The Annals of Statistics
1994, Vol. 22, No. 1, 540–562

WEAK CONVERGENCE OF RANDOMLY WEIGHTED DEPENDENT RESIDUAL EMPIRICALS WITH APPLICATIONS TO AUTOREGRESSION[1]

BY HIRA L. KOUL AND MINA OSSIANDER

Michigan State University and Oregon State University

This paper establishes the uniform closeness of a randomly weighted residual empirical process to its natural estimator via weak convergence techniques. The weights need not be independent, bounded or even square integrable. This result is used to yield the asymptotic uniform linearity of a class of rank statistics in pth-order autoregression models. The latter result, in turn, yields the asymptotic distributions of a class of robust and Jaeckel-type rank estimators. The main result is also used to obtain the asymptotic distributions of the least absolute deviation and certain other robust minimum distance estimators of the autoregression parameter vector.

1. Introduction. This paper establishes the uniform closeness of a randomly weighted residual empirical process to its natural estimator via weak convergence techniques. This result is shown to form a basis for the large sample investigation of an analogue of Jaeckel's 972) rank and some robust estimators of the autoregression parameters in pth-order autoregression [AR(p)] models.

More precisely, let $p \geq 1$ be a fixed integer, let $\mathbb{R}^p$ denote the p-dimensional Euclidean space, $\mathbb{R} = \mathbb{R}^1$, and let $\mathbf{t}'$ denote the transpose of a $p \times 1$ vector $\mathbf{t} \in \mathbb{R}^p$. Let F be a distribution function (d.f.) on $\mathbb{R}$; let $\varepsilon_1, \varepsilon_2, \dots$ be independent and identically distributed (i.i.d.) F random variables (r.v.'s); let $\mathbf{Y}_0 = (X_0, \dots, X_{1-p})'$ be an observable random vector independent of $\{\varepsilon_i, i \geq 1\}$. In an AR($p$) model one observes $\{X_i\}$ satisfying the relation

$$X_i = \rho_1 X_{i-1} + \cdots + \rho_p X_{i-p} + \varepsilon_i, \qquad i \geq 1, \tag{1.1}$$

where $\boldsymbol{\rho} := (\rho_1, \dots, \rho_p)' \in \mathbb{R}^p$ is the parameter vector of interest.

Akin to linear regression, a class of estimators of $\boldsymbol{\rho}$ that are robust against outliers in the errors and have desirable asymptotic efficiency properties can be obtained among rank estimators. To define these, let $\mathbf{Y}_{i-1} := (X_{i-1}, \dots, X_{i-p})'$, $1 \leq i \leq n$, let φ be a nondecreasing real-valued function on $(0, 1)$, let $R_{i\mathbf{t}}$ denote the rank of $X_i - \mathbf{t}'\mathbf{Y}_{i-1}$ among $\{X_j - \mathbf{t}'\mathbf{Y}_{j-1},\ 1 \leq j \leq n\}$, $1 \leq i \leq n$, $\mathbf{t} \in \mathbb{R}^p$, and let $\mathbf{g} := (g_1, g_2, \dots, g_p)'$ be a p-vector of measurable functions from $\mathbb{R}^p$ to

Received March 1992; revised January 1993.
[1]Research partly done while authors were visiting MSRI, Berkeley in autumn 1991; partly supported by NSF Grants DMS-85-05550, DMS-88-06206, DMS-91-02041 and DMS-91-03738.
AMS 1991 *subject classifications.* Primary 60F17; secondary 62M10.
Key words and phrases. Asymptotic uniform linearity of rank statistics, Jaeckel, generalized and minimum distance rank estimators.

$\mathbb{R}$. Define, for $0 \le u \le 1, \mathbf{t} \in \mathbb{R}^p$,

$$
\begin{aligned}
\mathcal{J}(\mathbf{t}) &:= \sum_i \varphi\Big(\frac{R_{it}}{n+1}\Big)(X_i - \mathbf{t}'\mathbf{Y}_{i-1}),\\
S_j(\mathbf{t}) &:= n^{-1}\sum_i g_j(\mathbf{Y}_{i-1})\varphi\Big(\frac{R_{it}}{n+1}\Big),\\
Z_j(u,\mathbf{t}) &:= n^{-1}\sum_i g_j(\mathbf{Y}_{i-1}) I(R_{it} \le nu),\\
\mathcal{Z}_j(u,\mathbf{t}) &:= Z_j(u,\mathbf{t}) - \bar{g}_j u,\\
\bar{g}_j &:= n^{-1}\sum_i g_j(\mathbf{Y}_{i-1}), \qquad 1 \le j \le p.
\end{aligned} \tag{1.2}
$$

Here and in the sequel, the index i in the summation and the maximum varies from 1 to n, unless specified otherwise. Let $\mathbf{S}_g(\mathbf{t}) := (S_1(\mathbf{t}), \dots, S_p(\mathbf{t}))'$, $\mathbf{Z}_g(u,\mathbf{t}) := (Z_1(u,\mathbf{t}), \dots, Z_p(u,\mathbf{t}))'$, $\boldsymbol{\mathcal{Z}}_g(u,\mathbf{t}) := (\mathcal{Z}_1(u,\mathbf{t}), \dots, \mathcal{Z}_p(u,\mathbf{t}))'$, $0 \le u \le 1, \mathbf{t} \in \mathbb{R}^p$. Write $\mathbf{S}, \mathbf{Z}$ and so on for $\mathbf{S}_g, \mathbf{Z}_g$ and so on whenever $\mathbf{g}$ is the identity vector, that is, $\mathbf{g}(\mathbf{y}) \equiv \mathbf{y}, \mathbf{y} \in \mathbb{R}^p$.

Jaeckel (1972) used an analogue of $\mathcal{J}$ suitable in linear regression models, where $\mathbf{Y}_{i-1}'$ is replaced by the ith row of the design matrix, to define a class of rank estimators of the slope parameters. Hettmansperger (1984) has shown that these estimators and the corresponding $\mathcal{J}$ play a fundamental role in the development of the analysis of variance based on ranks in linear regression models. Using Jaeckel's arguments, one can prove the following facts in the present case.

If $(\mathbf{Y}_0', X_1, \dots, X_n)$ has a continuous joint distribution, and if φ satisfies

$$\sum_i \varphi\big(i/(n+1)\big) = 0, \tag{φ1}$$

then $\mathcal{J}$ is nonnegative, continuous and convex on $\mathbb{R}^p$. In addition, the almost everywhere differential of $\mathcal{J}$ is $-n\mathbf{S}$ and, for every $0 \le b < \infty$, the set $\{\mathbf{t} \in \mathbb{R}^p; \mathcal{J}(\mathbf{t}) \le b\}$ is bounded for all those data points $(\mathbf{Y}_0', X_1, \dots, X_n)$ for which $\sum_i (\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})'$ is positive definite, where $\overline{\mathbf{Y}} := (\overline{X}_1, \dots, \overline{X}_p)'$, $\overline{X}_j := n^{-1}\sum_i X_{i-j}$, $1 \le j \le p$. Consequently, the estimator

$$\hat{\rho}_J := \arg\min_{\mathbf{t}} \mathcal{J}(\mathbf{t}) \tag{1.3}$$

is well defined.

Unlike in linear regression, the estimators $\{\hat{\rho}_J\}$ are not robust against outliers in the errors in (1.1). A class of such rank estimators, akin to G-M estimators of Denby and Martin (1979), is given by generalized rank (G-R) estimators

$$\hat{\rho}_{\mathbf{g}} := \arg\min_{\mathbf{t}} \|\mathbf{S}_{\mathbf{g}}(\mathbf{t})\|, \tag{1.4}$$

where $\mathbf{g}$ is typically a vector of bounded functions. An example of $\mathbf{g}$ would be the Huber function [Huber (1981)]: $\mathbf{g}(\mathbf{y}) := \mathbf{y}I(\|\mathbf{y}\| \le k) + \mathbf{y}\,\|\mathbf{y}\|^{-1}I(\|\mathbf{y}\| > k)$,

$\mathbf{y} \in \mathbb{R}^p$, where k and c are some constants. The choice of $\mathbf{g}(\mathbf{y}) \equiv \mathbf{y}$ makes $\widehat{\rho}_{\mathbf{g}}$ asymptotically equal to $\widehat{\rho}_J$.

Another class of estimators of ρ that has desirable efficiency and robustness properties is obtained by minimizing certain L_2-norms based on $\mathcal{Z}_{\mathbf{g}}$-processes of (1.2). More precisely, let G be a d.f. on [0, 1] and define

$$(1.5)\quad K_{\mathbf{g}}(\mathbf{t}) := \int_0^1 \left\|n^{1/2}\mathcal{Z}_{\mathbf{g}}(u,\mathbf{t})\right\|^2 dG(u), \qquad \mathbf{t} \in \mathbb{R}^p; \qquad \widetilde{\rho}_{\mathbf{g}} := \arg\min_{\mathbf{t}} K_{\mathbf{g}}(\mathbf{t}).$$

Analogues of $\widehat{\rho}_{\mathbf{g}}$ and $\tilde{\rho}_{\mathbf{g}}$ in linear regression appear in Koul (1992).

It is clear from the above definitions that the investigation of the asymptotic distributions of these estimators is facilitated by the weak convergence properties of $\{n^{1/2}\mathcal{Z}_{\mathbf{g}}(u, \rho + n^{-1/2}\mathbf{s}); 0 \le u \le 1, \mathbf{s} \in \mathbb{R}^p\}$. As will be seen in Section 3, the processes whose weak convergence properties in turn facilitate this investigation are of the type

$$(1.6)\quad \begin{aligned} W_h(y,\mathbf{t}) &:= n^{-1}\sum_i h(\mathbf{Y}_{i-1})I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \le y), \\ \nu_h(y,\mathbf{t}) &:= n^{-1}\sum_i h(\mathbf{Y}_{i-1})F(y + (\mathbf{t}-\rho)'\mathbf{Y}_{i-1}), \\ \mathcal{W}_h(y,\mathbf{t}) &:= n^{1/2}[W_h(y,\mathbf{t}) - \nu_h(y,\mathbf{t})], \qquad y \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p, \end{aligned}$$

where h is a measurable function from $\mathbb{R}^p$ to $\mathbb{R}$.

