The Annals of Statistics
1996, Vol. 24, No. 1, 380–404

ASYMPTOTICS OF SOME ESTIMATORS AND SEQUENTIAL RESIDUAL EMPIRICALS IN NONLINEAR TIME SERIES[1]

BY HIRA L. KOUL

Michigan State University

This paper establishes the asymptotic uniform linearity of M- and R-scores in a family of nonlinear time series and regression models. It also gives an asymptotic expansion of the standardized sequential residual empirical process in these models. These results are, in turn, used to obtain the asymptotic normality of certain classes of M-, R- and minimum distance estimators of the underlying parameters. The classes of estimators considered include analogs of Hodges–Lehmann, Huber and LAD (least absolute deviation) estimators. Some applications to the change point and testing of the goodness-of-fit problems in threshold and amplitude-dependent exponential autoregression models are also given. The paper thus offers a unified functional approach to some aspects of robust inference for a large class of nonlinear time series models.

1. Introduction. This paper establishes asymptotic uniform linearity of M- and R- scores and an asymptotic expansion of the standardized randomly weighted sequential residual empirical process in a family of nonlinear time series and regression models. These results are then used to derive the large-sample distributions of certain classes of M-, R- and minimum distance (M.D.) estimators in these models. Section 2 gives some applications of these results to the estimation, goodness-of-fit testing and the change point problems in some threshold and amplitude-dependent exponential autoregression (EXPAR) models.

More precisely, let $m \wedge p \wedge q \geq 1$ be fixed integers, $n \geq m$ be an integer, Ω be an open subset of the m-dimensional Euclidean space $\mathbb{R}^m$, $\mathbb{R} = \mathbb{R}^1$, $\mathbf{t}'$ denote the transpose of a $p \times 1$ vector $\mathbf{t} \in \mathbb{R}^p$ and $\|\mathbf{t}\|$ its Euclidean norm. Let F be a distribution function (d.f.) on $\mathbb{R}$; $\{\varepsilon_i,\ i = 1, 2, \ldots\}$ be independent and identically distributed (i.i.d.) F random variables (r.v.'s); $\{\mathbf{Z}_{ni},\ i = 1, 2, \ldots, n\}$ be q-dimensional observable independent r.v.'s taking values in $\mathbb{R}^q$, independent of $\{\varepsilon_i\}$; and $\mathbf{Y}_0 := (X_0, \ldots, X_{1-p})'$ be an observable r.v., independent of both $\{\varepsilon_i,\ i = 1, 2, \ldots\}$ and $\{\mathbf{Z}_{ni},\ i = 1, 2, \ldots, n\}$; and let $\mathscr{F}_{n1} :=$ σ-field$\{\mathbf{Y}_0; \mathbf{Z}_{n1}\}$, $\mathscr{F}_{ni} := \sigma$-field$\{\mathbf{Y}_0; \varepsilon_j,\ 1 \leq j < i; \mathbf{Z}_{nk},\ 1 \leq k \leq i\}$, $2 \leq i \leq n$.

Received September 1993; revised July 1995.
[1]Research partially supported by NSF Grants DMS-91-02041 and DMS-94-02904.
AMS 1991 *subject classifications*. Primary 60F17; secondary 62M10.
Key words and phrases. Asymptotic uniform linearity, minimum distance estimators, threshold autoregression models, change point, goodness-of-fit, asymptotically distribution free.

In a pth-order nonlinear time series model considered here, one observes an array of the process $\{X_{ni},\ i = 1, 2, \dots, n\}$ satisfying the relation

$$X_{ni} = h(\boldsymbol{\theta}, \mathbf{Y}_{n,i-1}, \mathbf{Z}_{ni}) + \varepsilon_i, \qquad i = 1, 2, \dots, n, \tag{1.1}$$

for some $\boldsymbol{\theta} \in \Omega$, where $\mathbf{Y}_{n0} := \mathbf{Y}_0$, $\mathbf{Y}_{n,i-1} := (X_{n,i-1}, \dots, X_{n,i-p})'$, and h is a known function from $\Omega \times \mathbb{R}^p \times \mathbb{R}^q$ to $\mathbb{R}$ that is measurable in the last $p + q$ coordinates so that $h_{ni}(\mathbf{t}) := h(\mathbf{t}, \mathbf{Y}_{n,i-1}, \mathbf{Z}_{ni})$ is $\mathscr{F}_{ni}$-measurable, $\mathbf{t} \in \Omega$, $1 \le i \le n$. The r.v.'s $\{\mathbf{Z}_{ni}\}$ may represent regression or trend variables in these models.

The above models include classes of nonlinear regression (NLR) models and nonlinear autoregression (NLAR) models. In NLR models, $h(\boldsymbol{\theta}, \mathbf{y}, \mathbf{z}) = \gamma(\boldsymbol{\theta}, \mathbf{z})$, and in NLAR models, $h(\boldsymbol{\theta}, \mathbf{y}, \mathbf{z}) \equiv H(\boldsymbol{\theta}, \mathbf{y})$, where γ (H) is a known function from $\Omega \times \mathbb{R}^q$ $(\Omega \times \mathbb{R}^p)$ to $\mathbb{R}$ that is measurable in the last q (p) coordinates. In NLR models, $\{\mathbf{Z}_{ni}\}$ represent the known design variables that can be either random or nonrandom. Upon choosing $h(\boldsymbol{\theta}, \mathbf{y}, \mathbf{z}) \equiv H(\boldsymbol{\theta}, \mathbf{y}) + \gamma(\boldsymbol{\theta}, \mathbf{z})$ in (1.1), one obtains a class of models where autoregression and regression is present in an additive and nonlinear fashion, rather an important class of models in statistics and econometrics. A class of submodels this paper shall focus on in some detail is the family of NLAR models where $X_{ni} \equiv X_i$ satisfies

$$X_i = H(\boldsymbol{\theta}, \mathbf{Y}_{i-1}) + \varepsilon_i, \qquad i = 0, \pm 1, \pm 2, \dots, \tag{1.2}$$

where $\{\varepsilon_i,\ i = 0, \pm 1, \pm 2, \dots\}$ are i.i.d. F.

Jennrich (1969) and Wu (1981) study the asymptotics of the least squares estimator (LSE) in NLR models. Hannan (1971) contains a similar study when the errors are generated by a stationary time series. See the monograph by Seber and Wild (1989) and references therein for more on NLR models. Tong (1983, 1990) discuss numerous useful examples of NLAR models and the asymptotics of the LSE in some of these models.

Hwang and Basawa (1993) study the local asymptotic normality of a variant of the model (1.1), where $\mathbf{Z}_{ni} \equiv \mathbf{Z}_i$, with $\{\mathbf{Z}_i\}$ being i.i.d. r.v.'s, not necessarily observable, and where $\{X_i\}$ are assumed to be stationary. They also discuss the asymptotics of some likelihood based tests in (1.1) and of the one-step MLE for a class of NLAR models (1.2).

The two examples we shall focus on in some detail are a self-exciting threshold first-order autoregression [SETAR(2; 1, 1)] and EXPAR models. The former is obtained from (1.2) upon taking

$$m = 2,\ p = 1 \text{ and } H(\boldsymbol{\theta}, y) \equiv \theta_1 y^+ + \theta_2 y^-, \qquad y \in \mathbb{R}, \boldsymbol{\theta} \in (0,1)^2, \tag{1.3}$$

while the choice of

$$\begin{aligned} &m = 3,\ p = 1 \text{ and } H(\boldsymbol{\theta}, y) = \{\theta_1 + \theta_2 \exp(-\theta_3 y^2)\} y, \\ &\qquad\qquad y \in \mathbb{R}, \boldsymbol{\theta} \in (-1, 1) \times \mathbb{R} \times (0, \infty), \end{aligned} \tag{1.4}$$

in (1.2) gives the latter model. Here, $x^+ := \max\{0, x\}$, $x^- := x^+ - x$, $x \in \mathbb{R}$. A general class of SETAR and EXPAR models of Tong (1990) are also included in (1.2).

To proceed further, fix a $\boldsymbol{\theta} \in \Omega$ and let $P_{\boldsymbol{\theta}}^n$ denote the probability distribution of $(\mathbf{Y}_0, X_{n1}, \dots, X_{nn})$ under (1.1) when $\boldsymbol{\theta}$ is the true parameter value. About h we shall assume the following:

(h1) *There exists a vector of functions* $\dot{\mathbf{h}}_n$ *from* $\Omega \times \mathbb{R}^p \times \mathbb{R}^q$ *to* $\mathbb{R}^m$ *such that* $\dot{\mathbf{h}}_{ni}(\mathbf{t}) := \dot{\mathbf{h}}_n(\mathbf{t}, \mathbf{Y}_{n,i-1}, \mathbf{Z}_{ni})$ *is* $\mathscr{F}_{ni}$*-measurable*, $\mathbf{t} \in \Omega$, $1 \le i \le n$, *and satisfies the following*: $\forall\ \alpha > 0$, $k < \infty$, $\mathbf{s} \in \Omega$,

$$\limsup_n P_{\boldsymbol{\theta}}^n \left(\sup_{1 \le i \le n,\ \|\mathbf{t}-\mathbf{s}\| \le kn^{-1/2}} \frac{\left| h_{ni}(\mathbf{t}) - h_{ni}(\mathbf{s}) - (\mathbf{t}-\mathbf{s})'\dot{\mathbf{h}}_{ni}(\mathbf{s}) \right|}{\|\mathbf{t}-\mathbf{s}\|} > \alpha \right) = 0.$$

Note that there is no loss of generality in assuming the $\mathscr{F}_{ni}$-measurability of $\dot{\mathbf{h}}_{ni}$ in (h1). Also, the differentiability of $h(\mathbf{t}, \mathbf{y}, \mathbf{z})$ in $\mathbf{t}$, for all $\mathbf{y}$ and $\mathbf{z}$, alone need not imply (h1). An example that illustrates this point is when $m = 1 = p$ and $h(t, y, z) \equiv t^2 y$. This h, even though differentiable in t, does not satisfy (h1) unless $\max\{n^{-1/2}|X_{i-1}|;\ 1 \le i \le n\}$ tends to 0 in probability. The above examples (1.3) and (1.4) are shown to satisfy (h1) in Section 2 provided the errors have finite variance and positive Lebesgue density.

