This note discusses the asymptotic behavior of a class of M-estimators in linear models when errors are Gaussian, or a function of Gaussian random variables, that are long range dependent. The asymptotics are discussed when the design variables are either i.i.d. or long range dependent, independent of the errors, or known constants. It is observed that the class of M-estimators of the regression parameter vector corresponding to skew symmetric scores and symmetric errors asymptotically behave like the least squares estimators. Moreover, in these cases, if the design variables are either i.i.d. or known constants then the limiting distributions are Normal. But if the design variables are also long range dependent then the limiting distributions are nonnormal.
M-Estimators in Linear Models With Long Range Dependent Errors
by
Hira L. Koul

February 1990
Technical Report No. 283


222. I. Mitoma. Weak solution of the Langevin equation on a generalized functional space. Feb. 88. (Revised as No. 238).


M – Estimators in Linear Models With Long Range Dependent Errors

by

Hira L. Koul

Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824

and

Center for Stochastic Processes
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260

Abstract. This note discusses the asymptotic behavior of a class of M – estimators in linear models when errors are Gaussian, or a function of Gaussian random variables, that are long range dependent. The asymptotics are discussed when the design variables are either i.i.d. or long range dependent, independent of the errors, or known constants. It is observed that the class of M – estimators of the regression parameter vector corresponding to skew symmetric scores and symmetric errors asymptotically behave like the least squares estimators. Moreover, in these cases, if the design variables are either i.i.d. or known constants then the limiting distributions are Normal. But if the design variables are also long range dependent then the limiting distributions are nonnormal.

AMS 1980 Subject Classification: Primary 62M10, 65G30.

Key words and phrases: Hermite rank and polynomials. Linearity.

Part of research supported by the Air Force Office of Scientific Research Contract No. F49620 85C 0144.
1. **Introduction and Summary.** A discrete time stationary stochastic process is said to be long range dependent if its covariances decrease to zero like a power of lag as the lag tends to infinity but their absolute sum diverges. Such processes arise in applications in Hydrology, Economics, Time Series Analysis and other sciences. See, e.g., the review paper by Mandelbrot and Taqqu (1979) and references therein for the importance of these processes. See Granger and Joyeux (1980), and Hosking (1981) for the usefulness of these processes in Economics and Time Series Analysis. For many technical results on these processes, see Taqqu (1975, 1979), Fox and Taqqu (1987) and Dehling and Taqqu (1989), and Yajima (1985, 1988), among others.

One of the popular class of estimators in linear models that has evolved over the last two and a half decades is the so called class of M-estimators. Most of the asymptotic literature on these estimators assumes either independent errors (Huber: 1981 and references therein) or weakly dependent errors, like strongly mixing, as in Koul (1977).

Because of the importance of both, M-estimators and the long range dependence, it is of interest to study the large sample behavior of these estimators in a linear regression setting when errors are either long range dependent Gaussian or functions of such random variables (r.v.'s). About the design variables in the linear model we shall assume that they are either r.v.'s or known constants. In the former case it will be further assumed that the design variables are independent of the errors and either i.i.d. or long range dependent. The case of the known constant designs will be discussed in Section 3. We shall for the time being restrict our attention to the case of random designs.

Accordingly, let \( \eta_1, \eta_2, \ldots \) be a sequence of strictly stationary mean zero unit variance Gaussian r.v.'s with \( \rho(k) := E\eta_k\eta_{k+1}, \quad k \geq 0 \). Let \( \xi_1, \xi_2, \ldots \) be a sequence of observable \( p \times 1 \) stationary mean zero random vectors with \( \Gamma(k) := E\xi_1\xi_{1+k}', \quad k \geq 0 \).

Consider the linear model

\[
Y_i = X_i'\beta + \epsilon_i, \quad X_i' = (1, \xi_i'), \quad \beta \in \mathbb{R}^{p+1}, \ i \geq 1
\]
where $\epsilon_i \equiv G(\eta_i), i \geq 1,$ $G$ a measurable function from $\mathcal{R}$ to $\mathcal{R}$.

Note that the marginal distribution of $\epsilon_1$ need not be Gaussian. In fact if one were to have a linear regression model with stationary errors whose marginal distribution function (d.f.) is $F$, then choosing $G = F^{-1}(\Phi)$ would yield the desired errors. Here $\Phi$ is the d.f. of a $N(0,1)$ r.v. and $F^{-1}(u) = \inf\{x; F(x) \geq u\}, 0 \leq u \leq 1$.

The class of $M$-estimators, one corresponding to each $\psi$, is defined as a solution $\hat{\beta}_N$ of the equation

$$S(t) := \sum_{i=1}^{N} X_i \psi(Y_i - X_i^t) = 0$$

where $\psi$ is a measurable function from $\mathcal{R}$ to $\mathcal{R}$ with

$$E\psi(\epsilon) = 0, 0 < E\psi^2(\epsilon) < \infty.$$  

Here, and in the sequel, $\eta, \epsilon, \xi$ etc. are copies of $\eta_1, \epsilon_1, \xi_1$ etc. Also for a $p \times 1$ vector $t \in \mathcal{R}^p$, $t'$ will denote its transpose and $\|t\|$ will stand for its Euclidean norm.

The present paper is concerned with investigating the large sample behavior of $M$-estimators when the r.v.'s $\{\eta_i\}$, in addition, satisfy

$$\rho(k) = k^{-D_1} L_1(k), 0 < D_1 < 1, k \geq 1$$

where $L_1(k)$ is positive for large $k$ and slowly varying at infinity, i.e., $L_1(tx)/L_1(t) \to 1$ as $t \to \infty$ for every $x \in \mathcal{R}$.

About $\{\xi_i\}$ we shall additionally assume that

$$\{\xi_i\} \text{ are independent of } \{\epsilon_i\}$$

and either

$$\xi_1, \xi_2, \ldots \text{ are i.i.d. r.v.'s}$$

or
are dependent with $\Gamma(k) = k^{-D_2} \mathcal{L}(k), \ 0 < D_2 < 1$,

where $\mathcal{L}$ is a $p \times p$ matrix of slowly varying functions at infinity and $\mathcal{L}(k)$ are positive definite for all large $k$.

