The Annals of Statistics
1977, Vol. 5, No. 3, 554–560

ASYMPTOTIC BEHAVIOR OF LEAST-SQUARES ESTIMATES FOR AUTOREGRESSIVE PROCESSES WITH INFINITE VARIANCES

BY VICTOR J. YOHAI AND RICARDO A. MARONNA

Universidad Nacional de Buenos Aires and Fundación Bariloche

Let y_t be an order p autoregressive process of the form $y_t + \sum_{s=1}^{p} \beta_s y_{t-s} = u_t$, where the u_t's are i.i.d. variables with a symmetric distribution F such that $E \log^+ |u_t| < \infty$. For the Yule–Walker version $\boldsymbol{\beta}_T^*$ of the least-squares estimate of $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$, it is shown that $T^{\frac{1}{2}}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta})$ is bounded in probability.

1. Introduction. In this paper we consider the order p autoregressive model.

$$y_t + \sum_{s=1}^{p} \beta_s y_{t-s} = u_t \qquad -\infty < t < +\infty \,, \tag{1.1}$$

where the u_t's are independent identically distributed (i.i.d.) random variables with a common distribution function F. In classical theory (Anderson, 1970) it is assumed that $Eu_t^2 < \infty$, and under this hypothesis it is proved that the least squares estimator $\boldsymbol{\beta}_T^*$ of the vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ is consistent; and, moreover, that $T^{\frac{1}{2}}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta})$ converges in law to a multivariate normal distribution.

In the last decade, several authors, especially Mandelbrot (1963 and 1967), have pointed out that some economic data (e.g., stock price changes) may be better represented by time series with infinite variances (see Granger and Orr (1972) for more complete references). This has raised the question of whether the classical estimators are still reliable when variances do not exist. A partial answer was obtained by Kanter and Steiger (1974), who showed the consistency of the least squares estimators when F is symmetric and satisfies

$$\lim_{t\to\infty} t^{\alpha}[1 - F(t)] = k > 0 \tag{1.2}$$

for some $\alpha \in (0, 2)$. It is well known that this condition is satisfied by stable laws, and that if F satisfies (1.2), then it belongs to the domain of attraction of a stable law with characteristic exponent α.

In this paper it is shown, more generally, that if the u_t's are symmetric and satisfy

$$E \log^+ |u_t| < \infty \,, \tag{1.3}$$

then $\boldsymbol{\beta}_T^*$ is consistent; and moreover, $T^{\frac{1}{2}}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta})$ is bounded in probability. It is easy to show that distributions satisfying (1.2) also satisfy (1.3), and hence our result includes the class considered by Kanter and Steiger. The condition

Received September 1975; revised July 1976.

AMS 1970 *subject classifications.* Primary 62M10; Secondary 62E20.

Key words and phrases. Autoregressive processes, least-squares estimates, infinite variance, asymptotic theory.

(1.3) is stated only to ensure the stationarity of the process $\{y_t\}$, and it seems difficult to relax it.

Finally we make some considerations about the order of convergence of the estimator $\boldsymbol{\beta}_T{}^*$.

2. Consistency theorems. Consider the autoregressive process $\{y_t, t \in Z\}$ described in (1.1). In classical theory (Anderson (1970), Section 5.2.1) it is assumed that $Eu_t^2 < \infty$; and under this hypothesis it is proved that if the roots $x_1, \cdots, x_p$ of the "associated equation"

$$x^p + \sum_{k=1}^p \beta^k x^{p-k} \tag{2.1}$$

have all modulus less than one, then there exists a representation of the process in the form

$$y_t = \sum_{r=0}^\infty \delta_r u_{t-r} , \tag{2.2}$$

where the series converges in L_2, and a fortiori, the process y_t is stationary. The sequence $\{\delta_r\}$ is a solution of the difference equation.

$$\beta_r + \sum_{k=1}^p \beta_k \delta_{r-k} = 0 \qquad r = p, p+1, \cdots \tag{2.3}$$

with $\delta_0 = 1$. If (2.1) has no multiple roots, then

$$\delta_r = \sum_{i=1}^p k_i x_i^r \qquad r = 0, 1, 2, \cdots \tag{2.4}$$

where the k_i's are constant coefficents depending on the β_i's.