We now define various processes and scores that are useful for inference in these models. From now on the dependence of X_{ni}, h_{ni}, etc., on n will not be exhibited, for convenience. Thus, we shall write X_i for X_{ni} and so on. Let ψ and φ be bounded nondecreasing real-valued functions on $\mathbb{R}$ and $(0,1)$, respectively; $R_{i\mathbf{t}}$ denote the rank of $X_i - h_i(\mathbf{t})$ among $\{X_j - h_j(\mathbf{t}),\ 1 \le j \le n\}$, $1 \le i \le n$, $\mathbf{t} \in \Omega$; and G be a d.f. on $\mathbb{R}$. Define

$$\mathbf{M}(\mathbf{t}) := n^{-1/2} \sum_i \dot{\mathbf{h}}_i(\mathbf{t}) \psi(X_i - h_i(\mathbf{t})),$$

$$\mathbf{Z}(u, \mathbf{t}) := n^{-1/2} \sum_i \dot{\mathbf{h}}_i(\mathbf{t}) I(R_{i\mathbf{t}} \le nu),$$

$$\mathbb{Z}(u, \mathbf{t}) := \mathbf{Z}(u, \mathbf{t}) - n^{-1/2} \sum_i \dot{\mathbf{h}}_i(\mathbf{t}) u, \qquad 0 \le u \le 1,$$

$$\mathbf{S}(\mathbf{t}) := \int_0^1 \varphi(u) \mathbf{Z}(du, \mathbf{t}) = n^{-1/2} \sum_i \dot{\mathbf{h}}_i(\mathbf{t}) \varphi(R_{i\mathbf{t}}/(n+1)),$$

$$\mathbf{K}(\mathbf{t}) := \int_0^1 \|\mathbb{Z}(u, \mathbf{t})\|^2 \, dG(u), \qquad \mathbf{t} \in \Omega.$$

Here, and in the sequel, the index i in the summation and the maximum varies from 1 to n, unless specified otherwise. The M-, R- and M.D.-estimators of $\boldsymbol{\theta}$ to be considered are, respectively,

$$\hat{\boldsymbol{\theta}}_M := \operatorname{argmin}_{\mathbf{t}} \|\mathbf{M}(\mathbf{t})\|, \qquad \hat{\boldsymbol{\theta}}_R := \operatorname{argmin}_{\mathbf{t}} \|S(\mathbf{t})\|, \qquad \hat{\boldsymbol{\theta}}_{\mathrm{md}} := \operatorname{argmin}_{\mathbf{t}} K(\mathbf{t}).$$

The name R-estimator for $\hat{\boldsymbol{\theta}}_R$ is borrowed from linear regression with known design where its analogue is a measurable function of the residual ranks only. But here $\mathbf{S}$ is a measurable function of observations and the residual ranks; strictly speaking, $\hat{\boldsymbol{\theta}}_R$ is thus not a rank estimator in the same sense as in the linear regression setup.

Analogues of the above estimators and their asymptotic distribution theory in linear regression and autoregression (LRAR) models appear in Jurečková (1971), Huber (1981), Bustos (1982), Koul (1992) and Koul and Ossiander (1994), among others. Asymptotically efficient estimators at various error distributions can be obtained from these classes of estimators in these models. Well-known examples of *M*-estimators are obtained by choosing $\psi(x) \equiv xI(|x| \le k) + c\,\mathrm{sgn}(x)I(|x| > k)$, k a constant, or $\psi(x) \equiv \mathrm{sgn}(x)$, thereby giving the Huber(k) and the least absolute deviation (LAD) estimators, respectively. A useful example of an *R*-estimator is obtained by choosing $\varphi(u) \equiv u$, thereby giving the Hodges–Lehmann type estimator which is known to be asymptotically efficient at the logistic errors in LRAR models. An interesting example of an M.D.-estimator is obtained upon taking $G(u) \equiv u$. Koul (1992) and Koul and Ossiander (1994) observed that this estimator is asymptotically more efficient than the Hodges–Lehmann type (LAD) estimator in an LRAR model with double exponential (logistic) errors. It was also noted that a large class of estimators $\{\hat{\boldsymbol\theta}_{\mathrm{md}}\}$ are asymptotically more efficient than the least squares estimator at heavy tail distributions in these models.

In view of these desirable properties, it is only natural to investigate the large-sample distributions of the above three classes of estimators for the class of models given in (1.1) and (h1). One of the goals of this paper is to do precisely that. We shall now state additional assumptions and the results to that effect. In what follows, $\|g\|_\infty := \sup\{|g(x)|;\ x \in \mathbb{R}\}$ for any $g\colon \mathbb{R}$ to $\mathbb{R}$; $N_b := \{\mathbf{t} \in \Omega;\ \|\mathbf{t}\| \le b\}$, $0 < b < \infty$; $o_p(1)$ $[O_p(1)]$ stands for a sequence of r.v.'s that converges to 0 (is bounded) in probability under $P_{\boldsymbol\theta}^{n\cdot}$, $E_{\boldsymbol\theta}^n$ stands for expectation under $P_{\boldsymbol\theta}^{n\cdot}$, and all limits are taken as $n \to \infty$, unless specified otherwise. With $\boldsymbol\theta \in \Omega$ fixed throughout, consider the following assumptions.

(F) F has a uniformly continuous density f which is positive *a.e.*

(h2) $n^{-1}\sum_i \dot{\mathbf{h}}_i(\boldsymbol\theta)\dot{\mathbf{h}}_i'(\boldsymbol\theta) = \boldsymbol\Sigma_{\boldsymbol\theta} + o_p(1)$, where $\boldsymbol\Sigma_{\boldsymbol\theta}$ is a positive-definite matrix.

(h3)
$$n^{-1/2} \max_i \left\| \dot{\mathbf{h}}_i(\boldsymbol\theta) \right\| = o_p(1).$$

(h4)
$$n^{-1} \sum_i E_{\boldsymbol\theta}^n \left\| \dot{\mathbf{h}}_i(\boldsymbol\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_i(\boldsymbol\theta) \right\|^2 = o(1), \qquad \mathbf{t} \in \Omega.$$

(h5)
$$n^{-1/2} \sum_i \left\| \dot{\mathbf{h}}_i(\boldsymbol\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_i(\boldsymbol\theta) \right\| = O_p(1), \qquad \mathbf{t} \in \Omega.$$

(h6) For every $\alpha > 0$, there exists a $\delta > 0$ and an $N < \infty$, such that

$$P_{\boldsymbol\theta}^n\left(\sup_{\|\mathbf{t}-\mathbf{s}\| \le \delta} n^{-1/2} \sum_i \left\| \dot{\mathbf{h}}_i(\boldsymbol\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_i(\boldsymbol\theta + n^{-1/2}\mathbf{s}) \right\| \le \alpha \right) \ge 1 - \alpha,$$

for all $n > N$.

REMARK 1.1. If the underlying process and h are such that $\{\dot{\mathbf{h}}_i(\boldsymbol{\theta})\}$ does not depend on n and is stationary and ergodic, then the distribution of $\mathbf{Y}_0$ will depend on $\boldsymbol{\theta}$ and the following hold: conditions (h2) and (h3) are implied by

$$(\mathrm{h_s}1)\qquad E_{\boldsymbol{\theta}}\left\|\dot{\mathbf{h}}_1(\boldsymbol{\theta})\right\|^2 < \infty,$$

while (h4) is equal to

$$(\mathrm{h_s}2)\qquad E_{\boldsymbol{\theta}}\left\|\dot{\mathbf{h}}_1(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_1(\boldsymbol{\theta})\right\|^2 = o(1), \qquad \mathbf{t} \in \Omega,$$

and assumption (h5) is implied by

$$(\mathrm{h_s}3)\qquad n^{1/2}E_{\boldsymbol{\theta}}\left\|\dot{\mathbf{h}}_1(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_1(\boldsymbol{\theta})\right\| = O(1), \qquad \mathbf{t} \in \Omega,$$

where $E_{\boldsymbol{\theta}}$ denotes the expectation under the stationary distribution.

Tong (1990) contains various sufficient conditions for a nonlinear autoregression model to be stationary and ergodic. Some of these are discussed in Section 2 for examples (1.3) and (1.4) mentioned above.

We shall now state a lemma which is basic to the proof of the asymptotic uniform linearity of the above scores. Let

$$(1.5)\qquad \begin{aligned} \mathbf{V}(y, \mathbf{t}) &:= n^{-1/2} \sum_i \dot{\mathbf{h}}_i(\mathbf{t}) I(X_i - h_i(\mathbf{t}) \le y), \\ d_{ni}(\mathbf{t}) &:= h_i(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - h_i(\boldsymbol{\theta}), \qquad \dot{\mathbf{h}}_{ni}(\mathbf{t}) := \dot{\mathbf{h}}_i(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}), \qquad 1 \le i \le n, \\ \boldsymbol{\nu}(y, \mathbf{t}) &:= n^{-1/2} \sum_i \dot{\mathbf{h}}_{ni}(\mathbf{t}) F(y + d_{ni}(\mathbf{t})), \\ \mathbb{W}(y, t) &:= \mathbf{V}(y, \boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \boldsymbol{\nu}(y, \mathbf{t}), \qquad y \in \mathbb{R}, \mathbf{t} \in \Omega. \end{aligned}$$

LEMMA 1.1. *Suppose* (1.1) *and* (h1)–(h6) *hold. Then, for every y at which F is continuous and for every* $0 < b < \infty$,

$$(1.6)\qquad \sup_{\mathbf{t} \in N_b} \|\mathbb{W}(y, \mathbf{t}) - \mathbb{W}(y, \mathbf{0})\| = o_p(1).$$

If, in addition, (F) *holds, then, for every* $0 < b < \infty$,

$$(1.7)\qquad \sup \|\mathbb{W}(y, \mathbf{t}) - \mathbb{W}(y, \mathbf{0})\| = o_p(1)$$

and

$$(1.8)\qquad \begin{aligned} \sup\Big\| &\mathbf{V}(y, \boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \mathbf{V}(y, \boldsymbol{\theta}) - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{t} f(y) \\ &- n^{-1/2} \sum_i \left[\dot{\mathbf{h}}_i(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_i(\boldsymbol{\theta})\right] F(y) \Big\| = o_p(1), \end{aligned}$$

where the supremum in (1.7) *and* (1.8) *is over* $(y, \mathbf{t}) \in \mathbb{R} \times N_b$.

The result (1.8) directly follows from (1.7) and (F). Proofs of (1.6) and (1.7) appear in Section 3, soon after Lemma 3.2. To state other results, we also need

$$\Psi := \Big\{\psi\colon \mathbb{R} \text{ to } \mathbb{R}, \text{ nondecreasing right continuous}, \int \psi\, dF = 0,\ \psi(\infty) - \psi(-\infty) \le 1\Big\},$$

$$\Phi := \{\varphi\colon [0,1] \text{ to } \mathbb{R}, \text{ nondecreasing right continuous}, \varphi(1) - \varphi(0) \le 1\}.$$

We are now ready to state the asymptotic uniform linearity result of the **M**-scores.

THEOREM 1.1. *In addition to* (1.1), *assume that* (F) *and* (h1)–(h6) *hold. Then, for every* $0 < b < \infty$,

$$\sup_{\psi\in\Psi,\, \mathbf{t}\in N_b} \left\| \mathbf{M}(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \mathbf{M}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{t}\int f\, d\psi \right\| = o_p(1). \tag{1.9}$$

The result (1.9) follows from (1.8), the fact that $\int F\, d\psi = \psi(\infty)$ for all $\psi \in \Psi$ and that

$$\mathbf{M}(\mathbf{t}) - \mathbf{M}(\mathbf{s}) = n^{-1/2} \sum_i \big[\dot{\mathbf{h}}_i(\mathbf{t}) - \dot{\mathbf{h}}_i(\mathbf{s})\big]\psi(\infty) - \int [\mathbf{V}(y,\mathbf{t}) - \mathbf{V}(y,\mathbf{s})]\, d\psi(y),$$

uniformly in $\mathbf{t}, \mathbf{s}$ in Ω and $\psi \in \Psi$, with probability 1.

To use (1.9) in establishing the asymptotic normality of $\hat{\boldsymbol{\theta}}_M$, one must first ensure that $\|n^{1/2}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta})\| = O_p(1)$. In view of the fact that $\|\mathbf{M}(\boldsymbol{\theta})\| = O_p(1)$, a sufficient condition for this is as follows:

(B1) *For every* $\varepsilon > 0$, $0 < \alpha < \infty$, *there exists an* N_ε *and a* k (*depending on* ε *and* α) *such that*

$$P^n_{\boldsymbol{\theta}}\Big(\inf_{\|\mathbf{t}\|>k} \|\mathbf{M}(\boldsymbol{\theta} + n^{-1/2}\mathbf{t})\| \ge \alpha\Big) \ge 1 - \varepsilon, \qquad \forall\, n > N_\varepsilon.$$

A condition that, in turn, implies (B1) is given by (M1) below.

