Limiting Behavior of Recursive $M$-Estimators in Multivariate Linear Regression Models

B. Q. Miao*

University of Science and Technology of China, Hefei, Anhui, China

and

Y. Wu†

York University, North York, Ontario, Canada

In this paper, several recursive algorithms for computing $M$-estimates in multivariate linear regression models are discussed. It is shown that the recursive $M$-estimators of regression coefficient and scatter parameters are strongly consistent. In particular, the asymptotic normality of the recursive $M$-estimators of regression coefficients is established.

1. INTRODUCTION

Consider the multivariate linear regression model

$$ y_i = X_i \beta + e_i, \quad i = 1, 2, ..., \quad (1.1) $$

where $X_i$, $i = 1, 2, ...$ are $m \times p$ matrices, $\beta$ is a $p$-vector of unknown regression coefficients, and $e_i$, $i = 1, 2, ...$ are $m \times 1$ random errors. In the literature, there are many papers devoted to the theory of consistency and asymptotic normality for $M$-estimates of $\beta$. References may be made to Huber (1964, 1973, 1981), Bickel (1975), Yohai and Maronna (1979), Maronna and Yohai (1981), Portnoy (1984, 1985), Heller and Willers

Received December 13, 1995; revised April 1996.


Key words and phrases: $M$-estimation, recursive algorithm, robust estimation, regression coefficients, scatter parameters, strong consistency, asymptotic normality.

* Research partially supported by the National Natural Science Foundation of China and the National Sciences and Engineering Research Council of Canada.

† Research supported by the National Sciences and Engineering Research Council of Canada.

0047-259X/96 $18.00$

Copyright © 1996 by Academic Press, Inc.

All rights of reproduction in any form reserved.
(1988), Chen and Wu (1988), Bai, Rao and Wu (1992), Bai and Wu (1994a, b) and others.

Except for the least squares estimates, all other M-estimates have no explicit expressions and usually are not easy to compute. Therefore, there is a great need to develop a recursive algorithm for computing an M-estimate. The first attempt was made in Bickel (1975) by so called “one-step approximation.” Englund et al. (1988) proposed a recursive algorithm for simultaneously computing the location and scale parameters, which was generated to the multivariate location models in Englund (1993). Bai and Wu (1993) extended the algorithm to general multivariate linear models.

In the following, we will use $c$ or $M$ to denote positive numbers which may take different values in different formulas, $I$ to denote the identity matrix, $\|K\|$ to denote the Euclidean norm of $K$, $\|y\|^2_V$ to denote $y^T V^{-1} y$, $\lambda_{\text{min}}(B)$ and $\lambda_{\text{max}}(B)$ to denote the smallest and largest eigenvalues of the square matrix $B$ of product of two positive definite matrices, respectively, $G(i,j)$ denotes the $ij$th element of the matrix $G$, $\chi(A)$ to denote the indicator of $A$.

Following Moranna (1976), a class of M-estimators of the regression coefficient and scatter parameters may be defined as the solutions of the following equations:

$$\begin{align*}
\sum_{i=1}^{n} X_i' V^{-1} (y_i - X_i \beta) \ u_1(\|y_i - X_i \beta\|_V) &= 0, \\
\sum_{i=1}^{n} [(y_i - X_i \beta)(y_i - X_i \beta)' \ u_2(\|y_i - X_i \beta\|_V^2) - V] &= 0,
\end{align*}
$$

(1.2)

where $u_1$ and $u_2$ are suitably chosen functions.

A recursive algorithm for (1.2) can be given as follows:

$$\begin{align*}
\tilde{\beta}_{n+1} &= \tilde{\beta}_n + (n + 1)^{-1} a_n h_1(\tilde{\beta}_n, \tilde{V}_n, X_{n+1}, y_{n+1}), \\
\tilde{V}_{n+1} &= V_n + (n + 1)^{-1} H_2(\tilde{\beta}_n, \tilde{V}_n, X_{n+1}, y_{n+1}),
\end{align*}
$$

(1.3)

where

$$\begin{align*}
h_1(\beta, V, X, e) &= X' V^{-1} (e - X \beta) \ u_1(\|e - X \beta\|_V), \\
H_2(\beta, V, X, e) &= (e - X \beta)(e - X \beta)' \ u_2(\|e - X \beta\|_V^2) - V,
\end{align*}
$$

$\tilde{\beta}_0$ and $V_0 > 0$ are arbitrary, $u_1$ and $u_2$ are suitably chosen functions, $\{a_n\}$ satisfies certain conditions, $\tilde{V}$ is defined as follows:

Let $\lambda_1$ and $\alpha$ be the eigenvalues and orthonormal eigenvectors of $V$ respectively. Properly choose two positive constants $0 < \delta_1 < \delta_2 < \infty$ and define $\tilde{\lambda}_j = (\delta_1 \vee \lambda_j) \wedge \delta_2$.

Then $\tilde{V} = \sum_{j=1}^{m} \tilde{\lambda}_j \alpha \alpha'$.
It is obvious that
\[ 0 < \delta_1 \leq \lambda_{\min}(P) \leq \lambda_{\max}(P) \leq \delta_2 < \infty. \]  
(1.4)

**Remark 1.1.** When \( X_i = I \) for \( i = 1, 2, \ldots \), (1.3) is the same as the recursive algorithm proposed in Englund (1993). For this special case, Englund (1993) showed that the recursive estimators of location and scatter were strongly consistent under some conditions.

Let \( \bar{S}_n = \sum_{i=1}^{n} X_i'X_i \). Bai and Wu (1993) proposed the following recursive algorithm:

\[
\begin{align*}
\hat{\beta}_{n+1} &= \hat{\beta}_n + \frac{S_{n+1}^{-1}}{n+1} a_n \{ \bar{h}(\hat{\beta}_n, \hat{\sigma}_n, X_{n+1}, y_{n+1}) \}, \\
V_{n+1} &= V_n + (n+1)^{-1} H_2(\hat{\beta}_n, \hat{\sigma}_n, X_{n+1}, y_{n+1}),
\end{align*}
\]  
(1.5)

where \( \hat{\beta}_0 \) and \( V_0 > 0 \) are arbitrary, \( u_1, u_2, h_1, H_2 \), and \( \{a_n\} \) are defined as in (1.3). Under some conditions, Bai and Wu showed that the recursive estimators based on the algorithm (1.5) are strongly consistent.

