The Annals of Statistics
2008, Vol. 36, No. 5, 2453–2470
DOI: 10.1214/07-AOS543

RESIDUAL EMPIRICAL PROCESSES FOR LONG AND SHORT MEMORY TIME SERIES[1]

BY NGAI HANG CHAN AND SHIQING LING

Chinese University of Hong Kong and
Hong Kong University of Science and Technology

This paper studies the residual empirical process of long- and short-memory time series regression models and establishes its uniform expansion under a general framework. The results are applied to the stochastic regression models and unstable autoregressive models. For the long-memory noise, it is shown that the limit distribution of the Kolmogorov–Smirnov test statistic studied in Ho and Hsing [*Ann. Statist.* **24** (1996) 992–1024] does not hold when the stochastic regression model includes an unknown intercept or when the characteristic polynomial of the unstable autoregressive model has a unit root. To this end, two new statistics are proposed to test for the distribution of the long-memory noises of stochastic regression models and unstable autoregressive models.

1. Introduction. Let the time series $\{y_t\}$ be generated by the model

$$y_t = \beta' X_t + \varepsilon_t \quad \text{and} \quad \varepsilon_t = \sum_{i=0}^{\infty} a_i e_{t-i}, \tag{1.1}$$

where X_t's are a sequence of p-dimensional time series which are measurable with respect to $\mathcal{F}_{t-1} = \sigma\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ or independent of $\{\varepsilon_t\}$. The coefficients a_i satisfy $\sum_{i=1}^{\infty} a_i^2 < \infty$; $a_0 = 1$ and $a_k = k^{H-3/2} L_0(k)$ for some slowly varying function L_0 [see Feller (1971)] with $H < 1$; and $\{e_t\}$ is a sequence of i.i.d. mean zero random variables with $\sigma_e^2 = Ee_t^2 < \infty$. The process $\{\varepsilon_t\}$ exhibits a long-memory (short-memory) phenomenon when $H \in (1/2, 1)$ $(H < 1/2)$, which has been considerably studied in the literature; see, for example, Robinson (1995a, 1995b) and the references therein. When model (1.1) is used to construct forecasting intervals or value-at-risk (VaR), knowledge on the distribution function $F(x)$ of ε_t is of crucial importance. This motivates the study on testing of $F(x)$ and on related empirical processes of $\{\varepsilon_t\}$.

Received June 2006; revised August 2007.
[1]Supported in part by CERG from the Hong Kong Research Grants Council Grant Numbers: CUHK400305, CUHK400306, HKUST6022/05P, HKUST6428/06H and HKUST6016/07.
AMS 2000 subject classifications. Primary 62G30; secondary 62M10.
Key words and phrases. Empirical process, long-memory time series, residuals, unit root, weak convergence.

When $H \in (1/2, 1)$, Ho and Hsing (1996) established a strong expansion for the empirical process of $\{\varepsilon_t\}$ in (1.1). Specifically, let

$$K_n(x) = \frac{1}{\sigma_n} \sum_{t=1}^{n} [I(\varepsilon_t \le x) - F(x)], \tag{1.2}$$

where $I(\cdot)$ is the indicator function and $\sigma_n^2 = \mathrm{var}(\sum_{t=1}^n \varepsilon_t)$. They proved that

$$\sup_x \left| K_n(x) + \frac{1}{\sigma_n} F'(x) \sum_{t=1}^{n} \varepsilon_t \right| = o(1) \qquad \text{a.s.}, \tag{1.3}$$

$$\sigma_n^2 \sim \kappa(H) n^{2H} L_0^2(n) \quad \text{and} \quad \sigma_n^{-1} \sum_{t=1}^{n} \varepsilon_t \xrightarrow{\mathcal{L}} N(0, 1); \tag{1.4}$$

see also Taqqu (1975) and Hosking (1996). Herein, $\sup_x = \sup_{x \in R}$, $\kappa(H) = \int_0^\infty (x + x^2)^{H-3/2}\, dx$, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$ and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution as $n \to \infty$. By (1.3),

$$\left[\sup_x F'(x) \right]^{-1} \sup_x |K_n(x)| \xrightarrow{\mathcal{L}} |N(0, 1)|, \tag{1.5}$$

if $\sup_x |F'(x)| < \infty$. This is the Kolmogorov–Smirnov test statistic of Ho and Hsing (1996) for testing the distribution $F(x)$. Contrary to the standard weak convergence of the empirical process in the short-memory case, the result (1.5) is somewhat striking as $\sup_x |K_n(x)|$ does not converge to the maximum of a Brownian bridge as in the traditional case. Weak convergence of $\{K_n(x)\}$ was established in Dehling and Taqqu (1989) when $\{\varepsilon_t\}$ is a long-range dependent Gaussian process. Koul and Surgailis (1997) obtained some related results when $H \in (1/2, 1)$. Wu (2003) showed that (1.3) holds in probability under a weaker condition and a general setup and characterized the limit behavior of $K_n(x)$ when $H \le 1/2$; see also Ho and Hsing (1997).

Note that since $\{\varepsilon_t\}$ is unobservable in model (1.1), the Kolmogorov–Smirnov test has to be evaluated based on the residual process of $\{\varepsilon_t\}$. In this situation, a key issue of interest is to determine the validity of (1.5) for the Kolmogorov–Smirnov statistic when $\{\varepsilon_t\}$ is replaced by its corresponding residual process. Furthermore, when (1.5) becomes invalid, how can one test for the distribution of $\{\varepsilon_t\}$? These two issues have been studied extensively when $\{\varepsilon_t\}$ is i.i.d.; see Bai (1994, 1996, 2003), Ling (1998), Lee and Wei (1999), Koul (2002), Lee and Taniguchi (2005) and Koul and Ling (2006) for further discussions. But for model (1.1) and for the Kolmogorov–Smirnov statistic studied in Ho and Hsing (1996), these two important issues still remain unresolved. When $\beta' X_t$ is a constant and ε_t is an ARFIMA(p, d, q) model, the distribution of $\{\varepsilon_t\}$ can be determined by $\{e_t\}$ once the parameters of the ARFIMA model are estimated. In this case, it would be sufficient to test for the distribution of $\{e_t\}$, for which standard procedures for residuals from a model with i.i.d. noises, such as those given in Bai (1994) and Lee and Wei (1999), can be adopted. To study the general residual process of $\{\varepsilon_t\}$, however, substantially different arguments need to be employed which rely heavily on the results of Ho and Hsing (1996, 1997) and Wu (2003).

This paper first establishes a uniform expansion of the residual empirical process of $\{\varepsilon_t\}$ under a general framework. The result is used to study the stochastic regression model of Robinson and Hidalgo (1997) and the unstable AR model of Chan and Terrin (1995), Truong-Van and Larramendy (1996) and Wu (2006). It is shown that the test statistic (1.5) of Ho and Hsing (1996) is no longer valid when the stochastic regression model includes an unknown intercept or when the characteristic polynomial of the unstable AR model has a unit root. Our results not only encompass the long-memory $\{\varepsilon_t\}$, but also the short-memory $\{\varepsilon_t\}$. Furthermore, two new statistics are constructed to test the distribution of the long-memory noises in the stochastic regression model and the unstable AR model.

This paper is organized as follows. A general result is given in Section 2. The residual processes of stochastic regression and unstable time series are presented in Sections 3 and 4, respectively.