COROLLARY 1.1. *In addition to the assumptions of Theorem* 1.1, *assume that* $\int f\, d\psi > 0$ *and*

(M1) $\mathbf{e}'\mathbf{M}(\boldsymbol{\theta} + n^{-1/2}r\mathbf{e})$ *is monotonic in* $r \in \mathbb{R}$, $\forall\, \mathbf{e} \in \mathbb{R}^m$, $\|\mathbf{e}\| = 1$, $n \ge 1$, *a.s.*

Then, $\forall\, \psi \in \Psi$,

$$n^{1/2}\big(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\big) = \Big\{\int d\psi\, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\Big\}^{-1} \mathbf{M}(\boldsymbol{\theta}) + o_p(1) \tag{1.10}$$

and

$$n^{1/2}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}) \Rightarrow N_q(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\nu(\psi, F)), \tag{1.11}$$

where $\nu(\psi, F) := \int \psi^2 \, dF / (\int f \, d\psi)^2$.

REMARK 1.2. As mentioned above, (M1) implies $\|n^{1/2}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta})\| = O_p(1)$. This and (1.9) together then imply (1.10) in a routine fashion. The claim (1.11) follows from (1.10) and the CLT for martingales as given in Corollary 3.1 in Hall and Heyde (1980).

The assumption (M1) is, for example, always satisfied by those models where h of (1.1) or H of (1.2) is linear in parameters and nonlinear in the remaining arguments. Tong (1990) contains numerous examples of these types of nonlinear time series models. It is a useful condition when ψ is not differentiable as is the case for the Huber(k) and the least absolute deviation estimators.

If ψ is twice differentiable and $\{\dot{\mathbf{h}}_i\}$ are differentiable and satisfy certain additional integrability conditions, then using standard arguments *à la* Cramér, one can verify (B1) directly. See, for example, Tjøstheim (1986) in connection with the maximum likelihood and the least squares estimators in nonlinear times series models of the type (1.2). Obviously, this method does not work if ψ is not differentiable.

The following corollary gives an analogue of Corollary 1.1 for the least absolute deviation estimator under weaker conditions. Its proof uses (1.6) and a routine argument.

COROLLARY 1.2. *Assume* (1.1) *and* (h1)–(h6) *hold. In addition, if* (M1) *with* $\psi(x) \equiv \operatorname{sgn}(x)$ *holds and if the d.f. F has density f in an open neighborhood of* 0 *such that f is positive and continuous at* 0, *then* $n^{1/2}(\hat{\boldsymbol{\theta}}_{\text{lad}} - \boldsymbol{\theta}) \Rightarrow N(0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}/4f^2(0))$.

To state analogous results for R- and M.D.-scores and estimators, we need to define

$$\bar{\varphi} := n^{-1} \sum_i \varphi(i/(n+1)), \qquad \dot{\mathbf{h}}^*(\boldsymbol{\theta}) := n^{-1} \sum_i \dot{\mathbf{h}}_i(\boldsymbol{\theta}),$$

$$\hat{\mathbf{Z}}(u) := n^{-1/2} \sum_i \big(\dot{\mathbf{h}}_i(\boldsymbol{\theta}) - \dot{\mathbf{h}}^*(\boldsymbol{\theta})\big)\big[I(F(\varepsilon_i) \le u) - u\big],$$

$$q(u) := f(F^{-1}(u)), \qquad 0 \le u \le 1,$$

$$\hat{\mathbf{S}} := n^{-1/2} \sum_i \big(\dot{\mathbf{h}}_i(\boldsymbol{\theta}) - \dot{\mathbf{h}}^*(\boldsymbol{\theta})\big)\big[\varphi(F(\varepsilon_i)) - \bar{\varphi}\big].$$

THEOREM 1.2. *In addition to* (1.1), (F), (h1) *and* (h3)–(h6), *assume that*

$$\text{(h7)} \qquad n^{-1}\sum_i \big(\dot{\mathbf{h}}_i(\boldsymbol\theta)-\dot{\mathbf{h}}^*(\boldsymbol\theta)\big)\big(\dot{\mathbf{h}}_i(\boldsymbol\theta)-\dot{\mathbf{h}}^*(\boldsymbol\theta)\big)' = \boldsymbol\Gamma_{\boldsymbol\theta} + o_p(1)$$

for some positive-definite matrix $\boldsymbol\Gamma_{\boldsymbol\theta}$. *Then, for every* $0 < b < \infty$,

$$\sup_{0\le u\le 1,\,\mathbf{t}\in N_b} \big\|\mathbb{Z}(u,\boldsymbol\theta+n^{-1/2}\mathbf{t}) - \hat{\mathbb{Z}}(u) - \boldsymbol\Gamma_{\boldsymbol\theta}\mathbf{t}q(u)\big\| = o_p(1), \tag{1.12}$$

$$\sup_{\varphi\in\Phi,\,\mathbf{t}\in N_b} \bigg\|\mathbf{S}(\boldsymbol\theta+n^{-1/2}\mathbf{t}) - n^{1/2}\dot{\mathbf{h}}^*(\boldsymbol\theta)\overline{\varphi} - \hat{\mathbf{S}} + \Gamma_{\boldsymbol\theta}\mathbf{t}\int f\,d\varphi(F)\bigg\| = o_p(1) \tag{1.13}$$

and

$$\sup_{G\in\Phi,\,\mathbf{t}\in N_b} \big|K(\boldsymbol\rho+n^{-1/2}\mathbf{t}) - \hat{K}(\mathbf{t})\big| = o_p(1), \tag{1.14}$$

where $\hat{K}(\mathbf{s}) := \int_0^1 \|\hat{\mathbb{Z}}(u) + \boldsymbol\Gamma_{\boldsymbol\theta} q(u)\mathbf{s}\|^2\, dG(u)$, $\mathbf{s}\in\mathbb{R}^q$.

REMARK 1.3. Note that (1.14) follows from (1.12) trivially, while (1.13) follows from (1.12) and the fact that $\mathbf{S}(\mathbf{t}) \equiv \dot{\mathbf{h}}^*(\boldsymbol\theta)\overline{\varphi} - \int \mathbb{Z}(u(n+1)/n, \mathbf{t})\, d\varphi(u)$. The result (1.13) gives the asymptotic uniform linearity of the **S**-score, while (1.14) gives the asymptotic uniform quadraticity of the dispersion K. The proof of (1.12) uses (1.7) in a crucial way and appears in Section 3.

COROLLARY 1.3. (a) *In addition to the assumptions of Theorem* 1.2, *assume that*

(S1) $\mathbf{e}'\mathbf{S}(\boldsymbol\theta + n^{-1/2} r\mathbf{e})$ *is monotonic in* $r\in\mathbb{R}$, $\mathbf{e}\in\mathbb{R}^m$, $\|\mathbf{e}\| = 1$, $n \ge 1$, *a.s.*,

and that

$$(\varphi 1) \qquad \overline{\varphi} \equiv 0 \quad \text{for all sufficiently large } n \ge 1.$$

Then

$$n^{1/2}\big(\hat{\boldsymbol\theta}_R - \boldsymbol\theta\big) = \Big\{\int f\,d\varphi(F)\boldsymbol\Gamma_{\boldsymbol\theta}\Big\}^{-1}\hat{\mathbf{S}} + o_p(1) \tag{1.15}$$

and

$$n^{1/2}\big(\hat{\boldsymbol\theta}_R - \boldsymbol\theta\big) \Rightarrow N_q\big(\mathbf{0}, \boldsymbol\Gamma_{\boldsymbol\theta}^{-1}\mu(\varphi, F)\big), \tag{1.16}$$

where $\mu(\varphi, F) := \int \varphi^2(u)\,du/\{\int f\,d\varphi(F)\}^2$.

COROLLARY 1.4. (a) *In addition to the assumptions of Theorem* 1.2, *assume that, for some* $0 < g \in L_2([0,1], G)$,

(K1) $\int_0^1 \mathbf{e}'\mathbb{Z}(u, \boldsymbol\theta + n^{-1/2} r\mathbf{e}) g(u)\, dG(u)$ *is monotonic in* $r\in\mathbb{R}$, $\forall$ $\mathbf{e}\in\mathbb{R}^m$, $\|\mathbf{e}\| = 1$, $n\ge 1$, *a.s.*

Then

$$n^{1/2}(\hat{\boldsymbol\theta}_{\mathrm{md}} - \boldsymbol\theta) = \left\{\int_0^1 q^2\, dG\, \boldsymbol\Gamma_{\boldsymbol\theta}\right\}^{-1} \int_0^1 n^{1/2}\hat{\mathbb{Z}}(u) q(u)\, dG(u) + o_p(1) \tag{1.17}$$

and

$$n^{1/2}\big(\hat{\boldsymbol\theta}_{\mathrm{md}} - \boldsymbol\theta\big) \Rightarrow N_q\big(\mathbf{0}, \boldsymbol\Gamma_{\boldsymbol\theta}^{-1} m(G,F)\big), \tag{1.18}$$

where

$$m(G,F) := \int_0^1\int_0^1 [u \wedge v - uv] q(u) q(v)\, dG(u)\, dG(v) \Big/ \left(\int_0^1 q^2\, dG\right)^2.$$

REMARK 1.4. Assumptions (S1) and (K1) ensure the $n^{1/2}$-consistency of the R- and M.D.-estimators, respectively. They may be replaced by any other assumptions that will imply an analogue of (B1) for R- and M.D.-scores. In any case, (1.13) and (1.14) together with (S1) and (K1) yield (1.15) and (1.17) in a routine fashion, while (1.16) and (1.18) follow from these results and the CLT for martingales as given in Corollary 3.1 in Hall and Heyde (1980). See, for example, Koul (1992), Sections 5.4 and 7.4, for the type of argument needed.

Akin to (M1), conditions (S1) and (K1) are always satisfied by those models where h of (1.1) or H of (1.2) is linear in parameters and nonlinear in the remaining arguments. See Section 2 for illustrations.

Next, we shall give an asymptotic expansion of the sequential residual empirical process. We, in fact, prove a somewhat general result of broader applicability from which this will follow. To that effect, let $\{g_{ni}, 1 \le i \le n\}$ be an array of r.v.'s, with g_{ni} being $\mathscr{F}_{ni}$-measurable and independent of ε_i, $1 \le i \le n$. Define, for $y \in \mathbb{R}$, $\mathbf{t} \in \Omega$, $u \in [0,1]$,

$$\begin{aligned} T_g(y,\mathbf{t},u) &:= n^{-1/2} \sum_{i=1}^{[nu]} g_{ni} I\big(\varepsilon_i \le y + h_i(\mathbf{t}) - h_i(\boldsymbol\theta)\big), \\ \tilde{T}_g(y,\mathbf{t},u) &:= n^{-1/2} \sum_{i=1}^{[nu]} g_{ni}\big[I\big(\varepsilon_i \le y + h_i(\mathbf{t}) - h_i(\boldsymbol\theta)\big) \\ &\qquad - F\big(y + h_i(\mathbf{t}) - h_i(\boldsymbol\theta)\big)\big], \end{aligned} \tag{1.19}$$

where $[x] \equiv$ the greatest integer in x. Write $T_g(y,\mathbf{t})$ and so on for $T_g(y,\mathbf{t},1)$ and so on. Also, all probability statements in the following theorem are understood to be under the joint distribution of $\{\mathbf{Y}_0, X_i, g_{ni}, 1 \le i \le n\}$. We have the following result.

