The Annals of Statistics
1986, Vol. 14, No. 3, 1180–1193

A MINIMUM DISTANCE ESTIMATOR FOR FIRST-ORDER AUTOREGRESSIVE PROCESSES

By Chamont W. H. Wang[1]

University of Montana

In this paper we construct a class of minimum distance Cramér–von Mises-type estimators for the parameter in the first-order stationary autoregressive time series. The estimator is proved to be asymptotically normal under appropriate assumptions. The proofs involve some results of independent interest.

1. Introduction. Consider the first-order stationary autoregressive model

$$X_k = \beta X_{k-1} + U_k, \qquad |\beta| < 1, \quad -\infty < k < \infty, \tag{1.1}$$

where $\{U_k\}$ are independent and identically distributed random shocks with $E(U) = 0$ and $0 < \sigma^2 = \mathrm{Var}(U) < \infty$. This model has been widely used in applications for forecasting and control [see Box and Jenkins (1976)]. The constant β in (1.1) is the unknown parameter which we would like to estimate. Based on the observations $\{X_0, X_1, \dots, X_n\}$, the commonly used least-squares estimate of β is given by

$$\hat{\beta}_{\mathrm{LS}} = \sum_{k=1}^{n} X_{k-1} X_k \Big/ \sum_{k=1}^{n} X_{k-1}^2. \tag{1.2}$$

It can be proved [see, e.g., Akritas and Johnson (1980)] that, within a wide class of density functions f of U, the best attainable asymptotic variance for the estimates of β under model (1.1) is

$$(1 - \beta^2)\sigma^{-2}(I(f))^{-1}, \tag{1.3}$$

where $I(f)$ is the Fisher information of f. Note that the asymptotic variance of the least-squares estimate is $1 - \beta^2$ [Theorem 4.3, Anderson (1959)] which equals (1.3) for Gaussian disturbances but not for more general situations [see, e.g., Hájek and Šidák (1967), pages 16 and 17].

Akritas and Johnson proposed an alternative test statistic to compete with the normal theory least squares tests under the contiguity framework. Lai and Siegmund (1983) showed that the asymptotic normality of $\hat{\beta}_{\mathrm{LS}}$ is not especially good when $n = 50$, even for small $|\beta|$ and normal shocks, and it deteriorates quite noticeably for β near 1. Lai and Siegmund proposed an estimator under a sequential sampling scheme to improve the fixed sample size least-squares estimator.

In this paper we consider a minimum distance Cramer–von Mises-type estimator of β under model (1.1). This approach has been successful in the usual

Received March 1984; revised December 1985.

[1] Now at Trenton State College.

AMS 1980 *subject classification*. Primary 62L10.

Key words and phrases. Randomly weighted empirical processes, stationary and ergodic processes, bounded functionals on L^2-spaces, Lipschitz condition, conditional expectation.

regression model. See, for example, Williamson (1982) and Millar (1982). Millar considered a general regression model and proved that his estimator of the regression function is asymptotically normal, efficient, and qualitatively robust against certain contaminated errors. Veitch (1983) considered an estimator $\hat{\theta}$ which minimizes the distance (in a certain Hilbert space) between the empirical distribution $\hat{F}$ and the true distribution F^{θ} of a general stationary time series $\{X_0, X_1, \dots, X_n\}$. The estimator is proved to be asymptotically normal and qualitatively robust. Veitch's result includes all Gaussian ARMA models with a finite number of parameters.

Instead of assuming the knowledge of the error distribution F, we propose a nonparametric estimator for the first-order autoregressive processes. The motivation of this estimator is as follows. Consider the randomly weighted empirical process

$$S(t, \Delta) = \sum_{k=1}^{n} X_{k-1} I(X_k \le t + \Delta X_{k-1}) \tag{1.4}$$

where $I(A)$ is the indicator function of set A. Note that, by the completeness of L^2-spaces, (1.1) can be rewritten as [see, e.g., Fuller (1976), page 29]

$$X_k = \sum_{j=0}^{\infty} \beta^j U_{k-j} = \beta^k X_0 + \sum_{j=0}^{k-1} \beta^j U_{k-j}. \tag{1.5}$$

Hence X_{k-1} and U_k are independent, $E(X_k) = 0$, for all k, and $ES(t, \beta) = 0$. Furthermore, by denoting

$$F(t) = P(U_k \le t), \qquad G(t) = P(X_k \le t) \tag{1.6}$$

for all k, we have

$$ES(t, \Delta) = n \int x F(t + (\Delta - \beta)x)\, dG(x). \tag{1.7}$$

We now claim

$$ES(t, \Delta) \ge 0 \quad \text{if } \Delta > \beta \quad \text{and} \quad ES(t, \beta) \le 0 \quad \text{if } \Delta < \beta. \tag{1.8}$$

To prove this, consider the case $\Delta > \beta$ first. Note that

$$x > 0 \text{ implies } F(t + (\Delta - \beta)x) \ge F(t) \text{ and}$$
$$xF(t + (\Delta - \beta)x) \ge xF(t),$$
$$x < 0 \text{ implies } F(t + (\Delta - \beta)x) \le F(t) \text{ and}$$
$$xF(t + (\Delta - \beta)x) \ge xF(t).$$

This completes the proof of the first half of (1.8). The second part is done by the same argument. Similarly, we can prove that $S(t, \Delta)$ is nondecreasing in Δ for any fixed t. Hence, if we define

$$Q(\Delta) = \int_{-\infty}^{\infty} S^2(t, \Delta)\, dH(t), \tag{1.9}$$

where H is a finite measure on $(R, \mathscr{B})$, then similar to a rank test for testing H_0:

$\beta = \beta_0$ [Hájek and Šidák (1967), page 103], H_0 is rejected for large values of $Q(\beta_0)$. Therefore it is reasonable to define an estimator $\hat{\beta}$ for β by the relation

$$Q(\hat{\beta}) = \inf_{\Delta} Q(\Delta). \tag{1.10}$$

In Section 2 we obtain the asymptotic quadratic approximation of $Q(\Delta)$ and then use this result to prove the asymptotic normality of $\hat{\beta}$. We also discuss a method to find H to minimize $\mathrm{Var}(\hat{\beta})$.

