Statistica Sinica **18**(2008), 313-333

EMPIRICAL PROCESSES OF STATIONARY SEQUENCES

Wei Biao Wu

University of Chicago

Abstract: The paper considers empirical distribution functions of stationary causal processes. Weak convergence of normalized empirical distribution functions to Gaussian processes is established and sample path properties are discussed. The Chibisov-O'Reilly Theorem is generalized to dependent random variables. The proposed dependence structure is related to the sensitivity measure, a quantity appearing in the prediction theory of stochastic processes.

Key words and phrases: Empirical process, Gaussian process, Hardy inequality, linear process, martingale, maximal inequality, nonlinear time series, prediction, short-range dependence, tightness, weak convergence.

1. Introduction

Let ε_k, $k \in \mathbb{Z}$, be independent and identically distributed (i.i.d.) random variables and define

$$X_n = J(\ldots, \varepsilon_{n-1}, \varepsilon_n), \tag{1}$$

where J is a measurable function such that X_n is a proper random variable. The framework (1) is general enough to include many interesting and important examples. Prominent ones are linear processes and nonlinear time series arising from iterated random functions. Given the sample $X_i, 1 \le i \le n$, we are interested in the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \le x}. \tag{2}$$

When the X_i are i.i.d., the weak convergence of F_n and its sample path properties have been extensively studied (Shorack and Wellner (1986)). Various generalizations have been made to dependent random variables. It is a challenging problem to develop a weak convergence theory for the associated empirical processes without the independence assumption. One way out is to impose strong mixing conditions to ensure asymptotic independence; see Billingsley (1968), Gastwirth and Rubin (1975), Withers (1975), Mehra and Rao (1975), Doukhan, Massart and Rio (1995), Andrews and Pollard (1994), Shao and Yu (1996) and Rio (2000), among others. Other special processes that have been studied include linear processes and Gaussian processes; see Dehling and Taqqu (1989), Csörgő and Mielniczuk (1996), Ho and Hsing (1996) and Wu (2003). A collection of recent results is edited by Dehling, Mikosch and Sørensen (2002).

We now introduce some notation. Let the triple $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space. Let $F_\varepsilon(x|\xi_n) = \mathbb{P}(X_{n+1} \le x|\xi_n)$ be the conditional distribution function of X_{n+1} given the sigma algebra generated by $\xi_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$; let $F(x) = \mathbb{P}(X_1 \le x)$ and $R_n(s) = \sqrt{n}[F_n(s) - F(s)]$. Assume throughout the paper that the conditional density $f_\varepsilon(x|\xi_n) = (\partial/\partial x)F_\varepsilon(x|\xi_n)$ exists almost surely. Define the weighted measure $w_\lambda(du) = (1 + |u|)^\lambda(du)$. For a random variable Z write $Z \in \mathcal{L}^q$, $q > 0$, if $\|Z\|_q := [\mathbb{E}(|Z|^q)]^{1/q} < \infty$. Write the $\mathcal{L}^2$ norm as $\|Z\| = \|Z\|_2$. Define projections $\mathcal{P}_k Z = \mathbb{E}(Z|\xi_k) - \mathbb{E}(Z|\xi_{k-1})$, $k \in \mathbb{Z}$. Denote by C_q (resp. C_γ, C_μ, etc) generic positive constants which only depend on q (resp. γ, μ, etc). Their values may vary from line to line.

The rest of paper is structured as follows. Section 2 concerns sample path properties and weak convergence of R_n. In particular, in Section 2.1, the Chibisov-O'Reilly Theorem, which concerns the weak convergence of weighted empirical processes, is generalized to dependent random variables. Section 2.2 considers the weighted modulus of continuity of R_n. Applications to linear processes and nonlinear time series are made in Section 3. Section 4 presents some useful inequalities that may be of independent interest. The inequalities are applied in Section 5, where proofs of the main results are given.

2. Main results

It is certainly necessary to impose appropriate dependence structures on the process (X_i). We start by proposing a particular dependence condition which is quite different from the classical strong mixing assumptions. Let m be a measure on the 1-dimensional Borel space $(\mathbb{R}, \mathcal{B})$. For $\theta \in \mathbb{R}$, let $T_n(\theta) = \sum_{i=1}^n h(\theta, \xi_i) - n\mathbb{E}[h(\theta, \xi_1)]$, where h is a measurable function such that $\|h(\theta, \xi_1)\| < \infty$ for almost all θ (m). Denote by $h_j(\theta, \xi_0) = \mathbb{E}[h(\theta, \xi_j)|\xi_0]$ the j-step-ahead predicted mean, $j \ge 0$. Let (ε_i') be an i.i.d. copy of (ε_i), $\xi_k^* = (\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_k)$, $k \ge 0$, and define

$$\sigma(h, m) = \sum_{j=0}^{\infty} \Big[\int_{\mathbb{R}} \|h_j(\theta, \xi_0) - h_j(\theta, \xi_0^*)\|^2 m(d\theta) \Big]^{\frac{1}{2}}. \tag{3}$$

Let $f_\varepsilon'(\theta|\xi_k) = (\partial/\partial\theta) f_\varepsilon(\theta|\xi_k)$. In the case that $h(\theta, \xi_k) = F_\varepsilon(\theta|\xi_k)$ [resp. $f_\varepsilon(\theta|\xi_k)$ or $f_\varepsilon'(\theta|\xi_k)$], we write $\sigma(F_\varepsilon, m)$ [resp. $\sigma(f_\varepsilon, m)$ or $\sigma(f_\varepsilon', m)$] for $\sigma(h, m)$.

Our dependence conditions are expressed as $\sigma(F_\varepsilon, m) < \infty$, $\sigma(f_\varepsilon, m) < \infty$ and $\sigma(f_\varepsilon', m) < \infty$, where $m(dt) = (1 + |t|)^\delta dt$, $\delta \in \mathbb{R}$. These conditions are interestingly connected with the prediction theory of stochastic processes (Remark 1 and Section 2.3). The quantity $\sigma(h, m)$ can be interpreted as a cumulative weighted prediction measure. It is worthwhile to note that $\|h_j(\theta,\xi_0) - h_j(\theta,\xi_0^*)\|$ has the same order of magnitude as $\|\mathcal{P}_0 h(\theta,\xi_j)\|$. To see this, let $g(\ldots,e_{-1},e_0) = h(\theta,(\ldots,e_{-1},e_0))$. So $g(\xi_j^*) = h(\theta,\xi_j^*)$, $g_j(\xi_0^*) = \mathbb{E}[g(\xi_j^*)|\xi_0^*] = h_j(\theta,\xi_0^*)$, $\mathbb{E}[g(\xi_j)|\xi_{-1}] = \mathbb{E}[g(\xi_j^*)|\xi_0]$, and

$$\begin{aligned}\|\mathcal{P}_0 g(\xi_j)\| &= \|\mathbb{E}[g_j(\xi_0) - g_j(\xi_0^*)|\xi_0]\| \le \|g_j(\xi_0) - g_j(\xi_0^*)\| \\ &\le \|g_j(\xi_0) - g_{j+1}(\xi_{-1})\| + \|g_{j+1}(\xi_{-1}) - g_j(\xi_0^*)\| = 2\|\mathcal{P}_0 g(\xi_j)\|. \quad (4)\end{aligned}$$

Note also that $\|\mathcal{P}_0 g(\xi_j)\| \le \|g(\xi_j) - g(\xi_j^*)\|$. Therefore we have

$$\begin{aligned}\frac{1}{2}\sigma(h,m) \le \tilde{\sigma}(h,m) &:= \sum_{j=0}^{\infty}\Big[\int_{\mathbb{R}} \|\mathcal{P}_0 h(\theta,\xi_j)\|^2 m(d\theta)\Big]^{\frac{1}{2}} \le \sigma(h,m) \\ &\le 2\sum_{j=0}^{\infty}\Big[\int_{\mathbb{R}} \|h(\theta,\xi_j) - h(\theta,\xi_j^*)\|^2 m(d\theta)\Big]^{\frac{1}{2}} =: 2\hat{\sigma}(h,m). \quad (5)\end{aligned}$$

All results in the paper with the condition $\sigma(h,m) < \infty$ can be re-stated in terms of $\tilde{\sigma}(h,m) < \infty$, or the stronger version $\hat{\sigma}(h,m) < \infty$.

2.1. Weak convergence

The classical Donsker theorem asserts that, if X_k are i.i.d., then $\{R_n(s),\ s \in \mathbb{R}\}$ converges in distribution to an F-Brownian bridge process. The Donsker theorem has many applications in statistics. To understand the behavior at the two extremes $s = \pm\infty$, we need to consider the weighted version $\{R_n(s)W(s),\ s \in \mathbb{R}\}$, where $W(s) \to \infty$ as $s \to \pm\infty$. Clearly, if W is bounded, then by the Continuous Mapping Theorem, the weak convergence of the weighted empirical processes follows from that of R_n. By allowing $W(s) \to \infty$ as $s \to \pm\infty$, one can have the weak convergence of functionals of empirical processes, $t(F_n)$, for a wider class of functionals. The Chibisov-O'Reilly Theorem concerns weighted empirical processes of i.i.d. random variables; see Shorack and Wellner (1986, Sec. 11.5). The case of dependent random variables has been far less studied. For strongly mixing processes see Mehra and Rao (1975), Shao and Yu (1996) and Csörgő and Yu (1996).

