AISM (2007) 59: 367–384
DOI 10.1007/s10463-006-0051-y

Baiqi Miao · Yuehua Wu · Donghai Liu · Qian Tong

Asymptotic normality of the recursive M-estimators of the scale parameters

Received: 7 February 2005 / Revised: 28 November 2005 / Published online: 11 May 2006

Abstract In this paper, the limit distributions of the recursive M-estimators of scatter parameters in a multivariate linear model setting are studied. Under some mild conditions, the asymptotic normality of the recursive M-esimtators is established. Some Monte Carlo simulation results are presented to illustrate the performance of the recursive M-estimators.

Keywords M-estimation · Multivariate linear regression model · Recursive algorithm · Robust estimation · Scatter parameter · Diffusion process · Strong consistency · Asymptotic normality

1 Introduction

Consider the following multivariate linear regression model

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{e}_i, \quad i = 1, 2, \dots, \tag{1}$$

where X_i, $i = 1, 2, \dots$, are $m \times p$ matrices, $\boldsymbol{\beta}$ is a p-vector of unknown regression coefficients, and $\boldsymbol{e}_i$, $i = 1, 2, \dots$, are $m \times 1$ random errors.

There is a rich literature on estimation of the regression coefficients and scatter parameters for the model (1). The well-known method is the least squares. Even though this method is efficient for normal distributed errors and mathematically convenient, it is not resistant to outliers and stable with respect to deviations from

B. Miao · D. Liu
Department of Statistics and Finance, University of Science and Technology of China
Hefei, Anhui 230026, China

Y. Wu (✉) · Q. Tong
Department of Mathematics and Statistics, York University, 4700 Keele Street,
Toronto, ON, M3J 1P3, Canada
E-mail: wuyh@mathstat.yorku.ca

the given distributional model. Many robust statistical procedures have then been developed. The procedures based on M-estimation play an important and complementary role. However most of M-estimates have no explicit expressions. Often the Newton approach could not be applied to compute them. Hence there is a need to develop convenient algorithms for computing M-estimates, especially when the scatter parameters are also under consideration. The first attempt was made in Bickel (1975) by so called “one-step approximation”. Among other such algorithms proposed in the literature, one of them is a recursive algorithm. Note that in a recursive algorithm, a new estimate of a parameter is calculated based on the existing estimate and a new observation. In fact, recursive estimators are quick and easy to calculate and do not require extensive storage of data, which is very useful in the applications to prediction, control and target tracking. Therefore it is important to develop a recursive algorithm to compute M-estimate. Such efforts can be found in Englund et al. (1988), Englund (1993), Bai and Wu (1993), Miao and Wu (1996), Wu (1996), among others. Note that the recursive algorithms are related to the on-line learning in neural computation. See Hastie et al. (2001) for the discussion on the on-line learning.

Let $\rho(u)$ be a nonnegative function on $[0, \infty)$, and $\rho(u) = 0$ if and only if $u = 0$. The M-estimates of regression coefficients $\boldsymbol{\beta}$ and scatter parameters V for the model (1) are defined as the solutions of the following minimization problem:

$$\sum_{i=1}^{n} \left[\rho\left(\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_n\|_{\hat{V}_n}\right) + \log\left(\det\left(\hat{V}_n\right)\right)\right] = \min,$$

where $\det(V)$ denotes the determinant of a positive definite matrix V, $\|\boldsymbol{y}\|_V^2 = \boldsymbol{y}^T V^{-1}\boldsymbol{y}$, and $\boldsymbol{y}^T$ denotes the transpose of $\boldsymbol{y}$.

Suppose that ρ is continuously differentiable. Then $(\hat{\boldsymbol{\beta}}_n, \hat{V}_n)$ is the solution of the following equations

$$\begin{cases} \sum_{i=1}^{n} X_i^T V^{-1}(\boldsymbol{y}_i - X_i\boldsymbol{\beta})u_1\left(\|\boldsymbol{y}_i - X_i\boldsymbol{\beta}\|_V\right) = 0, \\ \sum_{i=1}^{n} \left[\left(\boldsymbol{y}_i - X_i\boldsymbol{\beta}\right)\left(\boldsymbol{y}_i - X_i\boldsymbol{\beta}\right)^T u_2\left(\|\boldsymbol{y}_i - X_i\boldsymbol{\beta}\|_V^2\right) - V\right] = 0, \end{cases} \tag{2}$$

where $u_1(t) = t^{-1}\rho'(t)$, $\rho'(\cdot)$ is the first order derivative of $\rho(\cdot)$, and $u_2(t) = u_1\left(\sqrt{t}\right)/2$.

When $u_1(t)$ and $u_2(t)$ are determined by the same ρ, it is difficult to keep the robustness of the M-estimates for both regression coefficients and scatter parameters simultaneously. In light of Maronna (1976), (2) may be extended to allow that u_1 and u_2 are chosen independently.

Motivated by Englund (1993), Bai and Wu (1993) proposed the following recursive algorithm for the multivariate linear regression model (1):

$$\begin{cases} \overline{\boldsymbol{\beta}}_{n+1} = \overline{\boldsymbol{\beta}}_n + S_{n+1}^{-1}a_n\boldsymbol{h}_1\left(\overline{\boldsymbol{\beta}}_n, \overline{V}_n, X_{n+1}, \boldsymbol{y}_{n+1}\right), \\ \overline{V}_{n+1} = \overline{V}_n + (n+1)^{-1}H_2\left(\overline{\boldsymbol{\beta}}_n, \overline{V}_n, X_{n+1}, \boldsymbol{y}_{n+1}\right), \end{cases} \tag{3}$$

where $S_n = \sum_{i=1}^{n} X_i^T X_i$,

$$\boldsymbol{h}_1(\boldsymbol{\beta}, V, X, \boldsymbol{e}) = X^T\widetilde{V}^{-1}(\boldsymbol{e} - X\boldsymbol{\beta})u_1\left(\|\boldsymbol{e} - X\boldsymbol{\beta}\|_{\widetilde{V}}\right),$$
$$H_2(\boldsymbol{\beta}, V, X, \boldsymbol{e}) = (\boldsymbol{e} - X\boldsymbol{\beta})(\boldsymbol{e} - X\boldsymbol{\beta})^T u_2\left(\|\boldsymbol{e} - X\boldsymbol{\beta}\|_{\widetilde{V}}^2\right) - V,$$

$\boldsymbol{\beta}_0$ is arbitrary, V_0 is a positive define matrix, write $V_0 > 0$ for simplicity, $\{a_n\}$ satisfies certain conditions, and $\widetilde{V}$ is a Lipschitz continuous $m \times m$ matrix function of V defined as follows:

Let λ_i and $\boldsymbol{\alpha}_i$ be the ith eigenvalue and corresponding eigenvector of V respectively. Then

$$\widetilde{V} = \sum_{i=1}^{m} \tilde{\lambda}_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T,$$

where $\tilde{\lambda}_i = (\delta_1 \vee \lambda_i) \wedge \delta_2$ and δ_1, δ_2 ($0 < \delta_1 < \delta_2 < \infty$) are two constants selected properly.

It is noted that when $X_i = I, i = 1, 2, \ldots,$ this recursive algorithm reduces to the one given by Englund (1993) for the multivariate location models. In Bai and Wu (1993), it was shown that under certain conditions, the recursive M-estimators given by (3) are strongly consistent.

For simplifying the recursive algorithm given in (3), Miao and Wu (1996) proposed the following algorithm:

$$\begin{cases} \boldsymbol{\beta}_{n+1} = \boldsymbol{\beta}_n + (n+1)^{-1} a_n \boldsymbol{h}_1 \left(\boldsymbol{\beta}_n, V_n, X_{n+1}, \boldsymbol{y}_{n+1}\right), \\ V_{n+1} = V_n + (n+1)^{-1} H_2 \left(\boldsymbol{\beta}_n, V_n, X_{n+1}, \boldsymbol{y}_{n+1}\right), \end{cases} \tag{4}$$

where S_{n+1} in (3) is replaced by $(n+1)$. When $X_i = I, i = 1, 2, \ldots,$ this recursive algorithm also reduces to the one given by Englund (1993) for the multivariate location models. Under certain conditions, Miao and Wu (1996) showed that the recursive M-estimators given by (4) are strongly consistent.

