Abstract. Let \( X_j = \sum_{r=0}^{\infty} a_r Z_{j-r} \) be a one-sided \( m \)-dimensional linear process, where \((Z_n)\) is a sequence of i.i.d. random vectors with zero mean and finite covariance matrix. The aim of this paper is to prove the moment inequalities of the form
\[
\mathbb{E}|S_n|^Q \leq C_n^{Q/2}
\]
for the sum
\[
S_n = \sum_{j=1}^{n} (G(X_j) - \mathbb{E}G(X_j)),
\]
where \( G \) is a real function defined on \( \mathbb{R}^m \). The form of the constant \( C \) in (0.1) plays an important role in applications concerning the problems of \( M \)-estimation, especially the Ghosh representation.

2000 AMS Mathematics Subject Classification: Primary: 62M10, 62F12; Secondary: 60G35.

Key words and phrases: Linear process, empirical processes, time series, short-range dependence, \( M \)-estimation, the Ghosh representation.

1. INTRODUCTION

The one-sided linear process (LP) is defined as follows:
\[
X_j = \sum_{r=0}^{\infty} a_r Z_{j-r}, \quad j = 1, 2, \ldots, n,
\]
where the innovations \((Z_n)\) are i.i.d. random variables with zero mean and unit variance, and \( a_r \) are constant coefficients satisfying the condition
\[
\sum_{r=0}^{\infty} a_r^2 < \infty.
\]
Linear processes have a wide range of applications in time series analysis. A large class of time series processes can be modelled in such a way, including
a subset of the fractional ARIMA processes (see Brockwell and Davis (1987)). In addition, we say that \((X_n)\) has a short memory (is short-range dependent) if the following condition is satisfied:

\[
\sum_{j=1}^{\infty} |\text{Cov}(X_1, X_{1+j})| < \infty.
\]

It is obvious that if

\[
\sum_{j=0}^{\infty} \sum_{r=0}^{\infty} |a_r a_{r+j}| < \infty,
\]

then the linear process \((X_n)\) is short-range dependent.

The generalization of the one-sided linear process is the multidimensional linear process (MLP), defined as follows:

\[
X_j = \sum_{r=0}^{\infty} A_r Z_{j-r},
\]

where the innovations \((Z_n) = (Z_n^{(1)}, \ldots, Z_n^{(m)})\) are i.i.d. random vectors in \(\mathbb{R}^m\) such that

\[
\mathbb{E}(Z_1) = 0 \quad \text{and} \quad \mathbb{E}|Z_1|^2 = 1.
\]

The \(A_r\)'s are the nonrandom matrices, where \(\sum_{r=0}^{\infty} \|A_r\|^2 < \infty\), \(A_0 = I\), and

\[
|A_r z| \leq \|A_r\| |z| \quad \text{for any} \ z \in \mathbb{R}^m,
\]

where \(|\cdot|\) is the usual Euclidean norm.

We will also consider the following assumptions:

\[(a_0)\]

\[
\sum_{r=0}^{\infty} \|A_r\| < \infty,
\]

\[(b_1)\]

the density of the vector \(Z_1\) is the Lipschitz function,

\[(b_2(t))\]

\[
\mathbb{E}|Z_1|^t < \infty \quad \text{for some} \ t \geq 2.
\]

In Section 2, we will prove the inequality for the moment bound of the sum of the functionals of linear processes. Namely, we will show that if

\[(1.2)\]

\[
Y_n = G(X_n) - \mathbb{E}G(X_n),
\]

then for different classes of functions \(G\)

\[(1.3)\]

\[
\mathbb{E}\left| \sum_{j=1}^{n} Y_j \right|^Q \leq Cn^{Q/2}, \quad n = 1, 2, \ldots,
\]
where the constant $C$ may depend on $Q$, $G$ and the distribution of $(X_n)$, but does not depend on $n$.