2. A general result.

Let $\hat{\beta}_n$ be an estimator of β in (1.1). Let $\hat{\varepsilon}_t = y_t - \hat{\beta}_n' X_t$ be the residual of model (1.1). Further, define the empirical process based on residuals $\{\hat{\varepsilon}_t\}$ by

$$\hat{K}_n^{\delta}(x) = \frac{1}{\sigma_n}\sum_{t=1}^{n}[I(\hat{\varepsilon}_t \le x) - F(x)].$$

For $H \in (1/2, 1)$, σ_n is given in (1.4). For $\sum_{j=0}^{\infty}|a_j| < \infty$, which implies $H \le 1/2$, Ho and Hsing (1997) show that $\sigma^2 \equiv \lim_{n\to\infty}\sigma_n^2/n$ exists and is finite; see also Wu (2003). Let G_0 be the common distribution of $\{e_t\}$. Write $\varepsilon_t = e_t + \xi_{t-1}$ and let $A_t(x) = G_0'(x - \xi_{t-1}) - E[G_0'(x - \xi_{t-1})]$, where $\xi_{t-1} = \sum_{i=1}^{\infty} a_i e_{t-i}$. Denote $\|\cdot\| = \mathrm{tr}(M'M)$ for some matrix or vector M. We need the following two assumptions.

ASSUMPTION 2.1. (a) $H < 1/2$ and $\sigma > 0$, or $H = 1/2$, $\sigma > 0$ and $\sum_{j=0}^{\infty}|a_j| < \infty$, or $1/2 < H < 1$, and (b) G_0 is three times differentiable with bounded, continuous and integrable derivatives such that $\int x^4\, dG_0(x) < \infty$.

ASSUMPTION 2.2. Let δ_n be a $p \times p$ constant matrix depending on n such that the following statements hold:

(a) $\delta_n^{-1}(\hat{\beta}_n - \beta) = O_p(1)$,
(b) $\sigma_n^{-1}\sum_{t=1}^{n} E\|\delta_n' X_t\| = O(1)$,
(c) $\sigma_n^{-1}\sum_{t=1}^{n} E\|\delta_n' X_t\|^2 = o(1)$,
(d) $\sigma_n^{-1}\sup_x \|\sum_{t=1}^{n} A_t(x)\delta_n' X_t\| = o_p(1)$.

Assumption 2.1(b) can be replaced by a general condition in Wu (2003). δ_n is the rate of convergence of $\hat{\beta}_n$. Assumptions 2.2(b) and (c) automatically hold if $\delta_n^{-1} = \sqrt{n} I_p$ and X_t is strictly stationary with $E\|X_t\|^2 < \infty$, where I_p is the $p \times p$ identity matrix. As will be seen in Sections 3 and 4, δ_n^{-1} may not always be equal to $\sqrt{n} I_p$. Assumptions 2.2(b)–(d) are sufficient for the remainder term in the following expansion to be negligible, although they may not be the weakest ones. We state a general result as follows.

THEOREM 2.1. *Assume that Assumption 2.1 and Assumption 2.2 hold. Then*

$$\sup_x |\hat{K}_n(x) - K_n(x) - R_n F'(x)| = o_p(1),$$

where $R_n = \sigma_n^{-1}(\hat{\beta}_n - \beta)' \sum_{t=1}^n X_t = O_p(1)$.

REMARK 2.1. According to this theorem, if $R_n = o_p(1)$, then $\sup_x |\hat{K}_n(x) - K_n(x)| = o_p(1)$ and, hence, $\sup_x |\hat{K}_n(x)|$ and $\sup_x |K_n(x)|$ have the same limit distribution. If $R_n \neq o_p(1)$, then the limit distribution of $\sup_x |\hat{K}_n(x)|$ may be different from that of $\sup_x |K_n(x)|$, as seen in Theorems 3.1 and 4.1. When $H \in (1/2, 1)$, $K_n(x)$ can be replaced by $-F'(x) \sum_{t=1}^n \varepsilon_t / \sigma_n$. When $H < 1/2$ with $EX_t = 0$ or when $H \in (1/2, 1)$, $\delta_n^{-1} = \sqrt{n} I_p$ and $\{X_t\}$ is strictly stationary, then $R_n = o_p(1)$.

REMARK 2.2. We require $\{a_k\}$ to have the form $k^{H-3/2} L_0(k)$ because we have to use the tightness condition of empirical processes of $\{\varepsilon_t\}$ of Ho and Hsing (1996) and Wu (2003) for $H \in (1/2, 1)$; and Theorem 3 and Corollary 2 of Wu (2003) for $H \leq 1/2$. Without this condition, Theorem 2.1 is still valid if $\sum_{i=0}^{\infty} |a_i| < \infty$ as long as the empirical process of $\{\varepsilon_t\}$ is tight on R.

PROOF OF THEOREM 2.1. Let $\hat{u}_n = \delta_n^{-1}(\hat{\beta}_n - \beta)$. Then $\hat{\varepsilon}_t = \varepsilon_t - \hat{u}_n' \delta_n' X_t$ and

$$\begin{aligned}\hat{K}_n(x) - K_n(x) &- \frac{1}{\sigma_n} \sum_{t=1}^n F'(x) \hat{u}_n' \delta_n' X_t \\ &= \frac{1}{\sigma_n} \sum_{t=1}^n [I(\varepsilon_t \leq x + \hat{u}_n' \delta_n' X_t) - I(\varepsilon_t \leq x) - F'(x) \hat{u}_n' \delta_n' X_t].\end{aligned}$$

To study the process $\hat{K}_n(x)$, consider the process

$$A_n(x, u) = \frac{1}{\sigma_n} \sum_{t=1}^n [I(\varepsilon_t \leq x + u' \delta_n' X_t) - I(\varepsilon_t \leq x) - u' F'(x) \delta_n' X_t]$$

for all $u \in R^p$ and $x \in R$. By Assumption 2.2(a), if we can show that

$$\sup_{u \in [-\Delta, \Delta]^p} \sup_x |A_n(x, u)| = o_p(1) \qquad \text{for every } \Delta \in (0, \infty), \tag{2.1}$$

then Theorem 2.1 is proved. Denote

$$Z_n(x,u)=\frac{1}{\sigma_n}\sum_{t=1}^{n}[I(\varepsilon_t\le x+u'\delta_n'X_t)-F(x+u'\delta_n'X_t)-I(\varepsilon_t\le x)+F(x)].$$

By the triangular inequality, $|A_n(x,u)|\le|Z_n(x,u)|+|H_n(x,u)|$, where

$$H_n(x,u)=\frac{1}{\sigma_n}\sum_{t=1}^{n}[F(x+u'\delta_n'X_t)-F(x)-u'\delta_n'X_tF'(x)].$$

Since $\sup_x|G_0''(x)|<\infty$, we have $\sup_x|F''(x)|<\infty$. Using this fact, Assumption 2.2(c) and the Taylor expansion, $\sup_{u\in[-\Delta,\Delta]^p}\sup_x|H_n(x,u)|=o_p(1)$. To prove (2.1), it is sufficient to show that the following equation holds:

$$\sup_{u\in[-\Delta,\Delta]^p}\sup_x|Z_n(x,u)|=o_p(1), \tag{2.2}$$

for every $\Delta>0$. For each $u\in R^p$ and $\lambda\in R$, let

$$\begin{aligned}\tilde{Z}_n(x,u,\lambda)=\frac{1}{\sigma_n}\sum_{t=1}^{n}\bigl[I\bigl(\varepsilon_t\le x+g_t(u,\lambda)\bigr)\\ -F\bigl(x+g_t(u,\lambda)\bigr)-I(\varepsilon_t\le x)+F(x)\bigr],\end{aligned} \tag{2.3}$$

where $g_t(u,\lambda)=u'\delta_n'X_t+\lambda\|\delta_n'X_t\|$. For every $\delta>0$, partition the rectangle $[-\Delta,\Delta]^p$ into m balls $\{C_1,\dots,C_m\}$ each with radius δ. Take one point in each C_r and denote it by u_r. For any $u\in C_r$, we have

$$|g_t(u,\lambda)-g_t(u_r,\lambda)|\le\|u-u_r\|\|\delta_n'X_t\|\le\delta\|\delta_n'X_t\|. \tag{2.4}$$