Another nonparametric minimum distance estimator was proposed by Koul (1985). First note that by using an extra assumption that the error distribution F is symmetric around 0, $F(t)$ can be estimated either by $n^{-1}\Sigma I[X_k \le t + \beta X_{k-1}]$ or by $n^{-1}\Sigma I[-X_k < t - \beta X_{k-1}]$. Koul then considered an estimator of β as a Δ which minimizes

$$\int \left[\sum X_{k-1}\{I(X_k \le t + \Delta X_{k-1}) - I(-X_k < t - \Delta X_{k-1})\}\right]^2 dH(t), \tag{1.11}$$

where H can be a σ-finite measure. For the asymptotic behavior of this estimator, see Koul (1985).

2. Main results. The following Proposition 1 concerns the asymptotic quadraticity of $Q(\Delta)$. The result will be used to study the asymptotic distribution of $\hat{\beta}$.

PROPOSITION 1. *If $I(f) < \infty$, H is a finite measure and $E|U|^5 < \infty$, then*

$$E\left\{\sup_{\sqrt{n}|\Delta - \beta| \le a} n^{-1}|Q(\Delta) - Q_1(\Delta)|\right\} \to 0 \quad \text{as } n \to \infty, \tag{2.1}$$

where a is a fixed number and

$$Q_1(\Delta) =: \int \left(S(t,\beta) + n(\Delta - \beta)\sigma_x^2 f(t)\right)^2 dH(t), \tag{2.2}$$

$$\sigma_x^2 =: \mathrm{Var}(X) = \sigma^2/(1 - \beta^2). \tag{2.3}$$

PROOF. See Section 3.

REMARKS ON THE ASSUMPTIONS. $I(f) < \infty$ implies

$$f(x) \to 0 \quad \text{as } x \to \pm\infty \tag{2.4}$$

[Hájek and Šidák (1967), page 20]. This in turn implies

$$f \text{ is bounded,} \tag{2.5}$$

since f is continuous, and

$$F \text{ satisfies a Lipschitz condition} \tag{2.6}$$

[Royden (1968), page 108] and

$$\eta(z) = \sup_y (f(y) - f(y - z))^2 \text{ is continuous at } 0, \tag{2.7}$$

where (2.7) holds by using (2.4) for large y and by using the fact that f is continuous and hence uniformly continuous on compact sets. (2.4)–(2.7) are crucial in the proof of (2.1). Indeed, we can replace $I(f) < \infty$ by the weaker assumptions (2.4) and (2.5) for the rest of the proof. We state these as assumptions:

$$f(x) \to 0 \quad \text{as } x \to \pm\infty, \tag{A.1}$$

$$f \text{ is continuous.} \tag{A.2}$$

Note that $Q_1(\Delta)$, as defined in (2.2), is a quadratic function and has its unique minimum at $\hat{\Delta}$, where $\hat{\Delta}$ is defined by

$$\begin{aligned}\sqrt{n}(\hat{\Delta} - \beta) &= \left(-1/\sigma_X^2 n^{1/2} \int f^2\,dH\right) \int S(t, \beta) f(t)\,dH(t) \\ &= \left(-1/\sigma_X^2 n^{1/2} \int f^2\,dH\right) \sum_{k=1}^{n} X_{k-1} \int I(U_k \le t) f(t)\,dH(t).\end{aligned} \tag{2.8}$$

Thus we conclude:

PROPOSITION 2. *Under the assumptions of Proposition* 1,

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \left(-1/\sigma_X^2 n^{1/2} \int f^2\,dH\right) \\ &\quad \times \sum_{k=1}^{n} X_{k-1} \int I(U_k \le t) f(t)\,dH(t) + o_p(1).\end{aligned} \tag{2.9}$$

PROOF. See Section 3.

THEOREM 1. *Assume that* (A.1) *and* (A.2) *hold, that* $E|U|^5 < \infty$, *and that* H *is a finite measure. Then* $\sqrt{n}(\hat{\beta} - \beta)$ *is asymptotically normal with mean* 0 *and asymptotic variance* $\sigma^{-2}(\int f^2\,dH)^{-2}\{(1 - \beta^2)E(\tilde{U}^2) + 2\beta(1 + \beta)E^2(\tilde{U})\}$, *where* $\tilde{U} = \int I(U \le t) f(t)\,dH(t)$.

Theorem 1 is a by-product of the following more general Lemma 1 and Theorem 2.

LEMMA 1. *For the model* (1.1), *let*

$$T_n = \sum_{k=1}^{n} X_{k-1} \tilde{U}_k, \tag{2.10}$$

where $\tilde{U}_k = \zeta(U_k)$, *while* ζ *is a measurable function such that* $E(\tilde{U}_k^2) < \infty$. *Then*

$$\lim_n n^{-1} E(T_n^2) = \sigma_X^2 \left\{E(\tilde{U}^2) + 2(E\tilde{U})^2 \beta/(1 - \beta)\right\}. \tag{2.11}$$

PROOF. See Section 3.