The basic idea in the proof of our result is the martingale decomposition of the process (1.2). This decomposition has the following form (see also Ho and Hsing (1996), (1997)):

(1.4) \[ Y_n = \sum_{s=0}^{\infty} U_{n,s}, \quad \text{where} \quad U_{n,s} = \mathbb{E}\{Y_n \mid \mathcal{F}_{n-s}\} - \mathbb{E}\{Y_n \mid \mathcal{F}_{n-s-1}\}, \]

where $\mathcal{F}_n := \sigma(\ldots, Z_{n-1}, Z_n)$ is the $\sigma$-field generated by the innovations in the “past” $\leq n$, $(X_n)$ denotes an $m$-dimensional process (MLP), and $G: \mathbb{R}^m \to \mathbb{R}$ stands for a real function such that $\mathbb{E}|Y_1|^Q < \infty$ for some $Q \geq 2$.

In Section 3, we will show how to use these moment bounds in order to obtain the Ghosh representation for $M$-estimation in the case of the short-range dependent observations. We will apply here Andrews and Pollard (1994) results (see Remark 3.2). The basic condition required for the proof of the Ghosh representation is the condition of the asymptotic stochastic equicontinuity (ASE). We obtain this condition from the Pisier result (see Lemma 3.2).

The Ghosh representation is especially useful for the proof of the asymptotic normality of $M$-estimators, provided the central limit theorem (CLT) holds for the sums $S_n$. Ho and Hsing (1997), Wu (2002), and Furmańczyk (2007) proved the CLT for the sums (0.2), where $(X_n)$ is a one-dimensional linear process or a multidimensional linear process, in the case of short-range dependence. In Subsection 3.1, we will prove Theorem 3.1 together with some conclusions from this theorem (see Propositions 3.1 and 3.2). At the end of Section 3, we will state a lemma about the consistence of $M$-estimators.

### 2. MOMENT INEQUALITIES

Let us put

\[ \|X\|_Q = \mathbb{E}^{1/Q}|X|^Q \]

and

\[ G_s(x) := \mathbb{E}G\left(\sum_{r=0}^{s-1} A_r Z_{k-r} + x\right) = \int G(z + x) dF_s(z), \]

where $F_s$ is the distribution function of $\sum_{r=0}^{s-1} A_r Z_{k-r}$. We will use the following condition:

(Lip) \[ |G_s(x) - G_s(y)| \leq \text{Lip}(G_s) |x - y| \]

for each $s = 1, 2, \ldots$ and for all $x, y \in \mathbb{R}^m$, where

\[ \sup_{s \geq 0} \text{Lip}(G_s) < C \quad \text{for some constant } C. \]
Theorem 2.1. Let $\|G(X_1)\|_Q < \infty$ for some $Q \geq 2$ and let $(X_n)$ be an $m$-dimensional linear process (MLP) satisfying $(a_0)$ and the assumption $(b_2(t))$ with $t = Q$. Then, if $(\text{Lip})$ holds, we have for every $n \in \mathbb{N}$

$$
(2.1) \quad \mathbb{E}\left| \sum_{j=1}^{n} (G(X_j) - \mathbb{E}G(X_j)) \right|^Q \leq C n^{Q/2}
$$

with $C = C_{A,Q} C(G)$, where

$$
C_{A,Q} = C_Q \left( \sum_{r=0}^{\infty} \|A_r\| \right)^Q, \quad C(G) = \left( \|G(X_1)\|_Q + \sup_{s \geq 1} \text{Lip}(G_s) \right)^Q,
$$

and $C_Q$ is a constant dependent on $Q$ and $\|Z_1\|_Q$.

For the proof of Theorem 2.1 we will need the following lemmas:

Lemma 2.1. Let $(X_n)$ be an $m$-dimensional linear process (MLP) and assume that $(\text{Lip})$ holds. Then

$$
(2.2) \quad |U_{k,s}| \leq \text{Lip}(G_s) \|A_s\| \left( 1 + |Z_{k-s}| \right) \text{ a.s. for } s = 1, 2, \ldots
$$

