Advances in Science, Technology and Environmentology
Special Issue on the "Financial & Pension Mathematical Science"
Vol. B10 (2014.3), 55–67

ASYMPTOTICS FOR M-ESTIMATORS IN TIME SERIES

By Yan LIU

Waseda University

It has been a half of a century since Hodges and Lehmann proposed a Lemma on the asymptotic normality for the estimation of location in 1963 before Huber, who changed the idea into the concept of M-estimation. In this paper, we give conditions which guarantee the asymptotic normality of M-estimator based on the observations from time series models by minimizing some convex objective functions. We do not assume the differentiability around the true parameter of the objective functions. The results are extended to M_m estimates. Examples with new aspects of the main results are also provided.

1. Introduction. Hodges and Lehmann (1963) proposed a Lemma on the asymptotic normality for the estimation of location of i.i.d. samples. Huber (1964) generalized the idea into M-estimation concept and investigated the robustness of the estimators.

In time series settings, the second order stationarity is usually considered. Under the conditions, the process $\{X_t\}$ has its own spectral density $f(\lambda)$ in the frequency domain. The parameter estimation is based on the periodogram

$$I_n(\lambda) = \frac{1}{2\pi n}\Big\{\sum_{t=1}^{n} X_t e^{it\lambda}\Big\}\Big\{\sum_{t=1}^{n} X_t e^{it\lambda}\Big\}^*, \tag{1.1}$$

which is defined on the observed stretch $\{X_1, \dots, X_n\}$. Then the Whittle estimator is defined as the minimizer of

$$D(f_\theta, I_n) = \int_{-\pi}^{\pi} [\log\det\{f_\theta(\lambda)\}] + \mathrm{tr}\{I_n(\lambda) f_\theta(\lambda)^{-1}\} d\lambda. \tag{1.2}$$

The asymptotic normality of the Whittle estimator under regular conditions was shown in Hosoya and Taniguchi (1982).

In this paper, we suppose the process$\{X_t\}$ has a unique one-sided autoregressive representation in time domain as

$$\sum_{j=0}^{\infty} b_j (X_{t-j} - \mu) = \epsilon_t, \tag{1.3}$$

with $b_0 = 1$. The process may be second order stationarity, with some "nice" structures for dependent data, such as ergodicity and mixing conditions. For example, the asymptotics of the process under mixing conditions is well investigated in Ibragimov and Linnik (1971).

Even if the the process is second order stationary nonlinear, the process can be decomposed into a linear part and a deterministic part (Wold's decomposition), and in turn it has AR(∞) representation since the process is invertible (See Brockwell and Davis (1991), p.90).

AMS 2000 subject classifications: Primary 62M10; Secondary 62F12, 62F10
Keywords and phrases: linear model, M estimation, M_m estimator, least absolute deviation estimation

For M-estimation, the problem may be generalized as follows. We assume that the AR(∞) representation of the model is characterized by a finite dimensional vector $\theta = (\theta_1, \eta)$. That is to say, the model is defined by

$$\sum_{j=0}^{\infty} b_j(\eta) X_{t-j} = \epsilon_t, \tag{1.4}$$

where $b_0 = 1$ and $\{\epsilon_t\}$ is i.i.d. $(0, \theta_1)$. The parameter θ_1 denotes the scale parameter of the model, and the remaining parameter $(\theta_2, \ldots, \theta_m)$ is denoted by η. The true parameter is represented by θ^0.

Define the residual process $e_t(\eta)$ and its corresponding form in the following way:

$$e_t(\eta) = \sum_{j=0}^{t-1} b_j(\eta) X_{t-j}, \qquad r_t(\theta) = \theta_1^{-1/2} e_t(\eta), \tag{1.5}$$

$$\epsilon_t(\eta) = \sum_{j=0}^{\infty} b_j(\eta) X_{t-j}, \qquad v_t(\theta) = \theta_1^{-1/2} \epsilon_t(\eta). \tag{1.6}$$

The approximate maximum likelihood estimator is equivalent to finding the solution of

$$\sum_{t=2}^{n} r_t(\theta) \dot{r}_t(\theta) - c(\theta) = 0,$$

which is considered in Beran (1994). They extend the estimator to a class of Z-estimators and investigate the robustness under Gaussian long-memory situation.

We look into the properties of M-estimators in both non-differentiable and differentiable cases of the objective functions. Even we showed the asymptotic normality of M-estimator with non-differentiable objective function under some regular conditions, the asymptotic variance matrix of the estimator is not explicitly obtained. For this reason, we give the asymptotic variance matrix of the estimator under the differentiable conditions. The objective function is considered to be convex, so the corresponding M-estimator includes LAD estimators (see Niemiro (1992)). We also extend the result to M_m estimator since the class is much richer. The class includes Oja's median and even Hodges-Lehmann's estimators of location (see Bose (1998)). The proof is similar to the method for U statistics in depend case, which is in order examined by Hoeffding (1948), Sen (1972), Yoshihara (1976) and Denker and Keller (1983).