The processes that have covariances like (4) or (6b) are called long range dependent. These covariances tend to zero but not fast enough so as to be summable.

In the case when errors are independent or weakly dependent, $A_N(\hat{\beta}_N - \beta)$ turns out to be asymptotically normally distributed where $A_N$ equals $N^{1/2}$ in the case \{\xi_i\} are i.i.d. r.v.'s or $A_N$ equals $(X'X)^{1/2}$ in the case \{\xi_i\} are the known constants. Here $X'X = \sum_{i=1}^{N} X_i'X_i$.

Recall that the way this result is proved is first to approximate $\hat{\beta}_N - \beta$ by $\left\{ \sum_{i=1}^{N} X_i'X_i \psi'(\epsilon_i) \right\}^{-1} S(\beta)$. Then, by the LLN's, the first term in this approximation is seen to be of the order $N^{-1}$ and this $N^{-1}$ is split so as to stabilize $S(\beta)$ and $\hat{\beta}_N - \beta$. In the case the errors are independent or weakly dependent and the design variables are random, the scores $S(\beta)$ are of the order $O_p(N^{1/2})$ and hence one must have $A_N = N^{1/2}$. Note that, in view of the Ergodic Theorem, the first term in the above approximation is $O_p(N^{-1})$ as long as the summands $\{X_i'X_i \psi'(\epsilon_i)\}$ are stationary, ergodic, have finite first moments and $\{E[X_i'X_i \psi'(\epsilon_i)]\}^{-1}$ exists, regardless of whether the r.v.'s are long range dependent or not. Hence, even in the present case, the magnitude of $S(\beta)$ determines that of $\hat{\beta}_N - \beta$. The exposition in Section 2 below uses this observation. A similar observation is used in Section 3 when the design variables are the known constants.

One of the observations of this note is that the class of M-estimators corresponding to the skew symmetric scores and symmetric errors (i.e. skew symmetric $G$) asymptotically behaves like the least squares estimator under (6a) or (6b) or the known constant design case. This result, in the cases of (6a) and (6b), is stated and proved in Section 2 and in the other case, in Section 3, below. A similar observation was made by Beran and Kunsch.
in connection with the one sample location model. We further observe that in these cases if the design variables are either i.i.d. or known constants then the limiting distributions are Normal. But if the design variables are also strongly dependent and there is no intercept parameter in the model then the limiting distributions are nonnormal and appear at the end of Section 2.

In what follows, L, with or without suffix is a generic notation for a slowly varying function. All limits are taken as $N \to \infty$, unless mentioned otherwise. Also in most of our discussion the design variables need not be Gaussian.

2. The Case of Random Designs. A preliminary result needed for obtaining a first order approximation to $M$ - estimators is the asymptotic uniform linearity of $S$. The following theorem gives a set of sufficient conditions for such a result to hold. It also gives the required approximation to $M$ - estimators. The statement of the Theorem is somewhat self contained.

**Theorem 1.** Let $(\xi_1, \epsilon_1), (\xi_2, \epsilon_2)$ be a strictly stationary sequence of random vectors with $\xi_1$ being $p \times 1$. Let $X_i = (1, \xi_i)$, $Y_i = X_i^T \beta + \epsilon_i$, for some $\beta \in \mathbb{R}^{p+1}, i \geq 1$.

In addition assume the following:

(a) The score function $\psi$ satisfies (1.3) and is absolutely continuous with a.e. derivative $\psi'$ satisfying $\mathbb{E}|\psi'| < \infty$, and,

$$\mathbb{E}\|X_i\|^2 | \psi'(\epsilon - z\|X_i\|) - \psi'(\epsilon) | \to 0 \quad \text{as} \quad z \to 0.$$  

(b) For $N \geq p+1$, there are sequences $\{A_N\}$ and $\{B_N\}$ of $(p+1) \times (p+1)$ matrices which are positive definite for sufficiently large $N$ and satisfy

(i) $\|B_N^{-1}\| \to 0$, $\|A_N^{-1}\| \to 0$, $N \cdot \|A_N^{-1}\| \cdot \|B_N^{-1}\| \to 1$.

(ii) $\|B_N^{-1} \cdot S(\beta)\| = O_P(1),$.  

Then, for every $0 < b < \infty$,

\[(1) \quad \mathbb{E} \sup_{\|\Delta\| \leq b} \| B_N^{-1} [S(\beta + A_N^{-1} \Delta) - S(\beta)] + B_N^{-1} \sum_{i} X_i X_i' \psi'(\epsilon_i) A_N^{-1} \Delta \| = o_p(1) \]

where $S$ is as in (1.2).

In addition, if $\{\xi_i\}$ are independent of $\{\epsilon_i\}$ and if $\mathbb{E} \|\xi\|^2 < \infty$, then the random coefficient of $\Delta$ in the linear expansion (1) may be replaced by $R \cdot \mathbb{E} \psi'(\epsilon)$ where

$R := \mathbb{E} X_1 X_1'$.

Furthermore, if

(c) $R^{-1}$ exists, and (d) $0 < \mathbb{E} \psi'(\epsilon)$,

then

\[(2) \quad A_N(\hat{\beta}_N - \beta) = [R \mathbb{E} \psi'(\epsilon)]^{-1} \cdot B_N^{-1} S(\beta) + o_p(1). \]

**Remark 1.** It is perhaps worth repeating that in the above theorem neither $\{\xi_i\}$ nor $\{\epsilon_i\}$ need be Gaussian or functions of Gaussian r.v.'s.

