



Large deviations for randomly weighted least squares estimator in a nonlinear regression model

Yi Wu¹ · Wei Yu² · Xuejun Wang²

Received: 21 September 2021 / Accepted: 8 September 2023 / Published online: 4 October 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

In this work, we introduce the random weighting method to the nonlinear regression model and study the asymptotic properties for the randomly weighted least squares estimator with dependent errors. The results reveal that this new estimator is consistent. Moreover, some simulations are also carried out to show the performance of the proposed estimator.

Keywords Random weighting method · Least squares estimator · Nonlinear regression model · Dependent errors

Mathematics Subject Classification 62J02 · 62F12

1 Introduction

Consider the nonlinear regression model:

$$X_n = g_n(\theta) + \xi_n, \quad n \geq 1, \quad (1.1)$$

where X_n is observed, $\{g_n(\theta)\}$ is a known sequence of continuous functions possibly nonlinear in $\theta \in \Theta$, a closed interval on the real line, and $\{\xi_n, n \geq 1\}$ is a sequence of random errors with zero mean. The nonlinear regression models have significant advantages over the linear models. The main one is that the nonlinear regression models have essentially fewer unknown parameters. Also, the parameters of nonlinear models have the meaning of physical variables while the linear parameters are usually

✉ Xuejun Wang
wangxj@ahu.edu.cn

¹ School of Big Data and Artificial Intelligence, Chizhou University, Chizhou 247000, People's Republic of China

² School of Big Data and Statistics, Anhui University, Hefei 230601, People's Republic of China

devoid of physical significance. Therefore, it is of great interest to study the nonlinear regression model. In most of studies devoted to the problems of regression analysis in the past decades, the central place is occupied by the least squares method of estimation of parameters, which has a protracted history. Let

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \omega_i^2 (X_i - g_i(\theta))^2,$$

where $\{\omega_i\}$ is a known sequence of positive numbers. An estimator θ_n is said to be an ordinary least squares estimator (OLSE, for short) of θ if it minimizes $Q_n(\theta)$, that is, $Q_n(\theta_n) = \inf_{\theta \in \Theta} Q_n(\theta)$.

The study of asymptotic properties of the OLSE for parameters in nonlinear regression models has been the main subject of investigation. It is challenging in investigating this model since the OLSE non-linearly entering in a regression function parameters can not be found in an explicit form, which complicates the description of its mathematical properties. Hence, on introducing into statistics the use of nonlinear regression analysis it is necessary to overcome a series of mathematical difficulties which do not have analogues in the linear theory. For the OLSE of the nonlinear model based on i.i.d. random errors, Jennrich (1969) established the asymptotic normality, Malinvaud (1970) investigated the consistency, and Wu (1981) established the necessary and sufficient condition for the strong consistency, and so on. In particular, Ivanov (1976) obtained the following result on large deviation for the OLSE with $\omega_i \equiv 1$ based on i.i.d. random errors.

Theorem 1.1 *Let $\{\xi_n, n \geq 1\}$ be i.i.d. random variables with $E|\xi_1|^p < \infty$ for some $p \geq 2$. Suppose there exist some constants $0 < c_1 \leq c_2 < \infty$ such that*

$$c_1(\theta_1 - \theta_2)^2 \leq \frac{1}{n} \sum_{i=1}^n (g_n(\theta_1) - g_n(\theta_2))^2 \leq c_2(\theta_1 - \theta_2)^2, \text{ for any } \theta_1, \theta_2 \in \Theta, n \geq 1.$$

Then for every $\rho > 0$ and for all $n \geq 1$, it has

$$P(n^{1/2}|\theta_n - \theta_0| > \rho) \leq c\rho^{-p},$$

where θ_0 is the true parameter such that $\theta_0 \in \text{interior of } \Theta$ and c is a positive constant independent of n and p .

Prakasa Rao (1984) extended Theorem 1.1 from i.i.d. case to some dependent cases such as φ -mixing and α -mixing assumptions. Hu (2002) extended Theorem 1.1 to martingale differences, φ -mixing and negative association (NA, for short) assumptions under $\sup_{i \geq 1} E|\xi_i|^p < \infty$ for some $p > 2$ without identical distribution. Hu (2004) further considered the large deviation result under the moment condition $\sup_{i \geq 1} E|\xi_i|^p < \infty$ for some $1 < p \leq 2$. Yang and Hu (2014) obtained some general results on the large deviation, which can also be available under some cases satisfying $\sup_{i \geq 1} E|\xi_i|^p = \infty$ for some $p > 1$; Yang et al. (2017) established some large deviation results under extended negatively dependent (END, for short) random errors, and

so on. However, a new challenge will emerge if the errors are heteroscedastic, i.e., estimating the variances of the errors is not an easy work.

It is well known that the bootstrap is a very excellent method, which has been used comprehensively in many statistical models including nonlinear regression model. One can see Staniewski (1984) for example. As an alternative approach, the random weighting method, or Bayesian bootstrap method, has received increasing attentions by scholars since it was originally suggested by Rubin (1981). The random weighting method is motivated by the bootstrap method and can be regarded as a kind of smoothing of bootstrap. Instead of re-sampling from the original data set, the random weighting method propels us to generate a group of random weights directly from the computer and then use them to weight the original samples. In comparison with the bootstrap method, the random weighting method has advantages such as the simplicity in computation, the suitability for large samples, and there is no need to know the distribution function. Therefore, this method was adopted in various statistical models. For more details, we refer the readers to Zheng (1987), Gao et al. (2003), Xue and Zhu (2005), Fang and Zhao (2006), Barvinok and Samorodnitsky (2007), Gao and Zhong (2010), and so forth.

However, to the best of our knowledge, there is no literature considering the randomly weighted estimation in nonlinear regression models. In this paper, the random weighting method is adopted for the first time to the least squares estimation in nonlinear regression models. Now we are at a position to present this method.

Definition 1.1 (cf. Ng et al. 2011) Let (W_1, \dots, W_n) be a random vector with $W_i \geq 0$ and $\sum_{i=1}^n W_i = 1$. Then the Dirichlet probability density function of (W_1, \dots, W_n) is defined as

$$f(w_1, \dots, w_n) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n w_i^{\alpha_i-1},$$

where $\alpha_i > 0, \alpha_0 = \sum_{i=1}^n \alpha_i, w_i \geq 0, \sum_{i=1}^{n-1} w_i \leq 1$ and $w_n = 1 - \sum_{i=1}^{n-1} w_i$. This distribution is denoted by $Dir(\alpha_1, \dots, \alpha_n)$.

By virtue of the concept of Dirichlet distribution, we can propose the randomly weighted least squares estimator of θ as follows. Let

$$H_n(\theta) = \sum_{i=1}^n W_i (X_i - g_i(\theta))^2, \tag{1.2}$$

where W_i 's are independent of ξ_i 's and the random vector $\mathbf{W} = (W_1, \dots, W_n)$ obeys the Dirichlet distribution $Dir(4, 4, \dots, 4)$, namely, $\sum_{i=1}^n W_i = 1$ and the joint density of W_1, \dots, W_{n-1} is

$$f(w_1, \dots, w_{n-1}) = \frac{\Gamma(4n)}{(\Gamma(4))^n} w_1^3 \dots w_{n-1}^3 (1 - w_1 - \dots - w_{n-1})^3,$$

where $(w_1, \dots, w_{n-1}) \in D_{n-1}$ and $D_{n-1} = \{(w_1, \dots, w_{n-1}) : w_i \geq 0, i = 1, \dots, n-1, \sum_{i=1}^{n-1} w_i \leq 1\}$. An estimator $\hat{\theta}_n$ is said to be a randomly weighted least squares estimator (RWLSE, for short) of θ if $\hat{\theta}_n = \arg \inf_{\theta \in \Theta} H_n(\theta)$.