This paper actually contains a more general result which, when specialized to the above model, yields the required weak convergence properties of $\{\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s});\ y \in \mathbb{R},\ \mathbf{s} \in \mathbb{R}^p\}$, under some growth conditions on $\{\mathbf{Y}_{i-1}, h(\mathbf{Y}_{i-1})\}$. To state this general result, let $(\Omega, \mathcal{A}, P)$, be a probability space and let H be a d.f. on $\mathbb{R}$. For each integer $n \ge 1$, let $(\eta_{ni}, \xi_{ni}, \gamma_{ni})$, $1 \le i \le n$, be an array of trivariate r.v.'s defined on $(\Omega, \mathcal{A})$ such that $\{\eta_{ni}, 1 \le i \le n\}$ are i.i.d. r.v.'s with d.f. H and η_{ni} is independent of (γ_{ni}, ξ_{ni}), $1 \le i \le n$. Furthermore, let $\{\mathcal{A}_{ni}\}$ be an array of sub-σ-fields such that $\mathcal{A}_{ni} \subset \mathcal{A}_{ni+1}$, $\mathcal{A}_{ni} \subset \mathcal{A}_{n+1i}$, $1 \le i \le n$, $n \ge 1$; (γ_{ni}, ξ_{ni}) is $\mathcal{A}_{n1}$-measurable, the r.v.'s $\{\eta_{n1}, \dots, \eta_{nj-1}; (\gamma_{n1}, \xi_{n1}), 1 \le i \le j\}$ are $\mathcal{A}_{nj}$-measurable, $2 \le j \le n$; and η_{nj} is independent of $\mathcal{A}_{nj}$, $1 \le j \le n$. Define, for an $x \in \mathbb{R}$,

$$(1.7)\quad \begin{aligned} V_n(x) &:= n^{-1}\sum_{i=1}^n \gamma_{ni}I(\eta_{ni} \le x + \xi_{ni}), \qquad V_n^*(x) := n^{-1}\sum_{i=1}^n \gamma_{ni}I(\eta_{ni} \le x), \\ J_n(x) &:= n^{-1}\sum_{i=1}^n E\{\gamma_{ni}I(\eta_{ni} \le x + \xi_{ni})|\mathcal{A}_{ni}\} \\ &= n^{-1}\sum_{i=1}^n \gamma_{ni}H(x + \xi_{ni}), \\ J_n^*(x) &:= n^{-1}\sum_{i=1}^n \gamma_{ni}H(x), \\ U_n(x) &:= n^{1/2}(V_n(x) - J_n(x)), \qquad U_n^*(x) := n^{1/2}(V_n^*(x) - J_n^*(x)). \end{aligned}$$

Note that if in (1.7) we take

$$\begin{aligned} &\gamma_{ni} = h(\mathbf{Y}_{i-1}), \qquad \eta_{ni} = \varepsilon_i, \qquad \xi_{ni} = n^{-1/2}\mathbf{s}'\mathbf{Y}_{i-1}, \\ &\mathcal{A}_{n1} = \sigma\text{-field}\{\mathbf{Y}_0\}, \quad \mathcal{A}_{ni} = \sigma\text{-field}\{\mathbf{Y}_0, \varepsilon_1, \dots, \varepsilon_{i-1}\}, \qquad i \geq 2, \end{aligned} \tag{1.8}$$

then $V_n(y)$, $V_n^*(y)$, $U_n(y)$ and $U_n^*(y)$ are, respectively, equal to $W_h(y, \rho + n^{-1/2}\mathbf{s})$, $W_h(y, \rho)$, $\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s})$ and $\mathcal{W}_h(y, \rho)$, for all $y \in \mathbb{R}$ and for each $\mathbf{s} \in \mathbb{R}^p$.

To state the main result of this paper, we also need to introduce a pseudometric

$$d_b(x,y) := \sup_{|a| \leq b} \left| H(x+a) - H(y+a) \right|^{1/2}, \qquad x, y \in \mathbb{R} \text{ and } b \geq 0;$$

the covering number of $\mathbb{R}$ with respect to d_b and the entropy integral, respectively, given by

$$\begin{aligned} N(\delta, b) &:= \text{cardinality of the minimal } \delta\text{-net of } (\mathbb{R}, d_b); \qquad \delta > 0, b \geq 0; \\ \mathcal{I}(b) &:= \int_0^1 \{\ln N(u,b)\}^{1/2} \, du, \qquad b \geq 0. \end{aligned}$$

The function $\ln N(\cdot, b)$ is known as the metric entropy of $\mathbb{R}$ with respect to the pseudometric d_b. This function has been used extensively in studying the weak convergence of empirical processes [see, e.g., Dudley (1978) and Ossiander (1987)].

In what follows, $o_p(1)$ $[O_p(1)]$ stands for a sequence of r.v.'s that converges to zero (is bounded) in probability, $\|h\|_\infty := \sup\{h(x); x \in \mathbb{R}\}$ for any function h from $\mathbb{R}$ to $\mathbb{R}$, $\ln x := \log_e(x)$, $x > 0$, and all limits are taken as $n \to \infty$, unless specified otherwise.

Since the processes $\{U_n\}$ and $\{U_n^*\}$, although right continuous with left limits, are possibly nonmeasurable in the uniform metric, we use here the Hoffman–Jorgensen definition of weak convergence: Let $L := \{f\colon \mathbb{R} \to \mathbb{R}, \|f\|_\infty < \infty\}$ be a subset of bounded functions in $\mathbb{R}$. A sequence of L-valued processes $\{T_n; n \geq 1\}$ is said to converge weakly in distribution to an L-valued process T, $(T_n \Rightarrow T)$, if $E^*g(T_n) \to Eg(T)$ for every $g \in \mathbb{C}(L, \|\cdot\|_\infty)$, where E^* denotes the outer expectation.

It is well known that in our context, weak convergence will follow from finite-dimensional convergence and asymptotic equicontinuity (eventual tightness) with respect to some totally bounded psuedometric, which we take to be d_b. Details of the argument involved can be found, for example, in Andersen and Dobrić (1987). See also Dudley (1984) and Giné and Zinn (1986) for some other related issues. We are now ready to state our main theorem.

THEOREM 1.1. *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) $$\max_{1\le i\le n}|\xi_{ni}| = o_p(1).$$

(A4) *H is continuous.*

(A5) *There exists a $b_0 > 0$ such that $\mathcal{I}(b_0) < \infty$.*

Then the processes $\{U_n\}$ and $\{U_n^*\}$ are eventually tight in the uniform metric and

(1.9) $$\|U_n - U_n^*\|_\infty = o_p(1).$$

Under the additional assumption (A6),

(1.10) $$U_n \Rightarrow \gamma B(H), \qquad U_n^* \Rightarrow \gamma B(H),$$

where B is a Brownian bridge in $C[0,1]$, independent of γ, and where the following holds:

(A6) *For each $n \ge 1$, $\{\gamma_{ni};\ 1 \le i \le n\}$ is square integrable.*

Conditions (A1), (A2) and (A6) suffice for the finite-dimensional convergence of U_n^* while (A3) and (A4) are added to yield the finite-dimensional convergence of U_n. The eventual tightness of U_n^* requires (A1)–(A4) only while that of U_n requires (A1)–(A5).

Here, we briefly discuss (A5). Under (A4), $N(\delta, 0) = [1/\delta^2]$. Moreover, it is not hard to see that

$$N(2^{1/2}\delta, b) \le N(\delta, 0)\big(1 + 2b/\omega(\delta)\big), \qquad b \ge 0,$$

where

$$\omega(\delta) := \min\left\{H^{-1}((k+1)\delta^2) - H^{-1}(k\delta^2);\ 1 \le k < 1/\delta^2\right\},$$

with $H^{-1}(u) := \inf\{x, H(x) \ge u\}$, $0 \le u \le 1$.

Now consider the following assumption:

(A4*) H has a uniformly continuous density h_1 which is positive a.e.

Under (A4*), h_1 is bounded and $\omega(\delta) \ge \delta^2/\|h_1\|_\infty$. Hence,

$$\ln N(u, b) \le -\ln(u^2/2) + \ln(1 + 4b\|h_1\|_\infty/u^2), \qquad 0 \le u \le 1,\ b \ge 0.$$

Thus, if (A4) is strengthened to (A4*), then (A5) holds a priori.

Our proof of Theorem 1.1 uses a truncation argument, an adaptive chaining argument similar to that of Ossiander (1987) and Greenwood and Ossiander (1991) and an exponential inequality of Freedman (1975). It is given in Section 2. In the case where the weights $\{\gamma_{ni}\}$ are *bounded*, a proof of a variant of Theorem 1 appears in Koul (1991). Applications of (1.9) to the empirical processes of the residuals, the weighted empirical processes with bounded weights used in defining the G-M estimators and some other inferential problems in the AR(p) model are also given in that paper. However, as is seen from (1.2) and (1.3) the weights involved in the definition of Jaeckel's estimators are necessarily *unbounded* and hence the result of Koul (1991) is *not* applicable. The weights involved in efficient G-R estimators $\hat{\rho}_{\mathbf{g}}$ or efficient minimum L_2-estimators $\tilde{\rho}_{\mathbf{g}}$ are also generally unbounded. To cover these cases we now have the following lemma.

