $\|\cdot\|_\infty$ stand for supnorm. Then*

$$\left|\mathrm{Cov}\left\{f\left(\sum_{i\in A}a_iX_i\right),g\left(\sum_{j\in A}a_jX_j\right)\right\}\right|\le\|f'\|_\infty\|g'\|_\infty\left|\mathrm{Cov}\left(\sum_{i\in A}a_iX_i,\sum_{j\in B}a_jX_j\right)\right|,$$

provided the covariation on the right hand side exists, where $\{a_i, 1 \leq i \leq n\}$ *is an array of real numbers.*

Proof For a pair of random variables $Z_1 = \sum_{i\in A} a_i X_i$, $Z_2 = \sum_{j\in B} a_j X_j$, we have

$$H(z_1, z_2) = P(Z_1 \leq z_1, Z_2 \leq z_2) - P(Z_1 \leq z_1)P(Z_2 \leq z_2).$$

Denote by $F(z_1, z_2)$ the joint distribution functions of (Z_1, Z_2), and $F_{Z_1}(z_2), F_{Z_2}(z_2)$ the marginal distribution function of Z_1, Z_2, one gets

$$\begin{aligned}\operatorname{Cov}(Z_1, Z_2) &= E(Z_1 Z_2) - E(Z_1)E(Z_2)\\ &= \int\int \left[F(z_1, z_2) - F_{Z_1}(z_1)F_{Z_2}(z_2)\right] dz_1\, dz_2 = \int\int H(z_1, z_2)\, dz_1\, dz_2,\end{aligned}$$

this relation was established in Lehmann [32] for any two random variables Z_1 and Z_2 with $\operatorname{Cov}(Z_1, Z_2)$ exist. Let f, g are complex valued function on $\mathbf{R}$ with derivatives $f', g' < \infty$, then we have

$$\begin{aligned}&\left|\operatorname{Cov}\bigl(f(Z_1), g(Z_2)\bigr)\right|\\ &\quad= \int\int f'(Z_1)g'(Z_2)H(z_1, z_2)\, dz_1\, dz_2\\ &\quad\leq \int\int \left|f'(Z_1)\right|\left|g'(Z_2)\right|\left|H(z_1, z_2)\right| dz_1\, dz_2 \leq \left\|f'\right\|_\infty \left\|g'\right\|_\infty \left|\operatorname{Cov}(Z_1, Z_2)\right|.\end{aligned}$$

The proof is completed. □

Lemma 3.5 *Let* $\{e_n, n \geq 1\}$ *be a sequence of NSD random variable with* $Ee_n = 0$. *Let* $\tilde{e}_{i_j} = \sum_{q=0}^m d_q e_{i_j+q}$, *and* $|i_j - i_k| > m$ *if* $j \neq k$. *Then*

$$\left|E\exp\left(\mathrm{i}\sum_{j=1}^n t_{i_j}\tilde{e}_{i_j}\right) - \prod_{j=1}^n E\exp(\mathrm{i}t_{i_j}\tilde{e}_{i_j})\right| \leq -\sum_{j=1}^n \sum_{k=j+1}^n \sum_{q1=0}^m \sum_{q2=0}^m t_0^2 \operatorname{Cov}(e_{i_j+q1}, e_{i_k+q2}), \qquad (16)$$

where $\mathrm{i} = \sqrt{-1}$, $\sum_{q=0}^m d_q = 0$ *and* $\sum_{q=0}^m d_q^2 = 1$, $t_{i_1}, t_{i_2}, \ldots, t_{i_n}$ *are real numbers with* $|t_{i_j}| \leq t_0$.

Proof Notice that the result is true for $n = 1$.

For $n = 2$, let $f(\tilde{e}_{i_1}) = \exp\{\mathrm{i}t_{i_1}\tilde{e}_{i_1}\}, g(\tilde{e}_{i_2}) = \exp\{\mathrm{i}t_{i_2}\tilde{e}_{i_2}\}$. Then, by Lemma 3.4 and $\sum_{q=0}^m d_q^2 = 1$,

$$\begin{aligned}&\left|E\exp\{\mathrm{i}t_{i_1}\tilde{e}_{i_1} + \mathrm{i}t_{i_2}\tilde{e}_{i_2}\} - E\exp\{\mathrm{i}t_{i_1}\tilde{e}_{i_1}\}E\exp\{\mathrm{i}t_{i_2}\tilde{e}_{i_2}\}\right|\\ &\quad= \left|\operatorname{Cov}\bigl(\exp\{\mathrm{i}t_{i_1}\tilde{e}_{i_1}\}, \exp\{\mathrm{i}t_{i_2}\tilde{e}_{i_2}\}\bigr)\right|\\ &\quad\leq t_0^2\left|\sum_{q1=0}^m \sum_{q2=0}^m d_{q1}\, d_{q2} \operatorname{Cov}(e_{i_1+q1}, e_{i_2+q2})\right|\\ &\quad\leq -t_0^2 \sum_{q1=0}^m \sum_{q2=0}^m \operatorname{Cov}(e_{i_1+q1}, e_{i_2+q2}).\end{aligned}$$

Hence, the result is true for $n = 2$.