Set
\[
S_n = \sum_{i=1}^{n} X_i' \Omega^{-1} X_i, \quad \bar{X}_n = X_n S_n^{-1/2}, \quad \hat{\beta}_n = S_n^{1/2} \hat{\beta}_n,
\]  
(1.6)

where \( \Omega \) is defined in (1.10). Then Model (1.1) can be written as

\[ y_i = \bar{X}_i \beta = e_i, \quad i = 1, 2, \ldots \]

We propose the following recursive algorithm:

\[
\begin{align*}
\hat{\beta}_{n+1} &= \hat{\beta}_n + a_n \{ \bar{h}(\hat{\beta}_n, \hat{\sigma}_n, \bar{X}_{n+1}, y_{n+1}) \} \\
V_{n+1} &= V_n + (n+1)^{-1} H_2(\hat{\beta}_n, \hat{\sigma}_n, \bar{X}_{n+1}, y_{n+1}),
\end{align*}
\]  
(1.7)

where \( \hat{\beta}_0 \) and \( V_0 > 0 \) are arbitrary, \( u_1, u_2, h_1, H_2 \), and \( \{a_n\} \) are defined as in (1.3). The corresponding recursive algorithm for Model (1.1) is as follows:

\[
\begin{align*}
\hat{\beta}_{n+1} &= D_{n+1}^{-1} \hat{\beta}_n + a_n S_n^{-1/2} \{ \bar{h}(D_{n+1} \hat{\beta}_n, \hat{\sigma}_n, \bar{X}_{n+1}, y_{n+1}) \} \\
V_{n+1} &= V_n + (n+1)^{-1} H_2(D_{n+1} \hat{\beta}_n, \hat{\sigma}_n, \bar{X}_{n+1}, y_{n+1}),
\end{align*}
\]  
(1.8)

where \( D_{n+1} = S_{n+1}^{-1/2} S_n^{1/2} \).

Let \( \mathcal{F}_n \) be the sigma field generated by \( (X_i, e_i), i = 1, \ldots, n \). It is obvious that \( \mathcal{F}_n, i = 1, 2, \ldots, \) are monotone.

In this paper, it will be shown that under the following assumptions (1.3) and (1.8) are strongly consistent and also asymptotically normal, and after little adjustment, the revised (1.5) will have the same properties.
Assumption 1.1. $u_1(t)$ is a nonnegative decreasing BL function such that $tu_1(t)$ is increasing for $t > 0$ (strictly increasing for all small $t$) and $u'_1(t)$ is continuous. $u_2(t)$ is a nonnegative decreasing BL function such that $tu_2(t)$ is increasing for $t > 0$ (strictly increasing for all small $t$) and $tu_2(t) > m$ for some $t > 0$ and $tu_2(t) < M < \infty$ for all $t$. The definition of BL functions is referred to in Bai and Wu (1993).

Assumption 1.2. $a_n$ is $\mathcal{F}_n$-measurable and there exist two constants $0 < v_1 < v_2 < \infty$ such that $v_1 \leq a_i \leq v_2$, $a_n \to a$ a.s., where $a$ is a constant.

Assumption 1.3. The $m \times (p + 1)$ random matrices $(X_i, e_i)$, $i = 1, 2, \ldots$, are independent and identically distributed (iid), and $X_i$ and $e_i$ are independent. $e_i$ has a density $|\theta|^\frac{v_1 - v_2}{2}f(|\theta|^\frac{v_1 - v_2}{2})$, where $f$ is decreasing on $[0, \infty)$ and strictly decreasing in a neighborhood of 0.

Let $\omega > 0$ be defined by the following equation:

$$m = \int \omega \, |z|^2 \, u_1(\omega \, |z|^2) \, f(|z|^2) \, dz$$

(1.9)

and let

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

Assumption 1.4. $EX_i \Omega^{-1}X_i = Q > 0,$ $E \|X_i\|^4 < \infty$.

Without loss of generality, we assume $Q = I$.

In Section 2, the strong consistency of recursive estimators given in (1.3) and (1.8) are proved. In Section 3, some notations are defined and some lemmas are given and proved. In Section 4, the theorems on the asymptotic normality if recursive $M$-estimators are stated and proved.

2. STRONG CONSISTENCY

Define $\zeta = \min\{\zeta^*, 0.1\}$ and $\zeta^*$ is defined by

$$\zeta^* = \frac{1}{4m} \left[ \omega \, |z|^2 \, u_2(\omega \, |z|^2) - \frac{\omega}{1.5} \, |z|^2 \, u_2 \left( \frac{\omega}{1.5} \, |z|^2 \right) \right] f(|z|^2) \, dz.$$

By Assumption 1.1, we know that $\zeta^*$ and hence $\zeta$ are positive constants.
Theorem 2.1. Suppose that the constants in (1.4) satisfy $\delta_1 < \lambda_{\min}(\Omega)$ and $\delta_2 > 3\lambda_{\max}(\Omega)$. Under Assumptions 1.1–1.4, $(\mathbf{p}_n, V_n)$ is a strong consistent estimate of $(\mathbf{p}, \Omega)$.

Proof. The difference between our algorithm and the algorithm (1.3) is that $D_{n+1}\mathbf{p}_n$ replaces $\mathbf{p}_n$. Using the technique as in the proof of Theorem 3.1 of Bai and Wu (1993), we can get

$$\mathbf{p}_{n(k+1)} = (I - G_k) \mathbf{p}_{n(k)} + \tilde{d}_k \tilde{h}_1(\mathbf{p}_{n(k)}, V_{n(k)}) + o(d_k),$$

where $G_k = I - S_{n(k+1)}^{-1/2} S_{n(k)}^{1/2}$ and the other notations are defined as in Bai and Wu (1993). Checking the proof of Lemma 2.2 of Bai and Wu (1993), we can similarly prove that $\mathbf{p}_k \rightarrow \mathbf{p}$, where $\mathbf{p}_k$ is recursively defined by

$$\mathbf{p}_{k+1} = (I - G_k) \mathbf{p}_k + d_k \tilde{h}_1(\mathbf{p}_k, V_k) + r_k,$$

$G_k$ is a $p \times p$ matrix, $q$ and $r_k$ are the same as in the Lemma 2.2 of Bai and Wu (1993), and $0 \leq \lambda_{\min}(G_k) \leq \lambda_{\max}(G_k) \rightarrow 0$. By the strong law of large numbers, the eigenvalues of $D_{n+1}$ locate in $[1 - \lambda_n, 1]$, a.s. where $\lambda_n \rightarrow 0$ and $\lambda_n \rightarrow 0$. Since $n(k+1)/n(k) \rightarrow 1$, the eigenvalues of $S_{n(k+1)}^{-1/2} S_{n(k)}^{1/2}$ locate in $[1 - \lambda_{n(k)}, 1]$, a.s. Therefore, $\mathbf{p}_n \rightarrow \mathbf{p}$, a.s. The proof of $V_n \rightarrow \Omega$ follows from Bai and Wu (1993) because it will not be affected when $X_{n+1} D_{n+1} \mathbf{p}_n$ is replaced by $X_{n+1} D_{n+1} \mathbf{p}_n$.

Remark 2.1. In the same way, it can be shown that the estimates computed by the recursive algorithm (1.3) or the recursive algorithm (1.5) with $\hat{S}_n$ is defined by (4.12) are strongly consistent.

Remark 2.2. For the strong consistency, assumptions made in the theorem can be weakened. See Bai and Wu (1993).