THEOREM 1.3. (a) *In addition to* (1.1), (F) *and* (h1)–(h3), *assume the following*:

(g1) $$\left(n^{-1}\sum_i g_{ni}^2\right)^{1/2} = g + o_p(1), \qquad g \quad \text{a positive r.v.}$$

(g2) $$n^{-1/2}\max_i |g_{ni}| = o_p(1).$$

(g3) $$n^{-1}\sum_i E g_{ni}^2 = O(1).$$

Then the process $\{\tilde{T}_g(y,\boldsymbol{\theta}),\, y \in \mathbb{R}\}$ *is eventually tight, and, for every* $0 < b < \infty$,

$$\sup_{y,\mathbf{t}} \left|\tilde{T}_g(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - \tilde{T}_g(y,\boldsymbol{\theta})\right| = o_p(1) \tag{1.20}$$

and

$$\sup_{y,\mathbf{t}} \left|T_g(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - T_g(y,\boldsymbol{\theta}) - \mathbf{t}'n^{-1}\sum_i g_{ni}\dot{\mathbf{h}}_i(\boldsymbol{\theta})f(y)\right| = o_p(1), \tag{1.21}$$

where the supremum is over $(y,\mathbf{t}) \in \mathbb{R} \times N_b$.

(b) *In addition to* (1.1), (F), (h1) *and* (h3), *assume that the following hold*:

(g4) $$n^{-1}\sum_i E g_{ni}^4 = O(1).$$

(g5) $$n^{-1}\sum_i E g_{ni}^4 \left\|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\right\|^2 = O(1).$$

(g6) $$\sum_i E g_{ni}^4 \left[h_i(\boldsymbol{\theta} + n^{-1/2}\mathbf{s}) - h_i(\boldsymbol{\theta})\right]^2 = O(1), \qquad \mathbf{s} \in \Omega.$$

Then, for every $0 < b < \infty$,

$$\sup_{y,\mathbf{t},u} \left|\tilde{T}_g(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}, u) - \tilde{T}_g(y,\boldsymbol{\theta}, u)\right| = o_p(1) \tag{1.22}$$

and

$$\begin{aligned}\sup_{y,\mathbf{t},u} \Bigg| & T_g(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}, u) - T_g(y,\boldsymbol{\theta}, u) \\ & - \mathbf{t}'n^{-1}\sum_{i=1}^{[nu]} g_{ni}\dot{\mathbf{h}}_i(\boldsymbol{\theta})f(y)\Bigg| = o_p(1),\end{aligned} \tag{1.23}$$

where now the supremum is over $(y,\mathbf{t},u) \in \mathbb{R} \times N_b \times [0,1]$.

A proof of the above theorem also appears in Section 3. Now let

$$F_n(y,\mathbf{t},u) := n^{-1}\sum_{i=1}^{[nu]} I(X_i - h_i(\mathbf{t}) \le y),$$

$$\nu(y,\mathbf{t},u) := n^{-1}\sum_{i=1}^{[nu]} F(y + h_i(\mathbf{t}) - h_i(\boldsymbol{\theta})),$$

$$W(y,\mathbf{t},u) := n^{1/2}[F_n(y,\mathbf{t},u) - \nu(y,\mathbf{t},u)], \qquad y \in \mathbb{R}, \mathbf{t} \in \Omega\ 0 \le u \le 1.$$

Upon choosing $g_{ni} \equiv 1$ in Theorem 1.3, one obtains an analogous result for F_n. Because of its importance we state it as a separate result, though at the cost of some repetition. Note that in this case (g1)–(g4) are trivially satisfied while (g5) and (g6) are, respectively, equal to

$$\text{(h3*)} \qquad n^{-1}\sum_i E_{\boldsymbol{\theta}}^n \|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\|^2 = O(1),$$

$$\text{(h8)} \qquad \sum_i E_{\boldsymbol{\theta}}^n [h_i(\boldsymbol{\theta} + n^{-1/2}\mathbf{s}) - h_i(\boldsymbol{\theta})]^2 = O_p(1), \qquad \mathbf{s} \in \Omega.$$

We are now ready to state the following corollary.

COROLLARY 1.5. (a) *Assume that* (1.1), (F), (h1) *and* (h3) *hold. Then*, $\forall\ 0 < b < \infty$,

$$\sup_{y,\mathbf{t}} |W(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - W(y,\boldsymbol{\theta})| = o_p(1). \tag{1.24}$$

If, in addition,

$$\text{(h9)} \qquad n^{-1}\sum_i \|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\| = O_p(1),$$

then, $\forall\ 0 < b < \infty$,

$$\sup_{y,\mathbf{t}} \Big| n^{1/2}[F_n(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t}) - F_n(y,\boldsymbol{\theta})] - \mathbf{t}' n^{-1}\sum_i \dot{\mathbf{h}}_i(\boldsymbol{\theta}) f(y) \Big| = o_p(1), \tag{1.25}$$

where the supremum here and in (1.24) *is over* $(y,\mathbf{t}) \in \mathbb{R} \times N_b$.

(b) *Assume that* (1.1), (F), (h1), (h3*) *and* (h8) *hold. Then* $\forall\ 0 < b < \infty$,

$$\sup_{y,\mathbf{t},u} |W(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t},u) - W(y,\boldsymbol{\theta},u)| = o_p(1) \tag{1.26}$$

and

$$\sup_{y,\mathbf{t},u} \Big| n^{1/2}[F_n(y,\boldsymbol{\theta} + n^{-1/2}\mathbf{t},u) - F_n(y,\boldsymbol{\theta},u)] - \mathbf{t}' n^{-1}\sum_{i=1}^{[nu]} \dot{\mathbf{h}}_i(\theta) f(y) \Big| = o_p(1), \tag{1.27}$$

where now the supremum is over $(y,\mathbf{t},u) \in \mathbb{R} \times N_b \times [0,1]$.

From (1.25) we readily obtain the following corollary.

COROLLARY 1.6. *In addition to* (1.1), (F), (h1), (h3) *and* (h9), *assume that there exists an estimator* $\hat{\boldsymbol{\theta}}$ *such that*

$$\left\| n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\| = O_p(1). \tag{1.28}$$

Then

$$\sup_{y \in \mathbb{R}} \left| n^{1/2}\left[F_n(y, \hat{\boldsymbol{\theta}}) - F_n(y, \boldsymbol{\theta})\right] - n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' n^{-1} \sum_i \dot{\mathbf{h}}_i(\boldsymbol{\theta}) f(y) \right| = o_p(1). \tag{1.29}$$

Perhaps it is worth mentioning that the underlying time series need not be stationary for the validity of all of the above results. The next section discusses some applications of these results to estimation, goodness-of-fit testing and the change point problems in SETAR and EXPAR models. The procedures discussed for the change point problem are similar to those discussed in Bai (1994) for ARMA models.

2. Applications. In what follows, (Hj) and so on stand for (hj) and so on, with h replaced by H and (1.1) replaced by (1.2). Also, (F^+) stands for condition (F), where now the density f is assumed to be positive *everywhere*. This condition is needed to ensure the stationarity and ergodicity of the processes considered here. To begin with, we shall focus on the:

SETAR(2; 1, 1) model of (1.2) and (1.3). Here $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, $H_i(\mathbf{t}) \equiv \mathbf{t}'\mathbf{W}_i$, where $\mathbf{W}_i \equiv (X_{i-1}^+, X_{i-1}^-)'$. Hence, $\dot{\mathbf{H}}_i(\mathbf{t}) \equiv \mathbf{W}_i$, and (H1) and (H4)–(H6) are trivially satisfied. Moreover, with $R_{ir\mathbf{e}}$ denoting the rank of $\varepsilon_i - rn^{-1/2}\mathbf{e}'\mathbf{W}_i$ among $\{\varepsilon_j - rn^{-1/2}\mathbf{e}'\mathbf{W}_j, 1 \le j \le n\}$,

$$\mathbf{e}'\mathbf{M}(\boldsymbol{\theta} + n^{-1/2}r\mathbf{e}) \equiv n^{-1/2} \sum_i \mathbf{e}'\mathbf{W}_i \psi\left(\varepsilon_i - rn^{-1/2}\mathbf{e}'\mathbf{W}_i\right),$$

$$\mathbf{e}'\mathbf{S}(\boldsymbol{\theta} + n^{-1/2}r\mathbf{e}) \equiv n^{-1/2} \sum_i \mathbf{e}'\mathbf{W}_i \varphi(R_{ir\mathbf{e}}/(n+1)),$$

$$\int_0^1 \mathbf{e}'\mathbb{Z}(u, \boldsymbol{\theta} + n^{-1/2}r\mathbf{e}) g(u)\, dG(u) \equiv -n^{-1/2} \sum_i \mathbf{e}'\left(\mathbf{W}_i - \overline{\mathbf{W}}\right)\tau(R_{ir\mathbf{e}}/n),$$

where $\tau(u) \equiv \int_0^u g\, dG$. Consequently, ψ nondecreasing implies (M1), while (S1) and (K1) follow from φ being nondecreasing and Theorem 2.7E, of Hájek (1969).

From Tong (1990), page 130, and Chan, Petruccelli, Tong and Woolford (1985), it follows that if $E\varepsilon^2 < \infty$ and (F^+) hold, then the SETAR(2; 1, 1) process specified at (1.2) and (1.3) is stationary and ergodic, and $E_{\boldsymbol{\theta}} X_0^2 < \infty$.

Hence, by the ergodic theorem, (h2), (h3) and (h7) hold with

$$\mathbf{\Sigma}_{\boldsymbol\theta} \equiv \begin{bmatrix} E_{\boldsymbol\theta}(X_0^+)^2 & 0 \\ 0 & E_{\boldsymbol\theta}(X_0^-)^2 \end{bmatrix}, \qquad \mathbf{\Gamma}_{\boldsymbol\theta} \equiv \begin{bmatrix} \mathrm{Var}_{\boldsymbol\theta}(X_0^+) & -\mu_0^+\mu_0^- \\ -\mu_0^+\mu_0^- & \mathrm{Var}_{\boldsymbol\theta}(X_0^-) \end{bmatrix},$$

where $\mu_0^+ := E_{\boldsymbol\theta}(X_0^+)$, $\mu_0^- := E_{\boldsymbol\theta}(X_0^-)$. We summarize the above discussion in the following corollary.

COROLLARY 2.1. *Assume that in the SETAR*(2; 1, 1) *model* (1.2) *and* (1.3), (F^+) *holds and that* $E\varepsilon^2 < \infty$. *Then the following hold*:

$$\text{(2.1)} \qquad \begin{aligned} n^{1/2}\big(\hat{\boldsymbol\theta}_M - \boldsymbol\theta\big) &\Rightarrow N_2\big(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol\theta}^{-1}\nu(\psi, F)\big) \quad \textit{for every } \psi \in \Psi, \\ n^{1/2}\big(\hat{\boldsymbol\theta}_R - \boldsymbol\theta\big) &\Rightarrow N_2\big(\mathbf{0}, \mathbf{\Gamma}_{\boldsymbol\theta}^{-1}\mu(\varphi, F)\big) \quad \textit{for every } \varphi \in \Phi, \\ n^{1/2}\big(\hat{\boldsymbol\theta}_{\mathrm{md}} - \boldsymbol\theta\big) &\Rightarrow N_2\big(0, \mathbf{\Gamma}_{\boldsymbol\theta}^{-1}\rho(G, F)\big) \quad \textit{for every } G \in \Phi. \end{aligned}$$

Since ν, μ and ρ functionals also appear in LRAR models, the above statement about the efficiency comparisons also holds for the above SETAR(2; 1, 1) model.

