RESEARCH   **Open Access**

Strong consistency of estimators in partially linear models for longitudinal data with mixing-dependent structure

Xing-cai Zhou[1,2] and Jin-guan Lin[1*]

* Correspondence: jglin@seu.edu.cn
[1]Department of Mathematics, Southeast University, Nanjing 210096, People's Republic of China
Full list of author information is available at the end of the article

Abstract

For exhibiting dependence among the observations within the same subject, the paper considers the estimation problems of partially linear models for longitudinal data with the ϕ-mixing and ρ-mixing error structures, respectively. The strong consistency for least squares estimator of parametric component is studied. In addition, the strong consistency and uniform consistency for the estimator of nonparametric function are investigated under some mild conditions.

Keywords: partially linear model, longitudinal data, mixing dependent, strong consistency

1 Introduction

Longitudinal data (Diggle et al. [1]) are characterized by repeated observations over time on the same set of individuals. They are common in medical and epidemiological studies. Examples of such data can be easily found in clinical trials and follow-up studies for monitoring disease progression. Interest of the study is often focused on evaluating the effects of time and covariates on the outcome variables. Let t_{ij} be the time of the jth measurement of the ith subject, $x_{ij} \in R^p$ and y_{ij} be the ith subject's observed covariate and outcome at time t_{ij} respectively. We assume that the full data-set $\{(x_{ij}, y_{ij}, t_{ij}), i = 1,..., n, j = 1,..., m_i\}$, where n is the number of subjects and m_i is the number of repeated measurements of the ith subject, is observed and can be modeled as the following partially linear models

$$y_{ij} = x_{ij}^T \beta + g(t_{ij}) + e_{ij}, \tag{1.1}$$

where β is a $p \times 1$ vector of unknown parameter, $g(\cdot)$ is an unknown smooth function, e_{ij} are random errors with $E(e_{ij}) = 0$. We assume without loss of generality that t_{ij} are all scaled into the interval $I = [0, 1]$. Although the observations, and therefore the e_{ij}, from the different subjects are independent, they can be dependent within each subject.

Partially linear models keep the flexibility of nonparametric models, while maintaining the explanatory power of parametric models (Fan and Li [2]). Many authors have studied the models in the form of (1.1) under some additional assumptions or restrictions. If the nonparametric component $g(\cdot)$ is known or not present in the models,

they become the general linear models with repeated measurements, which were studied under Gaussian errors in a amount of literature. Some works have been integrated into PROC MIXED of the SAS Systems for estimation and inference for such models. If $g(\cdot)$ is unknown but there are no repeated measurements, that is $m_1 = \cdots = m_n = 1$, the models (1.1) are reduced to non-longitudinal partially linear regression models, which were firstly introduced by Engle et al. [3] to study the effect of weather on electricity demand, and further studied by Heckman [4], Speckman [5] and Robinson [6], among others. A recent survey of the estimation and application of the models can be found in the monograph of Häardle et al. [7]. When the random errors of the models (1.1) are independent replicates of a zero mean stationary Gaussian process, Zeger and Diggle [8] obtained estimators of the unknown quantities and analyzed time-trend CD4 cell numbers among HIV sero-converters; Moyeed and Diggle [9] gave the rate of convergence for such estimators; Zhang et al. [10] proposed the maximum penalized Gaussian likelihood estimator. Introducing the counting process technique to the estimation scheme, Fan and Li [2] established asymptotic normality and rate of convergence of the resulting estimators. Under the models (1.1) for panel data with a one-way error structure, You and Zhou [11] and You et al. [12] developed the weighted semiparametric least square estimator and derived asymptotic properties of the estimators. In practice, a great deal of the data in econometrics, engineering and natural sciences occur in the form of time series in which observations are not independent and often exhibit evident dependence. Recently, the non-longitudinal partially linear regression models with complex error structure have attracted increasing attention by statisticians. For example, see Schick [13] with AR(1) errors, Gao and Anh [14] with long-memory errors, Sun et al. [15] with $MA(\infty)$ errors, Baek and Liang [16] and Zhou et al. [17] with negatively associated (NA) errors, and Li and Liu [18], Chen and Cui [19] and Liang and Jing [20] with martingale difference sequence, among others.

For longitudinal data, an inherent characteristic is the dependence among the observations within the same subject. Some authors have not considered the with-subject dependence to study the asymptotic behaviors of estimation in the semipara-metric models with assumption that the m_i are all bounded, see, for example, He et al. [21], Xue and Zhu [22] and the references therein. Li et al. [23] and Bai et al. [24] showed that ignoring the data dependence within each subject causes a loss of efficiency of statistical inference on the parameters of interest. Hu et al. [25] and Wang et al. [26] took into consideration within-subject correlations for analyzing longitudinal data and obtained some asymptotic results based on the assumption that $\max_{1 \leq i \leq n} m_i$ is bounded for all n. Chi and Reinsel [27] considered linear models for longitudinal data that contain both individual random effects components and with-individual errors that follow an (autoregressive) AR(1) time series process and gave some estimation procedures, but they did not investigate asymptotic properties of estimations. In fact, the observed responses within the same subject are correlated and may be represented by a sequence of responses $\{y_{ij}, j \geq 1\}$ for the i-individual with an intrinsic dependence structure, such as mixing conditions. For example, in hydrology, many measures may be represented by a sequence of responses $\{y_{ij}, j \geq 1\}$ for the ith year at t_{ij}, where t_{ij} represents the time elapsed from the beginning of the ith year, and $\{e_{ij}, j \geq 1\}$ are the measurements of the deviation from the mean $\{x_{ij}^T \beta + g(t_{ij}), j \geq 1\}$. It is not reasonable that $E(e_{ij_1} e_{ij_2}) = 0$ for $j_1 \neq j_2$. In practice, $\{e_{ij}, j \geq 1\}$ may be "weak error's structure",

such as mixing-dependent structure. In this paper, we consider the estimation problems for the models (1.1) with the ϕ-mixing and ρ-mixing error structures for exhibiting dependence among the observations within the same subject respectively and are mainly devoted to strong consistency of estimators.

Let $\{X_m, m \geq 1\}$ be a sequence of random variables defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{F}_k^l = \sigma(X_i, k \leq i \leq l)$ be σ-algebra generated by $X_k, \ldots, X_l$, and denote $L^2(\mathcal{F}_k^l)$ be the set of all $\mathcal{F}_k^l$ measurable random variables with second moments.

A sequence of random variables $\{X_m, m \geq 1\}$ is called to be ϕ-mixing if

$$\varphi(m) = \sup_{k \geq 1, A \in \mathcal{F}_1^k, P(A) \neq 0, B \in \mathcal{F}_{k+m}^\infty} |P(B|A) - P(B)| \rightarrow 0, \quad \text{as } m \rightarrow \infty.$$

A sequence of random variables $\{X_m, m \geq 1\}$ is called to be ρ-mixing if maximal correlation coefficient

$$\rho(m) = \sup_{k \geq 1, X \in L^2(\mathcal{F}_1^k), Y \in L^2(\mathcal{F}_{k+m}^\infty)} \frac{|\text{cov}(X, Y)|}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \rightarrow 0, \quad \text{as } m \rightarrow \infty.$$

The concept of mixing sequence is central in many areas of economics, finance and other sciences. A mixing time series can be viewed as a sequence of random variables for which the past and distant future are asymptotically independent. A number of limit theorems for ϕ-mixing and ρ-mixing random variables have been studied by many authors. For example, see Shao [28], Peligrad [29], Utev [30], Kiesel [31], Chen et al. [32] and Zhou [33] for ϕ-mixing; Peligrad [34], Peligrad and Shao [35,36], Shao [37] and Bradley [38] for ρ-mixing. Some limit theories can be found in the monograph of Lin and Lu [39]. Recently, the mixing-dependent error structure has also been used to study the nonparametric and semiparametric regression models, for instance, Roussas [40], Truong [41], Fraiman and Iribarren [42], Roussas and Tran [43], Masry and Fan [44], Aneiros and Quintela [45], and Fan and Yao [46].

The rest of this paper is organized as follows. In Section 2, we give least square estimator (LSE) $\hat{\beta}_n$ of β based on the nonparametric estimator of $g(\cdot)$ under the mixing-dependent error structure and state some main results. Section 3 is devoted to sketches of several technical lemmas and corollaries. The proofs of main results are given in Section 4. We close with concluding remarks in the last section.