The paper is organized as follows. In section 2, we review the sufficient conditions for asymptotic normality of M-estimators. Without the condition of differentiability for the convex objective function, we derive asymptotic normality of M-estimators in time series settings under the new class of conditions. Also, the asymptotic result is given in the detailed way if the objective function is differentiable in Section 4. In section 5, we extend the result to M_m estimators. Section 6 contains two important cases of the inference in time series analysis as examples of the main result.

2. Asymptotic Normality of M-estimators. First we revisit the work of Niemiro (1992) and Hodges and Lehmann (1963) in this section.

Suppose ϵ, ϵ_1, ..., ϵ_n are i.i.d. random variables. Let $\rho(\theta, \epsilon)$ be a real function defined for $\theta \in \mathbb{R}^m$ and $g(\theta, \epsilon)$ be a subgradient of $\rho(\theta, \epsilon)$. Define

$$Q(\theta) = E\rho(\theta, \epsilon). \tag{2.1}$$

The empirical analog of $Q(\theta)$ is defined by

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho(\theta, \epsilon_i). \tag{2.2}$$

Denote the corresponding score function $\xi_n(\theta)$. Usually, $\xi_n(\theta)$ is considered as $\nabla Q_n(\theta)$, where ∇ is an operation which means the differentiation with respect to θ. Set the ith element $\hat{\theta}_{ni}$ of minimizer $\hat{\theta}_n$ satisfying

$$\hat{\theta}_{ni} = \alpha\theta^*_{ni} + (1-\alpha)\theta^{**}_{ni} \tag{2.3}$$

for any $0 \le \alpha \le 1$, where

$$\begin{aligned} \theta^*_{ni} &= \sup\{r : \xi_{ni}(r) \ge 0\}, \\ \theta^{**}_{ni} &= \inf\{r : \xi_{ni}(r) \le 0\}. \end{aligned}$$

Assumption 2.1 (Niemiro (1992)).

(i) $\rho(\theta, \epsilon)$ is convex with respect to θ for each fixed ϵ.
(ii) $Q(\theta)$ is well defined, that is, the expectation exists and is finite for all θ.
(iii) θ^0 satisfying $Q(\theta^0) = \min_\theta Q(\theta)$ exists and is unique.
(iv) $E|g(\theta, \epsilon)|^2 < \infty$ for each θ in a neighborhood of θ^0.
(v) $Q(\theta)$ is twice differentiable at θ^0 and $\nabla^2 Q(\theta^0)$ is positive definite.

Assumption 2.2 (Hodges and Lehmann (1963), Inagaki and Kondo (1980)).

(i) $\xi_n(\theta)$ is a non-decreasing function of every element of θ.
(ii) For any vector valued u,

$$\sqrt{n}(\xi_n(\theta^0 + u/\sqrt{n}) - \xi_n(\theta^0)) \xrightarrow{\mathcal{P}} H'u,$$

where H is a positive definite matrix.
(iii) $\sqrt{n}\xi_n(\theta^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$.

Lemma 2.3. *Under each of Assumption 2.1 or Assumption 2.2, Then*

$$\sqrt{n}(\hat{\theta}_n - \theta^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, H^{-1}VH^{-1}), \tag{2.4}$$

where $H = \nabla^2 Q(\theta^0)$ *and* $V = \operatorname{Var} g(\theta^0)$.

Proposition 2.4. *If $\rho(\theta, \epsilon)$ is differentiable with respect to θ in a neighborhood of θ^0, then Assumption 2.1 implies Assumption 2.2.*

Proof. From Assumption 2.1.(i) and (ii), we see $\xi_n(\theta)$ is a non-decreasing function of every element of θ. Next, according to Niemiro's proof, we see for each θ in a neighborhood of θ^0,

$$Q_n(\theta + \frac{u}{\sqrt{n}}) - Q_n(\theta) - \frac{u'}{\sqrt{n}}\nabla Q_n(\theta) - Q(\frac{u}{\sqrt{n}}, \theta) \xrightarrow{\mathcal{P}} 0. \tag{2.5}$$

Differentiate the equation with respect to θ above and substitute $\theta = \theta^0$,

$$\xi_n(\theta^0 + \frac{u}{\sqrt{n}}) - \xi_n(\theta^0) - \frac{u'}{\sqrt{n}}\nabla^2 Q_n(\theta^0) \xrightarrow{\mathcal{P}} 0.$$

From Assumption 2.1.(v),

$$\nabla^2 Q_n(\theta^0) \xrightarrow{\mathcal{P}} \nabla^2 Q(\theta^0),$$

and by Theorem 4.1 of Billingsley (1968), we obtain the desirable result:

$$\sqrt{n}(\xi_n(\theta^0 + u/\sqrt{n}) - \xi_n(\theta^0)) \xrightarrow{\mathcal{P}} \nabla^2 Q(\theta^0)'u.$$

Under Assumption 2.1.(iv), we have the last result,

$$\sqrt{n}\xi_n(\theta^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

□

Remark 2.5. We mainly show the result by (2.5) in the subsequent section. It is sufficient for the asymptotics. (See Niemiro (1992).)