Moreover, suppose that (16) holds for $n-1$. By Lemma 3.4, we have, for n,

$$\begin{aligned}
&\left|E\exp\left\{\mathrm{i}\sum_{j=1}^{n} t_{i_j}\tilde{e}_{i_j}\right\} - \prod_{j=1}^{n} E\exp\{\mathrm{i}t_{i_j}\tilde{e}_{i_j}\}\right| \\
&\le \left|E\exp\left\{\mathrm{i}\sum_{j=1}^{n} t_{i_j}\tilde{e}_{i_j}\right\} - E\exp\left\{\mathrm{i}\sum_{j=1}^{n-1} t_{i_j}\tilde{e}_{i_j}\right\}E\exp\{\mathrm{i}t_{i_n}\tilde{e}_{i_n}\}\right| \\
&\quad + \left|E\exp\left\{\mathrm{i}\sum_{i=1}^{n-1} t_{i_j}\tilde{e}_{i_j}\right\}E\exp\{\mathrm{i}t_{i_n}\tilde{e}_{i_n}\} - \prod_{j=1}^{n-1} E\exp\{\mathrm{i}t_{i_j}\tilde{e}_{i_j}\}E\exp\{\mathrm{i}t_{i_n}\tilde{e}_{i_n}\}\right| \\
&\le \left|\mathrm{Cov}\left(\exp\left\{\mathrm{i}\sum_{i=1}^{n-1} t_{i_j}\tilde{e}_{i_j}\right\}, \exp\{\mathrm{i}t_{i_n}\tilde{e}_{i_n}\}\right)\right| + \left|E\exp\left\{\mathrm{i}\sum_{j=1}^{n-1} t_{i_j}\tilde{e}_{i_j}\right\} - \prod_{j=1}^{n-1} E\exp\{\mathrm{i}t_{i_j}\tilde{e}_{i_j}\}\right| \\
&\le \left|\mathrm{Cov}\left(\exp\left\{\mathrm{i}\sum_{i=1}^{n-1} t_{i_j}\tilde{e}_{i_j}\right\}, \exp\{\mathrm{i}t_{i_n}\tilde{e}_{i_n}\}\right)\right| + \sum_{j=1}^{n-1}\sum_{k=j+1}^{n-1}\sum_{q1=0}^{m}\sum_{q2=0}^{m} t_0^2\left|\mathrm{Cov}(e_{i_j+q1}, e_{i_k+q2})\right| \\
&\le -t_0^2\sum_{j=1}^{n}\sum_{k=j+1}^{n}\sum_{q1=0}^{m}\sum_{q2=0}^{m}\mathrm{Cov}(e_{i_j+q1}, e_{i_k+q2}),
\end{aligned}$$

which completes the proof. □

Lemma 3.6 (Hu et al. [29]) *If Condition* (C3) *holds, then*

(a1) $|E_0(t,s)| \le \frac{C_k}{(1+|t-s|)^k}$, $|E_{\tilde{m}}(t,s)| \le \frac{2^{\tilde{m}}C}{(1+2^{\tilde{m}}|t-s|)^k}$ *(where $k \in \mathbf{N}$ and $C = C(k)$ is a constant depending on k only).*

(a2) $\sup_{0\le s\le 1}|E_{\tilde{m}}(t,s)| = O(2^{\tilde{m}})$.

(a3) $\sup_t \int_0^1 |E_{\tilde{m}}(t,s)|\,ds \le C_2$.

(a4) $\int_0^1 E_{\tilde{m}}(t,s)\,ds \to 1, n \to \infty$.

Lemma 3.7 (Rao [33]) *Suppose that $\{X_n, n \ge 1\}$ are independent random variables with $EX_n = 0$ and $s_n^{-(2+\delta)}\sum_{j=1}^{n} E|X_j|^{2+\delta} \to 0$ for some $\delta > 0$. Then*

$$s_n^{-1}\sum_{j=1}^{n} X_j \xrightarrow{D} N(0,1),$$

where $s_n^2 = \sum_{j=1}^{n} EX_j^2 = \mathrm{Var}(\sum_{j=1}^{n} X_j)$.

Lemma 3.8 (Yu et al. [34]) *Let $\{e_n, n \ge 1\}$ be a sequence of NSD random variable satisfying $Ee_n = 0$, $\sup_{j\ge 1}\sum_{i:|i-j|\ge u}|\mathrm{Cov}(e_i,e_j)| \to 0$ as $u \to \infty$, and $\{a_{ni}, 1 \le i \le n, n \ge 1\}$ be an array of real numbers with $\max_{1\le i\le n}|a_{ni}| \to 0$ and $\sum_{i=1}^{n} a_{ni}^2 = O(1)$. Suppose that $\{e_n, n \ge 1\}$ is uniformly integral in L_2, then*

$$\sigma_n^{-1}\sum_{i=1}^{n} a_{ni}e_i \xrightarrow{D} N(0,1),$$

where $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^{n} a_{ni}e_i)$.

4 Main results and their proofs

Theorem 4.1 *Under Condition* (C1), *suppose that* $\{e_n, n \geq 1\}$ *is a sequence of NSD random variables with* $Ee_n = 0$ *and*

(i) $\sup_{n\geq 1} E|e_n|^{2+\delta} < \infty$ *for some* $\delta > 0$,

(ii) $\sup_{j\geq 1} \sum_{i:|i-j|\geq u} |\operatorname{Cov}(e_i, e_j)| \to 0$ *as* $u \to \infty$. *Then*

$$(n-m)^{-\frac{1}{2}} \tau_\beta^{-1} \Sigma_n(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I_d) \tag{17}$$

provided that

$$\tau_\beta^2 = \lim_{n\to\infty} (n-m)^{-1} \left\{ \sum_{i=1}^{n-m} \tilde{x}_i \tilde{x}_i^T \operatorname{Var}(\tilde{e}_i) + 2 \sum_{i=1}^{n-m} \sum_{j=i+1}^{n-m} \tilde{x}_i \tilde{x}_j^T \operatorname{Cov}(\tilde{e}_i, \tilde{e}_j) \right\} \tag{18}$$

is a positive definite matrix, where I_d *is the identity matrix of order* d.