LEMMA 1.1. *In addition to* (1.1), *assume that the following hold*:

(a1) $$\max_i n^{-1/2}\|\mathbf{Y}_{i-1}\| = o_p(1).$$

(h1) $$\max_i n^{-1/2}\left|h(\mathbf{Y}_{i-1})\right| = o_p(1).$$

(h2) $$\left(n^{-1}\sum_i h^2(\mathbf{Y}_{i-1})\right)^{1/2} = \alpha + o_p(1), \qquad \alpha \text{ a positive r.v.}$$

(h3) $$n^{-1}\sum_i \left\|h(\mathbf{Y}_{i-1})\mathbf{Y}_{i-1}\right\| = O_p(1).$$

(F1) *F has a uniformly continuous density f which is positive a.e.*

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

$$\sup_{y\in\mathbb{R},\ \|\mathbf{s}\|\le b}\left|\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s}) - \mathcal{W}_h(y,\rho)\right| = o_p(1), \tag{1.11}$$

and

$$\begin{aligned}\sup_{y\in\mathbb{R},\ \|\mathbf{s}\|\le b}\Big| & n^{1/2}\left[W_h(y, \rho + n^{-1/2}\mathbf{s}) - \mathcal{W}_h(y,\rho)\right] \\ & -\mathbf{s}' n^{-1}\sum_i h(\mathbf{Y}_{i-1})\mathbf{Y}_{i-1} f(y)\Big| = o_p(1).\end{aligned} \tag{1.12}$$

PROOF. The result (1.9) applied to the entities given in (1.8) readily yields that

$$\sup_y\left|\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s}) - \mathcal{W}_h(y,\rho)\right| = o_p(1), \qquad \forall\ \mathbf{s}\in\mathbb{R}^p.$$

The uniformity with respect to $\mathbf{s}$ is obtained by exploiting the monotonic property of the indicator function and the d.f. and the compactness of the ball $\{\mathbf{t} \in \mathbb{R}^p;\ \|\mathbf{t}\| \le b\}$. The details are similar to those in Koul [(1991), proof of Theorem 1.2.]. The only difference is that one uses (1.9) which allows h to be unbounded, wherever (1.7) of that paper was used. □

Using the asymptotic uniform linearity result of Jurečková (1971), Jaeckel (1972) showed that in the case of linear regression the suitably normalized analogue of $\widehat{\rho}_J$ is asymptotically normally distributed and asymptotically equivalent to the suitably normalized analogue of a solution $\mathbf{t}$ of the equation $\mathbf{S}(\mathbf{t}) = \overline{\mathbf{Y}}\overline{\varphi}$, $\overline{\varphi} := \int \varphi(u)\,du$. The corresponding result for $\widehat{\rho}_J$ under (1.1) is the second major contribution of this paper and is facilitated by the following theorem, which is general enough to be useful in obtaining the asymptotic distributions of $\{\widehat{\rho}_{\mathbf{g}}, \widetilde{\rho}_{\mathbf{g}};\ \mathbf{g} \text{ varies}\}$ of (1.4) and (1.5).

THEOREM 1.2. *Let* $\mathbf{g}(\mathbf{Y}_{i-1}) := (g_1(\mathbf{Y}_{i-1}), \dots, g_p(\mathbf{Y}_{i-1}))'$, $1 \le i \le n$. *In addition to* (1.1), (a1) *and* (F1), *assume that the following hold*:

(a2) $$n^{-1}\sum_i \|\mathbf{Y}_{i-1}\| = O_p(1).$$

(a3) $$n^{-1/2}\max_i \|\mathbf{g}(\mathbf{Y}_{i-1})\| = o_p(1).$$

(a4) $$n^{-1}\sum_i \left\|g_j(\mathbf{Y}_{i-1})\mathbf{Y}_{i-1}\right\| = O_p(1), \qquad 1 \le j \le p.$$

(a5) $$\left\{n^{-1}\sum_i g_j^2(\mathbf{Y}_{i-1})\right\}^{1/2} = \alpha_j + o_p(1), \qquad \alpha_j \text{ a positive r.v.}, 1 \le j \le p.$$

(a6) $n^{-1}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})' = \boldsymbol{\Sigma}_{\mathbf{g}} + o_p(1)$, *where* $\boldsymbol{\Sigma}_{\mathbf{g}}$ *is a positive definite matrix.*

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

$$\sup_{0 \le u \le 1,\ \|\mathbf{s}\| \le b} \left\| n^{1/2}\left[\mathcal{Z}_{\mathbf{g}}(u, \rho + n^{-1/2}\mathbf{s}) - \widehat{\mathcal{Z}}_{\mathbf{g}}(u)\right] - \boldsymbol{\Sigma}_{\mathbf{g}}\mathbf{s} f(F^{-1}(u)) \right\| = o_p(1), \tag{1.13}$$

where

$$\widehat{\mathcal{Z}}_{\mathbf{g}}(u) := n^{-1}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})\left[I(F(\varepsilon_i) \le u) - u\right], \qquad 0 \le u \le 1. \tag{1.14}$$

In addition, if $\varphi \in \mathcal{C}$, *where*

(φ2) $\mathcal{C} := \{\varphi\colon [0,1] \to \mathbb{R},\ \varphi \text{ nondecreasing, right continuous}, \varphi(1) - \varphi(0) = 1\}$, *then*, $\forall\, 0 < b < \infty$,

$$(1.15)\quad \sup_{\varphi\in\mathcal{C},\, \|\mathbf{s}\|\le b} \left\| n^{1/2}\left\{\mathbf{S_g}(\rho + n^{-1/2}\mathbf{s}) - \overline{\mathbf{g}}\overline{\varphi} - \widehat{\mathbf{S}}_\mathbf{g}\right\} + \mathbf{\Sigma_g s}\int f\, d\varphi(F)\right\| = o_p(1),$$

where

$$(1.16)\quad \widehat{\mathbf{S}}_\mathbf{g} = n^{-1}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})\left[\varphi(F(\varepsilon_i)) - \overline{\varphi}\right], \qquad \overline{\mathbf{g}} := n^{-1}\sum_i \mathbf{g}(\mathbf{Y}_{i-1}).$$

REMARK 1.1. The result (1.15) is an analogue of Jurečková's linearity result suitable under (1.1). It readily follows from (1.13) and the identity $\mathbf{S_g}(\mathbf{t}) \equiv \overline{\mathbf{g}}\,\overline{\varphi} - \int_0^1 \mathcal{Z}_\mathbf{g}(u(n+1)/n, \mathbf{t})\, d\varphi(u)$.

A proof of (1.13) appears in Section 3. In the case $p = 1$, $g_1(y) \equiv y$ and $E\varepsilon^4 < \infty$, its analogue was proved in Koul and Sen (1991), using a very cumbersome truncation argument, the AR(1) structure and the weak convergence techniques of Billingsley (1968). Such a proof does not extend easily to the AR(p) model (1.1) and the general processes $\mathcal{Z}_\mathbf{g}$. Theorem 1.2 gives an important extension of that result to a larger class of linear rank processes $\mathcal{Z}_\mathbf{g}$ and statistics $\mathbf{S_g}$ and to AR(p) models whose errors need not even have finite second moments. However, if $\mathbf{Y}_0$, $\{\varepsilon_i\}$ are so chosen as to make $\mathbf{Y}_0, X_1, X_2, \ldots$ stationary, ergodic and if $E\|\mathbf{g}(\mathbf{Y}_0)\|^2 < \infty$, $E\varepsilon_1^2 < \infty$, then (a1)–(a6) are a priori satisfied.

The asymptotic distributions of the above estimators is given in the following lemma and two corollaries. Their proofs appear in Section 3.

LEMMA 1.2. *In addition to* (1.1), (F1) *and* (φ2), *assume that the following hold*:

(a7) *All roots of the equation* $x^p - \rho_1 x^{p-1} - \cdots - \rho_p = 0$ *are inside the unit circle.*

(a8) $E\varepsilon_1^2 < \infty$.

(a9) $E\|\mathbf{g}(\mathbf{Y}_0)\|^2 < \infty$.

Then $\mathbf{\Gamma_g} := p\lim_n n^{-1}\Sigma_i(\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})(\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})'$ *exists and*

$$(1.17)\qquad n^{1/2}\widehat{\mathbf{S}}_\mathbf{g} \Rightarrow N_p(\mathbf{0}, \sigma_\varphi^2\, \mathbf{\Gamma_g}),$$

where $\sigma_\varphi^2 := \mathrm{Var}[\varphi F(\varepsilon_1))]$.

COROLLARY 1.1. *Assume that* (1.1), (φ1), (F1), (φ2), (a7) *and* (a8) *hold. Then*

$$\mathbf{\Sigma} := p\lim_n n^{-1}\sum_i (\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})'$$

exists, is positive definite and

$$n^{1/2}(\widehat{\rho}_J - \rho) \Rightarrow N\left(\mathbf{0}, \sigma_\varphi^2\left(\int f\, d\varphi(F)\right)^{-2}\mathbf{\Sigma}^{-1}\right). \tag{1.18}$$

If, in addition, (a9) *and*

(a10)
$$\begin{aligned} &\textit{either}\quad \boldsymbol{\theta}'(\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})'\boldsymbol{\theta} \geq 0,\\ &\qquad\qquad \forall\, \boldsymbol{\theta} \in \mathbb{R}^p \textit{ with } \|\boldsymbol{\theta}\| = 1 \textit{ and } \forall\, 1 \leq i \leq n, \textit{a.s.},\\ &\textit{or}\quad \boldsymbol{\theta}'(\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})'\boldsymbol{\theta} \leq 0,\\ &\qquad\qquad \forall\, \boldsymbol{\theta} \in \mathbb{R}^p \textit{ with } \|\boldsymbol{\theta}\| = 1 \textit{ and } \forall\, 1 \leq i \leq n, \textit{a.s.}, \end{aligned}$$

hold, then

$$n^{1/2}(\widehat{\rho}_{\mathbf{g}} - \rho) \Rightarrow N\left(\mathbf{0}, \sigma_\varphi^2\left(\int f\, d\varphi(F)\right)^{-2}\mathbf{\Sigma}_{\mathbf{g}}^{-1}\mathbf{\Gamma}_{\mathbf{g}}\mathbf{\Sigma}_{\mathbf{g}}^{-1}\right). \tag{1.19}$$

COROLLARY 1.2. *Assume that* (1.1), (F1), (a7), (a8), (a9) *and* (a10) *hold. Then*

$$n^{1/2}(\widehat{\rho}_{\mathbf{g}} - \rho) \Rightarrow N\left(\mathbf{0}, \tau^2\mathbf{\Sigma}_{\mathbf{g}}^{-1}\mathbf{\Gamma}_{\mathbf{g}}\mathbf{\Sigma}_{\mathbf{g}}^{-1}\right), \tag{1.20}$$

where

$$\tau^2 := \frac{\int_0^1\int_0^1[u \wedge v - uv]q(u)q(v)\, dG(u)\, dG(v)}{\left(\int_0^1 q^2\, dG\right)^2}, \qquad q := f(F^{-1}).$$

Another estimator whose asymptotic distribution can be obtained from some of the results in this paper is the *least absolute deviation* (LAD) estimator defined as a minimizer $\Sigma_i|X_i - \mathbf{t}'\mathbf{Y}_{i-1}|$ or, alternatively, the estimator $\widehat{\rho}_{\text{lad}} := \arg\min_{\mathbf{t}}\ \{\mathcal{S}(\mathbf{t})\}^2$, where $\mathcal{S}(\mathbf{t}) := n^{-1/2}\Sigma_i\mathbf{Y}_{i-1}\{I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \leq 0) - \frac{1}{2}\}$. Using Corollary 2.3 and certain monotonic structure that is present in $\mathcal{S}(\mathbf{t})$ we obtain Corollary 1.3:

COROLLARY 1.3. *In addition to* (1.1), (a7) *and* (a8), *assume that the following holds*:

(F2) *The error d.f. F has density f in an open neighborhood of 0 such that f is positive and continuous at 0 and $F(0) = \frac{1}{2}$.*

Then, $n^{-1}\Sigma_i\mathbf{Y}_{i-1}\mathbf{Y}_{i-1}' = \mathbf{\Sigma}_1 + o_p(1)$, where $\mathbf{\Sigma}_1$ is a positive definite matrix, and

$$n^{1/2}(\widehat{\rho}_{\text{lad}} - \rho) \Rightarrow N\left(\mathbf{0}, \mathbf{\Sigma}_1^{-1}/4f^2(0)\right).$$

The proof of this corollary also appears in Section 3. We shall now briefly discuss the asymptotic relative efficiencies of the above estimators. For convenience take $p = 1$ and assume that $E\varepsilon_1 = 0$. Let $\sigma^2 := E\varepsilon_1^2$. Then $\Sigma = (1-\rho^2)^{-1}\sigma^2 = \Sigma_1$. Let σ_r^2, σ_{lad}^2 and σ_{md}^2 denote, respectively, the asymptotic variance of $n^{1/2}(\widehat{\rho}_J - \rho)$, $n^{1/2}(\widehat{\rho}_{\text{lad}} - \rho)$ and $n^{1/2}(\widetilde{\rho} - \rho)$, where $\widetilde{\rho}$ is the $\widetilde{\rho}_{\mathbf{g}}$ with $\mathbf{g}(\mathbf{y}) \equiv \mathbf{y}$. Thus, with $\alpha^2 := (1-\rho^2)\sigma^{-2}$,

$$(1.21)\qquad \sigma_r^2 = \sigma_\varphi^2\left(\int f\,d\varphi(F)\right)^{-2}\alpha^2, \qquad \sigma_{\text{lad}}^2 = \left(2f(0)\right)^{-2}\alpha^2, \qquad \sigma_{\text{md}}^2 = \tau^2\alpha^2.$$

Now take $\varphi(u) \equiv u \equiv G(u)$ and denote the corresponding $\widehat{\rho}_J$ as $\widehat{\rho}_W$ and $\widetilde{\rho}$ as $\widetilde{\rho}_I$. Then, for $F(x) = 1/(1+\exp(-x))$, $x \in \mathbb{R}$, $\tau^2 = 3.0357$, $(2f(0))^{-2} = 4$ and $\sigma_\varphi^2(\int f\,d\varphi(F))^{-2} = 3$; for F with density f given by $f(x) = \frac{1}{2}\exp(-|x|)$, $x \in \mathbb{R}$, $\tau^2 = 1.2$, $(2f(0))^{-2} = 1$ and $\sigma_\varphi^2(\int f\,d\varphi(F))^{-2} = 1.333$; for F equal to the d.f. of $N(0,1)$ r.v., $\tau^2 = 1.0946$, $(2f(0))^{-2} = \pi/2 \approx 1.5708$ while $\sigma_\varphi^2(\int f\,d\varphi(F))^{-2} = \pi/3 \approx 1.0472$.

It thus follows that at the double exponential errors the m.d. estimator $\widetilde{\rho}_I$ is asymptotically more efficient than the Wilcoxon-type rank estimator $\widehat{\rho}_W$, and the two estimators have almost the same asymptotic variance at the logistic and the normal errors. Similarly, $\widetilde{\rho}_I$ is asymptotically more efficient than $\widehat{\rho}_{\text{lad}}$ at the logistic and normal errors. Moreover, from (1.21) and Lehmann (1975), it follows that $\widehat{\rho}_W$ ($\widehat{\rho}_{\text{lad}}$) is asymptotically most efficient at the logistic (double exponential) errors. Finally, compared to the least square estimator, all three are more efficient at heavy tail error distributions.

2. Proof of Theorem 1.1. The proof of Theorem 1.1 will be a consequence of the following several lemmas.

LEMMA 2.1. *If* (A1), (A2) *and* (A6) *hold, then the finite-dimensional distributions of* $\{U_n^*(x)\colon x \in \mathbb{R}\}$ *converge to those of* $\{U(x)\colon x \in \mathbb{R}\}$, *where* $U(\cdot) = \gamma B(H(\cdot))$ *with* B *denoting a Brownian bridge in* $C[0,1]$, *independent of* γ.

PROOF. We need only check that the conditions of Corollary 3.1 of Hall and Heyde (1980) hold. For simplicity, we will do the necessary calculation for a single point x. Note that the $\mathcal{F}_{n,i-1}$ of Hall and Heyde is our $\mathcal{A}_{ni}$. First, for any fixed $\varepsilon, \beta > 0$,

$$\limsup_n P\left(n^{-1}\sum_i \gamma_{ni}^2 I\left[|\gamma_{ni}| > \varepsilon n^{1/2}\right] > \beta\right) \le \limsup_n P\left(\max_i |\gamma_{ni}| > \varepsilon n^{1/2}\right) = 0.$$

Next, for each fixed $x \in \mathbb{R}$,

$$n^{-1}\sum_i E\left[\gamma_{ni}^2\{I[\eta_{ni} \le x] - H(x)\}^2 | \mathcal{A}_{ni}\right] = n^{-1}\sum_i \gamma_{ni}^2 H(x)(1-H(x))$$
$$= H(x)(1-H(x))\gamma^2 + o_p(1). \qquad \square$$

The next lemma, due to Freedman (1975), is used repeatedly to obtain an upper bound on the tail probability of the sum M_n of martingale differences on the set where the quadratic variation of M_n is bounded.

LEMMA 2.2. *Suppose that* $M_n = \sum_{i=1}^n D_i$ *is a sum of martingale differences defined on the increasing filtration* $\{\mathcal{G}_i\colon 1 \le i \le n\}$ *with* $|D_i| \le a$, *a.s.*, $1 \le i \le n$. *Then, for any* $\eta \wedge \alpha > 0$,

$$P\left([M_n > \eta] \cap \left[\sum_{i=1}^n E(D_i^2 | \mathcal{G}_{i-1}) \le \alpha\right]\right) \le \exp\left\{-\eta^2/2(a\eta + \alpha)\right\}. \tag{2.1}$$

To state the next lemma it is convenient to introduce the following events:

$$\begin{aligned} A_{na} &:= \left[\max_i |\gamma_{ni}| \le a n^{1/2}\right], \\ B_{nb} &:= \left[\max_i |\xi_{ni}| \le b\right], \qquad a \wedge b > 0, \\ C_{nc} &:= \left[n^{-1}\sum_i \gamma_{ni}^2 \le c\right], \qquad c > 0. \end{aligned} \tag{2.2}$$

LEMMA 2.3. *For any fixed* $x \in \mathbb{R}$ *and* $a \wedge b \wedge c \wedge \eta > 0$, *with* $H(x+b) - H(x-b) \le a$,

$$P\left(\left[|U_n(x) - U_n^*(x)| > \eta\right] \cap A_{na} \cap B_{nb} \cap C_{nc}\right) \le \exp\{-\eta^2/2a(\eta + c)\}. \tag{2.3}$$

PROOF. Fix an $x \in \mathbb{R}$ and choose $a \wedge b \wedge \eta > 0$, with $H(x+b) - H(x-b) \le a$. Using the monotonicity of H, on the set B_{nb}, the quadratic variation of $U_n(x) - U_n^*(x)$ is bounded above as follows:

$$\begin{aligned} &n^{-1}\sum_i \gamma_{ni}^2 E\left[\{I(\eta_{ni} \le x + \xi_{ni}) - I(\eta_{ni} \le x) - H(x+\xi_{ni}) + H(x)\}^2 | \mathcal{A}_{ni}\right] \\ &\quad \le n^{-1}\sum_i \gamma_{ni}^2 [H(x+b) - H(x-b)] \le a n^{-1}\sum_i \gamma_{ni}^2. \end{aligned}$$

Thus, applying Lemma 2.2 we obtain the following: the left-hand side of (2.3) is bounded above by

$$\begin{aligned} P\Bigg(\Bigg[\Bigg| &n^{-1/2}\sum_i \gamma_{ni} I\left[|\gamma_{ni}| \le a n^{1/2}\right] \\ &\times \{I(\eta_{ni} \le x + \xi_{ni}) - I(\eta_{ni} \le x) - H(x+\xi_{ni}) + H(x)\}\Bigg| > \eta\Bigg] \\ &\qquad \cap \left[n^{-1}\sum_i \gamma_{ni}^2 [H(x+b) - H(x-b)] \le ac\right]\Bigg) \\ &\le \exp\{-\eta^2/2a(\eta + c)\}. \end{aligned}$$

□

The finite-dimensional asymptotic equivalence of U_n and U_n^* follows quite easily from this lemma.

COROLLARY 2.1. *If* (A1)–(A4) *hold, then for each fixed* $x \in \mathbb{R}$,

$$|U_n(x) - U_n^*| = o_p(1). \tag{2.4}$$

PROOF. Fix $x \in \mathbb{R}$, $\varepsilon \wedge \eta > 0$. Choose $c > 0$ sufficiently large to have $P(\gamma^2 > c) < \varepsilon$; choose $a > 0$ sufficiently small to have $\exp\{-\eta^2/2a(\eta + c)\} < \varepsilon$; and choose $b > 0$ sufficiently small to have $[H(x+b) - H(x-b)] \le a$. Then

$$\begin{aligned}
&P\Big(|U_n(x) - U_n^*(x)| > \eta\Big)\\
&\quad \le P\Big(\Big[|U_n(x) - U_n^*(x)| > \eta\Big] \cap A_{na} \cap B_{nb} \cap C_{nc}\Big) + P((A_{na} \cap B_{nb} \cap C_{nc})^c)\\
&\quad \le \exp\{-\eta^2/2a(\eta + c)\} + P\Big(\max_i |\gamma_{ni}| > an^{1/2}\Big)\\
&\qquad + P\Big(\max_i |\xi_{ni}| > b\Big) + P\Big(n^{-1}\sum_i \gamma_{ni}^2 > c\Big).
\end{aligned}$$

The first term above is less than ε. As n tends to infinity, the second and third terms converge to 0, by (A2) and (A3), and by (A1) the fourth term tends to $P(\gamma^2 > c) < \varepsilon$. Thus

$$\limsup_n P\Big(|U_n(x) - U_n^*(x)| > \eta\Big) < 2\varepsilon.$$

Since ε is arbitrary, this completes the proof. □

COROLLARY 2.2. *If* (A1)–(A4) *and* (A6) *hold, then the finite-dimensional distributions of* $\{U_n(x)\colon x \in \mathbb{R}\}$ *converge weakly to those of* $\{U(x)\colon x \in \mathbb{R}\}$.