APPLICATIONS OF LEMMA 1. (a) If $\tilde{U} = I(U \le t)f(t)\,dH(t)$, then, by (2.9),

$$\operatorname{Var}(\sqrt{n}\,\hat{\beta}) = \sigma_x^{-2}\left(\int f^2\,dH\right)^{-2}\{E(\tilde{U}^2) + 2E^2(\tilde{U})\beta/(1-\beta)\} + o(1)$$

$$= \sigma^{-2}\left(\int f^2\,dH\right)^{-2}\{(1-\beta^2)E(\tilde{U}^2) + 2\beta(1+\beta)E^2(\tilde{U})\} + o(1),$$

since $(1-\beta^2)\sigma_X^2 = \sigma^2$. This gives the asymptotic variance of $\hat{\beta}$ in Theorem 1.

(b) Let $\tilde{U} = \int[I(U \le t) - I(-U < t)]f(t)\,dH(t)$. This yields (3.15) of Koul (1985).

(c) If $\tilde{U} = \int[I(U \le t) - I(-U < t)][f(t) + f(-t)]\,dH(t)$, then Lemma 1 yields ν^2 of Remark 5, Section 3 of Koul (1985).

(d) If $\tilde{U} = U$, then $E(\tilde{U}) = 0$, $E(\tilde{U}^2) = \sigma^2$, and $\tau^2 = \sigma_X^2\sigma^2 = \sigma^4/(1-\beta^2)$ [see Anderson (1959), Theorem 4.1].

(e) If $\tilde{U} \equiv 1$, then $E(\sqrt{n}\,\bar{X})^2 = n^{-1}E(T_n^2) \to \sigma_X^2(1 + 2\beta/(1-\beta)) = \sigma^2(1-\beta)^{-2}$ [see, e.g., Theorem 6.3.3 of Fuller (1976)].

(f) If $\tilde{U} = I[U \le t]$, then $E(\tilde{U}^2) = E(\tilde{U}) = F(t)$, and

$$n^{-1}ES^2(t,\beta) \to \sigma_X^2\beta(1-\beta)^{-1}F(t)(1-\beta+2\beta F(t)).$$

Note $\int F(t)(1-\beta+2\beta F(t))\,dt = \infty$ for any $-1 < \beta < 1$, since $F(t) \to 1$ as $t \to \infty$. Hence we confine ourselves to finite measure H. See the proof of Lemma 3.1.

The following Theorem 2 will be used to establish the asymptotic normality of $\hat{\beta}$ as well as that of some other commonly used estimators for β. Note also that the conclusion of Theorem 2 holds for all stationary ARMA (p, q) processes and for any $\tilde{X}_{k-1}$ where $\tilde{X}_{k-1}$ is a measurable function of X_{k-1} such that $E(\tilde{X}_{k-1}^2) < \infty$. For the sake of simplicity, we present the result and proof for the AR (1) model only.

THEOREM 2. *Let* $T_n = \sum X_{k-1}\tilde{U}_k$, *as defined in* (2.10), *and let*

$$\tau_n^2 = n^{-1}ET_n^2, \qquad \tau^2 = \lim_{n\to\infty}\tau_n^2. \tag{2.12}$$

Then under model (1.1), $T_n/\sqrt{n}\,\tau_n$ *converges in distribution to the standard normal law.*

REMARKS. (a) If $\tilde{U}_k = \int I(U_k \le t)f(t)\,dH(t)$, then Theorem 2 yields the asymptotic normality of $\hat{\beta}$ and $\hat{\Delta}$, see (2.8) and (2.9).

(b) For any t, let $\tilde{U}_k = I(U_k \le t)$. Then $T_n = S(t,\beta)$ in (1.4).

(c) If $\tilde{U} = U$, then Theorem 2 yields the asymptotic normality of the least-squares estimator [see Theorem 4.1, Anderson (1959)]. Note also in the least-squares case, $\{T_n\}$ is a martingale, while in our case it is unfortunately not.

(d) By a discussion analogous to that of (e), (b), and (c) in the applications of Lemma 1, we conclude the asymptotic normality of sample average [see, e.g., Theorem 6.3.3 of Fuller (1976)] and of Koul's estimators for symmetric and nonsymmetric F [see (3.5) and (3.1) of Koul (1985)].

To prove Theorem 2, we utilize the following convenient central limit theorem for stationary processes. Note that, in the following Lemma 2, T is an ergodic one-to-one measure-preserving transformation on the σ-field, $\mathscr{F}$, generated by Y_k, $-\infty < k < \infty$.