Since independence assumption is usually implausible in reality, we will adopt a relatively broad dependence, i.e., END assumption in the sequel. The concept of END random variables was introduced by Liu (2009) as follows.

Definition 1.2 A finite collection of random variables X_1, X_2, \dots, X_n is said to be END if there exists a constant $M > 0$ such that both

$$P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n) \leq M \prod_{i=1}^n P(X_i > x_i)$$

and

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \leq M \prod_{i=1}^n P(X_i \leq x_i)$$

hold for all real numbers x_1, x_2, \dots, x_n . An infinite sequence $\{X_n, n \geq 1\}$ is said to be END if every finite sub-collection is END.

Liu (2009) provided some examples satisfying END structure, one of which shows that if X_1, X_2, \dots, X_n are dependent according to a multivariate copula function $C(u_1, \dots, u_n)$ with absolutely continuous distribution functions F_1, \dots, F_n , the joint copula density $c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}$ exists and be uniformly bounded in the whole domain, then $\{X_n, n \geq 1\}$ are END. If we take $M = 1$, then the END structure degenerates to negatively orthant dependent (NOD, for short) structure which was introduced by Lehmann (1966) (cf. also Joag-Dev and Proschan 1983). The END structure can reflect not only a negative dependence structure but also a positive one to some extent. Liu (2009) pointed out that the END random variables can be regarded as negatively or positively dependent and provided some interesting examples to support this idea. Joag-Dev and Proschan (1983) also pointed out that negatively associated (NA, for short) random variables are NOD but the inverse is not necessarily true, thus NA random variables are also END. Hence, the consideration of END structure is reasonable and of great interest. Many applications have been found for END random variables. For example, Liu (2010) studied the sufficient and necessary conditions of moderate deviations for END random variables with heavy tails; Chen et al. (2010) established the strong law of large numbers for END random variables and gave their applications to risk theory and renewal theory; Shen (2011) established some exponential probability inequalities for END random variables and presented some applications; Wang and Wang (2013) investigated the precise large deviations for random sums of END real-valued random variables with consistent variation; Wang et al. (2014) proved some results on complete convergence of END random variables; Lita da Silva (2015) established the almost sure convergence for sequences of END random variables; Wang et al. (2015) and Yang et al. (2018) studied the complete consistency of the estimator of nonparametric regression models based on END errors;

Wu et al. (2019) investigated the complete f -moment convergence for END random variables, and so on.

For the proposed RWLSE above-mentioned, we establish two general results on large deviation for RWLSE of the parameter θ with $p > 2$, and respectively, $1 < p \leq 2$ under END errors. As direct corollaries, the rates of complete consistency, strong consistency, and weak consistency are obtained, which reflect that the proposed RWLSE is a consistent estimator of θ . The numerical analysis reveals that the RWLSE performs as well as the OLSE in heteroscedastic nonlinear regression models, where sometimes the former one is better. As we have pointed out earlier, it is not easy to estimate the variances of heteroscedastic errors, this paper provides an alternative method to estimate the parameters in a heteroscedastic nonlinear regression model.

Throughout this paper, the symbol C represents some positive constant which can be different in different places. $C(p), C'(p), C_1(p), C_2(p), \dots$ are some positive constants depending only on p . Let $I(A)$ be the indicator function of the event A and $\lfloor x \rfloor$ denote the integer part of x . Denote $x^+ = xI(x \geq 0), x^- = -xI(x < 0)$. $\log n = \ln \max(x, e)$, where $\ln x$ represents the natural logarithm of x .

The rest of this paper is organized as follows: The main results are stated in Sect. 2. The numerical analysis is provided in Sect. 3. The proofs of the main results are presented in Sect. 4. Some lemmas for proving the main results are given in Appendix.

2 Main results

The main results on large deviations are presented as follows.

Theorem 2.1 *Let $p > 2$. In model (1.1), assume that $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $E|\xi_n|^p < \infty$ for each $n \geq 1$. If there exist positive numbers $\lambda_n \leq \Lambda_n$ for each $n \geq 1$, such that*

$$\lambda_n |\theta_1 - \theta_2| \leq |g_i(\theta_1) - g_i(\theta_2)| \leq \Lambda_n |\theta_1 - \theta_2|, \text{ for any } \theta_1, \theta_2 \in \Theta, 1 \leq i \leq n, n \geq 1, \tag{2.1}$$

then there exists a positive constant $C(p)$ depending only on p such that for all $\rho > 0$ and each $n \geq 1$,

$$P(n^{1/2} |\hat{\theta}_n - \theta_0| > \rho) \leq C(p) \left[n^{-1} \lambda_n^{-p} \Delta_{np} + n^{-p/2} (\Lambda_n / \lambda_n^2)^p (\Delta_{np} + \nabla_{np}) \right] \rho^{-p}, \tag{2.2}$$

where $\Delta_{np} = \sum_{i=1}^n E|\xi_i|^p$ and $\nabla_{np} = (\sum_{i=1}^n E\xi_i^2)^{p/2}$.

Theorem 2.2 *Let $1 < p \leq 2$. In model (1.1), assume that $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $E|\xi_n|^p < \infty$ for each $n \geq 1$. If (2.1) holds, then there exists a positive constant $C'(p)$ depending only on p such that for all $\rho > 0$ and each $n \geq 1$,*

$$P(n^{1/2}|\hat{\theta}_n - \theta_0| > \rho) \leq C'(p)(\Lambda_n/\lambda_n^2)^p n^{-p/2} \Delta_{np} \rho^{-p}. \tag{2.3}$$

Remark 2.1 It is easy to see that if $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $p > 2$, Theorem 2.1 extends Theorem 1.1 from i.i.d assumption to END random errors with not necessarily identical distribution. Similarly, if $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $1 < p \leq 2$, then Theorem 2.2 also extends the corresponding result of Hu (2004).

Remark 2.2 Yang and Hu (2014) also established the similar results for OLSE of θ with NOD errors. By taking $\lambda_n = c_1, \Lambda_n = c_2$ for some $0 < c_1 \leq c_2$, we point out that the conditions in Theorem 2.1 and the corresponding result of Yang and Hu (2014) do not imply each other. For example, $n^{-1} \sum_{i=1}^n E|\xi_i|^p > n^{-p/2} \sum_{i=1}^n E|\xi_i|^p$ but $n^{-p/2} (\sum_{i=1}^n E\xi_i^2)^{p/2} \leq n^{-p/2} (\sum_{i=1}^n (E|\xi_i|^p)^{2/p})^{p/2}$. However, if $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $p > 2$, they are equivalent. Hence, our results extend the corresponding ones of Yang and Hu (2014).

By Theorem 2.1, we can obtain the result concerning the rate of complete consistency and strong consistency as follows.

Corollary 2.1 *In model (1.1), assume that $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $p > 2$. If (2.1) holds with $\lambda_n = c_1$ and $\Lambda_n = c_2$ for some $0 < c_1 \leq c_2 < \infty$, then for any $\epsilon > 0$,*

$$\sum_{n=1}^{\infty} P(|\hat{\theta}_n - \theta_0| > \epsilon n^{1/p-1/2} \sqrt{\log n}) < \infty,$$

and thus

$$|\hat{\theta}_n - \theta_0| = o\left(n^{1/p-1/2} \sqrt{\log n}\right) \text{ a.s., as } n \rightarrow \infty.$$

By Theorem 2.2, we can also obtain the following result on rate of weak consistency of the RWLSE $\hat{\theta}_n$.