Theorem 1. *Let $\gamma' \ge 0$, $q > 2$, and $\gamma = \gamma' q/2$. Assume $\mathbb{E}[|X_1|^\gamma + \log(1+|X_1|)] < \infty$ and*

$$\int_{\mathbb{R}} \mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)] w_{\gamma-1+\frac{q}{2}}(du) < \infty. \quad (6)$$

In addition assume that there exists $0 \le \mu \le 1$ and $\nu > 0$ such that

$$\sigma(F_\varepsilon, w_{\gamma'-\mu}) + \sigma(f_\varepsilon, w_{\gamma'+\mu}) + \sigma(f'_\varepsilon, w_{-\nu}) < \infty. \quad (7)$$

Then (i) $\mathbb{E}[\sup_{s\in\mathbb{R}} |R_n(s)|^2(1+|s|)^{\gamma'}] = \mathcal{O}(1)$, *and* (ii) *the process* $\{R_n(s)(1+|s|)^{\gamma'/2},\ s\in\mathbb{R}\}$ *converges weakly to a tight Gaussian process. In particular, if* $X_1 \in \mathcal{L}^{\gamma+q/2-1}$ *and*

$$\sup_u f_\varepsilon(u|\xi_0) \le C \tag{8}$$

holds almost surely for some constant $C<\infty$, *then* (6) *holds.*

An important issue in applying Theorem 1 is to verify (7), which is basically a short-range dependence condition (cf. Remark 1). For many important models including linear processes and Markov chains, (7) is easily verifiable (Section 3). If (X_n) is a Markov chain, then $\sigma(h,m)$ is related to the sensitivity measure (Fan and Yao (2003, p. 466)) appearing in nonlinear prediction theory (Section 2.3). If (X_n) is a linear process, then (7) reduces to the classical definition of the short-range dependence of linear processes.

Remark 1. If $h(\theta,\xi_k) = f_\varepsilon(\theta|\xi_k)$, then $h_k(\theta,\xi_0) = \mathbb{E}[f_\varepsilon(\theta|\xi_k)|\xi_0] = f_{k+1}(\theta|\xi_0)$, the conditional density of X_{k+1} at θ given ξ_0. Note that $\xi_0^* = (\xi_{-1},\varepsilon_0')$ is a coupled version of ξ_0 with ε_0 replaced by ε_0'. So $h_k(\theta,\xi_0) - h_k(\theta,\xi_0^*) = f_{k+1}(\theta|\xi_0) - f_{k+1}(\theta|\xi_0^*)$ measures the change in the $(k+1)$-step-ahead predictive distribution if ξ_0 is changed to its coupled version ξ_0^*. In other words, $h_k(\theta,\xi_0) - h_k(\theta,\xi_0^*)$ can be viewed as the contribution of ε_0 in predicting X_{k+1}. So, in this sense, the condition $\sigma(h,m)<\infty$ means that the cumulative contribution of ε_0 in predicting future values $X_k, k\ge 1$, is finite. It is then not unnatural to interpret $\sigma(h,m)$ as a cumulative weighted prediction measure. This interpretation seems in line with the connotation of short-range dependence.

We now compare Theorem 1 with the Chibisov-O'Reilly Theorem that concerns the weak convergence of weighted empirical processes for i.i.d. random variables. Note that $(\gamma+q/2-1)\downarrow\gamma'$ as $q\downarrow 2$. The moment condition $X_1\in\mathcal{L}^{\gamma+q/2-1}$ in Theorem 1 is almost necessary in the sense that it cannot be replaced by the weaker

$$\mathbb{E}\{|X_1|^{\gamma'}\log^{-1}(2+|X_1|)[\log\log(10+|X_1|)]^{-\lambda}\}<\infty \tag{9}$$

for some $\lambda>0$. To see this let X_k be i.i.d. symmetric random variables with continuous, strictly increasing distribution function F; let $F^{\#}$ be the quantile function and $m(u) = [1+|F^{\#}(u)|]^{-\gamma'/2}$. Then we have the distributional equality

$$\{R_n(s)(1+|s|)^{\frac{\gamma'}{2}},\ s\in\mathbb{R}\} =_{\mathcal{D}} \Big\{\frac{R_n(F^{\#}(u))}{m(u)},\ u\in(0,1)\Big\}.$$

Assume that $F(s)(1+|s|)^{\gamma'}$ is increasing on $(-\infty,G)$ for some $G<0$. Then $m(u)/\sqrt{u}$ is decreasing on $(0,F(G))$. By the Chibisov-O'Reilly Theorem,

$\{R_n(F^{\#}(u))/m(u), u \in (0,1)\}$ is tight if and only if $\lim_{t\downarrow 0} m(t)/\sqrt{t\log\log(t^{-1})} = \infty$, namely

$$\lim_{u\to-\infty} F(u)(1+|u|)^{\gamma'} \log\log|u| = 0. \tag{10}$$

Condition (10) controls the heaviness of the tail of X_1. Let $F(u) = |u|^{-\gamma'}(\log\log|u|)^{-1}$ for $u \le -10$. Then (9) holds while (10) is violated. It is unclear whether stronger versions of (9), such as $\mathbb{E}(|X_1|^{\gamma'}) < \infty$ or $\mathbb{E}[|X_1|^{\gamma'}\log^{-1}(2 + |X_1|)] < \infty$, are sufficient.

2.2. Modulus of continuity

Theorem 2 concerns the weighted modulus of continuity of $R_n(\cdot)$. Sample path properties of empirical distribution functions of i.i.d. random variables have also been extensively explored; see for example Csörgő et al. (1986), Shorack and Wellner (1986), and Einmahl and Mason (1988), among others. It is far less studied for the dependent case.

Theorem 2. *Let $\gamma' > 0$, $2 < q < 4$, and $\gamma = \gamma' q/2$; let $\delta_n < 1/2$ be a sequence of positive numbers such that $(\log n)^{2q/(q-2)} = \mathcal{O}(n\delta_n)$. Assume (8), $X_1 \in \mathcal{L}^\gamma$, and*

$$\sigma(f_\varepsilon, w_{\gamma'}) + \sigma(f'_\varepsilon, w_{\gamma'}) < \infty. \tag{11}$$

Then there exists a constant $0 < C < \infty$, independent of n and δ_n, such that for all $n \ge 1$,

$$\mathbb{E}\Big[\sup_{t\in\mathbb{R}}(1+|t|)^{\gamma'} \sup_{|s|\le\delta_n} |R_n(t+s) - R_n(t)|^2\Big] \le C\delta_n^{1-\frac{2}{q}}. \tag{12}$$

2.3. Sensitivity measures and dependence

Our basic dependence condition is that $\sigma(h,m) < \infty$. Here we present its connection with prediction sensitivity measures (Fan and Yao (2003, p.466)) special structure. Assume that (X_n) is a Markov chain expressed in the form of an iterated random function (Elton (1990) and Diaconis and Freedman (1999)):

$$X_n = M(X_{n-1}, \varepsilon_n), \tag{13}$$

where ε_k, $k \in \mathbb{Z}$, are i.i.d. random variables and $M(\cdot,\cdot)$ is a bivariate measurable function. For $k \ge 1$ let $f_k(\cdot|x)$ be the conditional (transition) density of X_k given $X_0 = x$. Then $f_\varepsilon(\theta|\xi_k) = f_1(\theta|X_k)$ is the conditional density of X_{k+1} at θ given X_k, and, for $k \ge 0$, $\mathbb{E}[f_\varepsilon(\theta|\xi_k)|\xi_0] = f_{k+1}(\theta|X_0)$. Fan and Yao (2003) argue that

$$D_k(x,\delta) := \int_{\mathbb{R}} [f_k(\theta|x+\delta) - f_k(\theta|x)]^2 d\theta \tag{14}$$

is a natural way to measure the deviation of the conditional distribution of X_k given $X_0 = x$. In words, D_k quantifies the sensitivity to initial values and it measures the error in the k-step-ahead predictive distribution due to a drift in the initial value. Under certain regularity conditions,

$$\lim_{\delta\to 0}\frac{D_k(x,\delta)}{\delta^2} = \int_{\mathbb{R}}\Big[\frac{\partial f_k(\theta|x)}{\partial x}\Big]^2 d\theta =: I_k(x).$$

Here I_k is called prediction sensitivity measure. It is a useful quantity in the prediction theory of nonlinear dynamical systems. Estimation of I_k is discussed in Fan and Yao (2003, p. 468). Proposition 1 shows the relation between $\sigma(h,m)$ and I_k. Since it can be proved in the same way as (i) of Theorem 3, we omit the details of its proof.

Proposition 1. *Let* $k \ge 0$. *For the process* (13), *we have*

$$\int_{\mathbb{R}} \|h_k(\theta,\xi_0) - h_k(\theta,\xi_0^*)\|^2 m(d\theta) \le 4\|\tau_k(X_0,X_0^*)\|^2,$$

where

$$\tau_k(a,b) = \int_a^b [I_{h_k}(x,m)]^{\frac{1}{2}} dx \quad\text{and}\quad I_{h_k}(x,m) = \int_{\mathbb{R}}\Big[\frac{\partial h_k(\theta,x)}{\partial x}\Big]^2 m(d\theta).$$

Consequently, $\sigma(h,m) < \infty$ *holds if* $\sum_{k=1}^\infty \|\tau_k(X_0,X_0^*)\| < \infty$.

In the special case $h(\theta,\xi_k) = f_\varepsilon(\theta|\xi_k) = f_1(\theta|X_k)$, $h_k(\theta,\xi_0) = \mathbb{E}[h(\theta,\xi_k)|\xi_0]$ $= f_{k+1}(\theta|X_0)$ and $I_{h_k}(x,m)$ reduces to Fan and Yao's sensitivity measure $I_{k+1}(x)$ provided $m(d\theta) = d\theta$ is Lebesgue measure. So it is natural to view $I_{h_k}(x,m)$ as a weighted sensitivity measure. Since the k-step-ahead conditional density $f_k(\theta|x)$, may have an intractable and complicated form, it is generally not very easy to apply Proposition 1. This is especially so in nonlinear time series where it is often quite difficult to derive explicit forms of $f_k(\theta|x)$. To circumvent such a difficulty, our Theorem 3 provides sufficient conditions which only involve 1-step-ahead conditional densities.