From the proof of Corollary 2.1, it is clear that the following also holds.

COROLLARY 2.3. *If* (A1)–(A3) *hold, then, for any fixed* $x \in \mathbb{R}$ *at which* H *is continuous*, $|U_n(x) - U_n^*(x)| = o_p(1)$.

To state the next result we need the following notation: For an $x \in \mathbb{R}$, $\delta \wedge b > 0$, let $\pi_{\delta b}(x)$ be a real number such that $\pi_{\delta b}(x) \ge x$, $d_b(\pi_{\delta b}(x), x) \le \delta$ and $\pi_{\delta b}(x)$ belongs to a minimal δ-net in $(\mathbb{R}, d_b) \cup \{\infty\}$. In addition, define

$$I_{n\eta}(\delta, b) := \int_{\delta(n)}^{\delta} (1 + \ln N(u,b))^{1/2}\, du, \quad \eta \ge 1, \quad \frac{\delta}{(1 + \ln N(\delta, b))^{1/2}} \ge 4\Big(\frac{\eta}{n}\Big)^{1/2},$$

where $2\delta(n) := \delta^{1/2}\{(1 + \ln N(\delta, b))\eta/n\}^{1/4}$. We are now ready to state and prove the following.

PROPOSITION 2.1 (Eventual tightness of U_n in d_b metric). *For any* $n \geq 1$, $b \geq 0$, $\eta \geq 1$ *and* $\delta > 0$, *with* $\delta/(1+\ln N(\delta,b))^{1/2} \geq 4(\eta/n)^{1/2}$,

$$P\Big(\Big[\sup_x |U_n(x) - U_n(\pi_{\delta b}(x))| > (c_1 + \eta c_2)(\delta + I_{n\eta}(\delta, b))\Big]$$

$$\cap \Big[\max_i |\gamma_{ni}| \leq \delta/(1+\ln N(\delta,b))^{1/2}\Big] \cap B_{nb} \cap C_{n\eta}\Big) \tag{2.5}$$

$$\leq c_3 \exp\{-\eta\},$$

where c_1, c_2 *and* c_3 *are universal constants.*

PROOF. Fix $n \geq 1$, $b \geq 0$, $\eta \geq 1$ and $\delta > 0$, with $\delta/(1+\ln N(\delta,b))^{1/2} \geq 4(\eta/n)^{1/2}$. For $k \geq 0$, let $\delta_k = 2^{-k}\delta$, and for $x \in \mathbb{R}$ let the pair (x_k', x_k) satisfy the following:

$$d_b(x, x_k) < \delta_k, \qquad d_b(x, x_k') < \delta_k, \qquad x_k' < x \leq x_k, \tag{2.6}$$

where x_k and x_k' are both either in the minimal δ_k-net in $(\mathbb{R}, d_b)$ or are $\pm\infty$. We also need to define the following: $L_k := 0$, $k < e$; $L_k := \ln k$, $k \geq e$;

$$h_k = (2L_k)^{1/2} + (1+\ln N(\delta_k, b))^{1/2}, \qquad a_k = \delta_k/h_k, \qquad \eta_k = \delta_k h_k, \qquad k \geq 0;$$

and

$$m = \min\{k: \delta_k^2 \leq 4\eta_0(\eta/n)^{1/2}\}.$$

Notice that $\{a_k\}$ is a strictly decreasing sequence with $a_k \to 0$ as $k \to \infty$. Moreover,

$$\delta_{m+1} = \delta_{m-1}/4 \geq \eta_0^{1/2}\eta^{1/4}/2n^{1/4}$$

and

$$\sum_{0\leq k\leq m} \eta_k = \sum_{0\leq k\leq m} \Big[\delta_k(2L_k)^{1/2} + 2\int_{\delta_{k+1}}^{\delta_k} (1+\ln N(\delta_k, b))^{1/2}\, du\Big]$$

$$\leq \sum_{0\leq k\leq m} \Big[2^{-k}\delta(2L_k)^{1/2} + 2\int_{\delta_{k+1}}^{\delta_k} (1+\ln N(u, b))^{1/2}\, du\Big]$$

$$\leq c\delta + 2I_{n\eta}(\delta, b),$$

where c is a fixed constant.

Next, define for $1 \leq i \leq n$,

$$\zeta_{ni}^{1k} = I[|\gamma_{ni}| \leq a_k n^{1/2}], \qquad 0 \leq k \leq m$$

and

$$\zeta_{ni}^{2k} = \begin{cases} I[a_{k+1}n^{1/2} < |\gamma_{ni}| \leq a_k n^{1/2}], & 0 \leq k < m, \\ \zeta_{ni}^{1k}, & k = m. \end{cases}$$

Also let

$$\begin{aligned} D_{ni}(x) &= I(\eta_{ni} \le x + \xi_{ni}) - H(x + \xi_{ni}), && 1 \le i \le n, \\ S_{nk}(x) &= n^{-1/2} \sum_i \gamma_{ni} \zeta_{ni}^{1k} [D_{ni}(x_{k-1}) - D_{ni}(x_k)], && 1 \le k \le m, \\ R_{nk}(x) &= n^{-1/2} \sum_i |\gamma_{ni}| \zeta_{ni}^{2k} [I(x_k' + \xi_{ni} < \eta_{ni} \le x_k + \xi_{ni}) \\ &\qquad + H(x_k + \xi_{ni}) - H(x_k' + \xi_{ni})], && 0 \le k \le m. \end{aligned}$$

We are finally ready to start our stratification and chaining. For any $x \in \mathbb{R}$,

$$\begin{aligned} \left|U_n(x_0) - U_n^*(x)\right| \le &\left|n^{-1/2} \sum_i \gamma_{ni} (1 - \zeta_{ni}^{10}) [D_{ni}(x_0) - D_{ni}(x)]\right| \\ &+ \left| \sum_{0 \le k \le m} n^{-1/2} \sum_i \gamma_{ni} \zeta_{ni}^{2k} [D_{ni}(x_0) - D_{ni}(x)] \right|. \end{aligned} \tag{2.7}$$

However,

$$D_{ni}(x_0) - D_{ni}(x) = \sum_{0 \le r < k} [D_{ni}(x_r) - D_{ni}(x_{r+1})] + [D_{ni}(x_k) - D_{ni}(x)],$$

$$\sum_{r < k \le m} \zeta_{ni}^{2k} = \zeta_{ni}^{1r+1}, \qquad 1 \le i \le n.$$

Moreover, because of the monotonicity of the d.f. H and the indicator function, and because $x_k' < x \le x_k$, for all i and k,

$$|D_{ni}(x_k) - D_{ni}(x)| \le I(x_k' + \xi_{ni} < \eta_{ni} \le x_k + \xi_{ni}) + H(x_k + \xi_{ni}) - H(x_k' + \xi_{ni}).$$

Combining these observations with (2.7), we readily obtain that the second term in the right-hand side of (2.7) is bounded above by

$$\begin{aligned} &\left| \sum_{0 \le k \le m} n^{-1/2} \sum_{1 \le i \le n} \gamma_{ni} \zeta_{ni}^{2k} \sum_{0 \le r < k} [D_{ni}(x_r) - D_{ni}(x_{r+1})] \right| \\ &\qquad + \sum_{0 \le k \le m} n^{-1/2} \sum_{1 \le i \le n} |\gamma_{ni}| \zeta_{ni}^{2k} \left| D_{ni}(x_k) - D_{ni}(x) \right| \\ &\le \sum_{0 \le r < m} \left| n^{-1/2} \sum_i \gamma_{ni} \zeta_{ni}^{1r+1} [D_{ni}(x_r) - D_{ni}(x_{r+1})] \right| + \sum_{0 \le k \le m} R_{nk}(x). \end{aligned}$$

The last inequality is obtained by summing the first term over k first. Now combine this with (2.7) finally obtaining that, for all $x \in \mathbb{R}$,

$$\begin{aligned} |U_n(x_0) - U_n(x)| \le &\, 2n^{-1/2} \sum_i |\gamma_{ni}| I\left[\max_i |\gamma_{ni}| > a_0 n^{1/2}\right] \\ &+ \sum_{1 \le k \le m} |S_{nk}(x)| + \sum_{0 \le k \le m} R_{n,k}(x). \end{aligned} \tag{2.8}$$

The first term on the right-hand of (2.8) is obviously only positive on the set $[\max_i |\gamma_{ni}| > a_0 n^{1/2}]$. The second term is bounded in probability using (essentially) Freedman's lemma as is shown next. Notice first that the quadratic variation of $S_{nk}(x)$ is given by

$$Q_{nk}(x) := n^{-1} \sum_i \gamma_{ni}^2 \zeta_{ni}^{1k} |H(x_k + \xi_{ni}) - H(x_{k-1} + \xi_{ni})| \times \left(1 - |H(x_k + \xi_{ni}) - H(x_{k-1} + \xi_{ni})|\right),$$

which, on the set B_{nb}, is bounded above by

$$n^{-1} \sum_i \gamma_{n,i}^2 (\delta_k^2 + \delta_{k-1}^2) = 5\delta_k^2 n^{-1} \sum_i \gamma_{ni}^2.$$

Thus, $\forall\, x \in \mathbb{R}, 1 \le k \le m$, $B_{nb} \cap C_{n\eta} \subset B_{nb} \cap [Q_{nk}(x) \le 5\eta\delta_k^2]$. We now apply Freedman's lemma to see that

$$\begin{aligned} &P\left(\left[\sum_{1\le k\le m} |S_{nk}(x)| > 10\eta \sum_{1\le k\le m} \eta_k, \quad \text{for some } x \in \mathbb{R}\right] \cap B_{nb} \cap C_{n\eta}\right) \\ &\quad\le \sum_{1\le k\le m} (N(\delta_k, b) + 1)(N(\delta_{k-1}, b) + 1) \\ &\qquad\qquad \times \sup_x P\left([|S_{nk}(x)| > 10\eta\,\eta_k] \cap B_{nb} \cap C_{n\eta}\right) \\ &\quad\le \sum_{1\le k\le m} 2\exp\left\{2\ln(1 + N(\delta_k, b)) - 100\eta^2\,\eta_k^2/2[5\eta\delta_k^2 + 10\eta\,\eta_k a_k]\right\} \\ &\quad= 2\sum_{1\le k\le m} \exp\left\{2\ln(1 + N(\delta_k, b)) - 10\eta\, h_k^2/3\right\} \\ &\quad\le 2\sum_{1\le k\le m} \exp\{-\eta - 2L_k\} \le c\exp\{-\eta\}, \quad \forall\, \eta \ge 1, \end{aligned} \tag{2.9}$$

where again c is a fixed constant.