2 Estimators and main results

For models (1.1), if β is known to be the true parameter, then by $Ee_{ij} = 0$, we have

$$g(t_{ij}) = E(y_{ij} - x_{ij}^T\beta), \quad 1 \leq i \leq n, \quad 1 \leq j \leq m_i.$$

Hence, a natural nonparametric estimator of $g(\cdot)$ given β is

$$g_n^*(t, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t)(y_{ij} - x_{ij}^T\beta), \tag{2.1}$$

where $W_{nij}(t) = W_{nij}(t, t_{11}, t_{12}, \ldots, t_{nm_n})$ is the weight function defined on I. Now, in order to estimate β, we minimize

$$SS(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(y_{ij} - x_{ij}^T\beta - g_n^*(t_{ij}, \beta)\right)^2.$$

The minimizer to the above equation is found to be

$$\hat{\beta}_n = \left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{x}_{ij} \tilde{x}_{ij}^T \right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{x}_{ij} \tilde{y}_{ij}, \tag{2.2}$$

where $\tilde{x}_{ij} = x_{ij} - \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij}) x_{kl}$ and $\tilde{y}_{ij} = y_{ij} - \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij}) y_{kl}$.

So, a plug-in estimator of the nonparametric component $g(\cdot)$ is given by

$$\hat{g}_n(t) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t)(y_{ij} - x_{ij}^T \hat{\beta}_n). \tag{2.3}$$

In this paper, let $\{e_{ij}, 1 \leq j \leq m_i\}$ be ϕ-mixing or ρ-mixing with $Ee_{ij} = 0$ for each $i(1 \leq i \leq n)$, and $\{e_i, 1 \leq i \leq n\}$ be mutually independent, where $e_i = (e_{i1}, \ldots, e_{im_i})^T$. For each i, denote $\phi_i(\cdot)$ and $\rho_i(\cdot)$ be the ith mixing coefficients of the sequence of ϕ-mixing and ρ-mixing, respectively. Define $S_n^2 = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{x}_{ij} \tilde{x}_{ij}^T$, $\tilde{g}(t) = g(t) - \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t) g(t_{kl})$, denote $I(\cdot)$ be the indicator function, $|| \cdot ||$ be the Euclidean norm, and set $\lfloor z \rfloor \leq z < \lfloor z \rfloor + 1$ for the integer part of z. In the sequence, C and C_1 denote positive constants whose values may vary at each occurrence.

For obtaining our main results, we list some assumptions:

A1 (i) $\{e_{ij}, 1 \leq j \leq m_i\}$ are ϕ-mixing with $Ee_{ij} = 0$ for each i;

(ii) $\{e_{ij}, 1 \leq j \leq m_i\}$ are ρ-mixing with $Ee_{ij} = 0$ for each i.

A2 (i) $\max_{1 \leq i \leq n} m_i = o(n^\delta)$ for some $0 < \delta < \dfrac{r-2}{2r}$ and $r > 2$;

(ii) $\lim_{n \to \infty} \dfrac{1}{N(n)} S_n^2 = \sum$, where $\sum$ is a positive definite matrix and $N(n) = \sum_{i=1}^{n} m_i$

(iii) $g(\cdot)$ satisfies the first-order Lipschitz condition on $[0, 1]$.

A3 For n large enough, the probability weight functions $W_{nij}(\cdot)$ satisfy

(i) $\sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) = 1$ for each $t \in [0, 1]$;

(ii) $\sup_{0 \leq t \leq 1} \max_{1 \leq i \leq n, 1 \leq j \leq m_i} W_{nij}(t) = O\left(n^{-\frac{1}{2}} \right)$;

(iii) $\sup_{0 \leq t \leq 1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) I(|t_{ij} - t| > \varepsilon) = o(1)$ for any $\epsilon > 0$;

(iv) $\max_{1 \leq k \leq n, 1 \leq l \leq m_i} || \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t_{kl}) x_{ij} || = O(1)$,

(v) $\sup_{0 \leq t \leq 1} \left\| \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) x_{ij} \right\| = O(1)$,

(vi) $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left| W_{nij}(s) - W_{nij}(t) \right| \leq C|s - t|$ uniformly for $s, t \in [0, 1]$.

Remark 2.1 For obtaining the asymptotic properties of estimators of the models (1.1), many authors often assumed that $\{m_i, 1 \leq i \leq n\}$ are bounded. Under the weak condition A2(i), we obtain the strong consistency of estimators of the models (1.1) with

mixing-dependent structure. The condition of $\{m_i, 1 \le i \le n\}$ being a bounded sequence is a special case of A2(i).

Remark 2.2 Assumption A2(ii) implies that

$$\frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} ||\tilde{x}_{ij}|| = O(1) \quad \text{and} \quad \max_{1 \le i \le n, 1 \le j \le m_i} ||\tilde{x}_{ij}|| = o\left(N(n)^{\frac{1}{2}}\right).$$

Remark 2.3 As a matter of fact, there exist some weights satisfying assumption A3. For example, under some regularity conditions, the following Nadaraya-Watson kernel weight satisfies assumption A3:

$$W_{nij}(t) = K\left(\frac{t - t_{ij}}{h_n}\right) \left[\sum_{k=1}^{n} \sum_{l=1}^{m_i} K\left(\frac{t - t_{kl}}{h_n}\right) \right]^{-1},$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth parameter. Assumption A3 has also been used by Hardle et al. [7], Baek and Liang [16], Liang and Jing [20] and Chen and You [47].

Theorem 2.1 *Suppose that A1(i) or A1(ii), and A2 and A3(i)-(iii) hold. If*

$$\max_{1 \le i \le n, 1 \le j \le m_i} E(|e_{ij}|^p) \le C, \quad \text{a.s.} \tag{2.4}$$

for $p > 3$, then

$$\hat{\beta}_n \to \beta, \quad \text{a.s.} \tag{2.5}$$

Theorem 2.2 *Suppose that A1(i) or A1(ii), and A2, A3(i-iv) and (2.4) hold. For any $t \in [0, 1]$, we have*

$$\hat{g}_n(t) \to g(t), \quad \text{a.s.} \tag{2.6}$$

Theorem 2.3 *Suppose that A1(i) or A1(ii), and A2, A3(i-iii), A3(v-vi) and (2.4) hold. We have*

$$\sup_{0 \le t \le 1} |\hat{g}_n(t) - g(t)| = o(1), \quad \text{a.s.} \tag{2.7}$$

3 Several technical lemmas and corollaries

In order to prove the main results, we first introduce some lemmas and corollaries. Let $S_j = \sum_{l=1}^{j} X_l$ for $j \ge 1$, and $S_k(i) = \sum_{j=k+1}^{k+i} X_j$ for $i \ge 1$ and $k \ge 0$.

Lemma 3.1. (**Shao** [28]) *Let $\{X_m, m \ge 1\}$ be a ϕ-mixing sequence.*

(1) If $EX_i = 0$, then

$$ES_k^2(i) \le 8000i \exp\left\{ 6 \sum_{j=1}^{\lfloor \log i \rfloor} \varphi^{1/2}(2^j) \right\} \max_{k+1 \le j \le k+i} EX_j^2.$$

(2) Suppose that there exists an array $\{c_{km}\}$ of positive numbers such that $\max_{1 \le i \le m} ES_k^2(i) \le c_{km}$ for every $k \ge 0$, $m \ge 1$. Then, for any $q \ge 2$, there exists a positive constant $C = C(q, \phi(\cdot))$ such that

$$E \max_{1 \le i \le m} |S_k(i)|^q \le C \left(c_{km}^{q/2} + E \max_{k<i\le k+m} |X_i|^q \right).$$

Lemma 3.2. (**Shao** [37]) *Let* $\{X_m, m \ge 1\}$ *be a* ρ-*mixing sequence with* $EY_i = 0$. *Then, for any* $q \ge 2$, *there exists a positive constant* $C = C(q, \rho(\cdot))$ *such that*

$$E \max_{1 \le j \le m} |S_j|^q \le C \left(m^{q/2} \exp \left\{ C \sum_{j=1}^{\lfloor \log m \rfloor} \rho(2^j) \right\} \max_{1 \le j \le m} (E|X_j|^2)^{q/2} \right.$$
$$\left. + m \exp \left\{ C \sum_{j=1}^{\lfloor \log m \rfloor} \rho^{2/q}(2^j) \right\} \max_{1 \le j \le m} E|Y_j|^q \theta \right).$$

Lemma 3.3. *Suppose that A1(i) or A1(ii) holds. Let* $\alpha > 1, 0 < r < \alpha$ *and*

$$e'_{ij} = e_{ij} I \left(|e_{ij}| \le \varepsilon i^{\frac{1}{r}} m_i \right), \tag{3.1}$$

$$e''_{ij} = e_{ij} - e'_{ij} = e_{ij} I \left(e_{ij} > \varepsilon i^{\frac{1}{r}} m_i \right) + e_{ij} I \left(e_{ij} < -\varepsilon i^{\frac{1}{r}} m_i \right) \tag{3.2}$$

for any $\varepsilon > 0$. *If*

$$\max_{1 \le i \le n} \max_{1 \le j \le m_i} E(|e_{ij}|^\alpha) \le C, \quad \text{a.s.,} \tag{3.3}$$

we have

$$\sum_{i=1}^{\infty} \sum_{j=1}^{m_i} |e''_{ij}| < \infty, \quad \text{a.s..}$$

Proof Note that $|e''_{ij}| = |e_{ij}| I \left(|e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i \right)$. Let