Next, consider the problem of *testing the goodness-of-fit hypothesis* H_0: $F \equiv F_0$ against the alternatives H_1: $F \neq F_0$, where F_0 is a known d.f. on $\mathbb{R}$, assumed to satisfy (F_0^+). The Kolmogorov test would reject H_0 for large values of $D_n := n^{1/2}\|\hat{F}_n - F_0\|_\infty$, where now $\hat{F}_n(\cdot) \equiv F_n(\cdot, \hat{\boldsymbol\theta})$, with $\hat{\boldsymbol\theta}$ an estimator of $\boldsymbol\theta$ satisfying (1.28). Then from the above discussion and (1.29) we readily obtain that

$$D_n = \big\| W + n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta)'\boldsymbol\mu f_0 \big\|_\infty + o_p(1),$$

where $W(\cdot) \equiv n^{1/2}[F_n(\cdot, \boldsymbol\theta) - F_0(\cdot)]$—the empirical of the i.i.d. r.v.'s—and $\boldsymbol\mu := (\mu_0^+, \mu_0^-)'$.

On the other hand, recall, say, from Koul (1992), Section 7.2, that in AR(p) models the analogous test satisfies $D_n = \|W\|_\infty + o_p(1)$, *provided* the mean of F_0 is 0. Thus, unlike in the AR models, even if F_0 has zero mean, the D_n-test is not asymptotically distribution free (A.D.F.) in general in the SETAR(2; 1, 1) model.

Next, let F_1, F_2 be two d.f.'s with $F_1 \neq F_2$. Consider the *change point* problem where one wishes to test H_0^*: $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are i.i.d. F, F not necessarily known, against the alternative H_1: $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_j$ are i.i.d. F_1 and $\varepsilon_{j+1}, \varepsilon_{j+2}, \dots, \varepsilon_n$ are i.i.d. F_2, for some $1 \le j < n$. To describe a test procedure for this problem, let $\hat{\boldsymbol\theta}$ be an estimator of $\boldsymbol\theta$ based on $\{X_i;\ 0 \le i \le n\}$ and satisfying (1.28). Also, let $\hat{F}_{nu}, \hat{F}_{n1-u}$ denote the residual empiricals based on the first $[nu]$ residuals $X_i - \hat{\boldsymbol\theta}'\mathbf{W}_i$, $1 \le i \le [nu]$, and the last $n - [nu]$ residuals $X_i - \hat{\boldsymbol\theta}'\mathbf{W}_i$, $[nu] + 1 \le i \le n$, respectively, $u \in [0, 1]$. The Kolmogorov–Smirnov type test of this hypothesis is based on $\mathscr{D}_n := \sup_{y,u}|\mathscr{D}_n(y, u)|$, where

$$\mathscr{D}_n(y, u) := ([nu]/n)\{1 - ([nu]/n)\}n^{1/2}\big\{\hat{F}_{nu}(y) - \hat{F}_{n1-u}(y)\big\}, \qquad y \in \mathbb{R},\ u \in [0, 1].$$

Now, for the sake of brevity, write $W(y,u)$ for $W(y,\boldsymbol{\theta},u)$ of (1.23). From (1.27), assuming that $E\varepsilon^2<\infty$ and the common F of H_0^* satisfies (F^+), we readily obtain that uniformly in y and u, under H_0^*,

$$\begin{aligned}\mathscr{D}_n(y,u) &= (1-u)\left[W(y,u)+n^{1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'n^{-1}\sum_{i=1}^{[nu]}\mathbf{W}_i f(y)\right]\\ &\quad -u\left[W(y,1)-W(y,u)+n^{1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'n^{-1}\sum_{i=1+[nu]}^{n}\mathbf{W}_i f(y)\right]\\ &\quad + o_p(1).\end{aligned}$$

From the ergodic theorem it follows that, under H_0^*,

$$\sup_u\left|n^{-1}\sum_{i=1}^{[nu]}\mathbf{W}_i-u\boldsymbol{\mu}\right|\to 0,\qquad \sup_u\left|n^{-1}\sum_{i=1+[nu]}^{n}\mathbf{W}_i-(1-u)\boldsymbol{\mu}\right|\to 0,\quad \text{a.s.}$$

Hence, one readily obtains that, under H_0^*,

$$\mathscr{D}_n=\sup_{y,u}|W(y,u)-uW(y,1)|+o_p(1).$$

Thus, it follows, from Bickel and Wichura (1971), that $\mathscr{D}_n\Rightarrow\sup\{|G(s,u)|, (s,u)\in[0,1]^2\}$, where G is a continuous Gaussian process on $[0,1]^2$ with $\mathrm{Cov}\{G(s,u),G(t,u)\}=(s\wedge t-st)\cdot(u\wedge v-uv)$. Consequently, the test based on $\mathscr{D}_n$ is A.D.F. for H_0^* against H_1. Clearly, a similar conclusion can be obtained for any other test of H_0^* based on $\{\mathscr{D}_n(y,u);\ y\in\mathbb{R},\ u\in[0,1]\}$, $n\ge 1$.

The results of the previous section are general enough to allow one to obtain similar conclusions in the SETAR(2; p,p) model where, for *known* (d,r), $1\le d\le p$, $r\in\mathbb{R}$,

$$H(\boldsymbol{\theta},\mathbf{Y}_{i-1})=\sum_{j=1}^{p}a_jX_{i-j}I(X_{i-d}\le r)+\sum_{j=1}^{p}b_jX_{i-j}I(X_{i-d}>r),$$

with $\boldsymbol{\theta}:=(a_1,a_2,\dots,a_p,b_1,b_2,\dots,b_p)'$. This is a version of the model in Tong (1990), (3.25), page 107. Under $\max\{\sum_{j=1}^{p}|a_j|,\sum_{j=1}^{p}|b_j|\}<1$, ($\mathrm{F}_0^+$) and $E|\varepsilon|<\infty$, this model is strictly stationary and geometrically ergodic.

Next, consider the **SETAR(2; 1, 1) + regression** model obtained from (1.1) upon choosing $m=3$, $p=1=q$ $\{Z_{ni}\}$ to be known constants and $h(\boldsymbol{\theta},y,z)=\theta_1y^++\theta_2y^-+\theta_3z$, $y,z\in\mathbb{R}$, $\boldsymbol{\theta}\in(0,1)^2\times\mathbb{R}$. The observations are no longer stationary now, but Theorems 1.1 to 1.3 can still be applied. Note that here $\dot{\mathbf{h}}_i(\boldsymbol{\theta})\equiv(\mathbf{W}_i',\mathbf{Z}_{ni})'$. Suppose

$$n^{-1}\sum_i\dot{\mathbf{h}}_i(\boldsymbol{\theta})\dot{\mathbf{h}}_i'(\boldsymbol{\theta})=n^{-1}\sum_i\begin{bmatrix}(X_{i-1}^+)^2 & 0 & X_{i-1}^+Z_{ni}\\ 0 & (X_{i-1}^-)^2 & X_{i-1}^-Z_{ni}\\ X_{i-1}^+Z_{ni} & X_{i-1}^-Z_{ni} & Z_{ni}^2\end{bmatrix}=\boldsymbol{\Sigma}_{\boldsymbol{\theta}}+o_p(1),$$

$$n^{-1/2}\max_i|X_{i-1}|=o_p(1)\quad\text{and}\quad n^{-1/2}\max_i|Z_{ni}|=o(1).$$

Then conditions (h1)–(h9) are satisfied as are (M1), (S1) and (K1) so that all the results of the previous section are applicable under (F). It is perhaps worth mentioning that even in this relatively simple model these results are new.

Next, consider the **EXPAR model of (1.2) and (1.4)**. Assume $E\varepsilon^2 < \infty$ and (F^+). Then it follows from Tong (1990), page 130, that this model is stationary and ergodic, and $E_{\boldsymbol\theta} X_0^2 < \infty$. Because $x^k \exp(-\alpha x^2)$ is a smooth function of α with all derivatives bounded in x, for all $k \geq 0$, $(\mathrm{H}_s 1)$–$(\mathrm{H}_s 3)$, (H6) and (H9) are easily verified here with

$$\dot{\mathbf{H}}_i(\boldsymbol\theta) \equiv \begin{bmatrix} X_{i-1} \\ X_{i-1}\exp\bigl(-\theta_3 X_{i-1}^2\bigr) \\ -\theta_2 X_{i-1}^3 \exp\bigl(-\theta_3 X_{i-1}^2\bigr) \end{bmatrix},$$

$$\operatorname{plim} n^{-1} \sum_i \dot{\mathbf{H}}_i(\boldsymbol\theta) = E_{\boldsymbol\theta} \dot{\mathbf{H}}_1(\boldsymbol\theta) = \boldsymbol\nu, \quad \text{say}.$$

Moreover, by stationarity, the left-hand side of (H8) is bounded above by

$$\begin{aligned} &8\Bigl\{ s_1^2 E_{\boldsymbol\theta} X_0^2 + E_{\boldsymbol\theta} X_0^2 \exp\bigl(-\theta_3 X_0^2\bigr)\Bigl[s_2^2 \exp\bigl(-2 s_3 n^{-1/2} X_0^2\bigr) \\ &\qquad\qquad + \theta_2^2 n \bigl\{1 - \exp\bigl(-s_3 n^{-1/2} X_0^2\bigr)\bigr\}^2 \Bigr]\Bigr\} \\ &\to 8\bigl\{ s_1^2 E_{\boldsymbol\theta} X_0^2 + E_{\boldsymbol\theta} X_0^2 \exp\bigl(-\theta_3 X_0^2\bigr)\bigl[s_2^2 + \theta_2^2 s_3 X_0^4 \bigr]\bigr\} = O(1). \end{aligned}$$

Thus, all asymptotic uniform linearity results of the previous section are valid.

Consider, for example, the *goodness-of-fit* problem of testing H_0: $F \equiv F_0$ versus the alternatives H_1: $F \neq F_0$, where F_0 is as above. Then, with D_n, $\hat{\boldsymbol\theta}$ and f_0 having the same meaning as in the SETAR example, from (1.29) we readily obtain that

$$D_n = \bigl\| W + n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta)' \boldsymbol\nu f_0 \bigr\|_\infty + o_p(1).$$

Now, if additionally, F_0 is such that the stationary distribution is symmetric around 0, then $\boldsymbol\nu = \mathbf{0}$ and the D_n-test would be A.D.F., but not in general.

Similarly, because of the fact

$$\sup_u \left\| n^{-1} \sum_{i=1}^{[nu]} \bigl(\dot{\mathbf{H}}_i(\boldsymbol\theta) - E_{\boldsymbol\theta} \dot{\mathbf{H}}_1(\theta) \bigr) \right\| \to 0 \quad \text{a.s.},$$

the conclusion obtained about the change point problem in the SETAR example is also valid here with obvious modifications.

Now, consider the *estimation* of $\boldsymbol\theta$ when θ_3 is *known*. Then, using arguments similar to those in the above SETAR model, one can verify that (M1), (S1) and (K1) are also satisfied. Thus, here the asymptotic normality results about M-, R- and M.D.-estimators of θ_1 and θ_2 are valid. For example, we have the following result.

COROLLARY 2.2. *Suppose that the EXPAR model* (1.2) *and* (1.4) *with a known* θ_3 *holds. In addition, suppose that* $E\varepsilon^2 < \infty$ *and* (F^+) *holds. Then the asymptotic normality conclusion of* (2.1) *hold here also with* $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\dot{\mathbf{H}}_{11}(\boldsymbol{\theta})\dot{\mathbf{H}}'_{11}(\boldsymbol{\theta})$, $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \boldsymbol{\nu}_1\boldsymbol{\nu}'_1$, *where*

$$\dot{\mathbf{H}}_{11}(\boldsymbol{\theta}) := \begin{bmatrix} X_0 & X_0^2 \exp(-\theta_3 X_0^2) \\ X_0^2 \exp(-\theta_3 X_0^2) & X_0^2 \exp(-2\theta_3 X_0^2) \end{bmatrix},$$

$$\boldsymbol{\nu}_1 := E_{\boldsymbol{\theta}}\begin{bmatrix} X_0 \\ X_0 \exp(-\theta_3 X_0^2) \end{bmatrix}.$$

Conditions (M1), (S1) and (K1) are not satisfied when estimating θ_3. The difficulty in estimating θ_3 is similar to that of estimating a scale parameter. However, one can prove a Cramér type result for *M*-estimators when ψ is smooth. See Tjøstheim (1986) for the case $\psi(x) \equiv x$.

