ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

On linear processes with dependent innovations

Wei Biao Wu*, Wanli Min[1]

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

Received 11 December 2003; received in revised form 7 December 2004; accepted 3 January 2005

Abstract

We consider asymptotic behavior of partial sums and sample covariances for linear processes whose innovations are dependent. Central limit theorems and invariance principles are established under fairly mild conditions. Our results go beyond earlier ones by allowing a quite wide class of innovations which includes many important nonlinear time series models. Applications to linear processes with GARCH innovations and other nonlinear time series models are discussed.

Keywords: Central limit theorem; Covariance; GARCH model; Invariance principle; Linear process; Nonlinear time series

1. Introduction

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be independent and identically distributed (i.i.d.) random elements and F be a measurable function such that

$$a_t = F(\ldots, \varepsilon_{t-1}, \varepsilon_t) \tag{1}$$

is a well-defined random variable. Then $\{a_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic process. Assume throughout the paper that a_t has mean 0, finite variance and that $\{\psi_i\}_{i\geqslant 0}$ is a sequence of real numbers such that $\sum_{i,j=0}^{\infty} |\psi_i\psi_j E(a_0 a_{i-j})| < \infty$. Then the

*Corresponding author. Tel.: +1 7737020958; fax: 1 7737029810.

E-mail addresses: wbwu@galton.uchicago.edu (Wei Biao Wu), wanlimin@us.ibm.com (Wanli Min).

[1]Present adress: IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA.

0304-4149/$ - see front matter

doi:10.1016/j.spa.2005.01.001

linear process

$$X_t = \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{2}$$

exists almost surely and $E(X_t^2) < \infty$. Our goal is to obtain a limit theory for the partial sum process $S_n = \sum_{i=1}^n X_i$ and the sample covariances $\hat{\gamma}_h = \sum_{t=1}^n X_t X_{t+h}/n$, $h \geqslant 0$. We will establish an invariance principle for the former and a central limit theorem (CLT) for $\hat{\gamma}_h$. Such results are needed in the related statistical inference.

In the classical time series analysis, the innovations a_t in the linear process X_t are often assumed to be i.i.d.; see for example [7,9]. In this case asymptotic properties of the partial sums have been extensively studied. It would be hard to compile a complete list. We only mention some representatives: Davydov [14], Gorodetskii [19], Hall and Heyde [20], Phillips and Solo [42], Yokoyama [52] and Hosking [29]. See references therein for further background. There are basically two types of results. If the coefficients ψ_t are absolutely summable, then the covariances of X_t are summable and we say that X_t is *short-range dependent* (SRD). Under SRD, the normalizing constant for the sum $\sum_{i=1}^n X_i$ is $\sqrt{n}$, which is of the same order as that in the classical CLT for i.i.d. observations. We generically say that $\{X_k\}$ is long-range dependent (LRD) if its covariances are not absolutely summable. A particularly interesting example is that $\psi_k = \ell(k)/k^\beta$, $k \geqslant 1$, where $\frac{1}{2} < \beta < 1$ and ℓ is a slowly varying function, namely $\lim_{n\to\infty} \ell(\lambda n)/\ell(n) = 1$ for all $\lambda > 0$ (cf. [17, p. 275]). Fractional autoregressive integrated moving average model (FARIMA, [28]) is an important class for LRD processes. Asymptotic normality for sample covariances has also been widely discussed; see for example, [7,9,18,20,22,29,42].

The asymptotic problem of partial sums and sample covariances becomes more difficult if dependence among a_t is allowed. Recently, FARIMA processes with GARCH innovations have been proposed to model econometric time series. The former feature allows LRD and the latter one allows that the conditional variance can change over time, namely heteroscedasticity. Financial time series often exhibit these two features. Hence, FARIMA models with GARCH innovations provide a natural vehicle for modelling processes with both features; see [2,25,32,34].

Romano and Thombs [43] point out that the traditional large sample inference on autocorrelations under the assumption of i.i.d. innovations is misleading if the underlying $\{a_t\}$ are actually dependent. Results so far obtained in this direction require that a_t are m-dependent [16] or martingale differences [24]. Recently, Wang et al. [45] considered invariance principles for i.i.d. or martingale differences a_t. However, it seems that the proof in the latter paper is not rigorous; see Remark 3. Chung [13] and He [26] considered linear processes with martingale difference innovations having constant conditional variance [cf. (16)], which is a quite restrictive assumption that excludes the widely used ARCH models.

The paper has two goals. The first goal is to obtain asymptotic distributions of S_n and $\hat{\gamma}_h$, while the second one is to introduce another type of dependence structure which is useful for asymptotic problems in econometrics time series analysis. With our dependence structure, martingales can be constructed to approximate the original sequences so that martingale theory can be applied; see [51]. The proposed dependence structure only involves the computation of conditional moments and it is easily verifiable. This feature is quite different from strong mixing conditions which might be too restrictive and hard to be verified. Our results go beyond earlier ones by allowing a large class of nonlinear processes, which substantially relaxes the i.i.d. or martingale differences assumptions. In particular, our conditions are satisfied if $\{a_t\}$ are GARCH, random coefficient AR, bilinear AR and threshold AR models etc under suitable conditions on model parameters. Recently, Wu and Mielniczuk [49], Hsing and Wu [30] and Wu [47,48] apply the idea of martingale approximations to some asymptotic problems.

The paper is organized as follows. Main results are presented in Section 2 and proved in Section 4. Applications to some nonlinear processes are given in Section 3.

2. Results

We first introduce some notation. Let $\{X_t\}$ be the linear process defined by (2) and recall $E(a_n) = 0$; let $\mathscr{F}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$ be the shift process. For a random variable ξ write $\xi \in \mathscr{L}^p$ $(p>0)$ if $\|\xi\|_p := [E(|\xi|^p)]^{1/p} < \infty$ and $\|\cdot\| = \|\cdot\|_2$. Define the projections $\mathscr{P}_t$ by $\mathscr{P}_t\xi = E(\xi|\mathscr{F}_t) - E(\xi|\mathscr{F}_{t-1})$, $\xi \in \mathscr{L}^1$. For two sequences of real numbers $\{c_n\}$ and $\{d_n\}$, we write $c_n \sim d_n$ if $\lim_{n\to\infty} c_n/d_n = 1$. For $n \geqslant 0$ let $A_n = \sum_{i=n}^{\infty} \psi_i^2$, $\Psi_n = \sum_{i=0}^{n} \psi_i$, $B_n^2 = \sum_{i=0}^{n-1} \Psi_i^2$ and

$$\sigma_n^2 = \sum_{i=-n}^{\infty} (\Psi_{n+i} - \Psi_i)^2, \text{ where } \Psi_k = 0 \quad \text{if } k<0. \tag{3}$$

A weak dependence condition based on $\mathscr{P}_t$ is introduced in Section 2.1. Section 2.2 presents invariance principles of the partial sum $S_k = \sum_{i=1}^{k} X_i$. In particular, it deals with the asymptotic behavior of the random function $W_n(t), 0 \leqslant t \leqslant 1$, which is continuous and piece-wise linear such that $W_n(t) = S_k/\|S_n\|$ at $t = k/n$, $k = 0, \ldots, n$, and $W_n(t) = S_k/\|S_n\| + (nt-k)X_{k+1}/\|S_n\|$ when $k/n \leqslant t \leqslant (k+1)/n$. A central limit theorem for sample covariances $n^{-1}\sum_{t=1}^{n} X_t X_{t+h}$ is given in Section 2.3.

2.1. $\mathscr{L}^p$ weak dependence

The $\mathscr{L}^p$ weak dependence condition is given in Definition 1. Unlike strong mixing conditions, it only involves conditional moments. In Section 3, we argue that many nonlinear time series models satisfy this condition. It provides a natural vehicle for the central limit theory for stationary processes; see [21,46].

Definition 1. The process $Y_n = g(\mathscr{F}_n)$, where g is a measurable function, is said to be *$\mathscr{L}^p$ weakly dependent with order r* ($p \geqslant 1$ and $r \geqslant 0$) if $E(|Y_n|^p) < \infty$ and

$$\sum_{n=1}^{\infty} n^r \|\mathscr{P}_1 Y_n\|_p < \infty. \tag{4}$$

If (4) holds with $r = 0$, then Y_n is said to be *$\mathscr{L}^p$ weakly dependent*.

The intuition of $\mathscr{L}^p$ weak dependence is that the projection of the "future" Y_n to the space $\mathscr{M}_1 \ominus \mathscr{M}_0 = \{Z \in \mathscr{L}^p\colon\ Z$ is $\mathscr{F}_1$ measurable and $E(Z|\mathscr{F}_0) = 0\}$ has a small magnitude, namely the future depends weakly on the current states. If Y_n are martingale differences, then (4) is automatically satisfied if $Y_0 \in \mathscr{L}^p$. Verification of (4) for nonlinear time series is discussed in Lemma 2 and Proposition 2. Hannan [21, p. 159] discussed the special case $r = 0$ and $p = 2$ which implies that $n^{-1/2}\sum_{i=1}^n [Y_i - E(Y_1)] \Rightarrow N(0, \sigma^2)$ for some $\sigma^2 < \infty$. Condition (4) together with the causality structure of $\{a_t\}$ and the linearity structure of $\{X_n\}$ provide a natural vehicle for the central limit theory.

Lemma 1. *Assume that* $\{a_n\}$ *defined by* (1) *is* $\mathscr{L}^2$ *weakly dependent and*

$$\sum_{i=0}^{\infty} |\psi_i| < \infty. \tag{5}$$

Then $\sum_{t=1}^{\infty} \|\mathscr{P}_1 X_t\| < \infty$ *and* $\sum_{t=1}^{n} X_t/\sqrt{n} \Rightarrow N(0, \|\xi\|^2)$, *where* $\xi = \sum_{t=1}^{\infty} \mathscr{P}_1 X_t$.