3. Estimation of Parameters in Linear Time Series Models. For time series model, we consider $\rho(\theta_1^{-1/2}, e_t(\eta))$, which sometimes is written as $\rho(\theta)$ for short. In this section, we only assume $\rho(\theta)$ is convex with respect to each parameter. Generally, consider $\theta = (\theta_1, \eta)$ in a compact set $\Theta \subset \mathbb{R}^m$, where $\eta = (\theta_2, \ldots, \theta_m)$. For simplicity, we write

$$Q(\theta) = E\rho(\theta_1^{-1/2}, e_t(\eta)). \tag{3.1}$$

The true value of θ, represented by θ^0 is defined by

$$Q(\theta^0) = \min_{\theta\in\Theta} Q(\theta). \tag{3.2}$$

The sample version corresponding to the objective function is

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \rho(\theta_1^{-1/2}, e_t(\eta)), \tag{3.3}$$

and

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta). \tag{3.4}$$

ASSUMPTION 3.1.

(i) $\rho(\theta)$ is convex with respect to θ.
(ii) $Q(\theta)$ is well defined.
(iii) θ^0 satisfying (3.1) exists and is unique.

Since $\rho(\theta)$ is a convex function, there exists a subgradient of $\rho(\theta)$, which is denoted by $g(\theta)$ satisfying

$$\rho(\alpha) + (\beta - \alpha)' g(\alpha) \leq \rho(\beta) \tag{3.5}$$

hold for all $\alpha, \beta \in \mathbb{R}^d$. Without loss of generality, we consider the case $\theta^0 = 0$ and $Q(0) = 0$. Then

$$\alpha' g(0) \leq \rho(\alpha) - \rho(0) \leq \alpha' g(\alpha), \tag{3.6}$$

$$0 \leq \rho(\alpha) - \rho(0) - \alpha' g(0) \leq \alpha'(g(\alpha) - g(0)). \tag{3.7}$$

For $\alpha = (\alpha_1, \alpha_2)$, we use the following symbols for the simplicity of the notation:

$$\begin{aligned} \rho(\frac{\alpha}{\sqrt{n}})_t &:= \rho((\theta_1^0 + \frac{\alpha_1}{\sqrt{n}})^{-1/2}, e_t(\eta^0 + \frac{\alpha_2}{\sqrt{n}})), \\ \rho(0)_t &:= \rho((\theta_1^0)^{-1/2}, e_t(\eta^0)) = \rho((\theta_1^0)^{-1/2}, \epsilon). \end{aligned}$$

$g(\cdot)_t$ is also defined in the same way. At neighborhood of θ^0, we have for each t,

$$0 \leq \rho(\frac{\alpha}{\sqrt{n}})_t - \rho(0)_t - \frac{\alpha'}{\sqrt{n}} g(0)_t \leq \frac{\alpha'}{\sqrt{n}} (g(\frac{\alpha}{\sqrt{n}})_t - g(0)_t). \tag{3.8}$$

ASSUMPTION 3.2.

(i) For all diagonal elements (i,i),

$$\sum_{k=1}^{\infty} E|(g(\frac{\alpha}{\sqrt{n}})_1 - g(0)_1)_i (g(\frac{\alpha}{\sqrt{n}})_k - g(0)_k)_i| < \infty.$$

(ii) $Q(\theta)$ is twice differentiable at θ^0 and $\nabla^2 Q(\theta^0)$ is positive definite.

REMARK 3.3. Assumption 3.2.(i) is a condition to control the correlation of the score.

THEOREM 3.4. *Let $\hat{\theta}_n$ be defined in (3.4) under Assumptions 3.1 and 3.2. Then we obtain*

(i) *$\hat{\theta}_n$ converges to the true value θ^0 in probability as $n \to \infty$.*
(ii) *$\sqrt{n}(\hat{\theta}_n - \theta^0)$ have a joint asymptotic normal distribution whose mean is 0 and the asymptotic covariance matrix is give by $H^{-1} V H^{-1}$, where*

$$\begin{aligned} H &= \nabla^2 Q(\theta^0) \\ V &= \text{Var } g(\theta^0). \end{aligned}$$

PROOF. First, we show (2.5) holds in this case. For fixed α, define Y_{nt} as

$$Y_{nt} = \rho(\frac{\alpha}{\sqrt{n}})_t - \rho(0)_t - \frac{\alpha'}{\sqrt{n}} g(0)_t.$$

Then

$$EY_{nt} = Q(\frac{\alpha}{\sqrt{n}}), \quad \sum_{t=1}^{n} Y_{nt} = \sum_{t=1}^{n} \rho(\frac{\alpha}{\sqrt{n}})_t - \rho(0)_t - \frac{\alpha'}{\sqrt{n}} g(0)_t.$$