Corollary 2.2 *In model (1.1), assume that $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $1 < p \leq 2$. If (2.1) holds with $\lambda_n = c_1$ and $\Lambda_n = c_2$ for some $0 < c_1 \leq c_2 < \infty$, then*

$$|\hat{\theta}_n - \theta_0| = O_P\left(n^{1/p-1}\right).$$

In particular, if $p = 2$, then for any positive sequence $\{\tau_n, n \geq 1\}$ satisfying $\tau_n = o(n)$,

$$\sqrt{\tau_n}|\hat{\theta}_n - \theta_0| \xrightarrow{P} 0.$$

3 Some examples and numerical analysis

3.1 Some examples

In this subsection, we present some examples for the RWLSE of nonlinear regression models.

Example 3.1 Consider the linear model

$$X_i = \theta + \xi_i, \quad i = 1, 2, \dots, n, \quad \theta \in \Theta, \tag{3.4}$$

where $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $E|\xi_n|^p < \infty$ for each $n \geq 1$. Obviously, (2.1) holds with $\lambda_n = \Lambda_n = 1$. Hence, Theorems 2.1 and 2.2 follows from $E|\xi_n|^p < \infty$ with $p > 2$ and respectively, $1 < p \leq 2$.

Example 3.2 Consider the Michaelis-Menten model (see Sieders and Dzhaparidze (1987)) or Miao and Tang (2021) for example)

$$V(v, L, N) = \frac{Lv}{N + v},$$

which is used to describe the relation between the velocity V of an enzyme reaction and the concentration v of the substrate. The parameter L denotes the maximal reaction velocity and the parameter N implies the chemical affinity. Based on the model above, for each concentration v_i , there is a measurement of the velocity V_i with error ξ_i , i.e.,

$$X_i = V_i(v_i, L, N) = \frac{Lv_i}{N + v_i} + \xi_i, \quad i \geq 1. \tag{3.5}$$

Assume that the parameter set $(L, N) \in \Theta$ is a bounded open set in the positive quadrant. Consider the following simple form of model (3.5)

$$X_i = g_i(N) + \xi_i = \frac{1}{N^{-1} + i^\mu} + \xi_i, \tag{3.6}$$

which follows from (3.5) by assuming N/L is known (without loss of generality, we may assume that $N/L = 1$) and letting $v_i = i^{-\mu}$, $0 < \mu < \min\{(p - 1)/(4p), 1/8\}$, where $p > 1$. It is easy to see that

$$\begin{aligned} c_3 n^{-2\mu} |N_1 - N_2| &\leq |g_i(N_1) - g_i(N_2)| = \frac{|N_1 - N_2|}{(1 + N_1 i^\mu)(1 + N_2 i^\mu)} \\ &\leq c_4 |N_1 - N_2|, \quad 1 \leq i \leq n \end{aligned}$$

for some $0 < c_3 \leq c_4 < \infty$. Assume further that $\{\xi_n, n \geq 1\}$ is a sequence of END random errors with zero mean and $E|\xi_n|^p < \infty$ for each $n \geq 1$, then Theorems 2.1 and 2.2 hold. Moreover, by choosing $\rho = n^{1/2}\epsilon$, we can obtain the weak consistency for $p > 1$ and strong consistency for some p large enough.

Table 1 Mean and variance of the RWLSE under END errors for homoscedastic Models 1 and 2

sample size	Model 1		Model 2	
	$E(\hat{\theta})$	$Var[\sqrt{n}(\hat{\theta} - \theta)]$	$E(\hat{N})$	$Var[\sqrt{n}(\hat{N} - N)]$
50	1.0011	0.6651	0.2213	1.8789
100	1.0016	0.6444	0.2237	1.7445
200	0.9981	0.6740	0.2117	1.7273
400	1.0011	0.5743	0.2000	1.8365
800	1.0016	0.6985	0.2024	1.7128
1600	0.9991	0.6811	0.2004	1.8366

3.2 Numerical analysis

In this section, we will carry out some simulations to study the finite sample performance of the RWLSE in the homoscedastic and heteroscedastic nonlinear regression models. The data are generated from model (3.4) (denoted as Model 1) and (3.6) (denoted as Model 2) respectively. For Model 1, set $\theta = 1$ and for Model 2, set $N = 1/5$. Set the sample size $n = 50, 100, 200, 400, 800, 1600$. Let $(\epsilon_1, \dots, \epsilon_n) \sim N_n(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & -0.3 & 0 & \dots & 0 \\ -0.3 & 1 & -0.3 & \ddots & \vdots \\ 0 & -0.3 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -0.3 \\ 0 & \dots & 0 & -0.3 & 1 \end{pmatrix}. \tag{3.7}$$

The weights $W \sim Dir(4, 4, \dots, 4)$, and the generation method of W is refer to Narayanan (1990).

We first use the RWLSE to estimate θ for Model 1 and N for Model 2 with homoscedasticity, i.e., $\xi_i = \epsilon_i$ for each $1 \leq i \leq n$. Repeat the procedure 1000 times and calculate the mean and variance of the estimator. The results are given in Table 1. We can see that $\hat{\theta}$ for Model 1 is unbiased, while \hat{N} for Model 2 is asymptotically unbiased. The trend that $Var[\sqrt{n}(\hat{\theta} - \theta)]$ and $Var[\sqrt{n}(\hat{N} - N)]$ are finite indicates that the convergence rate of the RWLSE is asymptotically $O(n^{-1/2})$. To compare the RWLSE to the OLSE under END errors, we further present the the mean and variance of the OLSE in Table 2. The results show that there are no intrinsic difference between the mean of the two estimators. The mean and the variance of the RWLSE are slightly inferior to those of the OLSE in both the two models.

We now consider the heteroscedastic case, i.e., $\xi_i = [1 + \frac{(-1)^i(i-1)}{n}] \epsilon_i$ for each $1 \leq i \leq n$. Other settings are the same as above. The results are given in Table 3 and Table 4. The mean and the variance of the RWLSE are better than those of the OLSE in Model 1 but slightly weaker in Model 2. However, the convergence rates of the two estimators are almost the same. Note that in our simulation, the heteroscedasticity is known. However, in many realistic applications, it is not easy to estimate the variance

Table 2 Mean and variance of the OLSE under END errors for homoscedastic Models 1 and 2

sample size	Model 1		Model 2	
	$E(\hat{\theta})$	$Var[\sqrt{n}(\hat{\theta} - \theta)]$	$E(\hat{N})$	$Var[\sqrt{n}(\hat{N} - N)]$
50	1.0001	0.4477	0.2229	1.1344
100	1.0030	0.4202	0.2037	0.9986
200	0.9995	0.4125	0.2065	1.1544
400	1.0003	0.4301	0.2023	1.0640
800	1.0007	0.3789	0.2020	0.9500
1600	1.0004	0.3770	0.2005	1.1707

Table 3 Mean and variance of the RWLSE under END errors for heteroscedastic Models 1 and 2

sample size	Model 1		Model 2	
	$E(\hat{\theta})$	$Var[\sqrt{n}(\hat{\theta} - \theta)]$	$E(\hat{N})$	$Var[\sqrt{n}(\hat{N} - N)]$
50	0.9867	1.2515	0.2811	6.2218
100	0.9975	1.2795	0.2273	4.4603
200	1.0038	1.2632	0.2202	3.6783
400	1.0015	1.5081	0.2078	4.3282
800	0.9989	1.4841	0.2078	4.3502
1600	1.0006	1.5372	0.2035	4.4674

Table 4 Mean and variance of the OLSE under END errors for heteroscedastic Models 1 and 2

sample size	Model 1		Model 2	
	$E(\hat{\theta})$	$Var[\sqrt{n}(\hat{\theta} - \theta)]$	$E(\hat{N})$	$Var[\sqrt{n}(\hat{N} - N)]$
50	1.0074	2.1632	0.2464	2.8356
100	1.0065	2.5308	0.2251	2.8470
200	0.9974	3.4748	0.2107	2.6467
400	1.0027	3.7136	0.2088	2.8344
800	1.0015	3.3483	0.2013	2.7371
1600	1.0021	4.7489	0.2039	2.6418

of the errors if they are heteroscedastic. Therefore, our simulation results show that the RWLSE has a good performance without estimating the variance of the errors first, which provide us an alternative choice when dealing with similar issues.