$\xi_i = \sum_{j=1}^{m_i} |e_{ij}|, \xi'_i = \sum_{j=1}^{m_i} |e_{ij}| \cdot I \left(\sum_{j=1}^{m_i} |e_{ij}| \le \varepsilon i^{\frac{1}{r}} m_i \right), \xi''_i = \xi_i - \xi'_i = \sum_{j=1}^{m_i} |e_{ij}| I \left(\sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i \right),$ and $|\xi''_i|_d = |\xi''_i| I (|\xi''_i| \le d)$ for fixed $d > 0$. First, we prove

$$\sum_{i=1}^{\infty} |\xi''_i| < \infty, \quad \text{a.s..} \tag{3.4}$$

Note that

$$\{|\xi''_i| > d\} = \left\{ \sum_{j=1}^{m_i} |e_{ij}| I \left(\sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i \right) > d \right\} = \left\{ \sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i \right\} \tag{3.5}$$

for i large enough. By Markov's inequality, C_r-inequality, and (3.3), we have

$$
\begin{aligned}
\sum_{i=1}^{\infty} P(|\xi''_i| \, d) &\leq C \sum_{i=1}^{\infty} P\left(\sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i\right) \\
&\leq C \sum_{i=1}^{\infty} i^{-\frac{\alpha}{r}} m_i^{-\alpha} E\left|\sum_{j=1}^{m_i} |e_{ij}|\right|^{\alpha} \leq C \sum_{i=1}^{\infty} i^{-\frac{\alpha}{r}} m_i^{-1} \sum_{j=1}^{m_i} E|e_{ij}|^{\alpha} \\
&\leq C \lim_{n \to \infty} \sum_{i=1}^{n} i^{-\frac{\alpha}{r}} \max_{1 \leq i \leq n} \max_{1 \leq j \leq m_i} E|e_{ij}|^{\alpha} \\
&\leq C \sum_{i=1}^{\infty} i^{-\frac{\alpha}{r}} < \infty,
\end{aligned}
\tag{3.6}
$$

From (3.5), $\{|\xi''_i| \leq d\} = \left\{ \sum_{j=1}^{m_i} |e_{ij}| \leq \varepsilon i^{\frac{1}{r}} m_i \right\}$ for i large enough. One gets

$$
\begin{aligned}
E(|\xi''_i| \, d) &= E(|\xi''_i| I(|\xi''_i| \leq d)) \\
&= E\left(\sum_{j=1}^{m_i} |e_{ij}| I\left(\sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i\right) I\left(\sum_{j=1}^{m_i} |e_{ij}| \leq \varepsilon i^{\frac{1}{r}} m_i\right)\right) = 0
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(|\xi''_i| \, d) &\leq E(|\xi''_i|^2_d) = E(|\xi''_i| I(|\xi''_i| \leq d))^2 \\
&= E\left(|\xi''_i|^2 I(|\xi''_i| \leq d)\right) \leq d E(|\xi''_i| I(|\xi''_i| \leq d)) = 0
\end{aligned}
$$

for i large enough. Therefore,

$$
\sum_{i=1}^{\infty} E(|\xi''_i| \, d) < \infty, \quad \sum_{i=1}^{\infty} \mathrm{Var}(|\xi''_i| \, d) < \infty.
\tag{3.7}
$$

Since $\{\xi''_i, 1 \leq i \leq n\}$ is a sequence of independent random variables, (3.4) holds from (3.6) and (3.7) by Three Series Theorem. Then,

$$
\begin{aligned}
\sum_{i=1}^{\infty} \sum_{j=1}^{m_i} |e''_{ij}| &= \sum_{i=1}^{\infty} \sum_{j=1}^{m_i} |e_{ij}| I\left(|e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i\right) \\
&\leq \sum_{i=1}^{\infty} \sum_{j=1}^{m_i} |e_{ij}| I\left(\sum_{j=1}^{m_i} |e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i\right) = \sum_{i=1}^{\infty} |\xi''_i| < \infty, \quad \text{a.s..}
\end{aligned}
$$

Thus, we complete the proof of Lemma 3.3.

Lemma 3.4. *Let $\{e_{ij}, 1 \leq j \leq m_i\}$ be the ϕ-mixing with $Ee_{ij} = 0$ for each i $(1 \leq i \leq n)$. Assume that $\{a_{nij}(\cdot), 1 \leq i \leq n, 1 \leq j \leq m_i\}$ is a function array defined on $[0, 1]$, satisfying $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(t)| = O\left(n^{-\frac{1}{2}}\right)$ and $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(t)| = O\left(n^{-\frac{1}{2}}\right)$ for any $t \in [0, 1]$, and A2(i) and (2.4) hold. Then, for any $t \in [0, 1]$ we have*

$$
\sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) e_{ij} = o(1), \quad \text{a.s..}
\tag{3.8}
$$

Proof Based on (3.1) and (3.2), we denote $\zeta_{nij} = e'_{ij} - E(e'_{ij})$, $\eta_{nij} = e''_{ij} - E(e''_{ij})$ and take r satisfying $2 < r < p - 1$. Since $e_{ij} = \zeta_{nij} + \eta_{nij}$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(t)e_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(t)\zeta_{nij} + \sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(t)e''_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(t)E(e''_{ij}) \tag{3.9}$$

$$=: A_{1n} + A_{2n} - A_{3n}.$$

First, we prove

$$A_{1n} \to 0, \quad \text{a.s..} \tag{3.10}$$

Denoting $\tilde{\zeta}_{ni} = \sum_{j=1}^{m_i} a_{nij}(t)\zeta_{nij}$, we know that $\{\tilde{\zeta}_{ni}, 1 \le i \le n\}$ *is a* sequence of independent random variables with $E\tilde{\zeta}_{ni} = 0$. By Markov's inequality, and Rosenthal's inequality, for any $\varepsilon > 0$ and $q \ge 2$, one gets

$$P\left(\left|\sum_{i=1}^{n}\tilde{\zeta}_{ni}\right| > \varepsilon\right) \le \varepsilon^{-q}E\left|\sum_{i=1}^{n}\tilde{\zeta}_{ni}\right|^{q}$$

$$\le C\left(\sum_{i=1}^{n}E|\tilde{\zeta}_{ni}|^{q} + \left(\sum_{i=1}^{n}E\tilde{\zeta}_{ni}^{2}\right)^{\frac{q}{2}}\right) \tag{3.11}$$

$$=: A_{11n} + A_{12n}.$$

Note that $\phi_i(m) \to 0$ as $m \to \infty$, hence $\sum_{k=1}^{\lfloor \log m_i \rfloor} \varphi_i^{1/2}(2^k) = o(\log m_i)$. Further, $\exp\left\{\lambda \sum_{k=1}^{\lceil \log m_i \rceil} \varphi_i^{1/2}(2^k)\right\} = o(m_i^\tau)$ for any $\lambda > 0$ and $\tau > 0$.

For A_{11n}, by Lemma 3.1, A2(i) and (2.4), and taking $q > p$, we have

$$A_{11n} = C\sum_{i=1}^{n}E\left|\sum_{j=1}^{m_i} a_{nij}(t)\zeta_{nij}\right|^{q}$$

$$\le C\sum_{i=1}^{n}\left[\left(m_i \exp\left\{6\sum_{k=1}^{\lfloor \log m_i \rfloor}\varphi_i^{1/2}(2^k)\right\}\max_{1 \le k \le m_i}E|a_{nik}(t)\zeta_{nik}|^2\right)^{q/2} + \sum_{j=1}^{m_i}E|a_{nij}\zeta_{nij}|^q\right]$$

$$\le C\sum_{i=1}^{n}\left[(m_i^{1+\tau}n^{-1})^{q/2} + \sum_{j=1}^{m_i}n^{-\frac{q}{2}}E|\zeta_{nij}|^p|\zeta_{nij}|^{q-p}\right]$$

$$\le Cn^{-\frac{q}{2}}\sum_{i=1}^{n}m_i^{\frac{(\tau+1)q}{2}} + Cn^{-\frac{q}{2}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(i^{\frac{1}{r}}m_i)^{q-p}$$

$$\le C_n^{-\left(\frac{q}{2}-\frac{(\tau+1)\delta q}{2}-1\right)} + Cn^{-\left(\frac{q}{2}-\frac{q}{r}+\frac{p}{r}-(q-p+1)\delta-1\right)}$$

Take $q > \max\left\{\dfrac{2r(2+\delta)}{r-2r\delta-2}, \dfrac{4}{1-\delta}, p\right\}$. We have $\dfrac{q}{2} - \dfrac{\delta q}{2} > 2$ and $\dfrac{q}{2} - \dfrac{q}{r} + \dfrac{p}{r} - (q-p+1)\delta > 2$.

