IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 143–152

DOI: 10.1214/10-IMSCOLL715

A class of minimum distance estimators in AR(p) models with infinite error variance*

Hira L. Koul[1] and Xiaoyu Li

Michigan State University

Abstract: In this note we establish asymptotic normality of a class of minimum distance estimators of autoregressive parameters when error variance is infinite, thereby extending the domain of their applications to a larger class of error distributions that includes a class of stable symmetric distributions having Pareto-like tails. These estimators are based on certain symmetrized randomly weighted residual empirical processes. In particular they include analogs of robustly weighted least absolute deviation and Hodges–Lehmann type estimators.

1. Introduction

When modeling extremal events one often comes across autoregressive time series with infinite variance innovations, cf. Embrecht, Küppelberg and Mikosch [10]. Assessing distributional properties of classical inference procedures in these time series models is thus important. Weak and strong consistency with some convergence rate of the least square (LS) estimator of the autoregressive parameter vector in such models are discussed in Kanter and Steiger [13], Hannan and Kanter [12], and Knight [14]) while Davis and Resnick [4] and [5] discuss its limiting distribution. Strong consistency and convergence rate of the least absolute deviation (LAD) estimator are considered separately by Gross and Steiger [11], and An and Chen [1]. Davis, Knight and Liu [6] and Davis and Knight [3] discuss consistency and asymptotic distributions of the LAD and M-estimators in autoregressive models of a known order p when error distribution is in the domain of attraction of a stable distribution of index $\alpha \in (0, 2)$. Knight [15] proves asymptotic normality of a class of M-estimators in a dynamic linear regression model where the errors have infinite variance but the exogenous regressors satisfy the standard assumptions. Ling [18] discusses asymptotic normality of a class of weighted LAD estimators.

Minimum distance (m.d.) estimation method consists of obtaining an estimator of a parameter by minimizing some dispersion or pseudo distance between the data and the underlying model. For a stationary autoregressive time series of a known order p with i.i.d. symmetric innovations a class of m.d. estimators was proposed in Koul [16]. This class of estimators is obtained by minimizing a class of certain integrated squared differences between randomly weighted empirical processes of residuals and negative residuals. More precisely, let p be a known positive integer

*Research in part supported by the USA NSF DMS Grant 0704130.
[1]Michigan State University, East Lansing, USA. e-mail: koul@stt.msu.edu
AMS 2000 subject classifications: Primary 62G05; secondary 62M10, 62G20.
Keywords and phrases: asymptotic normality, Pareto-like tails distributions.

and consider the linear autoregressive process $\{X_i\}$ obeying the model

$$X_i = \rho_1 X_{i-1} + \rho_2 X_{i-2} + \cdots + \rho_p X_{i-p} + \varepsilon_i, \quad i = 0, \pm 1, \pm 2, \cdots, \tag{1.1}$$

for some $\rho := (\rho_1, \cdots, \rho_p)' \in \mathbb{R}^p$, where the innovations $\{\varepsilon_i\}$ are i.i.d. r.v.'s from a continuous distribution function (d.f.) F, symmetric around zero, not necessarily known otherwise. We shall also assume $\{X_i\}$ is a strictly stationary solution of the equations (1.1). Some sufficient conditions for this to exist in the case of some heavy tail error distributions are given in the next section. Here, and in the sequel, by stationary we mean strictly stationary.

Let $Y_{i-1} := (X_{i-1}, \cdots, X_{i-p})'$. Because of the assumed symmetry of the innovation d.f. F, $X_i - \rho' Y_{i-1}$ and $-X_i + \rho' Y_{i-1}$ have the same distribution for each $i = 1, \cdots, n$. Using this fact, the following class of m.d. estimators was proposed in Koul [16].

$$\begin{aligned} K_h^+(t) &:= \int \Big\| n^{-1/2} \sum_{i=1}^n h(Y_{i-1}) \Big\{ I(X_i \le x + t'Y_{i-1}) \\ &\qquad - I(-X_i < x - t'Y_{i-1}) \Big\} \Big\|^2 \mathrm{d}G(x), \\ \rho_h^+ &:= \operatorname{argmin}\{K_h^+(t);\ t \in \mathbb{R}^p\}. \end{aligned}$$

Here h is a measurable function from $\mathbb{R}^p$ to $\mathbb{R}^p$ with its components h_k, $k = 1, \cdots, p$, G is a nondecreasing right continuous function on $\mathbb{R}$ having left limits, possibly inducing a σ-finite measure on $\mathbb{R}$, and $\|\cdot\|$ stands for the usual Euclidean norm.