Thus, $g_t(u_r,\lambda-\delta)\le g_t(u,\lambda)\le g_t(u_r,\lambda+\delta)$. Note that $Z_n(x,u)=\tilde{Z}_n(x,u,0)$. By the monotonicity of the indicator function, we obtain that

$$Z_n(x,u)\le\tilde{Z}_n(x,u_r,\delta)+\frac{1}{\sigma_n}\sum_{t=1}^{n}\bigl[F\bigl(x+g_t(u_r,\delta)\bigr)-F\bigl(x+g_t(u,0)\bigr)\bigr] \tag{2.5}$$

and a reverse inequality holds when δ is replaced by $-\delta$. Since $\sup_x|G_0'(x)|<\infty$, we have $\sup_x|F'(x)|<\infty$. By the mean value theorem, when $u\in C_r$,

$$\begin{aligned}&\left|\frac{1}{\sigma_n}\sum_{t=1}^{n}\bigl[F\bigl(x+g_t(u_r,\pm\delta)\bigr)-F\bigl(x+g_t(u,0)\bigr)\bigr]\right|\\ &\quad\le\frac{\sup_x|F'(x)|}{\sigma_n}\sum_{t=1}^{n}|g_t(u_r,\pm\delta)-g_t(u,0)|\\ &\quad\le\frac{O(1)\delta}{\sigma_n}\sum_{t=1}^{n}\|\delta_n'X_t\|=O_p(\delta),\end{aligned} \tag{2.6}$$

where the last equality follows from Assumption 2.2(b) and the $O_p(1)$ holds uniformly for all $x \in \tilde{R}$, all $u \in C_r$ and all $r = 1, \ldots, m$.

Given any $\varepsilon > 0$ and $\eta > 0$, by (2.6), there exists a $\delta_{1\varepsilon} > 0$ such that

$$P\left\{\frac{1}{\sigma_n} \max_r \max_{u \in C_r} \sup_x \left|\sum_{t=1}^n \left[F\big(x + g_t(u_r, \pm\delta)\big) - F\big(x + g_t(u, 0)\big)\right]\right| \geq \frac{\varepsilon}{3}\right\} \leq \frac{\eta}{6},$$

when $\delta \leq \delta_{1\varepsilon}$ and $n \to \infty$. By Lemma A.3, there exists a $\delta_{2\varepsilon} > 0$ such that

$$\begin{aligned} P\left\{\max_r \sup_x |\tilde{Z}_n(x, u_r, \pm\delta)| \geq \frac{\varepsilon}{3}\right\} &\leq P\left\{\max_r J_{3n}(u_r, \pm\delta) \geq \frac{\varepsilon}{6}\right\} + P\left\{\delta J_{4n} \geq \frac{\varepsilon}{6}\right\} \\ &\leq m \max_r P\left\{J_{3n}(u_r, \pm\delta) \geq \frac{\varepsilon}{6}\right\} + \frac{\eta}{6} \leq \frac{\eta}{3}, \end{aligned}$$

when $\delta \leq \delta_{2\varepsilon}$ and $n \to \infty$ because m is an integer depending on δ but not depending on n. By the preceding two inequalities, when $\delta \leq \min\{\delta_{1\varepsilon}, \delta_{1\varepsilon}\}$,

$$\begin{aligned} &P\left\{\sup_{u \in [-\Delta, \Delta]^p} \sup_x |Z_n(x, u)| \geq \varepsilon\right\} \\ &\quad \leq P\left\{\max_r \sup_x |\tilde{Z}_n(x, u_r, \delta)| \geq \frac{\varepsilon}{3}\right\} + P\left\{\max_r \sup_x |\tilde{Z}_n(x, u_r, -\delta)| \geq \frac{\varepsilon}{3}\right\} \\ &\qquad + P\left\{\frac{1}{\sigma_n} \max_r \max_{u \in C_r} \sup_x \left|\sum_{t=1}^n \left[F\big(x + g_t(u_r, \pm\delta)\big) - F\big(x + g_t(u, 0)\big)\right]\right| \geq \frac{\varepsilon}{3}\right\} \\ &\quad \leq \eta, \qquad \text{when } n \to \infty, \text{ proving (2.2).} \end{aligned}$$

□

3. Residual empirical process of stochastic regression models. In this section we apply the results in Section 2 to the stochastic regression model of Robinson and Hidalgo (1997):

$$y_t = \alpha_0 + \alpha' x_t + \varepsilon_t, \tag{3.1}$$

where ε_t is defined in model (1.1), x_t is a q-dimension vector time series independent of $\{\varepsilon_t\}$, and $\beta = (\alpha_0, \alpha')'$ is a $p = q + 1$ dimensional unknown parameter vector. The least squares estimator (LSE) or generalized LSE of α is not asymptotically normal when both x_t and ε_t exhibit long-range dependence; see Robinson (1994). Robinson and Hidalgo (1997) proposed a class of weighted LSE which is $\sqrt{n}$-consistent and asymptotically normal.

Let $f(\lambda)$ be the spectral density of ε_t and $\phi(\lambda)$ be a real-valued, even and integrable periodic function with period 2π such that $\psi(\lambda) = \phi^2(\lambda) f(\lambda)$ is continuous. Denote $\phi_j = (2\pi)^{-2} \int_{-\pi}^{\pi} \phi(\lambda) \cos j\lambda \, d\lambda$. Robinson–Hidalgo's weighted LSE of α is defined as

$$\hat{\alpha}_n = \left[\sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x})' \phi_{t-s}\right]^{-1} \left[\sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(y_s - \bar{y}) \phi_{t-s}\right],$$

where $\bar{x} = \sum_{t=1}^{n} x_t/n$ and $\bar{y} = \sum_{t=1}^{n} y_t/n$. Let $\gamma_j = E(\varepsilon_t \varepsilon_{t+j})$ and $\kappa_{abcd}(s, u, v, w)$ be the fourth cumulant of x_{as}, x_{bu}, x_{cv} and x_{dw}, where x_{as} is the ath element of x_s. Recall the assumptions of Robinson and Hidalgo (1997) as follows.

ASSUMPTION 3.1. (a) $\sum_{j=0}^{\infty} \tilde{\phi}_j < \infty$ and $(\sum_{j=0}^{n} |\gamma_j| + n\tilde{\gamma}_n)[(\sum_{j=0}^{n} \tilde{\phi}_j^{1/2})^2 + n\Phi_n] = O(n)$ as $n \to \infty$, where $\tilde{\gamma}_a = \max_{j \ge a} |\gamma_j|$, $\tilde{\phi}_a = \max_{j \ge a} |\phi_j|$ and $\Phi_a = \sum_{|j|>a} |\phi_j|$.

(b) $\{x_t\}$ is fourth-order stationary, $\Gamma_u = E[(x_1 - Ex_1)(x_{1+|u|} - Ex_1)'] \to 0$ and $\max_{|v|,|w|<\infty} |\kappa_{abcd}(0, u, v, w)| \to 0$ as $|u| \to \infty$, $1 \le a, b, c, d \le q$.

(c) Σ_ψ is finite and Σ_ϕ and Σ_ψ are nonsingular, where $\Sigma_\chi = \int_{-\pi}^{\pi} \chi(\lambda)\, dH(\lambda)/(2\pi)$ and $H(\lambda)$ is the Hermitian matrix such that $\Gamma_j = \int_{-\pi}^{\pi} e^{ij\lambda}\, dH(\lambda)$.