Next, take $\tau > 0$ small enough such that $\dfrac{q}{2} - \dfrac{(\tau+1)\delta q}{2} > 2$. Thus, we have

$$\sum_{n=1}^{\infty}A_{11n} < \infty. \tag{3.12}$$

For A_{12n}, by Lemma 3.1 and (2.4), we have

$$A_{12n} = C \left\{ \sum_{i=1}^{n} E \left| \sum_{j=1}^{m_i} a_{nij}(t) \zeta_{nij} \right|^2 \right\}^{\frac{q}{2}}$$

$$\leq C \left\{ \sum_{i=1}^{n} m_i \exp \left\{ 6 \sum_{k=1}^{\lfloor \log m_i \rfloor} \varphi_i^{1/2}(2^k) \right\} \max_{1 \leq j \leq m_i} E |a_{nij}(t) \zeta_{nij}|^2 \right\}^{\frac{q}{2}}$$

$$\leq C \left\{ \sum_{i=1}^{n} m_i^{\tau+1} \sum_{j=1}^{m_i} E |a_{nij}(t) \zeta_{nij}|^2 \right\}^{\frac{q}{2}}$$

$$\leq C n^{-\left(\frac{q}{4} - \frac{(\tau+1)\delta q}{2} \right)}$$

Note that $\delta < \frac{r-2}{2r} < \frac{1}{2}$. Taking $q > \frac{4}{1-2\delta}$, we have $\frac{q}{4} - \frac{\delta q}{2} > 1$. Next, take $\tau > 0$ small enough such that $\frac{q}{2} - \frac{(\tau+1)\delta q}{2} > 1$. Thus, we have

$$\sum_{n=1}^{\infty} A_{12n} < \infty. \tag{3.13}$$

Combining (3.11)-(3.13), we obtain (3.10).

By Lemma 3.3 and $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(t)| = O\left(n^{-\frac{1}{2}} \right)$ for any $t \in [0, 1]$, we have

$$|A_{2n}| \leq \max_{i \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(t)| \sum_{i=1}^{n} \sum_{j=1}^{m_i} |e_{ij}''| = O\left(n^{-\frac{1}{2}} \right). \tag{3.14}$$

Note that $\frac{p-1}{r} > 1$ and $\delta > 0$. From (2.4), we have

$$|A_{3n}| = \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) E(e_{ij}'') \right|$$

$$\leq n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} E \left(|e_{ij}| I(|e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i) \right)$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} E \left(|e_{ij}|^p |e_{ij}|^{1-p} I \left(|e_{ij}| > \varepsilon i^{\frac{1}{r}} m_i \right) \right) \tag{3.15}$$

$$\leq C n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(i^{\frac{1}{r}} m_i \right)^{1-p} \leq C n^{-\frac{1}{2}} \sum_{i=1}^{n} i^{-\frac{p-1}{r}} m_i^{2-p}$$

$$\leq C n^{-\left((p-2)\delta + \frac{1}{2} \right)} = o(1).$$

From (3.9), (3.10), (3.14) and (3.15), we have (3.8).

Corollary 3.1. *In Lemma 3.4, if $\{e_{ij}, 1 \leq j \leq m_i\}$ are ρ-mixing with $E e_{ij} = 0$ for each i $(1 \leq i \leq n)$, then (3.8) holds.*

Proof From the proof of Lemma 3.4, it is enough to prove that $\sum_{n=1}^{\infty} A_{11n} < \infty$ and $\sum_{n=1}^{\infty} A_{12n} < \infty$.

Note that $\rho_i(m) \to 0$ as $m \to \infty$, hence $\sum_{k=1}^{\lfloor \log m_i \rfloor} \rho_i^{2/q}(2^k) = o(\log m_i)$. Further, $\exp\left\{\lambda \sum_{k=1}^{\lceil \log m_i \rceil} \rho_i^{2/q}(2^k)\right\} = o(m_i^{\tau})$ for any $\lambda > 0$ and $\tau > 0$.

For A_{11n}, by Lemma 3.2 and (2.4), and taking $q > p$, we get

$$A_{11n} = C \sum_{i=1}^{n} E\left|\sum_{j=1}^{m_i} a_{nij}(t)\zeta_{nij}\right|^q$$

$$\leq C \sum_{i=1}^{n} \left(m_i^{\frac{q}{2}} \exp\left\{C_1 \sum_{k=1}^{\lfloor \log m_i \rfloor} \rho_1(2^k)\right\} \max_{1 \leq k \leq m_i} (E|a_{nik}\zeta_{nik}|^2)^{\frac{q}{2}}\right.$$

$$\left. + m_i \exp\left\{C_1 \sum_{k=1}^{\lfloor \log m_i \rfloor} \rho_i^{2/q}(2^k)\right\} \max_{1 \leq k \leq m_i} E|a_{nik}\zeta_{nik}|^q\right)$$

$$\leq C \sum_{i=1}^{n} \left(m_i^{\tau + \frac{q}{2}} n^{-\frac{q}{2}} + m_i^{\tau+1} n^{-\frac{q}{2}} \left(\frac{1}{i^{\frac{1}{r}} m_i}\right)^{q-p}\right)$$

$$\leq C n^{-\left(\frac{q}{2} - \left(r + \frac{q}{2}\right)\delta - 1\right)} + C n^{-\left(\frac{q}{2} - \frac{q}{r} + \frac{p}{r} - (q+p+r+1)\delta - 1\right)}$$

Take $q > \max\left\{\frac{2r(2+\delta)}{r - 2r\delta - 2}, \frac{4}{1-\delta}, p\right\}$. We have $\frac{q}{2} - \frac{q\delta}{2} > 2$ and $\frac{q}{2} - \frac{q}{r} + \frac{p}{r} - (q - p + 1)\delta > 2$.

Next, take $\tau > 0$ small enough such that $\frac{q}{2} - \left(\tau + \frac{q}{2}\right)\delta > 2$ and $\frac{q}{2} - \frac{q}{r} + \frac{p}{r} - (q + p + \tau + 1)\delta > 2$. Thus, $\sum_{n=1}^{\infty} A_{11n} < \infty$.

For A_{12n}, by Lemma 3.2 and (2.4), we have

$$A_{12n} = C\left\{\sum_{i=1}^{n} E\left|\sum_{j=1}^{m_i} a_{nij}(t)\zeta_{nij}\right|^2\right\}^{\frac{q}{2}}$$

$$\leq C\left(\sum_{i=1}^{n} m_i \exp\left\{C_1 \sum_{k=1}^{\lfloor \log m_i \rfloor} \rho_1(2^k)\right\} \max_{1 \leq j \leq m_i} E|a_{nik}\zeta_{nik}|^2\right)^{\frac{q}{2}}$$

$$\leq C\left(\sum_{i=1}^{n} m_i^{\tau+1} \sum_{j=1}^{m_i} E|a_{nij}\zeta_{nij}|^2\right)^{\frac{q}{2}}$$

$$\leq C n^{-\left(\frac{q}{4} - \frac{(\tau+1)\delta q}{2}\right)}$$

Note that $\delta < \frac{1}{2}$ from A2(i). Taking $q > \frac{4}{1-2\delta}$, we have $\frac{q}{4} - \frac{\delta q}{2} > 1$. Next, take $\tau > 0$ small enough such that $\frac{q}{2} - \frac{(\tau+1)\delta q}{2} > 1$. Thus, $\sum_{n=1}^{\infty} A_{12n} < \infty$.

So, we complete the proof of Lemma 3.4.

Remark 3.1 If the real function array $\{a_{nij}(t), 1 \leq i \leq n, 1 \leq j < m_i\}$ is replaced with the real constant array $\{a_{nij}, 1 \leq i \leq n, 1 \leq j \leq m_i\}$, the results of Lemma 3.4 and Corollary 3.1 hold obviously.