A large subclass of the estimators ρ_h^+, as h and G vary, is known to be robust against additive innovation outliers, cf. Dhar [8]. The class of estimators ρ_h^+, when $h(x) = x$ and as G varies, have desirable asymptotic relative efficiency properties. Moreover, for $h(x) = x$, ρ_h^+ becomes the LAD estimator when G is degenerate at zero while for $G(x) \equiv x$, it is an analog of the Hodges–Lehmann estimator.

Asymptotic normality of these estimators under a broad set of conditions on h, G and F was established in Koul (16, 17, chapter 7). These conditions included the condition of finite error variance. Main reason for having this assumption was to ensure stationarity of the underlying process $\{X_i\}$ satisfying (1.1). Given the importance of heavy tail error distributions and robustness properties of these m.d. estimators, it is desirable to extend the domain of their applications to autoregressive time series with heavy tail errors. We now establish asymptotic normality of these estimators here under similar general conditions in which not only the error variance is not finite but also even the first moment may not be finite.

In the next section, we first state general conditions for asymptotic normality of these estimators. Then we give a set of sufficient and easy to verify conditions that imply these general conditions. Among the new results is the asymptotic normality of a class of analogs of robust Hodges–Lehmann type estimators of the autoregressive parameters when error distribution has infinite variance. We also give examples of several functions h and G that satisfy the assumed conditions. In the last section another class of m.d. estimators based on residual ranks is discussed briefly to be used when errors may not have a symmetric distribution.

2. Main result

To describe our main result we now state the needed assumptions, most of which are the same as in Koul [17].

$$\text{Either} \quad e'h(y)y'e \ge 0, \quad \text{Or} \quad e'h(y)y'e \le 0, \quad \forall\, y, e \in \mathbb{R}^p,\ \|e\| = 1. \tag{2.1}$$

$$(2.2)\qquad (a)\ 0 < E(|h_k(Y_0)|\,\|Y_0\|) < \infty, \quad \forall\, 1 \le k \le p. \qquad (b)\ E\|h(Y_0)\|^2 < \infty.$$

In the following assumptions b is any positive finite real number.

$$(2.3)\qquad \int E\|h(Y_0)\|^2 |F(x + n^{-1/2}(v'Y_0 + a\|Y_0\|)) - F(x)|\,\mathrm{d}G(x) = o(1), \quad \forall\, \|v\| \le b, a \in \mathbb{R}.$$

There exists a constant $k \in (0, \infty)$, such that for all $\delta > 0$, $\|v\| \le b$ and $1 \le k \le p$,

$$(2.4)\qquad \liminf_n P\Big(\int \Big[n^{-1/2} \sum_{i=1}^n \sum_{k=1}^n h_k^{\pm}(Y_{i-1}) \Big\{ F(x + n^{-1/2} v'Y_{i-1} + \delta_{ni}) - F(x + n^{-1/2} v'Y_{i-1} - \delta_{ni}) \Big\} \Big]^2 \mathrm{d}G(x) \le k\delta^2 \Big) = 1,$$

where $\delta_{ni} := n^{-1/2}\delta\|Y_{i-1}\|$, $h_k^+ := \max(0, h_k)$, $h_k^- := h_k - h_k^+$.

$$(2.5)\qquad \int \Big\| n^{-1/2} \sum_{i=1}^n h(Y_{i-1}) \Big\{ F(x + n^{-1/2} v'Y_{i-1}) - F(x) - n^{-1/2} v'Y_{i-1} f(x) \Big\} \Big\|^2 \mathrm{d}G(x) = o_p(1), \quad \forall\, \|v\| \le b.$$

The d.f. F has Lebesgue density f satisfying the following.

$$(2.6)\ (\mathrm{a})\ 0 < \int f^2\,\mathrm{d}G < \infty, \quad (\mathrm{b})\ 0 < \int f\,\mathrm{d}G < \infty, \quad (\mathrm{c})\ \int_0^\infty (1 - F)\,\mathrm{d}G < \infty.$$

Assumption of stationarity replaces the assumption of finite error variance (7.4.7)(b) of Koul [17]. We are now ready to state our main result.