Proof. By the definition of $G_s$, we obtain

$$
U_{k,s} = \mathbb{E}\{G(X_k) \mid \mathcal{F}_{k-s}\} - \mathbb{E}\{G(X_k) \mid \mathcal{F}_{k-s-1}\} = G_s(R_{k,s}) - G_{s+1}(R_{k,s+1}),
$$

where $R_{k,s} = X_k - \sum_{r=0}^{s-1} A_r Z_{k-r}$. From independence of $A_s Z_{k-s}$ and $R_{k,s+1}$ we have

$$
G_{s+1}(x) = \int G_s(x+z) dF_{1,s}(z)
$$

and

$$
G_{s+1}(R_{k,s+1}) = \int G_s(R_{k,s+1} + z) dF_{1,s}(z),
$$

where $F_{1,s}$ is the distribution function of $A_s Z_{k-s}$. Hence, applying the condition (Lip), we get

$$
|U_{k,s}| \leq \int |G_s(R_{k,s}) - G_s(R_{k,s+1} + z)| dF_{1,s}(z)
\leq \text{Lip}(G_s) \int |R_{k,s} - R_{k,s+1} + z| dF_{1,s}(z)
\leq \text{Lip}(G_s) \left( |R_{k,s} - R_{k,s+1}| + \int |z| dF_{1,s}(z) \right)
\leq \text{Lip}(G_s) \left( |A_s Z_{k-s}| + \mathbb{E}|A_s Z_{k-s}| \right).
$$

From (1.1) and the fact that $\mathbb{E}|Z_1|^2 = 1$ we have

$$
(2.3) \quad |A_s Z_{k-s}| \leq \|A_s\| |Z_{k-s}|
$$

and

$$
(2.4) \quad \mathbb{E}|A_s Z_{k-s}| \leq \|A_s\|.
$$

The relations (2.3) and (2.4) imply the desired result (2.2).
LEMMA 2.2 (Burkholder (1966)). Let \( (Y_n, F_n) \) be a stationary sequence of the martingale differences such that \( \mathbb{E}|Y_1|^Q < \infty \) for some \( Q \geq 2 \). Then there exists a universal constant \( C'_Q \), dependent on \( Q \), such that

\[
\mathbb{E}\left| \sum_{j=1}^{n} Y_j \right|^Q \leq C'_Q n^{Q/2} \mathbb{E}|Y_1|^Q \quad \text{for every } n \in \mathbb{N}.
\]

Proof. Notice that from the Burkholder inequality (see Stout (1974), 3.3.14) we have

\[
\mathbb{E}\left| \sum_{j=1}^{n} Y_j \right|^Q \leq C'_Q \mathbb{E}\left| \sum_{j=1}^{n} Y_j^2 \right|^{Q/2}.
\]

In addition, by the Hölder inequality, we get

\[
\sum_{j=1}^{n} 1 \cdot Y_j^2 \leq \left( \sum_{j=1}^{n} |Y_j|^Q \right)^{2/Q} n^{1-2/Q} \quad \text{for } Q > 2.
\]

This together with (2.6) imply (2.5). ■

We now prove Theorem 2.1, the main result of our paper.

Proof of Theorem 2.1. We first use the martingale representation of the form (1.4) for the transformed sequence \( (Y_n) \):

\[
\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} \sum_{s=0}^{\infty} U_{k,s} = \sum_{s=0}^{\infty} W_{n,s},
\]

where \( W_{n,s} := \sum_{k=1}^{n} U_{k,s} \). Since for every \( s \geq 0 \) the sequence \( (U_{k,s}, F_{k-s})_k \) is a stationary sequence of the martingale differences, \( (W_{m,s}, F_{m-s})_m \) is a martingale. Applying the Burkholder inequality (see Lemma 2.2), we obtain

\[
\mathbb{E}|W_{n,s}|^Q \leq C'_Q n^{Q/2} \mathbb{E}|U_{1,s}|^Q.
\]

Moreover, it follows from Lemma 2.1 that

\[
|U_{1,s}| \leq \text{Lip}(G_s) \|A_s\| (1 + |Z_{1-s}|) \quad \text{for } s = 1, 2, \ldots
\]

and that

\[
\|W_{n,0}\|_Q \leq (C_Q)'^{1/Q} n^{1/2} \|U_{1,0}\|_Q \leq (C_Q)'^{1/Q} n^{1/2} \|G(X_1)\|_Q.
\]

Put \( S_n = \sum_{k=1}^{n} Y_k \). Then

\[
\|S_n\|_Q \leq \|W_{n,0}\|_Q + \sum_{s=1}^{\infty} \|W_{n,s}\|_Q \leq C_Q^{1/Q} n^{1/2} (\|G(X_1)\|_Q + \sum_{s=1}^{\infty} \|A_s\| \text{Lip}(G_s))
\]

for some constant \( C_Q \) dependent on \( Q \) and \( \|Z_1\|_Q \), which implies the desired result (2.1). ■
With reference to Theorem 2.1 and Lemma 2.1 the following property is worth proving.