Proof of Lemma 1. Since $\mathscr{P}_1 a_k = 0$ for $k \leqslant 0$, by (5),

$$\sum_{t=1}^{\infty} \|\mathscr{P}_1 X_t\| \leqslant \sum_{t=1}^{\infty}\sum_{i=0}^{\infty} |\psi_i| \|\mathscr{P}_1 a_{t-i}\| = \sum_{i=0}^{\infty} |\psi_i| \sum_{t=i+1}^{\infty} \|\mathscr{P}_1 a_{t-i}\| < \infty.$$

By Theorem 1(i) in Hannan [21], the lemma follows. □

2.2. Invariance principles

Invariance principle is a useful tool in statistical inference of econometric time series such as unit root testing problems, and it enables one to obtain limiting distributions for many statistics. It has a substantial history. The celebrated Donsker's theorem asserts invariance principles for i.i.d. sequence of X_n. For dependent sequences, see the survey by Bradley [8] and Peligrad [41] for strong mixing processes, McLeish [37,38] for mixingales. Other contributions are given in Billingsley [3] and Hall and Heyde ([20, abbreviated as HH hereafter]). In the classical theory of invariance principles for linear processes, it is often assumed that innovations $\{a_n\}$ are i.i.d. or martingale differences; see also the works of Davydov [14], Gorodetskii [19], HH [20, pp. 146] and Wang, Lin and Gulati ([45, abbreviated as WLG hereafter]) among others. Here our goal is to establish invariance principles for linear processes with innovations being weakly dependent in the sense of (4).

Let $\mathscr{C}[0,1]$ be the collection of continuous functions on $[0,1]$. For $f,\ g \in \mathscr{C}[0,1]$ define the distance $\rho(f,g) = \sup_{0 \leqslant t \leqslant 1} |f(t) - g(t)|$. Billingsley [3] provided a convergence theory on $\mathscr{C}[0,1]$. Recall $\Psi_n = \sum_{i=0}^n \psi_i$ for $n \geqslant 0$, $\Psi_n = 0$ for $n < 0$ and $B_n^2 = \sum_{i=0}^{n-1} \Psi_i^2$. Theorems 1 and 2 show two quite different asymptotic behaviors of the normalized processes W_n. Denote by W the standard Brownian motion and $W^H = \{W^H(t):\ t \geqslant 0\}$ the fractional Brownian motion (fBm) with Hurst index $H \in (0,1)$, which is a Gaussian process with zero mean and covariance function $E[W^H(t)W^H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}),\ s,t \geqslant 0$. See [36] for more details. Theorem 1 shows that, under (6), $W_n \Rightarrow W$ in $(\mathscr{C}[0,1], \rho)$ with the norming sequence $\|S_n\|$having the form $\ell^*(n)\sqrt{n}$ for some slowly varying function $\ell^*(n)$. In Theorem 2 we assume that the coefficients have the form $\psi_n = \ell(n)/n^\beta$ for $n \geqslant 1$, where $\frac{1}{2} < \beta < 1$ and ℓ is a slowly varying function. Then (6) is violated, the norming sequence $\|S_n\| \sim c_\beta n^{3/2-\beta}\ell(n)$ and $W_n \Rightarrow W^H$, which no longer have independent increments, while Brownian motions do have.

Theorem 1. *Assume that* $\{a_n\}$ *is* $\mathscr{L}^\alpha$ $(\alpha > 2)$ *weakly dependent with order* 1,

$$\sum_{i=0}^{\infty} (\Psi_{n+i} - \Psi_i)^2 = \mathrm{o}(B_n^2) \tag{6}$$

and $B_n \to \infty$. *Then* $\ell^*(n) := \|S_n\|/\sqrt{n}$ *is slowly varying*, $B_n/\sqrt{n} \sim \ell^*(n) \sim |\sum_{i=0}^{n-1} \Psi_i|/n$ *and* $W_n \Rightarrow W$ *in* $(\mathscr{C}[0,1], \rho)$.

Theorem 2. *Let* $\psi_n = \ell(n)/n^\beta$ *for* $n \geqslant 1$, *where* $\frac{1}{2} < \beta < 1$. *Assume that* a_n *is* $\mathscr{L}^2$ *weakly dependent with order* 1. *Then* $W_n \Rightarrow W^H$ *in* $(\mathscr{C}[0,1], \rho)$ *with Hurst index* $H = \frac{3}{2} - \beta$ *and* $\|S_n\| \sim n^{3/2-\beta}\ell(n)c_\beta\|\sum_{t=1}^{\infty} \mathscr{P}_1 a_t\|$, *where* $c_\beta = \{\int_0^\infty [x^{1-\beta} - \max(x-1,0)^{1-\beta}]^2\,\mathrm{d}x\}^{1/2}/(1-\beta)$.

Remark 1. In the case that $\{a_n\}$ is a stationary sequence of martingale differences with respect to the filter $\mathscr{F}_n$, Wu and Woodroofe [51] show that (6) with $B_n \to \infty$ is a necessary and sufficient condition for the conditional central limit theorem

$$E\{\Delta[\Phi, P(S_n^* \leqslant \cdot\,|\mathscr{F}_0)]\} \to 0,$$

where Φ is the standard normal distribution function, $S_n^* = S_n/\|S_n\|$ and Δ is the Levy distance between two distribution functions (see Example 1 therein).

Remark 2. The moment condition $a_n \in \mathscr{L}^\alpha$ with $\alpha > 2$ cannot be weakened to $a_n \in \mathscr{L}^2$. The following example is constructed based on Example 3 in [51]. Let a_t be i.i.d. symmetric innovations with $P(a_t \geqslant y) \sim y^{-2}(\log y)^{-3/2}$ as $y \to \infty$ and $E(a_t^2) = 1$; let $\phi_0 = \phi_1 = 0$, $\phi_2 = 1/\log 2$ and $\phi_k = 1/\log k - 1/\log(k-1)$, $k \geqslant 3$; let $X_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$ and $X'_t = a_t - a_{t-1} + X_t$; let $S'_k = \sum_{i=1}^{k} X'_k$ and W'_n be W_n withX_t replaced by X'_t; let $\tau_n = \sqrt{n}/\log n$. Elementary calculations show that (6) holds, $\|S_n\| \sim \tau_n$ and $\|S_n - S'_n\| = \mathrm{O}(1)$. W_n and W'_n cannot both converge to W. If so, then $\max_{k \leqslant n}|S_k| = \mathrm{O_P}(\tau_n)$ and $\max_{k \leqslant n}|S'_k| = \mathrm{O_P}(\tau_n)$. Hence $\max_{k \leqslant n}|a_k - a_0| \leqslant \max_{k \leqslant n}(|S_k| + |S'_k|) = \mathrm{O_P}(\tau_n)$, which contradicts the fact that $\max_{k \leqslant n}|a_k|/\tau_n \to \infty$ in probability.

Remark 3. WLG attempted to generalize previous results on invariance principle and wanted to establish $\{S_{k_n(t)}/\sigma_n,\ 0 \leqslant t \leqslant 1\} \Rightarrow W$ in $\mathscr{D}[0,1]$, where a_t in X_n are i.i.d. with $E(a_t^2) = 1$, $\mathscr{D}[0,1]$ is the collection of right continuous functions with left limits on $[0,1]$ and $k_n(t) = \sup\{m \leqslant n : B_m^2 \leqslant tB_n^2\}$ (cf. Theorem 2.1 in [45]). It seems that their derivation has a gap. Their key step is to apply their distributional equality (36), namely

$$\left\{\sum_{k=1}^{k_n(t)} a_k \Psi_{k_n(t)-k}, \quad 0 \leqslant t \leqslant 1\right\} \stackrel{\mathscr{D}}{=} \left\{\sum_{k=1}^{k_n(t)} a_k \Psi_{k-1}, \quad 0 \leqslant t \leqslant 1\right\}, \tag{7}$$

to establish the invariance principle

$$\{S_{k_n(t)}/B_n,\ 0\leqslant t\leqslant 1\} \Rightarrow W \tag{8}$$

via $\{B_n^{-1}\sum_{k=1}^{k_n(t)} a_k \Psi_{k-1},\ 0\leqslant t\leqslant 1\} \Rightarrow W$. It turns out that their claim (7) holds only for a *single* t and it fails to be valid *jointly* for $0\leqslant t\leqslant 1$. To see this, choose t_1 and t_2 such that $k_n(t_1)=1$ and $k_n(t_2)=2$. Then (7) fails since the random vectors $(a_1\psi_0, a_1(\psi_0+\psi_1)+a_2\psi_0)$ and $(a_1\psi_0, a_1\psi_0+a_2(\psi_0+\psi_1))$ generally have different distributions (even though the marginal distributions are the same). The invariance principle certainly requires the joint behavior over $0\leqslant t\leqslant 1$.

Remark 4. It would be interesting to compare our result with previous ones including HH and WLG even though the argument in the latter paper is not rigorous.

Our Theorem 1 differs from HH (p. 146) and WLG in several important aspects. Firstly, we allow a fairly general class of a_n which includes many nonlinear time series models. If a_n are i.i.d. or martingale differences, then a_n are automatically $\mathscr{L}^\alpha$ weakly dependent if $a_0\in\mathscr{L}^\alpha$. Our moment condition is slightly stronger since $\alpha>2$ is required (cf. Remark 2). HH and WLG assumed that a_n are i.i.d. or martingale differences with $E(a_n^2)<\infty$.

Secondly, WLG imposed the following condition on ψ_k

$$\frac{1}{B_n}\max_{1\leqslant j\leqslant n}|\Psi_j|\to 0 \quad\text{and}\quad \sum_{j=0}^{n} A_j^{1/2} = \mathrm{o}(B_n), \tag{9}$$

which is stronger than (6). To see this, let $\{a_n\}$ be i.i.d. with $\|a_n\|=1$. Then $\|E(S_n|\mathscr{F}_0)\|^2=\sum_{i=0}^{\infty}(\Psi_{n+i}-\Psi_i)^2$ and $\|E(X_j|\mathscr{F}_0)\|=A_j^{1/2}, j\geqslant 0$. Since $\|E(S_n|\mathscr{F}_0)\|\leqslant\sum_{j=1}^{n}\|E(X_j|\mathscr{F}_0)\|$, (9) implies (6). HH showed that the invariance principle holds if either

$$\sum_{j=1}^{\infty} A_j^{1/2}<\infty \tag{10}$$

or

$$\sum_{n=1}^{\infty}\left(\sum_{l=n}^{\infty}\psi_l\right)^2<\infty. \tag{11}$$

See Theorem 5.5, Corollary 5.4 and conditions (5.38) and (5.37) in HH. WLG's (9) weakens (10). However (11) cannot be derived from (9). For example, let $\psi_n=(-1)^n n^{-2/3}$, $n\geqslant 1$ and $\psi_0=1$. Then $|\sum_{l=n}^{\infty}\psi_l|=\mathcal{O}(n^{-2/3})$ and (11) holds, while (9) fails since $A_n=\sum_{m=n}^{\infty}\psi_n^2\sim 3n^{-1/3}$ and $\sum_{j=0}^{n}A_j^{1/2}\sim 1.2\sqrt{3}n^{5/6}$. Interestingly, (11) does imply our condition (6) since $\Psi_{n+i}-\Psi_i=\sum_{l=i+1}^{\infty}\psi_l-\sum_{l=i+1+n}^{\infty}\psi_l$. Thus (6) unifies (9)–(11).