LEMMA 2 [Heyde (1974)]. *Let $\{Y_k\}$ denote a stationary and ergodic sequence with $EY_0 = 0$, $EY_0^2 < \infty$, and put $S_n = \sum_{k=1}^{n} Y_k$. Suppose that $\mathscr{M}_0$ is a sub-σ-field of $\mathscr{F}$ and $\mathscr{M}_0 \subseteq T^{-1}(\mathscr{M}_0)$, and put $\mathscr{M}_k = T^{-k}(\mathscr{M}_0)$ and $y_j = E(Y_j|\mathscr{M}_0) - E(Y_j|\mathscr{M}_{-1})$, $-\infty < j < \infty$. If $\sum_{-\infty}^{\infty} y_j = Z \in L^2$, $EZ^2 = \sigma_Z^2 > 0$, and $n^{-1}ES_n^2 \to \sigma_Z^2$ as $n \to \infty$, then $S_n/\sigma_Z\sqrt{n}$ converges in distribution to the standard normal law.*

PROOF OF THEOREM 2. Denote

$$Y_k =: X_{k-1}\tilde{U}_k. \tag{2.13}$$

Then, by (1.5), Y_k is a measurable function of $\{U_{k-j}\colon j = 0, 1, 2, \dots\}$. Hence by the same arguments employed in Proposition 6.6, 6.31, and Corollary 6.33 of Breiman (1968), both $\{X_k\}$ and $\{Y_k\}$ are stationary and ergodic. The fact that $\{X_k\}$ is ergodic is also covered by Theorem 3 of Hannan (1970), page 204. Now let

$$\mathscr{M}_k =: \sigma\{U_k, U_{k-1}, U_{k-2}, \dots\} \tag{2.14}$$

and

$$y_k = E(Y_k|\mathscr{M}_0) - E(Y_k|\mathscr{M}_{-1}). \tag{2.15}$$

Note that

$$y_k = Y_k - Y_k = 0 \quad \text{for all } k \le -1. \tag{2.16}$$

If $k = 0$, then

$$y_0 = Y_0 - X_{-1}E(\tilde{U}). \tag{2.17}$$

For the case $k \ge 1$, we have

$$E(Y_k|\mathscr{M}_0) = E(\tilde{U})\left(\beta^{k-1}U_0 + \beta^k U_{-1} + \beta^{k+1}U_{-2} + \cdots\right),$$
$$E(Y_k|\mathscr{M}_{-1}) = E(\tilde{U})\left(\beta^k U_{-1} + \beta^{k+1}U_{-2} + \cdots\right),$$

and

$$y_k = E(\tilde{U})\beta^{k-1}U_0 \quad \text{for all } k \ge 1. \tag{2.18}$$

Hence

$$Z =: \sum_{-\infty}^{\infty} y_k = X_{-1}(\tilde{U}_0 - E\tilde{U}) + U_0 E(\tilde{U})/(1-\beta). \tag{2.19}$$

Note that the random variables $X_{-1}(\tilde{U}_0 - E\tilde{U})$ and U_0 are uncorrelated. Therefore

$$EZ^2 = \sigma_X^2 \mathrm{Var}(\tilde{U}) + (E\tilde{U})^2\sigma^2/(1-\beta)^2$$
$$= \sigma_X^2\left\{E(\tilde{U}^2) + 2(E\tilde{U})^2\beta/(1-\beta)\right\} = \tau^2.$$

Theorem 2 now follows by Lemmas 1 and 2. □

Note that Theorem 1 follows by Proposition 2, Lemma 1, and Theorem 2. Note that the asymptotic variance of $\hat{\beta}$ depends on H. It is natural to pursue the optimality of H. For this, we first define the following functionals:

$$J(h) =: \{(1-\beta^2)A(h) + 2\beta(1+\beta)B(h)\}/D(h), \tag{2.20}$$

where h is in $L^2(R)$ and

$$A(h) = \int\int F(t \wedge s) f(t) f(s) h(t) h(s)\, dt\, ds, \tag{2.21}$$

$$B(h) = \left(\int Ffh\right)^2, \tag{2.22}$$

$$D(h) = \left(\int f^2 h\right)^2. \tag{2.23}$$

THEOREM 3. *Assume f is bounded. Then $J(h)$ is continuous at $h \neq 0$ a.s. in L^2-norm. Furthermore, there exists a nonzero h in $L^2(R)$ to minimize $J(h)$.*

PROOF. Since f is bounded, Hölder's inequality implies that $D^{1/2}$ is a bounded functional on $L^2(R)$. Hence $D(h)$ is continuous in h [Royden (1968)]. A similar argument applies for $B(h)$. Next note that for any given $h_0 \in L^2(R)$, we have

$$\begin{aligned} |A(h) - A(h_0)| &\le \int\int f(x) f(y) |h(y) - h_0(y)| |h(x)|\, dx\, dy \\ &\quad + \int\int f(x) f(y) |h_0(y)| |h(x) - h_0(x)|\, dx\, dy \\ &\le b_1(\|h\| + \|h_0\|)\|h - h_0\| \to 0 \end{aligned}$$

as $h \to h_0$ in L^2-norm. Hence $J(h)$ is continuous at h where $h \neq 0$ a.s. in L^2-norm. Next note that $J(\alpha h) = J(h)$, for all $\alpha \neq 0$ and for all $\|h\| \neq 0$. But $\{\|h\| = 1\}$ is a compact set in $L^2(R)$. Hence the continuity of J implies the existence of h which minimizes J. This completes the proof of Theorem 3. □

Note that $L^2(R)$ is a separable Hilbert space. By Theorem 3, we can now apply Theorem 40.1 of Gelfand and Fomin (1968) to minimize $J(h)$. Note that, in view of the proof of the quoted theorem, it is not necessary for $J(h)$ to be continuous everywhere. The details of this approach are rather involved and will be reported elsewhere.