In statistical applications, one often needs to make statistical inferences on parameters. It is then worthwhile to find out the distribution or limit distribution of an estimator in advance. By the nature of a recursive algorithm, making a statistical inference on a parameter based on the recursive estimators is very appealing. Hence for the model (1), we need to find out the limit distributions of recursive M-estimators of regression coefficients and scatter parameters. In Miao and Wu (1996), the asymptotic normalities of $\boldsymbol{\beta}_n$ and $\overline{\boldsymbol{\beta}}_n$ are established for the recursive algorithms given by (3) and (4) respectively. Subsequently in Miao et al. (2005), optimal recursive M-estimators, asymptotic efficiencies of recursive M-estimators and asymptotic relative efficiencies between recursive M-estimators of regression coefficients are investigated. However the asymptotic distributions of the recursive M-estimators of the scatter parameters are still unsolved. In this paper, we shall tackle this problem. We shall show that under certain conditions, the recursive M-estimators of the scatter parameters are also asymptotically normal distributed.

The organization of this paper is as follows: In Sect. 2, we give notations and list the assumptions needed in this paper. The asymptotic normality of the recursive M-estimators of the scatter parameters is established in Sect. 3. Some simulation results are presented in Sect. 4. Technical lemmas are presented and proved in Appendix.

2 Notations and assumptions

In the section, we give the notations and a list of the assumptions which will be used in the paper.

Let $A = (a_{ij})$ be a $p_1 \times p_2$ matrix. Denote the Euclidean norm of A by $\|A\|$, i.e., $\|A\| = \left(\Sigma_{ij} a_{ij}^2\right)^{1/2}$ and the transpose of A by A^T. For $p_1 = p_2 = p$, denote the determinant of A by $\det(A)$, and the smallest and largest eigenvalues of A by $\lambda_*(A)$ and $\lambda^*(A)$ respectively if all the eigenvalues of A are real numbers, and the $p^2 \times 1$ column vector generalized by stacking the column vectors of A by $\mathrm{vec}(A)$. If A is a positive definite matrix, we write that $A > 0$.

In the following, $N_\varepsilon(0)$ denotes the ε-neighbor of zero; I_D denotes the indicator function for a set D, which assumes the value 1 on D and 0 on the complementary set of D. For convenience, c denotes a constant independent of the sample size and may represent different values in each appearance throughout the rest of this paper.

A function is said to be BLC($\boldsymbol{y}$) if it is bounded and Lipschitz continuous as a function of $\boldsymbol{y}$, where the distance is defined based on the Euclidean norm; and a function $u(t),\ t > 0$, is said to be a BL function if there exists a positive constant M such that

$$|u(t) - u(s)| \le M t^{-1} |t - s|, \quad \text{for any } t,\ s > 0\,.$$

Some discussion on the respective properties of the BLC and BL functions can be found in the appendix of Bai and Wu (1993).

The following assumptions are made in this paper.

(A1) $(\boldsymbol{X}_i, \boldsymbol{e}_i),\ i = 1, 2, \ldots,$ are independently and identically distributed and $\boldsymbol{X}_i$ is independent of $\boldsymbol{e}_i$. $\boldsymbol{e}_i,\ i = 1, 2, \ldots,$ have the same density function $(\det(\Sigma))^{-1/2} f\left(\|\boldsymbol{e}\|_\Sigma^2\right)$ with finite fourth moment, where $\Sigma > 0$, $f(t)$ is monotone decreasing on $[0, \infty)$, and strictly decreasing in $N_\varepsilon(0)$.

(A2) $u_1(t)$ is a non-negative decreasing BL function, $tu_1(t)$ is monotone increasing on $[0, \infty)$ and is strictly increasing in $N_\varepsilon(0)$, and $u_1{}'(t)$, the left or right derivative function of $u_1(t)$, is bounded.

(A3) $u_2(t)$ is a non-negative decreasing BL function, $tu_2(t)$ is increasing on $[0, \infty)$ and strictly increasing in $N_\varepsilon(0)$, and for some $t > 0$, $tu_2(t) > m$; there exists $M > 0$ such that $tu_2(t) \le M$ for all $t > 0$; $u_2{}'(t)$, the left or right derivative function of $u_2(t)$, is bounded, and for any vector $\boldsymbol{s} \in R^m$ and positive definite matrices $V_1, V_2 > 0$,

$$\mathrm{vec}\left[\boldsymbol{s}\boldsymbol{s}^T u_2\left(\|\boldsymbol{s}\|_{V_1}^2\right) - \boldsymbol{s}\boldsymbol{s}^T u_2(\|\boldsymbol{s}\|_{V_2}^2\right] = \Xi\,(\boldsymbol{s}, V_1, V_2)\,\mathrm{vec}(V_1 - V_2),$$

with

$$\|\,\Xi\,(\boldsymbol{s}, V_1, V_2)\| \le c < 1. \tag{5}$$

Such $u_1(t)$ and $u_2(t)$ exist. For example, consider the Huber's discrepancy function:

$$\rho_c(t) = \begin{cases} \frac{t^2}{2}, & |t| < c, \\ c|t| - c^2/2, & |t| \ge c. \end{cases}$$

Then

$$u_1(t) = \min\left\{1, \frac{c}{|t|}\right\}, \tag{6}$$

$$u_2(t) = \frac{\min\left\{1, c/\sqrt{|t|}\right\}}{2}. \tag{7}$$

where $c > 0$ is a constant.

(A4) $\{a_n\}$ is a non-negative random variable sequence and a_n is $\mathcal{F}_{n-1}$−measurable, $a_n \to a > 0$, a.s., and there exist constants v_1 and v_2 such that $0 < v_1 \leq a_n \leq v_2 < \infty$ for all n.

For describing the assumptions (A5)–(A7), we need the following notations: Suppose that f and u_2 are given as above. Let $\omega > 0$ be the solution of the equation

$$m = \int \omega z^T z u_2(\omega\|z\|^2) f(\|z\|^2)\mathrm{d}z,$$

where m is the dimension of dependent variable z. Denote

$$\Omega = \omega^{-1}\Sigma.$$

By Bai and Wu (1993), Ω satisfies that

$$\Omega = E\boldsymbol{e}\boldsymbol{e}^T u_2\left(\|\boldsymbol{e}\|^2_\Omega\right). \tag{8}$$

Define

$$\zeta = \frac{1}{4m}\int\left[\omega\|z\|^2 u_2\left(\omega\|z\|^2\right) - \frac{\omega}{1.5}\|z\|^2 u_2\left(\frac{\omega}{1.5}\|z\|^2\right)\right] f(\|z\|^2)\mathrm{d}z \wedge 0.1,$$

and

$$b_1(\Omega) = E\left[u_1(\|\boldsymbol{e}\|_\Omega) + \frac{1}{m}\|\boldsymbol{e}\|_\Omega u_1{}'(\|\boldsymbol{e}\|_\Omega)\right].$$

Assumptions (A5)–(A7) are given as follows:

(A5) $E\boldsymbol{X}_i^T\Omega^{-1}\boldsymbol{X}_i = Q > 0$, $E\|\boldsymbol{X}_1\|^4 < \infty$.

(A6) $2ab_1(\Omega) > 1$.

(A7) Suppose that $\delta_1 < \zeta\lambda_*(\Omega)$ and $\delta_2 > 3\lambda^*(\Omega)$, where δ_1 and δ_2 are used to define $\widetilde{V}_n$ in (4) and ζ is given above.

For convenience, we define the following notations, which will be used in the next section:

$$A(V, X, \boldsymbol{e}) = X^T\left[u_1(\|\boldsymbol{e}\|_V)V^{-1} + u_1{}'(\|e\|_V)\|\boldsymbol{e}\|_V^{-1}V^{-1}\boldsymbol{e}\boldsymbol{e}^T V^{-1}\right]X,$$

$$A(\Omega) = E\left[\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T \otimes \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\right)u_2{}'\left(\|\ \boldsymbol{e}_{n+1}\ \|^2_\Omega\right)\right]\left(\Omega^{-1}\otimes\Omega^{-1}\right),$$

$$B(\Omega) = E\left[\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T \otimes \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\right)u_2^2\left(\|\ \boldsymbol{e}_{n+1}\ \|^2_\Omega\right)\right] - \Omega\otimes\Omega,$$

where "$\otimes$" denotes the Kronecker product.