Thirdly, WLG posed the open problem whether B_n can be replaced by σ_n. Theorem 1 provides an affirmative answer. Let $\{a_n\}$ be i.i.d. with $\|a_n\|=1$. Since

$E(S_n|\mathscr{F}_0)$ and $S_n - E(S_n|\mathscr{F}_0)$ are orthogonal, by (6),

$$\sigma_n^2 = \|E(S_n|\mathscr{F}_0)\|^2 + \|S_n - E(S_n|\mathscr{F}_0)\|^2 = \sum_{i=0}^{\infty} (\Psi_{n+i} - \Psi_i)^2 + B_n^2 \sim B_n^2.$$

As a step further, our Theorem 1 asserts that σ_n necessarily has the form $\ell^*(n)\sqrt{n}$ and reveals the inner relations $B_n/\sqrt{n} \sim \ell^*(n) \sim |\sum_{i=0}^{n-1} \Psi_i|/n$.

Finally, the form of our result $W_n \Rightarrow W$ is a typical one for invariance principles. It is slightly different from the one in WLG's (8), which involves the function $k_n(\cdot)$ that is difficult to deal with. Our form seems more convenient for application and it actually implies the latter. To see this, we apply the strong approximation technique. Since $W_n \Rightarrow W$ in the metric space $(\mathscr{C}[0,1], \rho)$, there exists a probability space on which we can define processes $\hat{W}_n$ and $\hat{W}$ such that $\hat{W}_n \overset{\mathscr{D}}{=} W_n$, $\hat{W} \overset{\mathscr{D}}{=} W$ and $\rho(\hat{W}_n, \hat{W}) \to 0$ almost surely (cf. [31, p. 79]). So $\sup_{0\leqslant t\leqslant 1} |\hat{W}_n[n^{-1}k_n(t)] - \hat{W}[n^{-1}k_n(t)]| \to 0$ almost surely. Let $t \in (0,1)$. For every $0 < \delta < \min(t, 1-t)$, since B_n^2/n is slowly varying in n, we have $B^2_{\lfloor n(t+\delta)\rfloor} > tB_n^2 > B^2_{\lfloor n(t-\delta)\rfloor}$ for sufficiently large n, where $\lfloor z \rfloor$ denotes the integer part of z. Note that $k_n(t) < m$ if and only if $B_m^2 > tB_n^2$. So $|n^{-1}k_n(t) - t| \leqslant \delta$, which implies $\lim_{n\to\infty} n^{-1}k_n(t) = t$ since δ can be arbitrarily small. Since $k_n(\cdot)$ is nondecreasing, it is easily seen that the uniform convergence $\sup_{0\leqslant t\leqslant 1} |n^{-1}k_n(t) - t| \to 0$ holds and consequently $\sup_{0\leqslant t\leqslant 1} |\hat{W}[n^{-1}k_n(t)] - \hat{W}(t)| \to 0$ in probability since $\hat{W}$ has a version with continuous path. Therefore $\sup_{0\leqslant t\leqslant 1} |\hat{W}_n[n^{-1}k_n(t)] - \hat{W}(t)| \to 0$ in probability and the invariance principle of WLG follows.

Remark 5. Hannan [23] considered invariance principles for $\sum_{i=1}^n \theta_{n,i} a_i$, where $\theta_{n,i}$ is a sequence of sets of constants and $\{a_i\}$ is $\mathscr{L}^2$ weakly dependent; see condition (9) therein. Let $T_j = \sum_{i=1}^j a_i$ and $\tilde{W}_n(t) = T_{k_n(t)}/\|T_n\|$, $0 \leqslant t \leqslant 1$, where $k_n(t) = \sup\{j : \|T_j\| \leqslant t\|T_n\|\}$. If $\theta_{n,i} \equiv 1$, Hannan's theorem (p. 284) asserts $\tilde{W}_n \Rightarrow W$ in $D[0,1]$. In our setting we consider linear processes of the form (2) with a_t as innovations. The linear processes are not necessarily $\mathscr{L}^2$ weakly dependent even though a_t are. They can actually be long-range dependent, which may lead to fBm as limits. Hannan's result does not imply Theorem 1 either: X_t satisfying conditions of Theorem 1 may not be $\mathscr{L}^2$ weakly dependent.

Example 1. Let $\psi_k = \ell(k)/k, k \geqslant 1$, where ℓ is a slowly varying function such that $\sum_{k=1}^{\infty} |\psi_k| = \infty$. By Lemma 4, Ψ_n is slowly varying, $|\Psi_n| \sim \sum_{k=0}^n |\psi_k|$, $\sigma_n^2 \sim n\Psi_n^2$ and $\lim_{n\to\infty} \sigma_n^{-2} \sum_{j=n}^{\infty} (\Psi_j - \Psi_{j-n})^2 = 0$. Hence (6) is satisfied.

2.3. Sample covariances

For a fixed integer $h \geqslant 0$ let the column random vector $X_{t,h} = (X_{t-h}, \ldots, X_t)^{\mathrm{T}}$, where $^{\mathrm{T}}$ stands for transpose and $\Gamma(h) = E(X_0 X_{h,h}) = (\gamma(0), \ldots, \gamma(h))$.

Theorem 3. *Assume that $\{a_n\}$ is $\mathscr{L}^4$ weakly dependent,*

$$\sum_{i=0}^{\infty} |\psi_i| \sqrt{A_{i+1}} < \infty \qquad (12)$$

and

$$\sum_{i,j=0}^{\infty} \|\mathscr{P}_1(a_i a_j)\| < \infty. \tag{13}$$

Then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} [X_t X_{t+h,h} - \Gamma(h)] \Rightarrow N(0, \Sigma_h), \tag{14}$$

where $\Sigma_h = E(\xi_h \xi_h^{\mathrm{T}})$ *and* $\xi_h = \sum_{t=-\infty}^{\infty} \mathscr{P}_1(X_t X_{t+h,h}) \in \mathscr{L}^2$.

Proposition 1. *A sufficient condition for* (12) *is*

$$\sum_{t=1}^{\infty} \sqrt{t}\psi_t^2 < \infty. \tag{15}$$

Proof of Proposition 1. By Schwarz's inequality, the proposition follows from

$$\begin{aligned}\left[\sum_{t=1}^{\infty} |\psi_t| \sqrt{A_{t+1}}\right]^2 &\leqslant \left[\sum_{t=1}^{\infty} \sqrt{t}\psi_t^2\right] \left[\sum_{t=1}^{\infty} A_{t+1} t^{-1/2}\right] \\ &\leqslant \left[\sum_{t=1}^{\infty} \sqrt{t}\psi_t^2\right] \left[\sum_{i=2}^{\infty} \sum_{t=1}^{i-1} \psi_i^2 t^{-1/2}\right]\end{aligned}$$

in view of $\sum_{t=1}^{i-1} t^{-1/2} \leqslant 2\sqrt{i}$ for $i \geqslant 2$. □

Example 2. In Section 3.1, we will show that (13) holds for GARCH models. Note that (12) and (15) allow some nonsummable sequences ψ_k, for example, $\psi_k = k^{-\beta}\ell(k)$, where ℓ is slowly varying and $\frac{3}{4} < \beta < 1$. Consider the FARIMA$(0, d, 0)$ model $(1-B)^d X_n = a_n$, where B is the back-shift operator ($BX_n = X_{n-1}$) and $-\frac{1}{2} < d < \frac{1}{2}$. Then $X_n = (1-B)^{-d} a_n = \sum_{i=0}^{\infty} \psi_i a_{n-i}$ and $\psi_j = \Gamma(j+d)/[\Gamma(j)\Gamma(d)] \sim j^{d-1}/\Gamma(d)$ as $j \to \infty$. If $1/4 > d > 0$, then (15) holds. Asymptotic distribution for sample correlations can be easily obtained from Theorem 3.

Remark 6. Theorem 6.7 in HH (p. 188) asserts asymptotic normality of sample correlations under the condition (15) for martingale differences a_t for which

$$\mathbb{E}(a_t^2 | \mathscr{F}_{t-1}) = \text{a positive constant}. \tag{16}$$

In the literature the above condition is widely used; see, for example, [13,22,26]. However, (16) appears too restrictive and it excludes many important models. Among them the most interesting case is the ARCH model. To see this, let $a_t = \varepsilon_t \sqrt{\theta_1^2 + \theta_2^2 a_{t-1}^2}$ be the ARCH(1) model, where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ are i.i.d. with mean 0 and variance 1 and θ_1 and θ_2 are parameters. Then $\mathbb{E}(a_t^2 | \mathscr{F}_{t-1}) = \theta_1^2 + \theta_2^2 a_{t-1}^2$, which cannot be almost surely constant unless $\theta_2 = 0$. Thus, limit theorems by He [26], Chung [13] and HH cannot be directly applied to linear processes with ARCH innovations. Our results avoid this limitation.

On the other hand, since we are dealing with stationary causal processes, the filter $(\mathscr{F}_t)_{t\in\mathbb{Z}}$ is naturally chosen to be sigma algebras generated by the vectors $(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. Such a structural assumption is not imposed in [13,20,26].