**Proposition 2.1.** The following statements are true:

(i) If $G$ is Lipschitz, then (Lip) is satisfied with the constant $\text{Lip}(G_s)$ bounded as follows:

$$\text{Lip}(G_s) \leq \text{Lip}(G).$$

(ii) If $G$ is integrable, i.e. $G \in L^1(\mathbb{R}^m)$ and (b1) holds, then (Lip) is satisfied with the constant $\text{Lip}(G_s)$ bounded as follows:

$$\text{Lip}(G_s) \leq \text{Lip}(f_1) \int |G(s)| \, ds,$$

where $f_1$ is the density of $Z_1$.

**Proof.** The condition (i) is clear. In order to prove (ii), notice that

$$|G_s(x) - G_s(y)| = \left| \int G(z + x)f_s(z)dz - \int G(z + y)f_s(z)dz \right|,$$

where $f_s$ means the density of the random vector $\sum_{r=0}^{s-1} A_r Z_{k-r}$. Since $f_s = g_s * f_1$ is the convolution of $g_s$ and $f_1$, where $g_s$ is the density of $\sum_{r=1}^{s-1} A_r Z_{k-r}$, and $f_1$ is the density of $Z_1$, by (b1) we obtain

$$\left| \int G(z + x)f_s(z)dz - \int G(z + y)f_s(z)dz \right|$$

$$= \left| \int G(z)(f_s(z - x) - f_s(z - y))dz \right|$$

$$= \left| \int G(z)(g_s * f_1(z - x) - g_s * f_1(z - y))dz \right|$$

$$\leq |x - y| \text{Lip}(f_1) \int |G(z)| \, dz. \quad \blacksquare$$

3. AN APPLICATION TO $M$-ESTIMATION

Let $X_1, X_2, \ldots$ be an $m$-dimensional process (MLP) and let $P_0$, where $\theta \in \Theta \subset \mathbb{R}^k$, denote the marginal distribution of $X_1$. Let, in addition, $\psi: \Theta \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ be a map and

$$\Psi(\theta) := \mathbb{E}_\psi(\theta, X_1) = \int \psi(\theta, x) dP_0(x).$$

Moreover, we assume that $\Psi(\theta_0) = 0$ for some $\theta_0 \in \text{int}(\Theta)$. Consider the $M$-estimator $\theta_n$ of a parameter $\theta_0$ as a random function of the form

$$\Psi_n(\theta_n) = o_P(n^{-1/2}),$$

where

$$\Psi_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} \psi(\theta, X_i).$$
Moreover, define an empirical process

\( v_n \psi(\theta, \cdot) := \sqrt{n} \left( \Psi_n(\theta) - \Psi(\theta) \right) \)

indexed by functions \( \psi \). The following conditions will be used in this section:

\( \text{(C)} \quad |\theta_n - \theta_0| \xrightarrow{P} 0 \quad \text{as} \ n \to \infty, \)

where \( P \to \) denotes the convergence in probability \( P \).

\( \text{(V)} \quad \Psi(\theta) = V(\theta - \theta_0) + o(|\theta - \theta_0|) \quad \text{as} \ \theta \to \theta_0, \)

where \( V \) is certain nonsingular matrix.

\( \text{(B)} \quad v_n \psi(\theta_0, \cdot) = O_P(1). \)

\( \text{(ASE)} \) For all \( \varepsilon > 0 \) and for every \( \eta > 0 \) there exists \( \delta > 0 \) such that

\( \limsup_{n \to \infty} \mathbb{P} \left( \sup_{|\theta - \theta_0| < \delta} |v_n \psi(\theta, \cdot) - v_n \psi(\theta_0, \cdot)| > \eta \right) < \varepsilon. \)

**Remark 3.1.** Condition (B) holds if \( v_n \psi(\theta_0, \cdot) \xrightarrow{d} N_k(0, \Sigma) \) as \( n \to \infty \).

**Remark 3.2.** Under the conditions (C), (V), (B), and (ASE) Andrews and Pollard (1994) proved the Ghosh representation for \( M \)-estimators provided the observations \( (X_n) \) satisfy the strong mixing condition, i.e.