3. NOTATIONS AND LEMMAS

For convenience, we define the following notations:

$$\mathcal{G} = \{ V \in \mathbb{R}^{m \times m} : V' = V, 0 < \delta_1 \leq \lambda_{\min}(V) \leq \lambda_{\max}(V) \leq \delta_2 < \infty \},$$

$$A(V, X, \mathbf{e}) = X'(u_1(\|\mathbf{e}\|_V) V^{-1} + u'_1(\|\mathbf{e}\|_V) \|\mathbf{e}\|_V^{-1} V^{-1} \mathbf{e} \mathbf{e}' V^{-1}) X,$$

$$B_1(V) = E[u_1(\|\mathbf{e}\|_V) I + u'_1(\|\mathbf{e}\|_V) \|\mathbf{e}\|_V^{-1} V^{-1/2} \mathbf{e} \mathbf{e}' V^{-1/2}],$$

$$B_2(V) = E[u'_1(\|\mathbf{e}\|_V) \|\mathbf{e}\|_V^{-1/2} \mathbf{e} \mathbf{e}' V^{-1/2}],$$

$$b_1(\Omega) = E[u_1(\|\mathbf{e}\|_\Omega) + \frac{1}{m} \|\mathbf{e}\|_\Omega u'_1(\|\mathbf{e}\|_\Omega)],$$

$$b_2(\Omega) = \frac{1}{m} E[u'_1(\|\mathbf{e}\|_\Omega) \|\mathbf{e}\|_\Omega].$$

When the distribution of \( e \) is elliptically symmetric, we have

\[
B_1(\Omega) = b_1(\Omega) I,
\]

\[
B_2(\Omega) = b_2(\Omega) I,
\]

and \( B_i(V), i = 1, 2, \) are diagonal matrices.

The following lemmas are needed in the proofs of main results in Section 4.

**Lemma 3.1.** (i) If \( E(\|X_n\|^2) \leq M < \infty \) for some positive integer \( k \), then

\[
\sup_n n^k E[\|\tilde{X}_{n+1}\|^{2k} | \mathcal{F}_n] \leq c < \infty, \quad \text{a.s.}
\]

(ii) \( (n+1) E(\tilde{X}_{n+1}' \Omega^{-1} \tilde{X}_{n+1} | \mathcal{F}_n) \rightarrow I, \text{ a.s.} \)

(iii) \( (n+1) E[\tilde{X}_{n+1}' \tilde{P}_{n}^{-1/2} \tilde{B}_1(\tilde{P}_n) \tilde{P}_{n}^{-1/2} \tilde{X}_{n+1} | \mathcal{F}_n] \rightarrow b_i(\Omega) I \) a.s. \( i = 1, 2. \)

**Proof.** Noticing that \( \|\tilde{X}_{n+1}\|^{2k} = \|X_{n+1} S_{n+1}^{-1/2}\|^{2k} \leq \|S_{n}^{-1/2}\|^{2k} \|X_{n+1}\|^{2k}, \)

we have

\[
E(\|\tilde{X}_{n+1}\|^{2k} | \mathcal{F}_n) \leq \|S_{n}^{-1/2}\|^{2k} E \|X_{n+1}\|^{2k} \leq M \|S_{n}^{-1/2}\|^{2k} = M[\text{tr}(S_{n}^{-1})]^k.
\]

By the strong law of large numbers,

\[
\frac{S_n}{n} \rightarrow EX_1' \Omega^{-1} X_1 = I \quad \text{a.s.}
\]

Therefore,

\[
n^k E[\|\tilde{X}_{n+1}\|^{2k} | \mathcal{F}_n] \leq M[\text{tr}(S_{n}^{-1})]^k n^k \leq c \quad \text{a.s.}
\]

(ii) Since \( S_n \leq S_{n+1}, \)

\[
(n+1) E(\tilde{X}_{n+1}' \Omega^{-1} \tilde{X}_{n+1} | \mathcal{F}_n) \leq (n+1) E(S_{n+1}^{-1/2} X_{n+1}' \Omega^{-1} X_{n+1} S_{n+1}^{-1/2} | \mathcal{F}_n)
\]

\[
= \left( \frac{S_n}{n+1} \right)^{-1}.
\]

On the other hand, by the fact that \( X_n, i = 1, 2, ..., \) are iid and \( E(\|X_{n+1}\|^2) < \infty, \) we have that \( X_{n+1}/\sqrt{n+1} \rightarrow 0, \text{ a.s.} \) Since \( S_n/(n+1) \rightarrow I, \text{ a.s.} \), it follows that for any \( \varepsilon > 0, \)

\[
\frac{\varepsilon S_{n}}{n+1} > \frac{X_{n+1}' \Omega^{-1} X_{n+1}}{n+1} \quad \text{a.s.}
\]
\[
\left( \frac{S_{n+1}}{n+1} \right)^{-1/2} > (1 + \varepsilon)^{-1/2} \left( \frac{S_n}{n+1} \right)^{-1/2} \quad \text{a.s.}
\]

Therefore,
\[
(n + 1) \, \mathbb{E} \left( \hat{X}_{n+1} \, X^{-1}_{n+1} | \mathcal{F}_n \right) \\
\geq (1 + \varepsilon)^{-1} \, (n + 1) \, \mathbb{E} \left( S_{n+1}^{-1/2} \, X^{-1}_{n+1} X_n \, S_n^{-1/2} | \mathcal{F}_n \right) \\
= (1 + \varepsilon)^{-1} \left( \frac{S_n}{n+1} \right)^{-1} \quad \text{a.s.}
\]

i.e.
\[
(1 + \varepsilon)^{-1} \left( \frac{S_n}{n+1} \right)^{-1} \leq (n + 1) \, \mathbb{E} \left( \hat{X}_{n+1} \, X^{-1}_{n+1} | \mathcal{F}_n \right) \leq \left( \frac{S_n}{n+1} \right)^{-1} \quad \text{a.s.}
\]

Let \( n \to \infty \) and then \( \varepsilon \to 0 \), we get (ii)

(iii) Since \( \hat{P}_n \to \Omega \), a.s. and \( B_i(V), \quad i = 1, 2 \), are continuous functions of \( V \), by Egorov Theorem, for any \( \varepsilon > 0 \), there exists a set \( A \) such that \( P(A') < \varepsilon \), uniformly in \( A \)

\[
\| \hat{P}_n^{-1} - \Omega^{-1} \| < \varepsilon, \quad \| B_i(\hat{P}_n) - b_i(\Omega) I \| < \varepsilon, \quad i = 1, 2
\]

for \( n \geq n_0 \), where \( n_0 = n_0(\varepsilon) \). By the definition of \( \hat{P} \) and the continuity of \( u_1 \) and \( u_1^* \), there exists a constant \( M \) such that

\[
b_i(\Omega) \leq M, \quad \| \hat{P}_n^{-1} - \Omega^{-1} \| \leq M, \quad \| B_i(\hat{P}_n) - b_i(\Omega) I \| \leq M, \quad i = 1, 2.
\]