Proof By Condition (i), we have

$$\sup_{n\geq 1} Ee_n^2 < \infty \quad \text{and} \quad \lim_{x\to\infty} \sup_{n\geq 1} Ee_n^2 I\{|e_n| > x\} = 0,$$

from which it follows that

$$C_3 := \sup_{n>m} (n-m)^{-1} \sum_{i=1}^{n-m} \sum_{q=0}^{m} \operatorname{Var}(d_q e_{i+q}) < \infty,$$

and for all $\varepsilon > 0$

$$(n-m)^{-1} \sum_{i=1}^{n-m} \sum_{q=0}^{m} E(d_q e_{i+q})^2 I\{|d_q e_{i+q}| \geq \sqrt{n-m}\,\varepsilon\} \to 0 \quad \text{as } n \to \infty.$$

Then we can find a positive number sequence $\{\varepsilon_n, n \geq 1\}$ with $\varepsilon_n \to 0$ such that

$$(n-m)^{-1} \sum_{i=1}^{n-m} \sum_{q=0}^{m} E(d_q e_{i+q})^2 I\{|d_q e_{i+q}| \geq \sqrt{n-m}\,\varepsilon_n\} \to 0 \quad \text{as } n \to \infty.$$

Now, we define the integers: $m_0 = 0$, and, for each $j = 0, 1, 2, \ldots$, put

$$m_{2j+1} = \min\left\{ m' : m' \geq m_{2j}, (n-m)^{-1} \sum_{i=m_{2j}+1}^{m'} \sum_{q=0}^{m} \operatorname{Var}(d_q e_{i+q}) > \sqrt{\varepsilon_n} \right\},$$

$$m_{2j+2} = m_{2j+1} + \left\lfloor \frac{1}{\varepsilon_n} \right\rfloor + m.$$

Denote

$$I_j = \{k : m_{2j} < k \leq m_{2j+1}, j = 0, \ldots, l\} \quad \text{and}$$

$$J_j = \{k : m_{2j+1} < k \leq m_{2(j+1)}, j = 0, \ldots, l\},$$

where $l = l(n)$ is the number of blocks of indices I_j. Then

$$l\sqrt{\varepsilon_n} \le (n-m)^{-1}\sum_{j=1}^{l}\sum_{i\in I_j}\sum_{q=0}^{m}\operatorname{Var}(d_q e_{i+q}) \le (n-m)^{-1}\sum_{i=1}^{n-m}\sum_{q=0}^{m}E(d_q e_{i+q})^2 \le C_3, \tag{19}$$

and hence we have $l \le C_3/\sqrt{\varepsilon_n}$. If the number of the remainder term is not zero when the construction ends, then we put all the remainder terms into a block denoted by J_l. By (7), we have

$$\Sigma_n(\hat{\beta}_n - \beta_0) = \sum_{i=1}^{n-m}\tilde{x}_i\tilde{e}_i. \tag{20}$$

Then to prove (17), it is enough to prove that

$$(n-m)^{-1/2}\tau_\beta^{-1}\sum_{i=1}^{n-m}\tilde{x}_i\tilde{e}_i \xrightarrow{D} N(0, I_d). \tag{21}$$

Let u be an arbitrary d-dimensional column vector with $\|u\| = 1$, and set $a_i = u^T\tau_\beta^{-1}\tilde{x}_i$. Then, by the Cramér–Wold device, to prove (21) it suffices to prove that

$$\frac{1}{\sqrt{n-m}}\sum_{i=1}^{n-m}a_i\tilde{e}_i \xrightarrow{D} N(0,1). \tag{22}$$

Write

$$\begin{aligned}\frac{1}{\sqrt{n-m}}\sum_{i=1}^{n-m}a_i\tilde{e}_i &= \frac{1}{\sqrt{n-m}}\sum_{j=1}^{l}\sum_{i\in I_j}a_i\tilde{e}_i + \frac{1}{\sqrt{n-m}}\sum_{j=1}^{l}\sum_{i\in J_j}a_i\tilde{e}_i \\ &:= I + J.\end{aligned}$$

Moreover, note that $\max_{0\le q\le m}|d_q| \le 1$ and $\max_{1\le i\le n}|a_i| < \infty$ by Condition (C1), then applying Lemma 3.3 with $p = 2$ we have

$$\begin{aligned}&E\left(\frac{1}{\sqrt{n-m}}\sum_{j=1}^{l}\sum_{i\in J_j}a_i\tilde{e}_i\right)^2 \\ &\quad= \frac{1}{n-m}E\left(\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}a_i d_q e_{i+q}\right)^2 \\ &\quad\le \frac{4m}{n-m}\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}E|a_i d_q e_{i+q}|^2 \\ &\quad\le \frac{4m}{n-m}\Bigl(\max_{m_1\le i\le m_{2l+2}}a_i^2\Bigr)\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}E|d_q e_{i+q}|^2 \\ &\quad\le \frac{4m}{n-m}\Bigl(\max_{m_1\le i\le m_{2l+2}}a_i^2\Bigr)\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}E|d_q e_{i+q}|^2 I\bigl\{|d_q e_{i+q}| \ge \sqrt{n-m}\,\varepsilon_n\bigr\}\end{aligned}$$