3. Proofs. An important tool needed for the proofs of (1.6), (1.7) and (1.12) is a general result obtained in Koul and Ossiander (1994). For the sake of completeness, we restate it here. Let $(\Omega, \mathscr{A}, P)$ be a probability space. For each integer $n \geq 1$, let $(\eta_{ni}, \xi_{ni}, \gamma_{ni})$, $1 \leq i \leq n$, be an array of trivariate r.v.'s defined on $(\Omega, \mathscr{A})$ such that $\{\eta_{ni}, 1 \leq i \leq n\}$ are i.i.d. r.v.'s with d.f. F and η_{ni} is independent of (γ_{ni}, ξ_{ni}), $1 \leq i \leq n$. Furthermore, let $\{\mathscr{A}_{ni}\}$ be an array of sub σ-fields such that $\mathscr{A}_{ni} \subset \mathscr{A}_{ni+1}$, $1 \leq i \leq n$, $n \geq 1$; (γ_{n1}, ξ_{n1}) is $\mathscr{A}_{n1}$-measurable; the r.v.'s $\{\eta_{n1}, \dots, \eta_{nj-1}; (\gamma_{ni}, \xi_{ni}), 1 \leq i \leq j\}$ are $\mathscr{A}_{nj}$-measurable, $2 \leq j \leq n$; and η_{nj} is independent of $\mathscr{A}_{nj}$, $1 \leq j \leq n$. Define, for an $x \in \mathbb{R}$,

$$
\begin{aligned}
V_n(x) &:= n^{-1} \sum_{i=1}^n \gamma_{ni} I(\eta_{ni} \leq x + \xi_{ni}), \\
V_n^*(x) &:= n^{-1} \sum_{i=1}^n \gamma_{ni} I(\eta_{ni} \leq x), \\
J_n(x) &:= n^{-1} \sum_{i=1}^n E\{\gamma_{ni} I(\eta_{ni} \leq x + \xi_{ni}) | \mathscr{A}_{ni}\} \\
&= n^{-1} \sum_{i=1}^n \gamma_{ni} F(x + \xi_{ni}), \\
J_n^*(x) &:= n^{-1} \sum_{i=1}^n \gamma_{ni} F(x), \\
U_n(x) &:= n^{1/2}(V_n(x) - J_n(x)), \\
U_n^*(x) &:= n^{1/2}(V_n^*(x) - J_n^*(x)).
\end{aligned} \tag{3.1}
$$

The following lemma is obtained from Theorem 1.1 and Corollary 2.3 of Koul and Ossiander (1994).

LEMMA 3.1. *In addition to the above, assume that the following hold*:

(A1) $$\left(n^{-1}\sum_{i=1}^{n}\gamma_{ni}^2\right)^{1/2} = \gamma + o_p(1), \qquad \gamma \text{ a positive r.v.}$$

(A2) $$n^{-1/2}\max_{1\le i\le n}|\gamma_{ni}| = o_p(1).$$

(A3) $$\max_{1\le i\le n}|\xi_{ni}| = o_p(1).$$

Then, for every y at which F is continuous,

$$|U_n(y) - U_n^*(y)| = o_p(1). \tag{3.2}$$

If, in addition, (F) *holds, then the processes* $\{U_n\}$ *and* $\{U_n^*\}$ *are eventually tight in the uniform metric and*

$$\|U_n - U_n^*\|_\infty = o_p(1). \tag{3.3}$$

Now recall the notation from (1.5) and let

$$\mathbb{W}^*(y,\mathbf{t}) := n^{-1/2}\sum_i \dot{\mathbf{h}}_{ni}(\mathbf{t})[I(\varepsilon_i \le y) - F(y)], \qquad y\in\mathbb{R},\ \mathbf{t}\in\mathbb{R}^m. \tag{3.4}$$

Let $\dot{h}_{nij}$, V_j, $\mathbb{W}_j$ and so on denote the jth component of $\dot{\mathbf{h}}_{ni}$, $\mathbf{V}$, $\mathbb{W}$ and so on, $1\le i\le n$, $1\le j\le m$. Note that if in (3.1) we take

$$\gamma_{ni} \equiv \dot{h}_{nij}(\mathbf{s}), \qquad \eta_{ni}\equiv\varepsilon_i, \qquad \xi_{ni}\equiv d_{ni}(\mathbf{s}), \qquad \mathscr{A}_{ni}\equiv\mathscr{F}_{ni}, \tag{3.5}$$

then under (1.1) and (h1), $U_n(y)$ and $U_n^*(y)$ are, respectively, equal to $\mathbb{W}_j(y,\mathbf{s})$ and $\mathbb{W}_j^*(y,\mathbf{0})$, for all $y\in\mathbb{R}$ and $s\in\mathbb{R}^m$, $1\le j\le m$. We now state and prove the following result.

LEMMA 3.2. *Assume that* (1.1), (F) *and* (h1)–(h6) *hold. Then*, $\forall\ 0<b<\infty$,

$$\sup_{y\in\mathbb{R},\,\mathbf{t}\in N_b}\|\mathbb{W}(y,\mathbf{t}) - \mathbb{W}^*(y,\mathbf{t})\| = o_p(1). \tag{3.6}$$

PROOF. Fix a $b\in(0,\infty)$. In this proof the indices $i,\mathbf{t}$ in $\sup_{i,\mathbf{t}}$ and y in sup vary over $1\le i\le n$, $\mathbf{t}\in N_b$, $y\in\mathbb{R}$. Observe that by (h1), $\forall\ \alpha>0$, $\exists\ n_1$, $\ni\ \forall\ n>n_1$,

$$P_{\boldsymbol{\theta}}^n\left(\sup_{i,\mathbf{t}}\left|h_i(\boldsymbol{\theta}+n^{-1/2}\mathbf{t}) - h_i(\boldsymbol{\theta}) - n^{-1/2}\mathbf{t}'\dot{\mathbf{h}}_i(\boldsymbol{\theta})\right| \le b\alpha n^{-1/2}\right) \ge 1-\alpha. \tag{3.7}$$

Hence, from (h3), we readily obtain

$$\sup_{i,\mathbf{t}}|d_{ni}(\mathbf{t})| = o_p(1). \tag{3.8}$$

This verifies (A3) for the ξ_{ni} of (3.5).

Next, by (h2) and (h4), we readily obtain that $(n^{-1}\sum_i\dot{h}_{nij}^2(\mathbf{s}))^{1/2} = \sigma_{jj}(\boldsymbol{\theta}) + o_p(1)$, $\mathbf{s}\in\Omega$, $1\le j\le m$, where $\sigma_{jj}(\boldsymbol{\theta})$ is the jth diagonal term of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, so that (A1) is verified for the γ_{ni} of (3.5) for every $\mathbf{s}\in\Omega$, $1\le j\le m$. Finally, because for all $\mathbf{s} \in \Omega$, $1 \le j \le m$,

$$n^{-1/2} \max_i \left| \dot{h}_{nij}(\mathbf{s}) \right| \le n^{-1/2} \max_i \left| \dot{h}_{nij}(\mathbf{s}) - \dot{h}_{ij}(\boldsymbol{\theta}) \right| + n^{-1/2} \max_i \left| \dot{h}_{ij}(\boldsymbol{\theta}) \right|,$$

$$n^{-1/2} \max_i \left| \dot{h}_{nij}(\mathbf{s}) - \dot{h}_{ij}(\boldsymbol{\theta}) \right| \le \left(n^{-1} \sum_i \left\{ \dot{h}_{nij}(\mathbf{s}) - \dot{h}_{ij}(\boldsymbol{\theta}) \right\}^2 \right)^{1/2},$$

(A2) is implied by (h3) and (h4) for the γ_{ni} of (3.5), for every $\mathbf{s} \in \Omega$, $1 \le j \le m$. Hence (3.3) readily enables one to conclude that

$$\sup_y \| \mathbb{W}(y, \mathbf{s}) - \mathbb{W}^*(y, \mathbf{s}) \| = o_p(1), \qquad \mathbf{s} \in \Omega. \tag{3.9}$$

To complete the proof of (3.6), because of the compactness of N_b, it suffices to show that $\forall\ \alpha > 0$, $\exists\ \delta > 0$ and $n_0 < \infty$, $\ni\ \forall\ \mathbf{s} \in N_b$,

$$P_{\boldsymbol{\theta}}^n \left(\sup_{y \in \mathbb{R},\ \|\mathbf{t} - \mathbf{s}\| \le \delta} \| \mathbf{D}(y, \mathbf{t}) - \mathbf{D}(y, \mathbf{s}) \| > \alpha \right) \le \alpha, \qquad n > n_0, \tag{3.10}$$

where $\mathbf{D}(y, \mathbf{t}) \equiv \mathbb{W}(y, \mathbf{t}) - \mathbb{W}^*(y, \mathbf{t})$.

For the sake of brevity, let $\alpha_i(y, \mathbf{t}) := I(\varepsilon_i \le y + d_{ni}(\mathbf{t})) - F(y + d_{ni}(\mathbf{t}))$, $y \in \mathbb{R}$, $\mathbf{t} \in \Omega$, $1 \le i \le n$, and write $\alpha_i(y)$ for $\alpha_i(y, \mathbf{0})$. Then

$$\mathbb{W}(y, \mathbf{t}) \equiv n^{-1/2} \sum_i \dot{\mathbf{h}}_{ni}(\mathbf{t}) \alpha_i(y, \mathbf{t}), \qquad \mathbb{W}^*(y, \mathbf{t}) \equiv n^{-1/2} \sum_i \dot{\mathbf{h}}_{ni}(\mathbf{t}) \alpha_i(y)$$

and

$$\begin{aligned} \mathbf{D}(y, \mathbf{t}) - \mathbf{D}(y, \mathbf{s}) &= n^{-1/2} \sum_i \left[\dot{\mathbf{h}}_{ni}(\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\mathbf{s}) \right] [\alpha_i(y, \mathbf{t}) - \alpha_i(y)] \\ &\quad + n^{-1/2} \sum_i \dot{\mathbf{h}}_{ni}(\mathbf{s}) [\alpha_i(y, \mathbf{t}) - \alpha_i(y, \mathbf{s})] \\ &= \mathbf{D}_1(y, \mathbf{s}, \mathbf{t}) + \mathbf{D}_2(y, \mathbf{s}, \mathbf{t}), \quad \text{say}, \qquad y \in \mathbb{R}, \mathbf{s}, \mathbf{t} \in \Omega. \end{aligned}$$

To prove (3.10), it thus suffices to prove its analogue for $\mathbf{D}_1$ and $\mathbf{D}_2$. But (h6) obviously implies this for $\mathbf{D}_1$ because $|\alpha_i(y, \mathbf{t}) - \alpha_i(y)| \le 1$ for all i, y and $\mathbf{t}$.