Lemma 3.5. *Let $\{e_{ij}, 1 \leq j \leq m_i\}$ be the ϕ-mixing with $Ee_{ij} = 0$ for each i $(1 \leq i \leq n)$. Assume that $\{a_{nij}(\cdot), 1 \leq i \leq n, 1 \leq j \leq m_i\}$ is a function array defined on $[0, 1]$, satisfying $\sum_{i=1}^{n} \sum_{j=1}^{m_i} |a_{nij}(t)| = O(1)$ and $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(t)| = O\left(n^{-\frac{1}{2}}\right)$ uniformly for $t \in [0, 1]$, and $\max_{1 \leq i \leq n, 1 \leq j \leq m_i} |a_{nij}(s) - a_{nij}(t)| \leq C|s - t|$ uniformly for $s, t \in [0, 1]$, where C is a constant. If A2(i) and (2.4) hold, then*

$$\sup_{0 \leq t \leq 1} \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) e_{ij} \right| = o(1), \quad \text{a.s..} \tag{3.16}$$

Proof Based on (3.1) and (3.2), we denote $\zeta_{nij} = e'_{ij} - Ee'_{ij}$ and take r satisfying $2 < r < p$ - 1. Using the finite covering theorem, $[0, 1]$ is covered by $O\left(n^{2+\frac{1}{r}}\right)$'s neighborhoods D_n with center s_n and radius $n^{-\left(2+\frac{1}{r}\right)}$, and for each $t \in [0, 1]$, there exists some neighborhood $D_n(s_n(t))$ with center $s_n(t)$ and radius $n^{-\left(2+\frac{1}{r}\right)}$ such that $t \in D_n(s_n(t))$. Since $E(e_{ij}) = 0$, we have

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) e_{ij} \right| \leq \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) e''_{ij} \right| + \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} (a_{nij}(t) - a_{nij}(s_n(t))) e'_{ij} \right|$$

$$+ \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(s_n(t)) \zeta_{nij} \right| + \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} (a_{nij}(t) - a_{nij}(s_n(t))) E(e'_{ij}) \right|$$

$$+ \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} a_{nij}(t) E(e''_{ij}) \right|.$$

$$=: B_{1n}(t) + B_{2n}(t) + B_{3n}(t) + B_{4n}(t) + B_{5n}(t).$$

Denote $\sup \max_{t,i,j} = \sup_{0 \leq t \leq 1} \max_{1 \leq i \leq n, 1 \leq j \leq m_i}$. By Lemma 3.3 and the proof of (3.15), noting that $\delta < \frac{1}{2}$, we have

$$\sup_{0 \leq t \leq 1} B_{1n}(t) \leq \sup \max_{t,i,j} |a_{nij}(t)| \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left| e''_{ij} \right| = O\left(n^{-\frac{1}{2}}\right), \quad \text{a.s.,}$$

$$\sup_{0 \leq t \leq 1} B_{2n}(t) \leq \sup \max_{t,i,j} |a_{nij}(t) - a_{nij}(s_n(t))| \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left| e'_{ij} \right| \leq Cn^{-\left(2+\frac{1}{r}\right)} n^{2\delta} n^{1+\frac{1}{r}} = o(1),$$

$$\sup_{0 \leq t \leq 1} B_{4n}(t) \leq \sup \max_{t,i,j} |a_{nij}(t) - a_{nij}(s_n(t))| \sum_{i=1}^{n} \sum_{j=1}^{m_i} E\left(\left| e'_{ij} \right|\right) = o(1),$$

$$\sup_{0 \leq t \leq 1} B_{5n}(t) \leq \sup \max_{t,i,j} |a_{nij}(t)| \sum_{i=1}^{n} \sum_{j=1}^{m_i} E\left(|e_{ij}| I\left(|e_{ij}| > \varepsilon i^{\frac{1}{r}} \right) \right) = o(1).$$

Now, it is enough to show $\sup_{0 \le t \le 1} B_{3n}(t) = o(1)$, a.s..

From (3.11), A_{11n} and A_{12n}, for the given $t \in [0, 1]$ and $u \in D_n(s_n(t))$, we have

$$P\left(\left|\sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(u)\zeta_{nij}\right| > \varepsilon\right) \le C\left(n^{-\left(\frac{q}{2}-\frac{q}{r}+\frac{p}{r}-(q-p+1)\delta-1\right)}_{+} + n^{-\left(\frac{q}{2}-\frac{(r+1)\delta q}{2}-1\right)}_{+} + n^{-\left(\frac{q}{4}-\frac{(r+1)\delta q}{2}\right)}\right).$$

Then, we obtain

$$P\left(\sup_{0 \le t \le 1} B_{3n}(t) > \varepsilon\right) \le P\left(\sup_{0 \le t \le 1}\sup_{u \in D_n(s_n(t))}\left|\sum_{i=1}^{n}\sum_{j=1}^{m_i} a_{nij}(u)\zeta_{nij}\right| > \varepsilon\right)$$

$$\le Cn^{2+\frac{1}{r}}\left(n^{-\left(\frac{q}{2}-\frac{q}{r}+\frac{p}{r}-(q-p+1)\delta-1\right)}_{+} + n^{-\left(\frac{q}{2}-\frac{(r+1)\delta q}{2}-1\right)}_{+} + n^{-\left(\frac{q}{4}-\frac{(r+1)\delta q}{2}\right)}\right)$$

$$\le C\left(n^{-\left(\frac{q}{2}-\frac{q}{r}-\delta d-\delta-4\right)}_{+} + n^{-\left(\frac{q}{2}-\frac{(r+1)\delta q}{2}-4\right)}_{+} + n^{-\left(\frac{q}{4}-\frac{(r+1)\delta q}{2}-3\right)}\right).$$

Take $q > \max\left\{\dfrac{2r(5+\delta)}{r-2r\delta-2}, \dfrac{16}{1-2\delta}, p\right\}$. We have $\dfrac{q}{2}-\dfrac{q}{r}-\delta q-\delta > 5$, $\dfrac{q}{2}-\dfrac{\delta q}{2} > 5$ and $\dfrac{q}{4}-\dfrac{\delta q}{2} > 4$. Next, take $\tau > 0$ small enough such that $\dfrac{q}{2}-\dfrac{(r+1)\delta q}{2} > 5$ and $\dfrac{q}{4}-\dfrac{(r+1)\delta q}{2} > 4$. Thus, we have $\sum_{n=1}^{\infty} P\left(\sup_{0\le t\le1}B_{3n}(t) > \varepsilon\right) < \infty$. Thus, $\sup_{0\le t\le1}B_{3n}(t) = o(1)$,a.s.. Therefore, (3.16) holds.

Corollary 3.2. *In Lemma 3.5, if $\{e_{ij}, 1 \le j \le m_i\}$ are ρ-mixing with $Ee_{ij} = 0$ for each i $(1 \le i \le n)$, then (3.16) holds.*

Proof By Corollary 3.1, with arguments similar to the proof of Lemma 3.5, we have (3.16).

4 Proof of Theorems

Proof of Theorem 2.1 From (1.1) and (2.2), we have

$$\begin{aligned}
\hat{\beta}_n - \beta &= \left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} \tilde{x}_{ij}\tilde{x}_{ij}^T\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}(\tilde{y}_{ij} - \tilde{x}_{ij}^T\beta) \\
&= S_n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}\left[(y_{ij} - x_{ij}^T\beta) - \sum_{k=1}^{n}\sum_{l=1}^{m_i} W_{nkl}(t_{ij})(y_{kl} - x_{kl}^T\beta)\right] \\
&= S_n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}\left[(g(t_{ij}) + e_{ij}) - \sum_{k=1}^{n}\sum_{l=1}^{m_i} W_{nkl}(t_{ij})(g(t_{kl}) + e_{kl})\right] \\
&= S_n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}\left[e_{ij} - \sum_{k=1}^{n}\sum_{l=1}^{m_i} W_{nkl}(t_{ij})e_{kl} + \tilde{g}(t_{ij})\right] \\
&= S_n^{-2}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}e_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}\left(\sum_{k=1}^{n}\sum_{l=1}^{m_i} W_{nkl}(t_{ij})e_{kl}\right) + \sum_{i=1}^{n}\sum_{j=1}^{m_i}\tilde{x}_{ij}\tilde{g}(t_{ij})\right] \\
&= \left(\frac{S_n^2}{N(n)}\right)^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tilde{x}_{ij}}{N(n)}e_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tilde{x}_{ij}}{N(n)}\left(\sum_{k=1}^{n}\sum_{l=1}^{m_i} W_{nkl}(t_{ij})e_{kl}\right) + \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tilde{x}_{ij}}{N(n)}\tilde{g}(t_{ij})\right] \\
&=: D_{1n} + D_{2n} + D_{3n}.
\end{aligned} \tag{4.1}$$

From A2(ii), $\left\| \left\{ \dfrac{S_n^2}{N(n)} \right\}^{-1} \right\| = O(1)$. By Remark 2.2, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{||\tilde{x}_{ij}||}{N(n)} = O(1) \quad \text{and} \quad \max_{1 \leq i \leq n, 1 \leq j \leq m_i} \frac{||\tilde{x}_{ij}||}{N(n)} = o\left(n^{\frac{1}{2}} \right). \tag{4.2}$$