Now it is desired to ensure the stationarity of $\{y_t\}$ and the validity of (2.2) under conditions less restrictive than $Eu_t^2 < \infty$. It is clear that if the right side of (2.2) converges in probability, then the process $\{y_t\}$ defined by (2.2) and (2.3) is stationary. We shall now consider the convergence of (2.2). From now on it will be assumed without further notice that $|x_i| < 1$ $(i = 1, \cdots, p)$ and that (2.1) has no multiple roots (in the case of multiple roots, all results remain valid, but some modifications must be made in the proofs, related to changes in (2.4)).

LEMMA 1. (a) *In order for the series* $\sum_{r=0}^\infty \delta_r u_{t-r}$ (*where* δ_r *satisfies* (2.4)) *to converge absolutely with probability one, it is sufficient that*:

$$E \log^+ |u_t| < \infty . \tag{2.5}$$

(b) *If the series converges, then it is the* (a.s.) *only stationary process satisfying* (1.1).

PROOF. (a) From (2.4) it is enough to prove that for all x with $|x| < 1$, the series

$$\sum_{r=0}^\infty |u_r|\,|x|^r \tag{2.6}$$

converges a.s. Since (2.6) is a power series, according to Cauchy's root criterion it suffices to prove that $\limsup_{r\to\infty} |u_r|^{1/r} < 1$ a.s.; and hence by the Borel–Cantelli lemma, it is enough to prove that, for every $\varepsilon > 0$:

$$\infty > \sum_{r=0}^\infty P\{|u_r| > (1+\varepsilon)^r\} = \sum_{r=0}^\infty P\{\log^+ |u_1| > r \log(1+\varepsilon)\} ;$$

and the last condition is equivalent to (2.5).

(b) It is easy to show that if the series (2.2) converges, then it defines a stationary process satisfying (1.1). The unicity is proved as in Theorem 5.2.1 of Anderson (1970), replacing convergence in L_2 by convergence in probability.

Consider now a sequence of observations y_t with $t = 0, 1, \cdots, T$. We shall define the modified least-squares estimator (Yule–Walker estimator) $\boldsymbol{\beta}_T{}^*$ of the vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ in a form similar to equations 5.4.24—5.4.26 of Anderson (1970). For $h = 0, 1, \cdots, p$ let

$$C^*_{hT} = \textstyle\sum_{t=h}^{T} y_t y_{t-h} \tag{2.7}$$

$$r^*_{hT} = C^*_{hT}/C^*_{0T}\,, \tag{2.8}$$

and define $\mathbf{r}_T{}^*$ as the column vector with coordinates $(r^*_{1T}, \cdots, r^*_{pT})$. For any p-dimensional vector $\mathbf{r}$, define $\mathbf{R}(\mathbf{r})$ as the $p \times p$-matrix with (i, j)-elements equal to $r_{|i-j|}$. Put $\mathbf{R}_T{}^* = \mathbf{R}(r_T{}^*)$. Then the Yule–Walker estimator $\boldsymbol{\beta}_T{}^*$ is defined as the solution of

$$\mathbf{R}_T{}^*\boldsymbol{\beta}_T{}^* = -\mathbf{r}_T{}^*\,. \tag{2.9}$$

It is well known (Theorem 5.5.7 of Anderson (1970)) that in the case in which $Eu_t^2 < \infty$, the sequence $T^{\frac{1}{2}}(\boldsymbol{\beta}_T{}^* - \boldsymbol{\beta})$ has a limit normal distribution, with mean $\mathbf{0}$ and a certain covariance matrix $\mathbf{W}$ which depends on $\boldsymbol{\beta}$ but not on F.

