The Annals of Statistics
1986, Vol. 14, No. 3, 1194–1213

MINIMUM DISTANCE ESTIMATION AND GOODNESS-OF-FIT TESTS IN FIRST-ORDER AUTOREGRESSION[1]

By Hira L. Koul

Michigan State University

This paper gives a class of minimum L_2-distance estimators of the autoregression parameter in the first-order autoregression model when the errors have an unknown symmetric distribution. Within the class an asymptotically efficient estimator is exhibited. The asymptotic efficiency of this estimator relative to the least-squares estimator is the same as that of a certain signed rank estimator relative to the sample mean in the one sample location model. The paper also discusses goodness-of-fit tests for testing for symmetry and for a specified error distribution.

1. Introduction. Let $\{\varepsilon_i,\ i = 0, \pm 1, \pm 2, \dots\}$ be independent random variables (r.v.'s) that are identically distributed according to a distribution function (d.f.) F. Let $\{X_i\}$ be an observable process such that, for $|\rho| < 1$, X_{i-1} is independent of ε_i and

$$X_i = \rho X_{i-1} + \varepsilon_i, \qquad i = 0, \pm 1, \pm 2, \dots . \tag{1}$$

The above process $\{X_i\}$ is called the first-order autoregressive (AR(1)) process. This paper considers the minimum distance estimation of ρ based on the observations $\{X_0, X_1, \dots X_n\}$ when F is not necessarily known. Also considered are tests of symmetry of F and tests of the goodness-of-fit for F.

Of course the classical estimator $\hat{r} := \sum_{i=1}^n X_{i-1}X_i / \sum_{i=1}^n X_{i-1}^2$ is a minimum distance estimator. But this estimator is highly inefficient for non-Gaussian errors, including contaminated Gaussian errors [Fox (1972), Denby and Martin (1979), and Martin (1981)]. The minimum distance (m.d.) estimation methods that lead to efficient and robust estimators in models involving independent observations are those promoted by Wolfowitz (1957). These methods are further studied by Beran (1977, 1978), Williamson (1979), Boos (1981, 1982), Parr and Schucany (1980), Parr and DeWet (1981), Millar (1981, 1982), Koul (1980, 1985), and Koul and DeWet (1983), among others. See also Parr (1981) for a detailed bibliography prior to 1981.

The most common distance statistics used in the literature are the Cramér–von Mises type statistics. Some of the reasons for this are that the corresponding m.d. estimators are consistent, asymptotically normal, qualitatively robust against certain contaminated errors [Millar (1981, 1982) and Koul (1985)] and locally asymptotically minimax (Millar, op. cit.). In view of these properties of practical import it is highly desirable to seek m.d. estimators of ρ in (1), using a suitable Cramér–von Mises type statistic.

Received March 1984; revised October 1985.

[1] Research supported by NSF Grant 82-01291.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G20, 62G10.

Key words and phrases. Weighted empirical residual process, stationary, ergodic, influence curve.

To motivate the definition of m.d. estimators of ρ, let us recall a result from Koul and DeWet (K-D) (op. cit.). Consider a simple linear regression model through the origin: $Y_i = c_i\beta + \varepsilon_i$, $1 \le i \le n$, where $\{\varepsilon_i\}$ are i.i.d. F, F a known d.f., $\{c_i\}$ known constants. It was shown in K-D that a t minimizing

$$\int \left[\sum_{i=1}^{n} c_i\{I(Y_i \le y + tc_i) - F(y)\} \right]^2 dH(y), \tag{2}$$

with H as in (6) below, is asymptotically an optimal estimator of β among a class of estimators including the one based on an L_2-distance between the ordinary residual empirical process and F.

It is natural to consider estimation of ρ in the autoregression model (1) by formally replacing c_i in (2) by X_{i-1}. Thus, if we know the error d.f. F in (1), we may define an estimator of ρ as a t that minimizes

$$\int \left[\sum_{i=1}^{n} X_{i-1}\{I(X_i \le y + tX_{i-1}) - F(y)\} \right]^2 dH(y). \tag{3}$$

But in practice F is rarely known and we must now eliminate the centering F in (3). For that purpose we shall assume that

$$F \text{ is symmetric around } 0. \tag{4}$$

Then a way to eliminate F in (3) is to replace it by the indicator $I(-X_i < y - tX_{i-1})$ because at $t = \rho$, this indicator also estimates $F(y)$. We are thus motivated to introduce

$$\begin{aligned} S(y, t) &:= n^{-1/2} \sum_{i=1}^{n} X_{i-1}\{I(X_i \le y + tX_{i-1}) - I(-X_i < y - tX_{i-1})\}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad y, t \text{ in } \mathscr{R}, \\ M(t) &:= \int S^2(y, t)\, dH(y), \qquad t \text{ in } \mathscr{R}, \end{aligned} \tag{5}$$

where

(6) H is a known nondecreasing right continuous function from $\mathscr{R}$ to $\mathscr{R}$, inducing a σ-finite measure on $(\mathscr{R}, \mathscr{B})$—the Borel line.

Now define $\hat{\rho}$ by the relation

$$\inf_t M(t) = M(\hat{\rho}). \tag{7}$$

The dependence of $\hat{\rho}$ on H will be exhibited only occasionally. The proposed class of m.d. estimators is $\{\hat{\rho}(H),\ H \text{ varies}\}$. The optimality with respect to H is discussed in Remark 3.2.

This paper studies some finite and large sample properties of the class of estimators $\{\hat{\rho}(H)\}$. Section 2 discusses some finite sample properties including the computational and scale invariance aspects. In general the class of estimators $\{\hat{\rho}(H)\}$ is not scale invariant. These estimators can be made scale invariant by using $\{s^{-1}X_i,\ 0 \le i \le n\}$ in place of $\{X_i,\ 0 \le i \le n\}$ in (7), where s is a suitable scale invariant estimator of a scale parameter of F. See (2.7) below for an example of s. Section 2 also contains extensions of $\hat{\rho}$ to the AR(1) model with location parameter as well, and to the AR(2) model.

In Section 3, a class of estimators $\{\hat{\rho}_h\}$ is introduced and its asymptotic normality is proved. The estimator $\hat{\rho}$ is a member of this class. Theorem 3.3 asserts an asymptotic optimality of $\hat{\rho}$ among the class of estimators $\{\hat{\rho}_h\}$. This result is similar to Theorem 3.2 of K-D. In the same section, $\hat{\rho}$ is compared to other estimators. The asymptotic efficiency of $\{\hat{\rho}(H)\}$ relative to $\hat{r}$ is the same as that of certain signed rank estimators relative to the sample mean in the one sample location model; see Remark 3.3. Asymptotically, $\{\hat{\rho}_h(H)\}$ are like GM-estimators of Denby and Martin (1979) corresponding to a ψ that depends on F; see Remark 3.7.

In Remark 3.4 an estimator of the asymptotic variance of $n^{1/2}(\hat{\rho}(H) - \rho)$ is provided for $H(y) \equiv y$. The effect of asymmetry of F on $\hat{\rho}$ is evaluated in Remark 3.5. It is noted that if F is asymmetric but $E\varepsilon_1 = 0$ then $\hat{\rho}$ has no asymptotic bias but its asymptotic variance is larger than that under symmetry. In Remark 3.8 it is noted that the influence of contaminating $\{\varepsilon_i\}$ on $\hat{\rho}$ is zero at a symmetric F. Remark 3.9 discusses the asymptotic distribution of the scale invariant version of $\{\hat{\rho}(H)\}$. Finally, Section 3 briefly gives the asymptotic distributions of the extensions of $\hat{\rho}$ introduced in Section 2.

Section 4 discusses tests of the hypothesis of symmetry H_0 and tests of goodness of fit for a specified error distribution. In both situations the asymptotic null distributions of the proposed tests are shown to be the same as those of their counterparts in the one sample location model. See also Pierce (1985) for a similar observation with regards to tests of the normality of errors. The asymptotic null distribution of the Cramér–von Mises type tests based on the ordinary empirical for testing for a specified error distribution does not depend on the specified error d.f., as long as its mean is 0. Section 5 contains all the proofs.