For processes that are not necessarily in the form (13), we assume that there exists an $\sigma(\xi_n)$-measurable random variable Y_n such that

$$\mathbb{P}(X_{n+1} \le x|\xi_n) = \mathbb{P}(X_{n+1} \le x|Y_n) := F(x|Y_n). \tag{15}$$

Then there exists a similar bound as the one given in Proposition 1. Write $Y_n = I(\xi_n)$, $Y_n^* = I(\xi_n^*)$ and $h(\theta,\xi_n) = h(\theta,Y_n)$. For Markov chains, (15) is satisfied with $Y_n = X_n$. Let $X_n = \sum_{i=0}^\infty a_i\varepsilon_{n-i}$ be a linear process. Then (15) is also satisfied with $Y_n = \sum_{i=1}^\infty a_i\varepsilon_{n+1-i}$. Let $f(\theta|y)$ be the conditional density of

X_{n+1} given $Y_n = y$ and $f'(\theta|y) = (\partial/\partial\theta)f(\theta|y)$. Then $\sigma(h,m)$ is also related to a weighted distance between Y_n and Y_n^*. Define

$$\rho_{h,m}(a,b) = \int_a^b H_{h,m}^{\frac{1}{2}}(y)dy, \quad \text{where} \quad H_{h,m}(y) = \int_{\mathbb{R}} \Big|\frac{\partial}{\partial y} h(\theta,y)\Big|^2 m(d\theta). \tag{16}$$

Theorem 3. (i) *Let* $\Xi_{h,m} = \sum_{n=0}^{\infty} \|\rho_{h,m}(Y_n, Y_n^*)\|$. *Then* $\sigma(h,m) \le 2\Xi_{h,m}$ *and*

$$\int_{\mathbb{R}} \|h_n(\theta,\xi_0) - h_n(\theta,\xi_0^*)\|^2 m(d\theta) \le 4\|\rho_{h,m}(Y_n,Y_n^*)\|^2. \tag{17}$$

(ii) *Assume there exist* $C > 0$ *and* $q \in \mathbb{R}$ *such that* $H_{h,m}(y) \le C(1+|y|)^{q-2}$ *holds for all* $y \in \mathbb{R}$. *Then* $\|\rho_{h,m}(Y_n,Y_n^*)\| = O[\|Y_n - Y_n^*\|_q^{\min(1,q/2)}]$ *if* $q > 0$, *and* $\|\rho_{h,m}(Y_n,Y_n^*)\| = O[\|\min(1,|Y_n - Y_n^*|)\|]$ *if* $q < 0$.

Let $h(\theta, Y_n) = f(\theta|Y_n)$. Then $H_{h,m}(y)$ can be interpreted as a measure of "local dependence" of X_{n+1} on Y_n at y. As does $D_k(x,\delta)$ in (14), $H_{h,m}(y)$ measures the distance between the conditional densities of $[X_{n+1}|Y_n = y]$ and $[X_{n+1}|Y_n = y+\delta]$. In many situations $\|Y_n - Y_n^*\|_q$ is easy to work with since it is directly related to the data generating mechanisms.

3. Applications

3.1. Iterated random functions

Consider the process (13). The existence of stationary distributions has been widely studied and there are many versions of sufficient conditions; see Diaconis and Freedman (1999), Meyn and Tweedie (1993), Jarner and Tweedie (2001), Wu and Shao (2004), among others. Here we adopt the conditions given by Diaconis and Freedman (1999). The recursion (13) has a unique stationary distribution if there exist $\alpha > 0$ and x_0 such that

$$L_{\varepsilon_0} + |M(x_0,\varepsilon_0)| \in \mathcal{L}^\alpha \text{ and } \mathbb{E}[\log(L_{\varepsilon_0})] < 0, \text{ where } L_\varepsilon = \sup_{x \ne x'} \frac{|M(x,\varepsilon) - M(x',\varepsilon)|}{|x - x'|}. \tag{18}$$

Condition (18) also implies the *geometric-moment contraction* (GMC(β), see Wu and Shao (2004)) property: there exist $\beta > 0$, $r \in (0,1)$, and $C < \infty$ such that

$$\mathbb{E}[|J(\ldots,\varepsilon_{-1},\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_n) - J(\ldots,\varepsilon'_{-1},\varepsilon'_0,\varepsilon_1,\ldots,\varepsilon_n)|^\beta] \le Cr^n \tag{19}$$

holds for all $n \in \mathbb{N}$, where (ε'_k) is an i.i.d. copy of (ε_k). Hsing and Wu (2004) argued that (19) is a convenient condition for limit theorems.

Doukhan (2003) gave a brief review of weak convergence of R_n under various dependence structures including strong mixing conditions. Meyn and Tweedie

(1993) discussed alpha and beta-mixing properties of Markov chains. Doukhan and Louhichi (1999) introduced an interesting weak dependence condition based on the decay rate of the covariances between the past and the future values of the process, and obtained weak convergence of R_n. Wu and Shao (2004, pp.431-432) applied Doukhan and Louhichi's results and proved that, under (19), R_n converges weakly to a tight Gaussian process if, for some $\kappa > 5/2$, $\sup_x |F(x+\delta) - F(x)| = O[\log^{-\kappa}(|\delta|^{-1})]$ as $\delta \to 0$. The latter condition holds if X_k has a bounded density. Note that results based on Doukhan and Louhichi's dependence coefficients do not require that F has a density. In comparison, our Theorem 1 concerns weighted empirical processes under the assumption that conditional densities exist. They can be applied in different situations. In Section 3.1.1, we consider weak convergence of weighted empirical processes for autoregressive conditionally heteroskedastic (ARCH, Engle (1982))sequences. Berkes and Horváth (2001) discussed strong approximation for empirical processes of generalized ARCH sequences.

3.1.1. ARCH models

Consider the model

$$X_n = \varepsilon_n \sqrt{a^2 + b^2 X_{n-1}^2}, \tag{20}$$

where a and b are real parameters for which $ab \neq 0$ and ε_i are i.i.d. innovations with density f_ε. Let $M(x,\varepsilon) = \varepsilon\sqrt{a^2 + b^2 x^2}$. Then $L_\varepsilon = \sup_x |\partial M(x,\varepsilon)/\partial x| \leq |b\varepsilon|$. Assume that there exists $\beta > 0$ such that $r_0 := \mathbb{E}(|b\varepsilon_0|^\beta) < 1$. Then (18) holds with $\alpha = \beta$ and $\|M(x,\varepsilon_n) - M(x',\varepsilon_n)\|_\beta^\beta \leq r_0|x - x'|^\beta$. Iterations of the latter inequality imply (19) with $r = r_0$. For more details see Wu and Shao (2004).

Corollary 1. *Assume* $\mathbb{E}(|b\varepsilon_0|^\beta) < 1$, $\beta > 0$. *Let* $\phi \in (0,\beta)$ *and* $\nu > 0$ *satisfy*

$$\int_{\mathbb{R}} |f_\varepsilon(u)|^2 w_{1+\phi}(du) + \int_{\mathbb{R}} |f'_\varepsilon(u)|^2 w_{3+\phi}(du) + \int_{\mathbb{R}} |f''_\varepsilon(u)|^2 w_{2-\nu}(du) < \infty. \tag{21}$$

Then $\{R_n(s)(1+|s|)^{\phi/2},\ s \in \mathbb{R}\}$ *converges weakly to a tight Gaussian process.*

Proof. Let $t = t_y = \sqrt{a^2 + b^2 y^2}$. Then $F(\theta|y) = F_\varepsilon(\theta/t)$, $f(\theta|y) = f_\varepsilon(\theta/t)/t$, and $f'(\theta|y) = f'_\varepsilon(\theta/t)/t^2$. By (24) of Lemma 1 and (21), we have $\sup_x f_\varepsilon(x) < \infty$ and (8). Choose $q > 2$ such that $\tau = \phi q/2 + q/2 - 1 < \beta$. Then $X_1 \in \mathcal{L}^\tau$. By Theorem 1, it remains to verify (7). Let $\mu = 1$. Recall (16) for $H_{h,m}(y)$. By considering the cases $\phi \geq 1$ and $\phi < 1$ separately, we have, by (21) and the inequality $|ut| \leq 1 + |ut| \leq (1+|u|)(1+|t|)$, that

$$\int_{\mathbb{R}} \Big|\frac{\partial F_\varepsilon(\theta/t)}{\partial t}\Big|^2 w_{\phi-1}(d\theta) = t^{-1}\int_{\mathbb{R}} f_\varepsilon^2(u) u^2 (1+|ut|)^{\phi-1} du \leq C t^{\phi-2}$$

holds for some constant $C \in (0, \infty)$. Since $|dt/dy| \le |b|$, $\int_{\mathbb{R}} |\partial F(\theta|y)/\partial y|^2 w_{\phi-1}(d\theta) \le C(1+|y|)^{\phi-2}$. Similar but lengthy calculations show that $\int_{\mathbb{R}} |\partial f(\theta|y)/\partial y|^2 w_{\phi+1}(d\theta) \le C(1+|y|)^{\phi-2}$ and $\int_{\mathbb{R}} |\partial f'(\theta|y)/\partial y|^2 w_{-2-\nu}(d\theta) \le C(1+|y|)^{-5-\nu}$. By (19), $\|Y_n - Y_n^*\|_\phi = O(r^n)$ for some $r \in (0,1)$. By Theorem 3, simple calculations show that (7) holds.

