ORIGINAL PAPER

The large deviation results for the nonlinear regression model with dependent errors

Wenzhi Yang[1] · **Zhangrui Zhao**[1] ·
Xinghui Wang[2] · **Shuhe Hu**[1]

Abstract In this paper, we investigate the least squares (LS) estimator of the nonlinear regression model based on the extended negatively dependent errors which are widely dependent structures. Under the general conditions, we establish some large deviation results for the LS estimator of the nonlinear regression parameter, which can be applied to obtain a weak uniform consistency and a complete convergence rate for this estimator. In addition, some examples and simulations are presented for illustration.

Keywords Large deviation · Nonlinear regression models · Least squares estimator · END random variables

Mathematics Subject Classification 62J02 · 62F12

✉ Shuhe Hu
   hushuhe@263.net

   Wenzhi Yang
   wzyang827@163.com

   Zhangrui Zhao
   zruiz2958@163.com

   Xinghui Wang
   wangxinghuial@163.com

[1] School of Mathematical Sciences, Anhui University, Hefei, People's Republic of China

[2] School of Economics, Anhui University, Hefei, People's Republic of China

🖄 Springer

1 Introduction

1.1 The nonlinear regression model

First, we consider the nonlinear regression model for the observations $X^n :=$ $(X_1, X_2, \ldots, X_n)$:

$$X_t = f_t(\theta) + \varepsilon_t, \ \ t = 1, 2, \ldots, n, \tag{1}$$

where the f_t are known continuous functions on a parameter set $\Theta \subset \mathscr{R}^k$, the ε_t are random errors and the $\theta \in \Theta$ is the true value of the parameter. Denote

$$Q_n(\theta) = \sum_{t=1}^{n} (X_t - f_t(\theta))^2 .$$

Let $\hat{\theta}_n(X_1, X_2, \ldots, X_n)$ denote the least squares (LS) estimator of the parameter $\theta \in \Theta$ such as

$$Q_n\left(\hat{\theta}_n\right) = \inf_{\theta \in \Theta} \sum_{t=1}^{n} (X_t - f_t(\theta))^2 .$$

The LS method plays a central role in the inference of parameters in nonlinear regression models. The study of asymptotic properties of the LS estimator for parameters in nonlinear regression models has been the main subject of investigation, since it is, in general, difficult to obtain the exact distribution of the LS estimator for any fixed sample. For the LS estimator of the nonlinear model based on the independent identically distributed ($i.i.d.$) errors, Jennrich (1969) presented the asymptotic normality, Malinvaud (1970) obtained the consistency, and Wu (1981) established the necessary and sufficient condition for the strong consistency, etc. In addition, for the the nonlinear model based on the independent but not necessarily identically distributed errors, Bunke and Schmidt (1980) established the strong consistency and asymptotic normality for the weighted LS estimator, Ibragimov and Has'minskii (1981) obtained some large deviation results of the maximum likelihood (ML) estimator, Sieders and Dzhaparidze (1987) extended the results of Ibragimov and Has'minskii (1981) to the M-estimator and gave its application to the main results of large deviation for the LS estimator. For the LS estimator of the nonlinear model, Prakasa Rao (1984a) extended the result of Ibragimov and Has'minskii (1981) to the case of $i.i.d.$ Gaussian errors, Hu (1993) extended the results of Prakasa Rao (1984a) and Sieders and Dzhaparidze (1987) to the cases of locally generalized Gaussian errors, martingale differences, etc.

As far as we know, there is no large deviation result of the LS estimator of model (1) based on the extended negatively dependent (END, Liu 2009) errors. The END sequences are widely dependent structures which cover several negative dependent sequences such as negatively orthant dependent (NOD, Lehmann 1966), negatively superadditive dependent (NSD, Hu 2000) and negatively associated (NA, Joag-Dev and Proschan 1983). Based on the M estimator of Sieders and Dzhaparidze (1987), we obtain the main result of large deviation such as Theorems 2.1–2.4 and Corollary

2.1 for the LS estimator $\hat{\theta}_n$ of $\theta \in \mathcal{R}^k$ in model (1), which can be applied to establish a weak uniform consistency and a complete convergence rate.

Now, we recall the M-estimator. Let $\mathcal{E}^{(n)} = \{\mathcal{X}^{(n)}, \mathcal{U}^{(n)}, P_\theta^{(n)}, \theta \in \Theta\}$ be a family of probability spaces, where the $P_\theta^{(n)}$ does not necessarily have known form. The parameter set Θ is a Borel subset of k-dimensional Euclidean space. We shall consider the M-estimator maximizing an M-functional $C_n : \mathcal{X}^{(n)} \times \Theta \to [0, \infty)$, which is assumed to be, for all $X^n \in \mathcal{X}^{(n)}$, a positive continuous function of θ and, for all $\theta \in \Theta$, a measurable functional of X^n.

Throughout the paper, we assume that, for all $\theta \in \Theta$ and $P_\theta^{(n)}$-almost all X^n, a solution $\hat{\theta}_n$ to the equation

$$C_n\left(X^n, \hat{\theta}_n\right) = \sup_{\theta \in \Theta} C_n\left(X^n, \theta\right) \tag{2}$$

exists (this is certainly true if Θ is compact). So $\hat{\theta}_n$ is called the M-estimator of θ. Especially, the LS estimator $\hat{\theta}_n$ maximizes the M-functional

$$C_n\left(X^n, \theta\right) := \exp\left(-\frac{1}{2} \sum_{t=1}^{n} (X_t - f_t(\theta))^2\right).$$

For all $n \in \mathcal{N}$ and $\theta \in \Theta \subset \mathcal{R}^k$, let $u \in \mathcal{R}^k$, $\phi_n(\theta)$ be a nonsingular $k \times k$ matrix and define the normalized M-ratio

$$Z_{n,\theta}(u) := Z_{n,\theta}\left(X^n, u\right) = \frac{C_n\left(X^n, \theta + \phi_n(\theta)u\right)}{C_n\left(X^n, \theta\right)}, \tag{3}$$

which, for fixed observation X^n, is a continuous, nonnegative finite function on the set

$$U_{n,\theta} = \phi_n^{-1}(\theta)\left(\Theta - \theta\right).$$

Throughout the paper, for a matrix $A_{m \times n}$, $|A_{m \times n}|$ denotes its norm. Define

$$\Gamma_{n,\theta,R} := \bar{U}_{n,\theta} \cap \{u : R \le |u| \le R + 1\},$$

where $\bar{U}_{n,\theta}$ is a closure of $U_{n,\theta}$.

Similar to Theorem 1.5.1 of Ibragimov and Has'minskii (1981) and Theorem 2.1 of Sieders and Dzhaparidze (1987), we define the following sets of functions.

G is the set of all functions $g_n(\cdot)$ possessing the following properties:
(i) for fixed n, $g_n(\cdot)$ is a function on $[0, \infty)$ monotonically increasing to infinity;
(ii) for all $N > 0$,

$$\lim_{\substack{R \to \infty \\ n \to \infty}} R^N \exp(-g_n(R)) = 0. \tag{4}$$

Remark 1.1 If $g_n(R) = R^\alpha$ and $\alpha > 0$, then $g_n \in \mathbf{G}$.

Let K be a measurable subset of Θ and $\mathbf{H}_K$ be the set of all functions $\eta_{n,\theta}(\cdot)$ possessing the following properties:

(iii) for fixed n and $\theta \in \Theta$, $\eta_{n,\theta}(\cdot)$ is a function $U_{n,\theta} \to (0, \infty)$;

(vi) there exists a polynomial $\text{pol}_K(R)$ in R such that for R and n sufficiently large,

$$\sup_{\theta \in K; \, u \in \Gamma_{n,\theta,R}} \left(\eta_{n,\theta}(u)\right)^{-1} \leq \text{pol}_K(R). \tag{5}$$

For each n and θ, let $\zeta_{n,\theta} : [0, \infty) \to \mathscr{R}$ be a monotonically nondecreasing continuous function and define the random function

$$\zeta_{n,\theta}(u) := \zeta_{n,\theta}\left(Z_{n,\theta}(u)\right). \tag{6}$$

As a generalization of Theorem 1.5.1 of Ibragimov and Has'minskii (1981), Sieders and Dzhaparidze (1987, Theorem 2.1) obtained a large deviation result for the M-estimator as follows.

Theorem 1.1 *Let the functionals $\zeta_{n,\theta}(u)$ process the following properties: given a measurable subset $K \subset \Theta \subset \mathscr{R}^k$, there correspond to it numbers m and α, where $m \geq \alpha > k$, functions $g_n \in G$ and $\eta_{n,\theta} \in \mathbf{H}_K$, and a polynomial $\text{pol}_K(R)$ in R such that, for all R and n large enough,*

$$E_\theta^{(n)} |\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v)|^m \leq |u - v|^\alpha \text{pol}_K(R),$$
$$\textit{for all } \theta \in K \textit{ and } u, v \in \Gamma_{n,\theta,R}, \tag{7}$$

$$P_\theta^{(n)}(\zeta_{n,\theta}(u) - \zeta_{n,\theta}(0) \geq -\eta_{n,\theta}(u)) \leq \exp(-g_n(R))$$
$$\textit{for all } \theta \in K \textit{ and } u \in \Gamma_{n,\theta,R}. \tag{8}$$

Then there exist positive constants B_0 and b_0 such that, for all n and H large enough,

$$\sup_{\theta \in K} P_\theta^{(n)} \left\{ |\phi_n^{-1}(\theta) \left(\hat{\theta}_n - \theta\right)| \geq H \right\} \leq B_0 \exp\left(-b_0 g_n(H)\right). \tag{9}$$

The constant b_0 can be made arbitrarily close to $(\alpha - k)/(\alpha - k + mk)$ by choosing B_0 large enough.

