CrossMark

M-test in linear models with negatively superadditive dependent errors

Yuncai Yu[1*], Hongchang Hu[2], Ling Liu[3] and Shouyou Huang[2]

*Correspondence:
yuncaiyu20@163.com
[1] State Key Laboratory of Mechanics
and Control of Mechanical
Structures, Institute of Nano Science
and Department of Mathematics,
Nanjing University of Aeronautics
and Astronautics, Nanjing, 210016,
China
Full list of author information is
available at the end of the article

Abstract

This paper is concerned with the testing hypotheses of regression parameters in linear models in which errors are negatively superadditive dependent (NSD). A robust M-test base on M-criterion is proposed. The asymptotic distribution of the test statistic is obtained and the consistent estimates of the redundancy parameters involved in the asymptotic distribution are established. Finally, some Monte Carlo simulations are given to substantiate the stability of the parameter estimates and the power of the test, for various choices of M-methods, explanatory variables and different sample sizes.

MSC: 60F05; 60G10; 62F35; 62M10; 60G42

Keywords: NSD random sequences; linear regression models; M-test; asymptotic property; Monte Carlo simulations

1 Introduction

Consider the linear regression model:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + e_t, \quad t = 1, \ldots, n, \tag{1}$$

where $\{y_t\}$ and $\{\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})^\top\}$ are real-valued responses and real-valued random vectors, respectively. The superscript $\top$ represents the transpose throughout this paper, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a p-vector of the unknown parameter, and $\{e_t\}$ are random errors.

It is well known that linear regression models have received much attentions for their immense applications in various areas such as engineering technology, economics and social sciences. Unfortunately, there exists the problem that the classical maximum likelihood estimator for these models is sufficiently sensitive to outliers. To overcome this defect, Huber proposed the M-estimate which possesses the robustness (see Huber [1]) by minimizing

$$\sum_{t=1}^n \rho(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}), \tag{2}$$

where ρ is a convex function. It is obvious that many important estimates can be addressed easily. For instance, the least square (LS) estimate with $\rho(x) = x^2/2$, the least

absolute deviation (LAD) estimate with $\rho(x) = |x|$, and the Huber estimate with $\rho(x) = (x^2 I(|x| \leq k))/2 + (k|x| - k^2/2)I(|x| > k)$, $k > 0$, where $I(A)$ is the indicator function of A.

Let $\hat{\beta}_n$ be a minimizer of (2) and consequently $\hat{\beta}_n$ is a M-estimate of β. Some excellent results as regards the asymptotic properties of $\hat{\beta}_n$ with various forms of ρ have been reported in [2–5]. Most of the results rely on the independence errors. As Huber claimed in [1], the independence assumption on the errors is a serious restriction. It is practically essential and imperative to explore the case of dependent errors, which is a theoretically challenging. Under the dependence assumption of the errors, Berlinet *et al.* [6] proved the consistency of M-estimates for linear models with strong mixing errors. Cui *et al.* [7] obtained the asymptotic distributions of M-estimates for linear models with spatially correlated errors. Wu [8] investigated the weak and strong Bahadur representations of the M-estimates for linear models with stationary causal errors. Wu [9] established the strong consistency of M-estimates for linear models with negatively dependent (NA) random errors.

In the following we will introduce a wide random sequence, NSD random sequence, whose definition based on the superadditive functions.

Definition 1 (Hu [10]) A function $\phi : \mathbb{R}^n \to \mathbb{R}$, is called superadditive if

$$\phi(\mathbf{x} \vee \mathbf{y}) + \phi(\mathbf{x} \wedge \mathbf{y}) \geq \phi(\mathbf{x}) + \phi(\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where '$\vee$' is for a componentwise maximum and '$\wedge$' is for a componentwise minimum.

Definition 2 (Hu [10]) A random vector $(X_1, X_2, \ldots, X_n)$ is said to be NSD if

$$E\phi(X_1, X_2, \ldots, X_n) \leq E\phi\big(X_1^*, X_2^*, \ldots, X_n^*\big), \tag{3}$$

where $\{X_t^*, t = 1, \ldots, n\}$ are independent random variables such that have same marginal distribution with $\{X_t, t = 1, \ldots, n\}$ for each t, and ϕ is a superadditive function such that the expectations in (3) exist.

Definition 3 (Wang *et al.* [11]) A sequence of random variables $(X_1, \ldots, X_n)$ is called NSD if for all $n \geq 1$, $(X_1, \ldots, X_n)$ is NSD.

The concept of NSD random variables, which generalizes the concept of NA, was proposed by Hu [1]. In such paper, author provided several essential properties and valuable theorems for NSD. It is realized that many multivariate distributions possess the NSD property exhibited in practical examples. Compared with NA, NSD contains more widely sequence [12], *i.e.*, NA means NSD, but not vise verse. Consequently, NSD has received an increasing attention for its enormous research value in comparison with NA both in copula theory and applications [13–19]. Specifically, a Kolmogorov and a Rosenthal inequality of NSD random variables are introduced in [16] and [17], respectively. Furthermore, Wang *et al.* [11] obtained the complete convergence for weighted sums of NSD random variables and investigated the complete consistency of LS estimates in the EV models. Wang *et al.* [19] established the strong consistency of M-estimates for linear models with NSD errors via improving the moment condition.

The purpose of this paper is to investigate the M-test problem of the regression parameters in the model (1) with NSD random errors, we consider a test for the following hypothesis:

$$H_0 : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) \neq \mathbf{0}, \tag{4}$$

where $\mathbf{H}$ is a known $p \times q$ matrix with the rank q ($0 < q \leq p$), $\mathbf{b}$ is a known p-vector.

A sequence of the local alternatives is considered as follows:

$$H_{2,n} : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) = \mathbf{H}^\top \boldsymbol{\omega}_n, \tag{5}$$

where $\boldsymbol{\omega}_n$ is a known p-vector such that

$$\left\| \mathbf{S}_n^{1/2} \boldsymbol{\omega}_n \right\| = O(1), \tag{6}$$

where $\mathbf{S}_n = \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t^T$, $\| \cdot \|$ is the Euclidean norm.

Denote

$$\min_{\mathbf{H}^\top(\boldsymbol{\beta}-\mathbf{b})=\mathbf{0}} \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}\right) = \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \tilde{\boldsymbol{\beta}}\right),$$

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}\right) = \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\beta}}\right),$$

$$M_n = \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \tilde{\boldsymbol{\beta}}\right) - \sum_{t=1}^n \rho\left(y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\beta}}\right).$$

Actually, $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ are the M-estimates in the restricted and unrestricted model (1), respectively. To test the hypothesis (4), we adopt M-criterion which regards M_n as the criterion to measure the level of departure from the null hypothesis. Several classical conclusions have been presented in [20–22] when the errors are assumed to be independence, we will generalize the case to NSD random errors. Throughout this paper, let C be a positive constant. Put $|\boldsymbol{\tau}| = \max_{1 \leq t \leq p}\{|\boldsymbol{\tau}_1|, |\boldsymbol{\tau}_2|, \ldots, |\boldsymbol{\tau}_p|\}$ if $\boldsymbol{\tau}$ is a p-vector. Let $x^+ = xI(x \geq 0)$ and $x^- = -xI(x < 0)$. A random sequence $\{X_n\}$ is said to on L_q-norm, $q > 0$, if $E|X_n|^q < \infty$. Denote $a_n = o_P(b_n)$ if a_n/b_n converges to 0 in probability and $a_n = O_P(b_n)$ if a_n/b_n converges to a constant in probability.

The rest of the paper is organized as follows. In Section 2, the asymptotic distribution of M_n is obtained with the NSD random errors, and the consistence estimates of the redundancy parameters λ and σ^2 are constructed under the local hypothesis. Section 3 gives the theoretical proofs of main results. The simulations are presented to show the performances of parameter estimates and the M-test for the powers in Section 4, and the conclusions are given in Section 5.

2 Main results

In this paper, let ρ be a non-monotonic convex function on $\mathbb{R}$, and denote ψ_+ and ψ_- as the right and left derivatives of the function ρ, respectively. The derivative function ψ is chosen to satisfy $\psi_-(u) \leq \psi(u) \leq \psi_+(u)$, for all $u \in \mathbb{R}$.

Now, several assumptions are listed as follows:

(A1) The function $G(u) = E\psi(e_t + u)$ exists with $G(0) = E\psi(e_t) = 0$, and has a positive derivative λ at $u = 0$.

(A2) $0 < E\psi^2(e_1) = \sigma^2 < \infty$, and $\lim_{u \to 0} E|\psi(e_1 + u) - \psi(e_1)|^2 = 0$.

(A3) There exists a positive constant Δ such that for $h \in (0, \Delta)$, the function $\psi(u + h) - \psi(u)$ is monotonic on u.

(A4) $\sum_{t=2}^{\infty} |\text{cov}(\psi(e_1), \psi(e_t))| < \infty$.

(A5) Denote $\mathbf{S}_n = \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t^T$, assume that $\mathbf{S}_n > 0$ for sufficiently large n, and

$$d_n = \max_{1 \le t \le n} \mathbf{x}_t^T \mathbf{S}_n^{-1} \mathbf{x}_t = O(n^{-1}).$$

Remark 1 (A1)-(A4) are often applied in the asymptotic theory of M-estimate in regression models (see [20–30]). (A5) is reasonable because it is equivalent to the bound of $\max_{1 \le t \le n} |\mathbf{x}_t \mathbf{x}_t^T|$, and here is a particular case of the condition $d_n = O(n^{-\delta})$ for some $0 < \delta \le 1$, which was used in Wang *et al.* [19]. Those functions were mentioned in (2) whose 'derivative' function ψ correspond to least square (LS) estimate with $\psi(x) = x$, least absolute deviation (LAD) estimate with $\psi(x) = \text{sign}(x)$ and Huber estimate with $\psi(x) = -kI(x < -k) + xI(|x| \le k) + kI(x > k)$ are satisfied with the above conditions.