3 Asymptotic normality of V_n

Assuming that $X_i, i = 1, 2, \ldots$, are independently and identically distributed in Model (1), we have the following theorem:

Theorem 3.1 *Assume that* (A1)–(A7) *are satisfied. Then for the recursive algorithm (4),*

$$\sqrt{n}(\mathrm{vec}(V_n - \Omega)) \xrightarrow{\mathcal{L}} N\left(0, \int_0^\infty e^{-(I/2+A(\Omega))t} B(\Omega) e^{-(I/2+A(\Omega))^T t} \mathrm{d}t\right),$$

where $A(\Omega)$ *and* $B(\Omega)$ *are defined in Sect. 2.*

Proof Without loss of generality, we can assume that $\boldsymbol{\beta} = \mathbf{0}$. From (4) and the definition of $H_2(\boldsymbol{\beta}, V, X, e)$,

$$\begin{aligned} V_{n+1} = {} & (1-(n+1)^{-1})V_n + (n+1)^{-1} \\ & \times \left(\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\right)\left(\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\right)^T u_2\left(\|\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\|^2_{\widetilde{V}_n}\right). \end{aligned}$$

Subtracting Ω from both sides of the above equation, we have

$$\begin{aligned} V_{n+1} - \Omega &= \left(1-(n+1)^{-1}\right)(V_n - \Omega) + (n+1)^{-1} \\ &\quad \times ((\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n)(\boldsymbol{y}_{n+1} - X_{n+1}\boldsymbol{\beta}_n)^T \\ &\quad \times u_2\left(\|\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\|^2_{\widetilde{V}_n}\right) - \Omega) \\ &= \left(1-(n+1)^{-1}\right)(V_n - \Omega) + (n+1)^{-1} \\ &\quad \times \Big(\left(\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\right)\left(\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\right)^T \\ &\quad \times u_2\left(\|\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\|^2_{\widetilde{V}_n}\right) - \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\widetilde{V}_n}\right)\Big) \\ &\quad + (n+1)^{-1}\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\widetilde{V}_n}\right) - \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\Omega}\right)\right) \\ &\quad + (n+1)^{-1}\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\Omega}\right) - \Omega\right) \\ &= (1-(n+1)^{-1})(V_n - \Omega) + (n+1)^{-1}\left[\Phi_{1,n} + \Phi_{2,n} + \Phi_{3,n}\right], \end{aligned} \tag{9}$$

where

$$\begin{aligned} \Phi_{1,n} &= (\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n)\left(\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\right)^T u_2\left(\|\boldsymbol{e}_{n+1} - X_{n+1}\boldsymbol{\beta}_n\|^2_{\widetilde{V}_n}\right) \\ &\quad -\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\widetilde{V}_n}\right), \\ \Phi_{2,n} &= \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\widetilde{V}_n}\right) - \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\Omega}\right), \\ \Phi_{3,n} &= \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T u_2\left(\|\boldsymbol{e}_{n+1}\|^2_{\Omega}\right) - \Omega. \end{aligned}$$

Denote $\sqrt{n}(\mathrm{vec}(V_n - \Omega))$ by $\boldsymbol{v}_n$. Then by (9), we have

$$\boldsymbol{v}_{n+1} = (1-(n+1)^{-1})^{1/2}\boldsymbol{v}_n + (n+1)^{-1/2}\mathrm{vec}(\Phi_{1,n}) + (n+1)^{-1/2}\mathrm{vec}(\Phi_{2,n}) + (n+1)^{-1/2}\mathrm{vec}(\Phi_{3,n}). \tag{10}$$

By the definition of $\Phi_{2,n}$ and the assumption (A3), it follows that

$$\begin{aligned}\mathrm{vec}(\Phi_{2,n}) &= -\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\Omega^{-1}\otimes \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\Omega^{-1}\right)u_2{}'\left(\|\boldsymbol{e}_{n+1}\|_\Omega^2\right)\left(\mathrm{vec}(\widetilde{V}_n-\Omega)\right)\\ &\quad +o\left(\left\|\boldsymbol{e}_{n+1}\otimes\boldsymbol{e}_{n+1}(\mathrm{vec}(\widetilde{V}_n-\Omega))\right\|\right)\\ &= -\Psi_n(\boldsymbol{e}_{n+1},\Omega)\left(\mathrm{vec}(\widetilde{V}_n-\Omega)\right)+\Gamma_n(\boldsymbol{e}_{n+1},\Omega,\widetilde{V}_n)\left(\mathrm{vec}(\widetilde{V}_n-\Omega)\right).\end{aligned}$$

Based on Lemma A.1, $V_n \to \Omega$, a.s., which implies that we can select some constants $\delta_1 < \zeta\lambda_*(\Omega)$ and $\delta_2 > 3\lambda^*(\Omega)$ such that $V_n = \widetilde{V}_n$, a.s. Thus we have

$$\begin{aligned}\mathrm{vec}(\Phi_{2,n}) &= -\Psi_n(\boldsymbol{e}_{n+1},\Omega)(\mathrm{vec}(V_n-\Omega))\\ &\quad+\Gamma_n(\boldsymbol{e}_{n+1},\Omega,V_n)(\mathrm{vec}(V_n-\Omega))\end{aligned} \tag{11}$$

so that the Eq. (10) can be rewritten as

$$\begin{aligned}\boldsymbol{v}_{n+1} = \Bigg\{ I-(n+1)^{-1}\Bigg[\frac{1}{2}+o\left(\frac{1}{n+1}\right)I-\left(1+\frac{1}{n}\right)^{1/2}\Psi_n(\boldsymbol{e}_{n+1},\Omega)\\ +\left(1+\frac{1}{n}\right)^{1/2}\Gamma_n(\boldsymbol{e}_{n+1},\Omega,V_n)\Bigg]\Bigg\}\boldsymbol{v}_n+(n+1)^{-1/2}\mathrm{vec}(\Phi_{3,n}+\Phi_{1,n}).\end{aligned} \tag{12}$$

For deriving the limit distribution of $\boldsymbol{v}_n$, we need to simplify (12). Let

$$\begin{aligned}\widetilde{\boldsymbol{v}}_{n+1} = \Bigg\{ I-(n+1)^{-1}\Bigg[\left(\frac{1}{2}+o\left((n+1)^{-1}\right)\right)I-\left(1+\frac{1}{n}\right)^{1/2}\Psi_n(\boldsymbol{e}_{n+1},\Omega)\\ +\left(1+\frac{1}{n}\right)^{1/2}\Gamma_n(\boldsymbol{e}_{n+1},\Omega,V_n)\Bigg]\Bigg\}\widetilde{\boldsymbol{v}}_n+(n+1)^{-1/2}\mathrm{vec}(\Phi_{3,n}).\end{aligned}$$

By (5), we have

$$\|\Psi_n(\boldsymbol{e}_{n+1},\Omega)-\Gamma_n\,(\boldsymbol{e}_{n+1},\Omega,V_n)\|\le c_0<1$$

so that for any $n \ge 2$,

$$\begin{aligned}\lambda_*\Bigg(\left(\frac{n}{(n+1)}\right)^{1/2} I-(n(n+1))^{-1/2}\Psi_n(\boldsymbol{e}_{n+1},\Omega)\\ +(n(n+1))^{-1/2}\Gamma_n(\boldsymbol{e}_{n+1},\Omega,V_n)\Bigg)\ge\left(\frac{2}{3}\right)^{1/2}-c_0\sqrt{\frac{1}{6}}>0,\end{aligned}$$

and

$$\lambda^* \left(\left(\frac{(n+1)}{n} \right)^{1/2} \left[\Psi_n(\boldsymbol{e}_{n+1}, \Omega) - \Gamma_n(\boldsymbol{e}_{n+1}, \Omega, V_n) \right] \right) \leq \sqrt{1.5} c_0 < n+1. \tag{13}$$