Remark 1.2 In view of Sieders and Dzhaparidze (1987), the condition (8) can be replaced by the following condition

$$P_\theta^{(n)} \left(\zeta_{n,\theta}(u) - \zeta_{n,\theta}(0) \geq -\eta_{n,\theta}(u)\right) \leq C \exp\left(-g_n(R)\right) \tag{10}$$

for all $\theta \in K$ and $u \in \Gamma_{n,\theta,R}$, where C is a positive constant independent of n and θ.

Ibragimov and Has'minskii (1981, Theorem 1.5.1) obtained the large deviation (9) for the ML estimator. Under the $i.i.d.$ Gaussian errors, Prakasa Rao (1984a) obtained the result of LS estimator $\hat{\theta}_n$ such that for all $\rho > 0$ and $n \geq 1$

$$\sup_{\theta \in K} P_\theta \left(n^{1/2}|\hat{\theta}_n - \theta| > \rho\right) \leq B e^{-b\rho^2},$$

where K is a compact in $\Theta \subset \mathscr{R}$, B and b are some positive constants. Hu (1993) extended (9) to the locally generalized Gaussian and martingale differences cases. In addition, Ivanov (1976) investigated the LS estimator $\hat{\theta}_n$ of model (1) based on the $i.i.d.$ errors. Assume that there exist positive constants D_1 and D_2 such that, for all $\theta, \theta' \in \Theta \subset \mathscr{R}$,

$$D_1 n \left(\theta - \theta'\right)^2 \leq \sum_{t=1}^{n} \left(f_t(\theta) - f_t(\theta')\right)^2 \leq D_2 n \left(\theta - \theta'\right)^2. \tag{11}$$

By (11) and some other conditions, Ivanov (1976) presented that for all $\rho > 0$ and all $n \geq 1$,

$$P_\theta \left(n^{1/2}|\hat{\theta}_n - \theta| > \rho\right) \leq C\rho^{-p}, \tag{12}$$

where p is a positive constant such as $p \geq 2$, and C is a positive constant independent of n and ρ. Prakasa Rao (1984b) extended (12) to the dependent cases of φ-mixing and α-mixing errors. Under some general conditions and $\sup_{n \geq 1} E|\varepsilon_n|^p < \infty$ for some $p > 2$, Hu (2002) also obtained (12) and gave some applications to the dependent cases of martingale differences, φ-mixing sequence and NA sequence. By the condition $\sup_{n \geq 1} E|\varepsilon_n|^p < \infty$ for some $1 < p \leq 2$, Hu (2004) established that

$$P_\theta \left(n^{1/2}|\hat{\theta}_n - \theta| > \rho\right) \leq Cn^{1-p/2}\rho^{-p}, \tag{13}$$

for all $\rho > 0$, $n \geq 1$ and some $C > 0$, which was also applied to some dependent errors. In view of (12) and (13), by some moment information of errors, Yang and Hu (2014) obtained some similar results of (12) and (13), which can be used in some case satisfying $\sup_{n \geq 1} E|\xi_n|^p = \infty$ for some $p > 1$.

For more works on the nonlinear regression models, one can refer to Ivanov and Leonenko (1989) and Ivanov (1997) for some basic asymptotic theories, Midi (1999) for the robustness of weighted LS estimator under $i.i.d.$ errors with mean zero and unknown variance σ^2, Ivanov and Leonenko (2008) for the consistency and asymptotic distribution theory of LS estimator under long-range-dependent noise, etc.

Due to the importance of END random variables and the LS estimator of a nonlinear regression parameter, we investigate the LS estimator $\hat{\theta}_n$ for the model (1) based on the END errors which are not necessarily identically distributed. With the techniques of some exponential inequalities of END random variables giving by Sect. 5, we obtain the large deviation results for the LS estimator $\hat{\theta}_n$, which can be applied to get a weak uniform consistency and a complete convergence rate $\hat{\theta}_n - \theta = O(n^{-1/2}\log^{1/2} n)$, completely (see our results in Sect. 2). Some examples and simulations for the nonlinear models are illustrated in Sect. 3, and the conclusions are presented in Sect. 4. Last, we give the proofs in Sect. 5.

1.2 The concept of END random variables

In this subsection, let us recall the concept of END random variables which was introduced by Liu (2009).

Definition 1.1 *We call random variables $\{Z_n, n \geq 1\}$ to be END if there exists a constant $M > 0$ such that both*

$$P\left(Z_i > z_i, i = 1, 2, \ldots, n\right) \leq M \prod_{i=1}^{n} P\left(Z_i > z_i\right)$$

and

$$P\left(Z_i \leq z_i, i = 1, 2, \ldots, n\right) \leq M \prod_{i=1}^{n} P\left(Z_i \leq z_i\right)$$

hold for each $n \geq 1$ and all real numbers $z_1, z_2, \ldots, z_n$.

If $\{Z_n, n \geq 1\}$ is a sequence of END random variables, then for any fixed $m \geq 1$, $\{Z_{n+m}, n \geq 1\}$ is also a sequence of END random variables with the same dominating coefficient M. In fact, by Definition 1.1 and the continuity of probability, one can get this property.

Let $\{Z_n, n \geq 1\}$ be a sequence of random variables. For some $1 \leq i \leq n$, if $P\left(Z_i \leq z_i\right) = 0$, then $P\left(Z_1 \leq z_1, Z_2 \leq z_2, \ldots, Z_n \leq z_n\right) = 0$. Similarly, if for some $1 \leq i \leq n$, $P\left(Z_i > z_i\right) = 0$, then $P\left(Z_1 > z_1, Z_2 > z_2, \ldots, Z_n > z_n\right) = 0$. Define $\frac{0}{0} = 1$. If

$$M_1 = \sup_{n \geq 1} \sup_{z_i \in (-\infty, \infty), 1 \leq i \leq n} \frac{P\left(Z_i > z_i, i = 1, 2, \ldots, n\right)}{\prod_{i=1}^{n} P(Z_i > z_i)} < \infty$$

and

$$M_2 = \sup_{n \geq 1} \sup_{z_i \in (-\infty, \infty), 1 \leq i \leq n} \frac{P\left(Z_i \leq z_i, i = 1, 2, \ldots, n\right)}{\prod_{i=1}^{n} P(Z_i \leq z_i)} < \infty$$

then we take $M = \max\{M_1, M_2\}$ in Definition 1.1 and obtain that $\{Z_n, n \geq 1\}$ are END random variables. Obviously, for all $1 \leq i \leq n$, let $z_i = -\infty$ or $z_i = +\infty$ in Definition 1.1, it is easy to see that the dominating coefficient $M \geq 1$.

Moreover, for any $n \geq 1$, let $Z_1, Z_2, \ldots, Z_n$ be dependent according to a multivariate copula function $C(u_1, \ldots, u_n)$ with absolutely continuous distribution functions $F_1, \ldots, F_n$. Assume that the joint copula density

$$C_{1,\ldots,n}(u_1, \ldots, u_n) = \frac{\partial^n}{\partial u_1 \ldots \partial u_n} C(u_1, \ldots, u_n)$$

exists and is uniformly bounded in the whole domain. Then random variables $\{Z_n, n \geq 1\}$ are END (see Example 4.2 of Liu 2009). By Remark 3.1 of Ko and Tang (2008), the copulas in the Frank family of the form

$$C(u_1, \ldots, u_n; \theta) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1) \ldots (e^{-\theta u_n} - 1)}{(e^{-\theta} - 1)^{n-1}}\right), \quad \theta < 0$$

belong to this category. Meanwhile, Chen et al. (2010) showed that every *n*-dimensional Farlie–Gumbel–Morgenstern (FGM) distribution described a specific END structure.

If $M = 1$, then END random variables reduce to NOD random variables (see Lehmann 1966), which contain NA random variables and NSD random variables (see Joag-Dev and Proschan 1983; Hu 2000; Wang et al. 2015a). Joag-Dev and Proschan (1983) established that a permutation distribution is NA. Recall that a family of real-valued random variables $Z = \{Z_t, t \in T\}$ is called normal (or Gaussian) system if all its finite-dimensional distributions are Gaussian. Let $Z = (Z_1, \ldots, Z_n)$ be a normal random vector, $n \geq 2$. Then Joag-Dev and Proschan (1983) proved that it is NA if and only if its components are non-positively correlated. They also pointed out that NA random variables are NOD random variables but the converse statement cannot always be true. For various examples of NA random variables and the related fields, we can refer to the studies by Bulinski and Shaskin (2007), Prakasa Rao (2012), Oliveira (2012) and so on. Since END random variables are widely dependent random variables, many researchers pay attention to the study of the properties of END. For example, Liu (2009, 2010) studied the precise large deviations and moderate deviations of END sequence with heavy tails; Chen et al. (2010) obtained strong law of large numbers of END sequence. They also established some large deviation inequalities and applications to risk theory and renewal theory; Shen (2011) obtained some moment inequalities of END sequence; Wang et al. (2013) and Hu et al. (2015) investigated the complete convergence for END sequences; Wang et al. (2015b) investigated the application of the nonparametric regression model under END errors, etc.