Finally, we would like to point out that another method to attain the lower bound in (1.3) for the large-sample estimation can possibly be done by using a general adaptive procedure [see, e.g., Fabian and Hannan (1982)]. The disadvantage of this method is that it may require an extremely large sample to construct a nonparametric estimator for β. This is fine for the earth-sciences data [see, e.g., Tukey (1978)], but it causes big problems for business or social-sciences data. Because this kind of data is usually subject to heavy outside influences, it is then difficult to find a good model for an extremely long series. This is one of the reasons that the more direct approaches like minimum distance method are highly desired.

3. Proofs of Proposition 1, Proposition 2, and Lemma 1. We split the proof of Proposition 1 into the following lemmas. This approach is close to the one used by Koul and DeWet (1983) for the usual regression model.

LEMMA 3.1. *If f is bounded and H is finite, then*

$$\limsup_n E \sup_{\sqrt{n}|\Delta-\beta|\le a} n^{-1}\int\{S(t,\beta)+n(\Delta-\beta)\sigma_X^2 f(t)\}^2\,dH(t)<\infty. \tag{3.1}$$

PROOF. By using the fact that H is a finite measure and the inequality $(a+b)^2 \le 2(a^2+b^2)$, the left-hand side of (3.1) is less than or equal to

$$\limsup_n 2n^{-1}E\int S^2(t,\beta)\,dH(t) + \limsup_n \sup_{\sqrt{n}|\Delta-\beta|\le a} 2n(\Delta-\beta)^2\sigma_X^4\int f^2\,dH.$$

The proof of the lemma is then completed by the assumptions and application (f) of Lemma 1, Section 2, which also rules out the consideration of H to be the Lebesgue measure over the whole real line. □

LEMMA 3.2. *Assume* (A.1), (A.2), $E(U^4)<\infty$, *and H is finite. Then*

$$\sup_{\sqrt{n}|\Delta-\beta|\le a}\int n^{-1}\{\bar{S}(t,\Delta)-n(\Delta-\beta)\sigma_X^2 f(t)\}^2\,dH(t)=o(1), \tag{3.2}$$

where

$$\bar{S}(t,\Delta) =: ES(t,\Delta)-ES(t,\beta)=n\int xF(t+(\Delta-\beta)x)\,dG(x). \tag{3.3}$$

PROOF. By the Schwarz inequality, we have, for all t,

$$\begin{aligned} &n^{-1}\{\bar{S}(t,\Delta)-n(\Delta-\beta)\sigma_X^2 f(t)\}^2\\ &\quad\le n(\Delta-\beta)^2\int\left(\frac{F(t+(\Delta-\beta)x)-F(t)}{(\Delta-\beta)x}-f(t)\right)^2 x^4\,dG(x) \qquad (3.4)\\ &\quad=O(1),\end{aligned}$$

since $n(\Delta-\beta)^2\le a$, F is Lipschitz, f is bounded, and $E(X^4)<\infty$. (3.4) also implies that lhs (3.4) $=o(1)$. In detail, note that for any $\varepsilon>0$, there exists $b>0$ such that

$$\text{lhs (3.4)} \le \varepsilon + a^2\int_{-b}^{b}\{f(t+\xi_n(x))-f(t)\}^2x^4\,dG(x),$$

where $|\xi_n(x)|\le|\Delta x-\beta x|\le n^{-1/2}ab\to 0$ as $n\to\infty$. Note ε does not depend on

n. Therefore,

$$\text{lhs (3.4)} \le \varepsilon + a^2E(X^4) \sup_{x\in[-b,\,b]} \{f(t+\xi_n(x)) - f(t)\}^2$$
$$= \varepsilon + o(1)$$

by (2.7), which is true under (A.1) and (A.2). We further have

$$\sup_t n^{1/2}(\Delta - \beta)\sigma_X^2 f(t) = O(1) = \sup_t n^{-1/2}\bar{S}(t, \Delta).$$

Hence the dominated convergence theorem implies

$$\int n^{-1}\{\bar{S}(t, \Delta) - n(\Delta - \beta)\sigma_X^2 f(t)\}^2\, dH(t) = o(1). \tag{3.5}$$

This in turn implies (3.2) by using the monotonicity of $\bar{S}(t, \Delta)$ in Δ [see, e.g., the proof of Theorem 5.1 of Koul and DeWet (1983)]. □

LEMMA 3.3. *Assume* (*A*.1), (*A*.2), $E|U|^5 < \infty$, *and* H *is finite. Then*

$$\sup_{\sqrt{n}|\Delta-\beta|\le a} \int n^{-1}E\{S(t, \Delta) - S(t, \beta) - \bar{S}(t, \Delta)\}^2\, dH(t) = o(1). \tag{3.6}$$

PROOF. In view of the previous proof, it suffices to verify

$$n^{-1}E\{S(t, \Delta) - S(t, \beta) - \bar{S}(t, \Delta)\}^2 = o(1). \tag{3.7}$$

Note that, by Lemma 3.2,

$$\text{lhs (3.7)} = n^{-1}E\{S(t, \Delta) - S(t, \beta)\}^2 - n(\Delta - \beta)^2\sigma_X^4 f^2(t) + o(1). \tag{3.8}$$

Denote

$$D_k = X_{k-1}(I[U_k - (\Delta - \beta)X_{k-1} \le t] - I[U_k \le t]). \tag{3.9}$$

By (1.5) and $|\beta| < 1$, D_k is a measurable function of $\{U_j\colon -\infty < j \le k\}$. Hence $\{D_k\}$ is a stationary process and the first term of rhs (3.8) reduces to

$$ED_1^2 + 2n^{-1}\{(n-1)ED_1D_2 + (n-2)ED_1D_3 + \cdots + ED_1D_n\}. \tag{3.10}$$