Remark 7. If a_n are martingale differences, then the $\mathscr{L}^p$ weak dependence condition trivially holds if $a_n \in \mathscr{L}^p$. Since $\mathscr{P}_1 a_i a_j = 0$ for $i > j \geqslant 1$, condition (13) is reduced to

$$\sum_{i=1}^{\infty} \|\mathscr{P}_1 a_i^2\| < \infty.$$

On the other hand, if (16) holds, then for $i \geqslant 2$, $E(a_i^2 | \mathscr{F}_1) = E[E(a_i^2 | \mathscr{F}_{i-1}) | \mathscr{F}_1]$ is almost surely a constant and hence $\mathscr{P}_1 a_i^2 = 0$ almost surely.

3. Applications

To apply Theorems 1–3, an important issue is to verify $\mathscr{L}^p$ weak dependence conditions. Proposition 2 below provides easily verifiable and mild conditions for a huge class of time series models that (1) represents. An important special class of (1) is the so-called *iterated random functions*. Let $G(\cdot,\cdot)$ be a bivariate measurable function with Lipschitz constant $L_\varepsilon = \sup_{x' \neq x} |G(x,\varepsilon) - G(x',\varepsilon)|/|x - x'|$ and Z_n be defined recursively by

$$Z_n = G(Z_{n-1}, \varepsilon_n). \tag{17}$$

Diaconis and Freedman [15] show that $\{Z_n\}$ has a unique stationary distribution if

$$E(\log L_\varepsilon) < 0, E(L_\varepsilon^\alpha) < \infty \quad \text{and} \quad E[|z_0 - G(z_0,\varepsilon)|^\alpha] < \infty \tag{18}$$

hold for some $\alpha > 0$ and z_0.

Example 3. Threshold autoregressive models (TAR, Tong [44]). Let the TAR(1) $a_n = \phi_1 \max(a_{n-1}, 0) + \phi_2 \max(-a_{n-1}, 0) + \varepsilon_n$. Then (18) is satisfied if $L_\varepsilon = \max(|\phi_1|, |\phi_2|) < 1$ and $E(|\varepsilon_0|^\alpha) < \infty$ for some $\alpha > 0$.

Example 4. Bilinear models [39]. Let $a_n = (\alpha_1 + \beta_1 \varepsilon_n) a_{n-1} + \varepsilon_n$, where α_1 and β_1 are real parameters and $E(|\varepsilon_0|^\alpha) < \infty$ for some $\alpha > 0$. Then the Lipschitz constant $L_\varepsilon = |\alpha_1 + \beta_1 \varepsilon|$ and (18) holds if $E(L_\varepsilon^\alpha) < 1$.

Example 5. Random coefficient autoregressive models (RCA, [40]). Let $a_n = (\phi_1 + \eta_n) a_{n-1} + \varepsilon_n$, where η_n are i.i.d., then the Lipschitz constant $L_\varepsilon = |\phi_1 + \eta_n|$ and (18) holds if $E(L_\varepsilon^\alpha) < 1$.

Lemma 2. (i) *Assume* (18). *Let* $a_n' = F(\ldots, \varepsilon_{-1}', \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$, *where* $\{\varepsilon_t'\}_{t\in\mathbb{Z}}$ *is an i.i.d. copy of* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. *Then there exists* $C, \alpha > 0$ *and* $\rho \in (0,1)$ *such that for all* $n \geqslant 0$,

$$E(|a_n - a_n'|^\alpha) \leqslant C\rho^n. \tag{19}$$

(ii) *Assume that* (19) *holds for some* $\alpha > 0$. *Let* $a_n \in \mathscr{L}^q$, $q > 0$. *Then for every* $\alpha \in (0, q)$, *there exist* $C_\alpha > 0$ *and* $\rho_\alpha \in (0,1)$ *such that* (19) *holds.*

Proof of Lemma 2. (i) follows from Lemma 3 in [50]. (ii) Let $\lambda_n = \rho^{n/(2\alpha)}$ and $0<p<q$. By (19), $P(|a_n - a'_n| \geqslant \lambda_n) \leqslant C\lambda_n^{\alpha}$. By Hölder's inequality,

$$E[|a_n - a'_n|^p(\mathbf{1}_{|a_n - a'_n| \geqslant \lambda_n} + \mathbf{1}_{|a_n - a'_n| < \lambda_n})] \leqslant \|a_n - a'_n\|_q^p [E(\mathbf{1}_{|a_n - a'_n| \geqslant \lambda_n})]^{1-p/q} + \lambda_n^p.$$

Hence $E(|a_n - a'_n|^p) = \mathrm{O}(\lambda_n^{\alpha(1-p/q)} + \lambda_n^p)$ since $a_n \in \mathscr{L}^q$. $\square$

We say that $\{a_t\}$ is *geometrically moment contracting* (GMC) if (19) holds. Besides (17), it also holds for GARCH models; see Section 3.1. The GMC property implies that the process $\{a_t\}$ forgets the past $\mathscr{F}_0$ exponentially fast in terms of the Euclidean distance between a_n and its coupled version a'_n. It is easily verifiable since it is directly related to the data generating mechanism of the process $\{a_n\}$. Recently, Hsing and Wu [30] obtained an asymptotic theory for U-statistics of processes satisfying (19). It turns out that (19) implies $\mathscr{L}^p$ weak dependence, and moreover, $\|\mathscr{P}_1 a_n\|_p$ decays to 0 exponentially fast.

Proposition 2. (i) *If* (19) *holds with some* $\alpha \geqslant 1$, *then* $\|E(a_n|\mathscr{F}_0)\|_\alpha = \mathrm{O}(\rho^n)$ *and hence* $\|\mathscr{P}_1 a_n\|_\alpha = \mathrm{O}(\rho^n)$ *for some* $\rho \in (0,1)$. (ii) *Assume that* $a_t \in \mathscr{L}^4$ *and* (19) *holds with* $\alpha = 4$. *Then there exist* $C>0$ *and* $\rho \in (0,1)$ *such that for all* $t,\ k \geqslant 0$,

$$\|\mathscr{P}_1(a_t a_{t+k})\| \leqslant C\rho^{t+k}. \tag{20}$$

Hence $\{a_t\}$ *satisfies* (13).

Proof of Proposition 2. In the proof let $C>0$ and $\rho \in (0,1)$ denote constants may vary from line-to-line. (i) Since $E(a'_n|\mathscr{F}_0) = 0$, $\|E(a_n|\mathscr{F}_0)\|_\alpha = \|E(a_n - a'_n|\mathscr{F}_0)\|_\alpha \leqslant \|a_n - a'_n\|_\alpha$. So $\|E(a_n|\mathscr{F}_1) - E(a_n|\mathscr{F}_0)\|_\alpha \leqslant \|E(a_{n-1}|\mathscr{F}_0)\|_\alpha + \|E(a_n|\mathscr{F}_0)\|_\alpha$ implies $\|\mathscr{P}_1 a_n\|_\alpha \leqslant C\rho^n$.

(ii) Let $t, k \geqslant 0$. Observe that $\gamma_k = E(a_t a_{t+k}) = E(a'_t a'_{t+k}|\mathscr{F}_0)$,

$$\begin{aligned}\|E(a_t a_{t+k}|\mathscr{F}_0) - \gamma_k\| &= \|E(a_t a_{t+k} - a'_t a'_{t+k})|\mathscr{F}_0\| \leqslant \|a_t a_{t+k} - a'_t a'_{t+k}\| \\ &\leqslant \|a_t(a_{t+k} - a'_{t+k})\| + \|(a_t - a'_t)a'_{t+k}\| \\ &\leqslant \|a_t\|_4 \|a_{t+k} - a'_{t+k}\|_4 + \|a_t - a'_t\|_4 \|a'_{t+k}\|_4 \\ &\leqslant C\rho^{t+k} + C\rho^t \leqslant C\rho^t,\end{aligned}$$

which, combined with a similar inequality $\|E(a_t a_{t+k}|\mathscr{F}_1) - \gamma_k\| \leqslant C\rho^t$, yields $\|\mathscr{P}_1 a_t a_{t+k}\| \leqslant C\rho^t$ via the triangle inequality. On the other hand, by Cauchy's inequality,

$$\begin{aligned}\|E(a_t a_{t+k}|\mathscr{F}_0)\| &= \|E(E(a_t a_{t+k}|\mathscr{F}_t)|\mathscr{F}_0)\| \leqslant \|E(a_t a_{t+k}|\mathscr{F}_t)\| \\ &= \|E(a_0 a_k|\mathscr{F}_0)\| = \|E(a_0(a_k - a'_k)|\mathscr{F}_0)\| \leqslant \|a_0\|_4 \|a_k - a'_k\|_4 \leqslant C\rho^k.\end{aligned}$$

Similarly $\|E(a_t a_{t+k}|\mathscr{F}_1)\| \leqslant C\rho^k$. Thus $\|\mathscr{P}_1 a_t a_{t+k}\| \leqslant C\rho^k$ and consequently $\|\mathscr{P}_1 a_t a_{t+k}\| \leqslant C\,\min(\rho^k, \rho^t) \leqslant C\rho^{(k+t)/2}$. $\square$

3.1. GARCH models

Let $\varepsilon_t,\ t \in \mathbb{Z}$, be i.i.d. random variables with mean 0 and variance 1; let

$$a_t = \sqrt{h_t}\varepsilon_t \text{ and } h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_q a_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \tag{21}$$

be the generalized autoregressive conditional heteroscedastic model GARCH(p, q), where $\alpha_0 > 0$, $\alpha_j \geqslant 0$ for $1 \leqslant j \leqslant q$ and $\beta_i \geqslant 0$ for $1 \leqslant i \leqslant p$. Then $\{a_t\}$ is stationary if

$$\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1, \tag{22}$$

See [5]. Notice that a_t form martingale differences. Hence $\mathscr{P}_1 a_t = 0$ for $t \geqslant 2$, $\mathscr{P}_1 a_1 = a_1$ and (4) holds for any $r \geqslant 0$ if $a_1 \in \mathscr{L}^p$. The existence of moments for GARCH models has been widely studied; see [11,27,33,35] and references therein.