\( \sqrt{n}(\theta_n - \theta_0) = -V^{-1}v_n \psi(\theta_0, \cdot) + o_P(1). \)

The condition (ASE) (asymptotic stochastic equicontinuity) plays the key role in our applications. Let \( \Theta_\delta = \{\theta \in \Theta : |\theta - \theta_0| \leq \delta\} \) for some \( \delta > 0 \).

Assume that the increments of the empirical process satisfy the following condition:

\( \text{(d}_Q\text{)} \) For all \( n \geq n_0 \) and \( \theta_1, \theta_2 \in \Theta_\delta, \)

\( \|v_n \psi(\theta_1, \cdot) - v_n \psi(\theta_2, \cdot)\|_Q \leq C |\theta_1 - \theta_2|, \)

where \( C \) is some positive constant.

**Lemma 3.1.** The condition \( \text{(d}_Q\text{)} \) for \( Q = k + 1 \) implies (ASE).

We first prove some auxiliary lemma. Denote by \( N(C, d, r) \) the covering number as the smallest number of closed balls with radius less than or equal to \( r \), whose union covers the set \( C \), where the metric space \( (C, d) \) is totally bounded.

**Lemma 3.2.** Let \( X = (X_t)_{t \in T} \) be a stochastic process in \( \mathbb{R}^k \), indexed by the totally bounded metric space \( (T, d) \). Assume that the process \( X \) has continuous sample paths, its increments are such that
(i) for all \( t, s \in T \) the inequality

\[
\|X_t - X_s\|_Q \leq C d(t, s)
\]

is satisfied, where \( C \) is some constant independent of \( t \) and \( s \), and

(ii) the following condition holds:

\[
\int_{0}^{\delta} \sqrt{N(T, d, r)} dr < \infty.
\]  

Then

\[
E \sup_{t \in T} |X_t| \leq E |X_{t_0}| + 8C \int_{0}^{\delta} \sqrt{N(T, d, r)} dr,
\]

where \( t_0 \) is some point of \( T \) and \( \delta = \sup_{t \in T} d(t, t_0) \).

Proof of Lemma 3.2. By Theorem 11.1 of Ledoux and Talagrand (1991), putting the Young function \( \psi(x) := x^Q \), we have

\[
E \sup_{s, t \in T} |X_s - X_t| \leq 8C \int_{0}^{\delta} \sqrt{N(T, d, r)} dr.
\]

We also have, for every \( t_0 \) in \( T \),

\[
E \sup_{t \in T} |X_t| \leq E |X_{t_0}| + E \sup_{s, t \in T} |X_s - X_t|.
\]

Hence we obtain (3.6). ■

Proof of Lemma 3.1. Since the metric space \((\Theta_\delta, d)\) is totally bounded, where \( d(\theta_1, \theta_2) = |\theta_1 - \theta_2| \), putting

\[
t_0 = \theta_0, \quad t = \theta, \quad T = \Theta_\delta \quad \text{and} \quad X_t := v_n \psi(t, \cdot) - v_n \psi(t_0, \cdot)
\]

we see that if \( \dim(\Theta_\delta) = k, \) then \( N(T, d, r) = O(r^{-k}) \). Consequently, assumption (ii) of Lemma 3.2 holds for \( Q = k + 1 \). Condition \((d_Q)\) for \( Q = k + 1 \) implies that for all \( n \geq n_0 \) and every \( t, s \in \Theta_\delta \)

\[
\|X_t - X_s\|_{k+1} = \|v_n \psi(t, \cdot) - v_n \psi(s, \cdot)\|_{k+1} \leq C |t - s|
\]

for some constant \( C \). Thus, assumption (i) of Lemma 3.2 is fulfilled for all \( n \geq n_0 \).