Let \( B \) be the set of the sample space such that \( n^k \mathbb{E}[\| \hat{X}_{n+1} \|^{2k} | \mathcal{F}_n] \leq c < \infty \). By (i), \( P(B) = 1 \). Therefore, by \( \delta^{-1} I \leq \hat{P}_n^{-1} \leq \delta^{-1} I \)

\[
\| (n + 1) \, \mathbb{E} \left[ \hat{X}_{n+1}^{-1/2} B_i(\hat{P}_n) \, \hat{P}_n^{-1/2} \hat{X}_{n+1}^{-1} | \mathcal{F}_n \right] \\
- b_i(\Omega) \, \mathbb{E} \left[ \hat{X}_{n+1}^{-1} \, X^{-1} \hat{X}_{n+1}^{-1} | \mathcal{F}_n \right] \| \\
\leq \| (n + 1) \, \mathbb{E} \left[ \hat{X}_{n+1}^{-1/2} B_i(\hat{P}_n) - b_i(\Omega) I \right] \, \hat{P}_n^{-1/2} \hat{X}_{n+1}^{-1} | \mathcal{F}_n \right] \| \\
+ \| b_i(\Omega) (n + 1) \, \mathbb{E} \left[ \hat{X}_{n+1}^{-1} \, X^{-1} \hat{X}_{n+1}^{-1} | \mathcal{F}_n \right] \|
\]
\[
\leq c \delta^{-1} \left\| \left( n+1 \right) \mathbb{E}[\widehat{X}_{n+1} | \mathcal{F}_n] \right\|
+ \delta_2^{-1} M \left\| \left( n+1 \right) \mathbb{E}[\widehat{X}_{n+1} | \mathcal{F}_n] \right\|
+ M \delta(n+1) \mathbb{E}[\| \widehat{X}_{n+1} \|^2 | \mathcal{F}_n] + \delta_2^{-1} \delta M + \delta e c M + M \delta_2^{-1} c e
\]

in \( B \), which and (ii) imply that (iii) is true.

**Lemma 3.2.** Let \( \eta \) and \( V \) be \( \mathcal{F}_n \)-measurable, \( V \in \mathscr{S} \), \( \widehat{X}_{n+1} \) and \( \eta_{n+1} \) are the same as in Theorem 2.1. Define

\[
\phi(\eta, V, \widehat{X}_{n+1}, \eta_{n+1}) = h(\eta, \widehat{X}_{n+1}, \eta_{n+1}) - h(0, V, \widehat{X}_{n+1}, \eta_{n+1})
+ A(\eta, \widehat{X}_{n+1}, \eta_{n+1}) \eta.
\]

Then

\[
n[\phi(\eta, V, \widehat{X}_{n+1}, \eta_{n+1}) = o(n \| \widehat{X}_{n+1} \|^2) \eta \quad a.s.\]

as \( n \to \infty \).

**Proof.** Since \( u_1(t) \) is BL function, \( y u_1(\| y \|) \) is BLC(\( y \)) (see Bai and Wu (1993)). Therefore, there exists \( M > 0 \) such that

\[
\| (e - a) u_1(\| e - a \|) - eu_1(\| e \|) \| \leq M \| a \|.
\]

Since

\[
\frac{\partial}{\partial e}(e u_1(\| e \|)) = u_1(\| e \|) I + \| e \|^{-1} u_1(\| e \|) ee',
\]

and \( u_1(t) \) is continuous,

\[
\| (e - a) u_1(\| e - a \|) - eu_1(\| e \|) \|
+ \| u_1(\| e \|) I + \| e \|^{-1} u_1(\| e \|) ee' \| a \|
\]

\[= o(\| a \|). \quad (3.4)\]

By the fact that \( \widehat{X}_{n+1} \to 0 \), a.s., for fixed \( \eta \), substituting \( e \) and \( a \) of (3.4) by \( V^{-1/2} e_{n+1} \) and \( V^{-1/2} \eta_{n+1} \), we have

\[
|n \phi(\eta, V, \widehat{X}_{n+1}, \eta_{n+1})| = o(n \| \widehat{X}_{n+1} \|^2) \| \eta \| \quad a.s. \quad n \to \infty.
\]
Therefore, 

\[ n\phi(\eta, V, \bar{X}_{n+1}, e_{n+1}) = o(n \|\bar{X}_{n+1}\|^2) \eta, \]

for every fixed \( \eta \).

**Lemma 3.3.** Let \( \tau > 0, 0 < \alpha \leq 1, \varepsilon_n \leq \varepsilon, 0 \leq r_n \leq r \) and \( b_n \geq c \). If 

\[ b_{n+1} \leq (1 - \tau n^{-\alpha}) b_n + \varepsilon_n n^{-\alpha}, \]  

(3.5) 

or 

\[ b_{n+1} \leq (1 - \tau n^{-\alpha}) b_n + \varepsilon_n n^{-\alpha} + n^{-3\alpha/2} \sqrt{|b_n| r_n}, \]  

(3.6) 

then 

\[ \lim_{n \to \infty} b_n = b \leq \frac{\varepsilon}{\tau}. \]

**Proof.** Without loss of generality, we assume that \( \varepsilon_n \leq 0 \) in (3.5). Otherwise, let \( l_n = b_n - \varepsilon / \tau \) and \( \eta_n = \varepsilon_n - \varepsilon \). Then \( l_n, n = 1, 2, \ldots, \) satisfy (3.5). By (3.5), we have

\[ b_{n+1} \leq \prod_{i=n_0}^{n} (1 - \tau i^{-\alpha}) b_m. \]

Since \( \sum_{i=n_0}^{n} i^{-\alpha} = \infty \), for large \( n_0 \),

\[ \left| \prod_{i=n_0}^{n} (1 - \tau i^{-\alpha}) b_m \right| < 1, \]

which implies \( b_{n+1} \leq 1 \). Therefore, \( \{b_n\} \) is bounded. Denote

\[ \lim \inf b_n = b_{(0)} = \lim \sup b_n. \]

It follows that there exists a subsequence \( \{n_1\} \) of \( \{n\} \) such that \( \tau n_1^{-\alpha} < 1, b_{n_1} > (b_{(1)} + b_{(0)})/2, \) and \( b_{n_1-1} \leq (b_{(1)} + b_{(0)})/2. \) By (3.5),

\[ b_{n_1} \leq (1 - \tau n_1^{-\alpha}) b_{n_1-1} \leq (1 - \tau n_1^{-\alpha})(b_{(1)} + b_{(0)})/2 < (b_{(1)} + b_{(0)})/2, \]

which is contradictory to \( b_{n_1} > (b_{(1)} + b_{(0)})/2. \) Therefore, \( \lim_{n \to \infty} b_n = b \), which implies that

\[ \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} j^{-\alpha} b_j = \lim_{n \to \infty} b_n = b, \quad 0 < \alpha < 1, \]
and

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} b_j = b. \]

By (3.5), it is easy to see that

\[
\begin{cases}
  n^{-(1-s)}(b_{n+1} - b_1) + \tau n^{-(1-s)} \sum_{j=1}^{n} j^{-\gamma} b_j \leq 0, & 0 < \alpha < 1, \\
  b_{n+1} + (\tau - 1) \sum_{j=1}^{n} b_j/n \leq 0, & \alpha = 1.
\end{cases}
\]

(3.7)

Taking limits in (3.7), we have \( \tau b \leq 0 \), which implies that \( b \leq 0 \).