Since F is Lipschitz, the first term of (3.10) reduces to

$$ED_1^2 \le O(1)|\Delta - \beta|E|X^3| = o(1). \tag{3.11}$$

For the second term of (3.10), note first

$$\begin{aligned} ED_1D_k &= EX_0X_{k-1}\{I[U_1 - (\Delta-\beta)X_0 \le t] - I[U_1 \le t] - (\Delta - \beta)X_0 f(t)\} \\ &\qquad \times \{I[U_k - (\Delta - \beta)X_{k-1} \le t] - I[U_k \le t]\} \\ &\quad + (\Delta - \beta)f(t)EX_0^2X_{k-1} \\ &\qquad \times \{I[U_k - (\Delta - \beta)X_{k-1} \le t] - I[U_k \le t] - (\Delta - \beta)X_{k-1}f(t)\} \\ &\quad + (\Delta - \beta)^2 f^2(t)E(X_0^2X_{k-1}^2) \\ &= B_1 + B_2 + B_3, \end{aligned} \tag{3.12}$$

say. We claim

$$nB_1 + nB_2 = o(1). \tag{3.13}$$

To prove the above claim, in view of the proof of Lemma 3.2, it suffices to verify

$$nB_1 + nB_2 = O(1). \tag{3.14}$$

To this end, note F is Lipschitz and hence

$$\begin{aligned} &E\big(X_{k-1}I[t \le U_k \le t + (\Delta - \beta)X_{k-1}]I[(\Delta - \beta)X_k \ge 0]\,|\mathscr{F}_{k-1}\big) \\ &\quad \le X_{k-1}^2|\Delta - \beta|O(1), \end{aligned} \tag{3.15}$$

where $\mathscr{F}_{k-1}$ is the σ-algebra generated by $\{X_0, U_1, \dots, U_{k-1}\}$, and

$$\begin{aligned} &E\big(U_1 I[t \le U_1 \le t + (\Delta - \beta)X_0]I[(\Delta - \beta)X_0 \ge 0]\,|\mathscr{F}_0\big) \\ &\quad \le O(1)(\Delta - \beta)X_0 I[(\Delta - \beta)X_0 \ge 0]. \end{aligned} \tag{3.16}$$

By (3.15), $EX^4 < \infty$, and the boundedness of f, we then have $nB_2 = O(1)$. Using this, we further have

$$\begin{aligned} n|B_1| &\le n \cdot O(1)(\Delta - \beta)EX_0X_{k-1}^2 \\ &\qquad \times \{I[U_1 - (\Delta - \beta)X_0 \le t] - I[U_1 \le t]\} + O(1). \end{aligned} \tag{3.17}$$

Let $I_{0,1} =: I[U_1 - (\Delta - \beta)X_0 \le t] - I[U_1 \le t]$, then the rhs (3.17) is bounded by

$$\begin{aligned} &n \cdot O(1)(\Delta - \beta)E(X_0 I_{0,1}) \\ &\quad + n \cdot O(1)(\Delta - \beta)E\Big\{X_0 I_{0,1}\big(\beta^{k-1}X_0 + \beta^{k-2}U_1\big)^2\Big\} + O(1) \\ &\quad \le O(1), \end{aligned}$$

where the inequality holds by (3.15), (3.16), and $E|X^5| < \infty$. This completes the proof of (3.14) and therefore (3.13) holds. The proof of the lemma is then completed by using (3.10)–(3.13) and by the facts that $E(X_0^2X_{k-1}^2) = \beta^{2k-2}E(X^4) + \sigma_X^4(1 - \beta^{2k})$ and that

$$n^{-1}\sum_{k=1}^{n-1}(n - k)\beta^k = o(1) + \beta/(1 - \beta) \quad \text{for all } |\beta| < 1. \qquad \square$$

Proof of Proposition 1. To simplify the notation, we denote $\int g^2(t, \Delta)\,dH(t) =: \|g_\Delta\|_H^2$, for any measurable function g. Then, by Minkowski's inequality,

$$\begin{aligned} Q(\Delta) - Q_1(\Delta) &= \|S_\Delta\|_H^2 - \big\|S_\beta + n(\Delta - \beta)f\big\|_H^2 \\ &\le \Big\{\|S_\Delta - S_\beta - \bar{S}_\Delta\|_H + \big\|\bar{S}_\Delta - n(\Delta - \beta)\sigma_X^2 f\big\|_H\Big\} \\ &\quad \times \Big\{\|S_\Delta - S_\beta - \bar{S}_\Delta\|_H + \big\|\bar{S}_\Delta - n(\Delta - \beta)\sigma_X^2 f\big\|_H \\ &\qquad + 2\big\|S_\beta + n(\Delta - \beta)\sigma_X^2 f\big\|_H\Big\}, \end{aligned}$$

where the conditions for Minkowski's inequality are assured by the above lemmas which further imply (2.1). This completes the proof of Proposition 1. □

PROOF OF PROPOSITION 2. In view of (2.8), it suffices to prove

$$\sqrt{n}(\hat{\beta} - \hat{\Delta}) \to 0 \quad \text{in probability.} \tag{3.18}$$

