The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

On Linear Processes with Dependent Innovations

Wei Biao Wu, Wanli Min

TECHNICAL REPORT NO. 549

May 24, 2004

5734 S. University Avenue
Chicago, IL 60637

On linear processes with dependent innovations

Wei Biao Wu, Wanli Min [1]

May 24, 2004

Department of Statistics, The University of Chicago, Chicago, IL 60637, USA

Abstract

We investigate asymptotic properties 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 non-linear time series models. Applications to linear processes with GARCH innovations and other non-linear time series models are discussed.

1 Introduction

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

$$a_t = F(\ldots, \epsilon_{t-1}, \epsilon_t) \tag{1}$$

is a well-defined random variable. Then $\{a_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic infinite-order nonlinear moving average process. Assume that a_t has mean 0, finite variance and that $\{\psi_i\}_{i\geq 0}$ 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 linear process

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

exists and $E(X_t^2) < \infty$. In this paper we consider asymptotic distributions of the properly normalized partial sum process $S_n = \sum_{i=1}^{n} X_i$ and the sample covariances $\hat{\gamma}_h = \sum_{t=1}^{n} X_t X_{t+h}/n$, $h \geq 0$. We will establish an invariance principle for the former and a central limit theorem (CLT) for $\hat{\gamma}_h$. Such results are important and useful in statistical inference of stationary processes.

[1] *E-mail address*: wbwu,wmin@galton.uchicago.edu

In the classical time series analysis, the innovations $\{a_t\}_{t\in\mathbb{Z}}$ in the linear process X_t are often assumed to be iid; see for example Brockwell and Davis (1991) and Box and Jenkins (1976). In this case asymptotic behaviors of the sample means and partial sum processes have been extensively studied in the literature. It would be hard to compile a complete list. Here we only mention some representatives: Davydov (1970), Gorodetskii (1977), Hall and Heyde (1980), Phillips and Solo (1992), Yokoyama (1995) and Hosking (1996). 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 iid observations. When the coefficients ψ_t are not summable, then X_t is *long-range dependent* (LRD) and the normalizing constant for $\sum_{i=1}^n X_i$ is typically larger than $\sqrt{n}$. Fractional ARIMA model (Hosking, 1981) is an important class which may exhibit LRD. Asymptotic normality for sample covariances has also been widely discussed; see for example, Box and Jenkins (1976), Hannan (1976), Hall and Heyde (1980), Giraitis and Surgailis (1990), Brockwell and Davis (1991), Phillips and Solo (1992) and Hosking (1996).

The asymptotic problem of partial sums and sample covariances becomes more difficult if dependence among a_t is allowed. Recently, fractional ARIMA (FARIMA) processes with GARCH innovations have been proposed to model some 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 such features; see Baillie, Chung and Tieslau (1995), Hauser and Kunst (2001) and Lien and Tse (1999).

Romano and Thombs (1996) point out that the traditional large sample inference on autocorrelations under the assumption of iid 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 (Dianana, 1953) or martingale differences (Hannan and Heyde, 1972). Recently, Wang, Lin and Gulati (2002) consider invariance principles for iid or martingale differences a_t. However, the proof in the latter paper contains a serious error [cf Remark 1]. Chung (2002) considers long-memory processes under the assumption that a_t are martingale differences with a prohibitively strong restriction [cf. (18)] that exclude the widely-used ARCH models.

In this paper, we shall apply the central limit theory and the idea of martingale approximations developed in Woodroofe (1992) and Wu and Woodroofe (2003) to obtain asymptotic distributions of the sample means and covariances of (2) under sufficiently mild conditions. Our results go beyond earlier ones by allowing a large class of important nonlinear process a_t, which substantially weakens the iid or martingale differences assumptions. In particular, our conditions are satisfied if a_t are the GARCH, random coefficient AR, bilinear AR and threshold AR models etc.