Let $Y_t = (a_t^2, \ldots, a_{t-q+1}^2, h_t, \ldots, h_{t-p+1})^{\mathrm{T}}$, $b_t = (\alpha_0 \varepsilon_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^{\mathrm{T}}$ and $\theta = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$; let $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$ be the unit column vector with ith element being 1, $1 \leqslant i \leqslant p + q$. It is well known that GARCH models admits the following representation [6]:

$$Y_t = M_t Y_{t-1} + b_t, \text{ where } M_t = (\theta \varepsilon_t^2, e_1, \ldots, e_{q-1}, \theta, e_{q+1}, \ldots, e_{p+q-1})^{\mathrm{T}}. \tag{23}$$

For a square matrix M let $\rho(M)$ be its largest eigenvalue of $(M^{\mathrm{T}} M)^{1/2}$. Let $\otimes$ be the usual Kronecker product; let $|Y|$ be the Euclidean length of the vector Y. Assume $E(\varepsilon_t^4) < \infty$. Ling (1999) shows that if $\rho[E(M_t^{\otimes 2})] < 1$, then $\{a_t\}$ has a stationary distribution and $E(a_t^4) < \infty$. Ling and McAleer [35] argue that the condition $\rho[E(M_t^{\otimes 2})] < 1$ is also necessary for the finiteness of the fourth moment. Our Proposition 3 asserts that the same condition actually implies (19) as well.

Proposition 3. *For the* GARCH *model* (21), *assume that* ε_t *are i.i.d. with mean* 0 *and variance* 1, $E(\varepsilon_t^4) < \infty$ *and* $\rho[E(M_t^{\otimes 2})] < 1$. *Then* $E(|a_n - a_n'|^4) \leqslant C\rho^n$ *for some* $C < \infty$ *and* $\rho \in (0, 1)$. *Therefore* (19), *and consequently* (20) *hold.*

Proof of Proposition 3. Let Y_0',independent of $\{\varepsilon_t,\ t \in \mathbb{Z}\}$, be an i.i.d. copy of Y_0 and define recursively $Y_t' = M_t Y_{t-1}' + b_t$, $t \geqslant 1$; let $Y_n^* = Y_n - Y_n'$. Then $Y_n^* = M_n Y_{n-1}^*$. Note that M_t are i.i.d. $(p+q) \times (p+q)$ matrices. Using $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$, we have

$$Y_n^{*\otimes 2} = M_n^{\otimes 2} Y_{n-1}^{*\otimes 2} = \ldots = M_n^{\otimes 2} \ldots M_1^{\otimes 2} Y_0^{*\otimes 2}.$$

Hence $E(Y_n^{*\otimes 2}) = [E(M_1^{\otimes 2})]^n E(Y_0^{*\otimes 2})$ and (19) easily follows since $\rho[E(M_t^{\otimes 2})] < 1$. □

Under the conditions of Proposition 3, it is clear that Proposition 2 and hence Theorem 3 are applicable since $\|\mathscr{P}_0 a_i a_j\|$ and $\|\mathscr{P}_0 a_i\|$ decays to zero exponentially fast. To derive asymptotic distributions related to stationary processes, traditional approaches normally require strong mixing conditions. However, it is difficult to show that GARCH processes are strong mixing; see [10] for an recent attempt. If ε_t has a discrete distribution, then the results in the latter paper are not applicable and it is unclear whether $\{a_t\}$ is strong mixing. Andrews [1] showed that AR processes with Bernoulli innovations are not strong mixing. Our approach, however, provides a framework that completely avoids strong mixing conditions.

4. Proofs

In this section, we shall prove Theorems 1–3. Recall $A_n = \sum_{i=n}^{\infty} \psi_i^2$, $\Psi_n = \sum_{i=0}^{n} \psi_i$ and $B_n^2 = \sum_{i=0}^{n-1} \Psi_i^2$ and (3) for σ_n^2.

Lemma 3. *Let $\{D_i,\ i \in \mathbb{N}\}$ be a martingale difference sequence and $D_i \in \mathscr{L}^p$ for some $p \geqslant 2$. Then $\|\sum_{i=1}^{\infty} D_i\|_p \leqslant C_p[\sum_{i=1}^{\infty} \|D_i\|_p^2]^{1/2}$, where $C_p = 18p^{3/2}/(p-1)^{1/2}$.*

Proof of Lemma 3. It is a straightforward consequence of Burkholder's inequality (cf. Theorem 11.2.1 in [12]) and Minkowski's inequality

$$\left\|\sum_{i=1}^{\infty} D_i\right\|_p^p \leqslant C_p^p E\left(\sum_{i=1}^{\infty} D_i^2\right)^{p/2} \leqslant C_p^p \left[\sum_{i=1}^{\infty} \|D_i^2\|_{p/2}\right]^{p/2}$$

since $\|\cdot\|_{p/2}$ becomes a norm when $p/2 \geqslant 1$. □

(i) of Lemma 4 is also used in [45]. For the sake of completeness, we provide a proof here. (ii) is well-known and it is an easy consequence of Karamata's theorem [17].

Lemma 4. *Let ℓ be a slowly varying function.* (i) *Let $\psi_k = \ell(k)/k, k \geqslant 1$ and assume that $\sum_{k=1}^{\infty} |\psi_k| = \infty$. Then Ψ_n is slowly varying, $\ell(n)/\Psi_n \to 0$, $|\Psi_n| \sim \sum_{k=0}^{n} |\psi_k|$, $\sigma_n^2 \sim n\Psi_n^2$ and* $\lim_{n\to\infty} \sigma_n^{-2} \sum_{j=n}^{\infty} (\Psi_j - \Psi_{j-n})^2 = 0$. (ii) *Let $\psi_k = \ell(k)/k^{\beta}, k \geqslant 1$ and $\frac{1}{2} < \beta < 1$. Then $\Psi_n \sim n^{1-\beta}\ell(n)/(1-\beta)$ and $\sigma_n \sim n^{3/2-\beta}\ell(n)c_\beta$, where c_β is given in Theorem* 2.

Proof of Lemma 4. (i) Since ℓ is slowly varying, there exists $N_0 \in \mathbb{N}$ such that either $\ell(n) > 0$ for all $n \geqslant N_0$ or $\ell(n) < 0$ for all $n \geqslant N_0$. Without loss of generality we assume the former. So $\sum_{k=1}^{\infty} |\psi_k| = \infty$ implies $|\Psi_n| \sim \sum_{k=0}^{n} |\psi_k|$. For any $0 < \delta < 1$ and $G > 1$,

$$\begin{aligned} 0 &\leqslant \limsup_{n\to\infty} \frac{\Psi_n}{\Psi_{\delta n}} - 1 \leqslant \limsup_{n\to\infty} \frac{\sum_{k=\delta n}^{n} |\psi_k|}{\sum_{k=\delta n/G}^{\delta n-1} |\psi_k|} \\ &= \limsup_{n\to\infty} \frac{|\ell(n)| \sum_{k=\delta n}^{n} 1/k}{|\ell(\delta n)| \sum_{k=\delta n/G}^{\delta n-1} 1/k} = \frac{\log \delta^{-1}}{\log G} \end{aligned}$$

which approaches 0 as $G \to \infty$. Thus by definition Ψ_n is a slowly varying function. The same argument also implies $\lim_{n\to\infty} \ell(n)/\Psi_n = 0$.

By Karamata's theorem, $\sigma_n^2 \geqslant \sum_{j=0}^{n-1} \Psi_n^2 \sim n\Psi_n^2$. For any fixed $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n\Psi_n^2} \sum_{j=(1+\delta)n}^{\infty} (\Psi_j - \Psi_{j-n})^2 = \limsup_{n\to\infty} \frac{1}{n\Psi_n^2} \sum_{j=(1+\delta)n}^{\infty} \mathcal{O}(n\psi_j)^2 = 0$$

since $\ell(n)/\Psi_n \to 0$ and both $\ell(n)$ and Ψ_n are slowly varying. Thus

$$\limsup_{n\to\infty} \frac{1}{n\Psi_n^2} \sum_{j=n}^{\infty} (\Psi_j - \Psi_{j-n})^2 = \limsup_{n\to\infty} \frac{1}{n\Psi_n^2} \sum_{j=n}^{(1+\delta)n} (\Psi_j - \Psi_{j-n})^2$$
$$\leqslant \limsup_{n\to\infty} \frac{1}{n\Psi_n^2} \sum_{j=n}^{(1+\delta)n} 2(\Psi_j^2 + \Psi_{j-n}^2) \leqslant 4\delta,$$

which completes the proof of (i) since $\delta > 0$ is arbitrarily chosen. □

Lemma 5. *Let $\gamma > 0$ and $\ell(n)$ be a slowly varying function. Then*

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} (k/n)^\gamma |\ell(k)/\ell(n) - 1| = 0. \tag{24}$$

In particular, $\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} (k/n)^\gamma \ell(k)/\ell(n) = 1$.

Proof of Lemma 5. Let $\ell(m)$ have the representation $\ell(m) = c_m e^{\int_1^m \eta(u)/u \, du}$ [4] with $c_m \to c > 0$. Choose $K_0 \in \mathbb{N}$ be sufficiently large such that $i^{-\gamma/4} \leqslant \ell(i) \leqslant i^{\gamma/4}$ holds for all $i \geqslant K_0$. Then

$$\limsup_{n\to\infty} \max_{1\leqslant k\leqslant K_0} (k/n)^\gamma |\ell(k)/\ell(n) - 1| = 0, \tag{25}$$

$$\limsup_{n\to\infty} \max_{K_0\leqslant k\leqslant \sqrt{n}} \left(\frac{k}{n}\right)^\gamma \left|\frac{\ell(k)}{\ell(n)} - 1\right| = \limsup_{n\to\infty} \max_{K_0\leqslant k\leqslant \sqrt{n}} \left(\frac{k}{n}\right)^\gamma k^{\gamma/4} n^{\gamma/4} = 0 \tag{26}$$

and for any $\delta \in (0,1)$,

$$\limsup_{n\to\infty} \max_{n\delta\leqslant k\leqslant n} \left(\frac{k}{n}\right)^\gamma \left|\frac{\ell(k)}{\ell(n)} - 1\right| \leqslant \limsup_{n\to\infty} \max_{n\delta\leqslant k\leqslant n} \left|\frac{\ell(k)}{\ell(n)} - 1\right| = 0. \tag{27}$$