Discussions on this assumption, the choice of ϕ and its computational procedures can be found in Robinson and Hidalgo (1997). Under Assumption 3.1, Robinson and Hidalgo (1997) showed that

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{\mathcal{L}} N(0, \Sigma_\phi^{-1} \Sigma_\psi \Sigma_\phi^{-1}). \tag{3.2}$$

The intercept term α_0 is estimated by

$$\hat{\alpha}_{0n} = \bar{y} - \hat{\alpha}_n' \bar{x} = \alpha_0 + \bar{\varepsilon} - (\hat{\alpha}_n - \alpha)' \bar{x},$$

where $\bar{\varepsilon} = \sum_{t=1}^{n} \varepsilon_t/n$. When $H \in (1/2, 1)$ or $H \le 1/2$ with $Ex_t = 0$, we see that $n\sigma_n^{-1}(\hat{\alpha}_n - \alpha)'\bar{x} = o_p(1)$ and hence, in these cases, we have

$$n\sigma_n^{-1}(\hat{\alpha}_{0n} - \alpha_0) \xrightarrow{\mathcal{L}} N(0, 1). \tag{3.3}$$

The results of Robinson and Hidalgo (1997) hold not only for long-memory $\{\varepsilon_t\}$ but also for short-memory $\{\varepsilon_t\}$. The following result entails the residual empirical process for both long- and short-memory cases.

THEOREM 3.1. *If Assumptions 2.1 and 3.1 hold, then the results of Theorem 2.1 hold with* $\hat{\beta}_n = (\hat{\alpha}_{0n}, \hat{\alpha}_n')'$, $\delta_n = \mathrm{diag}(\sigma_n n^{-1}, n^{-1/2} I_q)$ *and* $X_t = (1, x_t')'$.

PROOF. It is readily seen that Assumptions 2.2(a)–(c) hold. Note that

$$\frac{1}{\sigma_n} \sup_x \left\| \sum_{t=1}^{n} A_t(x) \delta_n' X_t \right\| \le \sup_x \left\| \frac{1}{n} \sum_{t=1}^{n} A_t(x) \right\| + \frac{1}{\sqrt{n}\sigma_n} \sup_x \left\| \sum_{t=1}^{n} A_t(x) x_t \right\|.$$

To check Assumption 2.2(d), we only need to show that

$$\sup_x \frac{1}{\sqrt{n}\sigma_n} \sup_x \left\| \sum_{t=1}^{n} A_t(x) x_t \right\| = o_p(1). \tag{3.4}$$

Similarly, it can be proved that $\sup_x |\sum_{t=1}^{n} A_t(x)|/n = o_p(1)$. Since $\sup_x |G_0''(x)| < \infty$ implies $\lim_{|x|\to\infty} G_0'(x) = 0$ [see Lee and Wei (1999)], we see that $E\sup_{|x|>M}\{G_0'(x-\xi_{t-1})\|x_t\|\} \to 0$ as $M \to \infty$. Since $\sqrt{n}/\sigma_n = O(1)$, for any given $\epsilon > 0$, there exists a constant $M > 0$ such that

$$\begin{aligned} &P\left(\sup_{|x|>M}\left\|\frac{1}{\sqrt{n}\sigma_n}\sum_{t=1}^{n} A_t(x)x_t\right\| > \eta\right) \\ &\qquad \le \frac{2\sqrt{n}}{\sigma_n\eta} E\sup_{|x|>M}\{G_0'(x-\xi_{t-1})\|x_t\|\} < \epsilon, \end{aligned} \tag{3.5}$$

uniformly in n. Partition $[-M, M]$ into $m = [4M\delta^{-1}]$ subintervals such that $-M = c_0 \le c_1 \le \cdots \le c_m = M$ with $c_{r+1} - c_r < \delta$ for any given constant $\delta > 0$. Let $U_{nr} = (\sqrt{n}\sigma_n)^{-1}\sum_{t=1}^{n} A_t(c_r)x_t$. When $H \in (1/2, 1)$, $\|U_{nr}\| \le 2n^{-1/2-H} \times \sum_{t=1}^{n}\|x_t\| = o_p(1)$. When $H \le 1/2$, since $A_t(c_r)$ and x_t are independent for each c_r, we can show that $U_{nr} = o_p(1)$. Thus, we have

$$\begin{aligned} &\sup_{|x|\le M}\left\|\frac{1}{\sqrt{n}\sigma_n}\sum_{t=1}^{n} A_t(x)x_t\right\| \\ &\quad \le \max_r \sup_{x\in[c_r,c_{r+1}]}\left\|\frac{1}{\sqrt{n}\sigma_n}\sum_{t=1}^{n}[A_t(x)-A_t(c_r)]x_t\right\| + \max_r\|U_{nr}\| \\ &\quad \le 2\delta\sup_x|G_0''(x)|O_p(1) + o_p(1) \\ &\quad = O_p(\delta) + o_p(1). \end{aligned} \tag{3.6}$$

Using (3.5)–(3.6), (3.4) is established. $\square$

We see that $R_n = O_p(1)$ and $K_n(x) = O_p(1)$. When $Ex_t = 0$, we have $R_n(x) = n\sigma_n^{-1}(\hat{\alpha}_{0n} - \alpha_0) \ne o_p(1)$ by virtue of (3.3). In this case, the estimated mean affects the limit distribution of $K_n(x)$ by Theorem 3.1. By (1.3) and (3.3), we have the following result.

COROLLARY 3.1. *If Assumptions 2.1 and 3.1 hold and $H \in (1/2, 1)$, then*

$$\left[\sup_x F'(x)\right]^{-1}\sup_x|\hat{K}_n(x)| \xrightarrow{\mathcal{L}} |N(0,4)|.$$

REMARK 3.1. This corollary gives a statistic for testing the distribution of the long-memory noises in model (3.1) when α_0 is unknown. The asymptotic variance of this test statistic is four times bigger than that in (1.5), which reflects the effects of the slower convergence rate of the estimated parameter $\hat{\alpha}_{0n}$. When α_0 is known, the test statistic (1.5) is still valid, however. As pointed out by the reviewer, when $F = F(x,\theta)$ involves an unknown parameter θ, one should consider $\hat{K}_n$ with $F(x)$ being replaced by $F(x,\hat{\theta}_n)$. Under such circumstances, the limit distribution of the statistic is usually different from that of Corollary 3.1. This fact serves as a reminiscence of the classical Kolmogorov–Smirnov statistics problem when the underlying parameters are estimated; see Durbin (1976). When $H \le 1/2$, it can be shown that the limit distribution of the statistic exists by means of the result of Wu (2003). The closed form of such a limit distribution is rather complicated and does not possess a simple expression, however, and is not presented here.