NOTATION. All limits are taken as $n \to \infty$, unless specified otherwise. By $o_p(1)$ ($O_p(1)$) is meant a sequence of r.v.'s that tends to zero (stays bounded) in probability. The dependence of various entities on n is not exhibited, for the sake of convenience. The real line is denoted by $\mathscr{R}$.

2. Some finite sample properties of $\hat{\rho}$ and extensions. For the purpose of computation of $\hat{\rho}$ the following representation of M is useful. Fix a t in $\mathscr{R}$ and let $Z_i = X_i - tX_{i-1}$, $i = 1, \dots, n$. Let $c = \max\{Z_i, -Z_i;\ 1 \le i \le n\}$. Observe that in (1.5), $S(y, t) \equiv 0$ for all $y > c$. Now use the fact that for any reals a, b, $2\max(a, b) = a + b + |a - b|$, and the nondecreasing nature of H to conclude that for any real t

$$M(t) = n^{-1}\sum_i \sum_j h_i h_j \Big[|H(Z_i) - H(-Z_j)| - \tfrac{1}{2}\big\{ |H(Z_i) - H(Z_j)| + |H(-Z_i) - H(-Z_j)| \big\} \Big], \tag{1}$$

where $h_i = X_{i-1}$. If

$$|H(a) - H(b)| = |H(-a) - H(-b)|, \qquad a, b \text{ in } \mathscr{R}, \tag{2}$$

then, for $t \in \mathscr{R}$,

$$M(t) = n^{-1}\sum_i \sum_j X_{i-1}X_{j-1}\Big[\big|H(X_i - tX_{i-1}) - H(-X_j + tX_{j-1})\big| - \big|H(X_i - tX_{i-1}) - H(X_j - tX_{j-1})\big|\Big]. \tag{3}$$

In the derivation of (1), H is assumed to be continuous. If H is not continuous (1) continues to hold with probability 1 as long as F is continuous. In any case the representation (1) or (3) make it clear that the computation of $\hat{\rho}$ is similar to that of maximum likelihood estimators.

The assumption (2) is that H is symmetric around 0, which is natural when F is symmetric. Useful examples are $H(y) \equiv y$ and H given by $dH = \{F_0(1 - F_0)\}^{-1} dF_0$, F_0 a known symmetric d.f.

To overcome the difficulty due to the possible nonuniqueness, modify $\hat{\rho}$ as follows. First observe that by the Cauchy–Schwarz inequality, for all t in $\mathscr{R}$,

$$M(t) \geq \left\{\int S(y,t) g^{1/2}(y)\, dH(y)\right\}^2 \Big/ \int g\, dH, \tag{4}$$

where g is a nonnegative function on $\mathscr{R}$ such that

$$0 < \int g\, dH < \infty. \tag{5}$$

Let

$$\begin{aligned} L(t) &:= \int S(y,t) g^{1/2}(y)\, dH(y) \\ &= \int n^{-1/2} \sum_i X_{i-1}\{I(X_i \leq y + tX_{i-1}) - I(-X_i < y - tX_{i-1})\} \\ &\quad \times g^{1/2}(y)\, dH(y). \end{aligned} \tag{6}$$

Clearly $L(t)$ is a nondecreasing function of t. Therefore, by (4), $M(t)$ is bounded below by a nonnegative function which is nonincreasing on $(-\infty, b_0)$ and nondecreasing on $[b_0, \infty)$ for some finite b_0. Consequently, $\hat{\rho}$ may be uniquely defined as an average of the two quantities at which $M(t)$ is minimized for the first time and for the last time as t moves from the left to the right.

Now consider the question of scale invariance of $\hat{\rho}$. Write $\hat{\rho}(\lambda)$ for $\hat{\rho}(H)$ when $H(y) \equiv y$, (λ for the Lebesgue measure). Note that $\hat{\rho}(\lambda)$ is scale invariant in the sense that $\hat{\rho}(\lambda)$ based on $\{bX_i, 0 \leq i \leq n\}$ is the same as the $\hat{\rho}(\lambda)$ based on $\{X_i, 0 \leq i \leq n\}$ for all b in $\mathscr{R}$. In general $\{\hat{\rho}(H)\}$ are not scale invariant in this sense. One way to make these estimators scale invariant is to base them on $\{s^{-1}X_i, 0 \leq i \leq n\}$, where $s = s(\mathbf{X}) = s(X_0, X_1, \dots X_n)$ is a scale estimator such that $s(b\mathbf{X}) = |b| s(\mathbf{X})$ for all real b. We mention one such estimator:

$$s = \operatorname{med}\{|X_i - \hat{\rho}_0 X_{i-1}|, 1 \leq i \leq n\}, \tag{7}$$

where $\hat{\rho}_0$ is a scale invariant estimator of ρ, e.g. $\hat{\rho}(\lambda)$ or $\hat{r}$. The effect of making $\hat{\rho}(H)$ scale invariant on its asymptotic distribution is discussed in Remark 3.9.

Next, consider extensions of $\hat{\rho}$. First, consider the model where, for θ in $\mathscr{R}$, $|\rho| < 1$,

$$X_i = \theta + \rho X_{i-1} + \varepsilon_i, \qquad i = 0, \pm 1, \pm 2 \dots, \tag{8}$$

with ε_i, X_{i-1} as in (1.1). To define m.d. estimators of (θ, ρ), let

$$S_0(y; a, t) := n^{-1/2} \sum_{i=1}^{n} \{I(X_i \le y + a + tX_{i-1}) - I(-X_i < y - a - tX_{i-1})\},$$

$$S_1(y; a, t) := n^{-1/2} \sum_{i=1}^{n} X_{i-1}\{I(X_i \le y + a + tX_{i-1}) - I(-X_i < y - a - tX_{i-1})\}, \tag{9}$$

$$M(a, t) := \int \{S_0^2(y; a, t) + S_1^2(y; a, t)\}\, dH(y), \qquad a, t, y \text{ in } \mathscr{R}.$$

Now define $(\hat{\theta}, \hat{\rho})$ by the relation

$$\inf_{a, t} M(a, t) = M(\hat{\theta}, \hat{\rho}). \tag{10}$$

Next, consider the stationary AR(2) process, where for ρ_1, ρ_2 real,

$$X_i = \rho_1 X_{i-1} + \rho_2 X_{i-2} + \varepsilon_i, \qquad i = 0, \pm 1, \pm 2, \dots, \tag{11}$$

with ε_i, X_{i-1} as in (1.1). To define m.d. estimators of (ρ_1, ρ_2), define for t_1, t_2 in $\mathscr{R}$,

$$S_j(y; t_1, t_2) := n^{-1/2} \sum_{i=j}^{n} X_{i-j}\{I(X_i \le y + t_1 X_{i-1} + t_2 X_{i-2}) - I(-X_i < y - t_1 X_{i-1} - t_2 X_{i-2})\}, \qquad j = 1, 2. \tag{12}$$

Let

$$M(t_1, t_2) := \int \{S_1^2(y; t_1, t_2) + S_2^2(y; t_1, t_2)\}\, dH(y), \qquad t_1, t_2 \text{ in } \mathscr{R}. \tag{13}$$

Define $(\hat{\rho}_1, \hat{\rho}_2)$ by the relation

$$\inf_{t_1, t_2} M(t_1, t_2) = M(\hat{\rho}_1, \hat{\rho}_2). \tag{14}$$

The asymptotic distributions of the estimators defined in (10) and (14) are summarized at the end of Section 3. Extension of $\hat{\rho}$ to the pth order stationary autoregressive series with location θ is now apparent from (8)–(14).

3. Asymptotic behavior of $\hat{\rho}$ and $\hat{\rho}_h$. This section studies the asymptotic behavior of $\hat{\rho}$ under fairly general assumptions on the underlying quantities. Actually, we first study the asymptotic distribution of a class of estimators $\{\hat{\rho}_h\}$, to be defined shortly, of which $\hat{\rho}$ is a member. Then we deduce various asymptotic results about $\hat{\rho}$ including its optimality within the class $\{\hat{\rho}_h\}$.