Since \( X_{t_0} \equiv 0 \), by (3.6) we infer that for all \( n \geq n_0 \)

\[
E \sup_{t \in T} |X_t| = E \sup_{d(\theta, \theta_0) < \delta} |v_n \psi(\theta, \cdot) - v_n \psi(\theta_0, \cdot)|
\]

\[
\leq 8C \int_{0}^{\delta} N^{1/(k+1)}(\Theta_\delta, d, r) dr.
\]
Hence
\[
\limsup_{n \to \infty} \mathbb{E} \sup_{d(\theta, \theta_0) < \delta} |\psi_n(\theta, \cdot) - \psi_n(\theta_0, \cdot)| \leq C \int_0^\delta N^{1/(k+1)}(\Theta_\delta, d, r) dr.
\]

Therefore, it follows that for any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
\limsup_{n \to \infty} \mathbb{E} \sup_{d(\theta, \theta_0) < \delta} |\psi_n(\theta, \cdot) - \psi_n(\theta_0, \cdot)| < \varepsilon,
\]
which yields the desired condition (ASE).

\section{Asymptotic properties of \( M \)-estimators.} Let us introduce the following notation:
\[
\psi_{\theta_1, \theta_2}(x) := \psi(\theta_1, x) - \psi(\theta_2, x),
\]
\[
\Psi_s(\theta, x) := \mathbb{E} \psi(\theta, \sum_{r=0}^{s-1} A_r Z_1 - r + x),
\]
\[
\Psi_s(x, \theta_1, \theta_2) := \Psi_s(\theta_1, x) - \Psi_s(\theta_2, x).
\]

In our further considerations we will apply the following conditions:
\begin{enumerate}
\item[(L_Q)] There exists a constant \( L \) such that
\[
\|\psi_{\theta_1, \theta_2}(X_1)\|_Q \leq L |\theta_1 - \theta_2| \quad \text{for all } \theta_1, \theta_2 \in \Theta_\delta.
\]
\item[(Lip_1)] There exists some constant \( C \) such that
\[
\sup_{s \geq 1} |\Psi_s(x, \theta_1, \theta_2) - \Psi_s(y, \theta_1, \theta_2)| \leq C |\theta_1 - \theta_2| |x - y|
\]
for all \( \theta_1, \theta_2 \in \Theta_\delta \) and all \( x, y \in \mathbb{R}^m \).
\end{enumerate}

\begin{theorem}
Suppose that \( \mathbb{E}|\psi(\theta, X_1)|^{k+1} < \infty \) for all \( \theta \in \Theta_\delta \), the conditions (B), (C), (V), (a_0), (b_2(t)) with \( t = k + 1 \) are satisfied. Assume also that (L_{k+1}) and (Lip_1) are fulfilled. Then we obtain the Ghosh representation (3.3). Additionally, if \( \Psi_s(\theta_0, \cdot) \) is Lipschitz, then
\[
\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} \mathcal{N}_k(0, \Sigma^{-1}(\mathbf{V}^{-1})^T),
\]
where the elements of matrix \( \Sigma \) have the form
\[
\sigma_{i,j} = \mathbb{E}(\psi_i(\theta_0, X_1)\psi_j(\theta_0, X_1)) + 2 \sum_{t=1}^\infty \mathbb{E}(\psi_i(\theta_0, X_1)\psi_j(\theta_0, X_{1+t}))
\]
and \( \psi = (\psi_1, \ldots, \psi_k) \).
In order to get the Ghosh representation, by applying our assumptions, we only need to show that \((d_Q)\) holds for \(Q = k + 1\) (see Remark 3.2 and Lemma 3.1). It follows from Theorem 2.1 that

\[
\|v_n \psi(\theta_1, \cdot) - v_n \psi(\theta_2, \cdot)\|_{k+1} \leq C' \left( \|\psi_{\theta_1, \theta_2}(X_1)\|_{k+1} + \sup_{s \geq 1} \text{Lip} \left( \Psi_s(\cdot, \theta_1, \theta_2) \right) \right),
\]

where \(C'\) is some constant independent of \(\psi, \theta_1, \theta_2, n\). Using \((L_{k+1})\) and \((\text{Lip}_1)\) we state that the condition \((d_Q)\) for \(Q = k + 1\) holds.

The property of asymptotic normality of \(\theta_n\) can be obtained from Furmanczyk (2007), Theorem 2.1. ■

**Remark 3.3.** Notice that if \(\|A_r\| \sim r^{-\alpha} L(r)\) for some \(\alpha > 1\), where \(L(\cdot)\) is some slowly varying function, and if \(\Psi_s(\theta_1, \cdot)\) is the Lipschitz function for all sufficiently large \(s\), then \(\theta_n\) is asymptotically normal (see Wu (2002), Corollary 2).