Theorem 1 *In the model* (1), *assume that* $\{e_t, 1 \le t \le n\}$, *which is a sequence of identically distributed NSD random variables, is an uniformly integral family on L2-norm, and* (A1)-(A5) *hold. Then* $2\lambda\sigma^{-2}M_n$ *has an asymptotic non-central chi-squared distribution with p-degrees of freedom and a non-central parameter* $v(n)$, *namely,*

$$2\lambda\sigma^{-2}M_n \xrightarrow{\mathfrak{D}} \chi^2_{p,v(n)},$$

where $v(n) = \lambda^2\sigma^{-2}\|\boldsymbol{\omega}(n)\|^2$, $\boldsymbol{\omega}(n) = \mathbf{H}_n^{\top}\mathbf{S}_n^{1/2}\boldsymbol{\omega}_n$, $\mathbf{H}_n = \mathbf{S}_n^{-1/2}\mathbf{H}(\mathbf{H}^{\top}\mathbf{S}_n^{-1}\mathbf{H})^{-1/2}$. *In particular, when the local alternatives* $\|\mathbf{S}_n^{1/2}\boldsymbol{\omega}_n\| \to 0$, *which means that the true parameters deviate from the null hypothesis slightly, then* $2\lambda\sigma^{-2}M_n$ *has an asymptotic central chi-squared distribution with p degrees of freedom*

$$2\lambda\sigma^{-2}M_n \xrightarrow{\mathfrak{D}} \chi^2_p.$$

For a given significance level α, *we can determine the rejection region as follows*:

$$W = \left(0, \chi^2_p(1 - \alpha/2)\right) \cup \left(\chi^2_p(\alpha/2), +\infty\right), \tag{7}$$

where $\chi^2_p(1 - \alpha/2)$, $\chi^2_p(\alpha/2)$ *are the* $(1 - \alpha/2)$-*quantile and* $\alpha/2$-*quantile of central chi-squared distribution with p degrees of freedom, respectively.*

Theorem 2 *Denote*

$$\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} \psi^2\left(y_t - \mathbf{x}_t^{\top}\hat{\boldsymbol{\beta}}_n\right),$$

$$\hat{\lambda}_n = (2nh)^{-1} \sum_{t=1}^{n} \psi\left(y_t - \mathbf{x}_t^{\top}\hat{\boldsymbol{\beta}}_n + h\right) - \psi\left(y_t - \mathbf{x}_t^{\top}\hat{\boldsymbol{\beta}}_n - h\right),$$

where $h = h_n > 0$, *and* h_n *is a sequence chosen to satisfy*

$$h_n/d_n^{1/2} \to \infty, \qquad h_n \to 0, \qquad \lim_{n\to\infty} nh_n^2 > 0. \tag{8}$$

Under the conditions of Theorem 1, *we have*

$$\hat{\sigma}_n^2 \overset{\mathbb{P}}{\to} \sigma^2, \tag{9}$$

$$\hat{\lambda}_n^2 \overset{\mathbb{P}}{\to} \lambda. \tag{10}$$

Under the assumption $\|\mathbf{S}_n^{1/2}\boldsymbol{\omega}_n\| \to 0$, *replacing* λ, σ^2 *by their consistent estimates* $\hat{\lambda}_n$ *and* $\hat{\sigma}_n^2$, *then*

$$2\hat{\lambda}_n\hat{\sigma}_n^{-2}M_n \overset{\mathfrak{D}}{\to} \chi_p^2.$$

3 Proof of theorems

It is convenient to consider the rescaled model

$$y_{nt} = \mathbf{x}_{nt}^\top \boldsymbol{\beta}(n) + e_t, \quad t = 1, 2, \ldots, n, \tag{11}$$

where $\mathbf{x}_{nt} = \mathbf{S}_n^{-1/2}\mathbf{x}_t$, $\boldsymbol{\beta}(n) = \mathbf{S}_n^{1/2}(\boldsymbol{\beta} - \mathbf{b})$, $y_{nt} = y_t - \mathbf{x}_t^\top \mathbf{b}$. It is easily to check that

$$\sum_{t=1}^n \left\| \mathbf{x}_{nt}\mathbf{x}_{nt}^\top \right\| = p. \tag{12}$$

Assume that $q < p$, there exists a $p \times (p - q)$ matrix $\mathbf{K}$ with the rank $(p - q)$ such that $\mathbf{H}^\top \mathbf{K} = \mathbf{0}$ and $\mathbf{K}^\top \boldsymbol{\omega}_n = \mathbf{0}$, then, for some $\boldsymbol{\gamma} \in \mathbb{R}^{p-q}$, H_0 and $H_{2,n}$ can be written as

$$H_0 : \boldsymbol{\beta} - \mathbf{b} = \mathbf{K}\boldsymbol{\gamma}, \qquad H_{2,n} : \boldsymbol{\beta} - \mathbf{b} = \mathbf{K}\boldsymbol{\gamma} + \boldsymbol{\omega}_n. \tag{13}$$

Denote $\mathbf{H}_n = \mathbf{S}_n^{-1/2}\mathbf{H}(\mathbf{H}^\top \mathbf{S}_n^{-1}\mathbf{H})^{-1/2}$, $\mathbf{K}_n = \mathbf{S}_n^{1/2}\mathbf{K}(\mathbf{K}^\top \mathbf{S}_n\mathbf{K})^{-1/2}$, then

$$\mathbf{H}_n^\top \mathbf{H}_n = \mathbf{I}_q, \qquad \mathbf{K}_n^\top \mathbf{K}_n = \mathbf{I}_{p-q}, \qquad \mathbf{H}_n^\top \mathbf{K}_n = \mathbf{0}, \qquad \mathbf{H}_n\mathbf{H}_n^\top + \mathbf{K}_n\mathbf{K}_n^\top = \mathbf{I}_p. \tag{14}$$

Let $\boldsymbol{\gamma}_0(n) = (\mathbf{K}^\top \mathbf{S}_n\mathbf{K})^{1/2}\boldsymbol{\gamma}$. Under the null hypothesis, model (11) can be rewritten as

$$y_{nt} = \mathbf{x}_{nt}^\top \mathbf{K}_n\boldsymbol{\gamma}_0(n) + e_t, \quad t = 1, 2, \ldots, n.$$

Set $\boldsymbol{\omega}(n) = \mathbf{H}_n^\top \mathbf{S}_n^{1/2}\boldsymbol{\omega}_n$, $\boldsymbol{\gamma}(n) = \boldsymbol{\gamma}_0(n) + \mathbf{K}_n^\top \mathbf{S}_n^{1/2}\boldsymbol{\omega}_n$, under the local alternatives (13),

$$\boldsymbol{\beta}(n) = \mathbf{K}_n\boldsymbol{\gamma}(n) + \mathbf{H}_n\boldsymbol{\omega}(n). \tag{15}$$

Define $\hat{\boldsymbol{\beta}}(n) = \mathbf{S}_n^{1/2}(\hat{\boldsymbol{\beta}} - \mathbf{b})$, and $\hat{\boldsymbol{\gamma}}(n)$ satisfies

$$\min_{\boldsymbol{\varsigma} \in \mathbb{R}^{p-q}} \sum_{t=1}^n \rho\big(y_{nt} - \mathbf{x}_{nt}^\top \mathbf{K}_n\boldsymbol{\varsigma}\big) = \sum_{t=1}^n \rho\big(y_{nt} - \mathbf{x}_{nt}^\top \mathbf{K}_n\hat{\boldsymbol{\gamma}}(n)\big).$$

Obviously, $\hat{\boldsymbol{\beta}}(n)$, $\hat{\boldsymbol{\gamma}}(n)$ are the M-estimates of $\boldsymbol{\beta}(n)$ and $\boldsymbol{\gamma}(n)$, respectively. Thus

$$\tilde{\boldsymbol{\beta}} = \mathbf{b} + \mathbf{S}_n^{-1/2}\mathbf{K}_n\hat{\boldsymbol{\gamma}}(n).$$

Next, we will state some lemmas that are needed in order to prove the main results of this paper.

Lemma 1 (Hu [10]) *If $\{X_n, n \geq 1\}$ is a NSD random sequence, we have the following properties.*

(a) *For any $x_1, x_2, \ldots, x_n$,*

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \leq \prod_{t=1}^{n} P(X_t \leq x_t).$$

(b) *If $f_1, f_2, \ldots, f_n$ are non-decreasing functions, then $\{f_n(X_n), n \geq 1\}$ is still a NSD random sequence.*

(c) *The sequence $\{-X_1, -X_2, \ldots, -X_n\}$ is also NSD.*

Lemma 2 ((Rosenthal inequality) (Shen *et al.* [17])) *Let $\{X_n, n \geq 1\}$ be a NSD random sequence with $EX_t = 0$ and $E|X_n|^p < \infty$ for some $p \geq 2$, then, for all $n \geq 1$ and $p \geq 2$,*

$$E\left(\max_{1 \leq j \leq n}\left|\sum_{t=1}^{j} X_t\right|^p\right) \leq C\left\{\sum_{t=1}^{n} E|X_t|^p + \left(\sum_{t=1}^{n} EX_t^2\right)^{p/2}\right\}.$$

Lemma 3 (Anderson *et al.* [31]) *Let D be an open convex subset of $\mathbb{R}^n$ and $\{f_n\}$ are convex functions on D, for any $x \in D$,*

$$f_n(x) \xrightarrow{\mathbb{P}} f(x).$$

If f is a real function on D, then f is also convex, and for all compact subset $D_0 \subset D$,

$$\sup_{x \in D_0}\left|f_n(x) - f(x)\right| \xrightarrow{\mathbb{P}} 0. \tag{16}$$

Moreover, if f is a differentiable function on D, $g(x)$ and $g_n(x)$ represent the gradient and sub-gradient of f, respectively, then (16) implies that for all D_0

$$\sup_{x \in D_0}\left|g_n(x) - g(x)\right| \xrightarrow{\mathbb{P}} 0.$$

Lemma 4 *Assume that $\{X_n, n \geq 1\}$ is a sequence of identically distributed NSD random sequence with finite variance, and an array of real numbers $\{a_{nj}, 1 \leq j \leq n\}$ is satisfied $\sum_{j=1}^{n} a_{nj}^2 = O(1)$, $\max_{1 \leq j \leq n} |a_{nj}| \to 0$. Then, for any real $r_j, j = 1, \ldots, n$,*

$$\left|E\exp\left(i\sum_{j=1}^{n} r_j Z_{nj}\right) - \prod_{j=1}^{n} E\exp(ir_j Z_{nj})\right| \leq \frac{1}{2}\sum_{j \neq l, j, l = 1}^{n}\left|r_j r_l \operatorname{Cov}(Z_{nj}, Z_{nl})\right|,$$

where $Z_{nj} = \sum_{l \in \Upsilon_j} a_{nl} X_l$, $\{\Upsilon_j\}$ are disjoint subsets of $\{1, 2, \ldots, n\}$, i refers to imaginary unit.