We proceed to prove an analogue of (3.10) for $\mathbf{D}_2$. Let D_{2j} denote the jth component of $\mathbf{D}_2$. Write $\dot{h}_{nij}(\mathbf{s}) = \dot{h}^+_{nij}(\mathbf{s}) - \dot{h}^-_{nij}(\mathbf{s})$ and $D_{2j} \equiv D^+_{2j} - D^-_{2j}$, where $D^\pm_{2j}$ correspond to the D_{2j} with $\{\dot{h}_{nij}(\mathbf{s})\}$ replaced by $\{\dot{h}^\pm_{nij}(\mathbf{s})\}$. Thus, by the triangle inequality, it suffices to prove an analogue of (3.10) for $D^\pm_{2j}$, $1 \le j \le m$.

Now fix an $\alpha > 0$, $\mathbf{s} \in N_b$ and $\delta > 0$. Let $\Delta_{ni} := n^{-1/2}(\delta \|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\| + 2b\alpha)$ and

$$A_n := \left\{ \sup_{\mathbf{t} \in N_b,\ \|\mathbf{t} - \mathbf{s}\| \le \delta} |d_{ni}(\mathbf{t}) - d_{ni}(\mathbf{s})| \le \Delta_{ni}, 1 \le i \le n \right\}.$$

From (3.7), it follows that

$$P_{\boldsymbol{\theta}}^n(A_n) \ge 1 - \alpha, \qquad n > n_1. \tag{3.11}$$

Next, define, for $y, a \in \mathbb{R}$, $1 \le j \le m$,

$$\mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, \mathbf{a}) := n^{-1/2} \sum_i \dot{h}_{nij}^{\pm}(\mathbf{s}) \big[I(\varepsilon_i \le y + d_{ni}(\mathbf{s}) + a\Delta_{ni}) - F(y + d_{ni}(\mathbf{s}) + a\Delta_{ni}) \big].$$

By definition, $d_{ni}(\mathbf{s}) + a\Delta_{ni}$ is $\mathscr{F}_{ni}$-measurable, $a \in \mathbb{R}$, $1 \le i \le n$. Moreover, by (h3) and (3.8),

$$\begin{aligned} &\max_i |d_{ni}(\mathbf{s}) + a\Delta_{ni}| \\ &\le \max_i |d_{ni}(\mathbf{s})| + \max_i n^{-1/2}\big(\delta\|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\| + 2b\alpha\big) \qquad (3.12) \\ &= o_p(1). \end{aligned}$$

The rest of the argument being the same as for (3.9), one more application of (3.3) with $\xi_{ni} \equiv d_{ni}(\mathbf{s}) + a\Delta_{ni}$, and the other entities as in (3.5), yields that

$$\sup_y \big| \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, a) - \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, 0) \big| = o_p(1), \qquad a \in \mathbb{R}, 1 \le j \le m. \qquad (3.13)$$

Now, using the nonnegativity of $\{\dot{h}_{nij}^{\pm}(\mathbf{s})\}$ and the monotonicity of the indicator function and the d.f. F, we obtain, that on A_n, $\forall\, \mathbf{t} \in N_b$, $\|\mathbf{t} - \mathbf{s}\| \le \delta$,

$$\begin{aligned} &\big| D_{2j}^{\pm}(y, \mathbf{s}, \mathbf{t}) \big| \\ &\quad \le \big| \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, 1) - \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, 0) \big| + \big| \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, -1) - \mathscr{D}_{2j}^{\pm}(y, \mathbf{s}, 0) \big| \\ &\qquad + \Big| n^{-1/2} \sum_i \dot{h}_{nij}^{\pm}(\mathbf{s}) \big[F(y + d_{ni}(\mathbf{s}) + \Delta_{ni}) - F(y + d_{ni}(\mathbf{s}) - \Delta_{ni}) \big] \Big|. \end{aligned}$$

By (F) and the fact that $x^+ \le |x|$, $x \in \mathbb{R}$, the last term in this upper bound is no larger than

$$2\|f\|_\infty n^{-1} \sum_i \big| \dot{h}_{nij}(\mathbf{s}) \big| \big(\delta \|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\| + 2b\alpha \big),$$

which, in view of (h2) and (h4), can be made smaller than α with arbitrarily large probability for sufficiently large n by the choice of δ. This together with (3.11) and (3.13) completes the proof of an analogue of (3.10) for $\mathbf{D}_2$, and hence of (3.6). □

PROOF OF LEMMA 1.1. Rewrite $\mathbb{W}(y, \mathbf{t}) - \mathbb{W}(y, \mathbf{0}) \equiv \mathbb{W}(y, \mathbf{t}) - \mathbb{W}^*(y, \mathbf{t}) + \mathbf{U}(y, \mathbf{t})$, where $\mathbf{U}(y, \mathbf{t}) := \mathbb{W}^*(y, \mathbf{t}) - \mathbb{W}(y, \mathbf{0}) \equiv n^{-1/2} \sum_i [\dot{\mathbf{h}}_{ni}(\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\mathbf{0})] \alpha_i(y)$. Thus, in view of (3.6), it suffices to prove

$$\sup_{y, \mathbf{t}} \| \mathbf{U}(y, \mathbf{t}) \| = o_p(1). \qquad (3.14)$$

Fix a $1 \le j \le m$ and a $t \in N_b$. Let $\gamma_{ni} \equiv \dot{h}_{nij}(\mathbf{t}) - \dot{h}_{nij}(0)$. Write $\gamma_{ni} = \gamma_{ni}^+ - \gamma_{ni}^-$ so that the jth component of $\mathbf{U}$ is rewritten as $U_j \equiv U_j^+ - U_j^-$, where

$$U_j^{\pm}(y, \mathbf{t}) = n^{-1/2} \sum_i \gamma_{ni}^{\pm} \alpha_i(y).$$

Because $\gamma_{ni}^{\pm}$ is $\mathscr{F}_{ni}$-measurable, $1 \le i \le n$, from (h4) we obtain

$$\begin{aligned}\operatorname{Var}\Big(U_j^{\pm}(y,\mathbf{t})\Big) &= n^{-1}\sum_i E\big\{\gamma_{ni}^{\pm}\big\}^2 F(y)(1-F(y)) \\ &\le n^{-1}\sum_i E\big\{\gamma_{ni}^2\big\} = o(1), \qquad y \in \mathbb{R}.\end{aligned} \tag{3.15}$$

Next, fix an $\alpha > 0$, and let $-\infty = y_0 < y_1 < \cdots < y_r = \infty$ be a partition of $\mathbb{R}$ such that $[F(y_i) - F(y_{i-1})] \le \alpha$, $i = 0, 1, \dots, r$. Then, once again using the monotonicity of the indicator and F, we obtain

$$\sup_y \Big|U_j^{\pm}(y,\mathbf{t})\Big| \le 2 \max_{0 \le k \le r} \Big|U_j^{\pm}(y_k,\mathbf{t})\Big| + \alpha n^{-1/2} \sum_i |\gamma_{ni}|.$$

This, (3.15), (h5) and the arbitrariness of α enables one to conclude that

$$\sup_y \|\mathbf{U}(y,\mathbf{t})\| = o_p(1), \qquad \mathbf{t} \in N_b.$$

To obtain the uniformity with respect to $\mathbf{t}$, we need to show that an analogue of (3.10) holds for $\mathbf{U}(y,\mathbf{t})$-processes. But this is implied by (h6), because

$$\mathbf{U}(y,\mathbf{t}) - \mathbf{U}(y,\mathbf{s}) \equiv n^{-1/2} \sum_i \Big[\dot{\mathbf{h}}_{ni}(\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\mathbf{s})\Big]\alpha_i(y).$$

This completes the proof of (3.14), and hence that of (1.7) of Lemma 1.1. The proof of (1.6) uses (3.2) and is similar to the above proof. □

Our proof of (1.12) uses (1.7) and the tightness of the residual empirical process in the uniform metric which follows from Theorem 1.3(a). Therefore, we shall first prove Theorem 1.3. In the sequel all probability statements are understood to be under the joint distribution of $\{\mathbf{Y}_0, X_i, g_{ni}, 1 \le i \le n\}$ when $\boldsymbol{\theta}$ is the true parameter.

PROOF OF THEOREM 1.3(a). The proof of Theorem 1.3(a) follows from Lemma 3.1 applied to $\gamma_{ni} \equiv g_{ni}$, and the other entities as in (3.5), and by an argument similar to the one used in the proof of Lemma 3.2. □

Our proof of Theorem 1.3(b) is facilitated by the following two lemmas. Recall the definitions of $\{\Delta_{ni}\}$ from (3.11) and let $u_{ni} \equiv d_{ni}(\mathbf{s}) + a\Delta_{ni}$, $a \in \mathbb{R}$, $\|\mathbf{s}\| \le b$.

LEMMA 3.3. *Under* (F), (h1)–(h3) *and* (g1), *for some* $K < \infty$ *and for all* $\alpha > 0$,

$$\limsup_n P\bigg(\sup_{x,y} n^{-1/2} \sum_i \big| g_{ni}[F(y+u_{ni}) - F(x+u_{ni})]\big| \le K\alpha\bigg) = 1,$$

$$a \in \mathbb{R}, \|\mathbf{s}\| \le b,$$

where the supremum is taken over the set $\{x, y \in \mathbb{R}; |F(y) - F(x)| \le n^{-1/2}\alpha\}$.

PROOF. Let $u_n = \max_i u_{ni}$, $\tau_n := \sup\{|f(y) - f(x)|;\ |F(y) - F(x)| \le n^{-1/2}\alpha\}$, $\omega_n := \sup\{|f(z) - f(v)|; |z - v| \le u_n\}$. From (F) and (3.12), $\tau_n = o(1)$,

$\omega_n = o_p(1)$. Also, from (h2)–(h4) and (g1), we obtain that $n^{-1/2}\sum_i |g_{ni}u_{ni}| = O_p(1)$. Lemma 3.3 follows from these observations and the inequality

$$n^{-1/2} \sum_i |g_{ni}[F(y+u_{ni}) - F(x+u_{ni})]|$$
$$\leq n^{-1/2} \sum_i |g_{ni}| |F(y) - F(x)| + n^{-1/2} \sum_i |g_{ni}u_{ni}| \{\tau_n + 2\omega_n\}. \quad \square$$

LEMMA 3.4. *Let F be a continuous strictly increasing d.f., $\{\varepsilon_i\}$ be i.i.d. F, $\alpha > 0$, $n \geq 1$, $N := [N^{1/2}\alpha^{-1}]$ and $\{y_j\}$ be the partition of $\mathbb{R}$ such that $F(y_j) = j/N$, $1 \leq j \leq N$, $y_0 = -\infty$, $y_{N+1} = \infty$. Then, under (g4),*

$$\sup_{u,j} \left| n^{-1/2} \sum_{i=1}^{[nu]} g_{ni}^{\pm}\big[I(\varepsilon_i \leq y_{j+1}) - I(\varepsilon_i \leq y_j) - (1/N)\big]\right| = o_p(1), \tag{3.16}$$

where the supremum is taken over $0 \leq u \leq 1$, $0 \leq j \leq N+1$.