The goal of the paper is twofold: the first one is to obtain asymptotic distributions of $\bar{X}_n$ and $\hat{\gamma}_h$ which are certainly needed for statistical inferences of such processes. The second major goal is to introduce another type of dependence structure as well as to provide a unified methodology for asymptotic problems in econometrics models based on martingale approximations. With our dependence structure, martingales can be constructed which efficiently approximate the original sequences. Based on such martingale approximations, we can apply celebrated results in martingale theory such as martingale central limit theorems (MCLT) and martingale inequalities. 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. On the other hand, our condition is sufficiently mild as well. Recently Wu and Mielniczuk (2002), Hsing and Wu (2003), Wu (2003a, 2003b) apply the idea of martingale approximations and solve certain open asymptotic problems.

The paper is organized as follows. Main theorems are stated in Section 2 and comparisons with earlier results are elaborated. Proofs are given in the Section 4. In Section 3 we present applications to some nonlinear processes.

2 Results

Let X_t be the linear process defined by (2) and recall $E(a_n) = 0$; let $\mathcal{F}_t = (\ldots, \epsilon_{t-1}, \epsilon_t)$ be the shift process. Then $\mathcal{F}_t$ is a Markov Chain. For a random variable ξ define the projections $\mathcal{P}_t\xi = E(\xi|\mathcal{F}_t) - E(\xi|\mathcal{F}_{t-1})$ and its $\mathcal{L}^p$ $(p \geq 1)$ norm by $\|\xi\|_p = [E(|\xi|^p)]^{1/p}$. Let $\|\cdot\| = \|\cdot\|_2$ and write $\xi \in \mathcal{L}^p$ if $\|\xi\|_p < \infty$. We say $c_n \sim d_n$ for two sequences $\{c_n\}$ and $\{d_n\}$ if $\lim_{n\to\infty} c_n/d_n = 1$. For $n \geq 0$ write $A_n = \sum_{i=n}^{\infty} \psi_i^2$, $\Psi_n = \sum_{i=0}^{n} \psi_i$ and $B_n^2 = \sum_{i=0}^{n-1} \Psi_i^2$.

Let $\Psi_n = 0$ for $n < 0$ and define

$$\sigma_n^2 = \sum_{i=-n}^{\infty} (\Psi_{n+i} - \Psi_i)^2. \quad (3)$$

Before we state our main results, we now introduce the basic concept of $\mathcal{L}^p$ *weakly dependence* (cf Definition 1), which unlike strong mixing conditions, only involves the estimation of the decay rate of conditional moments. On the other hand, it appears to be very mild. In Section 3 we argue that many nonlinear time series models satisfy this condition. More importantly, it provides a natural vehicle for the central limit theory for stationary processes; see Woodroofe (1992) and Lemma 3 (Lemma 1 below) in Wu (2003a). An interesting observation of Lemma 1 is that there is a close form for the asymptotic variance in the limiting normal distribution.

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

$$\sum_{n=1}^{\infty} n^r \|\mathcal{P}_1 Y_n\|_p < \infty. \quad (4)$$

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

Lemma 1. *(Wu (2003a)) Let* $Y_n = g(\mathcal{F}_n)$ *be* $\mathcal{L}^2$ *weakly dependent. Then* $\sum_{i=1}^n (Y_i - EY_1)/\sqrt{n} \Rightarrow N(0, \|\xi\|^2)$*, where* $\xi = \sum_{j=1}^{\infty} \mathcal{P}_1 Y_j \in \mathcal{L}^2$ *and* $\Rightarrow$ *denotes convergence in distribution.*

The intuition of the definition weakly dependence is that the projection of the "future" Y_n to the space $\mathcal{M}_1 \ominus \mathcal{M}_0 = \{Z \in \mathcal{L}^p : Z \text{ is } \mathcal{F}_1 \text{ measurable and } E(Z|\mathcal{F}_0) = 0\}$ has a small magnitude. So the future depends weakly on the current states. If Y_n are martingale differences, then (4) is automatically satisfied for all $r \geq 0$ if $Y_0 \in \mathcal{L}^p$.