Denote by \(D_0\) and \(D_2\) the derivative of some function with respect to \(\theta\) and \(x\), respectively. Now we give the conditions which imply \((\text{Lip}_1)\).

**Proposition 3.1.** The condition \((\text{Lip}_1)\) holds if at least one of the following conditions is satisfied:

(i) \(\sup_{\theta \in \Theta_\delta} \sup_{x, s} \|D_{0} D_{2} \Psi_s(\theta, x)\| \leq C\) for some constant \(C\);

(ii) \(\sup_{\theta \in \Theta_\delta} \sup_{x} \|D_{0} D_{x} \Psi(\theta, x)\| \leq C\) for some constant \(C\);

(iii) there exists a function \(L \in L^1(R^m)\) such that

\[
|\psi_{\theta_1, \theta_2}(x)| \leq L(x) |\theta_1 - \theta_2| \quad \text{for all} \quad \theta_1, \theta_2 \in \Theta_\delta \quad \text{and} \quad x \in R^m
\]

and \((b_1)\) holds.

**Proof.** It is easily seen that from (i) and the theorem about the average value \((\text{Lip}_1)\) holds. Similarly, it follows from (ii) that

\[
|\psi_{\theta_1, \theta_2}(x) - \psi_{\theta_1, \theta_2}(y)| \leq C |\theta_1 - \theta_2| |x - y|,
\]

which immediately gives \((\text{Lip}_1)\). Now, let us consider the condition (iii). We have

\[
\Psi_s(x, \theta_1, \theta_2) - \Psi_s(y, \theta_1, \theta_2)
\]

\[
= E \psi_{\theta_1, \theta_2} \left( \sum_{r=0}^{s-1} A_r Z_{1-r} + x \right) - E \psi_{\theta_1, \theta_2} \left( \sum_{r=0}^{s-1} A_r Z_{1-r} + y \right)
\]

\[
= \int \psi_{\theta_1, \theta_2}(z) f_s(z - x) dz - \int \psi_{\theta_1, \theta_2}(z) f_s(z - y) dz
\]

\[
= \int \psi_{\theta_1, \theta_2}(z) (g_s * f_1(z - x) - g_s * f_1(z - y)) dz,
\]

where \(f_s\) is the density of \(\sum_{r=0}^{s-1} A_r Z_{1-r}\), \(g_s\) is the density of \(\sum_{r=1}^{s-1} A_r Z_{1-r}\), and \(f_1\) is the density of \(Z_1\). Since \(f_s\) is the convolution of \(g_s\) and \(f_1\), by condition \((b_1)\)
Linear processes with application to M-estimation

the function \( f_1 \) is Lipschitz, and

\[
|\Psi_s(x, \theta_1, \theta_2) - \Psi_s(y, \theta_1, \theta_2)| \leq \text{Lip}(f_1) |x - y| \int |\psi_{\theta_1, \theta_2}(z)| \, dz
\]

\[
\leq \text{Lip}(f_1) |x - y| |\theta_1 - \theta_2| \int L(z) \, dz,
\]

which implies (Lip1). ■

We will prove the following result for a one-dimensional case of linear processes where \( m = 1 \).

**Proposition 3.2.** Suppose that \( E |\psi(\theta, X_1)|^2 < \infty \) for all \( \theta \in \Theta \subset \mathbb{R} \), conditions (B), (C), (V) hold, and \((a_0), (b_1), (b_2(t))\) with \( t = 2 \) are fulfilled for \( m = 1 \). Assume in addition that (L2) holds and

(i) there exists a constant \( C \) such that

\[
\|\psi_{\theta_1, \theta_2}(\cdot)\|_{t_v} \leq C |\theta_1 - \theta_2| \quad \text{for all } \theta_1, \theta_2 \in \Theta,
\]

where \( \|\cdot\|_{t_v} \) denotes the total variation, and

(b0) there exists some constant \( C \) such that the condition

\[
\sup_{s \in \mathbb{N}} \text{Lip}(F_s) \leq C
\]

holds, where \( F_s \) is the distribution function of \( \sum_{r=0}^{s-1} a_r Z_{1-r} \).