4. Residual empirical process of unstable AR(p) models. This section considers the unstable AR(p) model with starting value $\{y_0, y_{-1}, \dots, y_{-p+1}\}$ independent of $\{\varepsilon_s : s < 0\}$ such that

$$y_t = \beta' X_t + \varepsilon_t, \tag{4.1}$$

where $X_t = (y_{t-1}, \dots, y_{t-p})'$, $\beta = (\phi_1, \dots, \phi_p)'$, and the characteristic polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ has the decomposition,

$$\phi(z) = (1-z)^a (1+z)^b \prod_{k=1}^{l} [(1 - ze^{i\theta_k})(1 + ze^{i\theta_k})]^{d_k}, \tag{4.2}$$

$a, b, l, d_k, k = 1, \dots, l$, are nonnegative integers, $p = a + b + 2(d_1 + \cdots + d_l)$, and $\{\varepsilon_t\}$ is defined in model (1.1). Here, a denotes the multiplicity of the root $z = 1$ for $\phi(z) = 0$. Same interpretations are given to b and l. We estimate β by the LSE:

$$\hat{\beta}_n = \left(\sum_{t=1}^{n} X_t X_t' \right)^{-1} \sum_{t=1}^{n} X_t y_t.$$

For the special case with $\phi(z) = 1 - z$, Wu (2006) obtained the limiting distribution of $\hat{\beta}_n$ under Assumption 2.1(a); see also Sowell (1990) and Wang, Lin and Gulati (2003). For the general case, the limit distribution of $\hat{\beta}_n$ was obtained by Chan and Terrin (1995) and Truong-Van and Larramendy (1996) under the following Assumption 4.1(a) and (b), respectively. It can be seen that Assumption 2.1(a) is much weaker than Assumption 4.1.

ASSUMPTION 4.1. (a) $L_0(j) \sim c$, c is a constant, $H \in (1/2, 1)$ and $e_t \sim N(0, \sigma_e^2)$, or (b) $\sum_{j=0}^{\infty} j|a_j| < \infty$ and $\sigma > 0$.

Let $\delta_n = G' J_n^{-1}$, where G is the constant matrix given in Chan and Wei (1988) and $J_n = \operatorname{diag}(N_1, N_2, \dots, N_{l+2})$ with $N_1 = \operatorname{diag}(n, n^2, \dots, n^a)$, $N_2 = \operatorname{diag}(n, n^2, \dots, n^b)$ and $N_{k+2} = \operatorname{diag}(nI_2, \dots, n^{d_k} I_2)$, $k = 1, \dots, l$. Define $\xi_H(\tau) = [f_0(\tau), \dots, f_{a-1}(\tau)]'$, $f_0(\tau) = B_H(\tau)$ and $f_j(\tau) = \int_0^\tau f_{j-1}(s)\,ds$, $j = 1, \dots, a$, where $B_H(\tau)$ is a fractional Brownian motion with covariances

$$E[B_H(\tau) B_H(s)] = \tfrac{1}{2}\{s^{2H} + \tau^{2H} - |s - \tau|^{2H}\} \qquad \text{for } 0 \le s, \tau \le 1.$$

We now state the results for model (4.1).

THEOREM 4.1. *For model* (4.1), *if Assumption* 2.1 *holds with* $\phi(z) = 1 - z$, *or if Assumption* 4.1(a) *holds*, *or if Assumptions* 2.1(b) *and* 4.1(b) *hold*, *then the result of Theorem* 2.1 *holds with* $R_n = o_p(1)$ *for* $a = 0$ *and*

$$R_n \xrightarrow{\mathcal{L}} \begin{cases} (\Gamma + \zeta_{1/2})' \Omega_{1/2}^{-1} \int_0^1 \xi_{1/2}(\tau)\, d\tau, & \text{if } H \le 1/2, \\ \zeta_H' \Omega_H^{-1} \int_0^1 \xi_H(\tau)\, d\tau, & \text{if } H \in (1/2, 1), \end{cases}$$

for $a \ge 1$, *where* $\Gamma = (\gamma, 0, \dots, 0)'_{a \times 1}$, $\gamma = 1/2(1 - E\varepsilon_t^2/\sigma^2)$, $\zeta_H = \int_0^1 \xi_H(\tau)\, dB_H(\tau)$, $\Omega_H = (\omega_{ij})_{a \times a}$ *and* $\omega_{ij} = \int_0^1 f_i(\tau) f_j(\tau)\, d\tau$.

Let $D[0, 1]$ be the Skorokhod space and $D^p = D \times D \times \cdots \times D$ denote the p-Cartesian product space of $D = D[0, 1]$. To prove Theorem 4.1, we need the following lemma. Using the results in Chan and Wei (1988), Truong and Larramendy (1996) and Wu (2006), its proof is similar to that of Lemma 2.1 in Ling (1998) and the details are omitted.

LEMMA 4.1. *Let* $\tilde{\xi} = \xi_H$ *if* $H \in (1/2, 1)$ *and* $\tilde{\xi} = \xi_{1/2}$ *if* $H \le 1/2$. *If the assumptions of Theorem* 4.1 *hold*, *then*:

(a) $$\frac{1}{\sigma_n} \sum_{t=1}^{[n\tau]} \delta_n' X_t \xrightarrow{\mathcal{L}} \left(\int_0^\tau \tilde{\xi}'(s)\, ds, 0 \right)' \qquad \text{in } D^p, \text{ if } a \ge 1,$$

(b) $$\frac{1}{\sigma_n} \sum_{t=1}^{[n\tau]} \delta_n' X_t = o_p(1) \qquad \text{uniformly for all } \tau \in [0, 1] \text{ if } a = 0,$$

(c) $$\frac{1}{\sigma_n} \sum_{t=1}^{n} E\|\delta_n' X_t\| = O(1),$$

(d) $$\frac{n}{\sigma_n^2} \sum_{t=1}^{n} E\|\delta_n' X_t\|^2 = O(1).$$

PROOF. For simplicity, we only prove Theorem 4.1 for $\phi(z) = (1 - z)$, that is, model (4.1) only has one unit root. The general case can similarly be proved by Lemma 4.1. When $\phi(z) = (1 - z)$, $\delta_n = n^{-1}$ and $X_t = y_{t-1} = \sum_{i=1}^{t-1} \varepsilon_i$. By Theorem 6.1 of Chan and Terrin (1995) and Theorem 3.1 of Truong-Van and Larramendy (1996) or Theorems 3 and 4 of Wu (2006), Assumption 2.2(a) holds. By Lemma 4.1(c) and (d), we see that Assumption 2.2(b) and (c) holds.

We now consider Assumption 2.2(d). First, note that $E \sup_{|x|>M} A_t^2(x) \to 0$ as $M \to \infty$ and $\max_{1 \le t \le n} \sigma_n^{-2} E X_t^2 = O(1)$. Thus, for any given $\epsilon > 0$ and $\eta > 0$, there exists a constant $M > 0$ such that

$$
\begin{aligned}
&P\left(\sup_{|x|>M}\left|\frac{1}{n\sigma_n}\sum_{t=1}^{n} A_t(x)X_t\right| > \eta\right) \\
&\qquad \le \frac{\sqrt{E\sup_{|x|>M}|A_t(x)|^2}}{\eta n\sigma_n}\sum_{t=1}^{n}\sqrt{E|X_t|^2} < \epsilon,
\end{aligned} \tag{4.3}
$$

uniformly in n. Partition $[-M, M]$ into $m = [4M\delta^{-1}]$ subintervals such that $-M = x_0 \le x_1 \le \cdots \le x_m = M$ with $x_{r+1} - x_r < \delta$ for any given $\delta > 0$. Thus,

$$
\begin{aligned}
&\sup_{|x|\le M}\left|\frac{1}{n\sigma_n}\sum_{t=1}^{n} A_t(x)X_t\right| \\
&\quad \le \max_r \sup_{x_{r-1}\le x\le x_r}\left|\frac{1}{n\sigma_n}\sum_{t=1}^{n} A_t(x)X_t\right| \\
&\quad \le \max_r \sup_{x_{r-1}\le x\le x_r}\left|\frac{1}{n\sigma_n}\sum_{t=1}^{n} [A_t(x) - A_t(x_r)]X_t\right| \\
&\qquad + \max_r\left|\frac{1}{n\sigma_n}\sum_{t=1}^{n} A_t(x_r)X_t\right| = J_{1n} + J_{2n}, \qquad \text{say.}
\end{aligned} \tag{4.4}
$$