**Proof.** From the definition of $S$, $S(\beta + A_N^{-1} \Delta) = \sum_{i} X_i \psi(\epsilon_i - X_i A_N^{-1} \Delta)$. Now, use the definition of absolute continuity and routine arguments to get that the

$$L.H.S.(1) \leq b \mathbb{E} \sum_{i} \| B_N^{-1} X_i \| \cdot \| X_i \| \cdot f \| \psi'(\epsilon - z \| X_i \| b) - \psi(\epsilon) \| dz \leq b \mathbb{E} \sum_{i} \| B_N^{-1} \| \cdot \| A_N^{-1} \| \cdot \| X_i \| \cdot f \| \psi'(\epsilon - z \| X_i \| b) - \psi(\epsilon) \| dz \to 0,$$

by (a) and (b)(i).

The claim about replacing $B_N^{-1} \sum_{i} X_i X_i' \psi'(\epsilon_i) A_N^{-1}$ by $R \cdot \mathbb{E} \psi'(\epsilon)$ follows from the Ergodic Theorem. The claim (2) is obtained from (1), (a)(ii), (c) and (d), with the help of Scheweder fixed point Theorem, just as in Huber (1981). \(\square\)
Remark 2. Observe that \( \psi(x) \equiv x \) a priori satisfies (a). Another example of \( \psi \) satisfying (a) is the Huber function \( \psi(x) := x I(|x| \leq c) + c \ \text{sgn}(x), \ c > 0 \), provided \( \{\xi_i\} \) are independent of \( \{c_i\} \), \( E\|\xi\|^2 < \infty \), and \( F \) is continuous at \( \pm c \). To see this observe that for this \( \psi \) the

\[
\text{L.H.S.}(a) \leq E\|X_1\|^2 \left\{ [F(c+z\|X_1\|) - F(c-z\|X_1\|)] + [F(-c+z\|X_1\|) - F(-c-z\|X_1\|)] \right\}.
\]

Now the Dominated Convergence Theorem gives the claim. \( \square \)

Observe that so far we have not used (1.4) or (1.6a) or (1.6b) or even the assumption about \( \{\eta_i\} \) being Gaussian. We shall now use these assumptions to determine the sequences of matrices \( \{A_N\} \) and \( \{B_N\} \). The main requirement on \( B_N \) is (b)(ii). Once \( B_N \) is determined, \( A_N \) can be determined from (b)(i).

In order to assess the magnitude of \( S \) (write \( S \) for \( S(\beta) \)) we shall use the Hermite expansion of \( L_2(\mathcal{R}; d\Phi) \) functions. What follows about Hermite expansions etc. is borrowed from Feller (1971) and Taqqu (1975). With \( \{H_q, q \geq 0\} \) denoting the Hermite polynomials, let \( J_q := E\psi_1(\eta)H_q(\eta) \), where \( \psi_1 = \psi(G) \). Let \( m := \min\{q \geq 1, J_q \neq 0\} \) denote the Hermite rank of \( \psi_1(\eta) \). The Hermite expansion of rank \( m \) of \( \psi_1(\eta) \) is given by

\[
\sum_{q=m}^{\infty} \frac{J_q}{q!} H_q(\eta).
\]

Recall from Feller (1971) that \( \{H_q(\eta_i)\} \) is a set of orthonormal r.v.'s in \( L_2(\mathcal{R}; d\Phi) \) satisfying

\[
H_0(x) \equiv 1, \ E H_q(\eta) = 0, \ q \geq 1;
\]

\[
E H_q(\eta_i) H_n(\eta_j) = \begin{cases} 0, & q \neq n \\ q! b^{q(i-j)}, & q = n \end{cases}, \qquad \forall \ i, j.
\]

Now, we begin the argument for determining \( B_N \) and \( A_N \). For a \( \lambda \in \mathbb{R}^{p+1} \), write \( \lambda = (\lambda_1, \lambda_2) \), \( \lambda_1 \in \mathcal{R} \), \( \lambda_2 \in \mathbb{R}^p \). From (1.5) and (3), \( \forall \ \lambda \in \mathbb{R}^{p+1} \),

\[
E[\lambda^\prime \sum_i X_i H_m(\eta)]^2 = m! \sum_i \sum_j [\lambda_1^2 + \lambda_2^2 \ G(i-j) \ \lambda_2] b^m(i-j).
\]
At this point we need to consider (1.6a) and (1.6b) separately.

Suppose that (1.6a) holds. Then the

\[
\text{RHS}(4) = m! \left[ \lambda_1^2 \sum_i \sum_j m^{i-j} + \lambda_2 \lambda_2 N \right]
\]

Now, if we restrict \( D_1 < 1/m \), then from Taqqu (1975; Lemma 3) it follows that the

\[
\text{RHS}(4) \approx c_1 \lambda_1^2 N^{2-mD_1} L(N) + m! \lambda_2 \lambda_2 N.
\]

where \( c_1 \) is a constant depending on \( D_1 \) and \( m \). Thus in this case if we choose

\[
B_N = \begin{bmatrix}
H_1 L(N) & 0_{1 \times p} \\
0_{p \times 1} & N_{1 \times p} \end{bmatrix} = \begin{bmatrix}
b_N & 0 \\
0 & B_{N2} \end{bmatrix}, \text{ say,}
\]

with \( 2H_1 = 2-mD_1 \), then we see that

\[
E(\lambda^2 B_N^{-1} \sum_i X_i H_m(\eta_i))^2 = O(1) \ \forall \ \lambda \in \mathbb{R}^{p+1}.
\]

We note that \( D_1 < 1/m \) implies that \( \{ \psi(\epsilon_i) \} \) are also long range dependent. The case \( D_1 \geq 1/m \) would yield that these r.v.'s are asymptotically weakly dependent and not interesting to us from the current point of view.