Theorem 2.1 *Assume the autoregressive process given at (1.1) exists and is strictly stationary. In addition, assume the functions h, G, F satisfy assumptions (2.1)–(2.6) and that G and F are symmetric around zero. Then,*

$$n^{1/2}(\rho_h^+ - \rho) = -\Big\{ \mathcal{B}_n \int f^2\,\mathrm{d}G \Big\}^{-1} \mathcal{S}_n^+ + o_p(1),$$

where $\mathcal{B}_n := n^{-1}\sum_{i=1}^n h(Y_{i-1})Y_{i-1}'$, and

$$\mathcal{S}_n^+ := \int n^{-1/2} \sum_{i=1}^n h(Y_{i-1}) \Big\{ I(\varepsilon_i \le x) - I(-\varepsilon_i < x) \Big\} f(x)\,\mathrm{d}G(x) = n^{-1/2} \sum_{i=1}^n h(Y_{i-1})[\psi(-\varepsilon_i) - \psi(\varepsilon_i)], \qquad \psi(x) := \int_{-\infty}^x f\,\mathrm{d}G.$$

Consequently,

$$n^{1/2}(\rho_h^+ - \rho) \to_d \mathcal{N}\Big(0, \frac{Var(\psi(\varepsilon))}{(\int f^2\,\mathrm{d}G)^2} \mathcal{B}^{-1}\mathcal{H}\mathcal{B}^{-1}\Big),$$

where $\mathcal{B} := Eh(Y_0)Y_0'$ and $\mathcal{H} := Eh(Y_0)h(Y_0)'$.

The existence of ρ_h^+ under the finite variance assumption has been discussed in Dhar [9]. Upon a close inspection one sees that this proof does not require the finiteness of any error moment but only the stationarity of the process and assumptions (2.1), (2.2)(b) and (2.6)(c). Also note that (2.1), (2.2)(a) and the Ergodic Theorem implies the existence of $\mathcal{B}^{-1}$, and $\mathcal{B}_n^{-1}$ for all n.

In view of the stationarity of the process $\{X_i\}$, the details of the proof of Theorem 2.1 are very similar to that of Theorem 7.4.5 in Koul [17] and are left out for an interested reader.

3. Some stronger assumptions and Examples

In this section we shall now discuss some easy to verify sufficient conditions for (2.3) to (2.5). In particular, we shall show that the above theorem is applicable to robust LAD and analogs of robust Hodges–Lehmann type estimators.

First, consider (2.4) and (2.5). As shown in Koul [17], under the finite error variance assumption, (2.2)(a), (2.4) and (2.5) are implied by (2.2)(b), (2.6)(a) and the assumption

$$\int |f(x+s)-f(x)|^2\,\mathrm{d}G(x)\to 0,\quad s\to 0. \tag{3.1}$$

We shall now show that (2.4) and (2.5) continue to hold under (2.2)(a), (2.6)(a) and (3.1) when $\{X_i\}$ is stationary, without requiring the error variance to be finite. First, consider (2.4). Recall $\delta_{ni}:=n^{-1/2}\delta\|Y_{i-1}\|$. Then, the r.v.'s inside the probability statement of (2.4) equals to

$$\begin{aligned}
&\int\Big[n^{-1/2}\sum_{i=1}^n h_k^{\pm}(Y_{i-1})\int_{-\delta_{ni}}^{\delta_{ni}} f(x+n^{-1/2}v'Y_{i-1}+s)\,\mathrm{d}s\Big]^2\,\mathrm{d}G(x)\\
&=\quad \sum_{i=1}^n\sum_{j=1}^n\int n^{-1}h_k^{\pm}(Y_{i-1})h_k^{\pm}(Y_{j-1})\\
&\qquad\cdot\int_{-\delta_{ni}}^{\delta_{ni}}\int_{-\delta_{nj}}^{\delta_{nj}} f(x+n^{-1/2}v'Y_{i-1}+s)f(x+n^{-1/2}v'Y_{j-1}+t)\,\mathrm{d}s\mathrm{d}t\,\mathrm{d}G(x)\\
&=\quad \delta^2 n^{-2}\sum_{i=1}^n\sum_{j=1}^n\|Y_{i-1}\|\|Y_{j-1}\|h_k^{\pm}(Y_{i-1})h_k^{\pm}(Y_{j-1})\\
&\qquad\frac{1}{\delta_{ni}\delta_{nj}}\int_{-\delta_{ni}}^{\delta_{ni}}\int_{-\delta_{nj}}^{\delta_{nj}}\int f(x+n^{-1/2}v'Y_{i-1}+s)\\
&\qquad\cdot f(x+n^{-1/2}v'Y_{j-1}+t)\,\mathrm{d}G(x)\,\mathrm{d}s\mathrm{d}t\\
&\le\quad \delta^2\Big(n^{-1}\sum_{i=1}^n\|Y_{i-1}\|\,|h_k(Y_{i-1})|\Big)^2\\
&\qquad\max_{1\le i,j\le n}\frac{1}{\delta_{ni}\delta_{nj}}\int_{-\delta_{ni}}^{\delta_{ni}}\int_{-\delta_{nj}}^{\delta_{nj}}\int f(x+n^{-1/2}v'Y_{i-1}+s)\\
&\qquad\cdot f(x+n^{-1/2}v'Y_{j-1}+t)\,\mathrm{d}G(x)\,\mathrm{d}s\mathrm{d}t\\
&\le\quad 4\delta^2\Big(n^{-1}\sum_{i=1}^n\|Y_{i-1}\|\,|h_k(Y_{i-1})|\Big)^2\\
&\qquad\times\Big[\max_{1\le i\le n}\frac{1}{2\delta_{ni}}\int_{-\delta_{ni}}^{\delta_{ni}}\Big(\int f^2(x+n^{-1/2}v'Y_{i-1}+s)\,\mathrm{d}G(x)\Big)^{1/2}\Big]^2\mathrm{d}s
\end{aligned}$$