According to (4.2) and Remark 3.1, we have

$$||D_{1n}|| \leq C \left| \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{||\tilde{x}_{ij}||}{N(n)} e_{ij} \right| = o(1), \quad \text{a.s.}. \tag{4.3}$$

By A3(i-ii), (4.2), Lemma 3.4 or Corollary 3.1, we have

$$||D_{2n}|| \leq C \max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left\| \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij}) e_{kl} \right\| \cdot \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{||\tilde{x}_{ij}||}{N(n)} = o(1). \quad \text{a.s.}. \tag{4.4}$$

From A2(iii) and A3(iii), we obtain

$$\begin{aligned}
\max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left| \tilde{g}(t_{ij}) \right| &= \max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left| g(t_{ij}) - \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij}) g(t_{kl}) \right| \\
&\leq \max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left| \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij})(g(t_{ij}) - g(t_{kl})) I(|t_{ij} - t_{kl}| > \varepsilon) \right| \\
&\quad + \max_{1 \leq i \leq n, 1 \leq j \leq m_i} \left| \sum_{k=1}^{n} \sum_{l=1}^{m_i} W_{nkl}(t_{ij})(g(t_{ij}) - g(t_{kl})) I(|t_{ij} - t_{kl}| \leq \varepsilon) \right| = o(1).
\end{aligned} \tag{4.5}$$

Together with (4.2), one gets

$$||D_{3n}|| \leq C \max_{1 \leq i \leq n, 1 \leq j \leq m_i} |\tilde{g}(t_{ij})| \cdot \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{||\tilde{x}_{ij}||}{N(n)} = o(1). \tag{4.6}$$

By (4.1), (4.3), (4.4) and (4.6), (2.5) holds.

Proof of Theorem 2.2 From (1.1) and (2.3), we have

$$\begin{aligned}
\hat{g}_n(t) - g(t) &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t)(\gamma_{ij} - x_{ij}^T \hat{\beta}_n) - g(t) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) \left((\gamma_{ij} - x_{ij}^T \hat{\beta}_n) - (\gamma_{ij} - x_{ij}^T \beta) \right) + \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t)(\gamma_{ij} - x_{ij}^T \beta) - g(t) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) x_{ij}^T (\beta - \hat{\beta}_n) + \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t)(g(t_{ij}) + e_{ij}) - g(t) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) x_{ij}^T (\beta - \hat{\beta}_n) + \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) e_{ij} - \tilde{g}(t) \\
&=: E_{1n} + E_{2n} + E_{3n}.
\end{aligned} \tag{4.7}$$

By A3(iv) and (2.5), one gets

$$|E_{1n}| \leq \left\| \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) x_{ij} \right\| ||\beta - \hat{\beta}_n|| = o(1), \quad \text{a.s.}. \tag{4.8}$$

By Lemma 3.4 or Corollary 3.1, $E_{2n} = o(1)$, a.s.; With arguments similar to (4.5), we have $E_{3n} = o(1)$. Therefore, together with (4.7) and (4.8), (2.6) holds.

Proof of Theorem 2.3 Here, we still use (4.7), but E_{in} in (4.7) are replaced by $E_{in}(t)$ for $i = 1,2$ and 3. By A3(v) and (2.5), we get

$$\sup_{0 \le t \le 1} |E_{1n}(t)| \le \sup_{0 \le t \le 1} \left\| \sum_{i=1}^{n} \sum_{j=1}^{m_i} W_{nij}(t) x_{ij} \right\| \|\beta - \hat{\beta}_n\| = o(1), \quad \text{a.s..}$$

By Lemma 3.5 or Corollary 3.2, $\sup_{0 \le t \le 1} |E_{2n}(t)| = o(1)$, a.s.; Similar to the arguments in (4.5), we have $\sup_{0 \le t \le 1} |E_{2n}(t)| = o(1)$. Hence, (2.7) is proved.

5 Simulation study

To evaluate the finite-sample performance of the least squares estimator $\hat{\beta}_n$ and the nonparametric estimator $\hat{g}_n(t)$, we respectively take two forms of functions for $g(\cdot)$:

$$\text{I.} \quad g(t) = \exp(3t); \quad \text{II.} \quad g(t) = \cos\left(\frac{3\pi}{2} t\right),$$

consider the case where $p = 1$ and $m_i = m = 12$, and take the design points $t_{ij} = ((i - 1)m + j)/(nm)$, $x_{ij} \sim N(1, 1)$ and the errors $e_{ij} = 0.2 e_{i, j-1} + \epsilon_{ij}$, where ϵ_{ij} are i.i.d. $N(0,1)$ random variables, and $e_{i,0} \sim N(0,1)$ for each i.

The kernel function is taken as the Epanechnikov kernel $K(t) = \frac{3}{4}(1 - t^2)I(|t| \le 1)$, and the weight function is given by Nadaraya-Watson kernel weight $W_{ij}(t) = K\left(\frac{t - t_{ij}}{h_n}\right) / \sum_{i=1}^{n} \sum_{j=1}^{m_i} K\left(\frac{t - t_{ij}}{h_n}\right)$. The bandwidth h is selected by a "leave-one-subject-out" cross validation method. In the simulations, we draw $B = 1000$ random samples of sizes 150,200,300 and 500 for $\beta = 2$, respectively. We obtain the estimators $\hat{\beta}_n$ and $\hat{g}_n(t)$ from (2.2) and (2.3), respectively. Let $\hat{\beta}_n^{(b)}$ be bth least squares estimator of β under the size n. Some numerical results for $\hat{\beta}_n$ are computed by

$$\bar{\beta}_n = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_n^{(b)}, \quad \widehat{\text{SD}}(\hat{\beta}_n) = \left(\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\beta}_n^{(b)} - \bar{\beta})^2 \right)^{1/2},$$

$$\widehat{\text{Bias}}(\hat{\beta}_n) = \bar{\beta} - \beta, \quad \widehat{\text{MSE}}(\hat{\beta}_n) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\beta}_n^{(b)} - \beta)^2,$$

which are listed in Table 1.

In addition, for assessing estimator of the nonparametric component $g(\cdot)$, we study the square root of mean-squared errors (RMSE) based on 1000 repetitions. Denote $\hat{g}_n^{(b)}(t)$ be the bth estimator of $g(t)$ under the size n, and $\bar{\hat{g}}_n(t) = \sum_{b=1}^{B} \hat{g}_n^{(b)}(t)/B$ be the average estimator of $g(t)$. We compute

$$\text{RMSE}_n = \left(\frac{1}{M} \sum_{s=1}^{M} (\bar{\hat{g}}_n(t_s) - g(t_s))^2 \right)^{1/2},$$
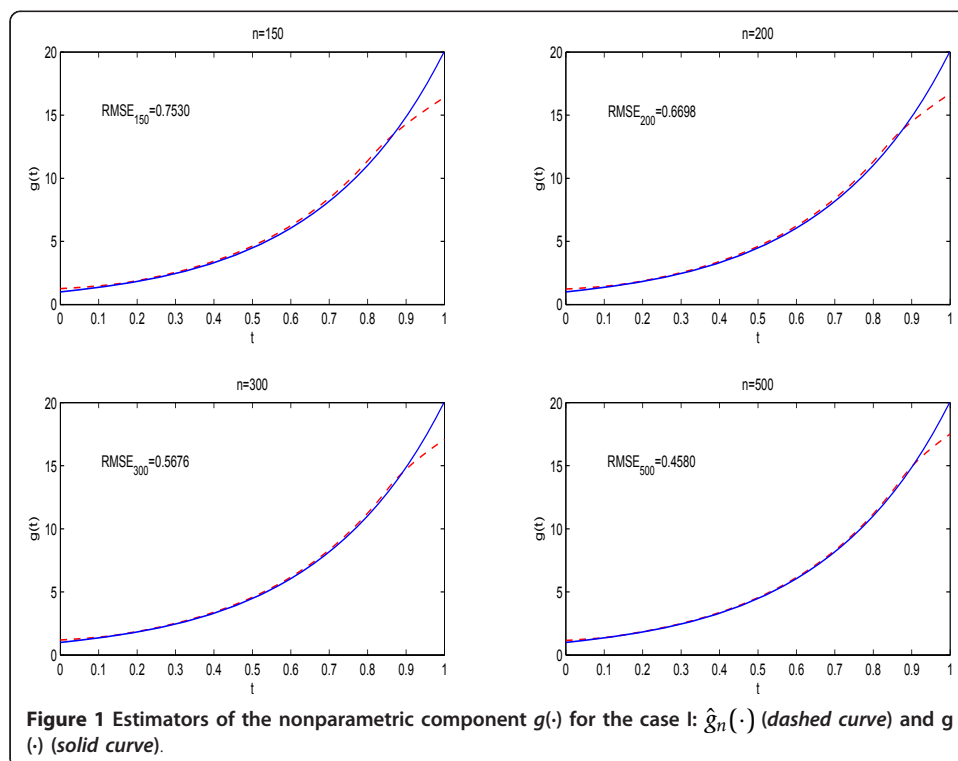
and