2 The large deviation results of the LS estimator

Let Θ be a Borel subset of $\mathscr{R}^k$, $f_t(\theta)$ be a continuous deterministic function from Θ to $\mathscr{R}$ for each $t \in \mathcal{N}$. Assume that $X^n := (X_1, X_2, \ldots, X_n)$ are the observed random variables of the nonlinear regression model (1).

The LS estimator $\hat{\theta}_n$, which we assume to exist (see (2)), maximizes the *M*-functional

$$C_n(X^n, \theta) := \exp\left(-\frac{1}{2} \sum_{t=1}^n (X_t - f_t(\theta))^2 \right). \tag{14}$$

Given a sequence of nonsingular $k \times k$ matrix norming factors $\phi_n(\theta)$, we define the ratio

$$Z_{n,\theta}(u) := \frac{C_n\left(X^n, \theta + \phi_n(\theta)u\right)}{C_n\left(X^n, \theta\right)}$$

$$= \exp\left(\sum_{t=1}^n d_{tn\theta}(u)\varepsilon_t - \frac{1}{2} \sum_{t=1}^n d_{tn\theta}^2(u) \right), \tag{15}$$

where

$$d_{tn\theta}(u) = f_t\left(\theta + \phi_n(\theta)u\right) - f_t(\theta).$$

Similar to Theorem 3.1 of Sieders and Dzhaparidze (1987), we assume that, for some Borel subset K of Θ, there exist functions $g_n(R) \in \mathbf{G}$, positive constant $r > 0$, $\Lambda_1 \in (0, \infty]$, $\delta \in (0, 1/2)$, $k_1 > 0$, $\rho \in (0, 1]$ and a polynomial $\mathrm{pol}(R)$ in R such that for all n and R large enough, the following inequalities hold:

(N.1) for all $t \in \mathcal{N}$ and $|\lambda| \le \Lambda_1$,

$$E \exp(\lambda \varepsilon_t) \le \exp\left(\frac{1}{2} r \lambda^2\right);$$ (16)

(N.2) for all $\theta \in K$ and $u, v \in \Gamma_{n,\theta,R}$, where $|u - v| \le k_1$, one has

$$\sum_{t=1}^{n} [f_t(\theta + \phi_n(\theta)u) - f_t(\theta + \phi_n(\theta)v)]^2 \le |u - v|^{2\rho} \mathrm{pol}(R)$$ (17)

and

$$\sum_{t=1}^{n} d_{tn\theta}^2(u) \le \mathrm{pol}(R);$$ (18)

(N.3) for all $\theta \in K$ and $u \in \Gamma_{n,\theta,R}$, one has

$$\sum_{t=1}^{n} d_{tn\theta}^2(u) \ge \max\left(\frac{8r}{\delta^2}, \frac{4}{\Lambda_1 \delta} \max_{1 \le t \le n} |d_{tn\theta}(u)|\right) g_n(R).$$ (19)

By (N.1)–(N.3), we have the following large deviation result.

Theorem 2.1 *Assume that the errors $\{\varepsilon_t\}$ in the nonlinear regression model (1) are END random variables with $E\varepsilon_t = 0$ for all $t \in \mathcal{N}$. For some $K \subset \Theta \subset \mathscr{R}^k$ and suitably chosen nonsingular $\phi_n(\theta)$, let the assumptions (N.1)–(N.3) be fulfilled. Then there exist positive constants B_0 and b_0 such that, for all n and H large enough,*

$$\sup_{\theta \in K} P_\theta^{(n)}\left\{|\phi_n^{-1}(\theta)(\hat{\theta}_n - \theta)| \ge H\right\} \le B_0 \exp(-b_0 g_n(H)).$$ (20)

Moreover, for any $\beta > 0$, we can choose B_0 large enough such that $b_0 \ge \frac{\rho}{\rho+k} - \beta$.

We list two assumptions (N.1)′ and (N.4) as follows:

(N.1)′ for some $r > 0$, condition (N.1) holds with $\Lambda_1 = \infty$;
(N.4) there exist positive constants D_1 and D_2 such that, for all $\theta, \theta' \in \Theta \subset \mathscr{R}^k$ and n large enough,

$$D_1 |\phi_n^{-1}(\theta - \theta')|^2 \le \sum_{t=1}^{n} (f_t(\theta) - f_t(\theta'))^2 \le D_2 |\phi_n^{-1}(\theta - \theta')|^2.$$ (21)

Replacing (N.1)–(N.3) by (N.1)′ and (N.4), we have a result as follows.

Theorem 2.2 *Assume that the errors $\{\varepsilon_t\}$ in the nonlinear regression model* (1) *are END random variables with $E\varepsilon_t = 0$ for all $t \in \mathcal{N}$. For a suitably chosen nonsingular $\phi_n(\theta)$, let the assumptions* (N.1)$'$ *and* (N.4) *be fulfilled. Then there exist positive constants B_0 and b such that, for all n and H large enough,*

$$\sup_{\theta \in \Theta} P_\theta^{(n)}\left\{ |\phi_n^{-1}(\theta)(\hat{\theta}_n - \theta)| \geq H \right\} \leq B_0 \exp(-bH^2). \tag{22}$$

For any $\beta > 0$, it can be chosen B_0 large enough such that $b \geq \frac{D_1}{32r(1+k)} - \beta$.

Similar to (N.1) and (N.3), we give the following assumptions:
(N.1)* for all $t \in \mathcal{N}$, suppose that there exists a positive number L such that

$$|E\varepsilon_t^m| \leq \frac{m!}{2}\sigma^2 L^{m-2},$$

for all positive integers $m \geq 2$, where $\sigma^2 = Var(\varepsilon_t)$;
(N.3)$'$ for all $\theta \in K$ and $u \in \Gamma_{n,\theta,R}$, it has

$$\sum_{t=1}^n d_{tn\theta}^2(u) \geq \max\left(\frac{16\sigma^2}{\delta^2}, \frac{8L}{\delta} \max_{1 \leq t \leq n} |d_{tn\theta}(u)| \right) g_n(R), \tag{23}$$

where $0 < \delta < 1/2$.
Therefore, similar to Theorems 2.1 and 2.2, we also establish the following results:

Theorem 2.3 *Assume that the errors $\{\varepsilon_t\}$ in the nonlinear regression model* (1) *are END random variables with $E\varepsilon_t = 0$ and $Var(\varepsilon_t) = \sigma^2$ for all $t \in \mathcal{N}$. For some $K \subset \Theta \subset \mathcal{R}^k$ and suitably chosen nonsingular $\phi_n(\theta)$, let the assumptions* (N.1)*, (N.2) and (N.3)$'$ be fulfilled. Then there exist positive constants B_0 and b_0 such that, for all n and H large enough,* (20) *holds. For any $\beta > 0$, it can be chosen B_0 large enough such that $b_0 \geq \frac{\rho}{\rho+k} - \beta$.*

Theorem 2.4 *Assume that the errors $\{\varepsilon_t\}$ in the nonlinear regression model* (1) *are END random variables with $E\varepsilon_t = 0$ and $Var(\varepsilon_t) = \sigma^2$ for all $t \in \mathcal{N}$. For a suitably chosen nonsingular $\phi_n(\theta)$, let the conditions* (N.1)* and (N.4) be fulfilled. Then there exist positive constants B_0 and C_0 such that, for all n and H large enough,*

$$\sup_{\theta \in \Theta} P_\theta^{(n)}\left\{ |\phi_n^{-1}(\theta)\left(\hat{\theta}_n - \theta\right)| \geq H \right\} \leq B_0 \exp(-C_0 H). \tag{24}$$

For all $\theta \in \Theta \subset \mathcal{R}^k$, all $\rho > 0$ and n large enough, by taking $\phi_n(\theta) = n^{-1/2}I_k$ and $H = n^{1/2}\rho$ in Theorem 2.2, we obtain the following corollary, where I_k is a $k \times k$ unit matrix.