$$
\begin{aligned}
&+\frac{4m}{n-m}\Big(\max_{m_1\le i\le m_{2l+2}} a_i^2\Big)\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}E|d_q e_{i+q}|^2 I\big\{|d_q e_{i+q}|<\sqrt{n-m}\varepsilon_n\big\}\\
&\le\frac{4m}{n-m}\Big(\max_{m_1\le i\le m_{2l+2}} a_i^2\Big)\sum_{j=1}^{l}\sum_{i\in J_j}\sum_{q=0}^{m}E|d_q e_{i+q}|^2 I\big\{|d_q e_{i+q}|\ge\sqrt{n-m}\varepsilon_n\big\}\\
&+\frac{4m}{n-m}\Big(\max_{m_1\le i\le m_{2l+2}} a_i^2\Big)l\big(\lfloor\varepsilon_n^{-1}\rfloor+m\big)(n-m)\varepsilon_n^2\\
&\le\frac{4m}{n-m}\Big(\max_{m_1\le i\le m_{2l+2}} a_i^2\Big)\sum_{i=1}^{n-m}\sum_{q=0}^{m}E(d_q e_{i+q})^2 I\big\{|d_q e_{i+q}|\ge\sqrt{n-m}\varepsilon_n\big\}\\
&+4m\Big(\max_{m_1\le i\le m_{2l+2}} a_i^2\Big)C_3{\varepsilon_n}^{-1/2}\big(\lfloor\varepsilon_n^{-1}\rfloor+m\big)\varepsilon_n^2\\
&\to 0\quad\text{as } n\to\infty,
\end{aligned}
\tag{23}
$$

which follows from

$$J\xrightarrow{P}0\quad\text{as } n\to\infty$$

by the Markov inequality. Therefore, to prove (22), it suffices to show that

$$\frac{1}{\sqrt{n-m}}\sum_{j=1}^{l}\sum_{i\in I_j}a_i\tilde{e}_i\xrightarrow{D}N(0,1). \tag{24}$$

On the one hand, by the definition of τ_β^2, it is easy to show that

$$\lim_{n\to\infty}\operatorname{Var}\Bigg(\frac{1}{\sqrt{n-m}}\sum_{i=1}^{n-m}a_i\tilde{e}_i\Bigg)=1.$$

Therefore by the above formula and (23),

$$\lim_{n\to\infty}\operatorname{Var}\Bigg(\frac{1}{\sqrt{n-m}}\sum_{j=1}^{l}\sum_{i\in I_j}a_i\tilde{e}_i\Bigg)=1. \tag{25}$$

On the other hand, by Lemma 3.5 and (ii), we have

$$
\begin{aligned}
&\Bigg|E\exp\Bigg(i\sum_{j=1}^{l}\sum_{i\in I_j}t_i\tilde{e}_i\Bigg)-\prod_{j=1}^{l}E\Bigg(\sum_{i\in I_j}\exp(it_i\tilde{e}_i)\Bigg)\Bigg|\\
&\le -t_0^2\sum_{p=1}^{l}\sum_{s=p+1}^{l}\sum_{i\in I_p}\sum_{j\in I_s}\sum_{q1=0}^{m}\sum_{q2=0}^{m}\operatorname{Cov}(e_{i+q1},e_{j+q2})\\
&=-t_0^2\sum_{q1=0}^{m}\sum_{q2=0}^{m}\sum_{i+q1-j-q2\ge\lfloor\frac{1}{\varepsilon_n}\rfloor+m}\operatorname{Cov}(e_{i+q1},e_{j+q2})\\
&\to 0\quad\text{as } n\to\infty,
\end{aligned}
\tag{26}
$$

which implies that the problem now is reduced to study the asymptotic behavior of independent and non-identically distribution random variables $\{\sum_{i\in I_j} a_i\tilde{e}_i\}$.

To complete the proof of (24), it is enough to show that random variables $\{\sum_{i\in I_j} a_i\tilde{e}_i\}$ satisfies the condition of Lemma 3.7. Set

$$C_4 = \max_{1\le i\le m_{2l+2}} |a_i|^{2+\delta} \quad \text{and} \quad \tau_n^2 = \operatorname{Var}\left(\frac{1}{\sqrt{n-m}}\sum_{i=1}^{n-m} a_i\tilde{e}_i\right).$$

By the definition of I_j,

$$\begin{aligned}
&(n-m)^{-1}\sum_{i\in I_j}\sum_{q=0}^{m} E(d_q e_{i+q})^2 \\
&= (n-m)^{-1}\sum_{m_{2j}}^{m_{2j+1}}\sum_{q=0}^{m} E(d_q e_{i+q})^2 \\
&= (n-m)^{-1}\sum_{m_{2j}}^{m_{2j+1}-1}\sum_{q=0}^{m} E(d_q e_{i+q})^2 + (n-m)^{-1}\sum_{d=0}^{m} E(d_q e_{m_{2j+1}+q})^2 \\
&\le \sqrt{\varepsilon_n} + (n-m)^{-1}\sum_{q=0}^{m} E(d_q e_{m_{2j+1}+q})^2 \\
&\le \sqrt{\varepsilon_n} + (n-m)^{-1}\sup_{n\ge 1} Ee_n^2. \qquad (27)
\end{aligned}$$