To define $\hat{\rho}_h$, let h be a measurable function from $\mathscr{R}$ to $\mathscr{R}$ and define, for y, t in $\mathscr{R}$,

$$\begin{aligned} S_h(y,t) &:= n^{-1/2}\sum_{i=1}^{n} h(X_{i-1})\{I(X_i \le y + tX_{i-1}) - I(-X_i < y - tX_{i-1})\}, \\ M_h(t) &:= \int S_h^2(y,t)\,dH(y). \end{aligned} \tag{1}$$

Now define $\hat{\rho}_h$ by the relation

$$\inf_t M_h(t) = M_h(\hat{\rho}_h). \tag{2}$$

Again, the dependence of $\hat{\rho}_h$ on H is suppressed. For each H we have a class of estimators $\{\hat{\rho}_h;\ h \text{ varies}\}$ which reduces to $\hat{\rho}$ of (1.7) upon taking $h(x) \propto x$.

In order to obtain the results, we shall need the following assumptions:

(A.1) F has a continuous density f with respect to λ—the Lebesgue measure on $(\mathscr{R}, \mathscr{B})$.
(A.2) $0 < \int f^r\,dH < \infty$, $r = 1, 2$.
(A.3) $0 < E\varepsilon_1^2 < \infty$; $0 < E|X_0|^r |h(X_0)|^2 < \infty$, for $r = 0, 1, 2$.
(A.4) $\lim_{s \to 0} \int E\{(X_0 h(X_0))[f(y + sX_0) - f(y)]\}^2\,dH(y) = 0$,
$\lim_{s \to 0} \int E h^2(X_0)|X_0| f(y + sX_0)\,dH(y) = E h^2(X_0)|X_0| \int f\,dH$.
(A.5) $\int_0^\infty (1 - F(y))\,dH(y) < \infty$ and H satisfies (2.2).

Next, define, for t in $\mathscr{R}$,

$$Q_h(t) := \int \left[S_h(y,\rho) + 2E(X_0 h(X_0)) f(y) n^{1/2}(t - \rho)\right]^2 dH(y). \tag{3}$$

We are now ready to state some results.

THEOREM 1. *Let $\{X_i,\ i = 0, \pm 1, \dots\}$ be as in* (1.1). *Assume that F satisfies* (1.4) *and that* (A.1)–(A.5) *hold. Then, for any $0 < b < \infty$,*

$$E \sup_{|n^{1/2}(t-\rho)| \le b} |M_h(t) - Q_h(t)| = o(1). \tag{4}$$

PROOF. See Section 5.

REMARK 1. If $E(X_0 h(X_0)) = 0$, from (3) and (4) it follows that M_h cannot be used to recover ρ asymptotically. In particular, if $h(x) \equiv 1$, then because of (1.4) and (A.3), M_1—the ordinary Cramér–von Mises statistic based on the residuals $\{X_i - tX_{i-1}\}$ and $\{-X_i + tX_{i-1}\}$—cannot be used to estimate ρ in the above fashion.

THEOREM 2. *In addition to the assumptions of Theorem* 1, *assume that*

(A.6) *either $xh(x) \ge 0$ for all x or $xh(x) \le 0$ for all x,*

$$E\{X_0 h(X_0)\} \ne 0. \tag{A.7}$$

Then

$$n^{1/2}(\hat{\rho}_h - \rho) = -\left\{2E(X_0 h(X_0)) \int f^2\, dH\right\}^{-1} \times \int S_h(y, \rho) f(y)\, dH(y) + o_p(1). \tag{5}$$

PROOF. See Section 5.

COROLLARY 1. *Under the assumptions of Theorem* 2

$$n^{1/2}(\hat{\rho}_h - \rho) \Rightarrow N(0, \tau_h^2)\ r.v., \tag{6}$$

where

$$\tau_h^2 := \left\{2E(X_0 h(X_0)) \int f^2\, dH\right\}^{-2} Eh^2(X_0) E\psi^2(\varepsilon_1),$$

$$\psi(x) := \psi_0(x) - \psi_0(-x), \qquad \psi_0(x) := \int_{-\infty}^{x} f\, dH, \qquad x \text{ in } \mathscr{R}.$$

PROOF. Set $Y_{ni} = n^{-1/2} h(X_{i-1})\psi(\varepsilon_i)$, $1 \le i \le n$. Then

$$-\int S_h(y, \rho) f(y)\, dH(y) = \sum_{i=1}^{n} Y_{ni} = T_{nn}, \tag{7}$$

where

$$T_{nj} := \sum_{i=1}^{j} Y_{ni}, \qquad 1 \le j \le n.$$

For $|\rho| < 1$, $\{X_i\}$ is strictly stationary ergodic and T_{nj} is $\mathscr{F}_j$ measurable, where $\mathscr{F}_j = \sigma$-field $\{\varepsilon_i, i \le j\}$. Moreover, the symmetry of F around 0 ensures $E\psi(\varepsilon_i) \equiv 0$, so that the independence of ε_i and X_{i-1} yields

$$\begin{aligned} &EY_{ni} = 0, \quad \text{for all } i; \qquad EY_{n1}Y_{nj} = 0, \quad \text{for all } j > 1, \\ &EY_{ni}^2 = n^{-1} Eh^2(X_0) E\psi^2(\varepsilon_1), \quad \text{for all } i. \end{aligned} \tag{8}$$

Thus $\{T_{nj}, \mathscr{F}_j;\ 1 \le j \le n\}$ is a mean zero square integrable martingale array. From (8), $ET_{nn} = 0$,

$$s_n^2 := \mathrm{Var}(T_{nn}) = Eh^2(X_0) E\psi^2(\varepsilon_1) = Eh^2(X_0)\mathrm{Var}(\psi(\varepsilon_1)) = \gamma^2, \quad \text{say}. \tag{9}$$

The ergodic theorem yields that

$$n^{-1} \sum_{i=1}^{n} h^2(X_{i-1}) \to Eh^2(X_0) \quad \text{a.s.} \tag{10}$$

These observations may be used to verify that sufficient conditions of Corollary 3.1 of Hall and Heyde (1980) are satisfied so that $T_{nn} \Rightarrow N(0, \gamma^2)$. The claim about $\hat{\rho}_h$ now follows from this and (5). □

Now, the Cauchy–Schwarz inequality implies that

$$\{EX_0h(X_0)\}^{-2}Eh^2(X_0) \geq \{EX_0^2\}^{-1}, \tag{11}$$

with equality if, and only if $h(x) \propto x$. Let τ^2 denote τ_h^2 when $h(x) \propto x$. Note that τ^2 denotes the asymptotic variance of $\hat{\rho}$. Thus (11) says that

$$\tau_h^2 \geq \tau^2 \text{ with equality if, and only if } h(x) \propto x. \tag{12}$$

We also have

$$EX_0^2 = \sigma^2\{1-\rho^2\}^{-1}, \qquad \sigma^2 = \operatorname{Var}(\varepsilon_1). \tag{13}$$

Moreover, the symmetry of F and H yields

$$E\psi^2(\varepsilon_1) = \operatorname{Var}\psi(\varepsilon_1) = 4K(F,H), \tag{14}$$

where

$$K(F,H) = \int\int[F(x\wedge y) - F(x)F(y)]\,f(x)f(y)\,dH(x)\,dH(y).$$

Consequently,

$$\begin{aligned}\tau^2 &= \{4EX_0^2\}^{-1}\operatorname{Var}\{\psi(\varepsilon_1)\}\left\{\int f^2\,dH\right\}^{-2}\\ &= (1-\rho^2)\sigma^{-2}K(F,H)\left\{\int f^2\,dH\right\}^{-2}.\end{aligned} \tag{15}$$

Summarizing the above results we have proved the following

THEOREM 3. *Among all estimators* $\{\hat{\rho}_h\}$ *of* ρ *in* (1.1), *where* h *satisfies* (A.3), (A.4), (A.6), *and* (A.7) *for every* F *and* H *satisfying* (A.1), (A.2), *and* (A.5), *the estimator that minimizes the asymptotic variance* τ_h^2 *is* $\hat{\rho}$—*the* $\hat{\rho}_h$ *when* $h(x) \propto x$. *Moreover*, $n^{1/2}(\hat{\rho}-\rho) \Rightarrow N(0,\tau^2)$ *r.v.*, τ^2 *as in* (15).