Lemma 1 readily entails the following corollary for the process (2).

Corollary 1. *Assume that* a_n *is* $\mathcal{L}^2$ *weakly dependent and*

$$\sum_{i=0}^{\infty} |\psi_i| < \infty. \quad (5)$$

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

2.1 Invariance Principles

Let $W_n(t), 0 \leq t \leq 1$ be a continuous, piece-wise linear function 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 $k/n \leq t \leq (k+1)/n$. Let space $\mathcal{C}[0,1]$ be the collection of all continuous function on $[0,1]$; let the distance between two functions $f,\ g \in \mathcal{C}[0,1]$ be $\rho(f,g) = \sup_{0\leq t\leq 1} |f(t) - g(t)|$. Billingsley (1968) contains an extensive treatment of the convergence theory on $\mathcal{C}[0,1]$. Recall $\Psi_n = \sum_{i=0}^n \psi_i$ for $n \geq 0$, $\Psi_n = 0$ for $n < 0$ and $B_n^2 = \sum_{i=0}^{n-1} \Psi_i^2$.

Theorem 1 asserts that the limiting distribution is the standard Brownian motion under (6). The norming sequence has the form $\|S_n\| = \ell^*(n)\sqrt{n}$ for some slowly varying function $\ell^*(n)$. If ψ_n decays sufficiently slowly, then (6) is violated and we may have fractional Brownian motions (Mandelbrot and Van Ness, 1969) as limits. In Theorem 2 we assume that the coefficients have the form $\psi_n = \ell(n)/n^\beta$ for $n \geq 1$, where $1/2 < \beta < 1$ and ℓ is a slowly varying function. Then ψ_n is not summable and norming sequences for S_n are different.

Theorem 1. *Assume that for some $\alpha > 2$, $\{a_n\}$ is $\mathcal{L}^\alpha$ weakly dependent with order 1,*

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

and $B_n \to \infty$. Then $\ell^*(n) := \|S_n\|/\sqrt{n}$ is a slowly varying function, $B_n/\sqrt{n} \sim \ell^*(n) \sim |\sum_{i=0}^{n-1} \Psi_i|/n$ and $\{W_n(t), 0 \leq t \leq 1\}$ converges in $\mathcal{C}[0,1]$ to the standard Brownian motion $W(t)$ on $[0,1]$.

Theorem 2. *Let $\psi_n = \ell(n)/n^\beta$ for $n \geq 1$, where $1/2 < \beta < 1$. Assume that a_n is $\mathcal{L}^2$ weakly dependent with order 1. Then $\{W_n(t), 0 \leq t \leq 1\}$ converges in $\mathcal{C}[0,1]$ to the standard fractional Brownian motion $\{W^H(t),\ 0 \leq t \leq 1]$ with Hurst index $H = 3/2 - \beta$, and $\|S_n\| \sim \sigma_n \|\eta\|$, with $\eta = \sum_{t=1}^\infty \mathcal{P}_1 a_t$ and σ_n defined in Lemma 3.*

Invariance principle is a powerful 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. This problem has a substantial history. The celebrated Donsker's theorem asserts invariance principles for iid sequence of X_n. For dependent sequences, see the survey by Bradley (1986) and Peligrad (1986) for strong mixing processes, McLeish (1975, 1977) for mixingale sequences, Billingsley (1968) and Hall and Heyde (1980).

In classical theory of invariance principles for linear processes, it is often assumed that innovations $\{a_n\}$ are iid or martingale differences; see Davydov (1970), Gorodetskii (1977), Hall and Heyde (HH, pp. 146, 1980) and the recent work by Wang, Lin and Gulati (WLG, 2002) among others. We believe that even for the special case that $\{a_n\}$ are martingale differences, Theorem 1 imposes the weakest sufficient conditions on the coefficients $\{\psi_n\}$.