4 Proofs of the main results

Proof of Theorem 2.1 Denote

$$\Psi_n(\theta_1, \theta_2) = \sum_{i=1}^n W_i (g_i(\theta_1) - g_i(\theta_2))^2, \quad V_n(\theta) = \sum_{i=1}^n W_i \xi_i (g_i(\theta) - g_i(\theta_0)),$$

and

$$U_n(\theta) = \frac{V_n(\theta)}{\Psi_n(\theta, \theta_0)}, \theta \neq \theta_0.$$

Note from $\sum_{i=1}^n W_i = 1$ and (2.1) that

$$\Psi_n(\theta, \theta_0) \geq \lambda_n^2(\theta - \theta_0)^2. \tag{4.1}$$

For all $\omega \in (|\hat{\theta}_n - \theta_0| > \varepsilon)$, where $\varepsilon > 0$ is arbitrary, we have that $\hat{\theta}_n \neq \theta_0$ and thus

$$\begin{aligned} \sum_{i=1}^n W_i \xi_i^2 &= \sum_{i=1}^n W_i (X_i - g_i(\theta_0))^2 \geq \sum_{i=1}^n W_i (X_i - g_i(\hat{\theta}_n))^2 \\ &= \sum_{i=1}^n W_i (X_i - g_i(\theta_0))^2 + 2 \sum_{i=1}^n W_i (X_i - g_i(\theta_0))(g_i(\theta_0) - g_i(\hat{\theta}_n)) \\ &\quad + \sum_{i=1}^n W_i (g_i(\theta_0) - g_i(\hat{\theta}_n))^2 \\ &= \sum_{i=1}^n W_i \xi_i^2 - 2U_n(\hat{\theta}_n)\Psi_n(\hat{\theta}_n, \theta_0) + \Psi_n(\hat{\theta}_n, \theta_0), \end{aligned}$$

which together with $\Psi_n(\hat{\theta}_n, \theta_0) > 0$ implies $U_n(\hat{\theta}_n) \geq 1/2$. Hence, $(|\hat{\theta}_n - \theta_0| > \varepsilon) \subseteq (U_n(\hat{\theta}_n) \geq 1/2)$. Via choosing $\varepsilon = \rho n^{-1/2}$, we have

$$\begin{aligned} P(n^{1/2}|\hat{\theta}_n - \theta_0| > \rho) &\leq P\left(\sup_{|\theta - \theta_0| > \rho n^{-1/2}} |U_n(\theta)| \geq 1/2\right) \\ &\leq P\left(\sup_{|\theta - \theta_0| > \rho} |U_n(\theta)| \geq 1/2\right) \\ &\quad + P\left(\sup_{\rho n^{-1/2} < |\theta - \theta_0| \leq \rho} |U_n(\theta)| \geq 1/2\right). \tag{4.2} \end{aligned}$$

By Cauchy’s inequality, we can see that for all $\theta \neq \theta_0$,

$$\frac{|V_n(\theta)|^2}{\Psi_n(\theta, \theta_0)} = \frac{[\sum_{i=1}^n W_i \xi_i (g_i(\theta) - g_i(\theta_0))]^2}{\sum_{i=1}^n W_i (g_i(\theta) - g_i(\theta_0))^2} \leq \sum_{i=1}^n W_i \xi_i^2. \tag{4.3}$$

Observing that $\sum_{i=1}^n W_i = 1$ and $f(x) = |x|^r$ is a convex function for all $r \geq 1$, we have by $p > 2$ and Lemma A.3 that

$$E\left(\sum_{i=1}^n W_i \xi_i^2\right)^{p/2} \leq E\left(\sum_{i=1}^n W_i |\xi_i|^p\right) = \sum_{i=1}^n E W_i E |\xi_i|^p = \frac{1}{n} \sum_{i=1}^n E |\xi_i|^p. \tag{4.4}$$

Moreover, we obtain by (4.1) that

$$\sup_{|\theta-\theta_0|>\rho} |U_n(\theta)| = \sup_{|\theta-\theta_0|>\rho} \frac{|V_n(\theta)|}{\Psi_n^{1/2}(\theta, \theta_0)\Psi_n^{1/2}(\theta, \theta_0)} \leq (\lambda_n\rho)^{-1} \sup_{|\theta-\theta_0|>\rho} \frac{|V_n(\theta)|}{\Psi_n^{1/2}(\theta, \theta_0)}. \tag{4.5}$$

Hence, it follows from (4.3)–(4.5) and Markov’s inequality that

$$\begin{aligned} P\left(\sup_{|\theta-\theta_0|>\rho} |U_n(\theta)| \geq 1/2\right) &\leq P\left(\sum_{i=1}^n W_i \xi_i^2 \geq \frac{\lambda_n^2 \rho^2}{4}\right) \\ &\leq \left(\frac{2}{\lambda_n \rho}\right)^p E\left(\sum_{i=1}^n W_i \xi_i^2\right)^{p/2} \\ &\leq C_1(p) \Delta_{np} n^{-1} \lambda_n^{-p} \rho^{-p}. \end{aligned} \tag{4.6}$$

For $m = 0, 1, 2, \dots, \lfloor n^{1/2} \rfloor$, let $\theta(m) = \theta_0 + \frac{\rho}{n^{1/2}} + \frac{m\rho}{\lfloor n^{1/2} \rfloor}$ and $\rho_m = \theta(m) - \theta_0 = \frac{\rho}{n^{1/2}} + \frac{m\rho}{\lfloor n^{1/2} \rfloor}$. It follows from (4.1) again that

$$\begin{aligned} \sup_{\rho_m < \theta - \theta_0 \leq \rho_{m+1}} |U_n(\theta)| &\leq \sup_{\rho_m < \theta - \theta_0 \leq \rho_{m+1}} \frac{|V_n(\theta)|}{\lambda_n^2 (\theta - \theta_0)^2} \leq \sup_{\rho_m < \theta - \theta_0 \leq \rho_{m+1}} \frac{|V_n(\theta)|}{\lambda_n^2 \rho_m^2} \\ &\leq \frac{|V_n(\theta(m))|}{\lambda_n^2 \rho_m^2} + \sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} \frac{|V_n(\theta_2) - V_n(\theta_1)|}{\lambda_n^2 \rho_m^2}. \end{aligned}$$

Hence, it yields that

$$\begin{aligned} P\left(\sup_{\rho_m < \theta - \theta_0 \leq \rho_{m+1}} |U_n(\theta)| \geq 1/2\right) &\leq P\left(|V_n(\theta(m))| \geq \frac{1}{4} \lambda_n^2 \rho_m^2\right) \\ &\quad + P\left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4} \lambda_n^2 \rho_m^2\right). \end{aligned} \tag{4.7}$$

By Lemma A.3 and Stirling’s approximation, we have that for each $1 \leq i \leq n$, when n is sufficiently large,

$$\begin{aligned} EW_i^p &= \frac{\Gamma(4n)\Gamma(4+p)}{\Gamma(4n+p)\Gamma(4)} \approx \frac{\Gamma(4+p)}{\Gamma(4)} \cdot \frac{\sqrt{2\pi(4n-1)} \left(\frac{4n-1}{e}\right)^{4n-1}}{\sqrt{2\pi(4n+p-1)} \left(\frac{4n+p-1}{e}\right)^{4n+p-1}} \\ &\leq C \left(\frac{4n-1}{4n+p-1}\right)^{4n-1} n^{-p} \leq Cn^{-p}. \end{aligned} \tag{4.8}$$