When \( \{b_n\} \) satisfies (3.6), we assume that \( b_n \geq 0 \) without loss of generality. Let \( \beta < 1 \) be any positive number. If \( b_n \geq (r/\beta \tau)^{3} n^{-3} \), then

\[ b_{n+1} \leq (1 - \tau n^{-3}) b_n + \varepsilon_n n^{-3} + \beta \tau n^{-3} b_n r_n / r \leq (1 - \tau (1 - \beta)) n^{-3} b_n + \varepsilon_n n^{-3}. \]

If \( b_n < (r/\beta \tau)^{3} n^{-3} \), then

\[ b_{n+1} \leq (1 - \tau n^{-3}) b_n + (\varepsilon_n + r^3 (\beta \tau)^{-1} n^{-3}) n^{-3}. \]

Let \( \xi_n = \varepsilon_n + r^3 (\beta \tau)^{-1} n^{-3} \). Then for any \( \eta > 0 \), \( \xi_n \leq \varepsilon + \eta \) for large \( n \). Therefore, in both situations, we have

\[ b_{n+1} \leq (1 - \tau (1 - \beta)) n^{-3} b_n + \xi_n n^{-3}. \]

By the first part of the lemma, we get that

\[ \lim_{n \to \infty} b_n = b \leq \frac{\varepsilon + \eta}{\tau (1 - \beta)}. \]

(3.8)

Let \( \eta \to 0 \), and then \( \beta \to 0 \) in (3.8) and the second part of the lemmas follows.

**Corollary 3.1.** If \( \tau \varepsilon = \varepsilon \), then \( \lim_{n \to \infty} b_n = \varepsilon / \tau \). Specially, when \( \varepsilon = \varepsilon = 0 \), \( \lim_{n \to \infty} b_n = 0 \).

**Corollary 3.2.** If \( \varepsilon_n \to \varepsilon \) and the equality in (3.5) or (3.6) holds, then \( \lim_{n \to \infty} b_n = \varepsilon / \tau \).
Proof. By Lemma 3.3, for any \( \eta > 0 \), \( \lim_{n \to \infty} b_n \leq (\varepsilon + \eta)/\tau \) when \( n \) is large enough. Therefore, for large \( n \),
\[
b_n \leq \frac{\varepsilon + \eta}{\tau}, \quad -b_n \geq -\frac{\varepsilon + \eta}{\tau}, \quad -\varepsilon \leq -\frac{\varepsilon - \eta}{\tau}.
\]
By Lemma 3.3, it follows that
\[
\lim_{n \to \infty} (-b_n) \leq -\frac{\varepsilon - \eta}{\tau}.
\]
(3.9)
The corollary follows by letting \( \eta \to 0 \) in (3.9).

Corollary 3.3. If \( b_n \geq 0 \), \( e_n \to 0 \), \( r_n \to 0 \) and
\[
b_{n+1} \leq (1 - \tau n^{-\alpha}) b_n + r_n n^{-\alpha} |b_n|^{1/2} + e_n n^{-\alpha},
\]
then \( \lim_{n \to \infty} b_n = 0 \).

Lemma 3.4. Suppose that \( u_n \) and \( \tilde{u}_n \) are two sequences of \( p \times 1 \) random vectors given as follows:
\[
u_{n+1} = (I - n^{-\alpha} \Psi) u_n + n^{-\alpha} v_n, \quad (3.10)
\]
\[
u_{n+1} = (I - n^{-\alpha} (\Psi + \Gamma_n)) \tilde{u}_n + n^{-\alpha} v_n, \quad (3.11)
\]
where \( \tilde{u}_1 = u_1 = c \), \( c \) is a constant vector, \( 0 < \alpha \leq 1 \), \( \Psi \in \mathbb{R}^{p \times p} \) and \( \Psi > 0 \), and \( p \times 1 \) random vectors \( v_n, n = 1, 2, ... \), and \( p \times p \) matrices \( \Gamma_n, n = 1, 2, ... \), satisfy the following conditions:
\( v_n \) is \( \mathbb{F}_{n+1} \)-measurable, \( E(v_n | \mathbb{F}_n) = 0 \), \( E(\|v_n\|^2 | \mathbb{F}_n) \leq c < \infty \),
and
\( \Gamma_n \) is \( \mathbb{F}_{n+1} \)-measurable, \( \|E(\Gamma_n | \mathbb{F}_n)\| \leq r_n \), \( E(\|\Gamma_n\|^2 | \mathbb{F}_n) \leq c < \infty \),
\( \{r_n\} \) is a sequence of constants such that \( \lim_{n \to \infty} r_n = 0 \). If \( \{u_n\} \) has an asymptotic distribution \( F \), so does \( \{\tilde{u}_n\} \).

Proof. Let \( \delta_n = u_n - \tilde{u}_n \). The difference of (3.10) and (3.11) is
\[
\delta_{n+1} = (I - n^{-\alpha} \Psi) \delta_n + n^{-\alpha} \Gamma_n \tilde{u}_n. \quad (3.12)
\]
Firstly, we prove that $E(\|\hat{u}_n\|^2) \leq M$ for some constant $M > 0$. Let $\hat{\lambda}_1 = \hat{\lambda}_{\text{min}}(\mathcal{Y})$. It follows from (3.11) that

$$E(\|\hat{u}_{n+1}\|^2) = E(\| (I - n^{-\gamma}(\mathcal{Y} + \Gamma_n)) \hat{u}_n \|^2)$$

$$+ n^{-x^2}E[\hat{u}_n(I - n^{-\gamma}(\mathcal{Y} + \Gamma_n)) \nu_n]$$

$$+ n^{-x^2}E[\nu_n(I - n^{-\gamma}(\mathcal{Y} + \Gamma_n)) \hat{u}_n] + n^{-x}E(\| \nu_n \|^2). \quad (3.13)$$

Noticing that $\|E(\Gamma_n | \mathcal{F}_n)\| \leq r_\gamma$, we get

$$|E(\hat{u}_n, \Gamma_n \hat{u}_n)| = |E[\hat{u}_n, E(\Gamma_n | \mathcal{F}_n) \hat{u}_n]| \leq r_\gamma E(\| \hat{u}_n \|^2).$$

and

$$E(\nu_n, \Gamma_n \nu_n) \leq E(\| \Gamma_n \|^2 \| \nu_n \|^2) \leq c E(\| \nu_n \|^2),$$

which imply that

$$E(\| (I - n^{-\gamma}(\mathcal{Y} + \Gamma_n)) \hat{u}_n \|^2) \leq (1 - n^{-\gamma}(\hat{\lambda}_1 - 2r_n)) E(\| \hat{u}_n \|^2)$$

$$\leq (1 - n^{-\gamma}\hat{\lambda}_1/2) E(\| \hat{u}_n \|^2) \quad (3.14)$$

for large $n$. The fact that $E(\nu_n | \mathcal{F}_n) = 0$ gives

$$n^{-x^2} |E[\hat{u}_n(I - n^{-\gamma}(\mathcal{Y} + \Gamma_n)) \nu_n]| = n^{-x^2} |E[\hat{u}_n \nu_n]|$$

$$\leq cn^{-3x^2}(E(\| \hat{u}_n \|^2))^{1/2} (E(\| \nu_n \|^2))^{1/2}$$

$$\leq cn^{-3x^2}(E(\| \nu_n \|^2))^{1/2}. \quad (3.15)$$

Combining (3.14) and (3.15), we get

$$E(\|\hat{u}_{n+1}\|^2) \leq (1 - n^{-\gamma}\hat{\lambda}_1/2) E(\| \hat{u}_n \|^2) + E(\| \nu_n \|^2) n^{-x}$$