Since $\boldsymbol{e}\boldsymbol{e}^T u_2 \left(\|\boldsymbol{e}\|_V^2 \right)$ is a BLC function, by Lipschitz continuity, the assumption (A5) and Lemma A.4, it follows that

$$E\|\text{vec}(\Phi_{1,n})\|^2 \leq cE\|\boldsymbol{X}_{n+1}\boldsymbol{\beta}_n\|^2 \leq cE\|\boldsymbol{X}_{n+1}\|^2 \cdot E\|\boldsymbol{\beta}_n\|^2 = o\left(\frac{1}{n} \right),$$

where c is a constant. By (8), we have

$$E(\text{vec}(\Phi_{3,n})|\mathcal{F}_n) = \boldsymbol{0}. \tag{14}$$

In view of (A1) and (A3),

$$E(\|\text{vec}(\Phi_{3,n})\|^2|\mathcal{F}_n) \leq E\left(\|\boldsymbol{e}_{n+1}\|^4 u_2^2 \left(\|\boldsymbol{e}_{n+1}\|_\Omega^2 \right) \right) \leq c < \infty. \tag{15}$$

Therefore by Lemma A.6, $\boldsymbol{v}_n$ and $\widetilde{\boldsymbol{v}}_n$ have the same limit distribution. It is apparent that $\widetilde{\boldsymbol{v}}_n$ needs to be further simplified. Since

$$E\Psi_n(\boldsymbol{e}_{n+1}, \Omega) = E\left(\boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\Omega^{-1} \otimes \boldsymbol{e}_{n+1}\boldsymbol{e}_{n+1}^T\Omega^{-1} \right) u_2{}' \left(\| \boldsymbol{e}_{n+1} \|_\Omega^2 \right) = A(\Omega),$$

we have

$$\widetilde{\boldsymbol{v}}_{n+1} = \left[I - (n+1)^{-1} \left(\frac{I}{2} + A(\Omega) + G_n \right) \right] \widetilde{\boldsymbol{v}}_n + (n+1)^{-1/2}\text{vec}(\Phi_{3,n}),$$

where

$$G_n = o(1) + \left[\Psi_n(\boldsymbol{e}_{n+1}, \Omega) - E\Psi_n(\boldsymbol{e}_{n+1}, \Omega) \right] + \Gamma_n\left(\boldsymbol{e}_{n+1}, \Omega, \widetilde{V}_n \right) \tag{16}$$

satisfying that

$$\begin{aligned} &\| E(G_n|\mathcal{F}_n) \| = o(1), \\ &E(\| G_n \|^2 |\mathcal{F}_n) \leq E(\| o(1) + c_0 \|^2 |\mathcal{F}_n) \leq c^2 < \infty. \end{aligned} \tag{17}$$

Note that $\lambda_* \left(I + A(\Omega) + (A(\Omega))^T \right) > 0$. Let

$$\approx{\boldsymbol{v}}_{n+1} = \left[I - (n+1)^{-1} \left(\frac{I}{2} + A(\Omega) \right) \right] \approx{\boldsymbol{v}}_n + (n+1)^{-1/2}\text{vec}(\Phi_{3,n}). \tag{18}$$

Hence, by (14), (15), (16), (17) and Lemma A.3, it follows that $\widetilde{\boldsymbol{v}}_n$ and $\overset{\approx}{\boldsymbol{v}}_n$ have the same limit distribution. □

Express (18) as

$$\begin{aligned}\overset{\approx}{\boldsymbol{v}}_{n+1} &= \overset{\approx}{\boldsymbol{v}}_n + (n+1)^{-1}\left(-\frac{I}{2} - A(\Omega)\right)\overset{\approx}{\boldsymbol{v}}_n \\ &\quad + (n+1)^{-1/2}B^{1/2}(\Omega)\left(B^{-1/2}(\Omega)\mathrm{vec}(\Phi_{3,n})\right) \\ &= \overset{\approx}{\boldsymbol{v}}_n + \ell_n^2\boldsymbol{b}(\overset{\approx}{\boldsymbol{v}}_n) + \ell_n Z(\overset{\approx}{\boldsymbol{v}}_n)\boldsymbol{\xi}_{n+1},\end{aligned}$$

where

$$\ell_n = (n+1)^{-1/2}, \quad \boldsymbol{b}(\boldsymbol{v}) = -\left(\frac{I}{2} + A(\Omega)\right)\boldsymbol{v},$$
$$Z(\boldsymbol{v}) = B^{1/2}(\Omega), \quad \boldsymbol{\xi}_{n+1} = B^{-1/2}(\Omega)\mathrm{vec}(\Phi_{3,n}).$$

It can be seen that ℓ_n, $\boldsymbol{b}(\boldsymbol{v})$, $Z(\boldsymbol{v})$ and $\boldsymbol{\xi}_n$ satisfy the assumptions of Lemma A.7, thus $\overset{\approx}{\boldsymbol{v}}_n$ weakly converges to invariability measure of a diffusion process with diffusion factor

$$L = \frac{1}{2}\sum_{i,j=1}^{m^2} B_{ij}(\Omega)\frac{\partial^2}{\partial x_i \partial x_j} - \sum_{i=1}^{m^2}\left[\left(\frac{I}{2} + A(\Omega)\right)\boldsymbol{x}\right]_i \frac{\partial}{\partial x_i},$$

where B_{ij} is the (i, j)−entry of $B(\Omega)$. Solving this diffusion process, we have

$$\boldsymbol{x}_t = \mathrm{e}^{-(I/2+A(\Omega))t}\boldsymbol{x}_0 + \int_0^t \mathrm{e}^{-(I/2+A(\Omega))(t-s)}B(\Omega)\mathrm{e}^{-(I/2+A(\Omega))^T(t-s)}d\boldsymbol{w}(s).$$

Here $\boldsymbol{x}_0$ is a random vector with finite covariance matrix and independent of $\{\boldsymbol{w}(t) : t > 0\}$, the Brownian motion. Therefore, for $n \to \infty$,

$$\overset{\approx}{\boldsymbol{v}}_n \xrightarrow{\mathcal{L}} N\left(0, \int_0^\infty \mathrm{e}^{-(I/2+A(\Omega))t}B(\Omega)\mathrm{e}^{-(I/2+A(\Omega))^T t}\mathrm{d}t\right),$$

which implies that $\sqrt{n}(\mathrm{vec}(V_n - \Omega)) \xrightarrow{\mathcal{L}} N(0, \int_0^\infty e^{-(I/2+A(\Omega))t}B(\Omega)\mathrm{e}^{-(I/2+A(\Omega))^T t}\mathrm{d}t)$.

Remark 3.1 Under the same conditions as given in Theorem 3.1, it can be shown that

$$\sqrt{n}\left(\mathrm{vec}(\overline{V}_n - \Omega)\right) \xrightarrow{\mathcal{L}} N\left(0, \int_0^\infty \mathrm{e}^{-(I/2+A(\Omega))t}B(\Omega)\mathrm{e}^{-(I/2+A(\Omega))^T t}\mathrm{d}t\right),$$

where $\overline{V}_n$ is given in (3).

Remark 3.2 By Assumptions (A2) and (A3), it can be seen that the recursive estimators computed by (3) or (4) inherit the breakdown properties of the initial estimator if there are no abnormality in $\{X_i,\ 1 \le i \le n\}$. To be detailed, the recursive estimators by (3) will not be bounded if $\lambda_*(S_n) = 0$, while the recursive estimators by (4) will not be bounded if some values of a X_i are infinite. A simulation study of the robustness of the recursive algorithm (4) is provided in the next section.