By Lemma 3.3 with $p = 2+\delta$ and (27), and recalling that $l \le C_3/\sqrt{\varepsilon_n}$,

$$\begin{aligned}
&\tau_n^{-(2+\delta)}\sum_{j=1}^{l} E\left|(n-m)^{-1/2}\sum_{i\in I_j} a_i\tilde{e}_i\right|^{2+\delta} \\
&= \tau_n^{-(2+\delta)}(n-m)^{-(2+\delta)/2}\sum_{j=1}^{l} E\left|\sum_{i\in I_j}\sum_{q=0}^{m} a_i d_q e_{i+q}\right|^{2+\delta} \\
&\le \tau_n^{-(2+\delta)}(n-m)^{-(2+\delta)/2} C_4 2^{\delta+3} m^{\delta+1}\left(\frac{15(2+\delta)}{\ln(2+\delta)}\right)^{2+\delta}\sum_{j=1}^{l}\sum_{i\in I_j}\sum_{q=0}^{m} E|d_q e_{i+q}|^{2+\delta} \\
&\quad + \tau_n^{-(2+\delta)} C_4 2^{\delta+3} m^{\delta+1}\left(\frac{15(2+\delta)}{\ln(2+\delta)}\right)^{2+\delta}\sum_{j=1}^{l}\left\{(n-m)^{-1}\sum_{i\in I_j}\sum_{q=0}^{m} E(d_q e_{i+q})^2\right\}^{(2+\delta)/2} \\
&\le \tau_n^{-(2+\delta)}(n-m)^{-\delta/2} C_4 2^{\delta+3} m^{\delta+2}\left(\frac{15(2+\delta)}{\ln(2+\delta)}\right)^{2+\delta}\sup_{n\ge 1} E|e_n|^{2+\delta} \\
&\quad + \tau_n^{-(2+\delta)} C_4 2^{\delta+3} m^{\delta+1}\left(\frac{15(2+\delta)}{\ln(2+\delta)}\right)^{2+\delta}\cdot C_3\varepsilon_n^{-1/2}\left\{\sqrt{\varepsilon_n} + (n-m)^{-1}\sup_{n\ge 1} Ee_n^2\right\}^{(2+\delta)/2} \\
&\to 0, \qquad (28)
\end{aligned}$$

since $\tau_n \to 1$ and (i).

Hence, by Lemma 3.7, (24) holds and the proof is completed. □

Corollary 4.1 *Under Condition* (C1), *let* $\{e_n, n \geq 1\}$ *be a sequence of independent random variables with* $Ee_n = 0$, *and suppose that* (i) *of Theorem* 4.1 *holds and* $Ee_n^2 = \sigma^2$ *for all* $n \geq 1$. *Then*

$$(n-m)^{-\frac{1}{2}} \tau_\beta^{-1} \Sigma_n (\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I_d),$$

provided that

$$\tau_\beta^2 = \lim_{n\to\infty} (n-m)^{-1} \left\{ \sum_{i=1}^{n-m} \tilde{x}_i \tilde{x}_i^T \sigma^2 + 2 \sum_{k=1}^{m} \sum_{i=1}^{n-m-k} \tilde{x}_i \tilde{x}_{i+k}^T (d_0 d_k + d_1 d_k + \cdots + d_{m-k} d_m) \sigma^2 \right\}$$

is a positive definite matrix.

Proof Since $\{e_n, n \geq 1\}$ is a sequence of independent random variables, we have $\mathrm{Cov}(e_i, e_j) = 0$ if $i \neq j$ and hence $\mathrm{Cov}(\tilde{e}_i, \tilde{e}_j) = 0$ if $|i-j| > m$. It follows that

$$\begin{aligned} \tau_\beta^2 &= \lim_{n\to\infty} (n-m)^{-1} \left\{ \sum_{i=1}^{n-m} \tilde{x}_i \tilde{x}_i^T \mathrm{Var}(\tilde{e}_i) + \sum_{i=1}^{n-m} \sum_{j=1, j\neq i}^{n-m} \tilde{x}_i \tilde{x}_j^T \mathrm{Cov}(\tilde{e}_i, \tilde{e}_j) \right\} \\ &= \lim_{n\to\infty} (n-m)^{-1} \left\{ \sum_{i=1}^{n-m} \tilde{x}_i \tilde{x}_i^T \sigma^2 + 2 \sum_{k=1}^{m} \sum_{i=1}^{n-m-k} \tilde{x}_i \tilde{x}_{i+k}^T (d_0 d_k + d_1 d_{k+1} + \cdots \right. \\ &\quad \left. + d_{m-k} d_m) \sigma^2 \right\} \end{aligned} \tag{29}$$

from the conditions of Corollary 4.1, we see that τ_β^2 is a positive definite matrix. Thus the result follows from (29). □

Theorem 4.2 *Assume the conditions of Theorem* 4.1, *and further assume that Conditions* (C2)–(C5) *hold. Then*

$$\sup_{0\leq t\leq 1} \left| \hat{f}_n(t) - f(t) \right| = O_P\left(n^{-\gamma}\right) + O_P(\tau_{\tilde{m}}) + O_P\left(n^{-1/3} M_n\right) \quad \text{as } n \to \infty, \tag{30}$$

where $M_n \to \infty$ *in arbitrary slowly rate, and* $\tau_{\tilde{m}} = 2^{-\tilde{m}(\alpha-1/2)}$ *if* $1/2 < \alpha < 3/2$, $\tau_{\tilde{m}} = \sqrt{\tilde{m}} 2^{-\tilde{m}}$ *if* $\alpha = 3/2$, *and* $\tau_{\tilde{m}} = 2^{-\tilde{m}}$ *if* $\alpha > 3/2$.