Now suppose that (1.6b) holds. Then the

\[
\text{RHS}(4) \approx m! \left[ \lambda_1^2 \sum_i \sum_j m^{i-j} L(i-j) + \lambda_2 \sum_{i,j} \mathcal{A}(i-j) \lambda_2 \lambda_2 |i-j|^{2-mD_1} \right]
\]

Note that \( \mathcal{A} \) being a matrix of slowly varying functions at infinity and that \( \mathcal{A}(k) \) being positive definite for all large \( k \), it follows that for every \( \lambda_2 \in \mathbb{R}^p \), \( \lambda_2 \mathcal{A}(k) \) is slowly varying at infinity and that \( \lambda_2 \mathcal{A}(k) \lambda_2 > 0 \) for all large \( k \) and for every \( \lambda_2 \in \mathbb{R}^p \).

Once again, use the arguments as in Taqqu (op cit.) to conclude that the

\[
\text{R.H.S.}(4) \approx c_1 \lambda_1^2 N^{2-mD_1} + c_2 \lambda_2 \mathcal{A}(N) \lambda_2 N^{2-mD_1-D_2}
\]
provided we assume

\[(7) \quad 0 < D_1 < 1/m, \quad mD_1 + D_2 < 1, \quad 0 < D_2 < 1.\]

Here \(c_1\) and \(c_2\) are constants depending on \(m, D_1\) and \(D_2\). Thus a choice of \(B_N\) here is

\[(8) \quad B_N = \begin{bmatrix} H_1 & 0_{1 \times p} \\ 0_{p \times 1} & H_2 \end{bmatrix}, \quad L(N) = \begin{bmatrix} b_{N1} & 0 \\ 0 & B_{N2} \end{bmatrix}, \quad \text{say}\]

with \(H_1\) as in (5) and \(2H_2 = 2 - mD_1 - D_2\).

With this \(B_N\), one can again verify that (6) above holds in this case. Note that (7) implies

\[(9) \quad 1/2 < H_1 < 1, \quad 1/2 < H_2 < 1.\]

Next, in view of (1.3), (1.5) and (3), \(\forall \lambda \in \mathbb{R}^{d+1},\)

\[E[\lambda \cdot B_N^{-1} \sum \psi_i(\eta_i) - \int_0^1 H_m(\eta_1)]^2 = E[\sum \lambda \cdot B_N^{-1} X_i \sum_{q=m+1}^\infty \frac{J_q}{q!} H_m(\eta_i)]^2 \]

\[= \sum_{q=m+1}^\infty \sum_{i,j} \frac{J^2}{q!} \sum E \lambda \cdot B_N^{-1} X_i \cdot B_N^{-1} \lambda \cdot \rho^q(i-j) \]

\[\leq \sum_{q=m+1}^\infty \frac{J^2}{q!} \sum \lambda \cdot B_N^{-1} \cdot 1 \cdot 0 \cdot \Gamma(i-j) \cdot B_N^{-1} \lambda \cdot \rho^{m+1}(i-j) \]

\[(10) \quad \rightarrow 0,\]

by arguing as in Taqqu (1975, p. 294), under both (1.6a) of (1.6b), using \(B_N\) as in (5) or in (8), as the case may be.

Combining (5), (6), (8), and (10), one sees that under either (1.6a) or (1.6b) (with \(B_N\) as in (5) or as in (8), respectively) one has, by (3),
\[ \text{Var}(\lambda^* B_N^{-1} \sum_i X_i \psi_1(\eta_i)) \]

\[ = \text{Var}[\sum_i (\lambda^* B_N^{-1} X_i)\{\psi_1(\eta_i) - \frac{J_m}{m} H_m(\eta_i)\}] + \text{Var}[\frac{J_m}{m} \sum_i (\lambda^* B_N^{-1} X_i) H_m(\eta_i)] \]

\[ = o(1) + O(1) = O(1). \]

This then determines \( B_N \) and verifies the assumption (b)(ii) of the Theorem 1 above when \( \{\xi_i\}, \{\eta_i\}, \{\epsilon_i\} \) are as in (1.1), (1.4), (1.5), (1.6a) or (1.6b). Now, if \( B_N \) is given by (5), then

\[ A_N = \begin{bmatrix} N^{1-H} L(N) & 0 \\ 0 & N^{\frac{1}{2}} I_{p \times p} \end{bmatrix} = \begin{bmatrix} a_{N1} & 0 \\ 0 & A_{N2} \end{bmatrix}, \text{ say,} \]

will satisfy (b)(i). If \( B_N \) is given by (8), then

\[ A_N = \begin{bmatrix} N^{1-H_1} \\ 0 \end{bmatrix} L(N) = \begin{bmatrix} a_{N1} & 0 \\ 0 & A_{N2} \end{bmatrix}, \text{ say,} \]

will satisfy (b)(i), with \( H_1 \) and \( H_2 \) as in (5) and (8) satisfying (9).

The above discussion is now summarized as

**Theorem 2.** Let \( \{Y_i\}, \{\xi_i\}, \{\eta_i\}, \beta, \psi \) satisfy (1.1), (1.3), (1.4), (1.5), and (1.6a) or (1.6b). In addition assume that \( \psi \) is nondecreasing satisfying (a) of Theorem 1 with \( 0 < E\psi'(\epsilon) \).

Then, with \( \hat{\beta}_N \) defined as a solution of (1.2),

\[ A_N(\hat{\beta}_N - \beta) = [R E\psi'(\epsilon)]^{-1} B_N^{-1} \sum_i X_i H_m(\eta_i) \frac{J_m}{m} + o_p(1). \]

where \( B_N, A_N \) are as in (5), (11), (8), (12)) in the case of (1.6a) ((1.6b)).
Remark 4. Hermite rank m of $\psi_1$. Often the function $\psi$ is chosen to be skew symmetric, viz. $\psi(-x) \equiv \psi(x)$. Thus if $G$ is also skew symmetric then $\psi_1(-x) \equiv \psi(G(-x)) \equiv \psi(-G(x)) \equiv -\psi(G(x)) \equiv -\psi_1(x)$. In such cases, using the fact that $H_q(-x) \equiv (-1)^q H_q(x)$ for all $q$, we have

$$J_q = E\psi_1(\eta)H_q(\eta) = \{1 + (-1)^{q+1}\} E\{\psi_1(\eta)H_q(\eta)I(\eta > 0)\} \neq 0, \quad q = 1.$$ 