Proof For a pair of NSD random variables X, Y, by the property (a) in Lemma 1, we have

$$H(x, y) = P(X \leq x, Y \leq y) - P(X \leq x)P(Y \leq y) \leq 0. \tag{17}$$

Denote by $F(x, y)$ the joint distribution functions of (X, Y), and $F_X(x)$, $F_Y(y)$ the marginal distribution function of X, Y, one gets

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = \iint \left[F(x, y) - F_X(x)F_Y(y) \right] \mathrm{d}x\,\mathrm{d}y$$

$$= \iint H(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{18}$$

Form (18), we obtain

$$\mathrm{Cov}\big(f(X), g(Y)\big) = \iint f'(X)g'(Y)H(x, y)\,\mathrm{d}x\,\mathrm{d}y,$$

where f, g are complex valued functions on $\mathbb{R}$ with f', $g' < \infty$. Combining (17) and (18) yields

$$\big|\mathrm{Cov}\big(f(X), g(Y)\big)\big| \leq \iint \big|f'(X)\big|\big|g'(Y)\big|\big|H(x, y)\big|\,\mathrm{d}x\,\mathrm{d}y \leq \big\|f'\big\|_\infty \big\|g'\big\|_\infty \big|\mathrm{Cov}(X, Y)\big|. \tag{19}$$

Taking $f(X) = \exp(irX)$, $g(Y) = \exp(iuY)$, it is easily seen that

$$\big\|f'(X)\big\|_\infty \leq 1 < \infty, \qquad \big\|g'(Y)\big\|_\infty \leq 1 < \infty,$$

thus for any real numbers r, u

$$\big|E\exp(irX + iuY) - E\exp(irX)E\exp(iuY)\big| \leq \big|ru\,\mathrm{Cov}(X, Y)\big|. \tag{20}$$

Next, we proceed the proof by induction on n. Lemma 4 for $n = 1$ is trivial and for $n = 2$ is true by (20). Assume that the result is true for all $n \leq M$ ($n \geq 3$). For $n = M + 1$, there exist some $\epsilon^2 = 1$, $\delta^2 = 1$, $k \in \{1, \ldots, M\}$ such that

$$\begin{cases} \epsilon r_j \geq 0, & 1 \leq j \leq k, \\ \delta r_j \geq 0, & 1 + k \leq j \leq M + 1. \end{cases}$$

Denote $X = \sum_{j=1}^{k} \epsilon r_j Z_{nj}$, $Y = \sum_{j=k+1}^{M+1} \delta r_j Z_{nj}$, then

$$\sum_{j=1}^{n} r_j Z_{nj} = \epsilon X + \delta Y.$$

Note that X, Y are non-decreasing functions, we have by the induction hypothesis that

$$\left| E\exp\left(i\sum_{j=1}^{n} r_j Z_{nj} \right) - \prod_{j=1}^{n} E\exp(ir_j Z_{nj}) \right|$$

$$\leq \big|E(T_1 T_2) - E(T_1)E(T_2)\big| + \left| E(T_1)E(T_2) - E(T_1) \prod_{j=1+k}^{M+1} E(R) \right|$$

$$+ \left| E(T_1) \prod_{j=1+k}^{M+1} E(R) - \prod_{j=1}^{k} E(R) \right|$$

$$\leq |\epsilon\delta| \left| \mathrm{Cov}(X,Y) \right| + \left| E(T_2) - \prod_{j=1+k}^{M+1} E(R) \right| + \left| E(T_1) - \prod_{j=1+k}^{M+1} E(R) \right|$$

$$\leq \left| \mathrm{Cov}\left(\sum_{j=1}^{k} \epsilon r_j Z_{nj}, \sum_{j=k+1}^{M+1} \delta r_l Z_{nl} \right) \right| + \frac{1}{2} \sum_{j\neq l,j,l=k+1}^{M+1} \left| r_j r_l \mathrm{Cov}(Z_{nj}, Z_{nl}) \right|$$

$$+ \frac{1}{2} \sum_{j\neq l,j,l=1}^{M+1} \left| r_j r_l \mathrm{Cov}(Z_{nj}, Z_{nl}) \right|$$

$$\leq \frac{1}{2} \sum_{j\neq l,j,l=1}^{n} \left| r_j r_l \mathrm{Cov}(Z_{nj}, Z_{nl}) \right|,$$

where $T_1 = \exp(i\epsilon X)$, $T_2 = \exp(i\delta Y)$, $R = \exp(ir_j Z_{nj})$. Thus, this completes the proof of Lemma 4. □

Lemma 5 (Billingsley [32]) *If $X_{nj} \xrightarrow{L} X_j$, $X_j \xrightarrow{L} X$ for each j, and uniform measure ϱ is satisfied that for all $\varepsilon > 0$,*

$$\lim_{j\to\infty} \lim_{n\to\infty} \sup\{\varrho(X_{nj}, Y_n) \geq \varepsilon\} = 0,$$

then

$$Y_n \xrightarrow{L} X.$$

Lemma 6 *Suppose that $\{X_n, n \geq 1\}$ and $\{a_{nj}, 1 \leq j \leq n\}$ satisfy the assumptions of Lemma 4. Further assume that $\{X_n, n \geq 1\}$ is an uniformly integral family on L_2-norm, then*

$$\sigma_n^{-1} \sum_{j=1}^{n} a_{nj} X_j \xrightarrow{\mathfrak{D}} N(0,1),$$

where $\sigma_n^2 = \mathrm{var}(\sum_{j=1}^{n} a_{nj} X_j)$.

Proof Without loss of generality, we suppose that $a_{nj} = 0$ for all $j > n$. By (18), we have $\mathrm{Cov}(X, Y) \leq 0$ because of the negativity of $H(x,y)$. Then, for $1 \leq m \leq n-1$,

$$\sum_{l,j=1,|l-j|\geq m}^{n} \left| a_{nl} a_{aj} \mathrm{Cov}(X_l, X_j) \right| \leq \sum_{j=1}^{n-u} \sum_{l=j+u}^{n} \left(a_{nj}^2 + a_{nl}^2 \right) \left| \mathrm{Cov}(X_l, X_j) \right|$$

$$\leq \sum_{j=1}^{n-m} a_{nj}^2 \sum_{l=j+m}^{n} \left| \mathrm{Cov}(X_l, X_j) \right| + \sum_{l=j+m}^{n} a_{nj}^2 \sum_{j=1}^{l-m} \left| \mathrm{Cov}(X_l, X_j) \right|$$

$$\leq \sum_{j=1}^{n} a_{nj}^2 \sum_{|l-j|\geq m}^{n} \left| \mathrm{Cov}(X_l, X_j) \right|$$

$$\leq \sup_{j} \left| \sum_{l=1,|l-j|\geq m}^{n} \mathrm{Cov}(X_l, X_j) \right| \left(\sum_{l=1}^{n} a_{nl}^2 \right).$$

Taking $\psi(x) = x$ in assumption (A4), we get, for all $l \geq 1$ and sufficiently large j,

$$\sum_{j:|l-j|\geq m}^{\infty} \left| \mathrm{Cov}(X_l, X_j) \right| \to 0.$$

Therefore, for a fixed small ε, there exists a positive integer $m = m_\varepsilon$ such that

$$\sum_{l,j=1,|l-j|\geq m} \left| a_{nl} a_{nj} \mathrm{Cov}(X_l, X_j) \right| \leq \varepsilon. \tag{21}$$

Denote $N_0 = [1/\varepsilon]$, where $[x]$ denotes the integer part of x, and $Y_{nj} = \sum_{k=mj+1}^{m(j+1)} a_{nk} X_k$, $j = 0, 1, 2, \ldots, n$,

$$\Upsilon_j = \left\{ j : 2N_0 l \leq j \leq 2N_0 l + N_0, \left| \mathrm{Cov}(Y_{nj}, Y_{nj+1}) \right| \leq \frac{2}{N_0} \sum_{j=2N_0 l}^{2N_0 l + N_0} \mathrm{Var}(Y_{nl}) \right\}.$$

We define $s_0 = 0$, $s_{j+1} = \min\{s : s > s_j, s \in \Upsilon_j\}$, and put

$$Z_{nj} = \sum_{l=s_j+1}^{s_{j+1}} Y_{nl}, \quad j = 0, 1, 2, \ldots, n,$$

$$\Lambda_j = \left\{ m(s_j + 1) + 1, \ldots, m(s_{j+1} + 1) \right\}.$$

Note that

$$Z_{nj} = \sum_{l \in \Lambda_j} a_{nl} X_l, \quad j = 0, 1, 2, \ldots, n,$$

it is easy to see that $\#\Lambda_j \leq 3N_0 m$, where $\#$ stands for the cardinality of a set. Next, we will proof that $\{Z_{nj}, 1 \leq j \leq n\}$ satisfies the Lindeberg condition.