Corollary 1 allows heavy-tailed ARCH processes. Tsay (2005) argued that in certain applications it is more appropriate to assume that ε_k has heavy tails. Let ε_k have a standard Student-t distribution with degrees of freedom ν, with density $f_\varepsilon(u) = (1+u^2/\nu)^{-(1+\nu)/2} c_\nu$, where $c_\nu = \Gamma((\nu+1)/2)/[\Gamma(\nu/2)\sqrt{\nu\pi}]$. Then (21) holds if $\phi < 2\nu$. Note that $\varepsilon_k \in \mathcal{L}^\phi$ if $\phi < \nu$, and consequently $X_k \in \mathcal{L}^\phi$ if $\mathbb{E}(|b\varepsilon_0|^\phi) < 1$.

3.2. Linear processes

Let $X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, where the ε_k are i.i.d. random variables with mean 0 and finite and positive variance, and the coefficients a_i satisfy $\sum_{i=0}^{\infty} a_i^2 < \infty$. Assume without loss of generality that $a_0 = 1$. Let F_ε and $f_\varepsilon = F_\varepsilon'$ be the distribution and density functions of ε_k. Then the conditional density of X_{n+1} given ξ_n is $f_\varepsilon(x - Y_n)$, where $Y_n = X_{n+1} - \varepsilon_{n+1}$ (cf. (15)).

Corollary 2. *Let $\gamma \ge 0$. Assume $\varepsilon_k \in \mathcal{L}^{2+\gamma}$, $\sup_u f_\varepsilon(u) < \infty$, and*

$$\sum_{n=1}^{\infty} |a_n| < \infty, \tag{22}$$

$$\int_{\mathbb{R}} |f_\varepsilon'(u)|^2 w_\gamma(du) + \int_{\mathbb{R}} |f_\varepsilon''(u)|^2 w_{-\gamma}(du) < \infty. \tag{23}$$

Then $\{R_n(s)(1+|s|)^{\gamma/2},\ s \in \mathbb{R}\}$ converges weakly to a tight Gaussian process.

Proof. Let $q = 2+\gamma$. Since $\varepsilon_1 \in \mathcal{L}^q$, we have $Y_1 \in \mathcal{L}^q$ and $\|Y_n - Y_n^*\|_q = \|a_{n+1}(\varepsilon_0 - \varepsilon_0')\|_q = O(|a_{n+1}|)$. Note that $f(x|y) = f_\varepsilon(x-y)$. Since $\int_{\mathbb{R}} f_\varepsilon(\theta) w_\gamma(d\theta) = \mathbb{E}[(1+|\varepsilon_1|)^\gamma] < \infty$ and $\sup_u f_\varepsilon(u) < \infty$, we have $\int_{\mathbb{R}} f_\varepsilon(u)^2 w_\gamma(du) < \infty$. Since $1+|v+y| \le (1+|v|)(1+|y|)$,

$$\begin{aligned}\int_{\mathbb{R}} [|f_\varepsilon(\theta-y)|^2 + |f_\varepsilon'(\theta-y)|^2] w_\gamma(d\theta) &\le (1+|y|)^\gamma \int_{\mathbb{R}} [|f_\varepsilon(\theta)|^2 + |f_\varepsilon'(\theta)|^2] w_\gamma(d\theta) \\ &\le C(1+|y|)^\gamma.\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_{\mathbb{R}} |f_\varepsilon''(\theta-y)|^2 (1+|\theta|)^{-\gamma} d\theta &\le \int_{\mathbb{R}} |f_\varepsilon''(\theta-y)|^2 (1+|\theta-y|)^{-\gamma}(1+|y|)^\gamma d\theta \\ &= (1+|y|)^\gamma \int_{\mathbb{R}} |f_\varepsilon''(u)|^2 w_{-\gamma}(du).\end{aligned}$$

By Theorems 1 and 3, the corollary follows.

Condition (22), together with $\varepsilon_1 \in \mathcal{L}^2$, implies that the covariances of (X_t) are absolutely summable, a well-known condition for a linear process being short-range dependent. If (22) is violated, then we say that the linear process (X_n) is long-range dependent. In this case properties of F_n have been discussed by Ho and Hsing (1996) and Wu (2003).

Remark 2. Strong mixing properties of linear processes have been discussed by Withers (1981) and Pham and Tran (1985), among others. It seems that for linear processes strong mixing conditions do not lead to results with best possible conditions. Let $\varepsilon_i \in \mathcal{L}^2$ have a density and $A_n = \sum_{i=n}^{\infty} a_i^2$. Withers (1981) showed that (X_i) is strongly mixing if $\sum_{n=1}^{\infty} \max(A_n^{1/3}, \sqrt{A_n|\log A_n|}) < \infty$. In the case that $a_n = n^{-\delta}$, the latter condition requires $\delta > 2$. In comparison, our condition (22) only requires that $\delta > 1$. More stringent conditions are required for β-mixing.

Remark 3. Let $\phi(t) = \mathbb{E}\exp(\sqrt{-1}u\varepsilon_0)$; let f_k (resp. ϕ_k) be the density (resp. characteristic) function of $\sum_{i=0}^{k} a_i\varepsilon_{-i}$. Doukhan (2003) mentioned the following result (see also Doukhan and Surgailis (1998)): R_n converges weakly if for some $0 < \alpha \leq 1$ and $s, C, \delta > 0$ such that $s\delta > 2\alpha$, $\varepsilon_k \in \mathcal{L}^s$, $\sum_{i=0}^{\infty} |a_i|^\alpha < \infty$ and $|\phi(t)| \leq C/(1 + |u|^\delta)$. (Doukhan noted that the proof of the latter result is unpublished.) These results allow heavy-tailed ε_k. Our Corollary 2 deals with weighted empirical processes, so there are different ranges of applications. Consider the special case of Corollary 2 in which $\gamma = 0$. A careful check of the proofs of Theorems 1 and 3 suggests that Corollary 2 still holds if f'_ε in (23) is replaced by f_k for some fixed k (see also Wu (2003)). Note that, by Parseval's identity, if $\gamma = 0$, then $2\pi \int_{\mathbb{R}} |f''_k(u)|^2 du = \int_{\mathbb{R}} |t\phi_k(t)|^2 dt$, which is finite if $|\phi_k(t)| \leq C/(1 + t^2)$. The latter condition holds if $|\phi(t)| \leq C/(1 + |u|^\delta)$ and $\#\{i : a_i \neq 0\} = \infty$.

Remark 4. Let g be a Lipschitz continuous function such that, for some $r < 1$, $|g(x) - g(x')| \leq r|x - x'|$ holds for all x, $x' \in \mathbb{R}$. Consider the nonlinear model $X_{n+1} = g(X_n) + \varepsilon_{n+1}$, where ε_n satisfy conditions in Corollary 2. Then X_n is GMC(q) and consequently $Y_n = g(X_n)$ is also GMC(q), $q = 2 + \gamma$. It is easily seen that the argument in Corollary 2 also applies to this model and that the same conclusion holds. Special cases include the threshold autoregressive model (Tong (1990)) $X_{n+1} = a\max(X_n, 0) + b\min(X_n, 0) + \varepsilon_{n+1}$ with $\max(|a|, |b|) < 1$, and the exponential autoregressive model (Haggan and Ozaki (1981)) $X_{n+1} = [a + b\exp(-cX_n^2)]X_n + \varepsilon_{n+1}$, where $|a| + |b| < 1$ and $c > 0$.

4. Inequalities

The inequalities presented in this section are of independent interest and may have wider applicability. They are used in the proofs of the results in other sections.

Lemma 1. *Let H be absolutely continuous.* (i) *If $\mu \le 1$ and $\gamma \in \mathbb{R}$, then*

$$\sup_{x\ge M}[H^2(x)(1+|x|)^{\gamma}] \le C_{\gamma,\mu}\int_M^{\infty} H^2(u)w_{\gamma-\mu}(du) + C_{\gamma,\mu}\int_M^{\infty}[H'(u)]^2 w_{\gamma+\mu}(du) \tag{24}$$

holds for all $M \ge 0$, where $C_{\gamma,\mu}$ is a positive constant. The above inequality also holds if $M = -\infty$ and $\sup_{x\ge M}$ is replaced by $\sup_{x\in\mathbb{R}}$. (ii) *If $\gamma > 0$, $\mu = 1$, and $H(0) = 0$, then*

$$\sup_{x\in\mathbb{R}}[H^2(x)(1+|x|)^{-\gamma}] \le \frac{1}{\gamma}\int_{\mathbb{R}}[H'(u)]^2 w_{-\gamma+1}(du), \tag{25}$$

$$\int_{\mathbb{R}} H^2(u)w_{-\gamma-1}(du) \le \frac{4}{\gamma^2}\int_{\mathbb{R}}[H'(u)]^2 w_{-\gamma+1}(du). \tag{26}$$

(iii) *If $\gamma > 0$ and $H(\pm\infty) = 0$, then $\sup_{x\in\mathbb{R}}[H^2(x)(1+|x|)^{\gamma}] \le \gamma^{-1}\int_{\mathbb{R}}[H'(u)]^2 w_{\gamma+1}(du)$ and $\int_{\mathbb{R}} H^2(u)w_{\gamma-1}(du) \le 4\gamma^{-2}\int_{\mathbb{R}}[H'(u)]^2 w_{\gamma+1}(du)$.*

Proof. (i) By Lemma 4 in Wu (2003), for $t \in \mathbb{R}$ and $\delta > 0$ we have

$$\sup_{t\le s\le t+\delta} H^2(s) \le \frac{2}{\delta}\int_t^{t+\delta} H^2(u)du + 2\delta\int_t^{t+\delta}[H'(u)]^2 du. \tag{27}$$

We first consider the case $\mu < 1$. Let $\alpha = 1/(1-\mu)$. In (27) let $t = t_n = n^{\alpha}$ and $\delta_n = (n+1)^{\alpha} - n^{\alpha}$, $n \in \mathbb{N}$, and $I_n = [t_n, t_{n+1}]$. Since $\lim_{n\to\infty}\delta_n/(\alpha n^{\alpha-1}) = 1$,

$$\begin{aligned}\sup_{x\in I_n}[H^2(x)(1+x)^{\gamma}] &\le 2\sup_{x\in I_n}(1+x)^{\gamma}\Big[\delta_n^{-1}\int_{I_n} H^2(u)du + \delta_n\int_{I_n}[H'(u)]^2 du\Big]\\ &\le C\int_{I_n} H^2(u)w_{\gamma-\mu}(du) + C\int_{I_n}[H'(u)]^2 w_{\gamma+\mu}(du). \end{aligned}\tag{28}$$

It is easily seen by (27) that (28) also holds for $n = 0$ by choosing a suitable C. By summing (28) over $n = 0, 1, \ldots$, we obtain (24) with $M = 0$. The case $M > 0$ can be similarly dealt with by letting $t_n = n^{\alpha} + M$.