$$\to_p \quad 4\delta^2\Big[E\big(\|Y_0\||h_k(Y_0)|\big)\Big]^2\Big(\int f^2\,\mathrm{d}G\Big).$$

The above last claim is implied by the Ergodic Theorem which uses (2.2)(a), and the fact that under (2.6)(a) and (3.1), the second factor in the last but one bound above tends, in probability, to a finite and positive limit $\int f^2\,\mathrm{d}G$.

The argument for verifying (2.5) is similar. Let $b_{ni} := n^{-1/2}b\|Y_{i-1}\|$. Then,

$$\begin{aligned}
&\int\Big\|n^{-1/2}\sum_{i=1}^n h(Y_{i-1})\Big\{F(x+n^{-1/2}v'Y_{i-1})-F(x)-n^{-1/2}v'Y_{i-1}f(x)\Big\}\Big\|^2\,\mathrm{d}G(x)\\
=&\ \sum_{k=1}^p\int\Big[n^{-1/2}\sum_{i=1}^n h_k(Y_{i-1})\int_0^{n^{-1/2}v'Y_{i-1}}\{f(x+s)-f(x)\}\,\mathrm{d}s\Big]^2\,\mathrm{d}G(x)\\
\le&\ b^2n^{-2}\sum_{k=1}^p\sum_{i=1}^n\sum_{j=1}^n\|Y_{i-1}\|\|Y_{j-1}\||h_k(Y_{i-1})||h_k(Y_{j-1})|\\
&\times\max_{1\le i,j\le n}\frac{1}{b_{ni}b_{nj}}\int_{-b_{ni}}^{b_{ni}}\int_{-b_{nj}}^{b_{nj}}\Big\{\int|f(x+s)-f(x)||f(x+t)-f(x)|\,\mathrm{d}G(x)\Big\}\,\mathrm{d}s\mathrm{d}t\\
\le&\ 4b^2\sum_{k=1}^p\Big(n^{-1}\sum_{i=1}^n\|Y_{i-1}\||h_k(Y_{i-1})|\Big)^2\\
&\times\Big[\max_{1\le i\le n}\frac{1}{2b_{ni}}\int_{-b_{ni}}^{b_{ni}}\Big(\int|f(x+s)-f(x)|^2\,\mathrm{d}G(x)\Big)^{1/2}\,\mathrm{d}s\Big]^2\\
\to_p&\ 0.
\end{aligned}$$

The last but one inequality follows from the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\int|f(x+s)-f(x)||f(x+t)-f(x)|\,\mathrm{d}G(x)\\
&\quad\le\ \Big\{\int|f(x+s)-f(x)|^2\,\mathrm{d}G(x)\Big\}^{1/2}\Big\{\int|f(x+t)-f(x)|^2\,\mathrm{d}G(x)\Big\}^{1/2},
\end{aligned}$$

while the last claim follows from (2.2)(a), Ergodic Theorem, and (3.1).

Now we turn to the verification of (2.3). First, consider the case when G is a finite measure. In this case, by the Dominated Convergence Theorem, (2.2)(b) and the continuity of F readily imply (2.3).

Of special interest among finite measures G is the measure degenerate at zero. Now assume that the distribution of Y_0 is continuous. Then, because F is continuous, the joint distribution of Y_{i-1}, X_i, $1\le i\le n$, is continuous for all n, and hence,

$$K_h^+(t) := \Big\|\sum_{i=1}^n h(Y_{i-1})\mathrm{sign}(X_i-t'Y_{i-1})\Big\|^2,\quad \forall\,t\in\mathbb{R}^p,\ \text{w.p. }1,$$

and the corresponding m.d. estimator, denoted by $\rho^+_{h,\mathrm{LAD}}$, becomes an analog of the LAD estimator. Note also that now (3.1) is equivalent to the continuity of f at zero, $\psi(x)$ of Theorem 2.1 equals $f(0)I(x>0)$ and $\mathrm{Var}(\psi(\varepsilon)) = f^2(0)/4$, where ε is the innovation variable having d.f. F. We summarize asymptotic normality result for $\rho^+_{h,\mathrm{LAD}}$ in the following