Since $\sup_x |A_t'(x)| < \infty$, by Lemma 4.1(c) and the Taylor expansion, we have

$$
J_{1n} \le O(\delta)\left[\frac{1}{n\sigma_n}\sum_{t=1}^{n}|X_t|\right] = O_p(\delta). \tag{4.5}
$$

For J_{2n}, we need the following decomposition:

$$
\begin{aligned}
\frac{1}{n\sigma_n}\sum_{t=1}^{n} A_t(x)X_t &= \frac{1}{n\sigma_n}\sum_{i=1}^{n}\left[\sum_{t=i+1}^{n} A_t(x)\right]\varepsilon_i \\
&= \frac{1}{n\sigma_n}\left(\sum_{i=1}^{n}\varepsilon_i\right)\left[\sum_{t=1}^{n} A_t(x)\right] - \frac{1}{n\sigma_n}\sum_{i=1}^{n}\left[\sum_{t=1}^{i} A_t(x)\right]\varepsilon_i \\
&= U_{1n}(x) - U_{2n}(x), \qquad \text{say.}
\end{aligned}
$$

By the ergodic theorem, $\sum_{t=1}^{n} A_t(x)/n = o_p(1)$ for each x. Furthermore, since $\sum_{i=1}^{n}\varepsilon_i/\sigma_n = O_p(1)$, we have $\max_r |U_{1n}(x_r)| = o_p(1)$ for a given $\delta > 0$.

We next consider $U_{2n}(x)$. When $H \le 1/2$, by Theorem 2 of Wu (2006), we know that $\sum_{t=1}^{[n\tau]} A_t(x)/\sigma_n \xrightarrow{\mathcal{L}} S(\tau)$ in D for each x and $\sum_{t=1}^{[n\tau]}\varepsilon_t/\sqrt{n} \xrightarrow{\mathcal{L}} \xi(\tau)$ in D, where $S(\tau)$ and $\xi(\tau)$ are standard Brownian motions. By Theorem 3.1 of Ling and Li (1998), $U_{2n}(x) = o_p(1)$ for each x and, hence, $\max_r |U_{2n}(x_r)| = o_p(1)$ for any given $\delta > 0$. Thus, Assumption 2.2(d) holds when $H \le 1/2$.

When $H \in (1/2, 1)$, we decompose $U_{2n}(x)$ as follows:

$$\frac{1}{n\sigma_n}\sum_{i=1}^{n}\left[\sum_{t=1}^{i} R_t(x)\right]\varepsilon_i + \frac{G_0''(x)}{n\sigma_n}\sum_{i=1}^{n}\left(\sum_{t=1}^{i}\xi_{t-1}\right)\varepsilon_i = U_{3n}(x) + U_{4n}(x), \tag{4.6}$$

say, where $R_t(x) = A_t(x) - G_0''(x)\xi_{t-1}$. For each x and any $\zeta > 0$, by Corollary 1 of Wu (2006) [see also Theorem 3.1 in Ho and Hsing (1997)], we have

$$E\left[\sum_{t=1}^{i} R_t(x)\right]^2 = O\big(i^{\max\{1,4(H-1/2)+2\zeta\}}\big). \tag{4.7}$$

By (4.7), for any $\eta > 0$ and $\delta > 0$, we have

$$\begin{aligned} P\Big(\max_r |U_{3n}(x_r)| > \eta\Big) &\le \frac{1}{\eta}\sum_{r=1}^{m} E|U_{3n}(x_r)| \\ &\le \frac{1}{\eta n\sigma_n}\sum_{r=1}^{m}\sum_{i=1}^{n}\left\{E\left[\sum_{t=1}^{i} R_t(x)\right]^2 E\varepsilon_i^2\right\}^{1/2} \\ &= O(n^{-\gamma}L_0^{-1}(n)) \to 0, \end{aligned} \tag{4.8}$$

when $n \to \infty$, where $\gamma = \min\{H - 1/2, 1 - H - \zeta\} > 0$. Note that

$$U_{4n}(x) = -\frac{G_0''(x)}{n\sigma_n}\sum_{i=1}^{n}\left(\sum_{t=1}^{i}\varepsilon_t\right)\varepsilon_i + \frac{G_0''(x)}{n\sigma_n}\sum_{i=1}^{n}\left(\sum_{t=1}^{i} e_t\right)\varepsilon_i.$$

By Theorems 3.2 and 3.3 of Chan and Terrin (1995) or Theorem 3 of Wu (2006),

$$\sum_{i=1}^{n}\left(\frac{1}{\sigma_n}\sum_{t=1}^{i}\varepsilon_t\right)\frac{\varepsilon_i}{\sigma_n} \xrightarrow{\mathcal{L}} \int_0^1 B_H(s)\,dB_H(s).$$

Thus, the first term in $U_{4n}(x)$ is $o_p(1)$ uniformly in $x \in R$. Note that $\sum_{t=1}^{n}|\varepsilon_t|/n = O_p(1)$ by the ergodic theorem and $\max_{1\le i\le n}|\sum_{t=1}^{i} e_t|/\sqrt{n} \xrightarrow{\mathcal{L}} \max_{0\le\tau\le1}|B_{1/2}(\tau)|$. Since $\sqrt{n}/\sigma_n = O(n^{-H+1/2}/L_0(n)) = o(1)$, the second term in $U_{4n}(x)$ is $o_p(1)$ uniformly in $x \in R$. Thus, we have $\max_x |U_{4n}(x)| = o_p(1)$. Furthermore, by (4.6) and (4.8), $\max_r |U_{2n}(x_r)| = o_p(1)$ for any given δ when $H \in (1/2, 1)$. Thus, Assumption 2.2(d) holds when $H \in (1/2, 1)$. $\square$

REMARK 4.1. From this theorem, we see that the empirical process of $\{\varepsilon_t\}$ is not affected if $\{\varepsilon_t\}$ is replaced by $\{\hat{\varepsilon}_t\}$ when $\phi(z)$ does not have a root equaling one. It has a profound effect when $\phi(z)$ has a unit root, however. In particular, using Theorem 3 of Wu (2006), we have the following corollary.

COROLLARY 4.1. *If $\phi(z) = (1 - z)$ and Assumption 2.1 holds with $H \in (1/2, 1)$, then it follows that*

$$\Big[\sup_x F'(x)\Big]^{-1} \sup_x |\hat{K}_n(x)| \xrightarrow{\mathcal{L}} \left| B_H(1) + \left[\int_0^1 B_H(\tau)\, dB_H(\tau)\right]\left[\int_0^1 B_H(\tau)\, d\tau\right]\left[\int_0^1 B_H^2(\tau)\, d\tau\right]^{-1}\right|.$$

REMARK 4.2. Corollary 4.1 gives the limit distribution of the Kolmogorov–Smirnov statistic. It can be used to test for the distribution of the long-memory noises in model (4.1). For instance, using $\hat{\varepsilon}_t$ as a proxy for ε_t, H may be estimated by Robinson's (1995a) semiparametric method. Although the asymptotic validity of such a procedure still needs to be examined, for a given $H \in (1/2, 1)$, the percentiles of the limit distribution can be tabulated by means of simulations. Corollary 4.1 thus provides a means to apply the Kolmogorov–Smirnov statistics to model (4.1).