REMARK 2. Theorem 3 above proves the asymptotic optimality of $\hat{\rho}(H)$ among a class of estimators $\{\hat{\rho}_h(H),\ h$ as in Theorem 3$\}$ for every fixed H. As far as finding an optimal $\hat{\rho}$ in the class $\{\hat{\rho}(H),\ H$ varies$\}$ is concerned, observe that H appears in the asymptotic variance of $\hat{\rho}(H)$ only through $V(F,H) := K(F,H)\{\int f^2\,dH\}^{-2}$. The term $V(F,H)$ is precisely the asymptotic variance of an M-estimator of the location parameter corresponding to the ψ_0 function. From the minimax theory of such estimators (Huber, 1981), one readily concludes that the optimal H is given by the equation

$$f\,dH = -I^{-1}d(f'/f), \qquad 0 < I := \int(f'/f)^2\,dF < \infty,$$

where now it is assumed that f' is a.e. derivative of f. A consequence of this observation is that if $H(y) = y$ or if $dH = \{F(1-F)\}^{-1}dF$ then the corresponding $\hat{\rho}$ are asymptotically efficient for logistic errors.

REMARK 3. *Comparison with* $\hat{r} := \sum_{i=1}^{n} X_{i-1}X_i / \sum_{i=1}^{n} X_{i-1}^2$. Recall that under (A.3), $n^{1/2}(\hat{r} - \rho) \Rightarrow N(0, 1 - \rho^2)$ r.v. Thus the asymptotic efficiency of $\hat{\rho}$ vs. $\hat{r}$ is

$$\begin{aligned} e := e(\hat{\rho}, \hat{r}) &= \{\text{asymptotic variance } n^{1/2}(\hat{r} - \rho)\} \\ &\qquad \times \{\text{asymptotic variance } (n^{1/2}(\hat{\rho} - \rho))\}^{-1} \\ &= \sigma^2\{V(F, H)\}^{-1} = \sigma^2\left\{\int f^2\, dH\right\}^2 \{K(F, H)\}^{-1}. \end{aligned} \tag{16}$$

Observe that the asymptotic efficiency e is similar to that of signed rank estimators corresponding to the score function $\psi_0(F^{-1}(u))$ of the location parameter vs. the sample mean in the one sample location model. Thus for example if $H(y) = y$, then $e = 12\sigma^2\{\int f^2\, dy\}^2$—the well-celebrated expression in connection with the Wilcoxon rank estimator—see Lehmann (1975). In this sense $\hat{\rho}(\lambda)$ may be said to be an extension of the Wilcoxon type rank estimator to the autoregression model (1.1).

REMARK 4. *An estimator of* τ^2 *when* $H(y) \equiv y$. In the case $H(y) = y$, the τ^2 of (15) becomes

$$\tau^2 = (1 - \rho^2)\sigma^{-2}\left\{(12)^{1/2}\int f^2(x)\, dx\right\}^{-2}. \tag{17}$$

Thus to estimate τ^2 one needs to estimate $(1 - \rho^2)\sigma^{-2}$ and $\int f^2(x)\, dx$. An estimator of $(1 - \rho^2)\sigma^{-2}$ is obviously $(n^{-1}\sum_{i=1}^{n} X_{i-1}^2)^{-1}$. Thus it remains to construct an estimator of

$$b(f) := \int f^2(x)\, dx = \int f(x)\, dF(x). \tag{18}$$

Define

$$b(y, F) := \int [F(y + x) - F(-y + x)]\, dF(x), \qquad y > 0. \tag{19}$$

For any $y > 0$, $(2y)^{-1}n^{1/2}b(yn^{-1/2}, F) \to b(f)$ as $n \to \infty$. This suggests that we first estimate $b(y, F)$. An estimator of $b(y, F)$ is

$$\hat{b}(y) := n^{-1}\int [V(y + x, \hat{\rho}_0) - V(-y + x, \hat{\rho}_0)]\, V(dx, \hat{\rho}_0), \tag{20}$$

where

$$V(y, t) := n^{-1/2}\sum_{i=1}^{n} I(X_i \le y + tX_{i-1}), \qquad y, t \text{ in } \mathscr{R}, \tag{21}$$

and $\hat{\rho}_0$ is an estimator of ρ, not necessarily $\hat{\rho}$. Observe that $\hat{b}$ is the empirical d.f. of $\{|X_i - X_j - \hat{\rho}_0(X_{i-1} - X_{j-1})|,\ 1 \le i, j \le n\}$. Let $s_{\alpha, n}$ be an αth quantile of this empirical d.f. Define

$$\hat{\tau}^2 := \left(n^{-1}\sum_{i=1}^{n} X_{i-1}^2\right)^{-1}\left\{12^{1/2}(2s_{\alpha, n})^{-1}n^{1/2}\hat{b}\left(n^{-1/2}s_{\alpha, n}\right)]\right\}^{-2}. \tag{22}$$

It is believed that if f is uniformly continuous and bounded, $0 < E\varepsilon_1^2 < \infty$ and $n^{1/2}(\hat{\rho}_0 - \rho) = O_p(1)$ then $\hat{\tau}^2$ is consistent for τ^2. See Koul (1984) or Sievers (1984) for the proof of a similar result in the linear regression model.

REMARK 5. *Effects of asymmetry on* $\hat{\rho}$. Suppose that F is asymmetric but has mean zero so that $EX_0 = 0$. Consequently, $ES(y, \rho) \equiv 0$ and one can still define $\hat{\rho}$ by (1.5) and (1.7). Such a $\hat{\rho}$ will not have any asymptotic bias. In fact, under suitably modified assumptions that compensate for the lack of symmetry one can prove that

$$n^{1/2}(\hat{\rho} - \rho) = -\left\{EX_0^2 \int g^2\, dH\right\}^{-1} \int S(y, \rho) g(y)\, dH(y) + o_p(1),$$

where

$$g(y) = f(y) + f(-y), \qquad y \text{ in } \mathscr{R}.$$

Similarly to the arguments of Corollary 1, one can then deduce that

$$n^{1/2}(\hat{\rho} - \rho) \Rightarrow N(0, \nu^2) \text{ r.v.}$$

where now

$$\nu^2 := \left\{EX_0^2 \int g^2\, dH\right\}^{-2} \lim_{n \to \infty} \operatorname{Var}\left(n^{-1/2} \sum_{i=1}^{n} X_{i-1} G(\varepsilon_i)\right),$$

and

$$G(x) = G_0(x) - G_0(-x),$$
$$G_0(x) = \int_{-\infty}^{x} g(y)\, dH(y), \qquad x \text{ in } \mathscr{R}.$$

Direct calculations show that

$$\begin{aligned} \nu^2 &= \left\{EX_0^2 \int g^2\, dH\right\}^{-2} EX_0^2 \left\{\operatorname{Var} G(\varepsilon_1) + (1 + \rho)(1 - \rho)^{-1} [EG(\varepsilon_1)]^2\right\} \\ &= \left(\int g^2\, dH\right)^{-2} \left\{(EX_0^2)^{-1} \operatorname{Var} G(\varepsilon_1) + (1 + \rho)^2 \sigma^{-2} [EG(\varepsilon_1)]^2\right\}. \end{aligned} \tag{23}$$

Now suppose that H is symmetric around 0. Then $G_0(x) = \psi_0(x) - \psi_0(-x) + \psi_0(\infty)$, $G(x) = \psi(x) - \psi(-x) = 2\psi(x)$, for all x; ψ as in (6).