Let $B_n^2 = \sum_{j=1}^{n} EZ_{nj}^2$, by Lemma 2, it yields

$$B_n^2 = \sum_{j=1}^{n} E \left(\sum_{l \in \Lambda_j} a_{nl} X_l \right)^2$$

$$\leq \sum_{j=1}^{n} a_{nl}^2 E \left(\sum_{l \in \Lambda_j} X_l \right)^2$$

$$\leq \sum_{j=1}^{n} \sum_{l \in \Lambda_j} a_{nl}^2 E(X_l)^2$$

$$\leq \sum_{j=1}^{n} E(a_{nl} X_j)^2 = \sigma_n^2 < \infty.$$

Clearly, $\{Z_{nj}^2\}$ is uniform integrable since $\{X_j^2, j \geq 1\}$ is uniform integrable. Hence, for any positive ε, $\{Z_{nj}, 1 \leq j \leq n\}$ is verified to satisfy the Lindeberg condition by

$$\frac{1}{B_n^2} \sum_{j=1}^n E Z_{nj}^2 I\{|Z_{nj}| \geq \varepsilon B_n\} \leq \sum_{j=1}^n \left(\sum_{l \in \Lambda_j} a_{nl}^2\right) \max_{l \in \Lambda_j} E X_l^2 I\left\{\sum_{l \in \Lambda_j} |X_l| \geq \varepsilon / \max_{l \in \Lambda_j} |a_{nl}|\right\}$$

$$\leq \left(\sum_{j=1}^n a_{nl}^2\right) \max_{1 \leq j \leq n} E X_l^2 I\left\{\sum_{l \in \Lambda_j} X_l^2 \geq \varepsilon^2 / \left(\max_{l \in \Lambda_l} |a_{nl}|\right)^2\right\}$$

$$\leq \left(\sum_{j=1}^n a_{nl}^2\right) \max_{1 \leq j \leq n} E \sum_{l \in \Lambda_j} X_l^2 I\left\{\sum_{l \in \Lambda_j} X_l^2 \geq \varepsilon^2 / \left(\max_{l \in \Lambda_j} |a_{nl}|\right)^2\right\}.$$

Now taking an independence random sequence $\{Z_{nj}^*, j = 1, 2, \ldots, n\}$ such that have same marginal distribution with Z_{nj} for each j. Let $F(Z_{n1}, Z_{n2}, \ldots, Z_{nn})$ and $G(Z_{n1}^*, Z_{n2}^*, \ldots, Z_{nn}^*)$ be the eigenfunctions of $\sum_{j=1}^n Z_{nj}$ and $\sum_{j=1}^n Z_{nj}^*$, respectively. Choosing $r = \max\{r_l, r_j\}$, we have by Lemma 4 and (21)

$$|F - G| = \left| E \exp\left(i \sum_{j=1}^n r_j Z_{nj}\right) - E \exp\left(i \sum_{j=1}^n r_j Z_{nj}^*\right) \right|$$

$$= \left| E \exp\left(i \sum_{j=1}^n r_j Z_{nj}\right) - \prod_{j=1}^n E \exp(i r_j Z_{nj}) \right|$$

$$\leq \frac{1}{2} \sum_{j \neq l, j, l=1}^n |r_j r_l \text{Cov}(Z_{nj}, Z_{nl})|$$

$$\leq r^2 \left(\sum_{1 \leq l < j \leq n, |l-j|=1}^n |\text{Cov}(Z_{nj}, Z_{nl})| + \sum_{1 \leq l < j \leq n, |l-j|>1}^n |\text{Cov}(Z_{nj}, Z_{nl})| \right)$$

$$\leq r^2 \left(\sum_{j=1}^n |\text{Cov}(Y_{ns_l}, Y_{ns_{l+1}})| + \sum_{1 \leq l < j \leq n, |l-j|>m}^n |a_{nl} a_{nj}| |\text{Cov}(X_{nj}, X_{nl})| \right)$$

$$\leq r^2 \left(\frac{C}{N_0} \sum_{j=1}^n \text{Var}(Y_{nj}) + \varepsilon \right) \leq \varepsilon (r^2 + C \sigma_n^2).$$

By Levy's theorem, Z_{nj}^* obtains the asymptotic normality, applying Lemma 5, then the identically distribution property of $\{X_j\}$ implies that

$$B_n^{-1} \sum_{j=1}^n Z_{nj} = \sigma_n^{-1} \sum_{j=1}^n a_{nj} X_j \xrightarrow{\mathfrak{D}} N(0, 1),$$

which completes the proof of Lemma 6. □

Lemma 7 *In the model* (1), *assume that* $\{e_t, 1 \leq t \leq n\}$ *is a sequence of NSD identically distributed random variables,* (A1)-(A4) *are satisfied, for any positive constant* δ *and sufficiently large* n, *then*

$$\sup_{|\tau| \leq \delta} \left| \sum_{t=1}^n \{\rho(e_t - \mathbf{x}_{nt}^\top \boldsymbol{\tau}) - \rho(e_t) + \psi(e_t) \mathbf{x}_{nt}^\top \boldsymbol{\tau}^\top\} - \frac{1}{2} \lambda \boldsymbol{\tau}^\top \boldsymbol{\tau} \right| \xrightarrow{\mathbb{P}} 0,$$

$$\sup_{|\boldsymbol{\tau}|\leq\delta}\left|\sum_{t=1}^{n}\{\psi(e_t-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau})-\psi(e_t)\}\mathbf{x}_{nt}^{\top}+\lambda\boldsymbol{\tau}\right|\overset{\mathbb{P}}{\to}0,$$

where $\boldsymbol{\tau}$ is a p-vector.

Proof Denote

$$f_n(\boldsymbol{\tau}) = \sum_{t=1}^{n}\{\rho(e_t-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau})-\rho(e_t)+\psi(e_t)\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}\}$$

$$= \sum_{t=1}^{n}\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}\{\psi(e_t+u)-\psi(e_t)\}\,du.$$

For fixed $\boldsymbol{\tau}$, it follows from (A5) that

$$\max_{1\leq t\leq n}\left|\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}\right|\to O(n^{-1/2}). \tag{22}$$

From (A1) and (22), there exist a sequence of positive numbers $\varepsilon_n\to 0$ and $\theta_{nt}\in(-1,1)$ such that, for sufficiently large n,

$$Ef_n(\boldsymbol{\tau}) = \sum_{t=1}^{n}\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}E(\psi(e_t+u)-\psi(e_t))\,du$$

$$= \sum_{t=1}^{n}\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}\{\lambda u+o(|u|)\}\,du$$

$$= \frac{1}{2}\lambda\sum_{t=1}^{n}(\mathbf{x}_{nt}^{\top}\boldsymbol{\tau})^2(1+\varepsilon_n\theta_{nt})\to\frac{1}{2}\lambda\boldsymbol{\tau}^{\top}\boldsymbol{\tau}.$$

In view of the monotonicity of $\psi(e_t+u)-\psi(e_t)$, the summands of $f_n(\boldsymbol{\tau})$ is also monotonous with respect to e_t from the property (b) in Lemma 1. We divide the summands of $f_n(\boldsymbol{\tau})$ into positive and negative two parts, by the property (c) in Lemma 1, they are still NSD. Therefore, applying Schwarz's inequality and (22), we obtain

$$\text{var}[f_n(\boldsymbol{\tau})] = E\Bigg\{\sum_{t=1}^{n}\Bigg[\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}(\psi(e_t+u)-\psi(e_t))\,du\Bigg]^{+}$$

$$-\sum_{t=1}^{n}\Bigg[\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}(\psi(e_t)+u)-\psi(e_t)\Bigg]^{-}\,du\Bigg\}^{2}$$

$$\leq E\Bigg\{\sum_{t=1}^{n}\Bigg[\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}(\psi(e_t+u)-\psi(e_t))\,du\Bigg]^{+}\Bigg\}^{2}$$

$$+E\Bigg\{\sum_{t=1}^{n}\Bigg[\int_{0}^{-\mathbf{x}_{nt}^{T}\boldsymbol{\tau}}(\psi(e_t+u)-\psi(e_t))\,du\Bigg]^{-}\Bigg\}^{2}$$

$$\leq \sum_{t=1}^{n}E\Bigg[\int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}}(\psi(e_t+u)-\psi(e_t))\,du\Bigg]^{2}$$

$$\leq \sum_{t=1}^{n} |\mathbf{x}_{nt}^{\top}| \left| \int_{0}^{-\mathbf{x}_{nt}^{\top}\boldsymbol{\tau}} E\big[\psi(e_t + u) - \psi(e_t)\big]^2 \, du \right|$$

$$= o(1) \sum_{t=1}^{n} \big(\mathbf{x}_{nt}^{\top}\big)^2 \to 0.$$

Hence for sufficiently large n, we have

$$f_n(\boldsymbol{\tau}) \xrightarrow{\mathbb{P}} \frac{1}{2}\lambda\boldsymbol{\tau}^{\top}\boldsymbol{\tau}. \tag{23}$$

Lemma 7 is proved by (23) and Lemma 3. □

Lemma 8 *Under conditions of Lemma 7 and the local alternatives* (5)-(6), *then we see that, for any positive constant δ and sufficiently large n,*

$$\sup_{|\boldsymbol{\xi}_1| \leq \delta} \left| \sum_{t=1}^{n} \big\{ \rho\big(y_{nt} - \mathbf{x}_{nt}^{\top}\boldsymbol{\eta}\big) - \rho(e_t) + \psi(e_t)\mathbf{x}_{nt}^{\top}\boldsymbol{\xi}_1 \big\} - \frac{1}{2}\lambda\|\boldsymbol{\xi}_1\|^2 \right| \xrightarrow{\mathbb{P}} 0, \tag{24}$$

$$\sup_{|\boldsymbol{\xi}_1| \leq \delta} \left| \sum_{t=1}^{n} \big[\psi\big(y_{nt} - \mathbf{x}_{nt}^{\top}\boldsymbol{\eta}\big) - \psi(e_t) \big]\mathbf{x}_{nt}^{\top} + \lambda\boldsymbol{\xi}_1 \right| \xrightarrow{\mathbb{P}} 0, \tag{25}$$

$$\sup_{|\boldsymbol{\xi}_2| \leq \delta} \left| \sum_{t=1}^{n} \big\{ \rho\big(y_{nt} - \mathbf{x}_{nt}^{\top}\mathbf{K}_n\boldsymbol{\zeta}\big) - \rho(e_t) \big\} + \sum_{t=1}^{n} \psi(e_t)\mathbf{x}_{nt}^{\top}\big(\mathbf{K}_n\boldsymbol{\zeta} - \boldsymbol{\beta}(n)\big) \right.$$

$$\left. - \frac{1}{2}\lambda\|\boldsymbol{\xi}_2\|^2 + \|\boldsymbol{\omega}(n)\|^2 \right| \xrightarrow{\mathbb{P}} 0, \tag{26}$$

$$\sup_{|\boldsymbol{\xi}_2| \leq \delta} \left| \sum_{t=1}^{n} \big[\psi\big(y_{nt} - \mathbf{x}_{nt}^{\top}\mathbf{K}_n\boldsymbol{\zeta}\big) - \psi(e_t) \big]\mathbf{x}_{nt} + \lambda\big(\mathbf{K}_n\boldsymbol{\zeta} - \boldsymbol{\beta}(n)\big) \right| \xrightarrow{\mathbb{P}} 0, \tag{27}$$

where $\boldsymbol{\xi}_1 = \boldsymbol{\eta} - \boldsymbol{\beta}(n)$, $\boldsymbol{\xi}_2 = \boldsymbol{\zeta} - \boldsymbol{\tau}(n)$.