For simplicity, let

$$G_{nt} = \frac{\alpha'}{\sqrt{n}}(g(\frac{\alpha}{\sqrt{n}})_t - g(0)_t).$$

Thus, we have

$$\begin{aligned}
\mathrm{Var} \sum Y_{nt} &\leq E(\sum Y_{nt})^2 \\
&\leq nEG_{n1}^2 + 2\sum_{k=1}^{n-1}(n-k)EG_{n1}G_{n,n+1-k} \\
&\leq E\{\alpha'(g(\frac{\alpha}{\sqrt{n}})_1 - g(0)_1)\}^2 \\
&\quad +2\sum_{k=1}^{n-1}\frac{n-k}{n}E\alpha'(g(\frac{\alpha}{\sqrt{n}})_1 - g(0)_1)(g(\frac{\alpha}{\sqrt{n}})_{n+1-k} - g(0)_{n+1-k})'\alpha
\end{aligned}$$

The first term in the right hand side can be approximated by

$$E\{\alpha'(g(\frac{\alpha}{\sqrt{n}})_1 - g(0)_1)\}^2 = \frac{1}{n}E(\alpha'\nabla g(0)_1\alpha)^2 + o(\frac{1}{n^2}), \tag{3.9}$$

and the second term can be approximated by

$$2\frac{n-k}{n^2}\sum_{k=1}^{n-1}E(\alpha'\nabla g(0)_1\nabla g(0)'_{n+1-k}\alpha)^2 + o(\frac{1}{n^2}). \tag{3.10}$$

These two terms in turn converge to 0 as $n \to \infty$. Therefore, we have

$$\frac{1}{n}\sum_{t=1}^{n} Y_{nt} - EY_{nt} \xrightarrow{\mathcal{P}} 0.$$

Asymptotic normality of $n^{-1/2}\sum g(\theta^0)_t$ follows from the classic central limit theorem under the Assumption 3.2.(ii). □

REMARK 3.5. Note that Y_{nt} is not independent, which is different from Niemiro (1992).

4. Differentiable Objective Functions. When the objective function satisfy the differentiable conditions, the result can be further shown easily and concretely by the martingale difference central limit theorem. In this section, we use Hodges-Lehmann's criteria to show the central limit theorem for θ_n. Consider $\rho(\theta_1^{-1/2}, e_t(\eta)) \equiv \rho(x, y)$. For simplicity of notation, we write $\rho(\theta)_t = \rho(\theta_1^{-1/2}, e_t(\eta))$ where $\theta = (\theta_1, \eta) \in \Theta \subset \mathbb{R}^m$: compact if it is not necessary to think θ_1 and η in a separate way. Here θ_1 is a scale parameter and $\eta = (\theta_2, \ldots, \theta_m)$.

The objective function is denoted by

$$Q(\theta) = E\rho(\theta_1^{-1/2}, e_t(\eta)). \tag{4.1}$$

The true value of θ, represented by θ^0 is defined by

$$Q(\theta^0) = \min_{\theta \in \Theta} Q(\theta). \tag{4.2}$$

On account of the simplicity of notation, we use σ and η^0 for the true value of θ_1 and η separately. The sample version corresponding to the objective function is

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \rho(\theta_1^{-1/2}, e_t(\eta)), \tag{4.3}$$

and

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta). \tag{4.4}$$

Let ρ_x, ρ_y be the partial derivative of $\rho(x, y)$ w.r.t x and y, $\mathcal{F}_t$ be the σ-field generated by the set of random variables $\{X_n;\, n \leq t\}$. As seen in the definition of $\rho(\theta_1^{-1/2}, e_t(\eta))$, it is $\mathcal{F}_t$-measurable. Also, by (1.5), we can see that $\frac{\partial}{\partial \eta} e_t(\eta)$ is $\mathcal{F}_{t-1}$-measurable.

Assumption 4.1.

(i) Let $\rho(\theta)$ be a measurable convex function with respect to θ from $\mathbb{R} \times \mathbb{R}^{m-1}$ to $\mathbb{R}$.
(ii) $Q(\theta)$ is well defined.
(iii) θ^0 satisfying (4.2) exists and is unique.
(iv) $E\rho_x(\theta^0) = 0$ and $E\rho_y(\theta^0) = 0$.
(v) $E\rho(\theta^0)^2 < \infty$, $E\rho_y(\theta^0)^2 < \infty$, $E(\frac{\partial^2}{\partial\eta\partial\eta'} e_t(\eta)|_{\theta=\theta_0})'(\frac{\partial^2}{\partial\eta\partial\eta'} e_t(\eta)|_{\theta=\theta_0})$ and $E(\frac{\partial}{\partial\eta} e_t(\eta) \frac{\partial}{\partial\eta'} e_t(\eta))'(\frac{\partial}{\partial\eta} e_t(\eta) \frac{\partial}{\partial\eta'} e_t(\eta))$ exist.
(vi) Define $\tilde{\rho}(\theta^0)_t \equiv \rho_y(\theta^0)_t \frac{\partial}{\partial\eta} e_t(\eta)\Big|_{\theta=\theta^0}$. To show central limit theorem for the martingale difference sequence $\{\tilde{\rho}(\theta^0)_t, \mathcal{F}_t\}$, we suppose the Lindeberg condition, by Euclidean norm $\|\cdot\|$ and indicator function $1(\,\cdot\,)$,