Consequently, from (13) and (15),

$$\nu^2 = \left(\int g^2\, dH\right)^{-2} \left\{\tau^2 \left(4 \int f^2\, dH\right)^2 + 4(1 + \rho)^2 \sigma^{-2} [E\psi(\varepsilon_1)]^2\right\}.$$

Therefore, from (15), and the inequality $\int g^2\, dH \le 4 \int f^2\, dH$, one gets

$$\frac{\nu^2}{\tau^2} \ge 1 + \frac{(1 + \rho)^2}{1 - \rho^2} a, \qquad a = \frac{[E\psi(\varepsilon_1)]^2}{4K(F, H)}.$$

Thus, as long as $E\psi(\varepsilon_1) \ne 0$, so that $a > 0$, one has

$$\nu^2 > \tau^2, \quad \text{for all } \rho \quad \text{and} \quad \sup_{|\rho| < 1} \frac{\nu^2}{\tau^2} = \infty! \tag{24}$$

Asymmetry of F and absolute continuity of H is enough to ensure $E\psi(\varepsilon_1) \neq 0$. Thus even though $\hat{\rho}$ has no asymptotic bias its asymptotic variance can be heavily influenced by the asymmetry of F even when $E\varepsilon_1 = 0$ and H is symmetric around 0.

REMARK 6. During the preparation of this paper the author received a preprint of the paper by Wang (1986) proposing a m.d. estimator $\hat{\rho}_w$ of ρ in (1.1) as a minimizer t of

$$W(t) = \int \left[n^{-1/2} \sum_{i=1}^{n} X_{i-1} I(X_i \le y + tX_{i-1}) \right]^2 dH(y).$$

Wang does not require symmetry of F but assumes $E\varepsilon_1 = 0$ and that H is bounded. The asymptotic variance σ_w^2 of $n^{1/2}(\hat{\rho}_w - \rho)$ turns out to be

$$\sigma_w^2 = \tau^2 + \left[(1+\rho)^2 E^2 \psi_0(\varepsilon_1) x \left\{ \sigma \int f^2 \, dH \right\}^{-2} \right], \qquad \tau^2 \text{ as in (15)}.$$

Thus, for *symmetric* F and H,

$$\tau^{-2}\sigma_w^2 = 1 + \left[(1+\rho)(1-\rho)^{-1} \left\{ [E\psi_0(\varepsilon_1)]^2 / \operatorname{Var} \psi_0(\varepsilon_1) \right\} \right],$$

which is arbitrarily large for ρ close to 1.

REMARK 7. *Connection with GM-estimators.* If in (2.5) of Denby and Martin (op. cit.) one takes $g = h$, $\psi(x) = \int_{-\infty}^{x} f \, dH$, then one has $|n^{1/2}(\hat{\rho}_h - \hat{\phi}_{\mathrm{GM}})| = o_p(1)$, where $\hat{\phi}_{\mathrm{GM}}$ is the GM-estimator. Now it is known that if $\psi(x)$ of $\hat{\phi}_{\mathrm{GM}}$ is x then $\hat{\phi}_{\mathrm{GM}} = \hat{r}$. But choosing $\psi(x) \equiv x$ would violate our assumption (A.2). Thus $\{\hat{\rho}(H)\}$ has no connection with the least-squares estimator $\hat{r}$ under the conditions of this paper.

REMARK 8. *Influence curve of $\hat{\rho}_h$.* Let

$$\begin{aligned} m(t, F, y) &:= E\{S_h(y, t)\} \\ &= Eh(X_0)[F(y + (t-\rho)X_0) + F(-y + (t-\rho)X_0) - 1], \end{aligned}$$

$$\mu(t, F) := \int m^2(t, F, y) \, dH(y).$$

Define $T(F)$ by the relation

$$\inf_t \mu(t, F) = \mu(T(F), F).$$

If F is symmetric around 0 then $T(F) = \rho$. Let L be a d.f. and define $F_s = F + s(L - F)$, $0 \le s \le 1$, T_s as a minimizer of $\mu(t, F_s)$ w.r.t. t. If F is symmetric then $T_0 = T(F) = \rho$. If $L(y) = \delta_z(y) = I(y \ge z)$ and if $\dot{T}_0 = (\partial/\partial s)T_s|_{s=0} = 0$ exists, then $\dot{T}_0$ is called the *influence curve* of $T(F)$ at F. Proceeding as in Huber (1981), one can derive, under some regularity conditions, that

$$IC(z, T, F) = \frac{Eh(X_0)}{2EX_0 h(X_0)} \frac{[\psi_0(z) - \psi_0(-z)]}{\int f \, d\psi_0}, \qquad \psi_0(y) = \int_{-\infty}^{y} f \, dH.$$

Under (A.2), ψ_0 is bounded. Thus the influence of z is bounded on $T(F)$ and hence on $\hat{\rho}_h$, at symmetric F. In particular if $h(x) = x$ and $E\varepsilon_1 = 0$, so that $EX_0 = 0$, then $\hat{\rho}$ is not influenced by the contamination of errors.

REMARK 9. *Asymptotics of scale invariant version of* $\hat{\rho}(H)$. Let $s = s(\mathbf{X}) = s(X_0, \ldots X_n)$ be an estimator such that s is positive and

$$s(b\mathbf{X}) = |b| s(\mathbf{X}), \quad \text{for all } b \text{ in } \mathscr{R}. \tag{25}$$

In addition, assume that there is a positive constant γ, possibly depending on F, such that

$$n^{1/2}(s - \gamma) = O_p(1). \tag{26}$$

Let $\rho^*(H)$ denote the scale invariant version of $\hat{\rho}(H)$ proposed near (1.7). Then, under some additional conditions on F and H, one can prove that

$$n^{1/2}(\rho^*(H) - \rho) \Rightarrow N(0, \tau_*^2) \text{ r.v.,}$$

where τ_*^2 is obtained from the τ^2 of (15) after $H(\cdot)$ there is replaced by $H(\cdot/\gamma)$. See Corollary 5.1 below for the proof of (26) for the estimator s of (2.7).

REMARK 10. *Asymptotic distributions of extensions of* $\hat{\rho}$. Recall the definition of $(\hat{\theta}, \hat{\rho})$ from (2.8)–(2.10). Under suitable assumptions one can show that the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta, \hat{\rho} - \rho)$ is bivariate normal with the asymptotic mean $\mathbf{0}$ and the asymptotic covariance matrix

$$\Sigma_1 := \begin{bmatrix} 1 & n^{-1} \sum_{i=1}^{n} EX_{i-1} \\ n^{-1} \sum_{i=1}^{n} EX_{i-1} & n^{-1} \sum_{i=1}^{n} EX_{i-1}^2 \end{bmatrix}^{-1} \times V_{\text{loc}}(\psi_0). \tag{27}$$

Next, recall the definition of $(\hat{\rho}_1, \hat{\rho}_2)$ from (2.12)–(2.14) in the stationary AR(2) model (2.11). Again, one can deduce under suitably modified assumptions that, for a symmetric F, $n^{1/2}(\hat{\rho}_1 - \rho_1, \hat{\rho}_2 - \rho_2) \Rightarrow N_2(\mathbf{0}, \Sigma_2)$ r.v.'s, where

$$\Sigma_2 := \begin{bmatrix} EX_0^2 & EX_0 X_1 \\ EX_0 X_1 & EX_0^2 \end{bmatrix}^{-1} \times V_{\text{loc}}(\psi_0). \tag{28}$$

In the above, $V_{\text{loc}}(\psi_0) = \text{Var}(\psi_0(\varepsilon_1)) / (\int \psi_0' \, dF)^2 = K(F, H) / (\int f^2 \, dH)^2$.