Proof Take the proofs of (24) and (25) as examples, the rest, equations (26) and (27), are the same. Note that (24) can be written as

$$\sup_{|\boldsymbol{\xi}_2| \leq \delta} \left| \sum_{t=1}^{n} \big\{ \rho\big(e_t - \mathbf{x}_{nt}^{\top}\big(\mathbf{K}_n\boldsymbol{\zeta} - \mathbf{H}_n\boldsymbol{\omega}(n)\big)\big) - \rho(e_t) \big\} + \sum_{t=1}^{n} \psi(e_t)\mathbf{x}_{nt}^{\top}\big(\mathbf{K}_n\boldsymbol{\zeta} - \mathbf{H}_n\boldsymbol{\omega}(n)\big) \right.$$

$$\left. - \frac{1}{2}\lambda\|\boldsymbol{\xi}_2\|^2 + \|\boldsymbol{\omega}(n)\|^2 \right| \xrightarrow{\mathbb{P}} 0.$$

On the other hand, $\|\mathbf{S}_n^{1/2}\boldsymbol{\omega}_n\| = O(1)$ and $\|\boldsymbol{\gamma}_2\| \leq \delta$, hence there exists a constant δ_1 such that

$$\big|\mathbf{K}_n\boldsymbol{\zeta} - \mathbf{H}_n\boldsymbol{\omega}(n)\big| \leq \delta_1. \tag{28}$$

Thus (24) and (25) follow naturally by (28) and Lemma 7. □

Lemma 9 *Under the conditions of Theorem 1, as $n \to \infty$, we have*

$$\hat{\boldsymbol{\beta}}(n) - \boldsymbol{\beta}(n) = \lambda^{-1} \sum_{t=1}^{n} \mathbf{x}_{nt} \psi(e_t) + o_P(1), \tag{29}$$

$$\hat{\boldsymbol{\gamma}}(n) - \boldsymbol{\gamma}(n) = \lambda^{-1} \sum_{t=1}^{n} \mathbf{K}_n^{\top} \mathbf{x}_{nt} \psi(e_t) + o_P(1). \tag{30}$$

Proof The estimate of (2) can be defined essentially as the solution of the following equation:

$$\left\| \mathbf{S}_n^{-1/2} \sum_{t=1}^{n} \psi(y_t - \mathbf{x}_t^{\top} \hat{\boldsymbol{\beta}}) \mathbf{x}_t \right\| = o_P(1). \tag{31}$$

Denote $\hat{\boldsymbol{\beta}}(n) = S_n^{1/2} \hat{\boldsymbol{\beta}}_n$, (31) can be rewritten to give

$$\left\| \sum_{t=1}^{n} \psi(e_t - \mathbf{x}_{nt}^{\top} \hat{\boldsymbol{\beta}}(n)) \mathbf{x}_{nt} \right\| = o_P(1). \tag{32}$$

By a routine argument, we shall prove that

$$\left| \hat{\boldsymbol{\beta}}(n) \right| = O_p(1). \tag{33}$$

Let U be a denumerable dense subset in the unit sphere of $\mathbb{R}^p$ such that

$$U = \left\{ \boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\| = 1 \right\}.$$

Write

$$D(\boldsymbol{\tau}, L) = \sum_{t=1}^{n} \psi(e_t - L\mathbf{x}_{nt}^T \boldsymbol{\tau}) \mathbf{x}_{nt}^T \boldsymbol{\tau},$$

where $L \geq 0$, $\boldsymbol{\tau} \in \mathbb{R}^p$.

Obviously, for a given $\boldsymbol{\tau}$, $D(\cdot, L)$ is non-decreasing on L since ψ is non-decreasing. For any $\varepsilon > 0$, let

$$L_0 = \sqrt{p}\sigma / (\lambda\sqrt{\varepsilon}).$$

Thus by (32), there exists a number n_1, as $n \geq n_1$,

$$\Pr\left\{ \left| D\left(\frac{\hat{\boldsymbol{\beta}}(n)}{\|\hat{\boldsymbol{\beta}}(n)\|}, \|\hat{\boldsymbol{\beta}}(n)\| \right) \right| \geq L_0\lambda \right\} < \varepsilon.$$

Note that $D(\boldsymbol{\tau}, \cdot)$ is a non-decreasing function on $\boldsymbol{\tau}$ for given L, then

$$\Pr\left\{ \|\hat{\boldsymbol{\beta}}(n)\| \geq L_0 \right\} < \Pr\left\{ \sup_{\boldsymbol{\tau} \in U} D(\tau, L_0) \leq -L_0\lambda \right\} + \varepsilon.$$

Based on Lemma 8 and $\max_{1 \le t \le n} |\mathbf{x}_{nt}^{\top} \boldsymbol{\tau}| = O(n^{-1/2})$, one can see that

$$\sup_{\boldsymbol{\tau} \in U} \left| \sum_{t=1}^{n} \left[\psi \left(e_t - L_0 x_{nt}^{\top} \boldsymbol{\tau} \right) - \psi(e_t) \right] \mathbf{x}_{nt}^{\top} \boldsymbol{\tau} + L_0 \lambda \right| \to 0. \tag{34}$$

On the other hand, by Schwarz's inequality, we have

$$\sup_{\boldsymbol{\tau} \in U} \left| \sum_{t=1}^{n} \psi(e_t) \mathbf{x}_{nt}^{\top} \boldsymbol{\tau} \right| \le \left\| \sum_{t=1}^{n} \psi(e_t) \mathbf{x}_{nt} \right\|. \tag{35}$$

Combining (34) and (35), there exists n_2 ($n_1 \le n_2 \le n$) such that

$$\Pr \left\{ \sup_{\boldsymbol{\tau} \in U} D(\boldsymbol{\tau}, L_0) < -L_0 \lambda + \left\| \sum_{t=1}^{n} \psi(e_t) \mathbf{x}_{nt} \right\| \right\} > 1 - \varepsilon. \tag{36}$$

Applying Chebyshev's inequality, the C_r inequality and Lemma 8, we have

$$\begin{aligned}
\Pr \left\{ \left\| \sum_{t=1}^{n} \psi(e_t) \mathbf{x}_{nt} \right\| \ge L_0 \lambda \right\} &\le \left(\frac{1}{L_0 \lambda} \right)^2 E \left\| \sum_{t=1}^{n} \psi(e_t) \mathbf{x}_{nt} \right\|^2 \\
&\le \left(\frac{1}{L_0 \lambda} \right)^2 E \sum_{t=1}^{n} \left\| \psi(e_t) \mathbf{x}_{nt} \right\|^2 \\
&\le \left(\frac{1}{L_0 \lambda} \right)^2 E \sum_{t=1}^{n} \sigma^2 \|\mathbf{x}_{nt}\|^2 \le \varepsilon.
\end{aligned} \tag{37}$$

From (36) and (37), it follows that

$$\Pr \left\{ \sup_{\boldsymbol{\tau} \in U} D(\boldsymbol{\tau}, L_0) < -L_0 \lambda \right\} > 1 - 2\varepsilon.$$

Likewise, when $n \ge n_2$, we obtain

$$\Pr \left\{ \left\| \hat{\boldsymbol{\beta}}(n) \right\| \ge L_0 \right\} < \varepsilon + 1 - (1 - 2\varepsilon) = 3\varepsilon. \tag{38}$$

Thus the result (33) follows from (38) and the arbitrariness of ε.

By Lemma 8 and $\max_{1 \le t \le n} |\mathbf{x}_{nt}^{\top} \boldsymbol{\tau}| = O(n^{-1/2})$, it follows that

$$I \left(\left\| \hat{\boldsymbol{\beta}}(n) \right\| \ge L_0 \right) \left| \sum_{t=1}^{n} \left[\psi \left(e_t - \mathbf{x}_{nt}^{\top} \hat{\boldsymbol{\beta}}(n) \right) - \psi(e_t) \right] \mathbf{x}_{nt} + \lambda \hat{\boldsymbol{\beta}}(n) \right| \xrightarrow{\mathbb{P}} 0,$$

which implies that

$$\hat{\boldsymbol{\beta}}(n) = \lambda^{-1} \sum_{t=1}^{n} \mathbf{x}_{nt} \psi(e_t) + o_p(1).$$

Consequently, (29) is proved.