Note that $0 = EW_i E\xi_i = EW_i \xi_i = EW_i \xi_i^+ - EW_i \xi_i^-$ for $1 \leq i \leq n$ and from Lemma A.1 that $\{W_n \xi_n^+ - EW_n \xi_n^+, n \geq 1\}$ and $\{W_n \xi_n^- - EW_n \xi_n^-, n \geq 1\}$ are

still sequences of END random variables with zero mean. Hence, applying Markov’s inequality, Lemma A.2, (2.1) and (4.8), one can easily obtain that

$$\begin{aligned}
 P\left(|V_n(\theta(m))| \geq \frac{1}{4}\lambda_n^2\rho_m^2\right) &\leq \left(\frac{4}{\lambda_n^2\rho_m^2}\right)^p E|V_n(\theta(m))|^p \\
 &= \left(\frac{4}{\lambda_n^2\rho_m^2}\right)^p E\left|\sum_{i=1}^n W_i\xi_i(g_i(\theta(m)) - g_i(\theta_0))\right|^p \\
 &\leq 2^{p-1}\left(\frac{4}{\lambda_n^2\rho_m^2}\right)^p E\left|\sum_{i=1}^n (W_i\xi_i^+ - EW_i\xi_i^+)(g_i(\theta(m)) - g_i(\theta_0))\right|^p \\
 &\quad + 2^{p-1}\left(\frac{4}{\lambda_n^2\rho_m^2}\right)^p E\left|\sum_{i=1}^n (W_i\xi_i^- - EW_i\xi_i^-)(g_i(\theta(m)) - g_i(\theta_0))\right|^p \\
 &\leq C\left(\frac{4}{\lambda_n^2\rho_m^2}\right)^p \left\{\sum_{i=1}^n E|W_i\xi_i|^p |g_i(\theta(m)) - g_i(\theta_0)|^p\right. \\
 &\quad \left.+ \left(\sum_{i=1}^n E|W_i\xi_i|^2 (g_i(\theta(m)) - g_i(\theta_0))^2\right)^{p/2}\right\} \\
 &\leq C\left(\frac{4\Lambda_n}{\lambda_n^2\rho_m^2}\right)^p |\theta(m) - \theta_0|^p \left\{\sum_{i=1}^n EW_i^p E|\xi_i|^p + \left(\sum_{i=1}^n EW_i^2 E\xi_i^2\right)^{p/2}\right\} \\
 &\leq C_2(p)(\Lambda_n/\lambda_n^2)^p \rho_m^{-p} n^{-p} (\Delta_{np} + \nabla_{np}). \tag{4.9}
 \end{aligned}$$

Similarly, we also obtain by Lemma A.2, (2.1) and (4.8) that for all $\theta_1, \theta_2 \in \Theta$ and n large enough,

$$\begin{aligned}
 E|V_n(\theta_2) - V_n(\theta_1)|^p &= E\left|\sum_{i=1}^n W_i\xi_i(g_i(\theta_2) - g_i(\theta_1))\right|^p \\
 &\leq C\Lambda_n^p |\theta_2 - \theta_1|^p \left\{\sum_{i=1}^n EW_i^p E|\xi_i|^p + \left(\sum_{i=1}^n EW_i^2 E\xi_i^2\right)^{p/2}\right\} \\
 &\leq C\Lambda_n^p n^{-p} (\Delta_{np} + \nabla_{np}) |\theta_2 - \theta_1|^p =: C(n, p) |\theta_2 - \theta_1|^p.
 \end{aligned}$$

Hence, taking $r = 1 + \alpha = p$, $C = C(n, p)$, $\varepsilon = \rho/\lfloor n^{1/2} \rfloor$, $a = \lambda_n^2\rho_m^2/4$, and $\gamma \in (2, p + 1)$ in Lemma A.4, we obtain

$$\begin{aligned}
 P\left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4}\lambda_n^2\rho_m^2\right) \\
 = P\left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m) + \rho/\lfloor n^{1/2} \rfloor} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4}\lambda_n^2\rho_m^2\right)
 \end{aligned}$$

$$\begin{aligned} &\leq \frac{8C\Lambda_n^p n^{-p}(\Delta_{np} + \nabla_{np})}{(p + 1 - \gamma)(p + 2 - \gamma)} \left(\frac{8\gamma}{\gamma - 2}\right)^p \left(\frac{\rho}{\lfloor n^{1/2} \rfloor}\right)^p \left(\frac{4}{\lambda_n^2 \rho_m^2}\right)^p \\ &\leq C_3(p)(\Lambda_n/\lambda_n^2)^p \rho^p \rho_m^{-2p} n^{-3p/2}(\Delta_{np} + \nabla_{np}). \end{aligned} \tag{4.10}$$

Noting that $\rho_0 = \rho n^{-1/2}$, $\rho_m > m\rho n^{-1/2}$ and $p > 2$, we obtain by (4.7), (4.9) and (4.10) that

$$\begin{aligned} &P\left(\sup_{\rho n^{-1/2} < \theta - \theta_0 \leq \rho} |U_n(\theta)| \geq 1/2\right) \\ &\leq \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} P\left(\sup_{\rho_m < \theta - \theta_0 \leq \rho_{m+1}} |U_n(\theta)| \geq 1/2\right) \\ &\leq \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} P\left(|V_n(\theta(m))| \geq \frac{1}{4}\lambda_n^2 \rho_m^2\right) \\ &\quad + \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} P\left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4}\lambda_n^2 \rho_m^2\right) \\ &\leq \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} \left[C_2(p)(\Lambda_n/\lambda_n^2)^p \rho_m^{-p} n^{-p}(\Delta_{np} + \nabla_{np}) \right. \\ &\quad \left. + C_3(p)(\Lambda_n/\lambda_n^2)^p \rho^p \rho_m^{-2p} n^{-3p/2}(\Delta_{np} + \nabla_{np}) \right] \\ &\leq [C_2(p) + C_3(p)](\Lambda_n/\lambda_n^2)^p (\Delta_{np} + \nabla_{np}) n^{-p/2} \rho^{-p} \\ &\quad + (\Lambda_n/\lambda_n^2)^p (\Delta_{np} + \nabla_{np}) n^{-p/2} \rho^{-p} \sum_{m=1}^{\lfloor n^{1/2} \rfloor - 1} \left(\frac{C_2(p)}{m^p} + \frac{C_3(p)}{m^{2p}}\right) \\ &\leq C_4(p)(\Lambda_n/\lambda_n^2)^p (\Delta_{np} + \nabla_{np}) n^{-p/2} \rho^{-p}. \end{aligned} \tag{4.11}$$

Similarly, we also have

$$P\left(\sup_{\rho n^{-1/2} < \theta_0 - \theta \leq \rho} |U_n(\theta)| \geq 1/2\right) \leq C_5(p)(\Lambda_n/\lambda_n^2)^p (\Delta_{np} + \nabla_{np}) n^{-p/2} \rho^{-p}. \tag{4.12}$$

The desired result (2.2) follows from (4.2), (4.6), (4.11) and (4.12) immediately. \square

Proof of Theorem 2.2 The proof is similar to that of Theorem 2.1. Thus, we only present the differences. It follows from (4.1) that

$$\begin{aligned} \sup_{|\theta-\theta_0|>\rho} |U_n(\theta)| &= \sup_{|\theta-\theta_0|>\rho} \frac{|V_n(\theta)|}{\Psi_n(\theta, \theta_0)} \leq \sup_{|\theta-\theta_0|>\rho} \frac{\sum_{i=1}^n W_i |\xi_i| \cdot \Lambda_n |\theta - \theta_0|}{\lambda_n^2 (\theta - \theta_0)^2} \\ &\leq \frac{\Lambda_n}{\lambda_n^2 \rho} \sum_{i=1}^n W_i |\xi_i|. \end{aligned} \quad (4.13)$$