Corollary 3.1 *Assume the stationary AR(p) model (1.1) and assumptions (2.1), (2.2) hold. In addition, assume that the symmetric error density f is continuous at 0 and $f(0) > 0$. Then,*

$$n^{1/2}(\rho^{+}_{h,\mathrm{LAD}} - \rho) \to_d \mathcal{N}\Big(0, \frac{\mathcal{B}^{-1}\mathcal{H}\mathcal{B}^{-1}}{4f^2(0)}\Big), \qquad \mathcal{B} := Eh(Y_0)Y_0', \quad \mathcal{H} := Eh(Y_0)h(Y_0)'.$$

Note that this result does not require finiteness of any error moment.

Examples of h that satisfy (2.1) and (2.2) include the weight function

$$h(y) = h_1(y) := yI(\|y\| \le c) + c(y/\|y\|^2)I(\|y\| > c), \quad c > 0, \tag{3.2}$$

$$h(y) = h_2(y) := y/(1 + \|y\|^2). \tag{3.3}$$

Note that both are bounded functions and trivially satisfy (2.1). Moreover, continuity of Y_0 implies that h_1 satisfies (2.2), because for all $1 \le k \le p$,

$$\begin{aligned} 0 < cE\Big(\frac{|Y_{0k}|}{\|Y_0\|}I(\|Y_0\| > c)\Big) &\le E\Big(|h_{1k}(Y_0)|\,\|Y_0\|\Big) \\ &\le E\Big(\|Y_0\|^2 I(\|Y_0\| \le c) + cI(\|Y_0\| > c)\Big) \le c^2 + c. \end{aligned}$$

Similarly, h_2 also satisfies (2.2), because for all $1 \le k \le p$,

$$0 < E\Big(\frac{|Y_{0k}|\|Y_0\|}{1 + \|Y_0\|^2}\Big) = E|h_{2k}(Y_0)|\,\|Y_0\| \le E\big(\|Y_0\|^2/(1 + \|Y_0\|^2\big) < 1.$$

Ling [18] considers weighted LAD estimators obtained by minimizing $\sum_{i=1}^n g(Y_{i-1})|X_i - t'Y_{i-1}|$ w.r.t. t, where g is a positive measurable function on $\mathbb{R}^p$ satisfying

$$E\{g(Y_0) + g^2(Y_0)\}(\|Y_0\|^2 + \|Y_0\|^3) < \infty. \tag{3.4}$$

This estimator corresponds to $\rho^{+}_{h,\mathrm{LAD}}$ with $h(y) = g(y)y$. Ling establishes asymptotic normality of this estimator under some assumptions that include f being differentiable everywhere.

Now note that with $h(y) = g(y)y$, (2.1) is a priori satisfied, and (2.2) becomes $0 < E\big(g(Y_0)|Y_{0k}|\|Y_0\|\big) < \infty$, $1 \le k \le p$ and $E\big(g^2(Y_0)\|Y_0\|^2\big) < \infty$. Positivity condition is again implied by the continuity of the distribution of Y_0 and g being positive. The finiteness of these two expectation is implied by $E\big[\big(g(Y_0)+g^2(Y_0)\big)\|Y_0\|^2\big] < \infty$, clearly a much weaker condition than (3.4). And the above corollary does not require differentiability of f. Thus for a large class of weighted LAD estimators, the above corollary provides a some what stronger result.

Bounded h and σ-finite G: Now we continue our discussion of assumption (2.3) for a general G that may not induce a finite measure. Note that because the second error moment is not necessarily finite, the identity function $h(x) \equiv x$ does not satisfy (2.2). Moreover, if h is unbounded then the corresponding ρ_h^{+} is known to be non-robust against innovation outliers, cf. Dhar [8]. This property is similar to that of M-estimators, cf. Denby and Martin [7]. We shall thus verify (2.3) only for a bounded h and a large class of G's. Accordingly, suppose for some $C < \infty$,

$$\sup_{y \in \mathbb{R}^p} \|h(y)\| \le C. \tag{3.5}$$

Additionally, suppose F is absolutely continuous with density f satisfying

$$\int\!\!\int f(x+s)P(\|Y_0\| > n^{1/2}\beta|s|)\,\mathrm{d}G(x)\,\mathrm{d}s \to 0, \quad \forall\, 0 < \beta < \infty. \tag{3.6}$$