As defined in (15), $\hat{\boldsymbol{\beta}}(n)$ can similarly be written as $\hat{\boldsymbol{\beta}}(n) = \mathbf{K}_n \hat{\boldsymbol{\gamma}}(n) + \mathbf{H}_n \boldsymbol{\omega}(n)$, replacing $\boldsymbol{\beta}(n)$, $\hat{\boldsymbol{\beta}}(n)$ by $\boldsymbol{\gamma}(n)$ and $\hat{\boldsymbol{\gamma}}(n)$, respectively, (30) is proved by $\mathbf{K}_n^{\top} \mathbf{K}_n = \mathbf{I}_{p-q}$. □

Proof of Theorem 1 According to (33) and Lemma 8, one gets

$$\sum_{t=1}^{n}\big[\rho\big(y_{nt}-\mathbf{x}_{nt}^{T}\hat{\boldsymbol{\beta}}(n)\big)-\rho(e_{t})\big]+\sum_{t=1}^{n}\psi(e_{t})\mathbf{x}_{nt}\big(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}(n)\big)-\frac{\lambda}{2}\big\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}(n)\big\|^{2}\xrightarrow{\mathbb{P}}0,$$

$$\sum_{t=1}^{n}\big[\rho\big(y_{nt}-\mathbf{x}_{nt}^{\top}\hat{\boldsymbol{\beta}}(n)\big)-\rho(e_{t})\big]+\frac{1}{2\lambda}\bigg\|\sum_{t=1}^{n}\psi(e_{t})\mathbf{x}_{nt}\bigg\|^{2}\xrightarrow{\mathbb{P}}0. \tag{39}$$

Similarly,

$$\sum_{t=1}^{n}\big[\rho\big(y_{nt}-\mathbf{x}_{nt}^{\top}\mathbf{K}_{n}\hat{\boldsymbol{\gamma}}(n)\big)-\rho(e_{t})\big]+\frac{1}{2\lambda}\bigg\|\sum_{t=1}^{n}\psi(e_{t})\mathbf{K}_{n}^{\top}\mathbf{x}_{nt}\bigg\|^{2}$$

$$-\sum_{t=1}^{n}\psi(e_{t})\mathbf{x}_{nt}^{\top}\mathbf{H}_{n}\boldsymbol{\omega}(n)-\frac{\lambda}{2}\big\|\boldsymbol{\omega}(n)\big\|^{2}\xrightarrow{\mathbb{P}}0. \tag{40}$$

From (14), (39) and (40), one can see that

$$2\lambda\sigma_{n}^{2}M_{n}=\bigg\|\sum_{t=1}^{n}\mathbf{H}_{n}^{\top}\mathbf{x}_{nt}\psi(e_{t})\bigg\|^{2}+\lambda^{2}\big\|\boldsymbol{\omega}(n)\big\|^{2}+2\lambda\boldsymbol{\omega}^{\top}(n)\sum_{t=1}^{n}\mathbf{H}_{n}^{\top}\mathbf{x}_{nt}\psi(e_{t})+o_{P}(1)$$

$$=\bigg\|\sum_{t=1}^{n}\mathbf{H}_{n}^{\top}\mathbf{x}_{nt}\psi(e_{t})+\lambda\boldsymbol{\omega}(n)\bigg\|^{2}+o_{P}(1). \tag{41}$$

Since $E\psi(e_{t})=0$, $E\psi^{2}(e_{t})=\sigma^{2}<\infty$, $\max_{1\leq t\leq n}\|\mathbf{x}_{nt}\boldsymbol{\tau}\|=O(n^{-1/2})$, we see by (A4) that σ_{n}^{2} is bounded by

$$\bigg\|\mathrm{var}\bigg(\sum_{t=1}^{n}\mathbf{x}_{nt}\psi(e_{t})\bigg)\bigg\|=\bigg\|\sum_{t=1}^{n}(\mathbf{x}_{nt})^{2}E\psi^{2}(e_{t})\bigg\|+2\bigg\|\sum_{t=1}^{n-1}\sum_{j=t+1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nj}^{\top}E\big(\psi(e_{t})\psi(e_{j})\big)\bigg\|$$

$$=p\sigma^{2}+2\big\|\mathbf{S}_{n}^{-1}\big\|\big\|\mathbf{x}_{t}\mathbf{x}_{j}^{\top}\big\|\sum_{t=1}^{n-1}\sum_{j=t+1}^{n}\big|E\big(\psi(e_{t})\psi(e_{j})\big)\big|$$

$$=p\sigma^{2}+2C\big\|\mathbf{x}_{t}\mathbf{x}_{j}^{\top}\big\|$$

$$=p\sigma^{2}+2pC=O(1).$$

In the view of $\mathbf{H}_{n}^{\top}\mathbf{H}_{n}=\mathbf{I}_{q}$ and Lemma 6,

$$\sigma^{-1}\sum_{t=1}^{n}\mathbf{H}_{n}^{\top}\mathbf{x}_{nt}\psi(e_{t})\xrightarrow{\mathfrak{D}}\mathrm{N}(\mathbf{0},\mathbf{I}_{p}). \tag{42}$$

Thus Theorem 1 follows immediately from (41) and (42). □

Proof of Theorem 2 Consider the model (11), without loss of generality, assume that the true parameter $\boldsymbol{\beta}(n)$ is equal to $\mathbf{0}$. For any $\delta>0$, write

$$V_{n}=E\big|\psi\big(e_{1}+\delta d_{n}^{1/2}\big)-\psi\big(e_{1}-\delta d_{n}^{1/2}\big)\big|.$$

By the monotonicity of ψ, Schwarz's inequality and (A2), we get, for sufficiently large n,

$$EI\big(\|\hat{\boldsymbol{\beta}}(n)\| \le \delta\big)\left|\hat{\sigma}_n^2 - n^{-1}\sum_{t=1}^{n}\psi^2(e_t)\right|$$

$$\le V_n + 2E\big|\psi(e_1)\big[\psi\big(e_1 + \delta d_n^{1/2}\big) - \psi\big(e_1 - \delta d_n^{1/2}\big)\big]\big|$$

$$\le V_n + 2\delta V_n^{1/2} \to 0.$$

By Lemma 9,

$$\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}\big\{\psi(e_t + h) - \psi(e_t - h)\big\} \xrightarrow{\mathbb{P}} E\psi^2(e_1) = \sigma^2.$$

Consequently, (9) is proved.

As mentioned in Chen *et al.* [21], in order to prove (10), it is desired to prove that

$$(2nh)^{-1}\sum_{t=1}^{n}\big\{\psi(e_t + h) - \psi(e_t - h)\big\} \xrightarrow{\mathbb{P}} \lambda.$$

Actually, by the monotonicity of $\psi(e_t + h) - \psi(e_t - h)$, and the assumption (8), applying Lemma 7 and Lemma 2, we get

$$\text{var}\left\{(2nh)^{-1}\sum_{t=1}^{n}\big[\psi(e_t + h) - \psi(e_t - h)\big]\right\}$$

$$\le \big(4n^2h^2\big)^{-1}E\left[\sum_{t=1}^{n}\big(\psi(e_t + h) - \psi(e_t - h)\big)\right]^2$$

$$\le \big(4nh^2\big)^{-1}E\big[\psi(e_t + h) - \psi(e_t - h)\big]^2 \to 0.$$

On the other hand, since $\lim_{n\to\infty}[G(h) - G(-h)]/(2h) = \lambda$,

$$(2nh)^{-1}\sum_{t=1}^{n}\big[\psi(e_t + h) - \psi(e_t - h)\big] = \big[G(h) - G(-h)\big]/(2h) + o_P(1) \to \lambda.$$

This completes the proof of Theorem 2. □

4 Simulation

We evaluate the parameter estimates and the M-test for the powers by Monte Carlo techniques. Under the null hypothesis, the estimators of regression coefficients and redundancy parameters are derived by some M-methods such as LS method, LAD method and Huber method. Under the local alternative hypothesis, the powers of the M-test is obtained with the rejection region given by Theorem 1. In this section, the case of the NSD sequence is raised as follows:

$$X_t = a_n Y_t + b_n Z_t, \quad t = 1, \ldots, n,$$

where a_n and b_n are positive sequences, Y_t and Z_t are negatively dependent (correspond to $\rho_0 < 0$) random variables with the distribution

$$(Y, Z) \sim \mathrm{N}\big(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_0\big).$$

Now, we will prove that $(X_1, X_2, \ldots, X_n)$ is a NSD sequence. Obviously, one may easily to check that

$$\mathrm{cov}(X_t, X_j) < 0, \quad 1 \le t < j \le n.$$

As stated in Hu [10], the superaddictivity of ϕ is equivalent to $\partial^2 \phi / \partial x_t \partial x_j \ge 0$, $1 \le t \ne j \le n$, if the function ϕ has continuous second partial derivatives. In which, $\phi(x_1, \ldots, x_n) = \exp\left(\sum_{t=1}^n X_t\right)^2$ can be chosen as a superadditive function. Note that the $\{X_t^*, 1 \le t \le n\}$ have same marginal distribution with $\{X_t, t = 1, \ldots, n\}$ for each t, by Jensen's inequality, the sequence $(X_1, X_2, \ldots, X_n)$ is proved to be NSD since

$$\frac{E\phi(X_1^*, X_2^*, \ldots, X_n^*)}{E\phi(X_1, X_2, \ldots, X_n)} = E \exp\left\{\left(\sum_{t=1}^n X_t^*\right)^2 - \left(\sum_{t=1}^n X_t\right)^2\right\}$$

$$\ge \exp E\left\{\left(\sum_{t=1}^n X_t^*\right)^2 - \left(\sum_{t=1}^n X_t\right)^2\right\} \ge 1.$$

Throughout the simulations, the Huber function is taken to be $\rho(x) = (x^2 I(|x| \le k))/2 + (k|x| - k^2/2)I(|x| > k)$, $k = 1.345\sigma_0$. The explanatory variables are generated from two random models and all of the simulations are run for 1,000 replicates and calculate the averages of the derived estimates to avoid the randomness impact.