APPENDIX: TECHNICAL LEMMAS

Let $x_r = r\epsilon\sigma_n^{-1}$ for any $r \in Z$ and some $\epsilon > 0$ and decompose the real line R as $R = \bigcup_{r\in Z}[x_r, x_{r+1}]$. Let $g_t(u, \lambda)$ be defined in (2.3) and

$$a_{nt}(x) = I\big(\varepsilon_t \le x + g_t(u, \lambda)\big) - F_{t-1}(x) - I(\varepsilon_t \le x) + G_0(x - \xi_{t-1}),$$

where $F_{t-1}(x) = E[I(e_t \le x - \xi_{t-1} + g_t(u, \lambda))|\mathcal{F}_{t-1}] = G_0[x - \xi_{t-1} + g_t(u, \lambda)]$, $u \in [-\Delta, \Delta]^p$ with $\Delta > 0$ and $\lambda \in [-1, 1]$. We have the following lemma.

LEMMA A.1. *Let $\tilde{Z}_{1n}(x, u, \lambda) = \sum_{t=1}^n a_{nt}(x)/\sigma_n$. For every u and λ, if Assumption 2.1 and Assumptions 2.2*(b) *and* (c) *hold, then*:

(a) $$\max_r \max_{x\in[x_r, x_{r+1}]} \frac{1}{\sigma_n}\sum_{t=1}^n \big|F\big(x_{r+1} + g_t(u, \lambda)\big) - F\big(x + g_t(u, \lambda)\big)\big| = O_p(\epsilon),$$

(b) $\sup_r |\tilde{Z}_{1n}(x_r, u, \lambda)| = o_p(1)$ *for any given $\epsilon > 0$.*

PROOF. By Assumption 2.1(b), $F'(x)$ exists and is bounded; see Ho and Hsing (1996). Since $n/\sigma_n^2 = O(1)$, by the Taylor expansion, part (a) holds.

For part (b), since $\sum_{t=1}^n a_{nt}(x)$ is a martingale array with respect to $\mathcal{F}_n = \sigma\{(e_t, X_t), t \le n\}$, by the Rosenthal inequality [see page 23 of Hall and Heyde (1980)],

$$\begin{aligned} E\left[\sum_{t=1}^n a_{nt}(x)\right]^4 &\le cE\left\{\sum_{t=1}^n E[a_{nt}^2(x)|\mathcal{F}_{t-1}]\right\}^2 + c\sum_{t=1}^n E[a_{nt}^4(x)] \\ &\le cn\sum_{t=1}^n E\{E[a_{nt}^2(x)|\mathcal{F}_{t-1}]\}^2 + 2c\sum_{t=1}^n E[a_{nt}^2(x)] \end{aligned} \tag{A.1}$$

for some constant c, where we use $a_{nt}^4(x) \le 2a_{nt}^2(x)$. Denote $g_t(u,\lambda)$ by g_t and let $H_t^{\pm}(x) = G_0(x - \xi_{t-1} \pm |g_t|)$. Since $E[I(e_t \le x - \xi_{t-1})|\mathcal{F}_{t-1}] = G_0(x - \xi_{t-1})$ and $G_0(x)$ is nondecreasing, we have

$$E[a_{nt}^2(x)|\mathcal{F}_{t-1}] \le |F_{t-1}(x) - G_0(x - \xi_{t-1})| \le H_t^+(x) - H_t^-(x).$$

Again, since $G_0(x)$ is nondecreasing, for any positive integer M, we have

$$\begin{aligned}
&\sum_{r=-M}^{M} E[H_t^+(x_r) - H_t^-(x_r)] \\
&\qquad \le \frac{\sigma_n}{\epsilon} \sum_{r=-M}^{M} E\left[\int_{x_r}^{x_{r+1}} H_t^+(x)\,dx - \int_{x_{r-1}}^{x_r} H_t^-(x)\,dx\right] \\
&\qquad = \frac{\sigma_n}{\epsilon} E\left\{\int_{x_M}^{x_{M+1}} H_t^+(x)\,dx + \int_{x_{-M-1}}^{x_{-M}} H_t^-(x)\,dx\right. \\
&\qquad\qquad\qquad \left. + \int_{x_{-M}}^{x_M} [H_t^+(x) - H_t^-(x)]\,dx\right\} \\
&\qquad \le 2 + \frac{\sigma_n}{\epsilon} E\left\{\int_{x_{-M}}^{x_M} \int_{-|g_t|}^{|g_t|} G_0'(x - \xi_{t-1} + y)\,dy\,dx\right\} \\
&\qquad \le 2 + \frac{\sigma_n}{\epsilon} E\left\{\int_{-|g_t|}^{|g_t|} \int_{-\infty}^{\infty} G_0'(x - \xi_{t-1} + y)\,dx\,dy\right\} \\
&\qquad = 2 + \frac{2\sigma_n}{\epsilon} E|g_t|.
\end{aligned} \tag{A.2}$$

Similarly, we have

$$\begin{aligned}
\sum_{r=-M}^{M} E[H_t^+(x_r) - H_t^-(x_r)]^2 &\le c \sum_{r=-M}^{M} E\{|g_t|[H_t^+(x_r) - H_t^-(x_r)]\} \\
&= 2cE|g_t| + \frac{2c\sigma_n}{\epsilon} Eg_t^2,
\end{aligned} \tag{A.3}$$

where $c = 2\sup_x G_0'(x)$. Using (A.2)–(A.3) and Assumptions 2.2(b)–(c),

$$\begin{aligned}
\frac{1}{\sigma_n^4} \sum_r \sum_{t=1}^{n} E[a_{nt}^2(x_r)] &\le \frac{1}{\sigma_n^4} \lim_{M\to\infty} \sum_{r=-M}^{M} \sum_{t=1}^{n} E[H_t^+(x_r) - H_t^-(x_r)] \\
&\le \frac{2n}{\sigma_n^4} + \frac{2}{\epsilon\sigma_n^3} \sum_{t=1}^{n} E|g_t| = o(1),
\end{aligned} \tag{A.4}$$

$$\frac{n}{\sigma_n^4} \sum_r \sum_{t=1}^{n} E\{E[a_{nt}^2(x_r)|\mathcal{F}_{t-1}]\}^2 \le \frac{2n}{\sigma_n^4} \sum_{t=1}^{n} E|g_t| + \frac{2}{\epsilon\sigma_n^3} \sum_{t=1}^{n} Eg_t^2 = o(1), \tag{A.5}$$

as $n/\sigma_n^2 = O(1)$. By the Markov inequality, (A.1), (A.4) and (A.5),

$$\begin{aligned}
P\Big(\sup_r |\tilde{Z}_{1n}(x_r, u, \lambda)| \geq \eta\Big) &\leq \sum_r P\big(|\tilde{Z}_{1n}(x_r, u, \lambda)| \geq \eta\big) \\
&\leq \frac{1}{\eta^4 \sigma_n^4} \sum_r E\left[\sum_{t=1}^n a_{nt}(x_r)\right]^4 \\
&= o(1),
\end{aligned}$$

as $n \to \infty$, for any given $\epsilon > 0$. Thus, part (b) is proved. □

LEMMA A.2. *Let* $\tilde{Z}_{2n}(x, u, \lambda) = \sum_{t=1}^n [F_{t-1}(x) - G_0(x - \xi_{t-1}) - F(x + g_t(u, \lambda)) + F(x)]/\sigma_n$. *If Assumptions* 2.1 *and* 2.2(b)–(d) *hold, then* $\tilde{Z}_{2n}(x, u, \lambda) = \lambda J_{1n}(x) + J_{2n}(x, u, \lambda)$ *such that* $\sup_x |J_{1n}(x)| = O_p(1)$ *and* $\sup_x \sup_u \sup_\lambda |J_{2n}(u, x, \lambda)| = o_p(1)$.