Now we shall show that (3.5) and (3.6) implies (2.3). Then, by the Fubini Theorem,

$$
\begin{aligned}
&\int E\|h(Y_0)\|^2|F(x+n^{-1/2}(v'Y_0+a\|Y_0\|))-F(x)|\,\mathrm{d}G(x)\\
&\le\quad C^2E\int_{-n^{-1/2}(b+|a|)\|Y_0\|}^{n^{-1/2}(b+|a|)\|Y_0\|}\int f(x+s)\,\mathrm{d}G(x)\,\mathrm{d}s\\
&=\quad C^2\int\int f(x+s)\,P(\|Y_0\|>n^{1/2}c^{-1}|s|)\,\mathrm{d}G(x)\,\mathrm{d}s\\
&\to\quad 0,\text{by (3.6)}.
\end{aligned}
$$

To summarize, we have shown (2.2)(a), (2.6)(a) and (3.1) imply (2.4) and (2.5) for general h and G, while (3.6) implies (2.3) for bounded h and a σ-finite G.

Verification of (3.6) is relatively easy if the following two assumptions hold.

(3.7) G is absolutely continuous with $\mathrm{d}G(x)=\gamma(x)\,\mathrm{d}x$, where γ is bounded, i.e., $\|\gamma\|_\infty:=\sup_{x\in\mathbb{R}}|\gamma(x)|<\infty$,

(3.8) $E\|Y_0\|<\infty$.

For, then, by Fubini's Theorem, the left hand side of (3.6) is bounded above by

$$
\|\gamma\|_\infty\int\Big\{\int f(x+s)\,\mathrm{d}x\Big\}P\big(\|Y_0\|\ge n^{1/2}c^{-1}|s|\big)\,\mathrm{d}s=2n^{-1/2}c\|\gamma\|_\infty E\|Y_0\|\to 0.
$$

Among the G satisfying (3.7) is the Lebesgue measure $\mathrm{d}G(x)\equiv\mathrm{d}x$, where $\gamma(x)\equiv 1$. For this G, (2.6) and (3.1) are implied by (2.6)(a) and $E|\varepsilon|<\infty$, and the $\psi(x)$ of Theorem 2.1 equals to $F(x)$, where F is the d.f. of ε, so that $\mathrm{Var}(\psi(\varepsilon)=\mathrm{Var}(F(\varepsilon))=1/12$. Moreover,

$$
\begin{aligned}
K_h^+(t)\quad&=\quad n^{-1}\sum_{k=1}^p\sum_{i=1}^n\sum_{j=1}^n h_k(X_{i-1})h_k(X_{j-1})\Big[\big|X_i+X_j-(Y_{i-1}+Y_{j-1})'t\big|\\
&\qquad -\big|X_i-X_j-(Y_{i-1}-Y_{j-1})'t\big|\Big],
\end{aligned}
$$

and the corresponding ρ_h^+, denoted by $\rho_{h,HL}^+$, is a robust analog of the Hodges–Lehmann type estimator, when h is bounded. Note that for bounded h, (2.2) is implied by (3.8). Because of the importance of this class of estimators we summarize their asymptotic normality result in the following corollary.

Corollary 3.2 *Assume the stationary AR(p) model (1.1) holds. In addition, suppose h is bounded and satisfies (2.1), and the error d.f. F is symmetric around zero and satisfies $\int f^2(x)\,\mathrm{d}x<\infty$, and $E|\varepsilon|<\infty$. Then, (3.8) holds, and*

$$
n^{1/2}(\rho_{h,HL}^+-\rho)\to_d\mathcal{N}\Big(0,\frac{\mathcal{B}^{-1}\mathcal{H}\mathcal{B}^{-1}}{12(\int f^2(x)\,\mathrm{d}x)^2}\Big),
$$

where $\mathcal{B}:=Eh(Y_0)Y_0'$ and $\mathcal{H}:=Eh(Y_0)h(Y_0)'$.

Perhaps it is worth emphasizing that none of the above mentioned literature dealing with the various estimators in AR(p) models with infinite error variance include this class of estimators.

It is thus apparent from the above discussion that asymptotic normality holds for some members of the above class of m.d. estimators without requiring finiteness of any moments, and for some other members requiring only the first error moment to be finite. If one still does not wish to assume (3.8), then it may be possible to verify (3.6) for some heavy tail error densities. We do not do this but now will give an example of a large class of strictly stationary processes satisfying (1.1) and for which this condition holds but which has infinite variance.