For sufficiently large n, we have $\max_{\sqrt{n}\leqslant u\leqslant \delta n} \eta(u) \leqslant \gamma/2$, and consequently by (25)–(27),

$$\limsup_{n\to\infty} \max_{1\leqslant k\leqslant n} (k/n)^\gamma |\ell(k)/\ell(n) - 1| \leqslant \limsup_{n\to\infty} \max_{\sqrt{n}\leqslant k\leqslant \delta n} (k/n)^\gamma |\ell(k)/\ell(n) - 1|$$
$$\leqslant \delta^\gamma + \limsup_{n\to\infty} \max_{\sqrt{n}\leqslant k\leqslant \delta n} (k/n)^\gamma c_k/c_n e^{\int_k^n \eta(u)/u \, du}$$
$$\leqslant \delta^\gamma + \limsup_{n\to\infty} \max_{\sqrt{n}\leqslant k\leqslant \delta n} (k/n)^\gamma e^{\max_{\sqrt{n}\leqslant u\leqslant \delta n} \eta(u) \int_k^n 1/u \, du} \leqslant \delta^\gamma + \delta^{\gamma/2}.$$

Thus the lemma follows since δ is arbitrarily chosen. □

Lemma 6. *Let $\psi_k = \ell(k)/k^\beta, k \geqslant 1$, where $\frac{1}{2} < \beta < 1$ and ℓ is a slowly varying function; let $\{\eta_k, k \in \mathbb{Z}\}$ be a stationary and ergodic process with mean 0 and $\Delta_n = \sigma_n^{-2} \sum_{j=0}^{\infty} (\Psi_j - \Psi_{j-n})^2 \eta_j$. Then* $\lim_{n\to\infty} E|\Delta_n| = 0$.

Proof of Lemma 6. Let $T_n = \sum_{k=0}^{n-1} \eta_k$ and $\delta_k = \sup_{n \geqslant k} E|T_n|/n$. By the ergodic theorem, $\delta_k \downarrow 0$. For any fixed $M > 1$ write

$$\Delta_n = \left[\sum_{j=0}^{n-1} + \sum_{j=n}^{(1+M)n} + \sum_{j=(1+M)n+1}^{\infty} \right] \frac{1}{\sigma_n^2} (\Psi_j - \Psi_{j-n})^2 \eta_j =: \Delta_{n,1} + \Omega_{n,M} + \Xi_{n,M}. \quad (28)$$

In the sequel we will show that $\lim_{n\to\infty} E|\Delta_{n,1}| = 0$ and $\lim_{n\to\infty} E|\Omega_{n,M}| = 0$. Using the Abelian summation technique, $\Delta_{n,1} = \sigma_n^{-2} \Psi_{n-1}^2 T_n + \sum_{j=1}^{n-1} \sigma_n^{-2} (\Psi_j^2 - \Psi_{j-1}^2) T_j$. So

$$\limsup_{n\to\infty} E|\Delta_{n,1}| \leqslant \limsup_{n\to\infty} \frac{\Psi_{n-1}^2 E|T_n|}{\sigma_n^2} + \limsup_{n\to\infty} \sum_{j=1}^{n-1} \frac{|\Psi_j^2 - \Psi_{j-1}^2| E|T_j|}{\sigma_n^2}. \quad (29)$$

Since $E|T_n| = o(n)$ and, by Lemma 4, $n\Psi_{n-1}^2 = \mathcal{O}(\sigma_n^2)$, the first term in the preceding display vanishes. For the second one, let $K \geqslant 1$ be a fixed integer. Then

$$\limsup_{n\to\infty} E|\Delta_{n,1}| \leqslant \limsup_{n\to\infty} \sum_{j=1}^{K-1} \frac{|\Psi_j^2 - \Psi_{j-1}^2| E|T_j|}{\sigma_n^2} + \delta_K \limsup_{n\to\infty} \sum_{j=K}^{n-1} \frac{j|\Psi_j^2 - \Psi_{j-1}^2|}{\sigma_n^2}. \quad (30)$$

Observe that as $j \to \infty$, $|\Psi_j^2 - \Psi_{j-1}^2| = |\psi_j||2\Psi_{j-1} + \psi_j| \sim 2j^{1-2\beta}\ell^2(j)/(1-\beta)$. Hence

$$\sum_{j=1}^{n-1} \frac{j|\Psi_j^2 - \Psi_{j-1}^2|}{\sigma_n^2} = \sum_{j=1}^{n-1} \frac{\mathcal{O}[j^{2-2\beta}\ell^2(j)]}{\sigma_n^2} = \frac{\mathcal{O}[n^{3-2\beta}\ell^2(n)]}{\sigma_n^2} = \mathcal{O}(1) \quad (31)$$

by Karamata's theorem. By (30), $\lim_{n\to\infty} E|\Delta_{n,1}| = 0$ since K is arbitrarily chosen and $\lim_{K\to\infty} \delta_K = 0$. The claim $\lim_{n\to\infty} E|\Omega_{n,M}| = 0$ can be similarly proved. Actually, by the same arguments in (29) and (30), it suffices to show the analogy of (31):

$$\frac{1}{\sigma_n^2} \sum_{j=1}^{nM} j|(\Psi_{n+j-1} - \Psi_{j-1})^2 - (\Psi_{n+j} - \Psi_j)^2| = \mathrm{O}(1). \quad (32)$$

Simple algebra shows that, for $1 \leqslant j \leqslant nM$,

$$\begin{aligned} |(\Psi_{n+j-1} - \Psi_{j-1})^2 - (\Psi_{n+j} - \Psi_j)^2| &\leqslant 2|\psi_{n+j} - \psi_j|\,|\Psi_{n+j-1} - \Psi_{j-1}| + |\psi_{n+j} - \psi_j|^2 \\ &= |\psi_{n+j} - \psi_j| \mathrm{O}[n^{1-\beta}\ell(n)] + |\psi_{n+j} - \psi_j|^2. \end{aligned} \quad (33)$$

Note that $\sum_{j=1}^{nM} |\psi_j| = \mathrm{O}[n^{1-\beta}\ell(n)]$. So the left-hand side of (32) is bounded by

$$\begin{aligned} &\frac{nM}{\sigma_n^2} \sum_{j=1}^{nM} |(\Psi_{n+j-1} - \Psi_{j-1})^2 - (\Psi_{n+j} - \Psi_j)^2| \\ &= \frac{n}{\sigma_n^2} \sum_{j=1}^{nM} |\psi_{n+j} - \psi_j| \mathrm{O}[n^{1-\beta}\ell(n)] = \mathrm{O}(1). \end{aligned}$$

Let $M \geqslant 1$. Since ℓ is slowly varying, there exists a constant $c_* > 0$ such that for all sufficiently large n, $|\Psi_j - \Psi_{j-n}| \leqslant c_* n j^{-\beta} \ell(j)$ holds for all $j \geqslant j_n = (1+M)n + 1$.

Therefore by (28) and Karamata's theorem,

$$\begin{aligned}\limsup_{n\to\infty} E|\Delta_n| &\leqslant \limsup_{n\to\infty} E|\Delta_{n,1}| + \limsup_{n\to\infty} E|\Omega_{n,M}| + \limsup_{n\to\infty} E|\Xi_{n,M}| \\ &\leqslant \limsup_{n\to\infty} \sum_{j=(1+M)n+1}^{\infty} \frac{1}{\sigma_n^2} [c_* n j^{-\beta} \ell(j)]^2 E|\eta_1| \\ &= \limsup_{n\to\infty} \frac{j_n^{1-2\beta}\ell^2(j_n)}{(2\beta-1)\sigma_n^2} c_*^2 E|\eta_1| = c_1 M^{1-2\beta}\end{aligned}$$

for some $c_1 < \infty$. Hence $\lim_{n\to\infty} E|\Delta_n| = 0$ by letting $M \to \infty$. $\square$

Lemma 7. *Assume that* $\{a_n\}$ *is* $\mathscr{L}^p$ $(p \geqslant 2)$ *weakly dependent with order* 1. *Then for* $S_n = \sum_{i=1}^n X_i$, *there exists a constant* C, *independent of* n, *such that for all* $n \in \mathbb{N}$,

$$\|S_n\|_p \leqslant C\sigma_n. \tag{34}$$

Proof of Lemma 7. Observe that $E(a_t|\mathscr{F}_0) = \sum_{k=-\infty}^{0} \mathscr{P}_k a_t$. Then

$$\sum_{t=0}^{\infty} \|E(a_t|\mathscr{F}_0)\|_p \leqslant \sum_{t=0}^{\infty} \sum_{k=-\infty}^{0} \|\mathscr{P}_k a_t\|_p = \sum_{t=0}^{\infty} \sum_{j=t+1}^{\infty} \|\mathscr{P}_1 a_j\|_p = \sum_{t=1}^{\infty} t\|\mathscr{P}_1 a_t\|_p < \infty. \tag{35}$$

Hence $b_k := \sum_{t=k}^{\infty} E(a_t|\mathscr{F}_k) \in \mathscr{L}^p$ and the Poisson equation $b_k = a_k + E(b_{k+1}|\mathscr{F}_k)$ holds. Let $d_k = b_k - E(b_k|\mathscr{F}_{k-1})$, which by definition are stationary and ergodic martingale differences. Let $X_t^* = \sum_{j=0}^{\infty} \psi_j d_{t-j}$, $X_t^\# = X_t - X_t^*$, $S_n^* = \sum_{i=1}^n X_t^*$ and $S_n^\# = \sum_{i=1}^n X_t^\#$. Then

$$S_n^* = \sum_{j=0}^{\infty} (\Psi_j - \Psi_{j-n}) d_{n-j} \text{ and } S_n^\# = \sum_{j=0}^{\infty} \psi_j [E(b_{1-j}|\mathscr{F}_{-j}) - E(b_{n+1-j}|\mathscr{F}_{n-j})]. \tag{36}$$