Corollary 2.1 *Assume that the errors $\{\varepsilon_t\}$ in the nonlinear regression model* (1) *are mean zero END random variables satisfying* (N.1)$'$. *Assume that there exist positive constants D_1 and D_2 such that, for all $\theta, \theta' \in \Theta \subset \mathcal{R}^k$ and n large enough,*

$$D_1 n |\theta - \theta'|^2 \leq \sum_{t=1}^{n} \left(f_t(\theta) - f_t(\theta') \right)^2 \leq D_2 n |\theta - \theta'|^2. \tag{25}$$

Then for all $\rho > 0$ and n large enough,

$$\sup_{\theta \in \Theta} P_\theta^{(n)} \left\{ |\hat{\theta}_n - \theta| \geq \rho \right\} \leq B_0 \exp(-b\rho^2 n), \tag{26}$$

where B_0 and b are defined by (22). *So it follows*

$$\hat{\theta}_n - \theta = O\left(n^{-1/2} \log^{1/2} n \right), \quad completely, \quad as \quad n \to \infty. \tag{27}$$

Remark 2.1 The (N.1), (N.1)′ and (N.1)* control the tails of errors ε_t for all $t \in \mathcal{N}$. Similar to Condition (III) of Ivanov (1976) (see (11)), Assumption A(ii) of Wu (1981), (2.5) of Prakasa Rao (1984a) and (N.2) of Sieders and Dzhaparidze (1987), the (N.2) is a Hölder-type continuity condition on the parametrization $\theta \to f(\theta)$. Similar to (N.3) of Sieders and Dzhaparidze (1987), the (N.3) and (N.3)′ prescribe the rate of asymptotic separation. The asymptotic separation is a necessary condition for the consistent estimation (see Theorem 1 of Wu 1981). The similar conditions can be found by Condition III of Ivanov (1976), (2.6) of Prakasa Rao (1984a), etc. In addition, by the proof of Theorem 2.2 in Sect. 5, (N.2) and (N.3) are fulfilled from the condition (N.4) and $\Lambda_1 = \infty$. Similarly, by the proof of Theorem 2.4, (N.2) and (N.3)′ are fulfilled from (N.4).

Assume that $\{a_n, n \geq 1\}$ is a positive constant sequence satisfying that $a_n \to 0$. For all $\theta \in \Theta \subset \mathcal{R}^k$, let $\phi_n(\theta) = a_n I_k$ and the conditions of Theorem 2.1 be fulfilled, where I_k is a $k \times k$ unit matrix. Then for all $\rho > 0$, taking $H = \frac{1}{a_n}\rho$ in (20), we obtain that

$$\sup_{\theta \in \Theta} P_\theta^{(n)} \left(|\hat{\theta}_n - \theta| > \rho \right) = \sup_{\theta \in \Theta} P_\theta^{(n)} \left(\frac{1}{a_n} |I_k \left(\hat{\theta}_n - \theta \right)| > \frac{\rho}{a_n} \right)$$
$$\leq B_0 \exp \left(-b_0 g_n(\rho/a_n) \right) \to 0,$$

in view of (*i*) of **G**. So the LS estimator $\hat{\theta}_n$ is a weak uniform consistency estimator of θ.

3 Some examples and simulations

In this section, some examples and simulations for the LS estimator of nonlinear regression models are illustrated.

Example 3.1 In the nonlinear model (1), let

$$f_t(\theta) = \frac{1}{\theta^{-1} + t^{1/4}}, \quad t = 1, 2, \ldots, n,$$

where $\theta \in \Theta = \{\theta : 0 < \delta_1 \leq \theta \leq \delta_2 < \infty\}$. Obviously, there exist some positive constants D_1 and D_2 such that, for all $\theta, \theta' \in \Theta$ and n large enough,

$$D_1 \left(\theta - \theta'\right)^2 \log n \leq \sum_{t=1}^{n} \left(f_t(\theta) - f_t(\theta')\right)^2 \leq D_2 \left(\theta - \theta'\right)^2 \log n,$$

where $D_1 < \frac{1}{\delta_2^4}$ and D_1 can be chosen arbitrarily close to $\frac{1}{\delta_2^4}$. Let the conditions of Theorem 2.2 hold. Then there exist some constants B_0 and b such that, for all n and H large enough,

$$\sup_{\theta \in \Theta} P_\theta^{(n)} \left\{ (\log n)^{1/2} |\hat{\theta}_n - \theta| \geq H \right\} \leq B_0 \exp(-bH^2), \tag{28}$$

where b can be chosen arbitrarily close (from below) to $\frac{1}{64r\delta_2^4}$.

If the errors $\{\varepsilon_t\}$ are $i.i.d.$ random variables and $\varepsilon_1 \sim N(0, \sigma^2)$, then by Theorem 5 of Wu (1981), it holds that

$$(\log n)^{1/2} (\hat{\theta}_n - \theta) \xrightarrow{\mathscr{L}} N(0, \sigma^2 \theta^4), \tag{29}$$

which yields

$$\lim_{\substack{H \to \infty \\ n \to \infty}} \left(-H^{-2} \log P_\theta^{(n)} \left\{ (\log n)^{1/2} |\hat{\theta}_n - \theta| \geq H \right\} \right) = \frac{1}{2\sigma^2 \theta^4}, \tag{30}$$

(see Example 1 of Sieders and Dzhaparidze 1987).
Moreover, by (28), we can get that

$$\liminf_{\substack{H \to \infty \\ n \to \infty}} \left(-H^{-2} \log P_\theta^{(n)} \left\{ (\log n)^{1/2} |\hat{\theta}_n - \theta| \geq H \right\} \right) \geq \frac{1}{64r\theta^4}. \tag{31}$$

By comparing (31) with (30), we obtain the large deviation result of (28) under END errors, which has the same order of optimal bound as that of independent case.

Example 3.2 Consider the linear model

$$X_t = \theta + \varepsilon_t, \quad t = 1, 2, \ldots, n, \ \theta \in \Theta \subset \mathscr{R},$$

where the errors $\{\varepsilon_t\}$ are mean zero END random variables satisfying (N.1)$'$. Applying Theorem 2.2 with $D_1 = D_2 = 1$ and $\phi_n(\theta) = n^{-1/2}$, we have that for all n and H large enough,

$$\sup_{\theta \in \Theta} P_\theta^{(n)} \left\{ n^{1/2} |\hat{\theta}_n - \theta| \geq H \right\} \leq B_0 \exp(-bH^2), \tag{32}$$

where B_0 and b are positive constants and b can be chosen arbitrarily close (from below) to $\frac{1}{64r}$. For all $\theta \in \Theta$, we take $H = \sqrt{\frac{2}{b}} \log^{\frac{1}{2}} n$ in (32) and obtain

$$\sum_{n=1}^{\infty} P_{\theta}^{(n)} \left\{ n^{1/2} |\hat{\theta}_n - \theta| \geq \sqrt{\frac{2}{b}} \log^{\frac{1}{2}} n \right\} \leq B_0 \sum_{n=1}^{\infty} \exp(-2 \log n) < \infty,$$

i.e.,

$$\hat{\theta}_n - \theta = O(n^{-\frac{1}{2}} \log^{\frac{1}{2}} n), \quad \text{completely, as } n \to \infty.$$

Example 3.3 Consider a power cure model

$$X_t = (t + \theta)^d + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{33}$$

where $d > 1/2$ and $\theta \in \Theta = \{\theta : 0 < \delta_1 \leq \theta \leq \delta_2 < \infty\}$. Let $f_t(\theta) = (t + \theta)^d$. Then there exist some positive constants D_1 and D_2 such that, for all $\theta, \theta' \in \Theta$,

$$D_1 n^{2d-1} \left(\theta - \theta' \right)^2 \leq \sum_{t=1}^{n} \left(f_t(\theta) - f_t(\theta') \right)^2 \leq D_2 n^{2d-1} \left(\theta - \theta' \right)^2.$$

Let the errors $\{\varepsilon_t\}$ be mean zero END random variables satisfying (N.1)'. Applying Theorem 2.2 with $\phi_n(\theta) = n^{1/2-d}$, we establish that for all n and H large enough,

$$\sup_{\theta \in \Theta} P_{\theta}^{(n)} \left\{ n^{d-1/2} |\hat{\theta}_n - \theta| \geq H \right\} \leq B_0 \exp(-bH^2), \tag{34}$$

where B_0 and b are positive constants and b can be chosen arbitrarily close (from below) to $\frac{D_1}{64r}$. For all $\theta \in \Theta$, taking $H = \sqrt{\frac{2}{b}} \log^{\frac{1}{2}} n$ in (34), we obtain that

$$\sum_{n=1}^{\infty} P_{\theta}^{(n)} \left\{ n^{d-\frac{1}{2}} |\hat{\theta}_n - \theta| \geq \sqrt{\frac{2}{b}} \log^{\frac{1}{2}} n \right\} \leq B_0 \sum_{n=1}^{\infty} \exp(-2 \log n) < \infty,$$

i.e.,

$$\hat{\theta}_n - \theta = O(n^{-(d-\frac{1}{2})} \log^{\frac{1}{2}} n), \quad \text{completely, as } n \to \infty,$$

where $d > 1/2$. Under the independent errors, Wu (1981) investigated the power cure model and obtained the strong consistency of the LS estimator $\hat{\theta}_n$ for θ. We extend the result of Wu (1981) to the END case and establish the complete convergence rate for the LS estimator $\hat{\theta}_n$ of θ.