In order to handle the third term on the right-hand side of (2.8), we need to center it conditionally. Let $R_n(x)$ denote the conditional centering of $\Sigma_{0\le k\le m} R_{nk}(x)$, that is,

$$\begin{aligned} R_n(x) &= n^{-1/2} \sum_i E\left(\left(\sum_{0\le k\le m} |\gamma_{ni}| \zeta_{ni}^{2k} \{I[x_k' + \xi_{ni} < \eta_{ni} \le x_k + \xi_{ni}] \right.\right. \\ &\qquad\qquad \left.\left. + H(x_k + \xi_{ni} - H(x_k' + \xi_{ni}))\}\right) \Big| \mathcal{A}_{ni}\right) \\ &= 2n^{-1/2} \sum_i \sum_{0\le k\le m} |\gamma_{ni}| \zeta_{ni}^{2k} [H(x_k + \xi_{ni}) - H(x_k' + \xi_{ni})]. \end{aligned}$$

On the set $B_{nb} \cap C_{n\eta}$ using the definition of m,

$$
\begin{aligned}
R_n(x) &\le 2n^{-1/2} \sum_i \sum_{0 \le k \le m} |\gamma_{ni}| \zeta_{ni}^{2k} 2\delta_k^2 \\
&\le 4n^{-1} \sum_i \sum_{0 \le k < m} \gamma_{ni}^2 \zeta_{ni}^{2k} \delta_k^2 / a_{k+1} + 4n^{1/2}\Big(n^{-1} \sum_i |\gamma_{ni}|\Big) \delta_m^2 \\
&\le 2^4 n^{-1} \sum_i \sum_{0 \le k < m} \gamma_{ni}^2 \zeta_{ni}^{2k} \eta_{k+1} + 4n^{1/2}\Big(n^{-1} \sum_i \gamma_{ni}^2\Big)^{1/2} \delta_m^2 \\
&\le 2^4 \Bigg(\sum_{0 \le k < m} \eta_{k+1} \Bigg) n^{-1} \sum_i \gamma_{ni}^2 + 4n^{1/2} \eta^{1/2} \delta_m^2 \qquad (2.10) \\
&= 2^4 \Bigg\{ \Bigg(\sum_{0 \le k < m} \eta_{k+1} \Bigg) n^{-1} \sum_i \gamma_{ni}^2 + \eta\, \eta_0 \Bigg\} \\
&\le 2^4 \eta \sum_{0 \le k \le m} \eta_k.
\end{aligned}
$$

The quadratic variation of the conditionally centered $R_{nk}(x)$ is bounded, on $B_{nb} \cap C_{n\eta}$, by

$$
n^{-1} \sum_i \gamma_{ni}^2 \zeta_{ni}^{2k} \big[H(x_k + \xi_{n,i}) - H(x_k' + \xi_{n,i}) \big] \le 2\,\delta_k^2 n^{-1} \sum_i \gamma_{ni}^2 \le 2\,\delta_k^2 \eta.
$$

Thus

$$
\begin{aligned}
P&\Bigg(\Bigg[\sum_{0 \le k \le m} R_{nk}(x) > 22\eta \sum_{0 \le k \le m} \eta_k, \text{ for some } x \in \mathbb{R} \Bigg] \cap B_{nb} \cap C_{n\eta} \Bigg) \\
&\le P\Bigg(\Bigg[\sum_{0 \le k \le m} R_{nk}(x) - R_n(x) > 6\eta \sum_{0 \le k \le m} \eta_k, \text{ for some } x \in \mathbb{R} \Bigg] \\
&\qquad\qquad \cap B_{nb} \cap C_{n\eta} \Bigg) \qquad (2.11) \\
&\le \sum_{0 \le k \le m} N(\delta_k, b) \exp\big\{ -36\eta^2 \eta_k^2 / 2(6\eta\, \eta_k a_k + 2\delta_k^2 \eta) \big\} \\
&\le \sum_{0 \le k \le m} \exp\{ \ln\, N(\delta_k, b) - 2\eta\, h_k^2 \} \\
&\le \sum_{0 \le k \le m} \exp\{ -\eta - 2L_k \} \le c\, \exp\{-\eta\}, \qquad \eta \ge 1.
\end{aligned}
$$

Combining (2.8)–(2.11) gives the result. □

We immediately have the following results.

COROLLARY 2.4. *If* (A1)–(A4) *hold, and if* (A5) *holds for some* $b_0 \geq 0$, *then, for any* $\eta \geq 1, 0 \leq b \leq b_0$ *and* $\delta > 0$,

$$\limsup_n P\Big(\sup_x \big| U_n(x) - U_n(\pi_{\delta b}(x)) \big| > \{c_1 + c_2\eta\} \Big\{ \delta + \textstyle\int_0^\delta (1 + \ln\ N(u,b))^{1/2} du \Big\} \Big) \leq P(\gamma^2 > \eta) + c_3 \exp\{-\eta\}$$

where c_1, c_2 *and* c_3 *are universal constants.*

COROLLARY 2.5 (Eventual tightness of U_n^* in d_b metric). *If* (A1), (A2) *and* (A4) *hold, and if* $\mathcal{I}(b_0) < \infty$ *for some* $b_0 \geq 0$, *then, for any* $\eta \geq 1, 0 \leq b \leq b_0$ *and* $\delta > 0$,

$$\limsup_n P\Big(\sup_x \big| U_n^*(x) - U_n^*(\pi_{\delta b}(x)) \big| > \{c_1 + c_2\eta\} \times \Big\{ \delta + \int_0^\delta (1 + \ln\ N(u,b))^{1/2} du \Big\} \Big) \leq P(\gamma^2 > \eta) + c_3 \exp\{-\eta\}, \tag{2.12}$$

where c_1, c_2 *and* c_3 *are universal constants.*

REMARK 2.1. Because $N(\delta, 0) = [1/\delta^2]$, the entropy integral $\mathcal{I}(0)$ is a a priori finite. Take $b = 0$ in (2.12) to see immediately that U_n^* is eventually tight in the d_0-metric under assumptions (A1), (A2) and (A4).

PROOF OF (1.9). Fix $\varepsilon > 0$. Take $\eta \geq 1$ large enough to have $P(\gamma^2 > \eta) < \varepsilon$ and $c_3 \exp\{-\eta\} < \varepsilon$. Fix $b_0 > 0$ so that (A5) holds. Pick $\delta > 0$ small enough to have

$$\{c_1 + c_2\eta\} \Big\{ \delta + \int_0^\delta (1 + \ln\ N(u, b_0))^{1/2} du \Big\} < \varepsilon, \tag{2.13}$$

and choose $a > 0$ small enough to have both

$$\exp\Big\{ -\frac{\varepsilon^2}{2a(\varepsilon + \eta)} \Big\} \leq \frac{\varepsilon}{N(\delta, b_0)} \quad \text{and} \quad a \leq \frac{\delta}{(1 + \ln\ N(\delta, b_0))^{1/2}}. \tag{2.14}$$

Since $N(u,b)$ increases in b for u fixed, if (2.13) and (2.14) hold for b_0, they also hold for all $b \leq b_0$. Pick $b \in [0, b_0]$ sufficiently small to have

$$\sup_x [H(x+b) - H(x-b)] \leq a.$$

Let $\Lambda_n := A_{na} \cap B_{nb} \cap C_{n\eta}$. Then $\limsup_n P(\Lambda_n^c) < \varepsilon$ and, for any

$n > 16\eta(1 + \ln N(\delta, b))/\delta^2$,

$$\begin{aligned} P([\|U_n - U_n^*\|_\infty > 3\varepsilon] \cap \Lambda_n) &\le N(\delta, b) \sup_x P([|U_n(x) - U_n^*(x)| > \varepsilon] \cap \Lambda_n) \\ &\quad + P\Big(\Big[\sup_x |U_n(x) - U_n(\pi_{\delta b}(x))| > \varepsilon\Big] \cap \Lambda_n\Big) \\ &\quad + P\Big(\Big[\sup_x |U_n^*(x) - U_n^*(\pi_{\delta b}(x))| > \varepsilon\Big] \cap \Lambda_n\Big) \\ &\le N(\delta, b) \exp\{-\varepsilon^2/2a(\varepsilon + \eta)\} + 2c_3 \exp\{-\eta\} \le 3\varepsilon. \end{aligned}$$

Thus

$$\limsup_n P(\|U_n - U_n^*\|_\infty > 3\varepsilon) \le 4\varepsilon.$$

Now let $\varepsilon \to 0$ in this to conclude (1.9). Note that this proof did not use (A6). □

PROOF OF (1.10). This follows from (1.9), Lemma 2.1 and Corollary 2.1. This also completes the proof of Theorem 1.1. □

3. Proofs of Theorem 1.2, Lemma 1.2 and Corollaries 1.1–1.3. Recall Lemma 1.1. We also need the following preliminaries.

LEMMA 3.1. *Assume that* (1.1), (a1) *and* (h1)–(h3) *hold. Then the following hold*:

For every y at which F is continuous and for every $0 < b < \infty$,

$$\sup_{\|\mathbf{s}\| \le b} \big|\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s}) - W_h(y, \rho)\big| = o_p(1). \tag{3.1}$$

Consequently, for every y at which F posseses a density f that is continuous in a neighborhood of y, and for every $0 < b < \infty$,

$$\sup_{\|\mathbf{s}\| \le b} \Big| n^{1/2}[W_h(y, \rho + n^{-1/2}\mathbf{s}) - W_h(y, \rho)] - \mathbf{s}' n^{-1} \sum_i h(\mathbf{Y}_{i-1}) \mathbf{Y}_{i-1} f(y) \Big| = o_p(1). \tag{3.2}$$

PROOF. Corollary 2.3 applied to the entities given in (1.8) readily yield that, $\forall\, y \in \mathbb{R}, \forall s \in \mathbb{R}^p$, $|\mathcal{W}_h(y, \rho + n^{-1/2}\mathbf{s}) - \mathcal{W}_h(y, \rho)| = o_p(1)$. Uniformity in $\mathbf{s}$ is obtained as in Lemma 1.1. □

Next, define

$$\begin{aligned} F_n(y, \mathbf{t}) &:= n^{-1} \sum_i I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \le y), \qquad y \in \mathbb{R}, \\ T_j(u, \mathbf{t}) &:= n^{-1} \sum_i g_j(\mathbf{Y}_{i-1}) I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \le F^{-1}(u)), \\ &\qquad 1 \le j \le p,\ 0 \le u \le 1, \mathbf{t} \in \mathbb{R}^p. \end{aligned} \tag{3.3}$$

Observe that if in (1.6) $h(\mathbf{y}) \equiv 1$, then $W_h \equiv F_n$ and if $h(\mathbf{Y}_{i-1}) \equiv g_j(\mathbf{Y}_{i-1})$, then $W_h(F^{-1}, \cdot) = T_j$. Let $\mathbf{T_g} := (T_1, \dots, T_p)'$. Write $\mathbf{T}$ for $\mathbf{T_g}$ whenever $\mathbf{g}(\mathbf{y}) \equiv \mathbf{y}$. From Lemmas 1.1 and 3.1, we thus readily obtain the following.