If $\mu = 1$, we let $t_n = 2^n$, $\delta_n = t_{n+1} - t_n = t_n$ and $I_n = [t_n, t_{n+1}]$, $n = 0, 1, \ldots$. The argument above yields the desired inequality.

(ii) Let $s \ge 0$. Since $H(s) = \int_0^s H'(u)du$, by the Cauchy-Schwarz Inequality,

$$\begin{aligned}H^2(s) &\le \int_0^s |H'(u)|^2(1+u)^{1-\gamma}du \times \int_0^s (1+u)^{\gamma-1}du\\ &\le \int_{\mathbb{R}}[H'(u)]^2 w_{-\gamma+1}(du) \times \frac{(1+s)^{\gamma}-1}{\gamma}.\end{aligned}$$

So (25) follows. Applying Theorem 1.14 in Opic and Kufner (1990, p. 13) with $p = q = 2$, the Hardy-type inequality (26) easily follows. The proof of (iii) is similar to that of (ii).

Lemma 2. *Let m be a measure on $\mathbb{R}$, $A \subset \mathbb{R}$ be a measurable set, and $T_n(\theta) = \sum_{i=1}^n h(\theta, \xi_i)$, where h is a measurable function. Then*

$$\sqrt{\int_A \|T_n(\theta) - \mathbb{E}[T_n(\theta)]\|^2 m(d\theta)} \leq \sqrt{n} \sum_{j=0}^{\infty} \sqrt{\int_A \|\mathcal{P}_0 h(\theta, \xi_j)\|^2 m(d\theta)}. \tag{29}$$

Proof. For $j = 0, 1, \ldots$ let $T_{n,j}(\theta) = \sum_{i=1}^n \mathbb{E}[h(\theta, \xi_i)|\xi_{i-j}]$ and $\lambda_j^2 = \int_A \|\mathcal{P}_0 h(\theta, \xi_j)\|^2 m(d\theta)$, $\lambda_j \geq 0$. By the orthogonality of $\mathbb{E}[h(\theta,\xi_i)|\xi_{i-j}] - \mathbb{E}[h(\theta,\xi_i)|\xi_{i-j-1}]$, $i = 1, 2, \ldots, n$,

$$\begin{aligned}\int_A \|T_{n,j}(\theta) - T_{n,j-1}(\theta)\|^2 m(d\theta) &= n \int_A \|\mathbb{E}[h(\theta,\xi_1)|\xi_{1-j}] - \mathbb{E}[h(\theta,\xi_1)|\xi_{-j}]\|^2 m(d\theta) \\ &= n \int_A \|\mathcal{P}_{1-j} h(\theta,\xi_1)\|^2 m(d\theta) = n\lambda_j^2.\end{aligned}$$

Note that $T_n(\theta) = T_{n,0}(\theta)$. Let $\Delta = \sum_{j=0}^{\infty} \lambda_j$. By the Cauchy-Schwarz Inequality,

$$\begin{aligned}\int_A \mathbb{E}|T_n(\theta) - \mathbb{E}[T_n(\theta)]|^2 m(d\theta) &= \int_A \mathbb{E}\Big\{\sum_{j=0}^{\infty} [T_{n,j}(\theta) - T_{n,j+1}(\theta)]\Big\}^2 m(d\theta) \\ &\leq \Delta \int_A \mathbb{E}\Big\{\sum_{j=0}^{\infty} \frac{[T_{n,j}(\theta) - T_{n,j+1}(\theta)]^2}{\lambda_j}\Big\} m(d\theta) = n\Delta^2,\end{aligned}$$

and (29) follows.

Lemma 3. *Let D_i be L^q ($q > 1$) martingale differences and $C_q = 18q^{3/2}(q-1)^{-1/2}$. Then*

$$\|D_1 + \cdots + D_n\|_q^r \leq C_q^r \sum_{i=1}^n \|D_i\|_q^r, \quad \textit{where} \quad r = \min(q, 2). \tag{30}$$

Proof. Let $M = \sum_{i=1}^n D_i^2$. By Burkholder's inequality, $\|\sum_{i=1}^n D_i\|_q \leq C_q \|M\|_{q/2}$. Then (30) easily follows by considering the cases $q > 2$ and $q \leq 2$ separately.

Lemma 4. (Wu (2005)) *Let $q > 1$ and Z_i, $1 \leq i \leq 2^d$, be random variables in $\mathcal{L}^q$, where d is a positive integer. Let $S_n = Z_1 + \cdots + Z_n$ and $S_n^* = \max_{i \leq n} |S_i|$. Then*

$$\|S_{2^d}^*\|_q \leq \sum_{r=0}^{d} \Big[\sum_{m=1}^{2^{d-r}} \|S_{2^r m} - S_{2^r(m-1)}\|_q^q\Big]^{\frac{1}{q}}. \tag{31}$$

5. Proofs of Theorems 1 and 2

To illustrate the idea behind our approach, let $\tilde{F}_n(x) = n^{-1}\sum_{i=1}^n F_\varepsilon(x|\xi_{i-1})$ be the conditional empirical distribution function and write

$$F_n(x) - F(x) = [F_n(x) - \tilde{F}_n(x)] + [\tilde{F}_n(x) - F(x)]. \quad (32)$$

The decomposition (32) has two important and useful properties. First, $n[F_n(x)-\tilde{F}_n(x)]$ is a martingale with stationary, ergodic and bounded martingale differences. Second, $\tilde{F}_n - F$ is differentiable with derivative $\tilde{f}_n(x) - f(x)$, where $\tilde{f}_n(x) = (\partial/\partial x)\tilde{F}_n(x) = n^{-1}\sum_{i=1}^n f_\varepsilon(x|\xi_i)$. The property of differentiability is useful in establishing tightness. The idea was applied in Wu and Mielniczuk (2002).

Following (32), let $G_n(s) = n^{1/2}[F_n(x) - \tilde{F}_n(x)]$ and $Q_n(s) = n^{1/2}[\tilde{F}_n(x) - F(x)]$. Then $R_n(s) = G_n(s) + Q_n(s)$. Sections 5.1 and 5.2 deal with G_n and Q_n, respectively. Theorems 1 and 2 are proved in Sections 5.3 and 5.4.

5.1. Analysis of G_n

The main result is this section is Lemma 7 which concerns the weak convergence of G_n.

Lemma 5. *Let $q > 2$ and $\alpha = \max(1, q/4) - q/2$. Then there is a constant C_q such that*

$$\|G_n(y)-G_n(x)\|_q^q \le C_q n^\alpha[F(y)-F(x)]+C_q(y-x)^{\frac{q}{2}-1}\int_x^y \mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du \quad (33)$$

holds for all $n \in \mathbb{N}$ and all $x < y$, and

$$\|G_n(x)\|_q^q \le C_q \min[F(x), 1 - F(x)]. \quad (34)$$

Proof. Let $q' = q/2$ and $p' = q'/(q'-1)$; let $d_i(s) = \mathbf{1}_{X_i \le s} - \mathbb{E}(\mathbf{1}_{X_i\le s}|\xi_{i-1})$, $d_i = d_i(y) - d_i(x)$, $D_i = d_i^2 - \mathbb{E}(d_i^2|\xi_{i-1})$, $K_n = \sum_{i=1}^n D_i$, and $L_n = \sum_{i=1}^n \mathbb{E}(d_i^2|\xi_{i-1})$. Then both (D_i) and (d_i) are martingale differences. By Burkholder's inequality (Chow and Teicher (1988)),

$$\begin{aligned}\|G_n(y) - G_n(x)\|_q^q &= n^{-\frac{q}{2}}\mathbb{E}(|d_1 + \cdots + d_n|^q)\\ &\le \frac{C_q}{n^{\frac{q}{2}}}\mathbb{E}[(d_1^2 + \cdots + d_n^2)^{\frac{q}{2}}] \le \frac{C_q}{n^{\frac{q}{2}}}(\|K_n\|_{q'}^{q'} + \|L_n\|_{q'}^{q'}). \end{aligned} \quad (35)$$