Simulation 3.1 For simplicity, we do the simulation for the power cure model (33) with the case $d = 2$, i.e.,

$$X_t = (t + \theta)^2 + \varepsilon_t, \quad t = 1, 2, \ldots, n,$$

where $\theta \in \Theta = \{\theta : 0 < \delta_1 \leq \theta \leq \delta_2 < \infty\}$. Let $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ be a normal random vector such as $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \sim N_n(\mathbf{0}, \Sigma)$, where $\mathbf{0}$ is zero vector, Σ is

$$\Sigma = \begin{bmatrix} 1+\rho & -\rho & -\rho^2 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -\rho & 1+\rho & -\rho & -\rho^2 & \cdots & 0 & 0 & 0 & 0 \\ -\rho^2 & -\rho & 1+\rho & -\rho & \cdots & 0 & 0 & 0 & 0 \\ 0 & -\rho^2 & -\rho & 1+\rho & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho & -\rho & -\rho^2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1+\rho & -\rho & -\rho^2 \\ 0 & 0 & 0 & 0 & \cdots & -\rho^2 & -\rho & 1+\rho & -\rho \\ 0 & 0 & 0 & 0 & \cdots & 0 & -\rho^2 & -\rho & 1+\rho \end{bmatrix}_{n \times n},$$

for $0 < \rho < 1$. By Joag-Dev and Proschan (1983), it can be seen that $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is an NA vector. So it is also an END vector. By (14), the LS estimator $\hat{\theta}_n$ is that $\hat{\theta}_n = \arg\min_{\theta \in \Theta} \sum_{t=1}^n (X_t - (t+\theta)^2)^2$. Let $\frac{d(\sum_{t=1}^n (X_t - (t+\theta)^2)^2)}{d\theta} = 0$, then

$$n\theta^3 + \left(3\sum_{t=1}^n t\right)\theta^2 + \left(3\sum_{t=1}^n t^2 - \sum_{t=1}^n X_t\right)\theta + \sum_{t=1}^n t^3 - \sum_{t=1}^n t X_t = 0. \qquad (35)$$

It is a cubic equation of θ. By choosing the solutions of cubic equation, one can get $\hat{\theta}_n$. For $\theta = 1$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5$ and sample size $n = 10, 50, 100, 200$,
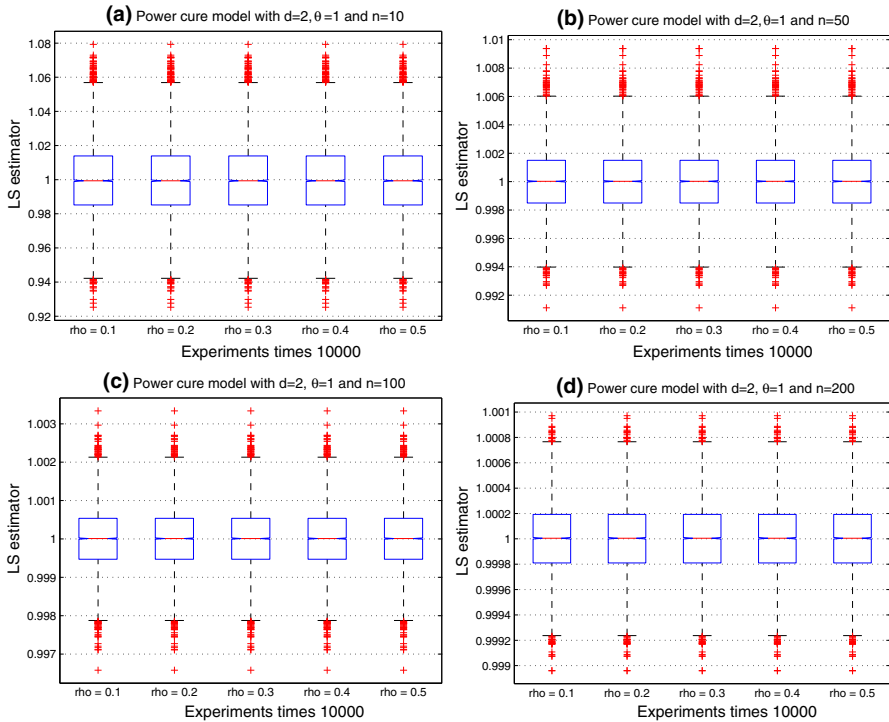

Fig. 1 Power cure model with $d = 2$, $\theta = 1$ and $n = 10, 50, 100, 200$, by 10000 times
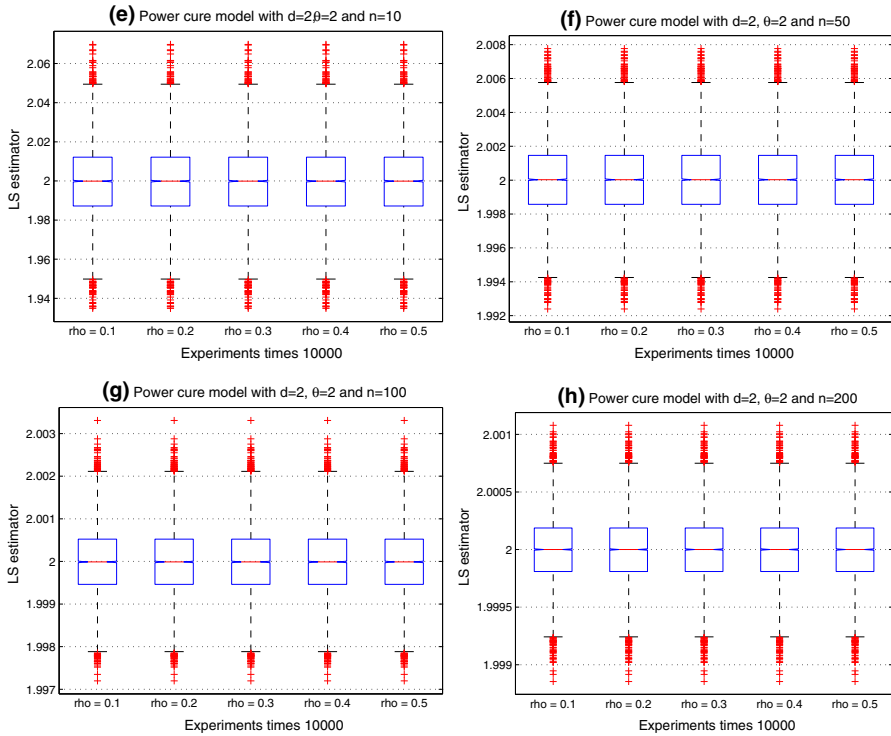
Fig. 2 Power cure model with $d = 2$, $\theta = 2$ and $n = 10, 50, 100, 200$, by 10000 times

we use MATLAB software to obtain the roots of cubic equation (35) by repeating the experiments 10000 times, and find that for each experiment, there are one real root and two complex roots. So we choose the real root as the LS estimator $\hat{\theta}_n$ and plot the Box plots as Fig. 1.

Similarly, for $\theta = 2$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5$ and sample size $n = 10, 50, 100, 200$, we do the simulation by repeating the experiments 10,000 times and plot the Box plots for LS estimator $\hat{\theta}_n$ as Fig. 2.

In Fig. 1a–d, with the same $\theta = 1$ but different $\rho = 0.1, \ldots, 0.5$, the median of LS estimator $\hat{\theta}_n$ is close to 1 and the variation range becomes smaller as the sample n increasing by 10, 50, 100 and 200. Likewise, in Fig. 2e–h, with the same $\theta = 2$ but different $\rho = 0.1, \ldots, 0.5$, the median of $\hat{\theta}_n$ is close to 2 and the variation range becomes smaller as the sample n increases.

We also give the Q–Q plot with $\theta = 1, 2$, $\rho = 0.1, 0.2$ and $n = 100$ by repeating the experiments 10,000 times, to test the normality of LS estimator $\hat{\theta}_n$ and obtain the Fig. 3. In Fig. 3, it can be found that the LS estimator $\hat{\theta}_n$ has a asymptotic normality based on this multivariate normal experiment.

4 Conclusion

In this paper, we investigate the LS estimator $\hat{\theta}_n$ of θ for the nonlinear model based on END errors which are not necessarily identically distributed. Under the general
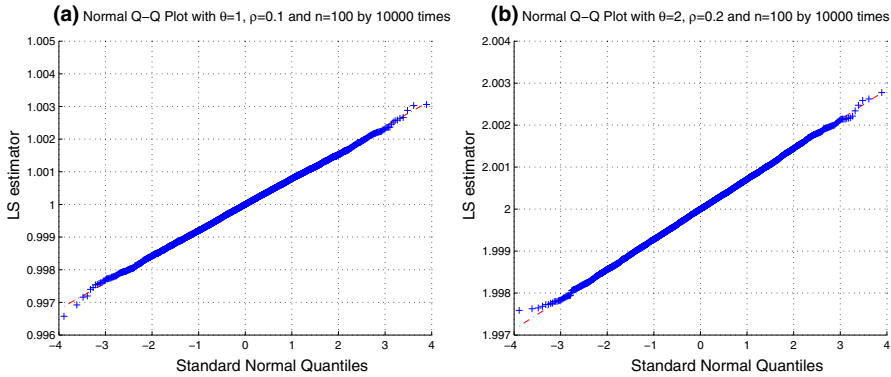
Fig. 3 Normal Q–Q plot with $\theta = 1, 2, \rho = 0.1, 0.2$ and $n = 100$, by 10000 times

conditions, we establish some large deviation results such as Theorems 2.1–2.4 for the LS estimator $\hat{\theta}_n$. As applications, by some simple conditions, a weak uniform consistency of $\hat{\theta}_n$ is established (see Remark 2.1), and a complete convergence rate of $\hat{\theta}_n$ is presented as $\hat{\theta}_n - \theta = O(n^{-1/2} \log^{1/2} n)$, completely, in Corollary 2.1. Some examples of nonlinear regression models and simulations are given to illustrate in Sect. 3. We extend the results of Sieders and Dzhaparidze (1987), Prakasa Rao (1984a) and Hu (1993) for the independence, Gaussian, locally generalized Gaussian and martingale differences to the case of END random variables. Since END random variables can be NOD random variables, NSD random variables and NA random variables, the results obtained in this paper also hold true for these dependent structures.