Note the measurability of $\hat{\beta}$ can be achieved by giving a definite rule of selection for $\hat{\beta}$. But the procedure is long and uninteresting and will be omitted without further discussion. The following proof of (3.18) is close to the one in Williamson (1982). Note first by (2.2) and (2.8) we have

$$Q_1(\Delta) = \|S_\beta\|_H^2 + n^2\sigma_X^4\|f\|_H^2\{(\Delta - \beta)^2 - 2(\Delta - \beta)(\hat{\Delta} - \beta)\},$$

$$Q_1(\hat{\beta}) - Q_1(\hat{\Delta}) = n^2\sigma_X^4\|f\|_H^2(\hat{\beta} - \hat{\Delta})^2,$$

and

$$n\sigma_X^4\|f\|_H^2(\hat{\beta} - \hat{\Delta})^2 = n^{-1}(Q_1(\hat{\beta}) - Q_1(\hat{\Delta})). \tag{3.19}$$

By Proposition 1 and Chebyshev's inequality, for any fixed a there exists $\varepsilon > 0$ such that

$$P\left\{\sup_{\sqrt{n}|\Delta - \beta| \le a} n^{-1}|Q(\Delta) - Q_1(\Delta)| \le \varepsilon\right\} \to 1. \tag{3.20}$$

This motivates us to consider the set

$$A_n(a) =: \left\{\sqrt{n}|\hat{\Delta} - \beta| \le a,\ \inf_{\sqrt{n}|\Delta - \beta| \le a} Q(\Delta) < \inf_{\sqrt{n}|\Delta - \beta| > a} Q(\Delta)\right\}. \tag{3.21}$$

It is reasonable to guess, for any small $\varepsilon > 0$,

$$P\{A_n(a)\} > 1 - \varepsilon \quad \text{for some } a \text{ and for sufficiently large } n. \tag{3.22}$$

This guess will be proved formally in Lemma 3.4 below. In view of (3.22), (3.20), and (3.19), to prove (3.18) it suffices to prove

$$\sup_{\sqrt{n}|\Delta - \beta| \le a} n^{-1}|Q(\Delta) - Q_1(\Delta)| < \varepsilon \tag{3.23}$$

implies $n^{-1}\{Q_1(\hat{\beta}) - Q_1(\hat{\Delta})\} < 2\varepsilon$ on $A_n(a)$. Indeed, by the assumption of (3.23), we have, on $A_n(a)$,

$$n^{-1}|Q(\hat{\beta}) - Q_1(\hat{\beta})| < \varepsilon \quad \text{and} \quad n^{-1}|Q(\hat{\Delta}) - Q_1(\hat{\Delta})| < \varepsilon; \tag{3.24}$$

hence

$$\begin{aligned} n^{-1}Q_1(\hat{\beta}) - n^{-1}Q_1(\hat{\Delta}) &\le \{n^{-1}Q(\hat{\beta}) + \varepsilon\} - n^{-1}Q_1(\hat{\Delta}) \\ &\le \{n^{-1}Q(\hat{\Delta}) + \varepsilon\} - n^{-1}Q_1(\hat{\Delta}) \le \varepsilon + \varepsilon = 2\varepsilon \end{aligned}$$

on $A_n(a)$, where the first and last inequalities hold by (3.24) and the second inequality holds by (1.10). This completes the proof of (3.23) and Proposition 2. □

The next lemma will make up the gap in the above proof. See (3.22).

LEMMA 3.4. *Under the assumptions of Proposition* 1, *we have*

$$P\{A_n(a)\} > 1 - \varepsilon \quad \textit{for some } a \textit{ and for sufficiently large } n, \tag{3.25}$$

where $A_n(a)$ *is defined in* (3.21).

PROOF. Define

$$L(\Delta) =: \int S(t, \Delta)\sqrt{f(t)}\, dH(t). \tag{3.26}$$

We will use the asymptotic linearity of $L(\Delta)$ as leverage to prove the lemma. By the asymptotic linearity of $L(\Delta)$ we mean

$$\sup_{\sqrt{n}|\Delta - \beta| \le a} n^{-1/2}\left|L(\Delta) - L(\beta) - (\Delta - \beta)n\sigma_X^2 \int f^{3/2}\, dH\right| \to 0 \tag{3.27}$$

in probability. Indeed,

$$\begin{aligned}\text{lhs (3.27)} &\le \sup n^{-1/2} \int |S(t, \Delta) - S(t, \beta) - \bar{S}(t, \Delta)|\sqrt{f(t)}\, dH(t)\\ &\quad + \sup n^{-1/2} \int |\bar{S}(t, \Delta) - (\Delta - \beta)n\sigma_X^2 f(t)|\sqrt{f(t)}\, dH(t)\\ &\to 0\end{aligned}$$

in probability by (3.6), (3.2), and the Schwarz inequality. Next, by the fact that

$$n^{-1}L^2(\beta) \le n^{-1}Q(\beta)\int f\, dH = O_p(1),$$

we extend (3.27) to

$$\sup_{\sqrt{n}|\Delta - \beta| \le a}\left|n^{-1}L^2(\Delta) - \left\{n^{-1/2}L(\beta) + (\Delta - \beta)n^{1/2}\sigma_X^2 \int f^{3/2}\, dH\right\}^2\right| \to_P 0. \tag{3.28}$$

Note also, by the Remark (a) of Theorem 2 and Lemma 3.1, we have

For all $\varepsilon > 0$ there exist $d > 0$ and $n_0 > 0$, such that for all $n \ge n_0$,

$$P\{\sqrt{n}|\Delta - \beta| \le d\} \ge 1 - \frac{\varepsilon}{2} \text{ and } P\left\{n^{-1}Q(\beta)\int f\, dH < d\right\} \ge 1 - \frac{\varepsilon}{4}. \tag{3.29}$$