Therefore, $m = 1$, $J_1 = 2 E\{\psi_1(\eta)\eta I(\eta > 0)\}$, $H_1(\eta) = \eta$ and, from Theorem 2,

$$A_N(\hat{\beta}_N - \beta) = [R \ E\psi'(\epsilon_1)]^{-1} \cdot B_N^{-1} \sum_i X_i \eta_i \cdot J_1 + o_p(1). \quad (13)$$

Now let $\beta_N$ be the least squares estimator of $\beta$ in (1.1). Then carrying out an analysis like the above one can derive the following:

If $EG(\eta) = 0$, $0 < EG^2(\eta) < \infty$ and $G$ is skew symmetric, then

$$A_N(\beta_N - \beta) = R^{-1} \cdot B_N^{-1} \sum_i X_i \eta_i \cdot \alpha_1 + o_p(1),$$

where $\alpha_1 = EG(\eta)\eta$ where $A_N$ and $B_N$ are the same as in (13).

The r.v. $\sum X_i \eta_i$ is precisely the leading term in the least squares estimator of the regression parameter with the errors $\{\eta_i\}$ and the design vectors $\{X_i\}$. Thus it follows that the above class of $M$ – estimators corresponding to the skew symmetric scores and symmetric errors are asymptotically like the least squares estimators regardless of whether the errors are Gaussian or not.

Now, suppose that there is no intercept parameter in (1.1). Then the result like (13), with $X_i$ replaced by $\xi_i$, $A_N$, $B_N$ replaced by $A_{N2}$, $B_{N2}$ of (5) and/or (8) remains valid. Of course now $\hat{\beta}_N$ is $p \times 1$ as is $\beta$. Note that in the case of (1.6a),

$$\lambda \cdot B_{N2}^{-1} \sum_i \xi_i \eta_i = N^{-1} \sum_i (\lambda' \xi_i) \eta_i = N_p(0, \lambda' \Gamma \lambda), \quad \forall \lambda \in \mathcal{R}^p,$$

where $\Gamma = \Gamma(0) = E\xi_1 \xi_1'$. 
But in the case of (1.6b) the limiting distribution is different. To determine this limiting distribution, we use Theorem 6.1 of Fox and Taqqu (1987). Observe that if \( \{ \xi_i \} \) are long range dependent and Gaussian then so are the r.v.'s \( \{ \lambda \xi_i \} \) for every \( \lambda \in \mathcal{R}^p \) with the same exponent \( D_2 \) as in (1.6b). Now, take \( X_i \) and \( Y_i \) in Theorem 6.1 of Fox and Taqqu to be \( \lambda \xi_i \) and \( \eta_i \), respectively. One then sees that (1.4), (1.5) and (1.6b) together with the Gaussianness assumptions imply all the conditions of that Theorem for every \( \lambda \). Hence,

\[
\lambda^* B_{N_2}^{-1} \sum_i \xi_i \eta_i = N^{-1} L(N) \sum_i (\lambda^* \xi_i) \eta_i = Z(1) \cdot (\lambda^* \Gamma \lambda)^{1/2}
\]

with \( Z(1) \) obtained from (6.1) of Fox and Taqqu after \( t \) is set equal to 1 in there. \( \square \)

3. The Case of Non-Random Designs. In order to separate this case from that of the random designs, we shall now denote an \( N \times p \) design matrix of known constants by \( C \) and its \( i \)th row by \( c_{Ni} \), \( 1 \leq i \leq N \). Consider the linear regression model where one observes \( \{ Y_{Ni} \} \) satisfying

\[
Y_{Ni} = c_{Ni} \beta + \epsilon_i, \quad 1 \leq i \leq N, \quad \beta \in \mathcal{R}^p,
\]

with \( \{ \epsilon_i \} \) as in (1.1).

Throughout we shall assume that

(L1) \( (C^* C)^{-1} \) exists for all \( N \geq p \).

The class of M-estimators \( \hat{\beta}_N \) is defined as a solution \( t \) of

\[
T(t) := \sum_i c_{Ni} \psi(Y_{Ni} - c_{Ni}^* t) = 0,
\]

where \( \psi \) is assumed to satisfy (1.3). Again, our objective here is to investigate the large sample behavior of these estimators when \( \{ \eta_i \} \) satisfy (1.4). Of course, conceptually the discussion that follows is similar to that in Section 2 above except for the difficulties
created by the nonstationarity that is introduced in the problem by \( \{c_{Ni}\} \). We begin by giving

**Theorem 1.** Let \( \epsilon_1, \epsilon_2, \ldots \) be a strictly stationary sequence of r.v's and \( C \) be as above satisfying (L1) and assume (1) above holds. In addition, assume that the following hold:

(L2) The score function \( \psi \) is absolutely continuous with its almost everywhere derivative \( \psi' \) satisfying \( E|\psi'(\epsilon)| < \infty \) and such that the function

\[
z - E|\psi'(\epsilon - z) - \psi'(\epsilon)| \text{ is continuous at zero.}
\]

(L3) There exists sequences \( \{A_N\} \) and \( \{B_N\} \) of \( p \times p \) matrices such that they are positive definite for sufficiently large \( N \) and satisfy

(i) \( \|A_N^{-1}\| \to 0, \|B_N^{-1}\| \to 0 \);

(ii) \( B_N^{-1}C'CA_N^{-1} = I_{p \times p} \)

(iii) \( \max_i \|A_N^{-1}c_{Ni}\| \to 0 \),

(iv) \( \|B_N^{-1}T(\beta)\| = O_p(1) \).