5 Proofs

Before proving our results, we give some technical preliminaries as follows.

Lemma 5.1 (cf. Liu 2010, Lemma 3.1) *Let random variables $\{Y_n, n \geq 1\}$ be a sequence of END random variables. We have that*

(1) *if $\{f_n, n \geq 1\}$ is a sequence of all nondecreasing (or nonincreasing) functions, then $\{f_n(Y_n), n \geq 1\}$ is also a sequence of END random variables;*

(2) *for each $n \geq 1$, there exists a positive constant M such that*

$$E\left(\prod_{i=1}^{n} Y_i^+\right) \leq M \prod_{i=1}^{n} EY_i^+.$$

Lemma 5.2 *Let $\{Y_n, n \geq 1\}$ be a sequence of END random variables and $\{r_n, n \geq 1\}$ be a sequence of positive numbers. For fixed $n \geq 1$, suppose that there exists a positive number Λ_1 such that*

$$E \exp\left(\lambda Y_i\right) \leq \exp\left(\frac{1}{2}r_i\lambda^2\right), \quad 0 \leq |\lambda| \leq \Lambda_1, \quad i = 1, 2, \ldots, n. \tag{36}$$

Denote $S_n = \sum_{i=1}^n Y_i$ and $G_n = \sum_{i=1}^n r_i$, $n \geq 1$. *Then there exists a positive constant M such that*

$$P(S_n \geq x) \leq \begin{cases} M \exp\left(-\frac{x^2}{2G_n}\right), & 0 \leq x \leq G_n \Lambda_1, \\ M \exp\left(-\frac{\Lambda_1 x}{2}\right), & x \geq G_n \Lambda_1, \end{cases} \tag{37}$$

and

$$P(S_n \leq -x) \leq \begin{cases} M \exp\left(-\frac{x^2}{2G_n}\right), & 0 \leq x \leq G_n \Lambda_1, \\ M \exp\left(-\frac{\Lambda_1 x}{2}\right), & x \geq G_n \Lambda_1. \end{cases} \tag{38}$$

Consequently,

$$P(|S_n| \geq x) \leq \begin{cases} 2M \exp\left(-\frac{x^2}{2G_n}\right), & 0 \leq x \leq G_n \Lambda_1, \\ 2M \exp\left(-\frac{\Lambda_1 x}{2}\right), & x \geq G_n \Lambda_1. \end{cases} \tag{39}$$

Proof For all x, by Markov's inequality, Lemma 5.1 and (36), we obtain that

$$P(S_n \geq x) \leq \exp(-\lambda x) E \exp(\lambda S_n) = \exp(-\lambda x) E\left(\prod_{i=1}^n \exp(\lambda Y_i)\right)$$

$$\leq M \exp(-\lambda x) \prod_{i=1}^n E \exp(\lambda Y_i)$$

$$\leq M \exp\left(\frac{G_n \lambda^2}{2} - \lambda x\right), \quad \text{for } 0 < \lambda \leq \Lambda_1.$$

So it has

$$P(S_n \geq x) \leq M \inf_{0 < \lambda \leq \Lambda_1} \exp\left(\frac{G_n \lambda^2}{2} - \lambda x\right) = M \exp\left(\inf_{0 < \lambda \leq \Lambda_1} \left(\frac{G_n \lambda^2}{2} - \lambda x\right)\right). \tag{40}$$

For the fixed $x \geq 0$, if $\Lambda_1 \geq \frac{x}{G_n} \geq 0$, then

$$\exp\left(\inf_{0 < \lambda \leq \Lambda_1} \left(\frac{G_n \lambda^2}{2} - \lambda x\right)\right) = \exp\left(-\frac{x^2}{2G_n}\right). \tag{41}$$

Meanwhile, for the fixed $x \geq 0$, if $\Lambda_1 \leq \frac{x}{G_n}$, $x \geq 0$, then

$$\exp\left(\inf_{0 < \lambda \leq \Lambda_1} \left(\frac{G_n \lambda^2}{2} - \lambda x\right)\right) = \exp\left(\frac{G_n \Lambda_1^2}{2} - \Lambda_1 x\right) \leq \exp\left(-\frac{\Lambda_1 x}{2}\right). \tag{42}$$

Consequently, (37) follows from (40) to (42) immediately.

According to Lemma 5.1 (1), $\{-Y_n\}$ are also END random variables. Therefore, by (37), it yields

$$P(S_n \leq -x) = P(-S_n \geq x) \leq \begin{cases} M \exp\left(-\frac{x^2}{2G_n}\right), & 0 \leq x \leq G_n \Lambda_1, \\ M \exp\left(-\frac{\Lambda_1 x}{2}\right), & x \geq G_n \Lambda_1, \end{cases}$$

which implies (38). Combining (37) with (38), we obtain (39) finally. □

Remark 5.1 Lemma 5.2 is an extension of exponential inequalities for the independent case (see Theorem 2.6 of Petrov 1995) and NOD case (see Theorem 2.1 of Wang et al. 2010) to the END structure case.

Corollary 5.1 *Let $\{Y_n, n \geq 1\}$ be a sequence of END random variables, $\{d_n, n \geq 1\}$ be a sequence of real numbers and $\{r_n, n \geq 1\}$ be a sequence of positive numbers. Suppose there exists a positive constant Λ_1 (Λ_1 possibly ∞) such that for all $|\lambda| \leq \Lambda_1$, (36) holds true. Denote $\tilde{S}_n = \sum_{i=1}^n d_i Y_i$, $\tilde{G}_n = \sum_{i=1}^n r_i d_i^2$ and $\Lambda = \Lambda_1 / \max_{1 \leq i \leq n} |d_i|$. Then for all $x \geq 0$, there exists a positive constant M such that*

$$P(\tilde{S}_n \geq x) \leq 2M \exp\left\{-\min\left(\frac{x^2}{8\tilde{G}_n}, \frac{\Lambda x}{4}\right)\right\}, \tag{43}$$

$$P(\tilde{S}_n \leq -x) \leq 2M \exp\left\{-\min\left(\frac{x^2}{8\tilde{G}_n}, \frac{\Lambda x}{4}\right)\right\}, \tag{44}$$

$$P(|\tilde{S}_n| \geq x) \leq 4M \exp\left\{-\min\left(\frac{x^2}{8\tilde{G}_n}, \frac{\Lambda x}{4}\right)\right\}. \tag{45}$$

Proof Obviously, for $|\lambda| \leq \Lambda = \Lambda_1 / \max_{1 \leq i \leq n} |d_i|$, we have $|\lambda d_i^+| \leq \Lambda |d_i| \leq \Lambda_1$. So by (36), it can be argued that

$$E \exp\left(\lambda d_i^+ Y_i\right) \leq \exp\left(\frac{1}{2} r_i \left(d_i^+\right)^2 \lambda^2\right) \leq \exp\left(\frac{1}{2} r_i d_i^2 \lambda^2\right), \text{ for } |\lambda| \leq \Lambda.$$

According to Lemma 5.1 (1), $d_1^+ Y_1, \ldots, d_n^+ Y_n$ are still END random variables. Denote $\tilde{S}_n(1) = \sum_{i=1}^n d_i^+ Y_i$ and $\tilde{G}_n = \sum_{i=1}^n r_i d_i^2$. Then, we apply Lemma 5.2 and establish that

$$P(\tilde{S}_n(1) \geq x) \leq M \exp\left\{-\min\left(\frac{x^2}{2\tilde{G}_n}, \frac{\Lambda x}{2}\right)\right\}. \tag{46}$$

Meanwhile, $d_1^- Y_1, \ldots, d_n^- Y_n$ are still END random variables. Denote $\tilde{S}_n(2) = \sum_{i=1}^n d_i^- Y_i$. Similar to the proof of (46), one has

$$P(\tilde{S}_n(2) \geq x) \leq M \exp\left\{-\min\left(\frac{x^2}{2\tilde{G}_n}, \frac{\Lambda x}{2}\right)\right\}. \tag{47}$$

Since

$$P(\tilde{S}_n \geq x) \leq P\left(\tilde{S}_n(1) \geq x/2\right) + P\left(\tilde{S}_n(2) \geq x/2\right), \tag{48}$$

by (46) to (48), we obtain the result (43). Combining with the proofs of (38), (39) and (43), we have the results of (44) and (45) immediately. □