Proof We can prove Theorem 4.2 by a similar argument to Theorem 3.2 of Hu et al. [12], so we omit the detail. □

Theorem 4.3 *Under the Conditions of Theorem* 4.2, *we have*

$$\frac{\hat{f}_n(t) - f(t)}{\tau_t} \xrightarrow{D} N(0,1), \tag{31}$$

where $\tau_t^2 = \mathrm{Var}(\sum_{i=1}^n e_i \int_{A_i} E_{\tilde{m}}(t,s)\, ds)$.

Proof Note

$$\begin{aligned}\hat{f}_n(t)-f(t) &= \sum_{i=1}^{n}\left(y_i-x_i^T\hat{\beta}_n\right)\int_{A_i}E_{\tilde{m}}(t,s)\,ds-f(t)\\ &= \sum_{i=1}^{n}\left(x_i^T\beta+f(t_i)+e_i-x_i^T\hat{\beta}_n\right)\int_{A_i}E_{\tilde{m}}(t,s)\,ds-f(t)\\ &= \sum_{i=1}^{n}x_i^T(\beta-\hat{\beta}_n)\int_{A_i}E_{\tilde{m}}(t,s)\,ds\\ &\quad+\left\{\sum_{i=1}^{n}f(t_i)\int_{A_i}E_{\tilde{m}}(t,s)\,ds-f(t)\right\}+\sum_{i=1}^{n}e_i\int_{A_i}E_{\tilde{m}}(t,s)\,ds\\ &:= I_1+I_2+I_3, \end{aligned} \tag{32}$$

from the proof of Theorem 3.2 in Hu et al. [12], we get $I_1 = O_P(n^{-1/2})$, $I_2 = O_P(n^{-\gamma}) + O_P(\tau_{\tilde{m}})$ and $I_3 = O_P(n^{-1/3}M_n)$, and it implies that

$$I_1 = o_P(I_3)$$

and

$$I_2 = o_P(I_3).$$

Then we should prove

$$\frac{I_3}{\tau_t} = \frac{\sum_{i=1}^{n} e_i\int_{A_i}E_{\tilde{m}}(t,s)\,ds}{\sqrt{\operatorname{Var}(\sum_{i=1}^{n} e_i\int_{A_i}E_{\tilde{m}}(t,s)\,ds)}} \xrightarrow{D} N(0,1). \tag{33}$$

Let $a_{ni} = \tau_t^{-1}\int_{A_i}E_{\tilde{m}}(t,s)\,ds$, then, by Lemma 3.6 and (C5), $\max_{1\le i\le n}|a_{ni}| \to 0$, and $\sum_{i=1}^{n} a_{ni}^2 = O(1)$, and condition (i) implies that $\{e_n, n \ge 1\}$ is a uniformly integral family on L_2, then, by Lemma 3.8 and (ii), we have

$$\tau_t^{-1}\left(\hat{f}_n(t)-f(t)\right) \xrightarrow{D} N(0,1). \tag{34}$$

The proof is completed. □

5 A simulation example

In this section, we perform a simulation example to verify the accuracy of Theorem 4.1 and Theorem 4.3. Consider the partially linear model

$$y_i = x_i\beta + f(t_i) + e_i, \quad i = 1,2,\ldots,n,$$

where $x_i = \cos(2\pi t_i)$, $f(t_i) = \sin(2\pi t_i)$, $\beta_0 = 5$, $t_i = i/n$, e_i is NSD sequence and raised as follows.

Let $\{e_1, e_2, \ldots, e_n\}$ be a sequence of independent and identically distributed random variables with common probability mass function $P(e_1 = 0) = 2P(e_1 = 1) = P(e_1 = 2) = 0.4$. Then $\{e_1, e_2, \ldots, e_n\}$ given $S_n = n$ is NSD by Theorem 3.1 in Hu [25], where $S_n = \sum_{i=1}^{n} e_i$.

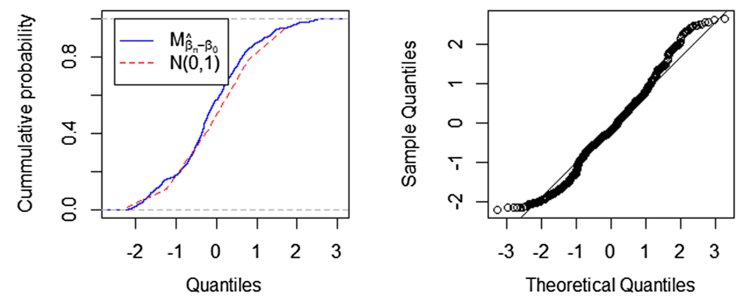

Figure 1 A comparison fitted distribution functions of $M_{\hat{\beta}_n-\beta_0}$ and $N(0,1)$, and QQ-plot of $M_{\hat{\beta}_n-\beta_0}$, where $n = 64$