$$\text{RMSE}_n^{(b)} = \left(\frac{1}{M} \sum_{s=1}^{M} (\hat{g}_n^{(b)}(t_s) - g(t_s))^2 \right)^{1/2}, \quad b = 1, 2, \ldots, B,$$
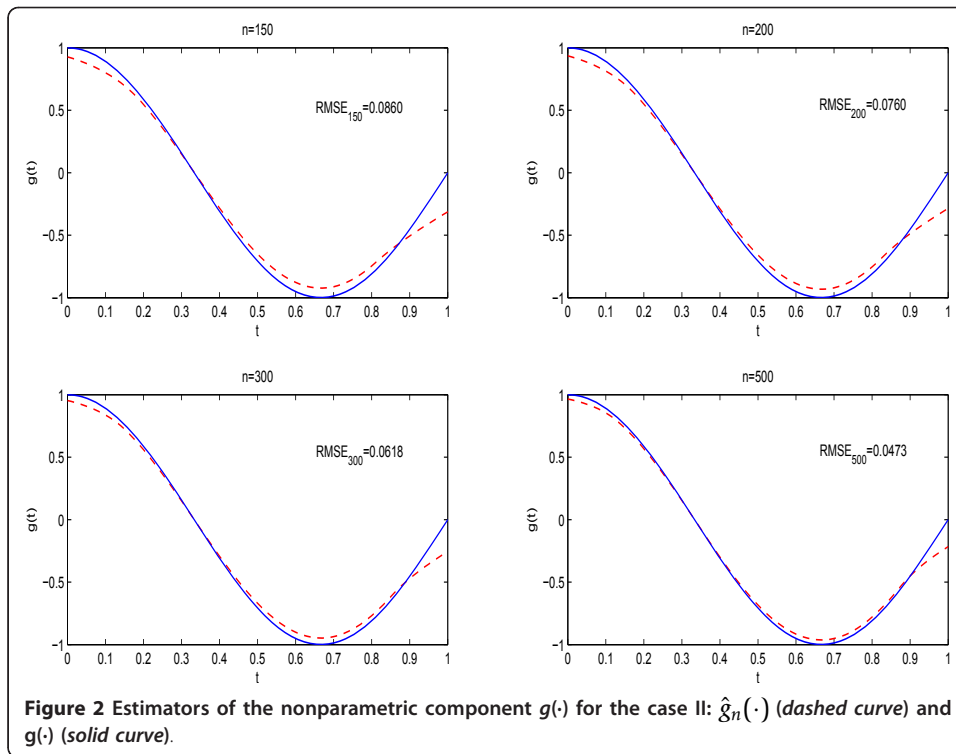
Table 1 The estimators of β and some indices of their accuracy for the different sample size n and nonparametric function $g(\cdot)$

$g(\cdot)$	n	$\bar{\beta}_n$	$\widehat{\text{Bias}}(\hat{\beta}_n)$	$\widehat{\text{SD}}(\hat{\beta}_n)$	$\widehat{\text{MSE}}(\hat{\beta}_n)$
exp(3t)	150	1.99938	-0.00062	0.0257	0.00066
	200	1.99960	-0.00040	0.0221	0.00049
	300	1.99965	-0.00035	0.0172	0.00030
	500	1.99963	-0.00037	0.0133	0.00018
$\cos\left(\dfrac{3\pi}{2}t\right)$	150	1.99908	-0.00092	0.0238	0.00056
	200	1.99945	-0.00055	0.0206	0.00043
	300	1.99950	-0.00050	0.0168	0.00028
	500	1.99969	-0.00031	0.0131	0.00017

where $\{t_s, s = 1,..., M\}$ is a sequence of regular grid points on [0, 1]. Figures 1 and 2 respectively provide the average estimators of the nonparametric function $g(\cdot)$ and RMSE_n values for Cases I and II, respectively. The boxplots for $\text{RMSE}_n^{(b)}(b = 1, 2, \ldots, B)$ values for Cases I and II are presented in Figure 3.

From Table 1, we see that (i) $|\widehat{\text{Bias}}(\hat{\beta}_n)|, \widehat{\text{SD}}(\hat{\beta}_n)$ and $\widehat{\text{MSE}}(\hat{\beta}_n)$ do decrease with increasing the sample size n; (ii) the larger the sample size n is, the closer the $\bar{\beta}_n$ is to the true value 2. From Figures 1, 2 and 3, we observe that the biases of estimators of the nonparametric component $g(\cdot)$ decrease as the sample size n increases. These show that, for semiparametric partially linear regression models for longitudinal data based on mixing error's structure, the least squares estimator of parametric component β and the estimator of nonparametric component $g(\cdot)$ work well.

Figure 1 Estimators of the nonparametric component $g(\cdot)$ for the case I: $\hat{g}_n(\cdot)$ (*dashed curve*) **and g** $(\cdot)$ (*solid curve*).

Figure 2 Estimators of the nonparametric component $g(\cdot)$ for the case II: $\hat{g}_n(\cdot)$ (*dashed curve*) **and $g(\cdot)$** (*solid curve*).

6 Concluding remarks

An inherent characteristic of longitudinal data is the dependence among the observations within the same subject. For exhibiting dependence among the observations within the same subject, we consider the estimation problems of partially linear models for longitudinal data with the ϕ-mixing and ρ-mixing error structures, respectively. The strong consistency for least squares estimator $\hat{\beta}_n$ of parametric component β is

Figure 3 The boxplots of $\mathrm{RMSE}_n^{(b)}(b = 1, 2, \ldots, B)$ values in the estimators of $g(\cdot)$.
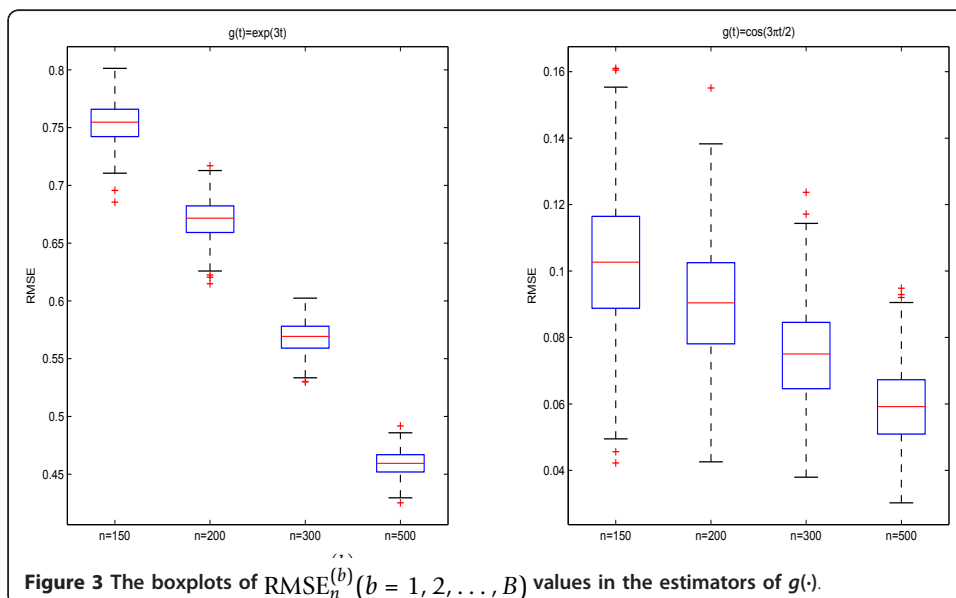
studied. In addition, the strong consistency and uniform consistency for the estimator $\hat{g}_n(\cdot)$ of nonparametric function $g(\cdot)$ are investigated under some mild conditions.

In the paper, we only consider (x_{ij}^T, t_{ij}) are known and nonrandom design points, as Baek and Liang [16], and Liang and Jing [20]. In the monograph of Hardle et al. [7], they respectively considered the two cases: the fixed design and the random design, to study non-longitudinal partially linear regression models. Our results can also be extended to the case of (x_{ij}^T, t_{ij}) being random. The interested readers can consider the work. In addition, we consider partially linear models for longitudinal data with only ϕ-mixing and ρ-mixing. In fact, our results with other mixing-dependent structures, such as α-mixing, ϕ^*-mixing and ρ^*-mixing, can also be obtained by the same arguments in our paper. At present, we have not given the asymptotic normality of estimators, since some details need further discussion. We will devote to establish the asymptotic normality of $\hat{\beta}_n$ and $\hat{g}_n(\cdot)$ in our future work.