4 Simulation results

In this section, we will study the finite sample performance of the recursive algorithm given by (4), where u_1 and u_2 are respectively given by (6) and (7) with $c = 3$. In our simulation, $a_n = 0.9, \delta_1 = 0.1, \delta_2 = 5$, and $\boldsymbol{\beta} = (-2, 6)'$. Let $\Psi = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. In our simulation, $\text{vec}(X_i')$, $i = 1, 2, \ldots$, are independently and identically distributed as $N((1, 1, 1, 2, 1, 1)', \text{diag}(\Psi, \Psi, \Psi))$. Let

$$A = \begin{pmatrix} 0.8 & 0 & 0 \\ 0.2 & 1.2 & 0 \\ 0.1 & 0.4 & 1.1 \end{pmatrix},$$

and $\Sigma = AA'$. The random error vectors $\boldsymbol{e}_i, i = 1, 2, \ldots$, are generated from $0.9N(\mathbf{0}, \Sigma) + 0.1N(\mathbf{0}, 36\Sigma)$. Then

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{e}_i, \quad i = 1, 2, \ldots.$$

Eight subplots in Fig. 1 show the behaviours of the recursive estimators computed by (4) for $n = 1, \ldots, 1{,}000$ against the M-estimators computed by Newton–Raphson iterative algorithm for $n = 80, 120, \ldots, 1{,}000$. By Figure 1, it can be seen that the recursive estimators perform comparably with the M-esimators by Newton–Raphson method. In the actual computation, the Newton–Raphson iterative algorithms are selective for the initial values, which have to be close to the true values, while the algorithm (4) is flexible in contrast. We have used $\boldsymbol{\beta}_0 = \mathbf{0}$ and $V_0 = I_3$ as initial values for the algorithm (4) in our simulation. The time for computing the recursive estimators are significantly less than the time for computing the M-estimators by Newton–Raphson iterative algorithm.

Further, based on the previous model setting for simulation, we select an element $\boldsymbol{y}_i$ randomly with equal probability and add 30 to it with probability 0.05. All the data in Fig. 2 are computed in this new model setting, which gives the performance of the recursive estimators against Lease Squares estimators. It can be seen that the recursive estimators outperform the Least Squares estimators in the simulation.

Back to the original model setting in the section, we carry out the simulation 1,000 times for $n = 100, 500, 1, 000$, respectively. Figs. 3, 4 and 5 provide the histograms of the recursive estimators. It reveals that the larger the sample size, the more bell-shaped the distributions of the recursive estimators become.

Appendix

Lemma A.1 (Miao and Wu 1996) *Assume that Conditions* $(A1)$–$(A7)$ *are satisfied. Then for* $(\boldsymbol{\beta}_n, V_n)$ *defined by (1.3) we have*

$$(\boldsymbol{\beta}_n, V_n) \to (\boldsymbol{\beta}, \Omega), \quad a.s.$$

Lemma A.2 *Assume that* $\tau > 0,\ 0 < q_n \to 0,\ \sum q_n = \infty,\ \varepsilon_n \leq \varepsilon,\ 0 \leq r_n \leq r,$ *and* $b_n \geq b$*, which satisfy the following recursive equations:*

$$b_{n+1} \leq (1 - \tau q_n)b_n + \varepsilon_n q_n,$$

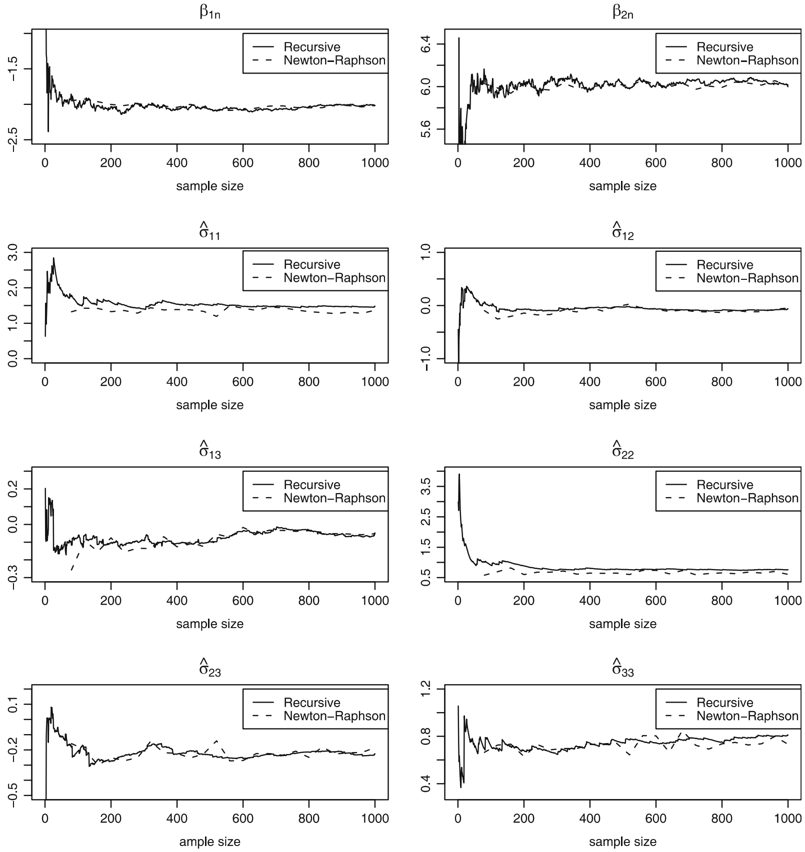

Fig. 1 The estimates by the recursive algorithm (4) and Newton–Raphson iterative algorithm

or

$$b_{n+1} \le (1-\tau q_n)b_n + \varepsilon_n q_n + q_n^{3/2}\sqrt{|b_n|}r_n,$$

where τ, r, ε *and* b *are constants. Then*

$$\lim_{n\to\infty} b_n = b_0 \le \frac{\varepsilon}{\tau}.$$

The proof is similar to the proof of Lemma 3.3 of Miao and Wu (1996).

Corollary A.1 *If* $\tau > 0$, $0 < q_n \to 0$, $\sum q_n = \infty$, $b_n > 0$, $\varepsilon_n \to 0$, $r_n \to 0$, *and* $b_{n+1} \le (1-\tau q_n)b_n + r_n q_n |b_n|^{1/2} + \varepsilon_n q_n$, *then* $\lim_{n\to\infty} b_n = 0$.

Lemma A.3 *Suppose that* $\boldsymbol{u}_n \in R^p$ *and* $\widetilde{\boldsymbol{u}}_n \in R^p$ *satisfy the following recursive equation*

$$\begin{aligned}\boldsymbol{u}_{n+1} &= (I - q_n\Psi)\boldsymbol{u}_n + q_n^{1/2}\boldsymbol{v}_n,\\ \widetilde{\boldsymbol{u}}_{n+1} &= (I - q_n(\Psi+\Gamma_n))\widetilde{\boldsymbol{u}}_n + q_n^{1/2}\boldsymbol{v}_n,\end{aligned}$$

and

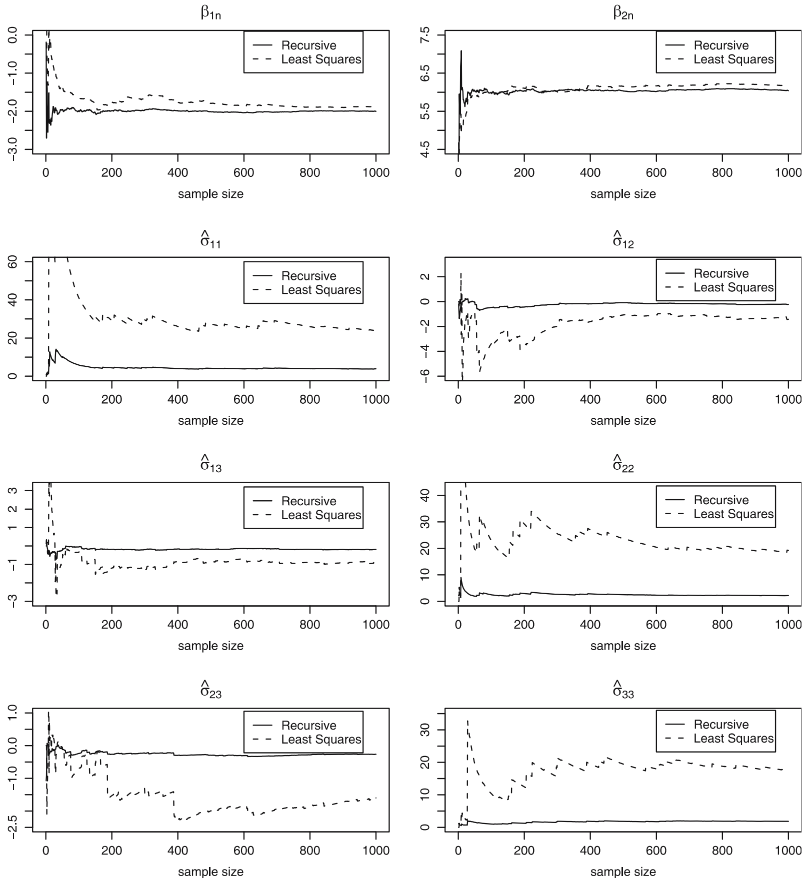

Fig. 2 The estimates of regression coefficients and scatter parameters by the recursive algorithm (4) and least squares method