Remark 1. WLG (2002) attempted to generalize previous results on invariance principle and wanted to establish $\{S_{k_n(t)}/\sigma_n,\ 0 \le t \le 1\} \Rightarrow \{W(t),\ 0 \le t \le 1\}$ in $\mathcal{D}[0,1]$, where X_n is defined by (2) with iid a_ts, $\mathcal{D}[0,1]$ is the collection of all right continuous functions with left limits on $[0,1]$ and $k_n(t) = \sup\{m:\ B_m^2 \le tB_n^2\}$ (cf Theorem 2.1 in the latter paper). However, there is a serious error in their argument. A key step in their paper is to use their equation (36), namely the distributional identity

$$\sum_{k=1}^{k_n(t)} a_k \Psi_{k_n(t)-k} =_D \sum_{k=1}^{k_n(t)} a_k \Psi_{k-1}, \quad 0 \le t \le 1, \tag{7}$$

to establish the invariance principle

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

via

$$\{B_n^{-1} \sum_{k=1}^{k_n(t)} a_k \Psi_{k-1},\ 0 \le t \le 1\} \Rightarrow \{W(t),\ 0 \le t \le 1\}. \tag{9}$$

The claim (7) holds only for a *single* t and it fails to be valid *jointly* for $0 \le t \le 1$. To see this, choose t_1 and t_2 such that $k_n(t_1) = 1$ and $k_n(t_2) = 2$. In this case it is easily seen that (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 $0 \le t \le 1$.

Remark 2. It would be interesting to compare our result with previous ones including WLG even though the latter paper contains an error.

Our Theorem 1 differs from HH (pp. 146, 1980) and WLG (2002) in several important aspects. First, we allow a fairly general class of a_n which includes many nonlinear time series models. If a_n are iid or martingale differences, then a_n are automatically $\mathcal{L}^\alpha$ weakly dependent with any order provided $a_0 \in \mathcal{L}^\alpha$. Our moment conditions is slightly stronger since $\alpha > 2$ is required. In HH (1980) and WLG (2002), a_n are assumed to iid or martingale differences and the existence of $E(a_n^2)$ is required.

Second, in WLG (2002), the conditions on ψ_k

$$\frac{1}{B_n} \max_{1\le j\le n} |\Psi_j| \to 0 \text{ and } \sum_{j=0}^{n} A_j^{1/2} = o(B_n) \tag{10}$$

are imposed, which are stronger than (6). To see this, assume that $\{a_n\}$ are stationary martingale differences with $\|a_n\| = 1$. Then $\|E(S_n|\mathcal{F}_0)\| = \sum_{i=0}^{\infty}(\Psi_{n+i}-\Psi_i)^2$, and for $j \ge 0$, $\|E(X_j|\mathcal{F}_0)\| = A_j^{1/2}$. Observe that $\|E(S_n|\mathcal{F}_0)\| \le \sum_{j=1}^{n} \|E(X_j|\mathcal{F}_0)\|$. Thus (10) implies (6). In HH (1980), two sufficient conditions for the invariance principle are presented. WLG's (2002) assumption (10) weakens one sufficient condition (cf Inequality (5.38) of HH's (1980))

$$\sum_{j=1}^{\infty} A_j^{1/2} < \infty \tag{11}$$

when one-sided linear processes is considered. However, the other sufficient condition (5.37) in HH (1980)

$$\sum_{n=1}^{\infty}(\sum_{l=n}^{\infty} \psi_l)^2 < \infty \tag{12}$$

cannot be derived from WLG's (10). For example, let $\psi_n = (-1)^n n^{-2/3}$, $n \ge 1$ and $\psi_0 = 1$. Then $|\sum_{l=n}^{\infty} \psi_l| = \mathcal{O}(n^{-2/3})$ and (12) holds. However, WLG's (10) does not hold 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, (12) does imply our condition (6) by noting that $\Psi_{n+i} - \Psi_i = \sum_{l=i+1}^{\infty} \psi_l - \sum_{l=i+1+n}^{\infty} \psi_l$. Thus (6) unifies (10), (11) and (12).