By Lemma 3,

$$\frac{\|K_n\|_{q'}^{q'}}{n^{\max(1,\frac{q'}{2})}} \le C_q\|D_1\|_{q'}^{q'} \le C_q 2^{q'-1}[\|d_1^2\|_{q'}^{q'} + \|\mathbb{E}(d_1^2|\xi_0)\|_{q'}^{q'}] \le C_q 2^{q'}\|d_1^2\|_{q'}^{q'}, \quad (36)$$

where we have applied Jensen's inequality in $\|\mathbb{E}(d_1^2|\xi_0)\|_{q'}^{q'} \le \|d_1^2\|_{q'}^{q'}$. Notice that $|d_1| \le 1$,

$$\|d_1^2\|_{q'}^{q'} \le \|d_1\|_{q'}^{q'} \le 2^{q'-1}[\|\mathbf{1}_{x\le X_i\le y}\|_{q'}^{q'} + \|\mathbb{E}(\mathbf{1}_{x\le X_i\le y}|\xi_0)\|_{q'}^{q'}] \le 2^{q'}[F(y)-F(x)]. \tag{37}$$

Since $\mathbb{E}(d_1^2|\xi_0) \le \mathbb{E}(\mathbf{1}_{x\le X_i\le y}|\xi_0)$ and $1/p' + 1/q' = 1$, we have by Hölder's Inequality that

$$\begin{aligned}\|L_n\|_{q'}^{q'} &\le n^{q'}\|\mathbb{E}(d_1^2|\xi_0)\|_{q'}^{q'} \le n^{q'}\mathbb{E}\Big\{\Big[\int_x^y f_\varepsilon(u|\xi_0)du\Big]^{q'}\Big\} \\ &\le n^{q'}\mathbb{E}\Big[(y-x)^{q'/p'}\int_x^y f_\varepsilon^{q'}(u|\xi_0)du\Big]. \end{aligned} \tag{38}$$

Combining (35), (36), (37) and (38), we have (33).

To show (34), in (35) we let $d_i = d_i(x) = \mathbf{1}_{X_i\le x} - \mathbb{E}(\mathbf{1}_{X_i\le x}|\xi_{i-1})$. Then

$$\mathbb{E}(|d_1+\cdots+d_n|^q) \le C_q n^{\frac{q}{2}}\|d_1\|_q^q \le C_q n^{\frac{q}{2}}\|d_1\|^2 \le C_q n^{\frac{q}{2}} F(x)[1-F(x)]$$

completes the proof.

Lemma 6. *Let $q > 2$ and $\alpha = \max(1, q/4) - q/2$. Then there exists a constant $C_q < \infty$ such that, for all $b > 0$, $a \in \mathbb{R}$ and $n, d \in \mathbb{N}$,*

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0\le s<b}|G_n(a+s)-G_n(a)|^q\Big] \\ &\quad\le C_q d^q n^\alpha[F(a+b)-F(a)] + C_q b^{\frac{q}{2}-1}[1+n^{\frac{q}{2}}2^{d(1-\frac{q}{2})}]\int_a^{a+b}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du. \end{aligned} \tag{39}$$

In particular, for $d = 1 + \lfloor(\log n)/[(1-2/q)\log 2]\rfloor$, we have

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\le s<b}|G_n(a+s)-G_n(a)|^q\Big] &\le C_q(\log n)^q n^\alpha[F(a+b)-F(a)] \\ &\quad + C_q b^{\frac{q}{2}-1}\int_a^{a+b}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du. \end{aligned} \tag{40}$$

Proof. Let $h = b2^{-d}$, $S_j = G_n(a+jh)-G_n(a)$ and $Z_j = S_j - S_{j-1}$. By Lemma 5,

$$\begin{aligned}\sum_{m=1}^{2^{d-r}}\|S_{2^r m} - S_{2^r(m-1)}\|_q^q &\le \sum_{m=1}^{2^{d-r}} C_q n^\alpha[F(a+2^r mh) - F(a+2^r(m-1)h)] \\ &\quad + \sum_{m=1}^{2^{d-r}} C_q(2^r h)^{\frac{q}{2}-1}\int_{a+2^r(m-1)h}^{a+2^r mh}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du \\ &= C_q n^\alpha[F(a+b)-F(a)] + C_q(2^r h)^{\frac{q}{2}-1}V, \end{aligned}$$

where $V = \int_a^{a+b} \mathbb{E}[f_\varepsilon^{q/2}(u|\xi_0)]du$. By Lemma 4,

$$\begin{aligned}\|S^*_{2^d}\|_q &\le \sum_{r=0}^{d}\{C_q n^\alpha [F(a+b)-F(a)]\}^{\frac{1}{q}} + \sum_{r=0}^{d}\{C_q (2^r h)^{\frac{q}{2}-1} V\}^{\frac{1}{q}} \\ &\le d\{C_q n^\alpha [F(a+b)-F(a)]\}^{\frac{1}{q}} + \{C_q (2^{d+1} h)^{\frac{q}{2}-1} V\}^{\frac{1}{q}}. \end{aligned} \tag{41}$$

Let $B_j = \sqrt{n}[\tilde{F}_n(a+jh) - \tilde{F}_n(a+(j-1)h)]$. Recall $\tilde{F}_n(x) = n^{-1}\sum_{i=1}^n F_\varepsilon(x|\xi_{i-1})$ and $\tilde{f}_n(x) = \tilde{F}_n'(x)$. Since $q' = q/2 > 1$, $\|\tilde{f}_n(x)\|_{q'} \le \|f_\varepsilon(x|\xi_0)\|_{q'}$. Note that $0 \le \tilde{F}_\varepsilon \le 1$, so by Hölder's Inequality,

$$\begin{aligned}\mathbb{E}\Big[\max_{j\le 2^d} B_j^q\Big] &\le \sum_{j=1}^{2^d}\mathbb{E}(B_j^q) = \sum_{j=1}^{2^d} n^{q'}\|\tilde{F}_n(a+jh) - \tilde{F}_n(a+(j-1)h)\|_q^q \\ &\le \sum_{j=1}^{2^d} n^{q'}\|\tilde{F}_n(a+jh) - \tilde{F}_n(a+(j-1)h)\|_{q'}^{q'} \\ &\le \sum_{j=1}^{2^d} n^{q'} h^{q'-1}\int_{a+(j-1)h}^{a+jh}\mathbb{E}[|\tilde{f}_n(x)|^{q'}]du \le n^{q'}h^{q'-1}V. \end{aligned} \tag{42}$$

Observe that

$$G_n(a+h\lfloor\frac{s}{h}\rfloor) - \max_{j\le 2^d} B_j \le G_n(a+s) \le G_n(a+h\lfloor\frac{s}{h}+1\rfloor) + \max_{j\le 2^d} B_j.$$

Hence (39) follows from (41), (42) and, since $S_j = G_n(a+jh) - G_n(a)$,

$$\begin{aligned}\sup_{0\le s<b}|G_n(a+s) - G_n(a)| &\le \sup_{0\le s<b}|G_n(a+h\lfloor\frac{s}{h}+1\rfloor) - G_n(a)| \\ &\quad + \sup_{0\le s<b}|G_n(a+h\lfloor\frac{s}{h}\rfloor) - G_n(a)| + 2\max_{j\le 2^d} B_j \\ &\le 2S^*_{2^d} + 2\max_{j\le 2^d} B_j \end{aligned}$$

by noticing that $h = 2^{-d}b$. For $d = 1 + \lfloor(\log n)/[(1-2/q)\log 2]\rfloor$, we have $n^{q/2}2^{d(1-q/2)} \le 1$, hence (40) is an easy consequence of (39).

Lemma 7. *Let $\gamma \ge 0$ and $q > 2$. Assume $\mathbb{E}[|X_1|^\gamma + \log(1+|X_1|)] < \infty$, and* (6). *Then* (i) $\mathbb{E}[\sup_{s\in\mathbb{R}}|G_n(s)|^q(1+|s|)^\gamma] = \mathcal{O}(1)$, *and* (ii) *the process* $\{G_n(s)(1+|s|)^{\gamma/q},\ s\in\mathbb{R}\}$ *is tight.*

Remark 5. In Lemma 7, the term $\log(1+|X_1|)$ is not needed if $\gamma > 0$.

Proof. (i) Without loss of generality we show that $\mathbb{E}[\sup_{s\ge 0}|G_n(s)|^q(1+|s|)^\gamma] = \mathcal{O}(1)$, since the case of $s < 0$ follows similarly. Let $\alpha_n = (\log n)^q n^{\max(1,q/4)-q/2}$.

By (6) and (40) of Lemma 6, with $a = b = 2^k$,

$$\begin{aligned}
&\sum_{k=1}^{\infty} 2^{k\gamma}\mathbb{E}\Big[\sup_{2^k\le s<2^{k+1}} |G_n(s) - G_n(2^k)|^q\Big] \\
&\quad \le C_q \sum_{k=1}^{\infty} 2^{k\gamma}\alpha_n[F(2^{k+1}) - F(2^k)] + C_q \sum_{k=1}^{\infty} 2^{k\gamma}(2^k)^{\frac{q}{2}-1}\int_{2^k}^{2^{k+1}} \mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du \\
&\quad \le C_{\gamma,q}\alpha_n \int_2^\infty f(u)(1+u)^\gamma du + C_{\gamma,q}\int_2^\infty (1+u)^\gamma u^{\frac{q}{2}-1}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du \\
&\quad \le C_{\gamma,q}\alpha_n + C_{\gamma,q} = \mathcal{O}(1). \qquad (43)
\end{aligned}$$

Observe that the function $\ell(x) = \sum_{k=1}^{\infty} 2^{k\gamma}\mathbf{1}_{2^k\le x}$, $x > 0$, is bounded by $C_\gamma[x^\gamma + \log(1+x)]$, where C_γ is a constant. Then by (34) of Lemma 5, we have

$$\sum_{k=1}^{\infty} 2^{k\gamma}\|G_n(2^k)\|_q^q \le \sum_{k=1}^{\infty} 2^{k\gamma} C\mathbb{E}(\mathbf{1}_{2^k\le X_1}) \le C\mathbb{E}[|X_1|^\gamma + \log(1+|X_1|)] < \infty. \qquad (44)$$

Simple calculations show that (i) follows from (43), (44) and (40), with $a = 0$ and $b = 2$.