Therefore, we have by (4.8), (4.13) and C_r -inequality that

$$\begin{aligned} P \left(\sup_{|\theta-\theta_0|>\rho} |U_n(\theta)| \geq 1/2 \right) &\leq P \left(\sum_{i=1}^n W_i |\xi_i| \geq \frac{\lambda_n^2 \rho}{2\Lambda_n} \right) \\ &\leq \left(\frac{2\Lambda_n}{\lambda_n^2 \rho} \right)^p E \left(\sum_{i=1}^n W_i |\xi_i| \right)^p \\ &\leq \left(\frac{2\Lambda_n}{\lambda_n^2 \rho} \right)^p n^{p-1} \sum_{i=1}^n E W_i^p E |\xi_i|^p \\ &\leq C_6(p) (\Lambda_n / \lambda_n^2)^p \Delta_{np} n^{-1} \rho^{-p}. \end{aligned} \quad (4.14)$$

Applying Markov's inequality, the Marcinkiewicz-Zygmund inequality in Lemma A.2, (2.1) and (4.8), we can also obtain that for all n large enough,

$$\begin{aligned} P \left(|V_n(\theta(m))| \geq \frac{1}{4} \lambda_n^2 \rho_m^2 \right) &\leq \left(\frac{4}{\lambda_n^2 \rho_m^2} \right)^p E \left| \sum_{i=1}^n W_i \xi_i (g_i(\theta(m)) - g_i(\theta_0)) \right|^p \\ &\leq C \left(\frac{4}{\lambda_n^2 \rho_m^2} \right)^p \sum_{i=1}^n E |W_i \xi_i|^p |g_i(\theta(m)) - g_i(\theta_0)|^p \\ &\leq C \left(\frac{4\Lambda_n}{\lambda_n^2 \rho_m^2} \right)^p |\theta(m) - \theta_0|^p \sum_{i=1}^n E W_i^p E |\xi_i|^p \\ &\leq C_7(p) (\Lambda_n / \lambda_n^2)^p \rho_m^{-p} n^{-p} \Delta_{np}, \end{aligned} \quad (4.15)$$

and

$$\begin{aligned} E |V_n(\theta_2) - V_n(\theta_1)|^p &= E \left| \sum_{i=1}^n W_i \xi_i (g_i(\theta_2) - g_i(\theta_1)) \right|^p \\ &\leq C \Lambda_n^p |\theta_2 - \theta_1|^p \sum_{i=1}^n E W_i^p E |\xi_i|^p \\ &\leq C \Lambda_n^p n^{-p} \Delta_{np} |\theta_2 - \theta_1|^p. \end{aligned}$$

Hence, analogous to the proof of (4.10), we have

$$\begin{aligned}
 & P \left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4} \lambda_n^2 \rho_m^2 \right) \\
 & \leq \frac{8C \Lambda_n^p n^{-p} \Delta_{np}}{(p+1-\gamma)(p+2-\gamma)} \left(\frac{8\gamma}{\gamma-2} \right)^p \left(\frac{\rho}{\lfloor n^{1/2} \rfloor} \right)^p \left(\frac{4}{\lambda_n^2 \rho_m^2} \right)^p \\
 & \leq C_8(p) (\Lambda_n / \lambda_n^2)^p \rho^p \rho_m^{-2p} n^{-3p/2} \Delta_{np}.
 \end{aligned} \tag{4.16}$$

Analogous to the proof of (4.11), we obtain by (4.7), (4.15) and (4.16) that

$$\begin{aligned}
 & P \left(\sup_{\rho n^{-1/2} < \theta - \theta_0 \leq \rho} |U_n(\theta)| \geq 1/2 \right) \\
 & \leq \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} P \left(|V_n(\theta(m))| \geq \frac{1}{4} \lambda_n^2 \rho_m^2 \right) \\
 & \quad + \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} P \left(\sup_{\theta(m) < \theta_1, \theta_2 \leq \theta(m+1)} |V_n(\theta_2) - V_n(\theta_1)| \geq \frac{1}{4} \lambda_n^2 \rho_m^2 \right) \\
 & \leq \sum_{m=0}^{\lfloor n^{1/2} \rfloor - 1} \left[C_7(p) (\Lambda_n / \lambda_n^2)^p \rho_m^{-p} n^{-p} \Delta_{np} + C_8(p) (\Lambda_n / \lambda_n^2)^p \rho^p \rho_m^{-2p} n^{-3p/2} \Delta_{np} \right] \\
 & \leq [C_7(p) + C_8(p)] (\Lambda_n / \lambda_n^2)^p \Delta_{np} n^{-p/2} \rho^{-p} \\
 & \quad + (\Lambda_n / \lambda_n^2)^p \Delta_{np} n^{-p/2} \rho^{-p} \sum_{m=1}^{\lfloor n^{1/2} \rfloor - 1} \left(\frac{C_7(p)}{m^p} + \frac{C_8(p)}{m^{2p}} \right) \\
 & \leq C_9(p) (\Lambda_n / \lambda_n^2)^p \Delta_{np} n^{-p/2} \rho^{-p}.
 \end{aligned} \tag{4.17}$$

Similarly, we also have

$$P \left(\sup_{\rho n^{-1/2} < \theta_0 - \theta \leq \rho} |U_n(\theta)| \geq 1/2 \right) \leq C_{10}(p) (\Lambda_n / \lambda_n^2)^p \Delta_{np} n^{-p/2} \rho^{-p}. \tag{4.18}$$

Combining (4.2), (4.14), (4.17), and (4.18), we obtain (2.3) immediately. □

Proof of Corollary 2.1 Taking $\rho = \epsilon n^{1/p} \sqrt{\log n}$ in Theorem 2.1, we have that

$$\begin{aligned}
 \sum_{n=1}^{\infty} P(n^{1/2} |\hat{\theta}_n - \theta_0| > \rho) &= \sum_{n=1}^{\infty} P(|\hat{\theta}_n - \theta_0| > \epsilon n^{1/p-1/2} \sqrt{\log n}) \\
 &\leq C \sum_{n=1}^{\infty} n^{-1} \log^{-p/2} n < \infty,
 \end{aligned}$$

which together with the Borel-Cantelli lemma yields the rate of strong consistency. □

Proof of Corollary 2.2 Noting that $\sup_{n \geq 1} E|\xi_n|^p < \infty$ for some $1 < p \leq 2$, we may assume that $\Delta_{np} \leq 1$ for each $n \geq 1$. Hence, for any $\epsilon > 0$, taking $\rho = \left(\frac{C'(p)}{\epsilon}\right)^{1/p} n^{1/p-1/2}$ in Theorem 2.2, we have

$$\begin{aligned} P\left(|\hat{\theta}_n - \theta_0| > \left(\frac{C'(p)}{\epsilon}\right)^{1/p} n^{1/p-1}\right) &= P\left(n^{1/2}|\hat{\theta}_n - \theta_0| > \left(\frac{C'(p)}{\epsilon}\right)^{1/p} n^{1/p-1/2}\right) \\ &\leq C'(p)n^{1-p/2} \frac{\epsilon}{C'(p)} n^{-1+p/2} = \epsilon. \end{aligned}$$

The second conclusion follows immediately by choosing $\rho = \epsilon\sqrt{n/\tau_n}$ in Theorem 2.2. This completes the proof. \square

5 Conclusions

In this work, we mainly consider the following nonlinear regression model:

$$X_n = g_n(\theta) + \xi_n, \quad n \geq 1, \quad (5.1)$$

where X_n is observed, $\{g_n(\theta)\}$ is a known sequence of continuous functions possibly nonlinear in $\theta \in \Theta$, and $\{\xi_n, n \geq 1\}$ is a sequence of random errors with zero mean.