$$+ cn^{-3x^2}(E(\| \hat{u}_n \|^2))^{1/2}. \quad (3.16)$$

Applying Lemma 3.3 to (3.16), we have

$$E(\|\hat{u}_{n+1}\|^2) \leq c < \infty. \quad (3.17)$$

By taking the Euclidean norms on both sides of (3.12), it follows that

$$E(\|\hat{\delta}_{n+1}\|^2) = E(\| (I - n^{-\gamma}\mathcal{Y}) \hat{\delta}_n \|^2) + n^{-x}E(\hat{\delta}_n(I - n^{-\gamma}\mathcal{Y}) \Gamma_n \hat{u}_n)$$

$$+ n^{-x}E(\hat{\delta}_n(I - n^{-\gamma}\mathcal{Y}) \hat{\delta}_n) + n^{-2x}E(\| \Gamma_n \hat{u}_n \|^2).$$
Since
\[ |E(\delta(I-n^{-\Psi}) \Gamma \tilde{u}_n)| = |E[\delta(I-n^{-\Psi}) E(\Gamma | \mathcal{F}_n) \tilde{u}_n]| \]
\[ \leq r_d(E(\|\delta\|^2))^{1/2} (E(\|\tilde{u}_n\|^2))^{1/2} \]
\[ \leq cr_d(E(\|\delta\|^2))^{1/2}, \]
and
\[ E(\|\Gamma \tilde{u}_n\|^2) \leq E(\|\Gamma_n\|^2) E(\|\tilde{u}_n\|^2) \leq cE(\|\tilde{u}_n\|^2) \leq c, \]
it follows that
\[ E(\|\delta_{n+1}\|^2) \leq (1-\lambda_n n^{-\Psi}) E(\|\delta_n\|^2) + cr_d(E(\|\delta_n\|^2))^{1/2} n^{-\Psi} + cn^{-2\Psi}. \] (3.18)
Applying Corollary 3.3 of Lemma 3.3 to (3.18), we have
\[ E(\|\delta_{n+1}\|^2) \to 0, \]
which implies that $\delta_{n+1} \to 0$ in probability.

**Lemma 3.5** (Fabian, 1968). Suppose that $p \times 1$ random vectors $u_n$, $n=1, 2, ..., \text{have recursive expression}$
\[ u_{n+1} = (I-n^{-\Psi}) u_n + n^{-\Psi} v_n, \] (3.19)
where $u_1 = c$, $c$ is $p \times 1$ constant vector, $0 < \alpha \leq 1$, $\Psi$ is a $p \times p$ positive definite matrix, and $p \times 1$ random vectors $v_n$, $n=1, 2, ...$, satisfy the following conditions:

1. $v_n$ is $\mathcal{F}_{n+1}$-measurable and $E(v_n | \mathcal{F}_n) = 0$;
2. There exist a constant $c$ and a positive definite matrix $\Sigma$ such that $c > E(v_n^2 | \mathcal{F}_n) - \Sigma \to 0$;
3. $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 = 0$ for every $r > 0$, where $\sigma_i^2 = E(\|v_i\|^2)$.

Then the asymptotic distribution of $u_n$ is normal with mean vector zero and covariance matrix $Y$, where
\[ Y(i, j) = (P^* \Sigma P)(i, j)(\Psi(i, j) + \Psi(j, j))^{-1} \]
and $P$ is orthogonal such that $P^* \Psi P$ is a diagonal matrix.
4. ASYMPTOTIC NORMALITY

Theorem 4.1. Under Assumptions 1.1–1.4,

\[
\tilde{\beta}_n - \hat{\beta} \xrightarrow{\mathcal{D}} N\left(0, \frac{ab_2(\Omega)}{\mathcal{D} b_1(\Omega)} I\right),
\]

where \( b_1(\Omega) \) and \( b_2(\Omega) \) are defined in (3.2) and (3.3).

Proof. Without loss of generality, assume that \( \beta = 0 \). Rewrite the recursive algorithm (1.7) as follows:

\[
\tilde{\beta}_{n+1} = \tilde{\beta}_n - n^{-1} \left[ E[n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) | \mathcal{F}_n] \tilde{\beta}_n + \frac{1}{n} A_n \tilde{\beta}_n - n^{-1} \Psi_n \tilde{\beta}_n \right] + n^{-1/2} \sigma_n \tilde{h}_1(0, \tilde{P}_n, \tilde{X}_{n+1}, e_{n+1})].
\]

It is easy to see that \( n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) \) can be written in the form of \( \Gamma_n \tilde{\beta}_n \) by Lemma 3.2. Therefore, (4.1) can be rewritten as

\[
\tilde{\beta}_{n+1} = (I - n^{-1} \Gamma_n \tilde{\beta}_n) \tilde{\beta}_n + n^{-1} \Gamma_n \tilde{\beta}_n - n^{-1} \Psi_n \tilde{\beta}_n + n^{-1/2} \sigma_n \tilde{h}_1(0, \tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}).
\]

where

\[
A_n = ab_1(\Omega) I - E[n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) | \mathcal{F}_n], \quad \Psi_n = n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) - E[n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) | \mathcal{F}_n],
\]

\[
\Gamma_n \tilde{\beta}_n = n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}), \quad \sigma_n = n^{-1/2} \sigma_n h_1(0, \tilde{P}_n, \tilde{X}_{n+1}, e_{n+1})
\]

Since \( e_{n+1} \) is independent of \( \{X_1, ..., X_{n+1}\} \) and \( a_n \to a \), by Lemma 3.1, we get

\[
E[n_a \phi(\tilde{P}_n, \tilde{X}_{n+1}, e_{n+1}) | \mathcal{F}_n] = n_a E[\tilde{X}_{n+1} \tilde{P}_n^{-1/2} B_n(\tilde{P}_n) \tilde{X}_{n+1}^{-1/2} \tilde{X}_{n+1} | \mathcal{F}_n] = ab_1(\Omega) I + o(1) \quad \text{a.s.}
\]

Therefore,

\[
A_n = o(1), \quad \|A_n\|^2 = o(1) \quad \text{a.s.}
\]
Note that
\[ 0 \leqslant u_1(\|e\|_V) V^{-1} + u_1'(\|e\|_V) \|e\|_V^{-1} V^{-1} ee' V^{-1} \leqslant M \|V^{-1}\|, \]
where \(M\) is the upbound of \(u_1(t)\). By Lemma 3.1, we have that
\[ E(\|\Psi_n\|^2 | \mathcal{F}_n) \leqslant 2 E(\|n\alpha A(\bar{p}_n, \bar{x}_{n+1}, e_{n+1})\|^2 | \mathcal{F}_n) \]
\[ \leqslant 2 M^2 \delta^2 n^2 \alpha^2 (E(\|\bar{x}_{n+1}\|^4 | \mathcal{F}_n)) \leqslant c \text{ a.s.} \quad (4.5) \]

In view of Lemma 3.2, it follows that
\[ \alpha_n, \phi(\bar{p}_n, \bar{x}_n, \bar{x}_{n+1}, e_{n+1}) = o(n \|\bar{x}_{n+1}\|^2) \tilde{b}_n, \]
which and Lemma 3.1 imply that
\[ \|E(F_n | \mathcal{F}_n)\| = o(1) \text{ a.s.} \quad E(\|F_n\|^2 | \mathcal{F}_n) = o(1) \text{ a.s.} \quad (4.6) \]