(ii) It is easily seen that the argument in (i) entails

$$\lim_{r\to\infty}\limsup_{n\to\infty}\mathbb{E}\Big[\sup_{|s|>r}|G_n(s)|^q(1+|s|)^\gamma\Big] = 0. \qquad (45)$$

Let $\delta \in (0,1)$. For $s, t \in [-r, r]$ with $0 \le s - t \le \delta$, we have

$$\begin{aligned}
&|G_n(s)(1+|s|)^{\frac{\gamma}{q}} - G_n(t)(1+|t|)^{\frac{\gamma}{q}}| \\
&\quad \le |(1+|s|)^{\frac{\gamma}{q}}[G_n(s) - G_n(t)]| + |G_n(t)[(1+|s|)^{\frac{\gamma}{q}} - (1+|t|)^{\frac{\gamma}{q}}]| \\
&\quad \le (1+r)^{\frac{\gamma}{q}}|G_n(s) - G_n(t)| + C_{r,\gamma,q}\delta \sup_{u\in[-r,r]}|G_n(u)|. \qquad (46)
\end{aligned}$$

By (i), $\|\sup_{u\in\mathbb{R}}|G_n(u)|\| = \mathcal{O}(1)$. Let $I_k = I_k(\delta) = [k\delta, (k+1)\delta]$. By Lemma 6,

$$\begin{aligned}
&\sum_{k=-\lfloor\frac{r}{\delta}\rfloor-1}^{\lfloor\frac{r}{\delta}\rfloor+1} \mathbb{P}\Big[\sup_{s\in I_k}|G_n(s) - G_n(k\delta)| > \epsilon\Big] \\
&\quad \le \epsilon^{-q}\sum_{k=-\lfloor\frac{r}{\delta}\rfloor-1}^{\lfloor\frac{r}{\delta}\rfloor+1}\Big\{C_q\alpha_n\mathbb{P}(X_1\in I_k) + C_q\delta^{\frac{q}{2}-1}\int_{I_k}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du\Big\} \\
&\quad \le \epsilon^{-q}C_q\alpha_n + \epsilon^{-q}C_q\delta^{\frac{q}{2}-1}\int_{\mathbb{R}}\mathbb{E}[f_\varepsilon^{\frac{q}{2}}(u|\xi_0)]du.
\end{aligned}$$

By (6), $\int_{\mathbb{R}} \mathbb{E}[f_\varepsilon^{q/2}(u|\xi_0)]du < \infty$. Hence

$$\limsup_{n\to\infty} \mathbb{P}\Big[\sup_{s,t\in[-r,r],\ 0\le s-t\le\delta} |G_n(s)-G_n(t)| > 2\epsilon\Big] \le \epsilon^{-q} C_q \delta^{\frac{q}{2}-1},$$

which implies the tightness of $\{G_n(s),\ |s|\le r\}$ for fixed r. So (45) and (46) entail (ii).

5.2. Analysis of Q_n

It is relatively easier to handle Q_n since it is a differentiable function. The Hardy-type inequalities (cf Lemma 1) are applicable.

Lemma 8. *Let $\gamma' \ge 0$ and assume* (7). *Then* (i) $\mathbb{E}[\sup_{s\in\mathbb{R}} |Q_n(s)|^2(1+|s|)^{\gamma'}] = \mathcal{O}(1)$, *and* (ii) *the process $\{Q_n(s)(1+|s|)^{\gamma'/2},\ s\in\mathbb{R}\}$ is tight.*

Proof. Let $r \ge 0$ and recall $0\le\mu\le 1$. (i) By Lemma 1,

$$\Lambda_r := \sup_{|s|\ge r} [Q_n^2(s)(1+|s|)^{\gamma'}] \le C\int_{|s|\ge r} Q_n^2(s) w_{\gamma'-\mu}(ds) + C\int_{|s|\ge r} [Q_n'(s)]^2 w_{\gamma'+\mu}(ds)$$

holds for some constant $C = C_{\gamma',\mu}$. By Lemma 1,

$$\frac{\|\Lambda_r^{\frac{1}{2}}\|}{\sqrt{C}} \le \sum_{j=0}^\infty \sqrt{\int_{|s|\ge r} \|\mathcal{P}_0 F_\varepsilon(\theta|\xi_j)\|^2 w_{\gamma'-\mu}(d\theta)} + \sum_{j=0}^\infty \sqrt{\int_{|s|\ge r} \|\mathcal{P}_0 f_\varepsilon(\theta|\xi_j)\|^2 w_{\gamma'+\mu}(d\theta)}. \quad (47)$$

So (i) follows by letting $r=0$ in (47).

(ii) The argument in (ii) of Lemma 7 in applicable. Let $0<\delta<1$. Then

$$\begin{aligned}
\Psi_{n,r}(\delta) &:= \sup_{s,t\in[-r,r],\ 0\le s-t\le\delta} |Q_n(s)(1+|s|)^{\frac{\gamma'}{2}} - Q_n(t)(1+|t|)^{\frac{\gamma'}{2}}| \\
&\le \sup_{s,t\in[-r,r],\ 0\le s-t\le\delta} |(1+|s|)^{\frac{\gamma'}{2}}[Q_n(s)-Q_n(t)]| \\
&\quad + \sup_{s,t\in[-r,r],\ 0\le s-t\le\delta} |Q_n(t)[(1+|s|)^{\frac{\gamma'}{2}} - (1+|t|)^{\frac{\gamma'}{2}}]| \\
&\le C_{r,\gamma'}\delta \sup_{u\in[-r,r]} |Q_n'(u)| + C_{r,\gamma'}\delta \sup_{u\in[-r,r]} |Q_n(u)|.
\end{aligned}$$

By (27), Lemma 2, and (7), there exists a constant $C = C(r,\gamma',\mu,\nu)$ such that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{|s|\le r} |Q_n'(s)|^2\Big] &\le C\int_{-r}^r \|Q_n'(s)\|^2 w_{\gamma'+\mu}(ds) + C\int_{-r}^r \|Q_n''(s)\|^2 w_{-\nu}(ds) \\
&\le C_{\gamma'}\sigma^2(f_\varepsilon, w_{\gamma'+\mu}) + C_{\gamma'}\sigma^2(f_\varepsilon', w_{-\nu}) = \mathcal{O}(1).
\end{aligned}$$

By (i), there exists $C_1 < \infty$ such that for all $n \in \mathbb{N}$, $\mathbb{E}[\Psi_{n,r}^2(\delta)] \le \delta^2 C_1$. Notice that the upper bound in (47) goes to 0 as $r \to \infty$. Hence (ii) obtains.

5.3. Proof of Theorem 1.

We need to verify finite-dimensional convergence and tightness. Let $j \ge 0$. Observe that $(\partial/\partial\theta)\mathcal{P}_0 F_\varepsilon(\theta|\xi_j) = \mathcal{P}_0 f_\varepsilon(\theta|\xi_j)$ and $\|\mathcal{P}_0 F_\varepsilon(\theta|\xi_j)\| = \|\mathcal{P}_0 \mathbf{1}_{X_{j+1}\le\theta}\|$. By Lemma 1(i)

$$\sup_{\theta\in\mathbb{R}} \|\mathcal{P}_0 F_\varepsilon(\theta|\xi_j)\| \le C\sqrt{\int_{\mathbb{R}} \|\mathcal{P}_0 F_\varepsilon(\theta,\xi_j)\|^2 w_{-\mu}(d\theta)} + C\sqrt{\int_{\mathbb{R}} \|\mathcal{P}_0 f_\varepsilon(\theta,\xi_j)\|^2 w_{\mu}(d\theta)}$$

which, by (7) and (5), implies $\sum_{i=0}^{\infty} \|\mathcal{P}_0 \mathbf{1}_{X_i\le\theta}\| < \infty$. Hence by Theorem 1(i) in Hannan (1973), $R_n(\theta)$ is asymptotically normal. The finite-dimensional convergence easily follows. Since $R_n(s) = G_n(s) + Q_n(s)$, the tightness and (i) follow from Lemmas 7 and 8.

Since $\mathbb{E}[f_\varepsilon(u|\xi_0)] = f(u)$, (8) and the moment condition $X_1 \in \mathcal{L}^{\gamma+q/2-1}$ imply (6).