The nonlinear regression model not only has essentially fewer unknown parameters, but also has the meaning of physical variables while the linear parameters are usually devoid of physical significance. Therefore, it is of great interest to study the nonlinear regression model.

In this work, in view of the concept of Dirichlet distribution, we introduce the random weighting method to the nonlinear regression model and propose the randomly weighted least squares estimator of θ as follows:

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} \sum_{i=1}^n W_i (X_i - g_i(\theta))^2, \quad (5.2)$$

where W_i 's are independent of ξ_i 's and the random vector $\mathbf{W} = (W_1, \dots, W_n)$ obeys the Dirichlet distribution $Dir(4, 4, \dots, 4)$.

In this work, we establish the asymptotic properties for the randomly weighted least squares estimator with END errors. The results reveal that this new estimator is consistent. Moreover, some simulations are also carried out to show the superiority to the ordinary least squares estimator, especially in a heteroscedastic nonlinear regression model.

Acknowledgements The authors are most grateful to the Editor and anonymous referee for carefully reading the manuscript and valuable suggestions which helped in improving an earlier version of this paper. Supported by the National Natural Science Foundation of China (12201004, 12201079, 12201600), the National Social Science Foundation of China (22BTJ059), and the Natural Science Foundation of Anhui Province (2108085MA06).

Declarations

Conflict of interests There is no conflict of interest.

6 Appendix

Lemma A.1 *Let $(W_1, \dots, W_n) \sim \text{Dir}(\alpha_1, \dots, \alpha_n)$ for each $n \geq 1$ and $\{\xi_n, n \geq 1\}$ is a sequence of nonnegative END random variables. Then $\{W_n \xi_n, n \geq 1\}$ is still a sequence of END random variables.*

Proof It follows from Example 5.4 of Block et al. (1982) that $\{W_n, n \geq 1\}$ is a sequence of nonnegative NOD random variables, which is independent of $\{\xi_n, n \geq 1\}$. Hence, for any real numbers z_1, \dots, z_n , we have by Definition 1.2 and the properties of NOD random variables (see in Bozorgnia et al. (1996) or Lemmas 1.1 and 1.2 of Wang et al. (2011) for example) that

$$\begin{aligned}
 &P(W_1 \xi_1 \leq z_1, \dots, W_n \xi_n \leq z_n) \\
 &= \int \dots \int I(w_1 y_1 \leq z_1, \dots, w_n y_n \leq z_n) dF_{W_1, \dots, W_n, \xi_1, \dots, \xi_n}(w_1, \dots, w_n, y_1, \dots, y_n) \\
 &= \int \dots \int I(w_1 y_1 \leq z_1, \dots, w_n y_n \leq z_n) dF_{W_1, \dots, W_n}(w_1, \dots, w_n) dF_{\xi_1, \dots, \xi_n}(y_1, \dots, y_n) \\
 &= \int \dots \int P(w_1 \xi_1 \leq z_1, \dots, w_n \xi_n \leq z_n) dF_{W_1, \dots, W_n}(w_1, \dots, w_n) \\
 &\leq M \int \dots \int P(w_1 \xi_1 \leq z_1) \dots P(w_n \xi_n \leq z_n) dF_{W_1, \dots, W_n}(w_1, \dots, w_n) \\
 &= ME \left[F_{\xi_1} \left(\frac{z_1}{W_1} \right) \dots F_{\xi_n} \left(\frac{z_n}{W_n} \right) \right] \leq ME \left[F_{\xi_1} \left(\frac{z_1}{W_1} \right) \right] \dots E \left[F_{\xi_n} \left(\frac{z_n}{W_n} \right) \right] \\
 &= M \int \int I(w_1 y_1 \leq z_1) dF_{W_1}(w_1) dF_{\xi_1}(y_1) \dots \int \int I(w_n y_n \leq z_n) dF_{W_n}(w_n) dF_{\xi_n}(y_n) \\
 &= M \int \int I(w_1 y_1 \leq z_1) dF_{W_1, \xi_1}(w_1, y_1) \dots \int \int I(w_n y_n \leq z_n) dF_{W_n, \xi_n}(w_n, y_n) \\
 &= MP(W_1 \xi_1 \leq z_1) \dots P(W_n \xi_n \leq z_n).
 \end{aligned}$$

Similarly, we also have that

$$P(W_1 \xi_1 > z_1, \dots, W_n \xi_n > z_n) \leq MP(W_1 \xi_1 > z_1) \dots P(W_n \xi_n > z_n).$$

Therefore, $\{W_n \xi_n, n \geq 1\}$ is still a sequence of END random variables by Definition 1.2 again. □

Lemma A.2 *Let $\{a_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of real numbers and $\{X_n, n \geq 1\}$ be a sequence of END random variables with $EX_n = 0$ and $E|X_n|^p < \infty$ for each $n \geq 1$ and some $p > 1$. Then there exist positive constants C_p and C'_p depending only on p such that*

$$E \left| \sum_{i=1}^n a_{ni} X_i \right|^p \leq C_p \left[\sum_{i=1}^n E|a_{ni} X_i|^p + \left(\sum_{i=1}^n E a_{ni} X_i^2 \right)^{p/2} \right], \text{ if } p \geq 2,$$

$$E \left| \sum_{i=1}^n a_{ni} X_i \right|^p \leq C'_p \sum_{i=1}^n E|a_{ni} X_i|^p, \text{ if } 1 < p < 2.$$

Proof Noting that $a_{ni} = a_{ni}^+ - a_{ni}^-$ for each $1 \leq i \leq n, n \geq 1$, the Rosenthal inequality above is a direct consequence of Corollary 3.2 in Shen (2011) by using C_r -inequality. The second inequality, i.e., Marcinkiewicz-Zygmund inequality can be obtained by the first one and the method used in the proof of Theorem 2.1 in Chen et al. (2014). The details are omitted. \square

Lemma A.3 Let $(Y_1, \dots, Y_n) \sim Dir(\alpha_1, \dots, \alpha_n)$ and $\alpha_0 = \sum_{i=1}^n \alpha_i$. Then for any $p > 0$ and each $1 \leq i \leq n$,

$$E Y_i^p = \frac{\Gamma(\alpha_0)\Gamma(\alpha_i + p)}{\Gamma(\alpha_0 + p)\Gamma(\alpha_i)}.$$

Proof Without loss of generality, we only need to show $E Y_1^p = \frac{\Gamma(\alpha_0)\Gamma(\alpha_1+p)}{\Gamma(\alpha_0+p)\Gamma(\alpha_1)}$. By Definition 1.1 and some standard calculation, we have that