The essence of our approach is to approximate S_n by S_n^*, which admits martingale structures. By (35), $\|d_0\|_p \leqslant 2\|b_0\|_p < \infty$. By Lemma 3, $\|S_n^*\|_p \leqslant C_p \sigma_n \|d_0\|_p$. To establish (34), it then remains to verify that $\sum_{j=0}^{\infty} \psi_j E(b_{1-j}|\mathscr{F}_{-j}) \in \mathscr{L}^p$, which entails $\|S_n^\#\|_p = \mathcal{O}(1)$. To this end, for $k \geqslant 0$ let $V_k = \sum_{i=-\infty}^{0} \psi_{-i} \mathscr{P}_{i-k} b_{1+i}$. Then $\sum_{j=0}^{\infty} \psi_j E(b_{1-j}|\mathscr{F}_{-j}) = \sum_{k=0}^{\infty} V_k$. Since $\mathscr{P}_{i-k} b_{1+i}$ forms martingale differences in i, by Lemma 3,

$$\|V_k\|_p^2 \leqslant C_p^2 \sum_{i=-\infty}^{0} \|\psi_{-i} \mathscr{P}_{i-k} b_{1+i}\|_p^2 = C_p^2 \left(\sum_{j=0}^{\infty} \psi_j^2 \right) \|\mathscr{P}_1 b_{2+k}\|_p^2.$$

Therefore,

$$\left\| \sum_{j=0}^{\infty} \psi_j E(b_{1-j}|\mathscr{F}_{-j}) \right\|_p \leqslant \sum_{k=0}^{\infty} \|V_k\|_p \leqslant C_p A_0^{1/2} \sum_{k=0}^{\infty} \|\mathscr{P}_1 b_{1+k}\|_p$$

which is finite in view of

$$\sum_{k=0}^{\infty} \|\mathscr{P}_1 b_{1+k}\|_p \leqslant \sum_{k=0}^{\infty} \sum_{t=k+1}^{\infty} \|\mathscr{P}_1 E(a_t|\mathscr{F}_{k+1})\|_p$$
$$= \sum_{k=0}^{\infty} \sum_{t=k+1}^{\infty} \|\mathscr{P}_1 a_t\|_p = \sum_{t=1}^{\infty} t\|\mathscr{P}_1 a_t\|_p < \infty.$$

Here we have applied $\mathscr{P}_1 E(a_t|\mathscr{F}_{k+1}) = \mathscr{P}_1 a_t$ for $t \geqslant k+1 \geqslant 1$. Thus $\|S_n^{\#}\|_p = \mathscr{O}(1)$. □

Proof of Theorem 1. By the weak convergence theory of random functions, it suffices to establish (i) finite-dimensional convergence and (ii) tightness of W_n.

For (i), let $b_k = \sum_{t=k}^{\infty} E(a_t|\mathscr{F}_k)$ and $d_k = b_k - E(b_k|\mathscr{F}_{k-1})$. By the proof of Lemma 7, $b_k, d_k \in \mathscr{L}^2$. Recall (36) for $S_n^{\#}$ and S_n^*. By (6), $\|E(S_n^*|\mathscr{F}_0)\| = \mathrm{o}(\|S_n^*\|)$ and $\|S_n\| \sim B_n\|d_1\| \to \infty$. By Theorem 1 in [51], $\ell_1(n) = \|S_n^*\|/\sqrt{n}$ is a slowly varying function, and moreover for $H_{ni} = \bar{\Psi}_n d_i$, where $\bar{\Psi}_n = \sum_{j=0}^{n-1} \Psi_j/n$, we have

$$\max_{1 \leqslant k \leqslant n} \left\| S_k^* - \sum_{i=1}^{k} H_{ni} \right\| = \mathrm{o}(\|S_n^*\|). \tag{37}$$

Then $\|S_n^*\|^2 \sim \|\sum_{i=1}^n H_{ni}\|^2 = n\|H_{n1}\|^2$ and since $\|S_n^{\#}\| = \mathscr{O}(1)$,

$$\ell^*(n) := \frac{\|S_n\|}{\sqrt{n}} \sim \frac{\|S_n^*\|}{\sqrt{n}} = \ell_1(n) \sim |\bar{\Psi}_n|.$$

So the finite-dimensional convergence follows from (37) since

$$\frac{S_{nt}^*}{\sqrt{n}\bar{\Psi}_n} = \frac{\sum_{i=1}^{nt} H_{ni}}{\sqrt{n}\bar{\Psi}_n} + \mathrm{o_P}(1) = \frac{\sum_{i=1}^{nt} d_i}{\sqrt{n}} + \mathrm{o_P}(1) \Rightarrow N(0, t\|d_1\|^2), \quad 0 < t < 1.$$

For the tightness, by Theorem 12.3 in [3], we need to show that there exists a constant $C < \infty$ and $\tau > 1$ such that for all $1 \leqslant k \leqslant n$,

$$E[|W_n(k/n)|^{\alpha}] \leqslant C(k/n)^{\tau}. \tag{38}$$

We claim that (38) holds for $\tau = (2+\alpha)/4 > 1$. By Lemma 7, (38) is reduced to

$$\limsup_{n\to\infty} \max_{1\leqslant k\leqslant n} \left(\frac{n}{k}\right)^{\frac{2+\alpha}{4\alpha}} \frac{\|S_k\|}{\|S_n\|} = \limsup_{n\to\infty} \max_{1\leqslant k\leqslant n} \left(\frac{n}{k}\right)^{\frac{2+\alpha}{4\alpha}} \frac{\sqrt{k}\ell^*(k)}{\sqrt{n}\ell^*(n)} < \infty.$$

By Lemma 5, the limit in the preceding display is actually 1. □

Proof of Theorem 2. As in the proof of Theorem 1, we need to verify the finite-dimensional convergence and the tightness of W_n.

Since $\{a_n\}$ is $\mathscr{L}^2$ weakly dependent with order 1, from the proof of Lemma 7, we can define $b_k = \sum_{t=k}^{\infty} E(a_t|\mathscr{F}_k) \in \mathscr{L}^2$ and $d_k = b_k - E(b_k|\mathscr{F}_{k-1})$. Recall (36) for $S_n^{\#}$ and S_n^*. Note that $\|S_n^{\#}\| = \mathscr{O}(1)$, it suffices to show that $S_n^*/\sigma_n \Rightarrow N(0, \|d_1\|^2)$. To this end, we shall apply the martingale central limit theorem. Let $\zeta_{n,i} = (\Psi_i - \Psi_{i-n})/\sigma_n$.

Then $\sum_{t=1}^{n} X_t^*/\sigma_n = \sum_{i=0}^{\infty} \zeta_{n,i} d_{n-i}$ and $\sum_{i=0}^{\infty} \zeta_{n,i}^2 = 1$ for each n. Note that

$$\sup_{i\geqslant 0} |\zeta_{n,i}| \leqslant \sup_{i>2n} |\zeta_{n,i}| + \sup_{0\leqslant i\leqslant 2n} |\zeta_{n,i}| = n\mathcal{O}\left(\sup_{j\geqslant n} |\psi_j|/\sigma_n\right) + \mathcal{O}(n^{1-\beta}\ell(n)/\sigma_n) = \mathcal{O}(n^{-1/2}).$$

Then the Lindeberg condition is satisfied. Let $\eta_{n,i} = E(d_{n-i}^2|\mathcal{F}_{i-1}) - E(d_{n-i}^2)$. By Lemma 6, $\lim_{n\to\infty} E|\sum_{i=0}^{\infty} \zeta_{n,i}^2 \eta_{n,i}| = 0$. So the convergence of conditional variance follows.

For the tightness, we shall show that (38) holds with $\alpha = 2$ and $\tau = H + \frac{1}{2}$. By (ii) of Lemma 4, $\ell^*(n) = \sigma_n/n^H$ is slowly varying. Then (38) is equivalent to

$$\limsup_{n\to\infty} \max_{1\leqslant k\leqslant n} \left(\frac{n}{k}\right)^{H-1/2} \left[\frac{\ell^*(k)}{\ell^*(n)}\right]^2 < \infty,$$

which is an easy consequence of Lemma 5. □

Lemma 8. *Under the conditions of Theorem* 3, $\{X_t^2\}$ *is* $\mathcal{L}^2$ *weakly dependent*.

Proof of Lemma 8. Let $\bar{z}_{t,1} = \sum_{i=0}^{t-1} \psi_i a_{t-i}$ and $\underline{z}_{t,0} = \sum_{i=t}^{\infty} \psi_i a_{t-i}$. Then $X_t = \bar{z}_{t,1} + \underline{z}_{t,0}$. Since $\underline{z}_{t,0}$ is measurable with respect to $\mathcal{F}_0$, $\mathcal{P}_1 \underline{z}_{t,0}^2 = 0$. By the triangle inequality,

$$\|\mathcal{P}_1 X_t^2\| \leqslant \|\mathcal{P}_1 \bar{z}_{t,1}^2\| + 2\|\underline{z}_{t,0}\mathcal{P}_1 \bar{z}_{t,1}\| \leqslant \|\mathcal{P}_1 \bar{z}_{t,1}^2\| + 2\|\underline{z}_{t,0}\|_4 \|\mathcal{P}_1 \bar{z}_{t,1}\|_4.$$

Let $\psi_j = 0$ if $j<0$. Note that $\sum_{i=0}^{\infty} |\psi_i \psi_{i+j}| \leqslant A_0^{1/2} [\sum_{i=0}^{\infty} \psi_{i+j}^2]^{1/2} \leqslant A_0$. By (13),

$$\begin{aligned}
\frac{1}{2}\sum_{t=1}^{\infty} \|\mathcal{P}_1 \bar{z}_{t,1}^2\| &\leqslant \sum_{t=1}^{\infty} \sum_{0\leqslant i'\leqslant i<t} |\psi_i \psi_{i'}| \|\mathcal{P}_1 a_{t-i} a_{t-i'}\| \\
&\leqslant \sum_{i'=0}^{\infty} \sum_{i=i'}^{\infty} \sum_{t=i+1}^{\infty} |\psi_i \psi_{i'}| \|\mathcal{P}_1 a_{t-i} a_{t-i'}\| \\
&= \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\psi_i \psi_{i+j}| \|\mathcal{P}_1 a_k a_{k+j}\| \leqslant \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} A_0 \|\mathcal{P}_1 a_k a_{k+j}\| < \infty.
\end{aligned}$$