Now we are ready to state our main result.

THEOREM 1. *Let the process* $\{y_t\}$ *verify equation* (1.1), *where the* u_t*'s are* i.i.d. *with a symmetric distribution* F *verifying* (2.5) *and not concentrated at* 0. *Then* $T^{\frac{1}{2}}(\boldsymbol{\beta}_T{}^* - \boldsymbol{\beta})$ *is bounded in probability.*

The theorem will be proved by showing that, for $T \to \infty$

$$T^{\frac{1}{2}}(\boldsymbol{\beta}_T{}^* - \boldsymbol{\beta}) = \mathbf{b}_T + o_p(1)\,, \tag{2.10}$$

where $E\mathbf{b}_T = \mathbf{0}$; the covariance matrix of $\mathbf{b}_T$ is of the form

$$\operatorname{Var}(\mathbf{b}_T) = (1 - \gamma_T)\mathbf{W}_T\,, \tag{2.10$'$}$$

where $\mathbf{W}_T$ depends on $\boldsymbol{\beta}$ but not on F; $\mathbf{W} = \lim_{T\to\infty} \mathbf{W}_T$ exists; and the sequence γ_T is defined as

$$\gamma_T = E\textstyle\sum_{t=0}^{T} u_t^4/(\sum_{t=0}^{T} u_t^2)^2\,, \tag{2.11}$$

and hence $0 \leqq \gamma_T \leqq 1$.

The proof of the theorem will be broken up in several lemmas. The hypothesis of the theorem will be assummed throughout.

LEMMA 2. *Define for* $h = 0, 1, \cdots, p$

$$r_h = C_h/C_0 \tag{2.12}$$

where

$$C_h = \textstyle\sum_{r=0}^{\infty} \delta_r\delta_{r+h} \tag{2.13}$$

where the δ_r*'s are the coefficients in* (2.2). *Define* $\mathbf{r}$ *as the vector with coordinates*

$(r_1, \cdots, r_p)$, *and put* $\mathbf{R} = \mathbf{R}(\mathbf{r})$. *Then* $\mathbf{R}$ *is nonsingular and*

$$\mathbf{R}\boldsymbol{\beta} = -\mathbf{r}\,. \tag{2.14}$$

PROOF. To prove (2.14), assume for a moment that $Eu_t^2 < \infty$. Then the correlation coefficient between y_t and y_{t+h} is equal to r_h as defined in (2.12), and it is then easy to verify (2.14) (the so-called Yule–Walker equations, in Section 5.2.2 of Anderson (1970)). But since the r_i's depend only on the δ's, which in turn depend only on $\boldsymbol{\beta}$, it results that relations (2.12)—(2.14) are valid independent of F.

The fact that $\mathbf{R}$ is nonsingular follows from Lemma 5.5.5 of Anderson (1970). This completes the proof.

LEMMA 3. *Define for* $h = 0, 1, \cdots, p$ *and* $i, j = 1, \cdots, p$ *and* $T \geqq 1$

$$g_{hijT} = \sum_{r=h+1}^{T} \sum_{s=1}^{T-h} u_r u_s x_i^{-r} x_j^{-(s+h)} (x_i x_j)^{\max(r,s+h)} I(r \neq s) \tag{2.15}$$

and

$$S_T = \sum_{t=0}^{T} u_t^2\,.$$

Let g_T *be the vector with coordinates* g_{hijT}. *Then for each* T

$$\operatorname{Var}(T^{\frac{1}{2}}\mathbf{g}_T/S_T) = (1-\gamma_T)\mathbf{V}_T \tag{2.16}$$

where γ_T *is defined in* (2.11) *and* $\mathbf{V}_T$ *depends only on* $\boldsymbol{\beta}$. *Moreover,* $\mathbf{V} = \lim_{T\to\infty} \mathbf{V}_T$ *exists.*

PROOF. In what follows the subscript T will be generally dropped from $\mathbf{g}$ and S. Note first that the Cauchy–Schwarz inequality implies that g_{hij}/S is bounded, so that second moments exist. The symmetry of the u_t's implies that $E(\mathbf{g}/S) = \mathbf{0}$ for all T.