Author details
[1]Department of Mathematics, Southeast University, Nanjing 210096, People's Republic of China [2]Department of Mathematics and Computer Science, Tongling University, Tongling 244000, Anhui, People's Republic of China

Authors' contributions
The two authors contributed equally to this work. All authors read and approved the final manuscript.

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

References
1. Diggle, LD, Heagerty, P, Liang, K, Zeger, S: Analysis of Longitudinal Data. Oxford University Press, New York, 2 (2002)
2. Fan, J, Li, R: New estimation and model selection procedure for semiparametric modeling in longitudinal data analysis. J Am Stat Assoc. **99**, 710–723 (2004). doi:10.1198/016214504000001060
3. Engle, R, Granger, C, Rice, J, Weiss, A: Nonparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc. **81**, 310–320 (1986). doi:10.2307/2289218
4. Heckman, N: Spline smoothing in a partly linear models. J R Stat Soc B. **48**, 244–248 (1986)
5. Speckman, P: Kernel smoothing in partial linear models. J R Stat Soc B. **50**, 413–436 (1988)
6. Robinson, PM: Root-n-consistent semiparametric regression. Econometrica. **56**, 931–954 (1988). doi:10.2307/1912705
7. Härdle, W, Liang, H, Gao, JT: Partial Linear Models. Physica-Verlag, Heidelberg (2000)
8. Zeger, SL, Diggle, PL: Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics. **50**, 689–699 (1994). doi:10.2307/2532783
9. Moyeed, RA, Diggle, PJ: Rate of convergence in semiparametric modeling of longitudinal data. Aust J Stat. **36**, 75–93 (1994). doi:10.1111/j.1467-842X.1994.tb00640.x
10. Zhang, D, Lin, X, Raz, J, Sowerm, MF: Semiparametric stochastic mixed models for longitudinal data. J Am Stat Assoc. **93**, 710–719 (1998). doi:10.2307/2670121
11. You, JH, Zhou, X: Partially linear models and polynomial spline approximations for the analysis of unbalanced panel data. J Stat Plan Inference. **139**, 679–695 (2009). doi:10.1016/j.jspi.2007.04.037
12. You, JH, Zhou, X, Zhou, Y: Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors. J Multivar Anal. **101**, 1079–1101 (2010). doi:10.1016/j.jmva.2010.01.003
13. Schick, A: An adaptive estimator of the autocorrelation coefficient in regression models with autocoregressive errors. J Time Ser Anal. **19**, 575–589 (1998). doi:10.1111/1467-9892.00109
14. Gao, JT, Anh, VV: Semiparametric regression under long-range dependent errors. J Stat Plan Inference. **80**, 37–57 (1999). doi:10.1016/S0378-3758(98)00241-9
15. Sun, XQ, You, JH, Chen, GM, Zhou, X: Convergence rates of estimators in partial linear regression models with MA(∞) error process. Commun Stat Theory Methods. **31**, 2251–2273 (2002). doi:10.1081/STA-120017224
16. Baek, J, Liang, HY: Asymptotics of estimators in semi-parametric model under NA samples. J Stat Plan Inference. **136**, 3362–3382 (2006). doi:10.1016/j.jspi.2005.01.008
17. Zhou, XC, Liu, XS, Hu, SH: Moment consistency of estimators in partially linear models under NA samples. Metrika. **72**, 415–432 (2010). doi:10.1007/s00184-009-0260-5

18. Li, GL, Liu, LQ: Strong consistency of a class estimators in partial linear model under martingale difference sequence. Acta Math Sci (Ser A). **27**, 788–801 (2007)
19. Chen, X, Cui, HJ: Empirical likelihood inference for partial linear models under martingale difference sequence. Stat Probab Lett. **78**, 2895–2910 (2008). doi:10.1016/j.spl.2008.04.012
20. Liang, HY, Jing, BY: Asymptotic normality in partial linear models based on dependent errors. J Stat Plan Inference. **139**, 1357–1371 (2009). doi:10.1016/j.jspi.2008.08.005
21. He, X, Zhu, ZY, Fung, WK: Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika. **89**, 579–590 (2002). doi:10.1093/biomet/89.3.579
22. Xue, LG, Zhu, LX: Empirical likelihood-based inference in a partially linear model for longitudinal data. Sci China (Ser A). **51**, 115–130 (2008). doi:10.1007/s11425-008-0020-4
23. Li, GR, Tian, P, Xue, LG: Generalized empirical likelihood inference in semipara-metric regression model for longitudinal data. Acta Math Sin (Engl Ser). **24**, 2029–2040 (2008). doi:10.1007/s10114-008-6434-7
24. Bai, Y, Fung, WK, Zhu, ZY: Weighted empirical likelihood for generalized linear models with longitudinal data. J Multivar Anal. **140**, 3445–3456 (2010)
25. Hu, Z, Wang, N, Carroll, RJ: Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. Biometrika. **91**, 251–262 (2004). doi:10.1093/biomet/91.2.251
26. Wang, S, Qian, L, Carroll, RJ: Generalized empirical likelihood methods for analyzing longitudinal data. Biometrika. **97**, 79–93 (2010). doi:10.1093/biomet/asp073
27. Chi, EM, Reinsel, GC: Models for longitudinal data with random effects and AR(1) errors. J Am Stat Assoc. **84**, 452–459 (1989). doi:10.2307/2289929
28. Shao, QM: A moment inequality and its application. Acta Math Sin. **31**, 736–747 (1988)
29. Peligrad, M: The r-quick version of the strong law for stationary φ-mixing sequences. Proceedings of the International Conference on Almost Everywere Convergence in Probability and Statistics. pp. 335–348.Academic Press, New York (1989)
30. Utev, SA: Sums of random variables with φ-mixing. Sib Adv Math. **1**, 124–155 (1991)
31. Kiesel, R: Summability and strong laws for φ-mixing random variables. J Theor Probab. **11**, 209–224 (1998). doi:10.1023/A:1021655227120
32. Chen, PY, Hu, TC, Volodin, A: Limiting behaviour of moving average processes under φ-mixing assumption. Stat Probab Lett. **79**, 105–111 (2009). doi:10.1016/j.spl.2008.07.026
33. Zhou, XC: Complete moment convergence of moving average processes under φ-mixing assumptions. Statist Probab Lett. **80**, 285–292 (2010). doi:10.1016/j.spl.2009.10.018
34. Peligrad, M: On the central limit theorem for p-mixing sequences of random variables. Ann Probab. **15**, 1387–1394 (1987). doi:10.1214/aop/1176991983
35. Peligrad, M, Shao, QM: Estimation of variance for p-mixing sequences. J Multivar Anal. **52**, 140–157 (1995). doi:10.1006/jmva.1995.1008
36. Peligrad, M, Shao, QM: A note on estimation of the variance of partial sums for p-mixing random variables. Stat Probab Lett. **28**, 141–145 (1996)
37. Shao, QM: Maximal inequalities for partial sums of p-mixing sequences. Ann Probab. **23**, 948–965 (1995). doi:10.1214/aop/1176988297
38. Bradley, R: A stationary rho-mixing Markov chain which is not "interlaced" rho-mixing. J Theor Probab. **14**, 717–727 (2001). doi:10.1023/A:1017545123473
39. Lin, ZY, Lu, CR: Limit Theory for Mixing Dependent Random Variables. Science Press/Kluwer Academic Publishers, Beijing/London (1996)
40. Roussas, GG: Nonparametric regression estimation under mixing conditions. Stoch Process Appl. **36**, 107–116 (1990). doi:10.1016/0304-4149(90)90045-T
41. Truong, YK: Nonparametric curve estimation with time series errors. J Stat Plan Inference. **28**, 167–183 (1991). doi:10.1016/0378-3758(91)90024-9
42. Fraiman, R, Iribarren, GP: Nonparametric regression estimation in models with weak error's structure. J Multivar Anal. **37**, 180–196 (1991). doi:10.1016/0047-259X(91)90079-H
43. Roussas, GG, Tran, LT: Asymptotic normality of the recursive kernel regression estimate under dependence conditions. Ann Stat. **20**, 98–120 (1992). doi:10.1214/aos/1176348514
44. Masry, E, Fan, JQ: Local polynomial estimation of regression functions for mixing processes. Scand J Stat. **24**, 165–179 (1997). doi:10.1111/1467-9469.00056
45. Aneiros, G, Quintela, A: Asymptotic properties in partial linear models under dependence. Test. **10**, 333–355 (2001). doi:10.1007/BF02595701
46. Fan, JQ, Yao, QW: Nonlinear Time Series–Nonparametric and Parametric methods. Springer, New York (2003)
47. Chen, GM, You, JH: An asymptotic theory for semiparametric generalized least squares estimation in partially linear regression models. Stat Pap. **46**, 173–193 (2005). doi:10.1007/BF02762967