Then, for every \( 0 < b < \infty \),

\[
E \sup_{\|\Delta\| \leq b} \|B_N^{-1}[T(\beta + A_N^{-1}\Delta) - T(\beta)] + B_N^{-1} \sum_i c_{Ni}c_{Ni}' \psi'(\epsilon_i)A_N^{-1}\Delta\| = o(1).
\]

If, in addition, \( \epsilon_i = G(\eta_i) \), with \( \{\eta_i\} \) satisfying (1.4),

(L4) \( \psi \) is nondecreasing, \( 0 < E\psi'(\epsilon), \ E(\psi'(\epsilon))^2 < \infty \), and

(L5) \( N^{1-(D/2)} \max_{1 \leq i \leq N} \|B_N^{-1}c_{Ni}\| \cdot \|A_N^{-1}c_{Ni}\| \to 0 \), with \( D = D_1 \) of (1.4)

then

\[
B_N^{-1} \sum_i c_{Ni}c_{Ni}' \psi'(\epsilon_i)A_N^{-1} = I_{p \times p} \cdot E\psi'(\epsilon) + o_p(1),
\]

and

\[
A_N(\beta_N - \beta) = [E\psi'(\epsilon)]^{-1} \cdot B_N^{-1}T(\beta) + o_p(1).
\]

**Remark 1.** Some comments about the assumptions are in order. The assumptions (L2) and (L3) are similar to the assumptions (a) and (b) of Theorem 2.1 above. Recall that in the
linear regression model with independent or weakly dependent errors and with the design matrix \( C \), the magnitude of \( T(\beta) \) is of the order

\[
\delta_N := (C^*C)^{\frac{1}{2}}
\]

However in the current situation, where \( \{\epsilon_i\} \) are functions of long range dependent r.v.'s, we can not expect this magnitude. But we must still have (L3)(ii) in order to stabilize the LHS(4).

In the case of random and stationary design variables, as in Section 2 above, an analogue of (4) is given by the Ergodic Theorem which does not require the second moment of the summands. But in the present situation, the LHS of (4) is neither stationary nor independent. The assumptions (L4), (L5) and (L3)(ii) together with the Gaussianness of \( \{\eta_i\} \) is used to conclude (4) below.

**Proof.** To simplify writing, let \( a_i := A_N^{-1}c_{Ni}, b_i := B_N^{-1}c_{Ni}, 1 \leq i \leq N. \) Now, by the absolute continuity of \( \psi \), the Fubini Theorem and the Cauchy–Schwarz inequality, the

\[
\text{LHS}(3) \leq 2b \sum_i \|b_i\| \|a_i\| \left\{2\|a_i\|\right\}^{-\frac{1}{2}} \int E|\psi'(\epsilon-zb) - \psi'(\epsilon)| \, dz
\]

\[
\leq 2b\left(\sum_i \|b_i\|^2 \sum_i \|a_i\|^2\right)^{\frac{1}{2}} \times \max_i \left(\sum_i \|a_i\|^2\right)^{-\frac{1}{2}} \int E|\psi'(\epsilon-zb) - \psi'(\epsilon)| \, dz \rightarrow 0,
\]

by (L2), (L3)((i)–(iii)). Note that by (L3)(ii),

\[
\sum_i \|b_i\|^2 \sum_i \|a_i\|^2 = \tr B_N^{-1}C^*CB_N^{-1} \cdot A_N^{-1}C^*CA_N^{-1} = p = O(1)
\]

where \( \tr A := \text{trace } A \) for any matrix \( A \).

Next, let \( \psi_2(\eta) := \psi'(\epsilon) = \psi'(G(\eta)) \) and \( \alpha_q := E\psi_2(\eta)H_q(\eta). \) In view of (L4), the Hermite expansion of \( \psi_2(\eta_i) - E\psi_2(\eta) \) is

\[
\sum_{q=1}^{\infty} \frac{\alpha_q}{q!} H_q(\eta_i). \]

Also note that the LHS(4) above is now \( \sum_i b_i a_i \psi_2(\eta_i). \) Hence \( \forall \lambda \in R^p, \)

\[
E\|\lambda^* \sum_i b_i a_i^* (\psi_2(\eta_i) - E\psi_2(\eta))\|^2 = E\|\lambda^* \sum_i b_i a_i^* \sum_{q=1}^{\infty} \frac{\alpha_q}{q!} H_q(\eta_i)\|^2
\]
\[
= \frac{\alpha^2}{q!} \sum_{i}^{\infty} \sum_{j}^{\lambda} \lambda, b_i a^i a_j b^j \lambda \cdot \rho^{q(i-j)} \quad \text{by (2.3)},
\]

\[
\leq \text{Var}(\psi_2(\eta)) \cdot \|\lambda\|^2 \max_{i,j} \|b_i a^i\|^2 \sum_{i}^{\lambda} \sum_{j} \rho(i-j).
\]

\[
(6) \quad \max_{i} \{\|B_{N}^{-1} c_{N_i}\| \cdot \|A_{N}^{-1} c_{N_i}\|^2 \} \cdot O(N^{2-D}),
\]

because \(\sum_{i,j} \rho(i-j) = O(N^{2-D})\) and because

\[
\|b_i a^i\|^2 = \text{tr} \cdot (b_i a^i a_j b^j) = \text{tr} \cdot (b_i b_i) \cdot (a_i a_j)
\]

\[
\leq \|b_i\|^2 \|a_i\|^2 \leq C_{N_i} \cdot \|A_{N}^{-1} c_{N_i}\|^2.
\]

Therefore (4) follows from the assumption (L4) and (6). The result (5) follows as in Huber (op cit.).