$$\frac{1}{n} \sum_{t=1}^{n} E(\|\tilde{\rho}(\theta^0)_t\|^2 1(\|\tilde{\rho}(\theta_0)_t\| \geq \epsilon)) \to 0, \tag{4.5}$$

$$\frac{1}{n} \sum_{t=1}^{n} E(\tilde{\rho}(\theta_0)_t \tilde{\rho}(\theta_0)'_t \Big| \mathcal{F}_{t-1}) \xrightarrow{\mathcal{P}} S, \tag{4.6}$$

where

$$S = E\tilde{\rho}(\theta^0)_t \tilde{\rho}(\theta_0)'_t.$$

REMARK 4.2. As a result of Assumption 4.1.(i), $\rho_x(\theta)$ and $\rho_y(\theta)$ are also functions from $\mathbb{R} \times \mathbb{R}^{m-1}$ to $\mathbb{R}$.

REMARK 4.3. In the L^2 theory, set $\rho : \mathbb{R} \times \mathbb{R}^{m-1} \mapsto \mathbb{R}$ as

$$\rho(x, y) = (xy)^2 - 2 \log x. \tag{4.7}$$

In this case, $\theta_1 = \sigma^2$, that is $\theta^{-1/2} = \sigma^{-1}$. We obtain the same result from the following asymptotics as that for the Whittle estimator in the second-order stationary case in the frequency domain.

REMARK 4.4. θ_1 can be a parameter for different scales. As an example, let θ_1 be defined as follows:

$$\rho(x, y) = (xy)^k - \frac{k}{k-2} x^{k-2}. \tag{4.8}$$

Then it is easy to see that θ_1^0 is $E\epsilon_t^k$. Note that in most case in time series analysis, $E\epsilon_t$ is assumed to be 0 or the symmetricity of ϵ_t is assumed. As an alternative, θ_1 can be defined by

$$\rho(x, y) = (x|y|)^k - \frac{k}{k-2} x^{k-2}, \tag{4.9}$$

then θ_1^0 is $E|\epsilon_t|^k$.

REMARK 4.5. Since the random structure of $\rho_x(\theta)$ is the same as $\rho(\theta)$, it is sufficient to only suppose $\rho(\theta^0) < \infty$.

From the definition,

$$\xi_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} (-\frac{1}{2}\theta_1^{-3/2} \rho_x(\theta), \rho_y(\theta) \frac{\partial}{\partial \eta'} e_t(\eta))'. \tag{4.10}$$

THEOREM 4.6. *Let $\hat{\theta}_n$ be defined in (4.4) under Assumption 4.1. Then we obtain*

(i) *$\hat{\theta}_n$ converges to the true value θ^0 in probability as $n \to \infty$.*
(ii) *$\sqrt{n}(\hat{\theta}_n - \theta^0)$ have a joint asymptotic normal distribution whose mean is 0 and the asymptotic covariance matrix is give by $H^{-1} V H^{-1}$, where*

$$\begin{aligned} H &= \begin{pmatrix} \frac{1}{2}\sigma^{-3} E\rho_{xx}(\theta^0) & 0 \\ 0 & E\rho_{yy}(\theta^0) E \frac{\partial}{\partial \eta} e_t(\eta) \frac{\partial}{\partial \eta'} e_t(\eta) \Big|_{\theta=\theta^0} \end{pmatrix}, \\ V &= \begin{pmatrix} E\rho_x(\theta^0)^2 & 0 \\ 0 & E\rho_y(\theta^0)^2 E \frac{\partial}{\partial \eta} e_t(\eta) \frac{\partial}{\partial \eta'} e_t(\eta) \Big|_{\theta=\theta^0}. \end{pmatrix}. \end{aligned}$$

In the case of Gaussian process with (4.7)*, the covariance matrix is*

$$V = 2D^{-1},$$

where

$$D_{ij} = (2\pi)^{-1} \Big\{ \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(\lambda) \frac{\partial}{\partial \theta_j} \log f(\lambda)\, dx \Big\} \Big|_{\theta=\theta^0},$$

and $f(\lambda)$ is the spectral density of the model.

PROOF. We will show Assumption 4.1 satisfies Assumption 2.2.

(i) If ρ is convex, then its derivative g is a non-decreasing function in each argument.