Corollary 5.2 *Let $\{Y_n, n \geq 1\}$ be a sequence of END random variables satisfying $EY_i = 0$ and $EY_i^2 = \sigma_i^2 < \infty$, $i = 1, 2, \ldots$, $\{d_n, n \geq 1\}$ be a sequence of real*

numbers. Denote $\tilde{S}_n = \sum_{i=1}^n d_i Y_i$ *and* $\tilde{B}_n^2 = \sum_{i=1}^n \sigma_i^2 d_i^2$. *For the fixed* $n \geq 1$, *suppose that there exists a positive number L such that*

$$|EY_i^m| \leq \frac{m!}{2}\sigma_i^2 L^{m-2}, \quad i = 1, 2, \ldots, n \tag{49}$$

for all positive integers $m \geq 2$. *Then there exists a positive constant M such that for all* $x \geq 0$,

$$P(\tilde{S}_n \geq x) \leq 2M \exp\left\{-\min\left(\frac{x^2}{16\tilde{B}_n^2}, \frac{x}{8L \max_{1\leq i\leq n} |d_i|}\right)\right\}, \tag{50}$$

$$P(\tilde{S}_n \leq -x) \leq 2M \exp\left\{-\min\left(\frac{x^2}{16\tilde{B}_n^2}, \frac{x}{8L \max_{1\leq i\leq n} |d_i|}\right)\right\}, \tag{51}$$

$$P(|\tilde{S}_n| \geq x) \leq 4M \exp\left\{-\min\left(\frac{x^2}{16\tilde{B}_n^2}, \frac{x}{8L \max_{1\leq i\leq n} |d_i|}\right)\right\}. \tag{52}$$

Proof It can be argued by $EY_i = 0$, $EY_i^2 = \sigma_i^2$ and (49) that

$$E \exp(\lambda Y_i) = 1 + \frac{\lambda^2}{2}\sigma_i^2 + \frac{\lambda^3}{6}EY_i^3 + \cdots + \frac{\lambda^k}{k!}EY_i^k + \cdots$$

$$\leq 1 + \frac{\lambda^2}{2}\sigma_i^2\left(1 + L|\lambda| + L^2\lambda^2 + \cdots + L^{k-2}|\lambda|^{k-2} + \cdots\right), \quad i = 1, 2, \ldots, n.$$

If $|\lambda| \leq \frac{1}{2L}$, then

$$E \exp(\lambda Y_i) \leq 1 + \frac{\lambda^2\sigma_i^2}{2}\frac{1}{1-L|\lambda|} \leq 1 + \lambda^2\sigma_i^2 \leq \exp\left(\lambda^2\sigma_i^2\right) := \exp\left(\frac{1}{2}r_i\lambda^2\right), \tag{53}$$

where $r_i = 2\sigma_i^2$ *and* $i = 1, 2, \ldots, n$. *Taking* $\Lambda_1 = \frac{1}{2L}$ *and* $\tilde{G}_n = \sum_{i=1}^n r_i d_i^2 = 2\sum_{i=1}^n \sigma_i^2 d_i^2 = 2\tilde{B}_n^2$ *in Corollary 5.1, we have the results* (50)–(52) *immediately.* □

Lemma 5.3 *For some* $m \geq 2$, *let* $\{Y_n, n \geq 1\}$ *be a sequence of END random variables with* $EY_n = 0$ *and* $E|Y_n|^m < \infty$, $n = 1, 2, \ldots$. *Assume that* $\{a_{ni}, 1 \leq i \leq n, n \geq 1\}$ *is a triangular array of real numbers. Denote* $S_n = \sum_{i=1}^n a_{ni}Y_i$. *Then there exists a positive constant C not dependent on n such that*

$$E|S_n|^m \leq C \max_{1\leq i\leq n} E|Y_i|^m \left(\sum_{i=1}^n a_{ni}^2\right)^{m/2}. \tag{54}$$

Proof Denote $S_n(1) = \sum_{i=1}^n a_{ni}^+ Y_i$ and $S_n(2) = \sum_{i=1}^n a_{ni}^- Y_i$. For $m \geq 1$, by C_r inequality, one has

$$E|S_n|^m = E|S_n(1) - S_n(2)|^m \leq 2^{m-1}(E|S_n(1)|^m + E|S_n(2)|^m). \tag{55}$$

Obviously, by Lemma 5.1 (1), $\{a_{ni}^+ Y_i, 1 \le i \le n\}$ and $\{a_{ni}^- Y_i, 1 \le i \le n\}$ are also END random variables. Then, for $m \ge 2$, by Corollary 3.2 of Shen (2011), it yields that

$$
\begin{aligned}
E|S_n(1)|^m &\le C_1 \left(\sum_{i=1}^{n} (a_{ni}^+)^m E|Y_i|^m + \left(\sum_{i=1}^{n} (a_{ni}^+)^2 E Y_i^2 \right)^{m/2} \right) \\
&\le C_1 \max_{1 \le i \le n} E|Y_i|^m \left(\sum_{i=1}^{n} |a_{ni}|^m + \left(\sum_{i=1}^{n} a_{ni}^2 \right)^{m/2} \right) \\
&\le C_2 \max_{1 \le i \le n} E|Y_i|^m \left(\sum_{i=1}^{n} a_{ni}^2 \right)^{m/2}.
\end{aligned}
\tag{56}
$$

Similarly, it has

$$
E|S_n(2)|^m \le C_3 \max_{1 \le i \le n} E|Y_i|^m \left(\sum_{i=1}^{n} a_{ni}^2 \right)^{m/2}.
\tag{57}
$$

Thus, (54) follows from (55) to (57) immediately. □

Proof of Theorem 2.1 Let $\theta \in K$ and $u, v \in \Gamma_{n,\theta,R}$. If $|u - v| \ge k_1$, then we obtain by (18) that

$$
\sum_{t=1}^{n} (d_{tn\theta}(u) - d_{tn\theta}(v))^2 \le 2\mathrm{pol}(R) \le 2|u - v|^{2\rho} k_1^{-2\rho} \mathrm{pol}(R).
$$

By (17), it can be found that

$$
\sum_{t=1}^{n} (d_{tn\theta}(u) - d_{tn\theta}(v))^2 \le |u - v|^{2\rho} \mathrm{pol}(R), \quad \text{for all } u, v \in \Gamma_{n,\theta,R}.
\tag{58}
$$

Taking $\zeta_{n,\theta}(u) := \log Z_{n,\theta}(u)$ in (15), we establish that

$$
\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v) = \sum_{t=1}^{n} (A_t \varepsilon_t - B_t),
\tag{59}
$$

where

$$
A_t = d_{tn\theta}(u) - d_{tn\theta}(v), \quad B_t = \frac{1}{2}(d_{tn\theta}^2(u) - d_{tn\theta}^2(v)).
\tag{60}
$$

For all $m \ge 1$, by C_r inequality, it yields

$$
E_\theta^{(n)} |\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v)|^m \le 2^{m-1} \left(E_\theta^{(n)} \left| \sum_{t=1}^{n} A_t \varepsilon_t \right|^m + \left| \sum_{t=1}^{n} B_t \right|^m \right).
\tag{61}
$$

Obviously, the condition (N.1) implies that $\{E|\varepsilon_t|^m, t \in \mathcal{N}\}$ is uniformly bounded. So, by (58) and Lemma 5.3 with (N.1) and $E\varepsilon_t = 0$, we obtain that for all $m \geq 2$

$$E_\theta^{(n)} \left| \sum_{t=1}^n A_t \varepsilon_t \right|^m \leq C \left(\sum_{t=1}^n A_t^2 \right)^{m/2} \leq |u - v|^{\rho m} \mathrm{pol}(R). \tag{62}$$

Meanwhile, by Cauchy–Schwarz inequality, (18) and (58), one has that

$$\left| \sum_{t=1}^n B_t \right| \leq \frac{1}{2} \sum_{t=1}^n |d_{tn\theta}(u) - d_{tn\theta}(v)| \cdot |d_{tn\theta}(u) + d_{tn\theta}(v)|$$

$$\leq \frac{1}{2} \left\{ \sum_{t=1}^n A_t^2 \cdot \sum_{t=1}^n (d_{tn\theta}(u) + d_{tn\theta}(v))^2 \right\}^{1/2} \leq |u - v|^\rho \mathrm{pol}(R). \tag{63}$$

So it follows from (61)–(63) that for all $\theta \in K$ and $u, v \in \Gamma_{n,\theta,R}$

$$E_\theta^{(n)} |\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v)|^m \leq |u - v|^{\rho m} \mathrm{pol}(R). \tag{64}$$

Taking $m > \max(2, k/\rho)$ in (64), we find that (7) is fulfilled with $\alpha = \rho m$.