Our next objective is to determine \(B_N\), using (L3)(iv). Again, to simplify exposition we shall write \(\{c_i\}\) for \(\{c_{N_i}\}\). Proceeding as in Section 2, we observe that \(\forall \lambda \in \mathcal{K}^p\),

\[
E\left(\lambda \cdot \sum c_i \psi_1(\eta)\right)^2 = E\left(\lambda \cdot c_i H_m(\eta) \cdot \frac{J_m}{m!}\right)^2 + E\left(\lambda \cdot c_i [\psi_1(\eta) - H_m(\eta) \cdot \frac{J_m}{m!}]\right)^2
\]

\[
= E\left(\lambda \cdot c_i H_m(\eta) \cdot \frac{J_m}{m!}\right)^2 + E\left[\lambda \cdot c_i \sum_{q \neq m+1} \frac{J_q}{q!} H_q(\eta)\right]^2
\]

\[
= \lambda \cdot \sum c_i c_j \rho^{m(i-j)} \cdot \frac{J_m}{m!} + \sum_{q \neq m+1} \frac{J_q}{q!} \sum c_i c_j \lambda \cdot \rho^{q(i-j)}
\]

\[
(7) \quad = \frac{J_m}{m!} \lambda \cdot K_{N1} \lambda + \lambda \cdot K_{N2} \lambda
\]

where

\[
K_{N1} := \sum_{i,j} c_i c_j \rho^{m(i-j)}, \quad K_{N2} := \sum_{q \neq m+1} \frac{J_q}{q!} \sum c_i c_j \rho^{q(i-j)}.
\]
At this point one is clearly persuaded to choose \( B_N \approx K_N^{\frac{1}{2}} \) and then try to show that 
\[ \|K_{N1}^{-\frac{1}{2}}K_{N2}K_{N1}^{-\frac{1}{2}}\| \to 0 \] 
so that we would have (L3)(iv) satisfied. Such a process, though feasible, appears to be quite involved for general \( \{c_i\} \). However, if we make some further assumptions on the design variable then this process is less involved and more transparent.

Accordingly, let \( \varphi^\perp := (\varphi_1, \ldots, \varphi_p) \) be a vector of measurable functions on \([0,1]\) to \( A \) satisfying the following conditions:

(a1) With \( D = D_1 \) and \( L \) as in (1.4), \( m \) as the Hermite rank of \( \psi_1(\eta) \),

(i) \[ \int_0^1 \int_0^{1-u} |\varphi_\ell(u) \varphi_k(u+v)v^{-mD}L(v)| \, dv \, du < \infty, \quad D < 1/m, \]

(ii) \[ \int_0^1 |\varphi_\ell(u) \varphi_k(u)| \, du < \infty, \quad \ell, k = 1, 2, \ldots, p. \]

(a2) (i) \( N^{-D/4} \max_{1 \leq i \leq N} \|\varphi(i/N)\| \to 0; \)  
(ii) \( N^{-1+mD} \max_{1 \leq i \leq N} \|\varphi(i/N)\|^2 \to 0. \)

(a3) The matrix \( \mathcal{G}^{-1} \) exists, where

\[ \mathcal{G} = ((g_{\ell k})), \quad g_{\ell k} = \int_0^1 \int_0^1 \varphi_\ell(u) \varphi_k(v) |v-u|^{-mD}L(|v-u|) \, du \, dv, \quad \ell, k = 1, \ldots, p. \]

Given such a collection of \( \varphi^\perp \)'s, choose

(8) \[ c_i := \varphi(i/N), \quad 1 \leq i \leq N. \]

Now observe that

\[ N^{-1} \sum_{i,j} c_i c_j^* \approx \int_{1/N}^{1-1/N} \varphi(u) \varphi^\perp(u) \, du \to \int_0^1 \varphi(u) \varphi^\perp(u) \, du, \]

so that

(9) \[ N^{-2+mD} \mathbf{C}^\top \mathbf{C} \to 0, \quad \text{because } -1 + mD < 0. \]

From (1.4), (8) and the slowly varying property of \( L \) it follows that

\[ \lambda^\top K_{N1} \lambda = \sum_{\ell=1}^p \sum_{k=1}^p \lambda_\ell \lambda_k \sum_{i,j} \varphi_\ell(i/N) \varphi_k(j/N) \rho^m(j-i) \]
\[= \lambda' C' C \lambda + 2 \sum_{\ell=1}^{P} \sum_{k=1}^{P} \lambda_{\ell} \lambda_{k} \sum_{i < j} \varphi(i/N) \varphi_k(j/N) \rho^{m(j-i)} \]
\[\approx N^{2-mD} 2 \sum_{\ell=1}^{P} \sum_{k=1}^{P} \lambda_{\ell} \lambda_{k} \int_{0}^{1-u} \varphi(u) \varphi_k(u+v) v^{-mD} L(v) dv \]
\[= N^{2-mD} \lambda' \lambda. \]

Now let
\[B_N := N^H \varphi^{1/2}, \quad H = 1-(mD/2), \quad D = D_1 \text{ of (1.4)}.\]

Our next objective is to show that the second term in the RHS(7) is \(O(N^{2H})\). To that effect, note that \(q \geq m+1, |\rho(k)| \leq 1, \forall k \geq 1, \) imply that

\[|\lambda' \sum_{i < j} 1 \varphi(i/N) \rho^{m+1(j-i)}| \leq \sum_{\ell=1}^{P} \sum_{k=1}^{P} |\lambda_{\ell} \lambda_{k}| \sum_{i < j} |\varphi(i/N) \varphi_k(j/N) \rho^{m+1(j-i)}|.\]

Now, since \(|\rho(k)| \rightarrow 0 \text{ as } k \rightarrow \infty, \forall \epsilon > 0 \exists N_{\epsilon} \text{ such that } |\rho(k)| \leq \epsilon \ \forall k > N_{\epsilon}. \) Hence, \(\forall N > N_{\epsilon},\)
\[\sum_{i < j} |\varphi(i/N) \varphi_k(j/N) \rho^{m+1(j-i)}| \leq \sum_{|j-i| \leq N_{\epsilon}} \sum_{i < j} |\varphi(i/N) \varphi_k(j/N)| \]
\[+ \epsilon \cdot \sum_{i < j; (j-i) > N_{\epsilon}} |\varphi(i/N) \varphi_k(j/N) \rho^{m(j-i)}| \]
\[= T_{N1} + \epsilon \cdot T_{N2}, \quad \text{say.} \]