Recall that a d.f. F of the error variable ε is said to have a Pareto-like tails of index α if for some $\alpha > 0$, $0 \le a \le 1$, $0 < C < \infty$,

$$x^{\alpha}(1-F(x)) \to aC, \quad x^{\alpha}F(-x) \to (1-a)C, \quad x \to \infty. \tag{3.9}$$

From Brockwell and Davis [2], p. 537, Proposition 13.3.2, it follows that if $1-\rho_1 x - \rho_2 x^2 - \cdots - \rho_p x^p \neq 0$, $|x| \le 1$, and if F satisfies (3.9), then $\{X_i\}$ satisfying (1.1) exists and is strictly stationary and invertible.

Now, (3.9) readily implies $x^{\alpha}P(|\varepsilon| > x) \to C$, as $x \to \infty$, and hence $E|\varepsilon|^{\delta} < \infty$, for $\delta < \alpha$, $E|\varepsilon|^{\delta} = \infty$, for $\delta \ge \alpha$. Suppose $1 < \alpha < 2$. Then $E|\varepsilon| < \infty$, and $\mathrm{Var}(\varepsilon) = \infty$. Thus we have a large class of strictly stationary $\mathrm{AR}(p)$ processes with finite first moment and infinite variance. In particular these processes satisfy (3.8). We summarize the above discussion in the following corollary.

Corollary 3.3 *Assume the autoregressive model (1.1) holds with the error d.f. F having Pareto-like tail of index $1 < \alpha < 2$. In addition, suppose (2.1) holds, G has a bounded Lebesgue density, h is bounded, F has square integrable Lebesgue density, and both F and G are symmetric around zero. Then, the conclusion of Theorem 2.1 holds for the class of m.d. estimators ρ_h^+.*

This still leaves open the problem of obtaining asymptotic distribution of a suitably standardized ρ_h^+ when a stationary solution to (1.1) exists with the error d.f. having Pareto-like tail of index $\alpha \le 1$.

4. M.D. estimators when F is not symmetric

Here we shall describe an asymptotic normality result of a class of minimum distance estimators when F may not be symmetric and when in (1.1) error variance may be infinity. Let $R_i(t)$ denote the rank of $X_i - t'Y_{i-1}$ among $X_j - t'Y_{j-1}$, $j = 1, \cdots, n$, $\bar{h}_n := n^{-1}\sum_{i=1}^n h(Y_{i-1})$, and define the randomly weighted empirical process of residual ranks

$$\begin{aligned} Z_h(t,u) &:= n^{-1/2}\sum_{i=1}^n (h(Y_{i-1}) - \bar{h}_n)[I(R_i(t) \le nu) - u], \quad u \in [0,1], \\ \mathcal{K}_h(t) &:= \int_0^1 \|Z_h(t,u)\|^2 \, \mathrm{d}L(u), \quad \widetilde{\rho}_h := \mathrm{argmin}\{\mathcal{K}_h(t);\, t \in \mathbb{R}^p\}, \end{aligned}$$

where L is a d.f. on $[0,1]$. See Koul [17] for a motivation on using the dispersion $\mathcal{K}_h$. It is an analog of the classical Cramér – von Mises statistic useful in regression and autoregressive models. The following proposition describes the asymptotic normality of $\widetilde{\rho}_h$.

Proposition 4.1 *Assume the process satisfying (1.1) is strictly stationary with the error d.f. F having uniformly continuous Lebesgue density f and finite first* *moment. In addition, assume L is a d.f. on $[0,1]$, (3.5) holds, and the following hold with* $\bar{Y}_{n-1} := n^{-1}\sum_{i=1}^{n} Y_{i-1}$.

(4.1) *Either* $e'(h(Y_{i-1}) - \bar{h}_n)(Y_{i-1} - \bar{Y}_{n-1})'e \geq 0$,
Or $e'(h(Y_{i-1}) - \bar{h}_n)(Y_{i-1} - \bar{Y}_{n-1})'e \leq 0, \quad \forall\, i = 1, \cdots, n,\ e \in \mathbb{R}^p,\ \|e\| = 1.$

Let $F^{-1}(u) := \inf\{x; F(x) \geq u\},\ q(u) := f(F^{-1}(u)),\ 0 \leq u \leq 1$. *Then,*

$$rn^{1/2}(\widetilde{\rho}_h - \rho) = -\left\{\mathcal{C}_n \int_0^1 q^2\,\mathrm{d}L\right\}^{-1} \widetilde{\mathcal{S}}_n + o_p(1),$$

where $\mathcal{C}_n := n^{-1}\sum_{i=1}^{n}(h(Y_{i-1}) - \bar{h}_n)(Y_{i-1} - \bar{Y}_{n-1})'$, *and*

$$\begin{aligned}\widetilde{\mathcal{S}}_n &:= \int_0^1 n^{-1/2}\sum_{i=1}^{n}(h(Y_{i-1}) - \bar{h}_n)\Big\{I(F(\varepsilon_i) \leq u) - u\Big\}q(u)\,\mathrm{d}L(u)\\ &= -n^{-1/2}\sum_{i=1}^{n}(h(Y_{i-1}) - \bar{h}_n)[\varphi(\varepsilon_i) - \int_0^1 \varphi(u)\,\mathrm{d}u], \qquad \varphi(u) := \int_0^u q\,\mathrm{d}L.\end{aligned}$$