Next, we verify that (10) holds true. For $0 < \delta < \frac{1}{2}$, let

$$\eta_{n,\theta}(u) = \left(\frac{1}{2} - \delta \right) \sum_{t=1}^n d_{tn\theta}^2(u). \tag{65}$$

By (19), it can be argued that

$$\sum_{t=1}^n d_{tn\theta}^2(u) \geq 32 r g_n(R), \tag{66}$$

which shows that $\eta_{n,\theta}(u) \in \mathbf{H}_K$, because from (5), $g_n^{-1}(R) \leq 1$ for all n and R large enough. Denote $d_{tn\theta}(u) = d_t$ and $\max\limits_{1 \leq t \leq n} |d_{tn\theta}(u)| = \max |d_t|$. Then, by (19), (59), (60), (65), (66) and Corollary 5.1, we get

$$P_\theta^{(n)} \left(\zeta_{n,\theta}(u) - \zeta_{n,\theta}(0) \geq -\eta_{n,\theta}(u) \right) = P_\theta^{(n)} \left(\sum_{t=1}^n d_t \varepsilon_t \geq \delta \sum_{t=1}^n d_t^2 \right)$$

$$\leq 2M \exp \left\{ -\sum_{t=1}^n d_t^2 \min \left(\frac{\delta^2}{8r}, \frac{\delta \Lambda_1}{4 \max |d_t|} \right) \right\}$$

$$= 2M \exp \left\{ -\sum_{t=1}^n d_t^2 \Big/ \max \left(\frac{8r}{\delta^2}, \frac{4 \max |d_t|}{\delta \Lambda_1} \right) \right\}$$

$$\leq 2M \exp(-g_n(R)),$$

which implies that (10) is fulfilled. Combining Theorem 1.1 with Remark 1.2, we obtain (20). Meanwhile, by choosing $\alpha = \rho m$ in Theorem 1.1, for all $\beta > 0$ and m large enough, there exists a positive B_0 such that (20) holds, where $b_0 \geq \frac{\rho}{\rho+k} - \beta$. $\quad\square$

Proof of Theorem 2.2 By (21), (17) is fulfilled with $\rho = 1$ and $\mathrm{pol}(R) = D_2$. Obviously, for all $u \in \Gamma_{n,\theta,R}, \theta \in \Theta$, it has

$$\sum_{t=1}^{n} d_{tn\theta}^2 \leq D_2 |u|^2 \leq D_2 (R+1)^2, \tag{67}$$

which implies that (18) is satisfied. Meanwhile, for all $0 < \delta < 1/2$ and $u \in \Gamma_{n,\theta,R}$, it can be argued that

$$\sum_{t=1}^{n} d_{tn\theta}^2 \geq D_1 |u|^2 \geq D_1 R^2. \tag{68}$$

Then, by (68) and $\Lambda_1 = \infty$, (19) is fulfilled with $g_n(R) = \frac{D_1 \delta^2}{8r} R^2$. Letting $\delta \to \frac{1}{2}$, we apply Theorem 2.1 and obtain the result (22) finally. $\quad\square$

Proof of Theorem 2.3 Combining Corollary 5.2 with the proof of Theorem 2.1, where r is replaced by $2\sigma^2$ and Λ_1 is replaced by $\frac{1}{2L}$, we have (20) immediately. $\quad\square$

Proof of Theorem 2.4 Together (21) with (67), for all $u \in \Gamma_{n,\theta,R}, \theta \in \Theta$, it can be checked that

$$d_{tn\theta}^2 \leq \sum_{t=1}^{n} d_{tn\theta}^2 \leq D_2 |u|^2 \leq D_2 (R+1)^2, \quad 1 \leq t \leq n,$$

which implies

$$\max_{1 \leq t \leq n} |d_{tn\theta}| \leq \sqrt{D_2}(R+1). \tag{69}$$

Let $g_n(R) = C_1 R$, where C_1 is a positive constant and is defined later. Next, we prove that (23) in (N.3)$'$ is fulfilled for all R large enough. By (21), (68) and (69), for all $u \in \Gamma_{n,\theta,R}, \theta \in \Theta$, we take a positive constant C_1 such that for all R large enough,

$$\sum_{t=1}^{n} d_{tn\theta}^2 \geq D_1 R^2 \geq \frac{16\sigma^2}{\delta^2} C_1 R = \frac{16\sigma^2}{\delta^2} g_n(R) \tag{70}$$

and

$$\sum_{t=1}^{n} d_{tn\theta}^2 \geq D_1 R^2 \geq \frac{8L}{\delta} \sqrt{D_2}(R+1)C_1 R \geq \frac{8L}{\delta} \max_{1 \leq t \leq n} |d_{tn\theta}| g_n(R). \tag{71}$$

Thus, by (70) and (71), (23) is fulfilled. Consequently, by the proofs of Theorems 2.1 and 2.3, we apply Theorem 2.3 and establish that for all n and H large enough,

$$\sup_{\theta \in \Theta} P_{\theta}^{(n)} \left\{ |\phi_n^{-1}(\theta)(\hat{\theta}_n - \theta)| \geq H \right\} \leq B_0 \exp(-C_0 H), \tag{72}$$

where $C_0 = b_0 C_1$, B_0 and b_0 are defined in Theorem 2.3. □

Proof of Corollary 2.1 For all $\theta \in \Theta \subset \mathscr{R}^k$ and all $\rho > 0$, taking $\phi_n(\theta) = n^{-1/2} I_k$ and $H = n^{1/2} \rho$ in (22), one establishes the result (26) immediately, where I_k is a $k \times k$ unit matrix. Taking C_1 large enough, we apply (26) and establish that

$$\sum_{n=1}^{\infty} P_\theta^{(n)} \left\{ |\hat{\theta}_n - \theta| \geq C_1 n^{-\frac{1}{2}} \log^{\frac{1}{2}} n \right\} \leq B_0 \sum_{n=1}^{\infty} \exp(-b C_1^2 \log n) < \infty.$$

This completes the proof of (27). □

References

Bulinski AV, Shaskin A (2007) Limit theorems for associated random fields and related systems. World Scientific, Singapore

Bunke H, Schmidt WH (1980) Asymptotic results on nonlinear approximation of regression functions and weighted least squares. Math Oper Stat Ser Stat 11(1):3–32

Chen YQ, Chen AY, NG KW (2010) The strong law of large numbers for extended negatively dependent random variables. J Appl Prob 47(4):908–922

Hu SH (1993) A large deviation result for the least squares estimators in nonlinear regression. Stoch Process Appl 47(2):345–352

Hu SH (2002) The rate of convergence for the least squares estimator in nonlinear regression model with dependent errors. Sci China Ser A 45(2):137–146

Hu SH (2004) Consistency for the least squares estimator in nonlinear regression model. Stat Probab Lett 67(2):183–192

Hu T-C, Rosalsky A, Wang KL (2015) Complete convergence theorems for extended negatively dependent random variables. Sankhya 77A(1):1–29

Hu TZ (2000) Negatively superadditive dependence of random variables with applications. Chin J Appl Probab Stat 16(2):133–144

Ibragimov IA, Has'minskii RZ (1981) Statistical estimation: asymptotic theory. Springer, New York Translated by Samuel Kotz

Ivanov AV (1976) An asymptotic expansion for the distribution of the least squares estimator of the nonlinear regression parameter. Theory Probab Appl 21(3):557–570

Ivanov AV (1997) Asymptotic theory of nonlinear regression. Kluwer Academic Publishers, Dordreht

Ivanov AV, Leonenko NN (1989) Statistical analysis of random fields. Kluwer Academic Publishers, Dordreht

Ivanov AV, Leonenko NN (2008) Semiparametric analysis of long-range dependence in nonlinear regression. J Stati Plan Inference 138(6):1733–1753

Jennrich RI (1969) Asymptotic properties of nonlinear least squares estimators. Ann Math Stat 40(2):633–643

Joag-Dev K, Proschan F (1983) Negative association of random variables with applications. Ann Stat 11(1):286–295

Ko B, Tang Q (2008) Sums of dependent nonnegative random variables with subexponential tails. J Appl Probab 45(1):85–94

Lehmann EL (1966) Some concepts of dependence. Ann Math Stat 37(5):1137–1153

Liu L (2009) Precise large deviations for dependent random variables with heavy tails. Stat Probab Lett 79(9):1290–1298

Liu L (2010) Necessary and sufficient conditions for moderate deviations of dependent random variables with heavy tails. Sci China Ser A 53(6):1421–434

Malinvaud E (1970) The consistency of nonlinear regression. Ann Math Stat 41(3):956–969

Midi H (1999) Preliminary estimators for robust non-linear regression estimation. J Appl Stat 26(5):591–600

Oliveira PE (2012) Asymptotics for associated random variables. Springer, Berlin

Petrov VV (1995) Limit theorems of probability theory: sequences of independent random variables. Clarendon Press, Oxford

Prakasa Rao BLS (1984a) On the exponential rate of convergence of the least squares estimator in the nonlinear regression model with Gaussian errors. Stat Probab Lett 2(3):139–142

Prakasa Rao BLS (1984b) The rate of convergence of the least squares estimator in a non-linear regression model with dependent errors. J Multivar Anal 14(3):315–322

Prakasa Rao BLS (2012) Associated sequences, demimartingales and nonparametric inference. Birkhäuser, Springer, Basel

Shen AT (2011) Probability inequalities for END sequence and their applications. J Inequal Appl 2011 2011(98):12

Sieders A, Dzhaparidze K (1987) A large deviation result for parameter estimators and its application to nonlinear regression analysis. Ann Stat 15(3):1031–1049

Wang XJ, Hu SH, Yang WZ, Ling NX (2010) Exponential inequalities and inverse moment for NOD sequence. Stat Probab Lett 80(5–6):452–461

Wang XJ, Hu T-C, Volodin AI, Hu SH (2013) Complete convergence for weighted sums and arrays of rowwise extended negatively dependent random variables. Commun Stat Theory Methods 42(13):2391–2401

Wang XJ, Shen AT, Chen ZY, Hu SH (2015a) Complete convergence for weighted sums of NSD random variables and its application in the EV regression model. Test 24(1):166–184

Wang XJ, Zheng LL, Xu C, Hu SH (2015b) Complete consistency for the estimator of nonparametric regression models based on extended negatively dependent errors. Statistics 49(2):396–407

Wu CF (1981) Asymptotic theory of nonlinear least squares estimation. Ann Stat 9(3):501–513

Yang WZ, Hu SH (2014) Large deviation for a least squares estimator in a nonlinear regression model. Stat Probab Lett 91:135–144