Obviously,

$$n^{-1/2}|L(\beta)| < \sqrt{d} \quad \text{on the set } \left\{n^{-1}Q(\beta)\int f\, dH < d\right\}. \tag{3.30}$$

Hence, on the set $\{n^{-1}Q(\beta)\int f\, dH < d\}$, by choosing $a \ge 2\sqrt{d}\{\int f^{3/2}\, dH\}^{-1}$ we have

$$
(3.31)\qquad
\begin{aligned}
&\min_{\sqrt{n}|\Delta-\beta|=a}\left\{n^{-1/2}L(\beta)+(\Delta-\beta)n^{1/2}\int f^{3/2}\,dH\right\}^2\\
&\quad\geq\left\{-n^{-1/2}|L(\beta)|+a\int f^{3/2}\,dH\right\}^2\\
&\quad\geq\{-\sqrt{d}+2\sqrt{d}\}^2=d>n^{-1}Q(\beta)\int f\,dH.
\end{aligned}
$$

Consequently, by (3.30), (3.29), and (3.28) we conclude

$$
(3.32)\qquad P\left\{n^{-1}Q(\beta)\int f\,dH<d \text{ and } Q(\beta)\int f\,dH<\min_{\sqrt{n}|\Delta-\beta|=a}L^2(\Delta)\right\}\geq 1-\frac{\varepsilon}{2}
$$

for large n. Finally, note

$$
(3.33)\qquad
\begin{aligned}
\inf_{\sqrt{n}|\Delta-\beta|>a}Q(\Delta)\int f\,dH&\geq\inf_{\sqrt{n}|\Delta-\beta|>a}L^2(\Delta)\\
&\geq\inf_{\sqrt{n}|\Delta-\beta|\geq a}L^2(\Delta)=\min_{\sqrt{n}|\Delta-\beta|=a}L^2(\Delta),
\end{aligned}
$$

where the last equality holds because $L(\Delta)$ is nondecreasing in Δ. (3.33), (3.32), and (3.29) together yield (3.25). □

Proof of Lemma 1. Note that

$$
(3.34)\qquad n^{-1}E(T_n^2)=\sigma_X^2E(\tilde{U}^2)+2n^{-1}\sum_{j<k}E(X_{j-1}X_{k-1}\tilde{U}_j)E(\tilde{U}).
$$

For all $j\geq 0$ let

$$
(3.35)\qquad h(j)=:E(X_0X_j\tilde{U}_1).
$$

Then, by induction

$$
(3.36)\qquad h(j)=\sigma_X^2E(\tilde{U})\beta^j.
$$

Since $\{X_{j-1}X_{k-1}\tilde{U}_j\colon 1\leq j\leq k\leq n\}$ are identically distributed, the second term of the rhs (3.34) equals

$$
\begin{aligned}
&2E(\tilde{U})n^{-1}\{(n-1)h(1)+(n-2)h(2)+\cdots+h(n-1)\}\\
&\quad=2E^2(\tilde{U})\sigma_X^2n^{-1}\sum_{j=1}^{n-1}(n-j)\beta^j=2E^2(\tilde{U})\sigma_X^2\beta/(1-\beta)+o(1).
\end{aligned}
$$

This completes the proof of the lemma. □

Acknowledgments. The author would like to thank the Associate Editor and the referees for their careful reading of the original manuscript and for their suggestions that led to substantial improvements in the presentation and the final result of Theorem 3.

REFERENCES

AKRITAS, M. G. and JOHNSON, R. A. (1980). Efficiencies of tests and estimators for pth-order autoregressive processes when the error distribution is nonnormal. Technical Report, Dept. of Statist., Univ. of Wisconsin, Madison.

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: forecasting and control.* Holden-Day, San Francisco.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. verw. Gebiete* **59** 459–478.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

GELFAND, I. M. and FOMIN, S. V. (1968). *Calculus of Variations.* Prentice Hall, New Jersey.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HEYDE, C. C. (1974). On the central limit theorem for stationary processes. *Z. Wahrsch. verw. Gebiete* **30** 315–320.

KOUL, H. L. (1985). Minimum distance estimation and goodness-of-fit tests in first-order autoregression. *Ann. Statist.* **14** 1194–1213.

KOUL, H. L. and DEWET, T. (1983). Minimum distance estimation in a linear regression model. *Ann. Statist.* **11** 921–932.

LAI, T. L. and SIEGMUND, D. (1983). Fixed accuracy estimation of an autoregressive parameter. *Ann. Statist.* **11** 478–485.

MILLAR, P. W. (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10** 717–740.

ROYDEN, H. L. (1968). *Real Analysis.* Macmillan, New York.

TUKEY, J. W. (1978). Can we predict where "time series" should go next? In *Directions in Time Series* (D. R. Brillinger and G. C. Tiao, eds.) 1–31. IMS, Hayward, Calif.

VEITCH, J. G. (1983). Minimum distance procedures in stationary time series. Ph.D. thesis, Univ. of California, Berkeley.

WILLIAMSON, M. A. (1982). Cramér–von Mises-type estimation of the regression parameter: The rank analogue. *J. Multivariate Anal.* **12** 248–255.

DEPARTMENT OF MATHEMATICS AND STATISTICS
TRENTON STATE COLLEGE
EWING TOWNSHIP, NEW JERSEY 08625