Noting that the distribution of \(e_{n+1}\) is elliptically symmetric and that \(e_{n+1}\) is independent of \(\{X_1, ..., X_{n+1}\}\), we get
\[ E(v_n v_n' | \mathcal{F}_n) = n \alpha_n^2 E[h_1(0, \bar{p}_n, \bar{x}_{n+1}, e_{n+1}) h_1(0, \bar{p}_n, \bar{x}_{n+1}, e_{n+1})' | \mathcal{F}_n] \]
\[ = n \alpha_n^2 E(\bar{x}_{n+1} \bar{p}_n^{-1} B_2(\bar{p}_n) \bar{p}_n^{-1/2} \bar{x}_{n+1} | \mathcal{F}_n). \]

By Lemma 3.1, it follows that
\[ E(v_n v_n' | \mathcal{F}_n) \to a^2 b_2(\Omega) I \text{ a.s.} \quad (4.7) \]

In view of (4.4)-(4.7), \(-A_n + \Psi_n - F_n\) and \(v_n\) in (4.2) satisfy the corresponding conditions assumed in Lemma 3.4 almost surely. If we define
\[ \eta_{n+1} = (1 - n^{-1} a b_1(\Omega)) \eta_n + n^{-1/2} v_n, \quad (4.8) \]
by Lemma 3.4, the asymptotic distribution of \(\tilde{\eta}_n\) is the same as the asymptotic distribution of \(\eta_n\). It is easy to verify that \(v_n\) satisfies the conditions of Lemma 3.5. By Lemma 3.4 and Lemma 3.5, the asymptotic distribution of \(\tilde{\eta}_n\) is normal, i.e.
\[ \tilde{\eta}_n \overset{d}{\to} N \left( 0, \frac{a b_2(\Omega)}{2 b_1(\Omega)} I \right). \]
Remark 4.1. If we replace the $X_{n+1}$ and $\hat{\beta}_n$ in (1.7) by $\tilde{X}_{n+1} = n^{-1/2}X_{n+1}$ and $\tilde{\beta}_n = n^{1/2} \hat{\beta}_n$, respectively, we have the following recursive algorithm for Model (1.1)

$$
\begin{align*}
\beta_{n+1} &= \left(1 - \frac{1}{n+1}\right)^{1/2} \beta_n + a_n(n(n+1))^{-1/2} h_1(\beta_n, V_n, X_{n+1}, y_{n+1}), \\
V_{n+1} &= V_n + (n+1)^{-1} H_2(\beta_n, V_n, X_{n+1}, y_{n+1}),
\end{align*}
$$

(4.9)

where $u_1$, $u_2$, $h_1$, $H_2$, $\beta_n$, $V_0 > 0$, $\{a_n\}$ and $\|y\|^2_T$ are defined as in (1.3). Carefully checking the proofs of Lemma 3.1 and Lemma 3.2, we see that the lemmas still hold if we replace $X_{n+1} S_{n+1}^{-1/2}$ by $n^{-1/2}X_{n+1}$. Therefore, following the same lines as in the proof of Theorem 4.1, we get

$$
\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathscr{D}} N(0, \frac{ab_2(\Omega)}{2b_1(\Omega)} I).
$$

It is obvious that (4.9) is equivalent to the following recursive algorithm

$$
\begin{align*}
\beta_{n+1} &= \left(1 - \frac{1}{2(n+1)}\right)^{1/2} \beta_n + a_n(n(n+1))^{-1} h_1(\beta_n, V_n, X_{n+1}, y_{n+1}), \\
V_{n+1} &= V_n + (n+1)^{-1} H_2(\beta_n, V_n, X_{n+1}, y_{n+1}),
\end{align*}
$$

(4.10)

Compared with (1.3), we only replace the coefficient of $\beta_n$, which is 1, by $1 - (1/2(n+1))$.

In the following, we discuss the asymptotic distributions of the recursive algorithms (1.3) and (1.5).

**Theorem 4.2.** Consider the recursive algorithm (1.3). Assume that the assumptions 1.1–1.4 hold. If $a_n \to a$ and $2ab_1(\Omega) > 1$, then

$$
\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathscr{D}} N(0, \frac{a^2b_2(\Omega)}{2ab_1(\Omega)} I).
$$

**Proof.** Without loss of generality, assume that $\beta = 0$. The recursive $M$-estimator in (1.3) is computed by

$$
(n+1)^{1/2} \hat{\beta}_{n+1} = \left(\frac{n+1}{n}\right)^{1/2} n^{1/2} \beta_n + a_n(n+1)^{-1/2} \times h_1(\beta_n, V_n, X_{n+1}, e_{n+1}).
$$

(4.11)
Let $\tilde{\beta}_n = n^{1/2} \tilde{\beta}$ and $\tilde{X}_{n+1} = n^{-1/2} X_{n+1}$. Then (4.11) can be written as

$$
\begin{align*}
\tilde{\beta}_{n+1} &= \left(1 + \frac{1}{n}\right)^{1/2} \tilde{\beta}_n + a_n \left( \frac{n}{n+1} \right)^{1/2} h_1(\tilde{\beta}_n, \tilde{V}_n, \tilde{X}_{n+1}, y_{n+1}) \\
&= \left(1 + \frac{1}{2n} + o\left(\frac{1}{n}\right)\right) \tilde{\beta}_n + b_n h_1(\tilde{\beta}_n, \tilde{V}_n, \tilde{X}_{n+1}, e_{n+1}),
\end{align*}
$$

where $b_n = a_n n/(n+1))^{1/2}$. It is easy to see that Lemma 3.1 and Lemma 3.2 still hold in this case. Therefore, we have

$$
\tilde{\beta}_{n+1} = \left[ I - n^{-1} \left[ (ab_1(\Omega) - 0.5) I + o(1) + \Psi_n - \Gamma_n \right] \right] \tilde{\beta}_n + n^{-1/2} b_n v_n
$$

where

$$
v_n = n^{1/2} b_n^{1/2} \tilde{X}_{n+1} \tilde{V}_n^{-1} e_{n+1} u_r(\|e_{n+1} \| r_n).
$$

Since $\Psi_n$ and $\Gamma_n$ satisfy the conditions of Lemma 3.4 and $V_n$ satisfies the conditions of Lemma 3.5, the proof follows from Lemma 3.5.

**Remark 4.2.** In Theorem 4.1, our $a$ is any positive number, but in Theorem 4.2, $a$ needs to satisfy an extra condition.

For the recursive algorithm (1.5), we make some changes in $\tilde{S}_n$, and we let

$$
\tilde{S}_n = \sum_{i=1}^n X_i \Omega^{-1} X_i = S_n.
$$

Then we have the following theorem for this revised recursive algorithm.

**THEOREM 4.3.** Consider the recursive algorithm (1.5), where $\tilde{S}_n$ is defined by (4.12). Assume that the assumptions 1.1–1.4 hold. If $a_n \rightarrow a$ and $2ab_1(\Omega) > 1$, then

$$
S_n^{1/2}(\tilde{\beta}_n - \beta) \xrightarrow{d} N\left(0, \frac{a^2b_1(\Omega)}{2ab_1(\Omega) - 1} I\right).
$$

In order to prove the theorem, we need the following lemma.