The linear model with NSD errors is given by $y_t = \beta_0 + \beta_1 x_t + e_t$, $e_t = Y_t + Z_t$, $t = 1, 2, \ldots, n$, where the NSD errors $\{e_t, t = 1, 2, \ldots, n\}$ are assumed to follow a multivariate mixture of normal distribution with joint distribution $(Y, Z) \sim \mathrm{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_0)$, $\rho_0 < 0$. The null hypothesis is $H_0 : (\beta_0, \beta_1)^\top = (1, 2)^\top$. The sample size is taken to be $n = 100$, $n = 500$, $n = 1,000$. The joint distribution is taken to be $(Y, Z) \sim \mathrm{N}(0, 0, 1, 16, -0.5)$. The explanatory variables x_t are generated by the following two random models: I. $x_t = 5u_t$, $1 \le t \le n$; II. $x_t = \sin(2t) + 1.5u_t$, $1 \le t \le n$, where u obeys a standard uniform distribution $U(0, 1)$.

Firstly, we generate a NSD sequence by the Gibbs sampling technique. Figure 1 shows the fitted distribution (full line) of NSD is close to the normal distribution, relatively speaking, the NSD distribution tends to behave a truncated distribution feature.

Next, we evaluate the estimators of regression coefficients and redundancy parameters under the null hypothesis, Table 1 illustrates that the M-methods are valid (the corresponded M-estimates are close to true parameters $\beta_0 = 1$, $\beta_1 = 2$) and the estimators of redundancy parameters are effective (one may easily to check that $\sigma^2 = 13$ and $\lambda = 1$ when the convex function is taken to LS function, for other estimates, although their values are different based on different methods, the sign and significance remain the same, so the general conclusions remain the same). Additionally, with the increasing sample size, the estimations are more and more accurate. In fact, the estimations behave well though the sample size is not large ($n = 100$). As excepted, the fitted residual densities are close
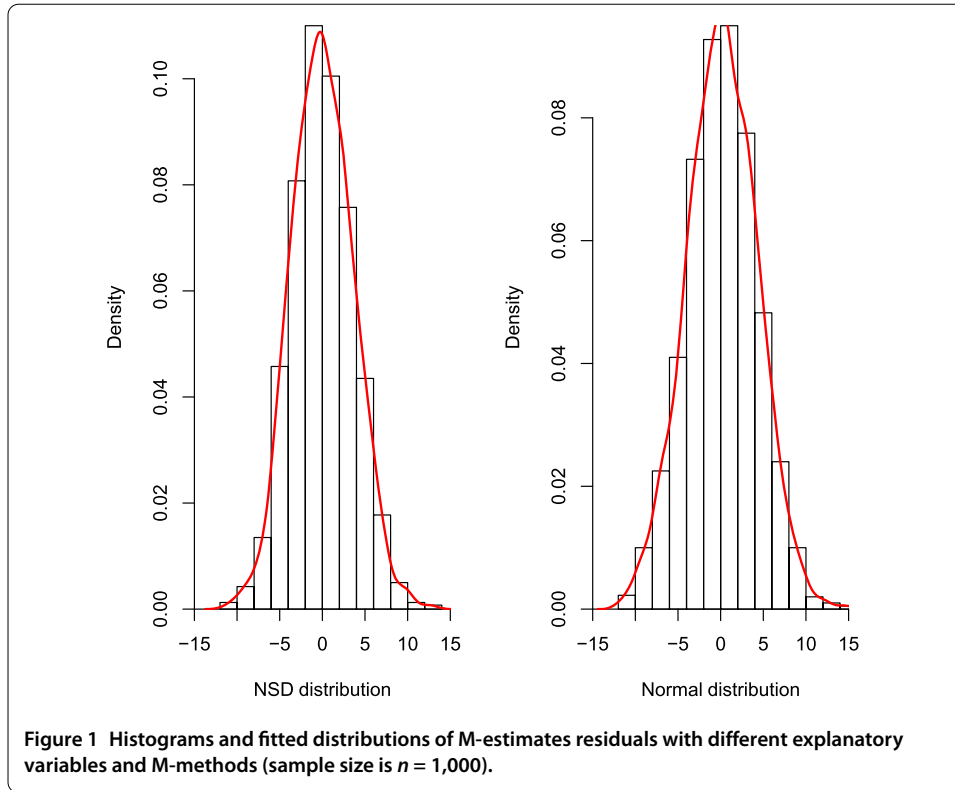
Figure 1 Histograms and fitted distributions of M-estimates residuals with different explanatory variables and M-methods (sample size is n = 1,000).

Table 1 The evaluations of regression coefficients and redundancy parameters

Estimates	n	LS		LAD		Huber	
		I	II	I	II	I	II
$\hat{\beta}_0$	100	1.031	0.985	1.016	0.978	1.006	1.007
	500	0.994	0.994	1.008	1.006	1.003	1.006
	1000	1.002	0.999	1.000	1.006	1.002	1.003
$\hat{\beta}_1$	100	1.983	2.008	1.992	2.016	2.131	2.131
	500	2.003	2.011	1.997	2.003	1.996	1.992
	1000	1.997	1.997	1.999	1.998	1.996	1.994
$\hat{\sigma}_n^2$	100	12.764	12.671	0.984	0.987	9.095	9.100
	500	12.965	12.941	0.997	0.997	9.193	9.206
	1000	12.967	12.956	0.998	0.998	9.208	9.291
$\hat{\lambda}_n$	100	1.000	1.000	0.282	0.282	0.825	0.825
	500	1.000	1.000	0.241	0.241	0.822	0.823
	1000	1.000	1.000	0.233	0.234	0.822	0.821

to the assumed NSD errors in Figure 2, and all of them still show a truncated distribution feature. Figure 3 checks the residuals are NSD by using the empirical distribution to approximate the distribution function, which supports the NSD errors assumption.

Finally, we study the empirical significant levels and the powers of M-test. Under the local hypothesis, $2\hat{\lambda}_n\hat{\sigma}_n^2 M_n$ has an asymptotic central chi-squared distribution with two degrees of freedom by Theorem 1 and Theorem 2, we may reject the null hypothesis if the simulative value $2\hat{\lambda}_n\hat{\sigma}_n^2 M_n \in W$ in (7). Table 2 presents the powers at significance levels $\alpha = 0.05$ and $\alpha = 0.01$ for various choices of M-methods, explanatory variables and different sample sizes $n = 100$, $n = 500$, $n = 1,000$. The result represents that the empirical significant levels are close to the nominal levels, consequently, the M-test is valid. Figure 4
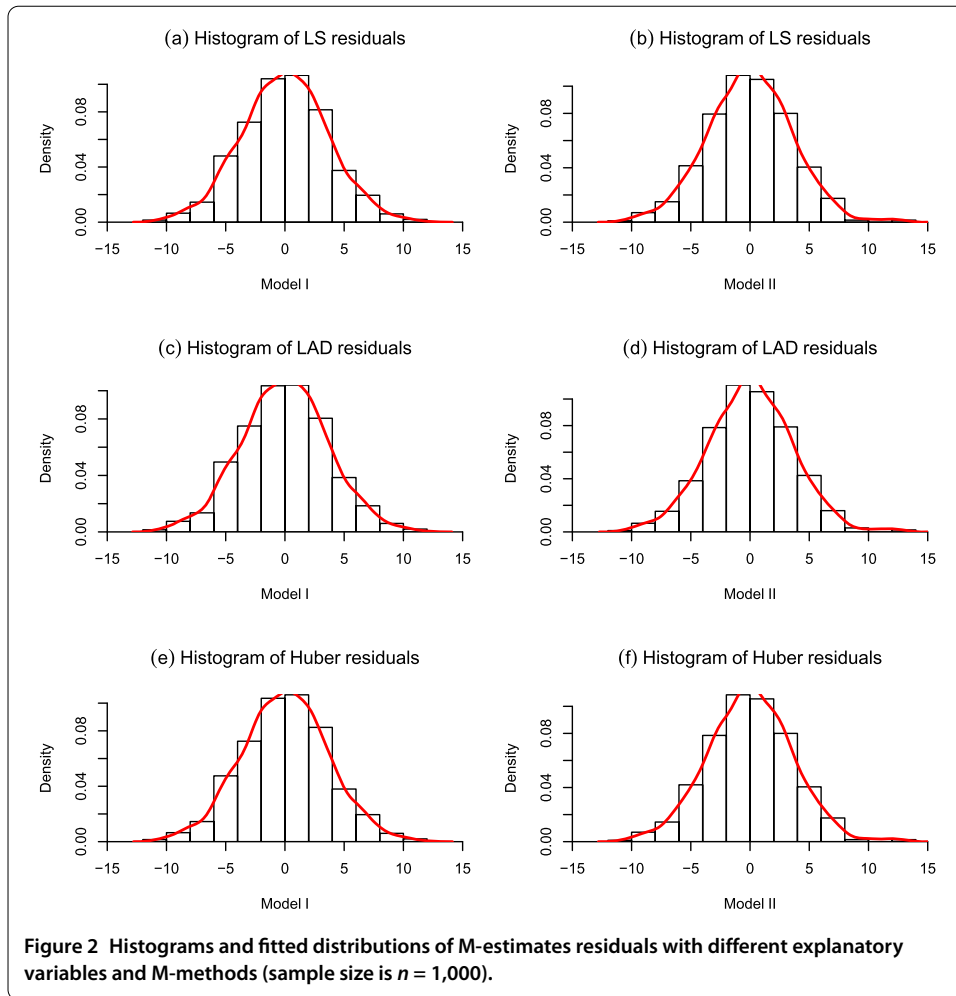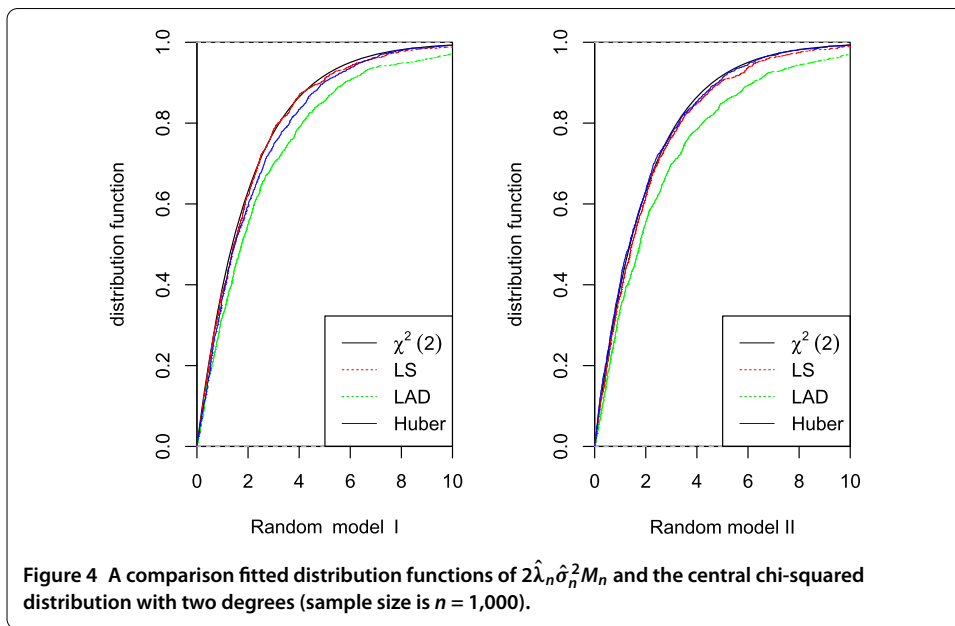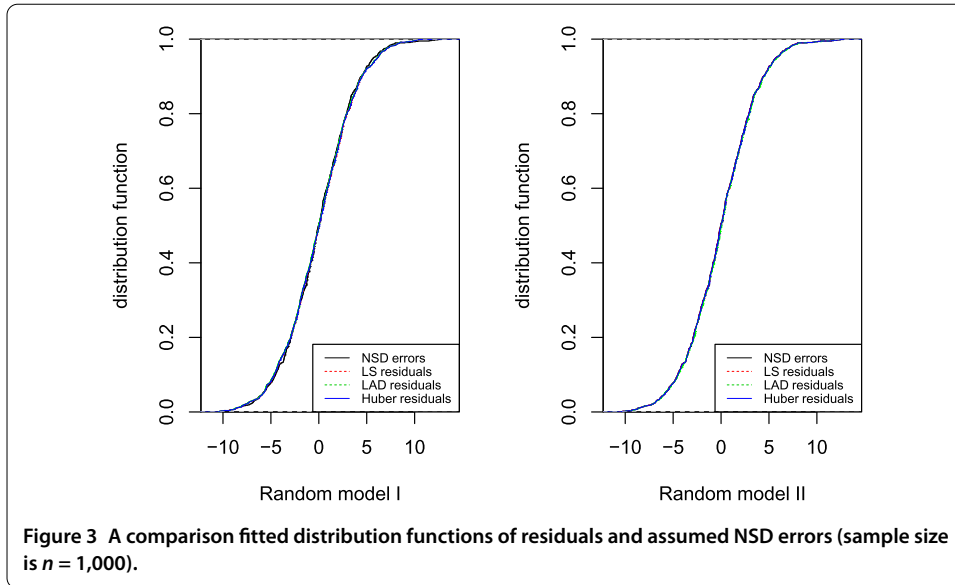
Figure 2 Histograms and fitted distributions of M-estimates residuals with different explanatory variables and M-methods (sample size is *n* = 1,000).