(i) $\boldsymbol{u}_1 = \widetilde{\boldsymbol{u}}_1 = \boldsymbol{c}$,
(ii) Ψ *is a* $p \times p$ *matrix, and* $\lambda_*(\Psi + \Psi^T) > 0$,
(iii) $\{q_n\}$ *is a sequence of constants,* $0 < q_n \to 0,\ \sum q_n = \infty$,
(iv) $\boldsymbol{v}_n \in R^p$ *is* $\mathcal{F}_{n+1}-$*measurable,* $E(\boldsymbol{v}_n|\mathcal{F}_n) = \mathbf{0},\ E\left(\|\boldsymbol{v}_n\|^2|\mathcal{F}_n\right) \le c_1 < \infty$,
(v) Γ_n, *a* $p \times p$ *matrix, is* $\mathcal{F}_{n+1}-$*measurable,* $\|E(\Gamma_n|\mathcal{F}_n)\| \le r_n,\ E\left(\|\Gamma_n\|^2|\mathcal{F}_n\right) \le c_2 < \infty, r_n \to 0$,

where $\boldsymbol{c}$ *is a constant vector and* $c_i > 0, i = 1, 2$, *are constants, then* $\boldsymbol{u}_n$ *and* $\widetilde{\boldsymbol{u}}_n$ *have the same limit distribution, and there exists a constant* $c_0 > 0$ *such that* $E\left(\|\widetilde{\boldsymbol{u}}_{n+1}\|^2\right) \le c_0 < \infty$.

Proof Denote $\lambda_*(\Psi + \Psi^T)/2$ by λ. Then $\lambda > 0$ by the assumption (ii). When n is large enough,

$$\begin{aligned} E\left(\|\widetilde{\boldsymbol{u}}_{n+1}\|^2\right) &= E\left(\|(I - q_n(\Psi + \Gamma_n))\widetilde{\boldsymbol{u}}_n\|^2\right) + q_n^{1/2} E\left[\left((I - q_n(\Psi + \Gamma_n))\,\widetilde{\boldsymbol{u}}_n\right)^T \boldsymbol{v}_n\right] \\ &\quad + q_n^{1/2} E\left[\boldsymbol{v}_n^T (I - q_n(\Psi + \Gamma_n))\widetilde{\boldsymbol{u}}_n\right] + q_n E(\|\boldsymbol{v}_n\|^2) \\ &\le \left(1 - \frac{\lambda}{2} q_n\right) E\left(\|\widetilde{\boldsymbol{u}}_n\|^2\right) + c q_n^{1/2} \cdot E\left(\|\widetilde{\boldsymbol{u}}_n\|^2\right)^{1/2} + q_n E\left(\|\boldsymbol{v}_n\|^2\right). \end{aligned}$$

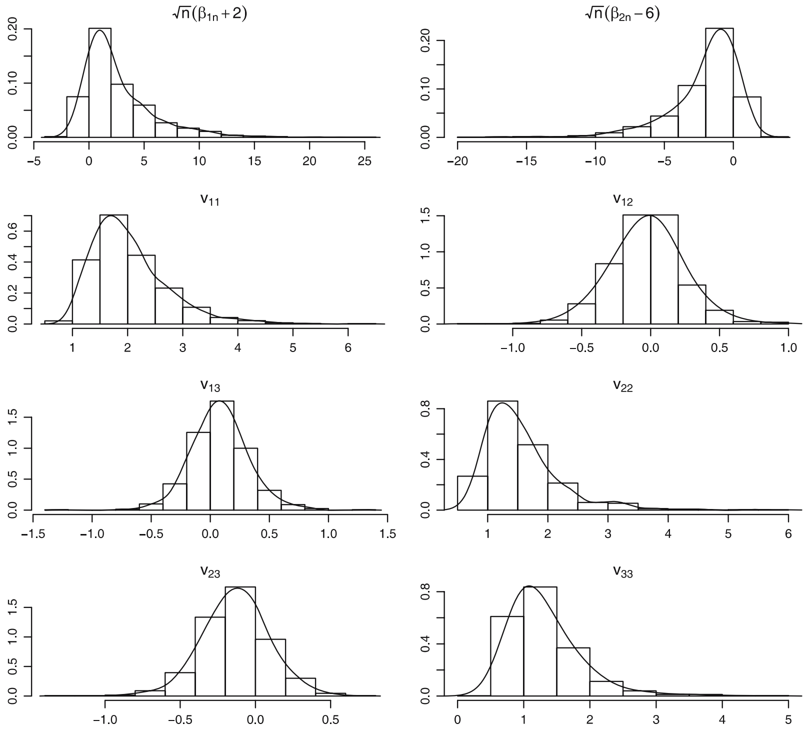

Fig. 3 Histogram of the estimates of regression coefficients and scatter parameters by the recursive algorithm (4) for $n = 100$

By Lemma A.2, there exists a constant $c_0 > 0$ such that $E\left(\|\widetilde{\boldsymbol{u}}_n\|^2\right) < c_0 < \infty$. Write $\boldsymbol{\delta}_{n+1} = \boldsymbol{u}_{n+1} - \widetilde{\boldsymbol{u}}_{n+1}$. Then

$$\boldsymbol{\delta}_{n+1} = (I - q_n\Psi)\boldsymbol{\delta}_n + q_n\Gamma_n\widetilde{\boldsymbol{u}}_n.$$

Hence,

$$\begin{aligned} E\left(\|\boldsymbol{\delta}_{n+1}\|^2\right) &= E\left(\|(I - q_n\Psi)\boldsymbol{\delta}_n\|^2\right) + q_n E\left[((I - q_n\Psi)\boldsymbol{\delta}_n)^T (\Gamma_n\widetilde{\boldsymbol{u}}_n)\right] \\ &\quad + q_n E\left[(\Gamma_n\widetilde{\boldsymbol{u}}_n)^T (I - q_n\Psi)\boldsymbol{\delta}_n\right] + q_n^2 E\left(\|\Gamma_n\widetilde{\boldsymbol{u}}_n\|^2\right) \\ &\leq (1 - \lambda q_n)E\left(\|\boldsymbol{\delta}_n\|^2\right) + cr_n\left(E\left(\|\boldsymbol{\delta}_n\|^2\right)\right)^{1/2} q_n + cq_n^2. \end{aligned}$$

Applying Corollary A.1, we have $E\left(\|\boldsymbol{\delta}_{n+1}\|^2\right) \to 0$, which implies that $\boldsymbol{u}_n$ and $\widetilde{\boldsymbol{u}}_n$ have the same limit distribution. □

Lemma A.4 *Suppose* $(A1)$–$(A7)$ *are satisfied, and* $2ab_1(\Omega) > 1$, *then*

$$E\|\sqrt{n}(\boldsymbol{\beta}_n - \boldsymbol{\beta})\|^2 \leq c < \infty\ .$$

Proof Write $\widetilde{\boldsymbol{\beta}}_n = \sqrt{n}(\boldsymbol{\beta}_n - \boldsymbol{\beta})$. By Miao and Wu (1996), $\widetilde{\boldsymbol{\beta}}_{n+1}$ can be written as follows:

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}_{n+1} &= \left\{I - (n+1)^{-1}\left[\left(ab_1(\Omega) - \frac{1}{2}\right) I + W_n\right]\right\}\widetilde{\boldsymbol{\beta}}_n \\ &\quad + (n+1)^{-1/2} a_n X_{n+1}^T \widetilde{V}_n^{-1}\boldsymbol{e}_{n+1} u_1\left(\|\boldsymbol{e}_{n+1}\|_{\widetilde{V}_n}\right), \end{aligned}$$

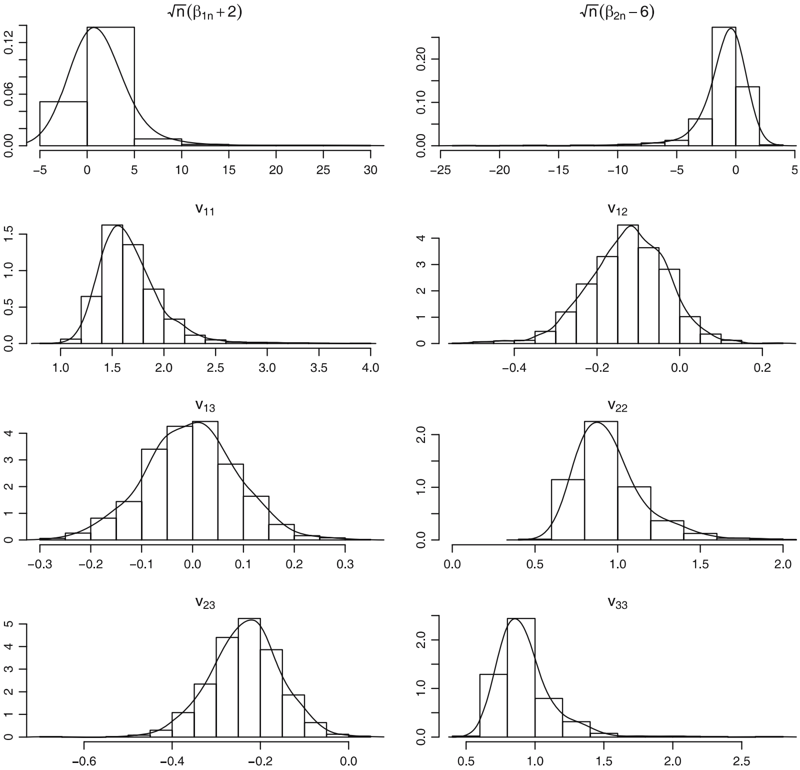

Fig. 4 Histogram of the estimates of regression coefficients and scatter parameters by the recursive algorithm (4) for $n = 500$

where W_n satisfies the condition (v) of Lemma A.3, $\boldsymbol{v}_n = a_n X_{n+1}^T \widetilde{V}_n^{-1} \boldsymbol{e}_{n+1} u_1 \left(\|\boldsymbol{e}_{n+1}\|_{\widetilde{V}_n}\right)$ satisfies the condition (iv) of Lemma A.3. Hence, by Lemma A.3,

$$E\|\widetilde{\boldsymbol{\beta}}_{n+1}\|^2 = nE\|\boldsymbol{\beta}_n - \boldsymbol{\beta}\|^2 \leq c < \infty .$$

□

Lemma A.5 *Assume that*

$$0 < q_n \to 0, \quad \text{and} \quad \sum_{n=1}^{\infty} q_n = \infty.$$

Then for any $\delta > 0$,

$$\sup_n \sum_{k=1}^{n-1} q_k \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) < \infty.$$

Proof Note that $\exp(-b) \leq 1 - b + b^2/2$ for $b > 0$. Hence for any $\delta > 0$, and $K < n$,

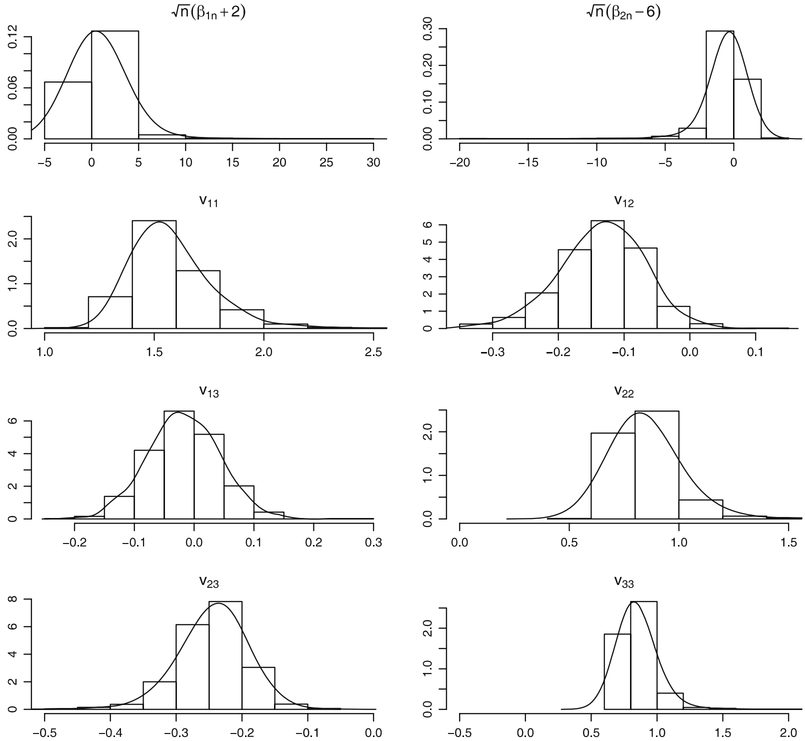

Fig. 5 Histogram of the estimates of regression coefficients and scatter parameters by the recursive algorithm (4) for $n = 1{,}000$

$$
\begin{aligned}
\sum_{k=K}^{n-1} q_k \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) &\le \frac{1}{\delta} \sum_{k=K}^{n} \left(1 - \mathrm{e}^{-\delta q_k} + \frac{(\delta q_k)^2}{2}\right) \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) \\
&= \frac{1}{\delta} \sum_{k=K}^{n-1} \left[\exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) - \exp\left(-\delta \sum_{j=k}^{n} q_j\right)\right] \\
&\quad + \frac{\delta}{2} \sum_{k=K}^{n-1} q_k^2 \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) \\
&\le \frac{1}{\delta} \left[\exp(-\delta q_n) - \exp\left(-\delta \sum_{j=K}^{n} q_j\right)\right] + \frac{\delta}{2} \sum_{k=K}^{n} q_k^2 \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right),
\end{aligned}
$$

which implies that

$$
\sum_{k=K}^{n-1} \left(1 - \frac{\delta}{2} q_k\right) q_k \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) \le \frac{1}{\delta}. \tag{19}
$$

In view of that $q_k \to 0$, there exists a positive integer K_2 such that $q_k\delta < 1$ and $q_k < 2q_k(1-\delta q_k/2)$ for $k \geq K_2$. Let $K_0 = \max\{K_1, K_2\}$. Therefore, for any $\delta > 0$, by (19),

$$\begin{aligned}\sum_{k=1}^{n-1} q_k \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) &\leq c \sum_{k=1}^{K_0-1} q_k + \sum_{k=K_0}^{n-1} q_k \exp\left(-\delta \sum_{j=k+1}^{n} q_j\right) \\ &\leq c \sum_{k=1}^{K_0-1} q_k + 2\sum_{k=K_0}^{n-1}\left(1-\frac{\delta}{2}q_k\right) q_k \exp\left(-\delta \sum_{i=k+1}^{n} q_i\right) \leq c\sum_{k=1}^{K_0-1} q_k + \frac{2}{\delta},\end{aligned}$$

where c is a constant. □

Lemma A.6 *Define* $\boldsymbol{\tau}_n \in R^d$ *and* $\widetilde{\boldsymbol{\tau}}_n \in R^d$ *as follows:*

$$\begin{aligned}\boldsymbol{\tau}_{n+1} &= \left(I - (n+1)^{-1}\Psi_n\right)\boldsymbol{\tau}_n + (n+1)^{-1/2}(\boldsymbol{\nu}_n + \boldsymbol{\gamma}_n),\\ \widetilde{\boldsymbol{\tau}}_{n+1} &= \left(I - (n+1)^{-1}\Psi_n\right)\widetilde{\boldsymbol{\tau}}_n + (n+1)^{-1/2}\boldsymbol{\nu}_n,\end{aligned}$$

where $\boldsymbol{\tau}_1 = \widetilde{\boldsymbol{\tau}}_1 = \boldsymbol{c}_0, \boldsymbol{\nu}_n \in R^d$, $\boldsymbol{\gamma}_n \in R^d$ *and* Ψ_n *is a* $d \times d$ *matrix. Assume that for* $n \geq n_0$ *and a constant* $\kappa > 0$, $\kappa \leq \lambda_*(\Psi_n) \leq \lambda^*(\Psi_n) \leq n+1$. *If* $E\left(\|\boldsymbol{\gamma}_n\|^2\right) = o(1)$, *then* $\boldsymbol{\tau}_n - \widetilde{\boldsymbol{\tau}}_n \xrightarrow{P} 0$.