PROOF. By Assumption 2.1(b) and Lemma 6.2 of Ho and Hsing (1996), $F''(x)$ exists and is bounded. By the Taylor expansion and Assumption 2.2(c),

$$\begin{aligned}
\tilde{Z}_{2n}(x, u, \lambda) &= \frac{1}{\sigma_n} \sum_{t=1}^n \left\{ A_t(x) g_t(u, \lambda) + \frac{1}{2} g_t^2(u, \lambda)[G_0''(\xi_{t-1}^*) - F''(\tilde{\xi}_{t-1}^*)] \right\} \\
&= \frac{1}{\sigma_n} \sum_{t=1}^n A_t(x) g_t(u, \lambda) + o_p(1) \\
&= \frac{\lambda}{\sigma_n} \sum_{t=1}^n A_t(x) \|\delta_n' X_t\| + \left[\frac{u}{\sigma_n} \sum_{t=1}^n A_t(x) \delta_n' X_t + o_p(1) \right] \\
&= \lambda J_{1n}(x) + J_{2n}(x, u, \lambda), \qquad \text{say,}
\end{aligned}$$

where we use $F'(x) = E G_0'(x - \xi_{t-1})$, $\xi_{t-1}^* = x - \xi_{t-1} + \theta g_t(u, \lambda)$ and $\tilde{\xi}_{t-1}^* = x + \tilde{\theta} g_t(u, \lambda)$ with $\theta, \tilde{\theta} \in (0, 1)$ and $o_p(1)$ being held uniformly in x, u, λ. Since $\sup_x |A_t(x)| \leq 2$, by Assumption 2.2(b), $\sup_x |J_{1n}(x)| = O_p(1)$. Since $u \in [-\Delta, \Delta]^p$, by Assumption 2.2(d), $\sup_x \sup_u \sup_\lambda |J_{2n}(x, u, \lambda)| = o_p(1)$. The desired conclusion follows. □

LEMMA A.3. *If Assumptions* 2.1 *and* 2.2(b)–(d) *hold, then it follows that*

$$\sup_x |\tilde{Z}_n(x, u, \lambda)| \leq J_{3n}(u, \lambda) + |\lambda| J_{4n},$$

where $\tilde{Z}_n(x, u, \lambda)$ *is defined in* (2.3), $0 < J_{3n}(u, \lambda) = o_p(1)$ *for each* u *and* λ, *and* $0 < J_{4n} = O_p(1)$ *is independent of* u.

PROOF. Since $I(\varepsilon_t \le x)$ and $F(x)$ are nondecreasing, for any $x \in [x_r, x_{r+1}]$,

$$\tilde{Z}_n(x,u,\lambda) \le \tilde{Z}_n(x_{r+1},u,\lambda) + \frac{1}{\sigma_n}\sum_{t=1}^{n}[F(x_{r+1}+g_t) - F(x+g_t)] + \frac{1}{\sigma_n}\sum_{t=1}^{n}[I(\varepsilon_t \le x_{r+1}) - F(x_{r+1}) - I(\varepsilon_t \le x) + F(x)],$$

where g_t denotes $g_t(u,\lambda)$ and a reverse inequality holds when x_{r+1} is replaced by x_r. Since $|\tilde{Z}_n(x_{r+1},u,\lambda)| \le |\tilde{Z}_{1n}(x_{r+1},u,\lambda)| + |\tilde{Z}_{2n}(x_{r+1},u,\lambda)|$, we have

$$\sup_x |\tilde{Z}_n(x,u,\lambda)| \le \max_r |\tilde{Z}_{2n}(x_r,u,\lambda)| + R_n(u,\lambda),$$

where

$$\begin{aligned} R_n(u,\lambda) = {} & \max_r |\tilde{Z}_{1n}(x_r,u,\lambda)| \\ & + \max_r \max_{x\in[x_r,x_{r+1}]} \frac{1}{\sigma_n}\sum_{t=1}^{n}|F(x_{r+1}+g_t) - F(x+g_t)| \\ & + \sup_{|x_1-x_2|\le\epsilon\sigma_n^{-1}} \frac{1}{\sigma_n}\Bigg|\sum_{t=1}^{n}[I(\varepsilon_t \le x_1) \\ & \qquad - F(x_1) - I(\varepsilon_t \le x_2) + F(x_2)]\Bigg|. \end{aligned} \tag{A.6}$$

For any $\varepsilon, \eta > 0$, by Lemma 4.1(a), we can take ϵ small enough such that the second term of (A.6) is less than η happens with probability being at least $1 - \varepsilon/4$. For this ϵ, the first term of (A.6) is $o_p(1)$ by Lemmas A.1(b), and the last term of (A.6) is $o_p(1)$ by the tightness of the empirical process of $\{\varepsilon_t\}$ of Ho and Hsing (1996) and Wu (2003). Thus, $R_n(u,\lambda) = o_p(1)$ for each u and λ. By virtue of Lemma A.2, the conclusion holds. □

Acknowledgments. The authors would like to thank two referees, an Associate Editor and the Co-editor, Professor M. L. Eaton, for their helpful and constructive comments, which substantially improved the presentation of this paper.

REFERENCES

BAI, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* **22** 2051–2061. MR1329182

BAI, J. (1996). Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica* **64** 597–622. MR1385559

BAI, J. (2003). Testing parametric conditional distributions of dynamic models. *Rev. Econom. Statist.* **85** 531–549.

CHAN, N. H. and TERRIN, N. (1995). Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Ann. Statist.* **23** 1662–1683. MR1370302

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312

DURBIN, J. (1976). Kolmogorov–Smirnov tests when parameters are estimated. In *Empirical Distributions and Processes* (P. Gaenssler and P. Révész, eds.). *Lecture Notes in Math.* **566** 33–44. Springer, Berlin. MR0436448

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*. **2**, 2nd ed. Wiley, New York.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic, San Diego. MR0624435

HO, H.-C. and HSING, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

HO, H.-C. and HSING, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636–1669. MR1487431

HOSKING, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econometrics* **73** 261–284. MR1410007

KOUL, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. *Lecture Notes in Statist.* **166**. Springer, New York. MR1911855

KOUL, H. L. and LING, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.* **34** 994–1012. MR2283401

KOUL, H. L. and SURGAILIS, D. (1997). Asymptotic expansion of M-estimators with long-memory errors. *Ann. Statist.* **25** 818–850. MR1439325

LEE, S. and TANIGUCHI, M. (2005). Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. *Statist. Sinica* **15** 215–234. MR2125729

LEE, S. and WEI, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.* **27** 237–261. MR1701109

LING, S. (1998). Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.* **26** 741–754. MR1626028

LING, S. and LI, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *Ann. Statist.* **26** 84–125. MR1611800

ROBINSON, P. M. (1994). Time series with strong dependence. In *Advances in Econometrics: Sixth World Congress* (C. Sims, ed.). Cambridge Univ. Press. MR1278267

ROBINSON, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301

ROBINSON, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214

ROBINSON, P. M. and HIDALGO, F. J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77–104. MR1429918

SOWELL, F. (1990). Fractional unit root distribution. *Econometrica* **50** 495–505. MR1046932

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329

TRUONG-VAN, B. and LARRAMENDY, I. (1996). Asymptotic distribution of least squares estimators for purely unstable ARMA(m, ∞). *Statistics* **28** 307–346. MR1430066

WANG, Q., LIN, Y. X. and GULATI, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* **19** 143–164. MR1965845

WU, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831. MR2047687

WU, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory* **22** 1–14. MR2212691

DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG,
SHATIN, NT
HONG KONG

DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE
AND TECHNOLOGY
HONG KONG
E-MAIL: maling@ust.hk