(ii) Stochastic expansion of $\xi_n(\theta)$ yields

$$\sqrt{n}(\xi_n(\theta^0 + u/\sqrt{n}) - \xi_n(\theta^0)) = \frac{1}{n}\sum_{t=1}^{n} \nabla\xi_n(\theta^0)'u + o_p(n^{-3/2}).$$

The (1,1)-element of $\nabla\xi_n(\theta)$ is

$$\nabla\xi_n(\theta^0)_{11} = \frac{1}{n}\sum_{t=1}^{n}(-\frac{1}{2}\sigma^{-3/2}\rho_{xx}(\theta^0)_t + \frac{3}{4}\sigma^{-5/2}\rho_x(\theta^0)_t).$$

Since $E\rho_x(\theta^0) = 0$ and $\rho_x(\theta^0)$ is $\mathcal{F}_t$-measurable, $\{\rho_x(\theta^0)\}_t$ is i.i.d. sequence with mean 0 and finite variance. Thus

$$\nabla\xi_n(\theta^0)_{11} \xrightarrow{\mathcal{P}} -\frac{1}{2}\sigma^{-3/2}E\rho_{xx}(\theta^0),$$

since

$$\sum_{t=1}^{n}\frac{3}{4}\sigma^{-5/2}\rho_x(\theta^0)_t \xrightarrow{\mathcal{P}} 0.$$

Similarly, we have (i,j)-element of $\nabla\xi_n(\theta)$ $(i \geq 2, j \geq 2)$,

$$\nabla\xi_n(\theta^0)_{ij} = \frac{1}{n}\sum_{t=1}^{n}\rho_{yy}(\theta^0)_t\frac{\partial}{\partial\eta}e_t(\eta)\frac{\partial}{\partial\eta'}e_t(\eta)\Big|_{\theta=\theta^0} + \frac{1}{n}\sum_{t=1}^{n}\rho_y(\theta^0)_t\frac{\partial^2}{\partial\eta\partial\eta'}e_t(\eta)\Big|_{\theta=\theta_0}.$$

The last term in the right hand side of the equation above forms a martingale since $b_0 = 1$ implies that $\frac{\partial}{\partial\eta}e_t(\eta)$ and $\frac{\partial^2}{\partial\eta\partial\eta'}e_t(\eta)$ is $\mathcal{F}_{t-1}$-measurable. Noting that

$$\begin{aligned}E\rho_y(\theta^0)\frac{\partial^2}{\partial\eta\partial\eta'}e_t(\eta)\Big|_{\theta=\theta_0} &= E(E\rho_y(\theta^0)\frac{\partial^2}{\partial\eta\partial\eta'}e_t(\eta)\Big|_{\theta=\theta_0}\Big| \mathcal{F}_{t-1}) \\ &= E\frac{\partial^2}{\partial\eta\partial\eta'}e_t(\eta)\Big|_{\theta=\theta_0}E\rho_y(\theta^0) = 0,\end{aligned}$$

under Assumption 4.1.(v), we obtain

$$\nabla\xi_n(\theta^0)_{ij} \xrightarrow{\mathcal{P}} E\rho_{yy}(\theta^0)E\frac{\partial}{\partial\eta}e_t(\eta)\frac{\partial}{\partial\eta'}e_t(\eta)\Big|_{\theta=\theta_0}$$

by Chebyshev's inequality to show the last term converges to 0 in probability.

(iii) Consider

$$\sqrt{n}\xi_n(\theta^0) = n^{-1/2}\sum_{t=1}^{n}(-\frac{1}{2}\sigma^{-3/2}\rho_x(\theta^0), \rho_y(\theta^0)\frac{\partial}{\partial\eta'}e_t(\eta)\Big|_{\theta=\theta^0})'.$$

As seen in (ii), $\{\xi_n(\theta^0), \mathcal{F}_n\}$ is a martingale with respect to $\mathcal{F}_n$. Under Assumption 4.1.(vi), we have

$$\sqrt{n}\xi_n(\theta^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V),$$

where

$$V = \begin{pmatrix} E\rho_x(\theta^0)^2 & 0 \\ 0 & E\rho_y(\theta^0)^2 E\frac{\partial}{\partial\eta}e_t(\eta)\frac{\partial}{\partial\eta'}e_t(\eta)\Big|_{\theta=\theta^0}. \end{pmatrix}$$

As a result, the asymptotic normality for $\hat{\theta}$ is shown and the asymptotic variance is given by $H^{-1}VH^{-1}$.

□

5. Asymptotics of M_m estimators. In this section, we give conditions for asymptotic normality of M_m estimators. Let

$$Q(\theta) = E\rho(\theta_1^{-1/2}, e_1(\eta), \ldots, e_m(\eta)). \tag{5.1}$$

The sample analog of $Q_n(\theta)$ is defined as

$$Q_n(\theta) = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} \rho(\theta_1^{-1/2}, e_{i_1}(\eta), \ldots, e_{i_m}(\eta)), \tag{5.2}$$

and

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta). \tag{5.3}$$

ASSUMPTION 5.1.