PROOF. We shall give the proof of (3.16) only for the case of g_{ni}^{+}, the other case being exactly similar. Let $\mathscr{F}_i := \sigma$-field$\{\varepsilon_j;\, 1 \leq j \leq i\}$. Fix a $0 \leq j \leq N+1$, and let

$$V_i := g_{ni}^{+}\big[I(\varepsilon_i \leq y_{j+1}) - I(\varepsilon_i \leq y_j) - (1/N)\big], \quad S_{ni} := \sum_{k=1}^{i} V_k, \quad 1 \leq i \leq n.$$

Note that $\{\mathscr{S}_{ni}, \mathscr{F}_{i-1};\, 1 \leq i \leq n\}$ is a mean-zero martingale. By Doob's and Rosenthal's inequalities [Hall and Heyde (1980), pages 15 and 23],

$$P\Big(\max_i |\mathscr{S}_{ni}| > \alpha\Big) \leq \alpha^{-4} E\{\mathscr{S}_{nn}\}^4,$$
$$E\{\mathscr{S}_{nn}\}^4 \leq C\left\{E\left[\sum_i E\big(V_i^2 | \mathscr{F}_{i-1}\big)\right]^2 + \sum_i EV_i^4\right\}, \quad \text{for some } C < \infty.$$

But, because $g_{ni}^{+} \leq |g_{ni}|$, for all i,

$$\sum_i EV_i^4 \leq \sum_i Eg_{ni}^4, \qquad E\big(V_i^2 | \mathscr{F}_{i-1}\big) \leq g_{ni}^2 (1/N) \leq n^{-1/2} g_{ni}^2 \alpha/(1-\alpha),$$
$$1 \leq i \leq n,$$
$$E\left[\sum_i E\big(V_i^2 | \mathscr{F}_{i-1}\big)\right]^2 \leq \{\alpha/(1-\alpha)\}^2 \sum_i Eg_{ni}^4.$$

Observe that these bounds do not depend on j. Therefore,

$$P(\text{l.h.s. of } (3.16) > \alpha)$$
$$\leq N \max_j P\Big(\max_i |\mathscr{S}_{ni}| > \alpha n^{1/2}\Big) \leq C(\alpha) n^{-2} N \sum_i Eg_{ni}^4 = O(n^{-1/2}),$$

where $C(\alpha)$ is a constant depending on C and α, and where the last equality follows from (g4) and the definition of N. This proves (3.16). $\square$

PROOF OF THEOREM 1.3(b). Since (1.23) follows from (F) and (1.22), it suffices to give the proof of (1.22).

PROOF OF (1.22). Write $g_{ni} \equiv g_{ni}^{+} - g_{ni}^{-}$, and $V^{+}(y, \mathbf{t}, u)$ and $V^{-}(y, \mathbf{t}, u)$ for the difference inside the absolute value in the l.h.s. of (1.22) with g_{ni} replaced by g_{ni}^{+} and g_{ni}^{-}, respectively. By the triangle inequality, it suffices to prove (1.22) for V^{+} and V^{-}. Since both $g_{ni}^{\pm}$ and $d_{ni}(\mathbf{t})$ are $\mathscr{F}_{ni}$-measurable, by (F), (g4), (3.8) and the dominated convergence theorem, we obtain that, $\forall\,(y, \mathbf{t}, u) \in \mathbb{R} \times N_b \times [0, 1]$,

$$\text{(3.17)}\quad \operatorname{Var}\{V^{\pm}(y, \mathbf{t}, u)\} \le n^{-1} \sum_{i=1}^{n} E g_{ni}^{2} \left| F(y + d_{ni}(\mathbf{t})) - F(y) \right| = o(1).$$

Now, fix an $\alpha > 0$, an $\mathbf{s} \in N_b$ and $\delta > 0$. Let A_n be as in (3.11). Define, for $a \in \mathbb{R}$, $(y, \mathbf{t}, u) \in \mathbb{R} \times N_b \times [0, 1]$,

$$\begin{aligned} U^{\pm}(a, y, \mathbf{t}, u) &:= n^{-1/2} \sum_{i=1}^{[nu]} g_{ni}^{\pm} \big[I(\varepsilon_i \le y + d_{ni}(\mathbf{t}) + a\Delta_{ni}) - I(\varepsilon_i \le y) \\ &\qquad\qquad - F(y + d_{ni}(\mathbf{t}) + a\Delta_{ni}) + F(y) \big]. \end{aligned}$$

Arguing as in the proof of Lemma 3.2, we obtain that on the set A_n, $\forall\, \mathbf{t} \in N_b$, $\|\mathbf{t} - \mathbf{s}\| \le \delta$,

$$\begin{aligned} &\left| V^{\pm}(y, \mathbf{t}, u) \right| \\ &\quad \le \sup_{y,u} \left| U^{\pm}(1, y, \mathbf{s}, u) \right| + \sup_{y,u} \left| U^{\pm}(-1, y, \mathbf{s}, u) \right| \\ &\qquad + \sup_{y,u} n^{-1/2} \sum_{i=1}^{[nu]} g_{ni}^{\pm} \big[F(y + d_{ni}(\mathbf{s}) + \Delta_{ni}) - F(y + d_{ni}(\mathbf{s}) - \Delta_{ni}) \big], \\ &\qquad\qquad\qquad\qquad\qquad\qquad y \in \mathbb{R},\ u \in [0, 1]. \end{aligned}$$

But by (F), the last term in the above bound is at the most equal to

$$2\|f\|_{\infty} \left(\delta n^{-1} \sum_{i} \left\| g_{ni} \dot{\mathbf{h}}_i(\boldsymbol{\theta}) \right\| + 2 n^{-1} \sum_{i} |g_{ni}| b\alpha \right),$$

which, in view of (g4) and (g5), can be made arbitrarily small by the choice of δ, with probability tending to 1 as $n \to \infty$. Thus, to finish the proof of (1.22), it suffices, in view of the compactness of N_b, to show that

$$\text{(3.18)}\qquad \sup_{y,u} \left| U^{\pm}(a, y, \mathbf{s}, u) \right| = o_p(1), \qquad a \in \mathbb{R},\ \mathbf{s} \in N_b.$$

Now let N and $\{y_j\}$ be as in Lemma 3.4. Then, once again using the monotonicity of the indicator function and F, we obtain

$$\begin{aligned}
&\sup_{y,u}\left|U^{\pm}(a,y,\mathbf{s},u)\right| \\
&\quad\le \sup_{j,u}\left|U^{\pm}(a,y_j,\mathbf{s},u)\right| \\
&\qquad + \sup_j n^{-1/2}\sum_{i=1}^{n}|g_{ni}|\left|F(y_{j+1}+u_{ni})-F(y_j+u_{ni})\right| \\
&\qquad + \sup_j\left[F(y_{j+1})-F(y_j)\right]n^{-1/2}\sum_i|g_{ni}| \\
&\qquad + \sup_{j,u}\left|n^{-1/2}\sum_{i=1}^{[nu]}g_{ni}^{\pm}\left[I(\varepsilon_i\le y_{j+1})-I(\varepsilon_i\le y_j)-(1/N)\right]\right|.
\end{aligned}$$

The last term in this upper bound is $o_p(1)$, by Lemma 3.4. The third term of this upper bound is at the most $\alpha/(1-\alpha)n^{-1}\sum_i|g_{ni}|$, for all n, while, by Lemma 3.3, the second term is at the most $K\alpha$, for all sufficiently large n, with probability tending to 1.

Now consider the first term. Let

$$\zeta_{ni} := g_{ni}^{\pm}\left[I(\varepsilon_i\le y_j+u_{ni})-I(\varepsilon_i\le y_j)-F(y_j+u_{ni})+F(y_j)\right], \qquad 1\le i\le n,$$

$$S_{nk} := \sum_{i=1}^{k}\zeta_{ni}, \qquad 1\le k\le n.$$

Then $U^{\pm}(a,y_j,\mathbf{s},k/n)\equiv n^{-1/2}S_{nk}$. Also, observe that for each n, $\{S_{nk},\mathscr{F}_{nk};\ 1\le k\le n\}$ is a mean-zero martingale array. Hence, as in the proof of Lemma 3.4, by Doob's and Rosenthal's inequalities, we obtain

$$\begin{aligned}
&P\left(\sup_{j,u}\left|U^{\pm}(a,y_j,\mathbf{s},u)\right|>\alpha\right) \\
&\quad\le N\max_j P\left(\max_{1\le k\le n}\left|U^{\pm}(a,y_j,\mathbf{s},k/n)\right|>\alpha\right)\le Nn^{-2}\max_j E\{S_{nn}^4\}/\alpha^4,
\end{aligned}$$

$$E\{S_{nn}\}^4\le C\left\{E\left[\sum_i E(\zeta_{ni}^2|\mathscr{F}_{ni})\right]^2+\sum_i E(\zeta_{ni})^4\right\}.$$

Because u_{ni} is $\mathscr{F}_{ni}$-measurable, we obtain, from (F), that

$$E(\zeta_{ni}^2|\mathscr{F}_{ni})\le g_{ni}^2|F(y_j+u_{ni})-F(y_j)|\le\|f\|_\infty g_{ni}^2|u_{ni}|, \qquad 1\le i\le n,\ 0\le j\le N+1,\ \text{a.s.},$$

and, using the definitions of u_{ni} and Δ_{ni},

$$\begin{aligned} E\Big[\sum_i E(\zeta_{ni}^2 | \mathscr{F}_{ni})\Big]^2 &\le \|f\|_\infty^2\, E\Big[\sum_i g_{ni}^2 |d_{ni}(\mathbf{s}) + a\Delta_{ni}|\Big]^2 \\ &\le 2\|f\|_\infty^2\, n\Big\{\sum_i E g_{ni}^4 d_{ni}^2(\mathbf{s}) + 2a^2\delta^2 n^{-1}\sum_i E g_{ni}^4 \|\dot{\mathbf{h}}_i(\boldsymbol{\theta})\|^2 \\ &\qquad + 4(b\alpha)^2 n^{-1}\sum_i E g_{ni}^4\Big\} \end{aligned}$$

for all $0 \le j \le N + 1$. Also, because $|\zeta_{ni}| \le |g_{ni}|$, $\sum_i E(\zeta_{ni})^4 = O(n)$, by (g4). From (g4)–(g6) and the definition of N, it thus follows that there exists an n^* and a constant B, depending on $\|f\|_\infty$, $\mathbf{s}$, a and α, but not on n, such that

$$P\Big(\sup_{j,u}\big|U^{\pm}(a, y_j, \mathbf{s}, u)\big| > \alpha\Big) \le BNn^{-1} \le B\{\alpha/(1-\alpha)\}n^{-1/2}, \qquad n > n^*,$$

thereby completing the proof of (1.22). □

PROOF OF (1.12) OF THEOREM 1.2. The proof of (1.12) uses (1.7) and representations similar to the ones used in AR models as in Koul and Ossiander (1992), Section 3. Use Corollary 1.3(a) above to obtain the tightness of the residual empirical process as and when needed. □

REFERENCES

BAI, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* **22** 2051–2061.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

BUSTOS, O. H. (1982). General *M*-estimates for contaminated *p*th order autoregressive processes: consistency and asymptotic normality. *Z. Wahrsch. Verw. Gebiete* **59** 491–504.

CHAN, K. S., PETRUCCELLI, J. D., TONG, H. and WOOLFORD, S. W. (1985). A multiple threshold AR(1) model. *J. Appl. Probab.* **22** 267–279.

HÁJEK, J. (1969). *A Course in Nonparametric Statistics*. Holden-Day, San Francisco.

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

HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* **8** 767–780.

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

HWANG, S. Y. and BASAWA, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series. *Stochastic Process. Appl.* **46** 91–114.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least square estimators. *Ann. Math. Statist.* **40** 633–643.

JUREČKOVÁ, J. (1971). Nonparametric estimation of regression coefficients. *Ann. Math. Statist.* **42** 1328–1338.

KOUL, H. L. (1992). *Weighted Empiricals and Linear Models*. IMS, Hayward, CA.

KOUL, H. L. and OSSIANDER, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.* **22** 540–562.

SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression.* Wiley, New York.

TJØSTHEIM, D. (1986). Estimation in nonlinear time series models I: stationary series. *Stochastic Process. Appl.* **21** 251–273.

TONG, H. (1983). *Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statist.* **21**. Springer, New York.

TONG, H. (1990). *Non-linear Time Series Analysis: A Dynamical System Approach.* Oxford Univ. Press.

WU, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824