To calculate the covariance matrix, note first that, if $r \neq s$ and $r' \neq s'$ then $E(u_r u_s u_{r'} u_{s'}/S^2)$ is nonnull only when $r = r'$ and $s = s'$, or $r = s'$ and $s = r'$. Hence the elements of Var $(\mathbf{g}/S)$ are

$$\begin{aligned} &E((g_{hij}/S)(g_{h'i'j'}/S)) \\ &\quad = \sum_{r=h+1}^{T} \sum_{s=1}^{T-h} (x_i x_{i'})^{-r} (x_j x_{j'})^{-(s+h)} (x_i x_j x_{i'} x_{j'})^{\max(r,s+h)} \\ &\qquad \times I(r \neq s) E(u_r^2 u_s^2/S^2)\,. \end{aligned} \tag{2.17}$$

Now by the interchangeability of the u's,

$$E(u_r^2 u_s^2/S^2) = (1-\gamma_T)/[T(T-1)]\,.$$

This proves (2.16). The convergence of $\mathbf{V}_T$ follows easily upon calculating the double geometric sum in (2.17).

LEMMA 4. *Under the same hypothesis as in Theorem* 1

$$T^{\frac{1}{2}}(r_h^* - r_h) = C_0^{-2} \sum_{i=1}^{p} \sum_{j=1}^{p} (1 - x_i x_j) k_i k_j T^{\frac{1}{2}} (C_0 g_{hij} - C_h g_{0ij})/S + O_p(1)\,.$$

PROOF. From the definition we may write

$$r_h^* - r_h = (C_0 C_0^*/S)^{-1}[C_0(C_h^*/S - C_h) - C_h(C_0^*/S - C_0)]\,. \tag{2.18}$$

Now (2.7), (2.2) and (2.4) yield

$$C_{hT}^{*} = \sum_{i=1}^{p} \sum_{j=1}^{p} k_i k_j D_{hij} \tag{2.19}$$

with

$$D_{hij} = \sum_{t=h}^{T} [\sum_{r=0}^{\infty} x_i^r u_{t-r}][\sum_{s=0}^{\infty} x_j^s u_{t-h-s}] . \tag{2.20}$$

It is easy to verify that

$$C_h = \sum_{i=1}^{p} \sum_{j=1}^{p} k_i k_j x_j^h/(1 - x_i x_j) .$$

Then by (2.19),

$$C_h^*/S - C_h = \sum_{i=1}^{p} \sum_{j=1}^{p} k_i k_j [D_{hij}/S - x_i^h/(1 - x_i x_j)] . \tag{2.21}$$

Rearrangement of the order of summation in (2.20) yields

$$\begin{aligned}(1 - x_i x_j)D_{hij} \\ &= \sum_{r=-\infty}^{T} \sum_{s=-\infty}^{T-h} u_r u_s x_i^{-r} x_j^{-(s+h)}[(x_i x_j)^{\max(h,r,s+h)} - (x_i x_j)^{T+1}] .\end{aligned} \tag{2.22}$$

Now it will be shown that

$$(1 - x_i x_j)D_{hij} = x_i^h \sum_{r=h+1}^{T-h} u_r^2 + g_{hij} + O_p(1) . \tag{2.23}$$