4. Tests of goodness of fit. Consider the model (1.1) with F as d.f. of $\{\varepsilon_i\}$. Consider the problem of testing

$$H_0\colon F(y) = 1 - F(-y), \quad \text{for all } y.$$

A natural class of tests of H_0 is given by $M_h(\hat{\rho}_h)$. The asymptotic null distribution of these statistics is derived from the following

THEOREM 1. *Under the assumptions of Theorem* 2.2

$$(1)\quad M_h(\hat{\rho}_h) = \int\left\{S_h(y,\rho) - \frac{\int S_h(y,\rho)\,d\psi_0(y)}{\int f\,d\psi_0} f(y)\right\}^2 dH(y) + o_p(1), (H_0).$$

PROOF. Follows from (3.4), (3.5), and (3.6). □

Now write $S(y)$ for $S_h(y,\rho)$. We shall throughout assume that H is symmetric around 0. For any process Y, define

$$(2)\qquad \|Y\|_H^2 := 2\int_0^\infty Y^2\,dH.$$

Then by the symmetry of H and F, the leading term on the right-hand side of (1) is $\|Z_n\|_H^2$ where, for all y real,

$$(3)\quad Z_n(y) := S(y) - g(y)\int_0^\infty S(t)\,d\psi_0(t), \qquad g(y) = f(y)\left(\int_0^\infty f\,d\psi_0\right)^{-1}.$$

Observe that

$$(4)\quad S(y) = n^{-1/2}\sum_{i=1}^n h(X_{i-1})\left[I(\varepsilon_i \le y) + I(\varepsilon_i \le -y) - 1\right], \quad \text{for all real } y.$$

Let

$$(5)\qquad \alpha(x,y) := I(x \le y) + I(x \le -y) - 1, \qquad x, y \text{ real}.$$

Then, under H_0, $E\alpha(\varepsilon_i, y) = 0$, all i, all y. Moreover, for $x, y \ge 0$, $ES(x) \equiv 0$,

$$(6)\qquad \begin{aligned} ES(x)S(y) &= Eh^2(X_0)E\alpha(\varepsilon_1,x)\alpha(\varepsilon_1,y) \\ &= b2\min(1-F(x), 1-F(y)) \qquad \left(b = Eh^2(X_0)\right), \\ &=: 2bC(x,y), \quad \text{say}. \end{aligned}$$

Then, by Fubini's theorem

$$(7)\qquad \begin{aligned} K_n(x,y) &:= \operatorname{Cov}(Z_n(x), Z_n(y)) = E(Z_n(x)Z_n(y)) \\ &= ES(x)S(y) - g(y)\int_0^\infty ES(x)S(t)\,d\psi_0(t) \\ &\quad -g(x)\int_0^\infty ES(y)S(t)\,d\psi_0(t) + g(x)g(y)E\left\{\int_0^\infty S(t)\,d\psi_0(t)\right\}^2 \\ &= 2b\left\{C(x,y) - g(y)\int_0^\infty C(x,t)\,d\psi_0(t) - g(x)\int_0^\infty C(y,t)\,d\psi_0(t)\right\} \\ &\quad +2bg(x)g(y)\int_0^\infty\int_0^\infty C(s,t)\,d\psi_0(s)\,d\psi_0(t) \\ &=: K(x,y), \quad \text{say}. \end{aligned}$$

Now, let W be a Wiener process on $[0,1]$, $W(0) = 0$, $EW = 0$, $\mathrm{Cov}(W(s), W(t)) = \min(s, t)$. Define

$$Z(x) = \left\{W(2^{1/2}(1 - F(x))) - g(x)\int_0^\infty W(2^{1/2}(1 - F(t)))\, d\psi_0(t)\right\}b^{1/2}, \qquad x \geq 0.$$

Note that $\mathrm{Cov}(Z(x), Z(y)) = K(x, y)$, $x, y \geq 0$.

Now use Theorem VI.2.1, Parthasarthy (1967), as in Millar (1981), to conclude that

$$\|Z_n\|_H^2 = 2\int_0^\infty Z_n^2(y)\, dH(y) \Rightarrow 2\int_0^\infty Z^2(y)\, dH(y) = \|Z\|_H^2, \quad \text{under } H_0.$$

Consequently, by (1),

$$M_h(\hat\rho_h) \Rightarrow \|Z\|_H^2, \quad \text{under } H_0. \tag{8}$$

But

$$\|Z\|_H^2 = 2b\left[\int_0^\infty W^2(2^{1/2}(1 - F))\, dH - \frac{\left\{\int_0^\infty W(2^{1/2}(1 - F))\, d\psi_0\right\}^2}{\int_0^\infty f\, d\psi_0}\right] \tag{9}$$
$$:= bB(W, F), \quad \text{say}.$$

Consequently, if we estimate b by

$$\hat b := n^{-1}\sum_{i=1}^n h^2(X_{i-1}) \tag{10}$$

and consider

$$\hat M_h(\hat\rho_h) := M_h(\hat\rho_h)/\hat b, \tag{11}$$

then by (3.10) and the above discussion we have

$$\hat M_h(\hat\rho_h) \Rightarrow B(W, F). \tag{12}$$

Thus, the first consequence of (12) is that the asymptotic level of $\hat M_h$-test is the same for every h. The second observation is that the limiting r.v. $B(W, F)$ is the same r.v. that arises when testing for symmetry in the one sample location model. Its distribution is accessible as in Martynov (1975, 1976) and Boos (1982).

Next, consider the problem of testing

$$H\colon F = F_0, \qquad F_0 \text{ a known d.f.} \tag{13}$$

A test of H is to reject H when $\hat T$ is large, where

$$\hat T := \int\left[V(y, \hat\rho_0) - n^{1/2}F_0(y)\right]^2 dF_0(y). \tag{14}$$

From the proof of (i) in Section 5 [see (5.8)–(5.18)] and Remark 5.1 one deduces the following

PROPOSITION. *Suppose that* (1.1) *and H hold. Assume that the error d.f. F_0 has bounded and continuous density f_0, $EX_0^2 < \infty$, $EX_0 = 0$, and that $\hat\rho_0$ is such* *that* $|n^{1/2}(\hat{\rho}_0 - \rho)| = O_p(1)$. *Then* $\hat{T} \Rightarrow \int_0^1 B^2(t)\,dt$, *where* B *is the continuous Brownian bridge on* $[0,1]$.

Consequently, the test of H based on $\hat{T}$ is asymptotically distribution free. A similar conclusion may be drawn about the test of H based on the Anderson–Darling statistic.

Pierce (1985) indicates without proof that the asymptotic null distribution of any test of the normality of errors in (1.1) is the same as that of its counterpart in the one sample location scale model. This observation is consistent with the results obtained here.

5. Proofs of Theorems 3.1 and 3.2. We shall first give a proof of Theorem 3.1. The basic proof of this theorem is similar to that of Theorem 5.1 of K-D (op. cit.), but because of the dependence of $\{X_i\}$, some calculations are intricate. We shall thus be brief, indicating only the differences.

For any functions g and k from $\mathscr{R} \times \mathscr{R}$ to $\mathscr{R}$, $|g_t|_H^2 := \int g^2(y,t)\,dH(y)$, $|g_t - k_s|_H^2 = \int [g(y,t) - k(y,s)]^2\,dH(y)$, for s, t in $\mathscr{R}$.

PROOF OF THEOREM 3.1. With $\{\rho, X_t, \varepsilon_i, F\}$ as in (1.1) and h as in Theorem 3.1, define

$$(1)\qquad J(y,t) := \int h(x)F(y + n^{-1/2}tx)\,dG(x) \qquad (G \text{ d.f. of } X_0),$$

$$(2)\qquad W(y,t) := n^{-1/2}\sum_{i=1}^{n}\{h(X_{i-1})I(\varepsilon_i \le y + n^{-1/2}tX_{i-1}) - J(y,t)\}, \qquad y, t \text{ in } \mathscr{R}.$$

Since h is held fixed, we shall now write S, Q, M for the S_h, Q_h, M_h of (3.1)–(3.3). Observe that

$$(3)\quad \text{the left-hand side of (3.4)} = E\sup_{|t|\le b}\left|M(tn^{-1/2} + \rho) - Q(n^{-1/2}t + \rho)\right|.$$