But, \(\forall \ell, k = 1, \ldots, p,\)
\[T_{N1} \leq N_{\epsilon} \cdot N \cdot \max_i \|\varphi(i/N)\|^2 = o(N^{2-mD}) , \quad \text{by (a2)(ii)}, \]
\[T_{N2} \leq \sum_{i < j} \sum |\varphi(i/N) \varphi_k(j/N)| |\rho^{m(j-i)}| \]
\[\approx N^{2-mD} \int_{0}^{1} \int_{0}^{1-u} |\varphi(u) \varphi_k(v+u) v^{-mD} L(v)| du dv = O(N^{2-mD}), \quad \text{by (a1)(i)}. \]
Hence, \( \forall \epsilon > 0, \exists N_\epsilon \) such that

\[
(12) \quad \text{LHS}(11) \leq o(N^{2-mD}) + \epsilon \cdot O(N^{2-mD}), \quad \forall N > N_\epsilon.
\]

From (9), (12) and the definition of \( K_{N2} \) it follows that \( \forall N > N_\epsilon \),

\[
N^{-2+mD} |\lambda \cdot K_{N2} \lambda| \leq \text{Var} \psi_1(\eta) \cdot \{ |\lambda \cdot C \cdot \lambda| + \text{LHS}(11) \}
\leq o(1) + \epsilon \cdot O(1) \rightarrow 0, \quad \text{by now letting} \ \epsilon \rightarrow 0.
\]

It thus follows that (L3)(iv) holds with \( B_N \) given by (10). From (L3)(ii) we get

\[
A_N = B_N^{-1} \cdot C \approx N^{1-H/2} f_0 f^1 \varphi^t.
\]

Note that \( m \geq 1 \Rightarrow \)

\[
\max_i ||A_N^{-1} c_i|| \approx \max_i N^{-mD/2} ||\varphi(i/N)|| \leq N^{-D/4} \max_i ||\varphi(i/N)||
\]

and

\[
N^{1-D/2} \max_i [||A_N^{-1} c_Ni|| \cdot ||B_N^{-1} c_Ni||] \approx [N^{-D/4} \max_i ||\varphi(i/N)||]^2
\]

so that (a2) implies (L3)(ii) and (L5). This shows that all the assumptions of Theorem 1 are satisfied. We now summarize the above discussion as

**Theorem 2.** Suppose that the linear regression model (1), with errors as in (1.1) and (1.4), holds. About the design variables \( \{c_Ni\} \) and the score function \( \psi \) assume that (8), (a1)-(a3), (1.3), (L2) and (L4) hold. Then M-estimators \( \{\hat{\beta}_N\} \) defined as solutions of (2) satisfy

\[
(14) \quad N^{1-H}(\hat{\beta}_N - \beta) = \{m! \cdot f_0^1 \varphi^t \cdot E\psi^*(\epsilon)\}^{-1} \cdot N^{-H} \Sigma_i \varphi(i/N)H_m(\eta_i)J_m + o_p(1),
\]

where \( H = (1-mD/2), \quad D = D_1 \) of(1.4).

**Remark 2.** Observe that if the design generating functions are bounded then \( 0 < D < 1/m \) guarantees the satisfaction of (a1) and (a2). In particular if \( \varphi_\ell(u) = u^\ell, \ell = 1, \ldots, p, \) then
(a1) – (a3) are all satisfied. That is, all of these conditions are satisfied in the case of the pth order polynomials.

An example of an unbounded design is obtained by taking p=1, \( \varphi_1(u) = u^{-r} \), \( r > 0 \). Then (a1)–(a3) are satisfied as long as \( r < (1-mD)/2 \).

**Remark 3.** An analogue of Remark 2.4 applies here also with obvious modifications.

Consequently, for skew symmetric \( \psi \) and symmetric errors the asymptotic distribution of 
\( N^{1-H}(\hat{\beta}_N - \beta) \) is \( p \)-variate Normal with mean vector 0 and the covariance matrix 
\[
\left[ \{J_0^{1} \psi \varphi^t \}^{-1} \cdot \{E \psi'(\varepsilon)\}^{-2} \cdot J_1^{1} \right].
\]

**Acknowledgement.** Author would like to thank Murad Taqqu for providing a generous supply of preprints and reprints of his and his co-authors work on long range dependent processes.
REFERENCES


256. E. Mayer-Wolf, A central limit theorem in nonlinear filtering, Apr. 89.

257. C. Houdré, Factorization algorithms and non-stationary Wiener filtering, Apr. 89.

258. C. Houdré, Linear Fourier and stochastic analysis, Apr. 89.


261. I. Herbst and L. Pitt, Diffusion equation techniques in stochastic monotonicity and positive correlations, Apr. 89.

262. R. Selukar, On estimation of Hilbert space valued parameters, Apr. 89. (Dissertation)

263. E. Mayer-Wolf, The noncontinuity of the inverse Radon transform with an application to probability laws, Apr. 89.

264. D. Monrad and W. Philipp, Approximation theorems for weakly dependent random vectors and Hilbert space valued martingales, Apr. 89.

265. K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes, Apr. 89.

266. S. Evans, Association and random measures, May 89.

267. H. Hurd, Correlation theory of almost periodically correlated processes, June 89.

268. O. Kallenberg, Random time change and an integral representation for marked stopping times, June 89.

269. O. Kallenberg, Some uses of point processes in multiple stochastic integration, Aug. 89.

270. W. Wu and S. Cambanis, Conditional variance of symmetric stable variables, Sept. 89.


272. O. Kallenberg, On an independence criterion for multiple Wiener integrals, Sept. 89.


274. G.W. Johnson and G. Kallianpur, Homogeneous chaos, p-forms, scaling and the Feynman integral, Sept. 89.

275. T. Hida, A white noise theory of infinite dimensional calculus, Oct. 89.

276. K. Benhenni, Sample designs for estimating integrals of stochastic processes, Oct. 89. (Dissertation)


283. H.L. Koul, M-estimators in linear models with long range dependent errors, Feb. 90.