COROLLARY 3.1. *Assume that* (1.1) *and* (a1) *hold.*
In addition, if (a2) *and* (F2) *hold, then,* $\forall\ 0 < b < \infty$,

$$\sup_{\|\mathbf{s}\| \le b} \left| n^{1/2}[F_n(0, \rho + n^{-1/2}\mathbf{s}) - F_n(0,\rho)] - \mathbf{s}' n^{-1} \sum_i \mathbf{Y}_{i-1} f(0) \right| = o_p(1). \tag{3.4}$$

In addition if, (a2) *and* (F1) *hold, then,* $\forall\ 0 < b < \infty$,

$$\sup_{y \in \mathbb{R},\, \|\mathbf{s}\| \le b} \left| n^{1/2}[F_n(y, \rho + n^{-1/2}\mathbf{s}) - F_n(y,\rho)] - \mathbf{s}' n^{-1} \sum_i \mathbf{Y}_{i-1} f(y) \right| = o_p(1). \tag{3.5}$$

In addition, if (a5) *with* $g_j(y) \equiv y$ *and* (F2) *hold, then,* $\forall 0 < b < \infty$,

$$\begin{aligned} \sup_{\|\mathbf{s}\| \le b} \Bigg\| n^{1/2} & \left[\mathbf{T}\left(\tfrac{1}{2}, \rho + n^{-1/2}\mathbf{s}\right) - \mathbf{T_g}\left(\tfrac{1}{2}, \rho\right)\right] \\ & - \mathbf{s}' n^{-1} \sum_i \mathbf{Y}_{i-1}\mathbf{Y}_{i-1}' f(0) \Bigg\| = o_p(1). \end{aligned} \tag{3.6}$$

In addition if (a2)–(a5) *and* (F1) *hold, then,* $\forall\ 0 < b < \infty$,

$$\begin{aligned} \sup_{0 \le u \le 1,\, \|\mathbf{s}\| \le b} \Bigg\| n^{1/2} & [\mathbf{T_g}(u, \rho + n^{-1/2}\mathbf{s}) - \mathbf{T_g}(u, \rho)] \\ & - n^{-1} \sum_i \mathbf{g}(\mathbf{Y}_{i-1}) \mathbf{Y}_{i-1}' \mathbf{s}\, q(u) \Bigg\| = o_p(1), \end{aligned} \tag{3.7}$$

where $q = f(F^{-1})$, $\mathbf{g}(\mathbf{Y}_{i-1}) \equiv (g_1(\mathbf{Y}_{i-1}), \dots, g_p(\mathbf{Y}_{i-1}))'$. $\square$

PROOF OF THEOREM 1.2. As mentioned in Remark 1.1, it suffices to prove (1.13). To that effect, let $F_{n\mathbf{s}}^{-1}(u) := \inf\{x; F_n(x, \rho + n^{-1/2}\mathbf{s}) \ge x\}$, and

$$\widetilde{\mathcal{Z}}_{\mathbf{g}}(u, \mathbf{s}) := n^{-1} \sum_i \mathbf{g}(\mathbf{Y}_{i-1}) I(\varepsilon_i \le F_{n\mathbf{s}}^{-1}(u) + n^{-1/2}\mathbf{s}'\mathbf{Y}_{i-1}), \qquad \mathbf{s} \in \mathbb{R}^p,\ 0 \le u \le 1.$$

Note that $\widetilde{\mathcal{Z}}_{\mathbf{g}}(u, \mathbf{s}) \equiv \mathbf{T_g}(FF_{n\mathbf{s}}^{-1}(u), \rho + n^{-1/2}\mathbf{s})$ and that

$$\begin{aligned} \sup_{0 \le u \le 1,\, \|\mathbf{s}\| \le b} & n^{1/2} \| \mathcal{Z}_{\mathbf{g}}(u, \rho + n^{-1/2}\mathbf{s}) - \mathcal{Z}_{\mathbf{g}}(u, \mathbf{s}) \| \\ & \le 2 \max_i n^{-1/2} \|\mathbf{g}(\mathbf{Y}_{i-1})\| = o_p(1), \quad \text{by (a3)}. \end{aligned}$$

Thus it suffices to prove (1.13) with $\mathcal{Z}_{\mathbf{g}}$ replaced by $\widetilde{\mathcal{Z}}_{\mathbf{g}}$ From (3.7), it readily follows that

$$n^{1/2} \widetilde{\mathcal{Z}}_{\mathbf{g}}(u, \mathbf{s}) = n^{1/2} \mathbf{T_g}(FF_{n\mathbf{s}}^{-1}(u), \rho) + \widehat{\Sigma}_n \mathbf{s}\, q(FF_{n\mathbf{s}}^{-1}(u)) + \bar{o}_p(1), \tag{3.8}$$

where $\widehat{\Sigma}_n := n^{-1}\Sigma_i \mathbf{g}(\mathbf{Y}_{i-1})\mathbf{Y}'_{i-1}$ and $\bar{o}_p(1)$ is a sequence of r.v.'s converging to zero uniformly in $0 \le u \le 1$, $\|\mathbf{s}\| \le b$, in probability. Now, let

$$\begin{aligned} \mu_{\mathbf{g}}(u) &:= n^{-1}\sum_i \mathbf{g}(\mathbf{Y}_{i-1})u, \\ \mu_{\mathbf{g}}(u,\mathbf{s}) &= n^{-1}\sum_i \mathbf{g}(\mathbf{Y}_{i-1})F(F^{-1}(u) + n^{-1/2}\mathbf{s}'\mathbf{Y}_{i-1}), \\ \mathbf{T}^*_{\mathbf{g}}(u) &:= n^{1/2}(\mathbf{T}_{\mathbf{g}}(u,\rho) - \mu_{\mathbf{g}}(u)), \qquad 0 \le u \le 1,\ \mathbf{s} \in \mathbb{R}^p. \end{aligned} \tag{3.9}$$

Note that $\mathbf{T}^*_{\mathbf{g}}$ is a vector of $U^*(F^{-1})$-processes. Corollary 2.5 applied with $\eta_{ni} \equiv F(\varepsilon_i)$ and the other entities given in (1.8) with $\mathbf{s} = \mathbf{0}$ readily implies, under (a1)–(a5) and (F1), that, $\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \ni$

$$\limsup_n P\bigg(\sup_{|u-v|<\delta} \|\mathbf{T}^*_{\mathbf{g}}(u) - \mathbf{T}^*_{\mathbf{g}}(v)\| > \varepsilon\bigg) < \varepsilon. \tag{3.10}$$

On the other hand, with $F_{n\mathbf{s}}(\cdot) \equiv F_n(\cdot, \rho + n^{-1/2}\mathbf{s})$,

$$FF^{-1}_{n\mathbf{s}}(u) - u = FF^{-1}_{n\mathbf{s}}(u) - F_{n\mathbf{s}}F^{-1}_{n\mathbf{s}}(u) + O(n^{-1}). \tag{3.11}$$

Hence, from (3.5), we obtain

$$\sup_{\|\mathbf{s}\| \le b,\, u \in [0,1]} |FF^{-1}_{n\mathbf{s}}(u) - u| = o_p(1). \tag{3.12}$$

Again, from (3.5), (3.11), (3.12), (a1) and (F1), we obtain

$$\begin{aligned} n^{1/2}[FF^{-1}_{n\mathbf{s}}(u) - u] &= -n^{1/2}[F_n(F^{-1}(u)) - u] \\ &\quad - \mathbf{s}'n^{-1}\sum_i \mathbf{Y}_{i-1}q(u) + \bar{o}_p(1), \end{aligned} \tag{3.13}$$

where F_n stands for $F_{n\rho}$, the empirical of $\{\varepsilon_i;\ 1 \le i \le n\}$.