**LEMMA 4.1.** Let $S_n$ defined by (4.12). If $E(\|X_1\|^4) < \infty$, then

$$
E(S_n^{1/2} | S_n^{-1/2} | \tilde{\beta}_n) = \left(1 + \frac{1}{2n} \right) I + o(1) \quad a.s.
$$

$$
E(\|m(S_n^{1/2} S_n^{-1/2} - I) \|^2 | \tilde{\beta}_n) \leq c < \infty \quad a.s.
$$
Proof. Since
\[
\frac{S_n}{n} = \frac{S_{n+1}}{n} - \frac{1}{n} X_n' \Omega^{-1} X_{n+1},
\]
then
\[
\left( \frac{S_n}{n} \right)^{-1} = \left( \frac{S_{n+1}}{n} \right)^{-1} + \frac{1}{n} X_n' \Omega^{-1} X_{n+1}
\]
\[
\times \left( \Omega - \frac{1}{n} X_n' \left( \frac{S_{n+1}}{n} \right)^{-1} X_{n+1} \right)^{-1} \left( \frac{S_{n+1}}{n} \right)^{-1} \Omega^{-1} X_{n+1} \left( \frac{S_{n+1}}{n} \right)^{-1}.
\]
(4.13)

By the strong law of large numbers, we have
\[
\frac{S_n}{n} = I + o(1) \quad \text{a.s.}
\]
Therefore,
\[
\frac{1}{n} X_n' S_n X_{n+1} = O \left( \frac{1}{n} X_n' X_{n+1} \right) \to 0 \quad \text{a.s.} \quad (4.14)
\]
and
\[
\frac{1}{n} X_n' \left( \Omega - \frac{1}{n} X_n' \left( \frac{S_{n+1}}{n} \right)^{-1} X_{n+1} \right)^{-1} X_{n+1}
\]
\[
= O \left( \frac{1}{n} X_n' \Omega^{-1} X_{n+1} \right) \quad \text{a.s.} \quad (4.15)
\]
Denote \((1/n) X_n' (\Omega - (1/n) X_n' (S_{n+1}/n)^{-1} X_{n+1})^{-1} X_{n+1}\) by \(B_n/n\). By (4.13)–(4.15), we have
\[
\left( \frac{S_n}{n} \right)^{-1} = \left( \frac{S_{n+1}}{n} \right)^{-1} + \left( I + o(1) \right) \frac{B_n}{n} (I + o(1))
\]
\[
= \left( \frac{S_{n+1}}{n} \right)^{-1} + \frac{B_n}{n} + o \left( \frac{1}{n} B_n \right) \quad \text{a.s.} \quad (4.16)
\]
It is easy to see that
\[
\left( \frac{S_n}{n} \right)^{-1/2} = \left( \frac{S_{n+1}}{n} \right)^{-1/2} + \frac{B_n}{2n} + o \left( \frac{1}{n} B_n \right) \quad \text{a.s.} \quad (4.17)
\]
Therefore, by (4.16) and (4.17), it follows that

\[
S_{n+1}^{1/2} S_n^{-1/2} - I = \left( \frac{S_{n+1}}{n} \right)^{1/2} \left( \frac{S_n}{n} \right)^{-1/2} - I
\]

\[
= \left( \frac{S_{n+1}}{n} \right)^{1/2} B_n + o \left( \frac{1}{n} B_n \right)
\]

\[
= \frac{B_n}{2n} + o \left( \frac{1}{n} B_n \right) \text{ a.s.}
\]

which implies that

\[
E[(S_{n+1}^{1/2} S_n^{-1/2} - I) | \mathcal{F}_n] = \frac{1}{2n} E[B_n (1 + o(1)) | \mathcal{F}_n]
\]

\[
= \frac{1}{2n} E[X_{n+1}' \Omega^{-1} X_{n+1} | \mathcal{F}_n] + o \left( \frac{1}{n} \right)
\]

\[
= \frac{1}{2n} I + o \left( \frac{1}{n} \right) \text{ a.s.}
\]

and

\[
E[\|m(S_{n+1}^{1/2} S_n^{-1/2} - I)\|^2 | \mathcal{F}_n] \leq \frac{1}{2} E[\|X_{n+1}' \Omega^{-1} X_{n+1} (1 + o(1))\|^2 | \mathcal{F}_n]
\]

\[
\leq c < \infty \text{ a.s.}
\]

**Proof of Theorem 4.3.** Without loss of generality, assume that \( \beta = 0 \). By (1.5), the recursive algorithm for \( M \)-estimator of \( \beta \) can be written as

\[
S_{n+1}^{1/2} \tilde{\beta}_{n+1} = (S_{n+1}^{1/2} S_n^{-1/2}) S_n^{1/2} \tilde{\beta}_n + a_n S_n^{-1/2} h(\tilde{\beta}_n, \tilde{\nu}_n, \tilde{X}_{n+1}, e_{n+1}). \tag{4.18}
\]

Set \( S_n^{1/2} \tilde{\beta}_n = \tilde{\nu}_n \) and \( X_{n+1} S_n^{-1/2} = \tilde{X}_{n+1} \). Then (4.18) can be rewritten as

\[
\tilde{\beta}_{n+1} = S_{n+1}^{1/2} \tilde{\nu}_n + a_n S_n^{-1/2} h(\tilde{\beta}_n, \tilde{\nu}_n, \tilde{X}_{n+1}, e_{n+1})
\]

\[
= \left[ I + \frac{1}{n} \left( \frac{1}{2} I + F_n \right) \right] \tilde{\nu}_n + a_n (I + o(1)) h(\tilde{\beta}_n, \tilde{\nu}_n, \tilde{X}_{n+1}, e_{n+1}) \text{ a.s.}
\]

where

\[
F_n = n \left[ S_{n+1}^{1/2} S_n^{-1/2} - \left( \frac{1}{2} + \frac{1}{2n} \right) I \right].
\]
Using the technique as in the proof of Theorem 4.1, we have

\[ h_1(\hat{\beta}_n, \hat{\Sigma}_n, \hat{\gamma}_n + e_{n+1}) = \frac{1}{n} (ab_1(\Omega) I - A_n + \Psi_n - \Gamma_n) \hat{\beta}_n + n^{-1/2} \nu_n \]

where \( \Psi_n, \Gamma_n \) and \( \nu_n \) are defined in Theorem 4.1 with \( \hat{\Sigma}_n \) replaced by \( \hat{\Sigma}_{n+1} \). By Lemma 4.1, we get

\[ \hat{\beta}_{n+1} = \left( I - \frac{1}{n} \left( ab_1(\Omega) - \frac{1}{2} I - A_n + \Psi_n - \Gamma_n - \hat{\nu}_n \right) \right) \hat{\beta}_n + n^{-1/2} \nu_n \text{ a.s.} \]

The theorem follows from Lemma 3.4 and Lemma 3.5.

Remark 4.3. Since \( ab_2(\Omega)/(2b_2(\Omega)) < a^2b_2(\Omega)/(2ab_1(\Omega) - 1) \) for \( 2ab_1(\Omega) > 1 \), the algorithm given by (4.10) would be a better choice, which is simpler and more efficient or at least as efficient as compared with the other algorithms.

REFERENCES