Consequently, $n^{1/2}(\widetilde{\rho}_h - \rho) \to_d \mathcal{N}\Big(0, \tau^2\mathcal{C}^{-1}\mathcal{G}\mathcal{C}^{-1}\Big)$, *where* $\tau^2 := Var(\varphi(\varepsilon))/(\int_0^1 q^2\,\mathrm{d}L)^2$ *and*

$$\mathcal{C} := E\big\{\big(h(Y_0) - Eh(Y_0)\big)Y_0'\big\}, \quad \mathcal{G} := E\big(h(Y_0) - Eh(Y_0)\big)\big(h(Y_0) - Eh(Y_0)\big)'.$$

The proof of this claim is similar to that of the asymptotic normality of an analogous estimator $\hat{\theta}_{md}$ discussed in chapter 8 of the monograph by Koul [17] in the case of finite variance, hence not given here. Note that again for bounded h, $\widetilde{\rho}_h$ are robust against innovation outliers.

A useful member of this class is obtained when $L(u) \equiv u$. In this case

$$\begin{aligned}\mathcal{K}_h(t) &= -2n^{-2}\sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}(h_k(Y_{i-1}) - \bar{h}_{nk})(h_k(Y_{j-1}) - \bar{h}_{nk})\big|R_i(t) - R_j(t)\big|,\\ \bar{h}_{nk} &:= n^{-1}\sum_{i=1}^{n} h_k(Y_{i-1}), \quad 1 \leq k \leq p.\end{aligned}$$

In the case of finite variance and when $h(x) \equiv x$, the asymptotic variance of the corresponding estimator is smaller than that of the LAD (Hodges–Lehmann) estimator at logistic (double exponential) errors. It is thus interesting to note that the above asymptotic normality of the robust analogs of this estimator holds even when error variance may be infinite. Note that when $L(u) \equiv u$, the corresponding $\tau^2 = 1/[12(\int f^3(x)\,\mathrm{d}x)^2]$.

Acknowledgement. Authors would like to thank the two anonymous referees for some constructive comments.

References

[1] An, H. Z. and Chen, Z. G. (1982). On convergence of LAD estimates in autoregression with infinite variance. *J. Multiv. Anal.* **12** 335–345.

[2] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Second edition. New York, Springer.

[3] Davis, R. A. and Knight, K. (1987). M-estimation for autoregressions with infinite variance. *Stoc. Proc. & Appl.* **40** 145–180 North-Holland.

[4] Davis, R. A. and Resnick, S. I. (1985). Limit theory for moving average of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195.

[5] Davis, R. A. and Resnick, S. I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533-558.

[6] Davis, R. A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stoc. Proc. Appl.* **40** 1 145–180.

[7] Denby, L. and Martin, D. (1979). Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74** 140–146.

[8] Dhar, S. K. (1991). Minimum distance estimation in an additive effects outliers model. *Ann. Statist.* **19** 205–228.

[9] Dhar, S. K. (1993). Computation of certain minimum distance estimators in AR[k] model. *J. Amer. Statist. Assoc.* **88** 278–283.

[10] Embrecht, P., Klüppelberg, C., and Mikosch, T. (2001). *Modelling extremal events for insurance and finance.* Springer-Verlag, Berlin – Heidelberg.

[11] Gross, S. and Steiger, W. L. (1979). Least absolute deviation estimates in autoregression with infinite variance. *J. Appl. Probab.* **16** 104–116.

[12] Hannan, E. J. and Kanter, M. (1977). Autoregressive processes with infinite variance. *Adv. Appl. Probab.* **6** 768–783.

[13] Kanter, M. and Steiger, W. L. (1974). Regression and autoregression with infinite variance. *Advances in Appl. Probability* **6** 768–783.

[14] Knight, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregression. *J. Time Ser. Anal.* **8** 51–60.

[15] Knight, K. (1993). Estimation in dynamic linear regression models with infinite variance errors. *Econometric Theory* **9** 570–588.

[16] Koul, H. L. (1986). Minimum distance estimation and goodness-of-fit tests in first-order autoregression. *Ann. Statist.* **14** 1194–1213.

[17] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Second edition. New York, Springer.

[18] Ling, S. Q. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J.R. Statist. Soc. B* **67** 381-393.