$$\begin{aligned} E Y_1^p &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \int_0^1 y_1^{\alpha_1+p-1} dy_1 \int_0^{1-y_1} y_2^{\alpha_1-1} dy_2 \dots \\ &\quad \times \int_0^{1-\sum_{i=1}^{n-2} y_i} y_{n-1}^{\alpha_{n-1}-1} \left(1 - \sum_{i=1}^{n-1} y_i\right)^{\alpha_n-1} dy_{n-1} \\ &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \int_0^1 y_1^{\alpha_1+p-1} dy_1 \int_0^{1-y_1} y_2^{\alpha_1-1} dy_2 \dots \\ &\quad \times \int_0^{1-\sum_{i=1}^{n-3} y_i} y_{n-2}^{\alpha_{n-2}-1} \left(1 - \sum_{i=1}^{n-2} y_i\right)^{\alpha_{n-1}+\alpha_n-1} dy_{n-2} \int_0^1 \eta_{n-1}^{\alpha_{n-1}-1} (1 - \eta_{n-1})^{\alpha_n-1} d\eta_{n-1} \\ &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \int_0^1 y_1^{\alpha_1+p-1} dy_1 \int_0^{1-y_1} y_2^{\alpha_1-1} dy_2 \dots \\ &\quad \times \int_0^{1-\sum_{i=1}^{n-3} y_i} y_{n-2}^{\alpha_{n-2}-1} \left(1 - \sum_{i=1}^{n-2} y_i\right)^{\alpha_{n-1}+\alpha_n-1} dy_{n-2} \cdot \frac{\Gamma(\alpha_{n-1})\Gamma(\alpha_n)}{\Gamma(\alpha_{n-1} + \alpha_n)} \\ &= \dots = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \cdot \frac{\Gamma(\alpha_1 + p) \prod_{i=2}^n \Gamma(\alpha_i)}{\Gamma(\alpha_0 + p)} = \frac{\Gamma(\alpha_0)\Gamma(\alpha_1 + p)}{\Gamma(\alpha_0 + p)\Gamma(\alpha_1)}, \end{aligned}$$

where the second equality above follows by letting $\eta_{n-1} = y_{n-1} / \left(1 - \sum_{i=1}^{n-2} y_i\right)$. \square

Lemma A.4 (cf. Hu 2004) Let (Ω, \mathcal{F}, P) be a probability space, $[T_1, T_2]$ be a closed interval on the real line. Assume that $V(\theta) = V(\omega, \theta)$ ($\theta \in [T_1, T_2], \omega \in \Omega$) is a stochastic process such that $V(\omega, \theta)$ is continuous for all $\omega \in \Omega$. If there exist numbers $\alpha > 0, r > 0$ and $C = C(T_1, T_2) < \infty$ such that

$$E|V(\theta_1) - V(\theta_2)|^r \leq C|\theta_1 - \theta_2|^{1+\alpha}, \text{ for any } \theta_1, \theta_2 \in [T_1, T_2],$$

then for any $\varepsilon > 0$, $a > 0$, $\theta_0, \theta_0 + \varepsilon \in [T_1, T_2]$, and $\gamma \in (2, 2 + \alpha)$, it has

$$P\left(\sup_{\theta_0 \leq \theta_1, \theta_2 \leq \theta_0 + \varepsilon} |V(\theta_1) - V(\theta_2)| \geq a\right) \leq \frac{8C}{(\alpha - \gamma + 2)(\alpha - \gamma + 3)} \left(\frac{8\gamma}{\gamma - 2}\right)^r \frac{\varepsilon^{\alpha+1}}{a^r}.$$

References

- Barvinok A, Samorodnitsky A (2007) Random weighting, asymptotic counting, and inverse isoperimetry. *Israel J Math* 158(1):159–191
- Block HW, Savits TH, Shaked M (1982) Some concepts of negative dependence. *Ann Prob* 10(3):765–772
- Bozorgnia A, Patterson RF, Taylor RL (1996) Limit theorems for dependent random variables. In: *Proceedings of the first world congress of nonlinear analysts 92 (II)*, Walter de Gruyter, Berlin, 1639–1650
- Chen PY, Bai P, Sung SH (2014) The von Bahr-Esseen moment inequality for pairwise independent random variables and applications. *J Math Anal Appl* 419(2):1290–1302
- Chen YQ, Chen AY, Ng KW (2010) The strong law of large numbers for extended negatively dependent random variables. *J Appl Prob* 47:908–922
- Fang YX, Zhao LC (2006) Approximation to the distribution of LAD estimators for censored regression by random weighting method. *J Stat Plan Inference* 136(4):1302–1316
- Gao SS, Zhang JM, Zhou T (2003) Law of large numbers for sample mean of random weighting estimate. *Inf Sci* 155(1–2):151–156
- Gao SS, Zhong YM (2010) Random weighting estimation of kernel density. *J Stat Plan Inference* 140(9):2403–2407
- 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 Prob Lett* 67(2):183–192
- Ivanov AV (1976) An asymptotic expansion for the distribution of the least squares estimator of the nonlinear regression parameter. *Theory Prob Appl* 21(3):557–570
- 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* 11:286–295
- Lita da Silva J (2015) Almost sure convergence for weighted sums of extended negatively dependent random variables. *Acta Math Hungarica* 146(1):56–70
- Liu L (2009) Precise large deviations for dependent random variables with heavy tails. *Stat Prob Lett* 79:1290–1298
- Liu L (2010) Necessary and sufficient conditions for moderate deviations of dependent random variables with heavy tails. *Sci China Ser Math* 53(6):1421–1434
- Lehmann E (1966) Some concepts of dependence. *Ann Math Stat* 37:1137–1153
- Malinvaud E (1970) The consistency of nonlinear regression. *Ann Math Stat* 41(3):956–969
- Miao Y, Tang YY (2021) Large deviation inequalities of LS estimator in nonlinear regression models. *Stat Prob Lett*, 168, Article ID 108930. <https://doi.org/10.1016/j.spl.2020.108930>
- Ng KW, Tian GL, Tang ML (2011) *Dirichlet and related distributions: theory, methods and applications*, vol 888. John Wiley & Sons, London (**Chapter 1, 2**)
- Narayanan A (1990) Computer generation of Dirichlet random vectors. *J Stat Comput Simul* 36(1):19–30
- Prakasa Rao BLS (1984) The rate of convergence of the least squares estimator in a non-linear regression model with dependent errors. *J Multivariate Anal* 14(3):315–322
- Rubin DB (1981) The Bayesian bootstrap. *Ann Stat* 9:130–134
- Shen AT (2011) Probability inequalities for END sequence and their applications. *J Inequal Appl* 2011:98, 12
- Sieders A, Dzharipidze K (1987) A large deviation result for parameter estimators and its application to nonlinear regression analysis. *The Annals of Statistics* 15(3):1031–1049

- Staniewski P (1984) The Bootstrap in nonlinear regression. In: Rasch D, Tiku ML (eds) Robustness of statistical methods and nonparametric statistics. Theory and decision library (series b: mathematical and statistical methods), vol 1. Springer, Dordrecht, pp 139–142
- Wang SJ, Wang XJ (2013) Precise large deviations for random sums of END real-valued random variables with consistent variation. *J Math Anal Appl* 402:660–667
- Wang XJ, Hu SH, Shen AT, Yang WZ (2011) An exponential inequality for a NOD sequence and a strong law of large numbers. *Appl Math Lett* 24:219–223
- Wang XJ, Li XQ, Hu SH, Wang XH (2014) On complete convergence for an extended negatively dependent sequence. *Commun Stat Theory Methods* 43(14):2923–2937
- Wang XJ, Zheng LL, Xu C, Hu SH (2015) 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
- Wu Y, Wang XJ, Hu T-C, Volodin A (2019) Complete f -moment convergence for extended negatively dependent random variables. *RACSAM* 113:333–351
- Xue LG, Zhu LX (2005) L_1 -norm estimation and random weighting method in a semiparametric model. *Acta Math Appl Sin* 21(2):295–302
- Yang WZ, Hu SH (2014) Large deviation for a least squares estimator in a nonlinear regression model. *Stat Prob Lett* 91:135–144
- Yang WZ, Xu HY, Chen L, Hu SH (2018) Complete consistency of estimators for regression models based on extended negatively dependent errors. *Stat Pap* 59(2):449–465
- Yang WZ, Zhao ZR, Wang XH, Hu SH (2017) The large deviation results for the nonlinear regression model with dependent errors. *TEST* 26(2):261–283
- Zheng ZG (1987) Random weighting method. *Acta Math Appl Sin* 10(2):247–253

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.