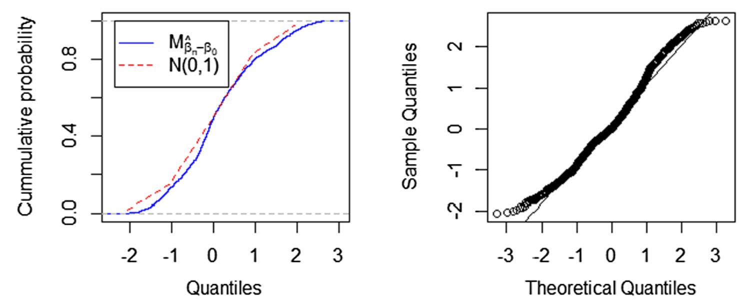

Figure 2 A comparison fitted distribution functions of $M_{\hat{\beta}_n-\beta_0}$ and $N(0,1)$, and QQ-plot of $M_{\hat{\beta}_n-\beta_0}$, where $n = 128$

Set $m = 3$ and the difference sequence $d_0 = \sqrt{3/4}, d_1 = d_2 = d_3 = -\sqrt{1/12}$ (Wang et al. [9]). We first evaluate the $M_{\hat{\beta}_n-\beta_0} = (n-m)^{-\frac{1}{2}}\tau_\beta^{-1}\Sigma_n(\hat{\beta}_n - \beta_0)$ approximation. Figures 1 and 2 show the results for two sample size specifications ($n = 64, n = 128$). Panel 1 in Fig. 1 compares the empirical distribution functions of $M_{\hat{\beta}_n-\beta_0}$ and $N(0,1)$. Panel 2 in Fig. 1 gives the QQ-plot of $M_{\hat{\beta}_n-\beta_0}$. Figure 1 shows that the distribution of $M_{\hat{\beta}_n-\beta_0}$ can approximate $N(0,1)$ well even if the sample size are not large ($n = 64$). Comparison of Fig. 2 with Fig. 1 indicates that the distribution approximation for the larger sample size is much more accurate than that for the small one.

Choose the Daubechies scaling function ${}_2\phi(t)$ as in Hu et al. [29]. Figures 3 and 4 show that the distribution of $M_{\hat{f}_n-f} = \tau_t^{-1}(\hat{f}_n(t) - f(t))$ is closer and closer to $N(0,1)$ with the increasing sample size.

6 Conclusions

In this paper, we use a difference-based and ordinary least square (DOLS) method to obtain the estimator of the unknown parametric component β of the partial linear model with dependent errors. In addition, we investigate the asymptotic normality for the DOLS estimator of β and wavelet estimator of $f(\cdot)$. Thus, we extend some results of Hu et al. [12] to the partially linear model with NSD errors. Furthermore, NSD random variables contain negatively associated random variables. Therefore, it is an interesting subject to investigate the limit properties of the difference-based estimator for a partially linear model with NSD errors in future studies.

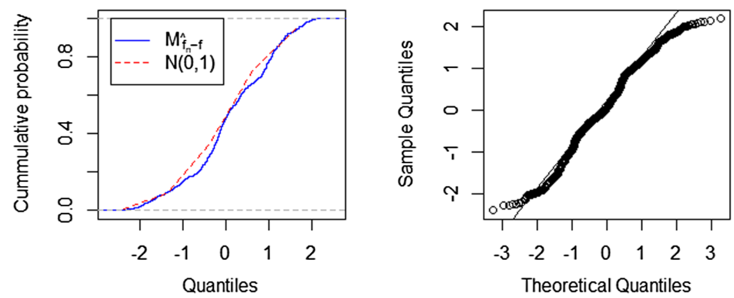

Figure 3 A comparison fitted distribution functions of $M_{\hat{f}_n - f}$ and $N(0,1)$, and QQ-plot of $M_{\hat{f}_n - f}$, where $n = 64$

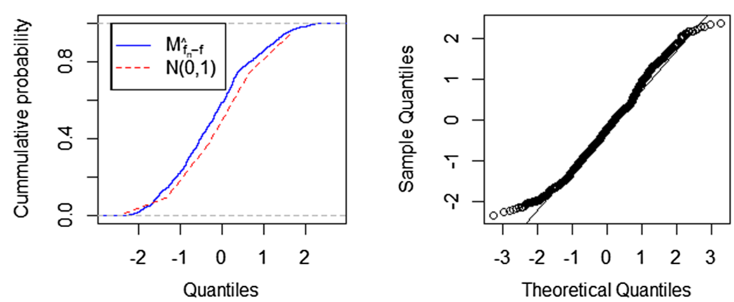

Figure 4 A comparison fitted distribution functions of $M_{\hat{f}_n - f}$ and $N(0,1)$, and QQ-plot of $M_{\hat{f}_n - f}$, where $n = 128$

Funding
The research is supported by National Natural Science Foundation of China [grant number 71471075].