From (3.1) and (1),

$$\begin{aligned} M(n^{-1/2}t + \rho) &= \int\Big[W(y,t) + W(-y,t) + n^{1/2}\{J(y,t) + J(-y,t)\} \\ &\qquad - n^{-1/2}\sum_i h(X_{i-1})\Big]^2\,dH(y) \\ (4)\qquad &= \int\big[\{W(y,t) - W(y,0)\} + \{W(-y,t) - W(-y,0)\} \\ &\qquad + \{S(y,\rho) + 2taf(y)\} \\ &\qquad + \{n^{1/2}[J(y,t) - J(y,0)] - taf(y)\} \\ &\qquad + \{n^{1/2}[J(-y,t) - J(-y,0)] - taf(-y)\}\big]^2\,dH(y), \end{aligned}$$

where $a := (\int xh(x)\,dG(x))$ and where we have used the fact that for all y,

$$S(y,\rho) = W(y,0) + W(-y,0) - n^{-1/2}\sum_i h(X_{i-1}) + n^{1/2}[J(y,0) + J(-y,0)],$$

which in turn follows from the definitions. Also note that

$$Q(n^{-1/2}t + \rho) = \int [S(y,\rho) + 2atf(y)]^2\,dH(y). \tag{5}$$

Using the quadratic expansion and the Cauchy–Schwarz inequality on the cross product terms one gets an inequality involving $L_2(H)$ norms of the differences $W(\cdot,t) - W(\cdot,0)$ and $J(\cdot,t) - J(\cdot,0) - taf$, just like the inequality (5.6) in K-D. Thus to prove (3.4) it suffices to prove

$$\text{(i)} \qquad E\sup_t |W_t - W_0|_H^2 = o(1),$$

$$\text{(ii)} \qquad \sup_t \left|n^{1/2}[J_t - J_0] - atf\right|_H^2 = o(1) \qquad \left(a := \int xh(x)\,dG(x)\right),$$

and

$$\text{(iii)} \qquad E\sup_t |S_\rho + 2atf|_H^2 = O(1),$$

where the sup over T is over $|t| \le b$. But (iii) readily follows from (A.2) and (A.5).

PROOF OF (ii). Define $h^+(x) := h(x)I(xh(x) > 0)$, $h^- = h - h^+$. Let $J^\pm, a^\pm$ stand for the J and a of (1) and (4) when h is replaced by $h^\pm$ so that $J = J^+ + J^-$, $a = a^+ + a^-$. By the inequality $(b+c)^2 \le 2b^2 + 2c^2$ for all reals b, c, (ii) will follow if we prove it for $J^\pm$. To begin with note that the Cauchy–Schwarz inequality and Fubini's theorem together with the fact that $(h^\pm)^2 \le h^2$, imply that for fixed t,

$$\begin{aligned}
&\left|n^{1/2}[J_t^\pm - J_0^\pm] - a^\pm tf\right|_H^2 \\
&\quad\le \int\!\!\int h^2(x)\{n^{1/2}[F(y + tn^{-1/2}x) - F(y)] - txf(y)\}^2\,dG(x)\,dH(y) \\
&\quad\le 4t^2(2n^{-1/2}t)^{-1}\int_{-|t|n^{-1/2}}^{|t|n^{-1/2}}\int E\{X_0h(X_0)[f(y + sX_0) - f(y)]\}^2\,dH(y)\,ds \\
&\quad\to 0, \quad \text{by (A.2)–(A.5).}
\end{aligned} \tag{6}$$

Now observe that J^+ (J^-) is a nondecreasing (nonincreasing) function of t. This fact together with the compactness of $[-b, b]$ and (6) yields (ii) in a standard fashion. This completes the proof of (ii). We now turn to the

PROOF OF (i). Write $W^\pm$ for W when h in W is replaced by $h^\pm$. Define

$$p(y,t;x) := F(y + tn^{-1/2}x) - F(y), \qquad y, x, t \text{ in } \mathscr{R}. \tag{7}$$

Observe that for all y, t

$$
\begin{aligned}
W^{\pm}(y,t) - W^{\pm}(y,0) &= n^{-1/2}\sum_i h^{\pm}(X_{i-1})\{I(\varepsilon_i \le y + tn^{-1/2}X_{i-1}) \\
&\qquad - I(\varepsilon_i \le y) - p(y,t;X_{i-1})\} \\
&\quad + n^{-1/2}\sum_i \{h^{\pm}(X_{i-1})p(y,t;X_{i-1}) - J^{\pm}(y,t) \\
&\qquad + J^{\pm}(y,0)\} \\
&= R_1^{\pm}(y,t) + R_2^{\pm}(y,t), \quad \text{say,}
\end{aligned}
\tag{8}
$$

where

$$
\begin{aligned}
R_1^{\pm}(y,t) &= n^{-1/2}\sum_i h^{\pm}(X_{i-1}) \\
&\quad \times \{I(\varepsilon_i \leqq y + tn^{-1/2}X_{i-1}) - I(\varepsilon_i \leqq y) - p(y,t;X_{i-1})\}.
\end{aligned}
\tag{9}
$$

Note that $R_1^{\pm}$ is the sum of conditionally centered r.v.'s so that the covariance of any two summands is zero. Consequently, by Fubini's theorem for every fixed t,

$$
\begin{aligned}
E|R_{1t}^{\pm}|_H^2 &\leqq \int E\{h^{\pm}(X_0)\}^2 |F(y + tn^{-1/2}X_0) - F(y)|\, dH(y) \\
&\leqq \int_{-bn^{-1/2}}^{bn^{-1/2}} \int Eh^2(X_0)|X_0| f(y + sX_0)\, dH(y)\, ds \\
&\to 0, \quad \text{by (A.4).}
\end{aligned}
\tag{10}
$$

Next, consider $R_{2t}^{\pm}$. Rewrite

$$
\begin{aligned}
R_2^{\pm}(y,t) &= n^{-1/2}\sum_i \{h^{\pm}(X_{i-1})[F(y + tn^{-1/2}X_{i-1}) - F(y)] \\
&\qquad - Eh^{\pm}(X_{i-1})[F(y + tn^{-1/2}X_{i-1}) - F(y)]\} \\
&= n^{-1/2}\sum_i h^{\pm}(X_{i-1})[F(y + tn^{-1/2}X_{i-1}) - F(y) - tn^{-1/2}X_{i-1}f(y)] \\
&\quad - n^{-1/2}\sum_i Eh^{\pm}(X_{i-1})[F(y + tn^{-1/2}X_{i-1}) \\
&\qquad - F(y) - tn^{-1/2}X_{i-1}f(y)] \\
&\quad + tn^{-1}\sum_i [h^{\pm}(X_{i-1})X_{i-1} - Eh^{\pm}(X_{i-1})X_{i-1}]\, f(y) \\
&= A_1(y,t) - A_2(y,t) + A_3(y,t), \quad \text{say.}
\end{aligned}
\tag{11}
$$

Fubini's theorem and (A.1)–(A.3) imply that

$$
\begin{aligned}
E \sup_{|t| \le b} |A_{1t}|_H^2 &\le 4b^2(2b/\sqrt{n})^{-1} \\
&\quad \times \int_{-b/\sqrt{n}}^{b/\sqrt{n}} \int E[h^{\pm}(X_0)X_0\{f(y + sX_0) - f(y)\}]^2\, dH(y)\, ds \\
&\to 0, \quad \text{by (A.4).}
\end{aligned}
\tag{12}
$$

Note that $A_2(y, t) \equiv EA_1(y, t)$. Therefore, (12) and Fubini's theorem imply that

$$E \sup_{|t| \le b} |A_{2t}|_H^2 \le E \sup_{|t| \le b} |A_{1t}|_H^2 \to 0. \tag{13}$$

Next, because $\{X_i\}$ is stationary and ergodic and because of (A.2) and (A.3),

$$\operatorname{Var}\left\{n^{-1/2} \sum_i h^{\pm}(X_{i-1}) X_{i-1}\right\} = O(1),$$

see, e.g., Hall and Heyde (1980). Therefore

$$E \sup_{|t| \le b} |A_{3t}|_H^2 \le bn^{-1} |f|_H^2 \operatorname{Var}\left\{n^{-1/2} \sum_i h^{\pm}(X_{i-1}) X_{i-1}\right\} \to 0. \tag{14}$$

From (11)–(14) it readily follows that

$$E \sup_{|t| \le b} |R_{2t}^{\pm}|_H^2 \to 0. \tag{15}$$

From (8), (10), and (15) it follows that for each fixed t in $[-b, b]$,

$$E|W_t^{\pm} - W_0^{\pm}|_H^2 \to 0. \tag{16}$$

To complete the proof of (i) observe that

$$|W_t - W_0|_H^2 \le 2\left\{|W_t^+ - W_0^+|_H^2 + |W_t^- - W_0^-|_H^2\right\}. \tag{17}$$