First note that

$$\sum_{r=-\infty}^{T} \sum_{s=-\infty}^{T-h} u_r u_s x_i^{-r} x_j^{-(s+h)}(x_i x_j)^{T+1} = O_p(1)$$

and hence the last term in brackets in (2.22) is already dealt with. Now decompose the remainder of (2.22) as

$$\sum_{r=-\infty}^{h} \sum_{s=-\infty}^{0} + \sum_{r=h+1}^{T} \sum_{s=-\infty}^{0} + \sum_{r=h+1}^{T} \sum_{s=1}^{T-h} . \tag{2.24}$$

Taking into account the absolute convergence of (2.6), it is easy to prove that the first two terms of (2.24) are $O_p(1)$. Finally, the last term of (2.24) is obviously equal to the right-hand side of (2.23). Hence (2.21) and (2.23) yield

$$C_h^*/S - C_h = \sum_{i=1}^{p} \sum_{j=1}^{p} (1 - x_i x_j)^{-1} g_{hij}/S + O_p(1)/S . \tag{2.25}$$

The desired result follows upon inserting (2.25) into (2.18) and recalling that the law of large numbers implies that $\lim_{T\to\infty} S/T^{\frac{1}{2}} = \infty$ a.s.

LEMMA 5. *Under the same hypothesis as Theorem* 1,

$$T^{\frac{1}{2}}(\mathbf{r}^* - \mathbf{r}) = \mathbf{a}_T + o_p(1) , \tag{2.26}$$

where $E\mathbf{a}_T = 0$ *and*

$$\operatorname{Var}(\mathbf{a}_T) = (1 - \gamma_T)\mathbf{U}_T , \tag{2.27}$$

where $\mathbf{U}_T$ *depends on* $\boldsymbol{\beta}$ *but not on* F; *and there exists a matrix* $\mathbf{U}$ *with*

$$\mathbf{U} = \lim_{T\to\infty} \mathbf{U}_T . \tag{2.28}$$

PROOF. Immediate from Lemmas 3 and 4.

PROOF OF THEOREM 1. Now to prove (2.10), define the function φ of R^p into

R^p as $\varphi(\mathbf{z}) = -\mathbf{R}(\mathbf{z})^{-1}\mathbf{z}$ ($\mathbf{z} \in R^p$), where the matrix function $\mathbf{R}$ is defined after (2.8). It follows from Lemmas 4 and 2 that for T large enough, $\mathbf{R}^*$ is nonsingular, and hence $\varphi(\mathbf{r}_T^*)$ is well defined, and φ is differentiable at $\mathbf{r}_T^*$. Hence $\boldsymbol{\beta}^* = \varphi(\mathbf{r}^*)$ and $\boldsymbol{\beta} = \varphi(\mathbf{r})$.

Call D the derivate operator of φ at the point $\mathbf{r}$. Then by the mean value theorem

$$\boldsymbol{\beta}^* - \boldsymbol{\beta} = \mathbf{D}(\mathbf{r}^* - \mathbf{r}) + o_p(\mathbf{r}^* - \mathbf{r}) .$$

Hence by (2.26)

$$T^{\frac{1}{2}}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) = \mathbf{D}\mathbf{a}_T + o_p(1)$$

and application of Lemma 5 completes the proof.

It seems difficult to give in general a more precise estimate of the order of convergence of $\boldsymbol{\beta}_T^*$ to $\boldsymbol{\beta}$. For the case in which F satisfies (1.2), it has been pointed out by a referee that the proof of Theorem 3.1 of Kanter and Steiger (1974) implies that

$$T^{1/\delta}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta}) = o_p(1) \tag{2.29}$$

for any $\delta > \alpha$. For the general situation, some information may be obtained from the limit behavior of the covariance matrix of the random vector $\mathbf{b}_T$ in (2.20), which is $(1 - \gamma_T)\mathbf{W}_T$. As respects $\mathbf{W}_T$, it was shown that it converges to a matrix $\mathbf{W}$, which depends only on $\boldsymbol{\beta}$. It will be shown in Theorem 2 that this $\mathbf{W}$ coincides with the covariance matrix of the limit normal distribution of $\boldsymbol{\beta}^*$ when $Eu_t^2 < \infty$. As to the constants γ_T, the result (2.29) implies that $\lim \gamma_T = 1$ when F satisfies (1.2); and it is easy to show that $\lim \gamma_T = 0$ when F has finite variance. It would be interesting to find distributions F such that $\lim \gamma_T \in (0, 1)$.