Put (3.10) together with (a1)–(a5), (3.12) and (3.13) to conclude that

$$\begin{aligned} &n^{1/2}[\mathbf{T}_{\mathbf{g}}(FF^{-1}_{n\mathbf{s}}(u),\rho) - \mu_{\mathbf{g}}(u)] \\ &\quad = \mathbf{T}^*_{\mathbf{g}}(FF^{-1}_{n\mathbf{s}}(u)) + n^{-1}\sum_i \mathbf{g}(\mathbf{Y}_{i-1})n^{1/2}[F(F^{-1}_{n\mathbf{s}}(u)) - u] \\ &\quad = \mathbf{T}^*(u) + \bar{o}_p(1) + n^{-1}\sum_i \mathbf{g}(\mathbf{Y}_{i-1})\bigg\{ - n^{1/2}[F_n(F^{-1}(u)) - u] \\ &\qquad\qquad - n^{-1}\sum_j \mathbf{Y}'_{j-1}\mathbf{s}\, q(u)\bigg\} + \bar{o}_p(1) \\ &\quad = n^{-1/2}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \bar{\mathbf{g}})\Big\{I(\varepsilon_i \le F^{-1}(u)) - u\Big\} - \bar{\mathbf{g}}\,\overline{\mathbf{Y}}'\mathbf{s}\, q(u) + \bar{o}_p(1). \end{aligned}$$

This together with (3.8) and (F1), which implies the uniform continuity of q, yields that

$$\begin{aligned} n^{1/2}[\widetilde{\mathcal{Z}}_{\mathbf{g}}(u,\mathbf{s}) - \mu_{\mathbf{g}}(u)] &= n^{-1/2}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1} - \bar{\mathbf{g}})\{I(\varepsilon_i \le F^{-1}(u)) - u\} \\ &\quad + n^{-1}\sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \bar{\mathbf{g}})(\mathbf{Y}_{i-1} - \overline{\mathbf{Y}})'\mathbf{s}\, q(u) + \bar{o}_p(1), \end{aligned}$$

which, in view of (a6), proves (1.13). □

PROOF OF LEMMA 1.2. Recall, from Anderson (1971) or Brockwell and Davis (1987), that under (a7) and (a8), $\mathbf{Y}_0, X_1, X_2, \dots$ can be chosen to be stationary and ergodic. Hence the ergodic theorem and (a8) and (a9) together imply that (a1)–(a6) are a priori satisfied, $\boldsymbol{\Gamma}_{\mathbf{g}}$ exists and that there exists a constant vector $\mathbf{c}$ such that $\overline{\mathbf{g}} \to \mathbf{c}$ a.s. Now, rewrite

$$\widehat{\mathbf{S}}_{\mathbf{g}} = \widehat{\mathbf{S}}_{\mathbf{g}1} + (\overline{\mathbf{g}} - \mathbf{c}) n^{-1} \sum_i [\varphi(F(\varepsilon_i)) - \overline{\varphi}],$$

$$\widehat{\mathbf{S}}_{\mathbf{g}1} := n^{-1} \sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \mathbf{c}) [\varphi(F(\varepsilon_i)) - \overline{\varphi}].$$

By the C.L.T., $|n^{-1/2} \Sigma_i [\varphi(F(\varepsilon_i)) - \overline{\varphi}]| = O_p(1)$, so that $n^{1/2} \|\widehat{\mathbf{S}}_{\mathbf{g}} - \widehat{\mathbf{S}}_{\mathbf{g}1}\| = o_p(1)$. Note that $\widehat{\mathbf{S}}_{\mathbf{g}1}$ is a mean-zero square-integrable martingale. By Hall and Heyde (1980), Corollary 3.1, it follows that under (φ2) and (a7)–(a9), $n^{1/2} \widehat{\mathbf{S}}_{\mathbf{g}1} \Rightarrow N_p(\mathbf{0}, \sigma_\varphi^2 \boldsymbol{\Gamma}_{\mathbf{g}})$. This proves (1.17). □

PROOF OF COROLLARY 1.1. The proof of (1.18) uses the same argument as in Jaeckel (1972) with the following modification: Use (1.15) and (1.17) with $g_j(y) \equiv y$ wherever he uses Jurečková's theorem and the asymptotic normality of simple linear rank statistics in linear regression. Note that for these g_j's (a7) and (a8) imply (a1)–(a6), (a9) and the claim about the existence and positive definiteness of Σ [see, e.g., Anderson (1971), page 193].

To prove (1.19), note that in view of (φ1), $\widehat{\rho}_{\mathbf{g}} = \arg\min_{\mathbf{t}} \|\widetilde{\mathbf{S}}_{\mathbf{g}}(\mathbf{t})\|$, where

$$\widetilde{\mathbf{S}}_{\mathbf{g}}(\mathbf{t}) := n^{-1} \sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}}) \varphi(\widetilde{R}_{it}/(n+1)),$$

with $\widetilde{R}_{it} \equiv R_{it} \equiv$ rank$(X_i - \mathbf{t}' \mathbf{Y}_{i-1} - \overline{X} + \mathbf{t}' \overline{\mathbf{Y}})$ among $\{X_j - \mathbf{t}' \mathbf{Y}_{j-1} - \overline{X} + \mathbf{t}' \overline{\mathbf{Y}}, 1 \le j \le n\}$, $1 \le i \le n, \mathbf{t} \in \mathbb{R}^p$. From Hájek [(1969), Theorem II.7E] it follows, under (a10), that, for every unit vector $\boldsymbol{\theta} \in \mathbb{R}$, $\boldsymbol{\theta}' \widetilde{\mathbf{S}}_{\mathbf{g}}(\rho + n^{-1/2} r \boldsymbol{\theta})$ is a monotonic function of $r \in \mathbb{R}$, a.s. Use this and argue as in Koul [(1985) proof of Theorem 3.1], with the help of (1.15), to show first that $\|n^{1/2}(\widehat{\rho}_{\mathbf{g}} - \rho)\| = O_p(1)$. Then, again use (1.15) to conclude that $n^{1/2}(\widehat{\rho}_{\mathbf{g}} - \rho) = \boldsymbol{\Sigma}_{\mathbf{g}}^{-1} n^{1/2} \widehat{\mathbf{S}}_{\mathbf{g}} + o_p(1)$. This together with (1.17) now readily yields (1.19). □

PROOF OF COROLLARY 1.2. Let

$$\widehat{K}_{\mathbf{g}}(\mathbf{s}) := \int_0^1 \|n^{1/2} \widehat{\mathcal{Z}}_{\mathbf{g}}(u) + \boldsymbol{\Sigma}_{\mathbf{g}} q(u) \mathbf{s}\|^2 \, dG(u), \qquad \mathbf{s} \in \mathbb{R}^p.$$

Observe that Theorem 1.1 applied p times, jth to $\gamma_{ni} \equiv g_j(\mathbf{Y}_{i-1})$, $\eta_{ni} \equiv F(\varepsilon_i)$, $\xi_{ni} \equiv 0$, $1 \le j \le p$, implies that $\sup\{\|n^{1/2} \widehat{\mathcal{Z}}_{\mathbf{g}}(u)\|; 0 \le u \le 1\} = O_p(1)$. Assumptions (a7)–(a9) imply (A1), (A2), (A3) and (A6) in this case. This, together with

(1.13), implies that, $\forall\, 0 < b < \infty$,

$$\sup_{\|\mathbf{s}\| \le b,\, G \in C} |K_{\mathbf{g}}(\rho + n^{-1/2}\mathbf{s}) - \widehat{K}_{\mathbf{g}}(\mathbf{s})| = o_p(1) \quad \text{and that} \quad K_{\mathbf{g}}(\rho) = O_p(1).$$

Use this, (a10) and an argument like the one given in [Koul (1985), proof of Theorem 3.1] to conclude that

$$\begin{aligned} n^{1/2}(\widehat{\rho}_{\mathbf{g}} - \rho) &= \left\{ \int_0^1 q\, dG\, \Sigma_{\mathbf{g}} \right\}^{-1} \int_0^1 n^{1/2} \widehat{\mathcal{Z}}_{\mathbf{g}}(u)\, q(u)\, dG(u) + o_p(1) \\ &= -\left\{ \int_0^1 q\, dG\, \Sigma_{\mathbf{g}} \right\}^{-1} n^{-1/2} \sum_i (\mathbf{g}(\mathbf{Y}_{i-1}) - \overline{\mathbf{g}}) \\ &\qquad \times \{\psi(F(\varepsilon_i)) - E\psi(F(\varepsilon_i))\} + o_p(1), \end{aligned}$$

where $\psi(u) := \int_0^u q\, dG, 0 \le u \le 1$. Now (1.20) follows by arguing as for the proof of (1.17) . □

PROOF OF COROLLARY 1.3. Write $\mathbf{T}^*$ and μ for $\mathbf{T}^*_{\mathbf{g}}$ and $\mu^*_{\mathbf{g}}$ of (3.9) when $g_j(y) \equiv y$. Note that $n^{1/2}\mathbf{T}^*(\frac{1}{2}) = \mathcal{S}(\rho)$. Because of (F2), $F^{-1}(\frac{1}{2}) = 0$, so that

$$\mathcal{S}(\rho + n^{-1/2}\mathbf{s}) = n^{1/2}\{\mathbf{T}(1/2, \rho + n^{-1/2}\mathbf{s}) - \mu(1/2)\}.$$

Hence from (3.6) it follows that

$$\mathcal{S}(\rho + n^{-1/2}\mathbf{s}) = \mathcal{S}(\rho) + n^{-1} \sum_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1}\mathbf{s} f(0) + \bar{o}_p(1). \tag{3.14}$$

Assumptions (a7) and (a8) imply that (a1)–(a5) hold, $n^{-1}\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1} = \Sigma_1 + o_p(1)$ and by Hall and Heyde [(1980), Corollary 3.1] that $\mathcal{S}(\rho) \Rightarrow N(\mathbf{0}, \Sigma_1/4)$.

Next, note that, for every $\theta \in \mathbb{R}^p$, $\theta e' \mathcal{S}(\rho + n^{-1/2} r\mathbf{e})$ is a nonincreasing function of $r \in \mathbb{R}$. These facts together with an argument like the one that appears in the proof of Corollary 1.1 yield that $\|n^{1/2}(\widehat{\rho}_{\text{lad}} - \rho)\| = O_p(1)$. Hence from (3.14) it follows that $n^{1/2}(\widehat{\rho}_{\text{lad}} - \rho) \Rightarrow N(\mathbf{0}, \Sigma_1/2f^2(0))$. See Pollard (1991) for another proof in the case $p = 1$. □

REFERENCES

ANDERSEN, N. T. and DOBRIĆ, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab.* **15** 164–177.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

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

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

DUDLEY, R. M. (1984). A course on empirical processes. *Lecture Notes in Math.* **1097** 1–142. Springer, New York.

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.

GINÉ, E. and ZINN. J. (1986). Lecture notes on central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50–113. Springer, New York.

GREENWOOD, P. E. and OSSIANDER, M. (1991). A central limit theorem for evolving random fields. In *Selected Proceedings of the Sheffield Symposium on Applied Probability* (I. V. Basawa and R. L. Taylor, eds.) IMS, Hayward, CA.

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, New York.

HETTMANSPERGER, T. P. (1984). *Statistical Inference Based On Ranks*. Wiley, New York.

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

JAECKEL, L. A. (1972). Estimating regression coefficient by minimizing the dispersion of residuals. *Ann. Math. Statist.* **43** 1449–1458.

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

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

KOUL, H. L. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference* **29** 291–308.

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

KOUL, H. L. and SEN, P. K. (1991). Weak convergence of a weighted residual empirical process in autoregression. *Statist. Decisions* **9** 235–262.

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

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919.

POLLARD, D. (1991). Asymptotics of least absolute deviation regression estimator. *Econometric Theory* **7** 186–199.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48823

DEPARTMENT OF MATHEMATICS
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331