(i) $\rho(\theta)$ is convex with respect to θ for each ϵ_t.
(ii) $Q(\theta)$ is well defined.
(iii) θ^0 satisfying (3.2) exists and is unique.
(iv) For all diagonal elements (i,i),

$$\sum_{k=1}^{\infty} E|(g(\frac{\alpha}{\sqrt{n}})_1 - g(0)_1)_i(g(\frac{\alpha}{\sqrt{n}})_k - g(0)_k)_i| < \infty.$$

(v) $Q(\theta)$ is twice differentiable at θ^0 and $\nabla^2 Q(\theta^0)$ is positive definite.

THEOREM 5.2. *Let $\hat{\theta}_n$ be defined in (5.3) under Assumption 5.1. Then we obtain*

(i) *$\hat{\theta}_n$ converges to the true value θ^0 in probability as $n\to\infty$.*
(ii) *$\sqrt{n}(\hat{\theta}_n - \theta^0)$ have a joint asymptotic normal distribution whose mean is 0 and the asymptotic covariance matrix is give by $m^2H^{-1}VH^{-1}$, where*

$$\begin{aligned} H &= \nabla^2 Q(\theta^0) \\ V &= \mathrm{Var}\, g(\theta^0). \end{aligned}$$

PROOF. Let J denote the set of all m element subsets of $\{1,\ldots,n\}$. For any $j = \{i_1,\ldots,i_m\} \in J$, let Y_j be the random vector $(e_{i_1},\ldots,e_{i_m})$. Accordingly, the notation $\rho(\frac{\alpha}{\sqrt{n}})_t$ and $\rho(0)_t$ are changed in the following way: for $\alpha = (\alpha_1, \alpha_2)$,

$$\begin{aligned} \rho(\frac{\alpha}{\sqrt{n}})_j &:= \rho((\theta_1^0 + \frac{\alpha_1}{\sqrt{n}})^{-1/2}, Y_j(\eta^0 + \frac{\alpha_2}{\sqrt{n}})), \\ \rho(0)_j &:= \rho((\theta_1^0)^{-1/2}, Y_j(\eta^0)). \end{aligned}$$

For any fixed α and j, define

$$Z_{nj} = \rho(\frac{\alpha}{\sqrt{n}})_j - \rho(0)_j - \frac{\alpha'}{\sqrt{n}} g(0)_j.$$

Note that $EZ_{nj} = Q(\frac{\alpha}{\sqrt{n}})$. For the same reason, we have

$$\text{Var} \sum Z_{nj} \xrightarrow{\mathcal{P}} 0.$$

Since $\text{Var}(n\binom{n}{m}^{-1} \sum Z_{nj}) \le m^2 \text{Var} \sum Z_{nj}$,

$$n\binom{n}{m}^{-1} \sum Z_{nj} - nEZ_{nj} \xrightarrow{\mathcal{P}} 0.$$

Thus the result of Theorem 5.2 depends on the asymptotics of $\sqrt{n}\binom{n}{m}^{-1} \sum g(0)_j$. Regard $g(0)_j$ as a kernel of U-statistics, we define the degenerate kernel of $g(0)$ by

$$g_0^c(x_1, \ldots, x_c) = \sum_{r=0}^{c} \binom{c}{r} (-1)^{c-r} \int \cdots \int_{\mathbb{R}^{m-r}} g(\theta^0, x_1, \ldots, x_m) \prod_{i=r+1}^{m} dF(x_i).$$

Suppose U_n is generated by $g(0)$ and U_n^c is generated by g_0^c, we have by Hoeffding's projection that

$$U_n = \sum_{c=1}^{m} \binom{m}{c} U_n^c = \frac{m}{n} \sum_{t=1}^{n} g(0)_t + R_n,$$

where $R_n \xrightarrow{\mathcal{P}} 0$. The conclusion is completed by the asymptotic normality of $n^{-1/2} \sum_{t=1}^{n} g(0)_t$. □

6. Examples. Suppose the second order stationary process $\{X_t\}$ is generated by the model

$$X_t - \beta_0 X_{t-1} = \epsilon_t, \tag{6.1}$$

where $\epsilon_t \sim$ i.i.d. $(0, \sigma^2)$. For the estimation of β_0, take

$$b_j(\eta) = \begin{cases} \beta & j = 1, \\ 0 & j \ge 2. \end{cases} \tag{6.2}$$

6.1. *Asymptotics of L^2 theory in AR(1) case.* From Remark 4.3, the objective function in L^2 theory is given by

$$Q_n(\theta) = \log \theta_1 + \frac{1}{n} \sum_{t=2}^{n} \theta_1^{-1} (X_t - \beta X_{t-1})^2 + o_p(1). \tag{6.3}$$