Now, let $-b = t_0 < t_1 < \cdots < t_r = b$ be a partition of $[-b, b]$ such that $\max_{1 \le i \le r}(t_j - t_{j-1}) \to 0$ as $r \to \infty$. Observe that $W_t^+ (W_t^-)$ is a difference of two nondecreasing (nonincreasing) functions of t. Therefore,

$$\sup_{|t| \le b} |W_t^{\pm} - W_0^{\pm}|_H^2 \le 4\left\{\max_{0 \le j \le r} |W_{t_j}^{\pm} - W_0^{\pm}|_H^2 + \max_{0 \le j \le r} \left| n^{1/2}\left[J_{tj}^{\pm} - J_{t_{j-1}}^{\pm}\right]\right|_H^2\right\},$$

so that (16) and (6) imply

$$\begin{aligned} &\limsup_{n \to \infty} E \sup_{|t| \le b} |W_t^{\pm} - W_0^{\pm}|_H^2 \\ &\quad \le \max_{0 \le j \le r} (t_j - t_{j-1})^2 |f|_H^2 \left(E\left|X_0 h^{\pm}(X_0)\right|\right)^2 \times 16 \\ &\quad \to 0 \quad \text{as } r \to \infty. \end{aligned} \tag{18}$$

This together with (17) entails (i) and hence Theorem 3.1. □

REMARK 1. It should be emphasized that the symmetry of F is *not* required for the proof of (i) and (ii). The symmetry of F is used only in proving (iii). In fact one can also prove an analogue of Theorem 3.1 when F is not symmetric. But calculations get involved. There will be an extra term in Q. In fact, the approximating Q will be

$$Q_1(n^{-1/2}t + \rho) := \int [S(y, \rho) - ES(y, \rho) + tag(y) + ES(y, \rho)]^2 \, dH(y),$$

$$g(y) = f(y) + f(-y).$$

We shall now sketch a proof of the asymptotic normality of s of (2.7). Accordingly, let

$$F_n^*(y) := n^{-1}\sum_i I(|X_i - \hat{\rho}_0 X_{i-1}| \leq y), \qquad F^*(y) = F(y) - F(-y),$$

and

$$T_n(y) := n^{1/2}(F_n^*(y) - F^*(y)), \qquad y > 0.$$

Assume that F^* has the unique median γ. Observe that for any real u,

$$\begin{aligned} T_n(\gamma + un^{-1/2}) = {} & W_1(\gamma_n, \hat{t}) - W_1(-\gamma_{\bar{n}}, \hat{t}) \\ & + n^{1/2}[J_1(\gamma_n, \hat{t}) - J_1(-\gamma_n, \hat{t}) - F^*(\gamma_n)], \end{aligned} \tag{19}$$

where $\hat{t} = n^{1/2}(\hat{\rho}_0 - \rho)$, $\gamma_n = \gamma + un^{-1/2}$, and W_1, J_1 are the W and J of (1) and (2) above with $h \equiv 1$. Using arguments like those used for the proof of (i) [see (8)–(18) above] one can show that if f is bounded and continuous and $0 < EX_0^2 < \infty$, then for any real u, $0 < b < \infty$,

$$E \sup_{|t| \leq b} |W_1(\gamma + un^{-1/2}, t) - W_1(\gamma, 0)| \to 0, \tag{20}$$

and

$$\sup_{|t| \leq b} |n^{1/2}[J_1(\gamma + un^{-1/2}, t) - F(\gamma + un^{-1/2})] - tEX_0 f(\gamma)| \to 0.$$

From (19) and (20), if $|\hat{t}| = O_p(1)$ then

$$T_n(\gamma + un^{-1/2}) = S_n(\gamma) + \hat{t}EX_0\{f(\gamma) - f(-\gamma)\} + o_p(1), \tag{21}$$

where

$$S_n(\gamma) = n^{-1/2}\sum_i \{I(|\varepsilon_i| \leq \gamma) - F^*(\gamma)\}.$$

Moreover, continuity of f yields that, for any real u,

$$n^{1/2}[F^*(\gamma) - F^*(\gamma + un^{-1/2})] \to -u(f(\gamma) + f(-\gamma)). \tag{22}$$

Observe that $S_n(\gamma)$ is the standardized sum of i.i.d. Bernoulli $(\frac{1}{2})$ r.v.'s. Consequently, $S_n(\gamma) \Rightarrow N(0, \frac{1}{4})$. Now use the standard argument of converting the events based on the sample median to the events based on the corresponding empiricals and the above discussion to conclude the following

COROLLARY. *Suppose that* (1.1) *holds; the error d.f. F has bounded and continuous density f; F^* has unique positive median γ; $n^{1/2}|\hat{\rho}_0 - \rho| = O_p(1)$ and that $0 < EX_0^2 < \infty$. Then*

$$|n^{1/2}(s - \gamma)| = O_p(1).$$

In addition, if either (i) $EX_0 = 0$ *or* (ii) *F is symmetric around* 0, *then $n^{1/2}(s - \gamma) \Rightarrow N(0, \nu^2)$ r.v., where*

$$\begin{aligned} \nu^2 &= \{f(\gamma) + f(-\gamma)\}^{-2}, \quad \textit{in case (i)} \\ &= \{2f(\gamma)\}^{-2}, \quad \textit{in case (ii)}. \end{aligned}$$

REFERENCES

BERAN, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, R. J. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313.

BOOS, D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* **76** 663–670.

BOOS, D. (1982). A test for asymmetry associated with the Hodges–Lehmann estimator. *J. Amer. Statist. Assoc.* **77** 647–651.

CHENG, K. F. and SERFLING, R. J. (1981). On estimation of a class of efficiency-related parameters. *Scand. Actuar. J.* 83–92.

DENBY L. and MARTIN, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *J. Amer. Statist. Assoc.* **74** 140–146.

FOX, A. J. (1972). Outliers in time series. *J. Roy. Statist. Soc. Ser. B* **34** 350–363.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic, New York.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

KOUL, H. L. (1980). Some weighted empirical inferential procedures for a simple regression model. *Colloq. Math. Soc. Janos Bolyai, Nonpar. Statist. Inf.* **32** 537–565.

KOUL, H. L. (1984). *Weighted Empiricals and Linear Regression Models.* A monograph. MSU RM Series 439.

KOUL, H. L. (1985). Minimum distance estimation in multiple linear regression. *Sankhyā Ser. A* **47** 57–74.

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

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

MARTIN, R. D. (1981). Robust methods for time series. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 683–759. Academic, New York.

MARTYNOV, G. V. (1975). Computation of distribution functions of quadratic forms of normally distributed r.v.'s. *Theory Probab. Appl.* **20** 782–793.

MARTYNOV, G. V. (1976). Computation of limit distributions of statistics for normality tests of type w^2. *Theory Probab. Appl.* **21** 1–13.

MILLAR, P. W. (1981). Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55** 73–89.

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

PARR, W. C. (1981). Minimum distance estimation: A bibliography. *Comm. Statist. A—Theory Methods* **10** 1205-1224.

PARR, W. C. and DEWET, T. (1981). On minimum Cramér–von Mises norm parameter estimation. *Comm. Statist. A—Theory Methods* **10** 1149–1166.

PARR, W. C. and SCHUCANY, W. R. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75** 616–624.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

PIERCE, D. A. (1985). Testing normality in autoregressive models. *Biometrika* **72** 293–297.

SIEVERS, G. L. (1982). A consistent estimate of a nonparametric scale parameter. Inst. Statist. Mimeo series 1501, Univ. of North Carolina, Chapel Hill.

WANG, C. W. H. (1986). A minimum distance estimator for first-order autoregressive processes. *Ann. Statist.* **14** 1180–1193.

WILLIAMSON, M. (1979). Weighted empirical type estimation of the linear regression parameter. Ph.D. thesis, Michigan State Univ.

WOLFOWITZ, J. (1957). Minimum distance method. *Ann. Math. Statist.* **28** 75–88.

DEPARTMENT OF STATISTICS
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824