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

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

Received: 10 June 2018 Accepted: 18 September 2018 Published online: 01 October 2018

References

1. Engle, R.F., Granger, C.W.J., Rice, J., Weiss, A.: Semiparametric estimates of the relation between weather and electricity sales. J. Am. Stat. Assoc. **81**, 310–320 (1986)
2. Silvapullé, M.J.: Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers. Ann. Stat. **13**(4), 1490–1497 (1985)
3. Kim, S., Cho, H.R.: Efficient estimation in the partially linear quantile regression model for longitudinal data. Electron. J. Stat. **12**(1), 824–850 (2018)
4. Härdle, W., Gao, J., Liang, H.: Partially Linear Models. Springer, New York (2000)
5. Chang, X.W., Qu, L.: Wavelet estimation of partially linear models. Comput. Stat. Data Anal. **47**, 31–48 (2004)
6. Liang, H., Wang, S.J., Carroll, R.J.: Partially linear models with missing response variables and error-prone covariates. Biometrika **94**(1), 185–198 (2007)
7. Holland, A.D.: Penalized spline estimation in the partially linear model. J. Multivar. Anal. **153**, 211–235 (2017)
8. Zhao, T., Cheng, G., Liu, H.: A partially linear framework for massive heterogeneous data. Ann. Stat. **44**(4), 1400–1437 (2016)
9. Wang, L., Brown, L.D., Cai, T.T.: A difference based approach to the semiparametric partial linear model. Electron. J. Stat. **5**, 619–641 (2011)
10. Tabakan, G., Akdeniz, F.: Difference-based ridge estimator of parameters in partial linear model. Stat. Pap. **51**, 357–368 (2010)
11. Duran, E.A., Hädle, W.K., Osipenko, M.: Difference based ridge and Liu type estimators in semiparametric regression models. J. Multivar. Anal. **105**(1), 164–175 (2012)
12. Hu, H.C., Yang, Y., Pan, X.: Asymptotic normality of DHD estimators in a partially linear model. Stat. Pap. **57**(3), 567–587 (2016)

13. Wu, J.: Restricted difference-based Liu estimator in partially linear model. J. Comput. Appl. Math. **300**, 97–102 (2016)
14. Shen, Y., Wang, X.J., Yang, W.Z., Hu, S.H.: Almost sure convergence theorem and strong stability for weighted sums of NSD random variables. Acta Math. Sin. Engl. Ser. **29**(4), 743–756 (2013)
15. Xue, Z., Zhang, L.L., Lei, Y.J., Chen, Z.J.: Complete moment convergence for weighted sums of negatively superadditive dependent random variables. J. Inequal. Appl. **2015**, Article ID 117 (2015)
16. Shen, Y., Wang, X.J., Hu, S.H.: On the strong convergence and some inequalities for negatively superadditive dependent sequences. J. Inequal. Appl. **2013**, Article ID 448 (2013)
17. Wang, X.J., Deng, X., Zheng, L.L., Hu, S.H.: Complete convergence for arrays of rowwise negatively superadditive dependent random variables and its applications. Statistics **48**(4), 834–850 (2014)
18. Wang, X.J., Shen, A.T., Chen, Z.Y., Hu, S.H.: Complete convergence for weighted sums of NSD random variables and its application in the EV regression model. Test **24**, 166–184 (2015)
19. Meng, B., Wang, D., Wu, Q.: On the strong convergence for weighted sums of negatively superadditive dependent random variables. J. Inequal. Appl. **2017**, Article ID 269 (2017)
20. Shen, A.T., Wang, X.H.: Kaplan–Meier estimator and hazard estimator for censored negatively superadditive dependent data. Statistics **50**(2), 377–388 (2016)
21. Shen, A.T., Xue, M.X., Volodin, A.: Complete moment convergence for arrays of rowwise NSD random variables. Stochastics **88**(4), 606–621 (2016)
22. Wang, X.J., Wu, Y., Hu, S.H.: Strong and weak consistency of LS estimators in the EV regression model with negatively superadditive-dependent errors. AStA Adv. Stat. Anal. **102**, 41–65 (2018)
23. Wang, X.J., Wu, Y., Hu, S.H.: Complete moment convergence for double-indexed randomly weighted sums and its applications. Statistics **52**(3), 503–518 (2018)
24. Kemperman, J.H.B.: On the FKG-inequalities for measures on a partially ordered space. Proc. Akad. Wet., Ser. A **80**, 313–331 (1977)
25. Hu, T.Z.: Negatively superadditive dependence of random variables with applications. Chinese J. Appl. Probab. Statist. **16**(2), 133–144 (2000)
26. Gannaz, I.: Robust estimation and wavelet thresholding in partially linear models. Stat. Comput. **17**, 293–310 (2007)
27. Hu, H.C., Wu, L.: Convergence rates of wavelet estimators in semiparametric regression models under NA samples. Chin. Ann. Math. **33**(4), 609–624 (2012)
28. Christophe, C., Isha, D., Hassan, D.: Nonparametric estimation of a quantile density function by wavelet methods. Comput. Stat. Data Anal. **94**, 161–174 (2016)
29. Hu, H.C., Cui, H.J., Li, K.C.: Asymptotic properties of wavelet estimators in partially linear errors-in-variables models with long memory errors. Acta Math. Appl. Sin. Engl. Ser. **34**(1), 77–96 (2018)
30. Yatchew, A.: An elementary estimator for the partial linear model. Econ. Lett. **5**, 135–143 (1997)
31. Zhao, H., You, J.: Difference based estimation for partially linear regression models with measurement errors. J. Multivar. Anal. **102**, 1321–1338 (2011)
32. Lehmann, E.L.: Some concepts of dependence. Ann. Math. Stat. **37**, 1137–1153 (1966)
33. Rao, B.L.S.P.: Asymptotic Theory of Statistical Inference. Wiley, New York (1987)
34. Yu, Y.C., Hu, H.C., Liu, L., Hung, S.Y.: M-test in linear models with negatively superadditive dependent errors. J. Inequal. Appl. **2017**, Article ID 235 (2017)