5.4. Proof of Theorem 2.

Let $\Theta_n(a,\delta) = \sup_{0\le s<\delta} |G_n(a+s) - G_n(a)|$ and $\alpha = \max(1, q/4) - q/2$. Note that $(\log n)^q n^\alpha = \mathcal{O}(\delta_n^{q/2-1})$. By (40) of Lemma 6, we have, uniformly in a, that

$$\begin{aligned}\mathbb{E}\Big[\Theta_n^q(a,\delta_n)\Big] &\le C_q(\log n)^q n^\alpha [F(a+\delta_n) - F(a)] + C_q \delta_n^{\frac{q}{2}-1} \tau^{\frac{q}{2}-1} \int_a^{a+\delta_n} f(u)du \\ &\le C\delta_n^{\frac{q}{2}-1}[F(a+\delta_n) - F(a)].\end{aligned}$$

Here the constant C only depends on τ, γ, q and $\mathbb{E}(|X_1|^\gamma)$. Hence

$$\begin{aligned}\sum_{k\in\mathbb{Z}} (1+|k\delta_n|)^\gamma \mathbb{E}\Big[\Theta_n^q(k\delta_n,\delta_n)\Big] &\le \sum_{k\in\mathbb{Z}} (1+|k\delta_n|)^\gamma C\delta_n^{\frac{q}{2}-1}[F(k\delta_n+\delta_n) - F(k\delta_n)] \\ &\le C\delta_n^{\frac{q}{2}-1}\mathbb{E}[(1+|X_1|)^\gamma].\end{aligned}$$

Let $I_k(\delta) = [k\delta, (1+k)\delta]$ and $c_\delta = \sup_{|u-t|\le\delta}[(1+|t|)/(1+|u|)]$. Then $c_\delta \le 2$ since $\delta < 1/2$. By the inequality $|G_n(a) - G_n(c)| \le |G_n(a) - G_n(b)| + |G_n(b) - G_n(c)|$,

$$\sup_{t\in I_k(\delta_n),\ 0\le s<\delta_n} |G_n(t+s) - G_n(t)| \le 2 \sup_{0\le u<2\delta_n} |G_n(k\delta_n + u) - G_n(k\delta_n)|.$$

Therefore,

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in\mathbb{R}} (1+|t|)^\gamma \Theta_n^q(t,\delta_n)\Big] &\le \sum_{k\in\mathbb{Z}} \mathbb{E}\Big[\sup_{t\in I_k(\delta_n)} (1+|t|)^\gamma \Theta_n^q(t,\delta_n)\Big] \\ &\le 2^q c_{\delta_n}^\gamma \sum_{k\in\mathbb{Z}} (1+|k\delta_n|)^\gamma \mathbb{E}\Big[\Theta_n^q(k\delta_n, 2\delta_n)\Big] \le C\delta_n^{\frac{q}{2}-1}.\end{aligned}$$

Note that $R_n(s) = G_n(s) + Q_n(s)$. Then (12) follows if it holds with R_n replaced by G_n and Q_n, respectively. The former is an easy consequence of the preceding inequality and Jensen's Inequality. To show that (12) holds with R_n replaced by Q_n, recall $\gamma' = 2\gamma/q$. By (24) of Lemma 1 and Lemma 2,

$$\mathbb{E}\Big[\sup_{x\in\mathbb{R}}(1+|x|)^{\gamma'}|Q_n'(x)|^2\Big] \le C\int_{\mathbb{R}} \|Q_n'(x)\|^2 w_{\gamma'}(dx) + C\int_{\mathbb{R}} \|Q_n''(x)\|^2 w_{\gamma'}(dx)$$
$$\le C\sigma^2(f_\varepsilon, w_{\gamma'}) + C\sigma^2(f_\varepsilon', w_{\gamma'}) < \infty,$$

which completes the proof in view of the fact that

$$(1+|t|)^{\gamma'}\sup_{|s|\le\delta_n}|Q_n(t+s)-Q_n(t)|^2 \le \delta_n^2(1+|t|)^{\gamma'}\sup_{|s|\le\delta_n}|Q_n'(t+s)|^2$$
$$\le c_{\delta_n}^{\gamma'}\delta_n^2 \sup_{u:|u-t|\le\delta_n}[(1+|u|)^{\gamma'}|Q_n'(u)|^2].$$

Remark 6. It is worthwhile to note that the modulus of continuity of G_n has the order $\delta_n^{1-2/q}$, while that of Q_n has the higher order δ_n.

6. Proof of Theorem 3.

(i) If (17) holds, then $\sigma(h,m) \le 2\Xi_{h,m}$. To prove (17), let $Z_n(\theta) = h_n(\theta,\xi_0) - h_n(\theta,\xi_0^*)$, $h(\theta,Y_n) = h(\theta,\xi_n)$, and $V_n(\theta) = h(\theta,Y_n^*) - h(\theta,Y_n) = \int_{Y_n}^{Y_n^*}\frac{\partial}{\partial y}h(\theta,y)dy$. Let $\lambda(y) = [H_{h,m}(y)]^{1/2}$ and $U = \int_{Y_n}^{Y_n^*}\lambda(y)dy$. By the Cauchy-Schwarz Inequality,

$$\int_{\mathbb{R}} V_n^2(\theta)m(d\theta) \le \int_{\mathbb{R}}\Big[\int_{Y_n}^{Y_n^*}\frac{1}{\lambda(y)}\Big|\frac{\partial}{\partial y}h(\theta,y)\Big|^2 dy \times \int_{Y_n}^{Y_n^*}\lambda(y)dy\Big]m(d\theta) = U^2.$$

Note that $\mathbb{E}[h(\theta,Y_n)|\xi_{-1}] = \mathbb{E}[h(\theta,Y_n^*)|\xi_0]$. By (4), $\|Z_n(\theta)\| \le 2\|\mathcal{P}_0 h(\theta,Y_n)\| \le 2\|V_n(\theta)\|$. So we have (17). (ii) Let $\delta = q/2-1$ and $W = \int_{Y_n}^{Y_n^*} w_\delta(dy)$. If $q<0$, then $\int_{\mathbb{R}} w_\delta(dy) = -4/q$ and $|W| \le \min(|Y_n - Y_n^*|, -4/q)$. If $\delta > 0$, by Hölder's Inequality,

$$\|W\| \le \|[(1+|Y_n|)^\delta + (1+|Y_n^*|)^\delta](Y_n - Y_n^*)\|$$
$$\le \|(1+|Y_n|)^\delta + (1+|Y_n^*|)^\delta\|_q^\delta \|Y_n - Y_n^*\|_q = O(\|Y_n - Y_n^*\|_q).$$

For the case $-1 < \delta \le 0$, we need to prove the inequality $|\int_v^u w_\delta(dy)| \le 2^{-\delta}|u-v|^{1+\delta}/(1+\delta)$. For the latter, it suffices to consider cases (a) $u \ge v \ge 0$, and (b) $u \ge 0 \ge v$. For (a),

$$\int_v^u w_\delta(dy) = \frac{(1+u)^{1+\delta} - (1+v)^{1+\delta}}{1+\delta} \le \frac{(u-v)^{1+\delta}}{1+\delta}.$$

For (b), let $t = (u - v)/2$. Then

$$\int_v^u w_\delta(dy) = \frac{(1+u)^{1+\delta} - 1 + (1+|v|)^{1+\delta} - 1}{1+\delta} \le \frac{2(1+t)^{1+\delta} - 2}{1+\delta} \le \frac{2t^{1+\delta}}{1+\delta}.$$

Therefore $\|W\| = O(\||Y_n - Y_n^*|^{\delta+1}\|) = O(\|Y_n - Y_n^*\|_q^{q/2})$.

Acknowledgements

The author is grateful to three referees for their valuable comments. The author also thanks Professors Sándor Csörgő and Jan Mielniczuk for useful suggestions. The work was supported in part by NSF grant DMS-0478704.

References

Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.* **62**, 119-132.

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

Berkes, I. and Horváth, L. (2001). Strong approximation of the empirical process of GARCH sequences. *Ann. Appl. Probab.* **11**, 789-809.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory.* 2nd edition. Springer, New York.

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31-85.

Csörgő, M. and Yu, H. (1996). Weak approximations for quantile processes of stationary sequences. *Canad. J. Statist.* **24**, 403-430.

Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields* **104**, 15-25.

Dehling, H., Mikosch, T. and Sørensen, M. (eds), (2002). *Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston.

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**, 1767-1786.

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 41-76.

Doukhan, P. (2003). Models, inequalities, and limit theorems for stationary sequences. In *Theory and applications of long-range dependence* (Edited by P. Doukhan, G. Oppenheim and M. S. Taqqu), 43-100, Birkhäuser, Boston.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 313-342.

Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **31**, 393-427.

Doukhan, P. and Surgailis, D. (1998). Functional central limit theorem for the empirical process of short memory linear processes. *C. R. Acad. Sci. Paris Ser. I Math.* **326**, 87-92.

Einmahl, J. H. J. and Mason, D. M. (1988). Strong limit theorems for weighted quantile processes. *Ann. Probab.* **16**, 1623-1643.

Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Process. Appl.* **34**, 39-47.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

Gastwirth, J. L. and Rubin, H. (1975). The asymptotic distribution theory of the empiric cdf for mixing stochastic processes. *Ann. Statist.* **3**, 809-824.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189-196.

Hannan, E. J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Gebiete* **26**, 157-170.

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.

Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.* **32**, 1600-1631.

Jarner, S. and Tweedie, R. (2001). Locally contracting iterated random functions and stability of Markov chains. *J. Appl. Probab.* **38**, 494-507.

Mehra, K. L. and Rao, M. S. (1975). Weak convergence of generalized empirical processes relative to d_q under strong mixing. *Ann. Probab.* **3**, 979-991.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

Opic, B. and Kufner, A. (1990). *Hardy-type Inequalities.* Longman Scientific & Technical. Wiley, New York.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19**, 297-303.

Rio, E. (2000). *Theorie Asymptotique des Processus Aleatoires Faiblement Dependants.* Mathématiques et Applications **31**. Springer, Berlin.

Shao, Q. M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24**, 2098-2127.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Tsay, R. S. (2005). *Analysis of Financial Time Series.* Wiley, New York.

Withers, C. S. (1975). Convergence of empirical processes of mixing rv's on [0, 1]. *Ann. Statist.* **3**, 1101-1108.

Withers, C. S. (1981). Conditions for linear process to be strongly mixing. *Z. Wahrsch. Verw. Gebiete* **57**, 477-480.

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

Wu, W. B. (2005). A strong convergence theory for stationary processes. Preprint.

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30**, 1441-1459.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41**, 425-436.

Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, U.S.A.

E-mail: wbwu@galton.uchicago.edu

(Received September 2005; accepted May 2006)