Proof Denote $\boldsymbol{\delta}_n = \boldsymbol{\tau}_n - \widetilde{\boldsymbol{\tau}}_n$, $\widetilde{\Psi}_{n,k} = \prod_{j=k}^{n}[I-(j+1)^{-1}\Psi_j]$ for $1 \leq k \leq n$ and $\widetilde{\Psi}_{n,n+1} = I$. It can be seen that

$$\begin{aligned}\boldsymbol{\delta}_{n+1} &= \left[I - (n+1)^{-1}\Psi_n\right]\boldsymbol{\delta}_n + (n+1)^{-1/2}\boldsymbol{\gamma}_n\\ &= \sum_{k=n_0}^{n}\widetilde{\Psi}_{n,k+1}(k+1)^{-1/2}\boldsymbol{\gamma}_k + \widetilde{\Psi}_{n,n_0}\boldsymbol{\delta}_{n_0}.\end{aligned}$$

Therefore, by the assumption and Lemma A.5, we have

$$\begin{aligned}E\left(\|\boldsymbol{\delta}_{n+1}\|^2\right) &\leq E\left(\sum_{k=n_0}^{n}\|\widetilde{\Psi}_{n,k+1}(k+1)^{-1/2}\boldsymbol{\gamma}_k\|^2\right) + E\left(\|\widetilde{\Psi}_{n,n_0}\boldsymbol{\delta}_{n_0}\|^2\right)\\ &\leq \sum_{k=n_0}^{n-1}\prod_{j=k+1}^{n}\left[1-\kappa(j+1)^{-1}\right]^2(k+1)^{-1}E\|\boldsymbol{\gamma}_k\|^2 + (n+1)^{-1}\|\boldsymbol{\gamma}_n\|^2\\ &\quad +1\prod_{j=n_0}^{n}\left[1-\kappa(j+1)^{-1}\right]^2 E\|\boldsymbol{\delta}_{n_0}\|^2\\ &\leq \sum_{k=n_0}^{n-1}\exp\left(-2\kappa\sum_{j=k+1}^{n}(j+1)^{-1}\right)(k+1)^{-1}E\|\boldsymbol{\gamma}_k\|^2 + (n+1)^{-1}\|\boldsymbol{\gamma}_n\|^2\\ &\quad + \exp\left(-2\kappa\sum_{j=n_0}^{n}(j+1)^{-1}\right)E\|\boldsymbol{\delta}_{n_0}\|^2 = o(1).\end{aligned}$$

Then the lemma follows. □

Lemma A.7 (Basak et al. 1997) *Let the stochastic process $\boldsymbol{x}_n$ be defined recursively by*

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \ell_n^2 \boldsymbol{b}(\boldsymbol{x}_n) + \ell_n Z(\boldsymbol{x}_n)\boldsymbol{\xi}_{n+1},$$

where $\{\boldsymbol{x}_n\}$ are d-dimensional random vectors, $\boldsymbol{b} : R^d \to R^d$ and $Z : R^d \to R^{d\times r}$ are locally Lipschitz continuous functions, $\{\boldsymbol{\xi}_n\}$ are r-dimensional random vectors, and $\{\ell_n\}$ is a sequence of constants. Assume that the following conditions are satisfied:

(i) $\|\boldsymbol{b}(\boldsymbol{x})\| + \|Z(\boldsymbol{x})\| \le K(1 + |\boldsymbol{x}|)$ *for some constant* $K > 0$, $Z(\boldsymbol{x})Z(\boldsymbol{x})^T > 0,\ \forall\, \boldsymbol{x} \in R^d$,

(ii) *There exist a positive-definite matrix D and positive constants c_0, c_1 such that* $2(D\boldsymbol{x})^T\boldsymbol{b}(\boldsymbol{x}) + \mathrm{tr}(Z(\boldsymbol{x})^T DZ(\boldsymbol{x})) \le -c_0|\boldsymbol{x}|^2 + c_1, \quad \forall \boldsymbol{x} \in R^d$,

(iii) $E(\boldsymbol{\xi}_n\boldsymbol{\xi}_n^T|\mathcal{F}_n) = I$ *for any n,* $E(\boldsymbol{\xi}_n|\mathcal{F}_n) = \boldsymbol{0}$, *and* $\sup_{n>1} E(\|\boldsymbol{\xi}_n\|^{2+\theta}|\mathcal{F}_{n-1}) < \infty$, *a.s. for some $\theta > 0$,*

(iv) $\sum_{n=1}^{\infty} \ell_n^2 = \infty$, *and* $\lim_{n\to\infty} \ell_n = 0$.

Then the stochastic process $\boldsymbol{x}_n$ weakly converges to invariability measure of a diffusion process with diffusion factor

$$L = \frac{1}{2}\sum_{ij}^{d} \alpha_{ij}(\boldsymbol{x})\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(\boldsymbol{x})\frac{\partial}{\partial x_i},$$

where $b_i(\boldsymbol{x})$ is the i-th element of $\boldsymbol{b}(\boldsymbol{x})$, $\alpha_{ij}(\boldsymbol{x}) = \sum_{k=1}^{r} z_{ik}(\boldsymbol{x})z_{jk}(\boldsymbol{x})$, and $Z(\boldsymbol{x}) = (z_{ij}(\boldsymbol{x}))$.

Acknowledgements The research of B.Q. Miao was partially supported by the Natural Sciences and Engineering Research Council of Canada, the National Natural Science Foundation of China and the Doctoral Fund of Education Ministry of China. The research of Y. Wu was partially supported by the Natural Sciences and Engineering Research Council of Canada. The research of D.H. Liu was partially supported by the National Natural Science Foundation of China. The authors would like to thank the referee for the helpful comments and suggestions, which has improved this manuscript.

References

Bai, Z. D., Wu, Y. (1993). Recursive algorithm for M-estimates of regression coefficients and scatter parameters in linear models. *Sankhyā, Ser. B* Part 2 *55*, 199–218.

Basak, G. K., Hu, I., Wei, C. Z. (1997). Weak convergence of recursions. *Stochastic Processes and their Applications 68*, 65–82.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association 70*, 428–434.

Englund, J. -E., Holst, U., Ruppert, D. (1988). Recursive M-estimate of location and scale for dependent sequences. *Scandinavian Journal of Statistics 15*, 147–159.

Englund, J. -E. (1993). Multivariate recursive M-estimate of location and scatter for dependent sequences. *Journal of Multivariate Analysis 45*, 257–273.

Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Berlin Heidelberg New York: Springer

Maronna, R. A. (1976). Robust estimators of multivariate location and scatter. *Annals of Statistics 4*, 51–67.

Miao, B. Q., Wu, Y. (1996). Limiting behavior of recursive M-estimate in multivariate linear regression models. *Journal of Multivariate Analysis 59*, 60–80.

Miao, B. Q., Wu, Y., Liu, D. (2005). Asymptotic efficiencies of recursive M-estimators of regression coefficients.

Wu, Y. (1996). On consistency of recursive multivariate M-estimators in linear models. In: H. Rieder (Ed.), *Robust statistics, data analysis, and computer intensive methods*. Lecture Notes in Statistics, Vol. 109. Berlin Heidelberg New York: Springer, pp. 411–424