Table 2 The powers of the M-test with NSD errors, '∗' is for the nominal significant levels

n	Significance levels	LS		LAD		Huber	
		I	II	I	II	I	II
100	0.05*	0.063	0.068	0.082	0.079	0.062	0.062
	0.01*	0.013	0.011	0.028	0.019	0.013	0.016
500	0.05*	0.059	0.057	0.064	0.052	0.054	0.059
	0.01*	0.009	0.013	0.020	0.013	0.009	0.012
1000	0.05*	0.056	0.056	0.062	0.052	0.048	0.057
	0.01*	0.012	0.015	0.013	0.011	0.010	0.013

illustrates that $2\hat{\lambda}_n\hat{\sigma}_n^2 M_n$ can approximate the central χ_2^2 well by comparing the empirical distributions of $2\hat{\lambda}_n\hat{\sigma}_n^2 M_n$ with χ_2^2, which implies that the M-test is valid under the local alternatives.

5 Conclusions

The results presented here generalize conclusions in [20–22]. In the simulations it turns out that the M-tests for the linear model with NSD errors are insensitive to different choices of M-methods and explanatory variables, therefore it shows robustness, which illustrates that the M-test is effective.

Figure 3 A comparison fitted distribution functions of residuals and assumed NSD errors (sample size is $n = 1,000$).

Figure 4 A comparison fitted distribution functions of $2\hat{\lambda}_n \hat{\sigma}_n^2 M_n$ and the central chi-squared distribution with two degrees (sample size is $n = 1,000$).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

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

Author details

[1] State Key Laboratory of Mechanics and Control of Mechanical Structures, Institute of Nano Science and Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China. [2] Department of Mathematics and Statistics, Hubei Normal University, Huangshi, 435002, China. [3] Department of Information Science and Technology, Donghua University, Shanghai, 201600, China.

Publisher's Note

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

References

1. Huber, PJ: Robust Statistics. Wiley, New York (1981)
2. Bai, ZD, Rao, CR, Wu, Y: M-estimation of multivariate linear regression parameters under a convex discrepancy function. Stat. Sin. **2**, 237-254 (1992)
3. Elsh, AH: Bahadur representations for robust scale estimators based on regression residuals. Ann. Stat. **14**, 1246-1251 (1986)
4. Yohai, VJ: Robust estimation in the linear model. Ann. Stat. **2**, 562-567 (1974)
5. Yohai, VJ, Maronna, RA: Asymptotic behavior of M-estimators for the linear model. Ann. Stat. **7**, 258-268 (1979)
6. Berlinet, A, Liese, F, Vajda, I: Necessary and sufficient conditions for consistency of M-estimates in regression models with general errors. J. Stat. Plan. Inference **89**, 243-267 (2000)
7. Cui, HJ, He, XM, Ng, KW: M-estimation for linear models with spatially-correlated errors. Stat. Probab. Lett. **66**, 383-393 (2004)
8. Wu, WB: M-estimation of linear models with dependent errors. Ann. Stat. **35**, 495-521 (2007)
9. Wu, QY: The strong consistency of M estimator in a linear model for negatively dependent random samples. Commun. Stat., Theory Methods **40**, 467-491 (2011)
10. Hu, TZ: Negatively superadditive dependence of random variables with applications. Chinese J. Appl. Probab. Statist. **16**(2), 133-144 (2000)
11. Wang, XJ, Shen, AT, Chen, ZY: Complete convergence for weighted sums of NSD random variables and its application in the EV regression model. Test **24**, 166-184 (2015)
12. Christofides, TC, Vaggelatou, E: A connection between supermodular ordering and positive/negative association. J. Multivar. Anal. **88**(1), 138-151 (2004)
13. Nelsen, RB: An Introduction to Copulas. Springer, New York (2006)
14. Jaworski, P, Durante, F, Härdle, WK: Copulas in Mathematical and Quantitative Finance. Springer, Heidelberg (2013)
15. Eghbal, N, Amini, M, Bozorgnia, A: Some maximal inequalities for quadratic forms of negative superadditive dependence random variables. Stat. Probab. Lett. **80**, 587-591 (2010)
16. Eghbal, N, Amini, M, Bozorgnia, A: On the Kolmogorov inequalities for quadratic forms of dependent uniformly bounded random variables. Stat. Probab. Lett. **81**, 1112-1120 (2011)
17. Shen, AT, Zhang, Y, Volodin, A: Applications of the Rosenthal-type inequality for negatively superadditive dependent random variables. Chemom. Intell. Lab. Syst. **78**, 295-311 (2015)
18. Wang, XJ, Deng, X, Zheng, LL, Hu, SH: Complete convergence for arrays of rowwise negatively superadditive dependent random variables and its applications. Statistics **48**, 834-850 (2014)
19. Wang, XH, Hu, SH: On the strong consistency of M-estimates in linear models for negatively superadditive dependent errors. Aust. N. Z. J. Stat. **57**(2), 259-274 (2015)
20. Zhao, LC, Chen, XR: Asymptotic behavior of M-test statistics in linear models. J. Comput. Inf. Syst. **16**, 234-248 (1991)
21. Chen, XR, Zhao, LC: M-Methods in Linear Model. Shanghai Sci. Technol., Shanghai (1996)
22. Bai, ZD, Rao, CR, Zhao, LC: MANOVA type tests under a convex discrepancy function for the standard multivariate linear model. J. Stat. Plan. Inference **36**, 77-90 (1993)
23. Rao, CR: Tests of significance in multivariate analysis. Biometriks **35**, 58-79 (1948)
24. Rao, CR, Zhao, LC: Linear representation of M-estimates in linear models. Can. J. Stat. **20**, 359-368 (1992)
25. Zhao, LC: Strong consistency of M-estimates in linear models. Sci. China Ser. A **45**(11), 1420-1427 (2002)
26. Bai, ZD, Rao, CR, Wu, Y: M-estimation of multivariate linear regression parameters under a convex discrepancy function. Stat. Sin. **2**, 237-254 (1992)
27. Newman, CM: Normal fluctuations and the FKG inequalities. Commun. Math. Phys. **74**, 119-128 (1980)
28. Liang, HY, Jing, BY: Asymptotic properties for estimates of nonparametric regression models based on negatively associated sequences. J. Multivar. Anal. **95**(2), 227-245 (2005)
29. Yu, YC, Hu, HC: A CLT for weighted sums of NSD random sequences and its application in the EV regression model. Pure Appl. Math. **32**(5), 525-535 (2016)
30. Hu, HC, Pan, X: Asymptotic normality of Huber-Dutter estimators in a linear EV model with AR(1) processes. J. Inequal. Appl. **2014**(1), 474 (2014)
31. Anderson, PK, Gill, RD: Cox's regression model for counting processes: a large sample study. Ann. Stat. **15**, 1100-1120 (1982)
32. Billingsley, P: Convergence of Probability Measures. Wiley, New York (1968)