By (12), it remains to show $\|\underline{z}_{t,0}\|_4 \leqslant C\sqrt{A_t}$ for some constant $C>0$ in view of

$$\begin{aligned}
\sum_{t=1}^{\infty} \|\underline{z}_{t,0}\|_4 \|\mathcal{P}_1 \bar{z}_{t,1}\|_4 &\leqslant \sum_{t=1}^{\infty} C\sqrt{A_t} \sum_{j=0}^{t-1} |\psi_j| \|\mathcal{P}_1 a_{t-j}\|_4 \\
&= C\sum_{j=0}^{\infty} \sum_{t=j+1}^{\infty} \sqrt{A_t} |\psi_j| \|\mathcal{P}_1 a_{t-j}\|_4 \\
&\leqslant C\sum_{j=0}^{\infty} |\psi_j| \sqrt{A_{j+1}} \left(\sum_{t=j+1}^{\infty} \|\mathcal{P}_1 a_{t-j}\|_4\right) < \infty.
\end{aligned}$$

To prove $\|\underline{z}_{t,0}\|_4 \leqslant C\sqrt{A_t}$, define $U_k = \sum_{j=t}^{\infty} \psi_j E(a_{t-j}|\mathcal{F}_{t-j-k})$, $k\geqslant 0$. Then $U_0 = \underline{z}_{t,0}$ and $U_k - U_{k+1} = \sum_{j=t}^{\infty} \psi_j \mathcal{P}_{t-j-k} a_{t-j} = \sum_{i=-\infty}^{-t} \psi_{-i} \mathcal{P}_{t+i-k} a_{t+i}$. Observe that the summands $\psi_{-i}\mathscr{P}_{t+i-k}a_{t+i}$ of $U_k - U_{k+1}$ form martingale differences in i. By Lemma 3,

$$\|U_k - U_{k+1}\|_4 \leqslant C_4\left\{\sum_{i=-\infty}^{-t}\|\psi_{-i}\mathscr{P}_{t+i-k}a_{t+i}\|_4^2\right\}^{1/2} = C_4A_t^{1/2}\|\mathscr{P}_0a_k\|_4.$$

So $\|z_{t,0}\|_4 = \mathcal{O}(A_t^{1/2})$ since $z_{t,0} = \sum_{k=0}^{\infty}(U_k - U_{k+1})$ and $\{a_k\}$ is $\mathscr{L}^4$ weakly dependent. □

Proof of Theorem 3. We first consider $h = 1$. Let $\psi'_0 = \psi_0$, $\psi'_k = \psi_k + \psi_{k-1}$ for $k \geqslant 1$ and $A'_k = \sum_{i=k}^{\infty}|\psi'_i|^2$. Since $\sum_{k=0}^{\infty}\psi_k^2 < \infty$, by (12) implies $\sum_{i=0}^{\infty}|\psi'_i|\sqrt{A'_{i+1}} < \infty$. Hence by Lemma 8, $\sum_{t=1}^{\infty}\|\mathscr{P}_1(X_t + X_{t+1})^2\| < \infty$ since $X_t + X_{t+1} = \sum_{k=0}^{\infty}\psi'_k a_{t+1-k}$. Similarly, $\sum_{t=1}^{\infty}\|\mathscr{P}_1(X_t - X_{t+1})^2\| < \infty$. Using $4uv = (u+v)^2 - (u-v)^2$, we have $\sum_{t=1}^{\infty}\|\mathscr{P}_1(X_tX_{t+1})\| < \infty$. So $\sum_{t=1}^{\infty}\|\mathscr{P}_1(X_tX_{t+h,h})\| < \infty$ and (14) holds by the Crámer–Wold device. □

Acknowledgements

The authors are grateful to the referees and the associate editor for many helpful comments.

References

[1] D.W.K. Andrews, Nonstrong mixing autoregressive processes, J. Appl. Probab. 21 (4) (1984) 930–934.
[2] R.T. Baillie, C.F. Chung, M.A. Tieslau, Analysing inflation by the fractionally integrated ARFIMA–GARCH model, J. Appl. Econom. 11 (1996) 23–40.
[3] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[4] N.H. Bingham, C.M. Goldie, J.M. Teugels, Regular Variation, Cambridge University Press, Cambridge, UK, 1987.
[5] T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, J. Econom. 31 (1986) 307–327.
[6] P. Bougerol, N. Picard, Stationarity of GARCH processes and of some nonnegative time series, J. Econom. 52 (1992) 115–127.
[7] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1976.
[8] R. Bradley, Basic properties of strong mixing conditions, in: E. Eberlein, M. Taqqu (Eds.), Dependence in Probability and Statistics: A Survey of Recent Results, Birkhauser, Boston, MA, 1986, pp. 165–192.
[9] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer, New York, 1991.
[10] M. Carrasco, X. Chen, Mixing and moment properties of various GARCH and stochastic volatility models, Econom. Theory 18 (2002) 17–39.
[11] M. Chen, H. An, A note on the stationarity and existence of moments of the GARCH models, Statist. Sinica 8 (1998) 505–510.
[12] Y.S. Chow, H. Teicher, Probability theory, second ed., Springer, New York, 1988.
[13] C.F. Chung, Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes, Econom. Theory 18 (2002) 51–78.
[14] Yu.A. Davydov, The invariance principle for stationary process, Theory Probab. Appl. 15 (1970) 487–498.

[15] P. Diaconis, D. Freedman, Iterated random functions, SIAM Rev. 41 (1999) 41–76.
[16] P.H. Diananda, Some probability limit theorems with statistical applications, Proc. Cambridge Philos. Soc. (1953) 239–246.
[17] W. Feller, An introduction to probability theory and its applications, vol. II, Wiley, New York, 1971.
[18] L. Giraitis, D. Surgailis, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, Probab. Theory Related Fields 86 (1990) 87–104.
[19] V.V. Gorodetskii, Convergence to semi-stable Gaussian processes, Theory Probab. Appl. 22 (1977) 498–508.
[20] P. Hall, C.C. Heyde, Martingale Limit Theorem and its Application, Academic Press, New York, 1980.
[21] E.J. Hannan, Central limit theorems for time series regression, Z. Wahrsch. Verw. Gebiete 26 (1973) 157–170.
[22] E.J. Hannan, The asymptotic distribution of serial covariances, Ann. Statist. 4 (1976) 396–399.
[23] E.J. Hannan, The central limit theorem for time series regression, Stochast. Process. Appl. 9 (1979) 281–289.
[24] E.J. Hannan, C.C. Heyde, On limit theorems for quadratic functions of discrete time series, Ann. Math. Statist. 43 (1972) 2058–2066.
[25] M.A. Hauser, R.M. Kunst, Forecasting high-frequency financial data with the ARFIMA–ARCH model, J. Forecasting 20 (2001) 501–518.
[26] S. He, A note on the asymptotic normality of sample autocorrelations for a linear stationary sequence, J. Multivariate Anal. 58 (1996) 182–188.
[27] C. He, T. Teräsvirta, Fourth moment structure of the GARCH(p, q) process, Econom. Theory 15 (1999) 824–846.
[28] J.R.M. Hosking, Fractional differencing, Biometrika 73 (1981) 217–221.
[29] J.R.M. Hosking, Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series, J. Econom. 73 (1996) 261–284.
[30] T. Hsing, W.B. Wu, On weighted U-statistics for stationary processes, Ann. Probab. 32 (2003) 1600–1631.
[31] O. Kallenberg, Foundations of Modern Probability, second ed., Springer, New York, 2002.
[32] D. Lien, Y.K. Tse, Forecasting the Nikkei spot index with fractional cointegration, J. Forecasting 18 (1999) 259–273.
[33] D. Lien, Y.K. Tse, Fractional cointegration and futures hedging, J. Futures Markets 19 (1999) 457–474.
[34] S. Ling, On the stationarity and the existence of moments of conditional heteroskedastic ARMA models, Statist. Sinica 9 (4) (1999) 1119–1130.
[35] S. Ling, M. McAleer, Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH(r, s) models, Econom. Theory 18 (2002) 722–729.
[36] B. Mandelbrot, W.J. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. 10 (1969) 422–437.
[37] D.L. McLeish, Invariance principles for dependent variables, Z. Wahrsch. Verw. Gebiete 32 (1975) 165–178.
[38] D.L. McLeish, On the invariance principle for non-stationary mixingales, Ann. Probab. 5 (1977) 616–621.
[39] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer, Berlin, 1994.
[40] D.F. Nicholls, B.G. Quinn, Random Coefficient Autoregressive Models: an Introduction, Springer, New York, 1982.
[41] M. Peligrad, Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (A survey), in: E. Eberlein, M. Taqqu (Eds.), Dependence in Probability and Statistics: A Survey of Recent Results, Birkhauser, Boston, MA, 1986, pp. 193–223.
[42] P.C.B. Phillips, V. Solo, Asymptotics for linear processes, Ann. Statist. 20 (1992) 971–1001.
[43] P.J. Romano, L.A. Thombs, Inference for autocorrelations under weak assumption, J. Amer. Statist. Assoc. 91 (1996) 590–600.

[44] H. Tong, Non-linear Time Series: A Dynamic System Approach, Oxford University Press, Oxford, 1990.
[45] Q.Y. Wang, Y.X. Lin, C.M. Gulati, The invariance principle for linear processes with applications, Econom. Theory 18 (2002) 119–139.
[46] M. Woodroofe, A central limit theorem for functions of a Markov chain with applications to shifts, Stochast. Process. Appl. 41 (1992) 33–44.
[47] W.B. Wu, Empirical processes of long-memory sequences, Bernoulli 9 (2003) 809–831.
[48] W.B. Wu, Additive functionals of infinite-variance moving averages, Statist. Sinica 13 (2003) 1259–1267.
[49] W.B. Wu, J. Mielniczuk, Kernel density estimation for linear processes, Ann. Statist. 30 (2002) 1441–1459.
[50] W.B. Wu, M. Woodroofe, A central limit theorem for iterated random functions, J. Appl. Probab. 37 (2000) 748–755.
[51] W.B. Wu, M. Woodroofe, Martingale approximations for sums of stationary processes, Ann. Probab. 32 (2004) 1674–1690.
[52] R. Yokoyama, On the central limit theorem and law of the iterated logarithm for stationary processes with applications to linear processes, Stochast. Process. Appl. 59 (1995) 343–351.