Then we obtain the Ghosh representation (3.3). Additionally, if \( \Psi_s(\theta_0, \cdot) \) is Lipschitz, then

\[
\sqrt{n}(\theta_n - \theta_0) \overset{d}{\to} \mathcal{N}(0, \gamma^{-2} \sigma^2),
\]

where \( \gamma = \gamma(\theta_0), \gamma(\theta) = D_\theta \Psi(\theta) \), and

\[
\sigma^2 = E\psi^2(\theta_0, X_1) + 2 \sum_{r=1}^\infty E(\psi(\theta_0, X_1)\psi(\theta_0, X_{1+r})).
\]

**Proof.** First, we will show that conditions (i) and (b0) imply (Lip1). Observe that for any \( x, y \)

\[
\psi_{\theta_1, \theta_2}(z)F_s(z - x) \bigg|_{z=-\infty}^{z=\infty} - \psi_{\theta_1, \theta_2}(z)F_s(z - y) \bigg|_{z=-\infty}^{z=\infty} = 0.
\]

By the formula of integration by parts for Stieltjes integrals, we have

\[
\Psi_s(x, \theta_1, \theta_2) - \Psi_s(y, \theta_1, \theta_2)
\]

\[
= \int \psi_{\theta_1, \theta_2}(z) dF_s(z - x) - \int \psi_{\theta_1, \theta_2}(z) dF_s(z - y)
\]

\[
= - \int (F_s(z - x) - F_s(z - y)) d\psi_{\theta_1, \theta_2}(z),
\]
where $F_s$ is the distribution function of $\sum_{r=0}^{s-1} A_r Z_{1-r}$. Therefore, from (b$_0$) we get $\text{Lip}(F_s) \leq C$ and

$$|\Psi_s(x, \theta_1, \theta_2) - \Psi_s(y, \theta_1, \theta_2)| \leq \text{Lip}(F_s) |x - y| \|\psi_{\theta_1, \theta_2}(\cdot)\|_{tv} \leq C |x - y| |\theta_1 - \theta_2|.$$ 

By Theorem 3.1 we obtain our assertion. ■

Some remarks on the consistence of $M$-estimators

**Lemma 3.3.** Suppose the conditions (a$_0$), (b$_1$) are satisfied, and (b$_2(t)$) holds with $t = k + 1$. Assume also that (Lip$_1$) is fulfilled. Then for each $\delta > 0$

$$\sup_{\theta \in \Theta_\delta} |\Psi_n(\theta) - \Psi(\theta)| \xrightarrow{P} 0.$$

**Proof of Lemma 3.3.** Obviously, we have

$$\sup_{\theta \in \Theta_\delta} |\Psi_n(\theta) - \Psi(\theta)| \leq \frac{1}{\sqrt{n}} |\nu_n \psi(\theta_0, \cdot)| + \frac{1}{\sqrt{n}} \sup_{d(\theta, \theta_0) < \delta} |\nu_n \psi(\theta, \cdot) - \nu_n \psi(\theta_0, \cdot)|.$$

Hence, reasoning as in the proof of Lemma 3.1 (see (3.7)), we have

$$\mathbb{E} \sup_{d(\theta, \theta_0) < \delta} |\nu_n \psi(\theta, \cdot) - \nu_n \psi(\theta_0, \cdot)| = O(1),$$

and the proof is completed. ■

**Remark 3.4.** One can be seen that if all the assumptions of Lemma 3.3 are satisfied and

$$\inf_{\theta \in \Theta \setminus \Theta_\delta} |\Psi(\theta)| > 0 \quad \text{for every } \delta > 0,$$

then condition (C) is fulfilled (see Van der Vaart (1998), Theorem 5.9).

**Acknowledgements.** The paper is a part of the author’s Ph.D. Thesis written under the supervision of Professor Wojciech Niemiro at Warsaw University. I would like to thank him for his help. I am also grateful to Professor Jan Mielniczuk for his comments and remarks.

**References**


Linear processes with application to $M$-estimation


Department of Applied Mathematics
Warsaw University of Life Sciences (SGGW)
ul. Nowoursynowska 159
02-776 Warszawa, Poland
E-mail: konfur@wp.pl

Received on 27.12.2006;
revised version on 30.9.2007