Note that the objective function is asymptotically equivalent to Whittle estimator. Also, the estimator is a modification of least square estimation since the scale parameter is estimated simultaneously. With

$$\rho_x(\theta^0) = \frac{2\epsilon^2}{\sigma} - 2\sigma, \qquad \rho_y(\theta^0) = \frac{2\epsilon}{\sigma^2},$$
$$\rho_{xx}(\theta^0) = 2\sigma^2 + 2\epsilon^2, \qquad \rho_{yy}(\theta^0) = \frac{2}{\sigma^2},$$

by Theorem 4.6 we obtain

$$H = \begin{pmatrix} 2\sigma^{-1} & 0 \\ 0 & 2(1-\beta_0^2)^{-1} \end{pmatrix}, \tag{6.4}$$

$$V = \begin{pmatrix} 4\sigma^{-2}(\mu_4 - \sigma^4) & 0 \\ 0 & 4(1-\beta_0^2)^{-1} \end{pmatrix}, \tag{6.5}$$

where μ_4 is the fourth moment of ϵ_t. As a result,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \operatorname{diag}(\mu_4 - \sigma^4, 1 - \beta_0^2)). \tag{6.6}$$

In the Gaussian case, $\mu_4 - \sigma^4 = 2\sigma^4$.

6.2. *Asymptotics of L^1 theory in AR(1) case (new aspects).* One parameterization for L^1 case is given in Remark 4.4. In this subsection, we are interested in another parameterization with convex objective function defined by Koenker and Bassett (1978)'s check function $\rho^\tau(u)$. That is,

$$\rho^\tau(u) = u(\tau - 1(u < 0)), \tag{6.7}$$

where $1(\cdot)$ is the indicator function. Suppose $\rho(x, y)$ is defined by

$$\rho(x, y) = \rho^\tau(xy) + x^{-1}. \tag{6.8}$$

Then the true parameters are given by

$$\theta_1^0 = \frac{1}{2}E|\epsilon|, \quad b_j(\eta^0) = \beta_0\delta(j, 1) \quad \text{for } j \geq 1, \tag{6.9}$$

from the result that $E\epsilon_t 1(\epsilon_t > 0) = \frac{1}{2}E|\epsilon|$ and $EX_{t-1}1(e_t(\eta) < 0) \neq 0$ if $b_j(\eta) \neq \beta_0$. Interestingly, the true parameters do not depend on τ even τ is included in the check function. To generalize the result, suppose more that

$$\kappa = P(\epsilon < 0), \tag{6.10}$$

where κ is different from τ in the check function. Then we obtain

$$\rho_x(\theta^0) = \epsilon(\tau - 1(\epsilon < 0)) - \theta_1^0, \qquad \rho_{xx}(\theta^0) = 2(\theta_1^0)^{3/2},$$
$$\rho_y(\theta^0) = \theta_1^0(\tau - 1(\epsilon < 0)), \qquad \rho_{yy}(\theta^0) = \delta(\epsilon),$$

where $\delta(\cdot)$ is the Dirac delta function. In conclusion, the asymptotic variance of the M-estimator defined by $\rho^\tau(u)$ is given by

$$H^{-1}VH^{-1} = \begin{pmatrix} \frac{1}{4}(\theta_1^0)^{-3}(\tau^2\sigma^2 + (1-2\tau)a) & 0 \\ 0 & \frac{\tau^2-2\tau\kappa+\kappa}{\sigma^2} f(0)^{-2}(1-\beta_0^2) \end{pmatrix}, \tag{6.11}$$

where $a = E\epsilon^2 1(\epsilon < 0)$. As a special case $\kappa = \tau$, the variance of the coefficient parameter is given by

$$\frac{\tau(1-\tau)}{\sigma^2} f(0)^{-2}(1-\beta_0^2). \tag{6.12}$$

Acknowledgements. The author thank the editor and an anonymous referee for their constructive comments that have substantially improved the earlier version of the paper.

References.

Beran, J. (1994). On a class of M-estimators for Gaussian long-memory models. *Biometrika* **81** 755–766.

Billingsley, P. (1968). *Convergence of probability measures* **493**. Wiley-Interscience.

Bose, A. (1998). Bahadur representation of M_m estimates. *The Annals of Statistics* **26** 771–777.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag.

Denker, M. and Keller, G. (1983). On U-statistics and v. mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64** 505–522.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic press New York.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics* 598–611.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics* 132–153. Correction: (1993). 21, 1115-1117.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35** 73–101.

Ibragimov, I. and Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen.

Inagaki, N. and Kondo, M. (1980). Several estimators of the autocorrelation based on limiter estimating functions for a stationary Gaussian process. *Journal of the Japan Statistical Society* 1–15.

Koenker, R. and Bassett, J. G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society* 33–50.

Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *The Annals of Statistics* 1514–1533.

Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary*-mixing processes. *Probability Theory and Related Fields* **25** 71–82.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics* **162**. Wiley. com.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. Springer New York.

Yoshihara, K.-i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields* **35** 237–252.

Yan Liu
Department of Applied Mathematics
School of Fundamental Science and Engineering
Waseda University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555
Japan
E-mail: great-rainbow@ruri.waseda.jp