THEOREM 2. *Let* $\mathbf{W} = \lim_{T\to\infty} \mathbf{W}_T$, *where* $\mathbf{W}_T$ *is defined in* (2.10′). *If* $Eu_t^2 < \infty$, *then the distribution of* $\mathbf{b}_T$ *converges to* $N(\mathbf{0}, \mathbf{W})$.

PROOF. Let $\mathbf{g}_T$ and $\mathbf{V}$ be the vector and matrix defined in the statement of Lemma 3. It will be first shown that when $Eu_t^2 = \sigma^2 < \infty$, $T^{\frac{1}{2}}\mathbf{g}_T/S_T$ converges in law to $N(\mathbf{0}, \mathbf{V})$. Since $\lim_{T\to\infty} S/T = \sigma^2$ (a.s.), it suffices to deal only with $g_{hij}/(\sigma T)$ in order to prove the asymptotic normality of $\mathbf{g}/S$.

For fixed k, define

$$g_{hij}^k = \sum_{r=h+1}^{T} \sum_{s=1}^{T-h} u_r u_s x_i^{-r} x_j^{-(s+h)} (x_i x_j)^{\max(r,s+h)} I(r \neq s) I(|r - s| \leqq k) .$$

Let $\mathbf{V}_T{}^k$ be the covariance matrix of $T^{\frac{1}{2}}\{g_{hij}^k\}$. Using Theorem 7.7.5 of Anderson (1970), it may be shown that when $T \to \infty$, $T^{\frac{1}{2}}g_{hij}^k/(\sigma T)$ converges in law to $N(\mathbf{0}, \mathbf{V}^k)$, where $\mathbf{V}^k = \lim_{T\to\infty} \mathbf{V}_T{}^k$. Besides, it is easy to show that

$$\lim_{k\to\infty} \sup_T TE(g_{hij} - g_{hij}^k)^2 = 0 .$$

Then Theorem 7.7.1 of Anderson (1970) yields that $T^{\frac{1}{2}}\mathbf{g}/(\sigma T)$ converges in law to $N(\mathbf{0}, \mathbf{V}_*)$, where $\mathbf{V}_* = \lim_{k\to\infty} \mathbf{V}^k$. Straightforward and tedious calculations show that $\mathbf{V}_* = \mathbf{V}$.

Then, from Lemmas 3 and 4 it is easy to show that the vector $\mathbf{a}_T$ in Lemma 5 is asymptotically $N(\mathbf{0}, \mathbf{U})$, where $\mathbf{U}$ is defined in (2.28). Finally, the same reasoning as in the final proof of Theorem 1 completes the proof.

Acknowledgment. We thank the referee for having pointed out to us the results contained in the paper by Kanter and Steiger.

REFERENCES

ANDERSON, T. W. (1970). *Statistical Analysis of Time Series*. Wiley, New York.

GRANGER, C. W. J. and ORR, D. (1972). Infinite variance and research strategy in time series analysis. *J. Amer. Statist. Assoc.* **67** 275–285.

KANTER, M. and STEIGER, W. L. (1974). Regression and autoregression with infinite variances. *Advances in Appl. Probability* **6** 768–783.

MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Business* **36** 394–419.

MANDELBROT, B. (1967). The variation of some other speculative prices. *J. Business* **40** 393–413.

DEPARTAMENTO DE MATEMATICAS
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD NACIONAL DE BUENOS AIRES
PABELLON 1, CIUDAD UNIVERSITARIA
BUENOS AIRES, ARGENTINA