Third, WLG (2002) posed the open problem whether B_n can be replaced by σ_n. Theorem 1 provides an affirmative answer to this problem. Let $\{a_n\}$ be martingale differences with $\|a_n\| = 1$. Since $E(S_n|\mathcal{F}_0)$ and $S_n - E(S_n|\mathcal{F}_0)$ are orthogonal,

$$\|S_n\|^2 = \|E(S_n|\mathcal{F}_0)\|^2 + \|S_n - E(S_n|\mathcal{F}_0)\|^2,$$

which implies that $\sigma_n^2 = \sum_{i=0}^{\infty}(\Psi_{n+i} - \Psi_i)^2 + B_n^2 \sim B_n^2$ by (6). Moreover, 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 $\{W_n(t),\ 0 \le t \le 1\} \Rightarrow \{W(t),\ 0 \le t \le 1\}$ in $\mathcal{C}[0,1]$ is a typical one for invariance principles. It is slightly different from the one in WLG's (8). Our form seems more convenient for application and it actually implies the latter. To see this, let $\hat{W}_n(t)$ and $\hat{W}(t)$ be defined on a (possibly richer) probability space such that $\hat{W}_n \stackrel{\mathcal{D}}{=} W_n$, $\hat{W} \stackrel{\mathcal{D}}{=} W$ and $\sup_{0\le t\le 1} |\hat{W}_n(t) - \hat{W}(t)| \xrightarrow{\mathcal{P}} 0$. Here $\eta_1 \stackrel{\mathcal{D}}{=} \eta_2$ denotes that η_1 and η_2 have the same distribution and $\xrightarrow{\mathcal{P}}$ means convergence in probability. Since B_n^2/n is a slowly varying function, by Lemma 4 with $\gamma = 1$,

$$\lim_{n\to\infty} \max_{1\le k\le n} \left| \frac{B_k^2}{B_n^2} - \frac{k}{n} \right| = 0, \tag{13}$$

and since $\hat{W}$ has a version with continuous path, $\max_{1\le k\le n} |\hat{W}(B_k^2/B_n^2) - \hat{W}(k/n)| \xrightarrow{\mathcal{P}} 0$. Hence $\max_{1\le k\le n} |\hat{W}_n(k/n) - \hat{W}(B_k^2/B_n^2)| \xrightarrow{\mathcal{P}} 0$, which implies the invariance principle of WLG (2002).

Example 1. Let $\psi_k = \ell(k)/k, k \ge 1$, where ℓ is a slowly varying function such that $\sum_{k=1}^\infty |\psi_k| = \infty$. By Lemma 3, Ψ_n is slowly varying, $|\Psi_n| \sim \sum_{k=0}^n |\psi_k|$, $\sigma_n^2 \sim n\Psi_n^2$ and $\lim_{n\to\infty}(1/\sigma_n^2)\sum_{j=n}^\infty (\Psi_j - \Psi_{j-n})^2 = 0$. Hence (6) is satisfied.

2.2 Sample Covariances

Covariances play a fundamental role in the study of stationary processes. In this section we present asymptotic results for sample covariances.

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

$$\sum_{i=0}^\infty |\psi_i| \sqrt{A_{i+1}} < \infty \tag{14}$$

and

$$\sum_{i,j=0}^\infty \|\mathcal{P}_1(a_i a_j)\| < \infty. \tag{15}$$

Then $\sum_{t=1}^\infty \|\mathcal{P}_1 X_t^2\| < \infty$*; namely* X_t^2 *is* $\mathcal{L}^2$ *weakly dependent.*

Theorem 3 is useful in proving Corollary 2 concerning asymptotic normality of sample covariances.

Proposition 1. *A sufficient condition for (14) is*

$$\sum_{t=1}^{\infty} \sqrt{t}\psi_t^2 < \infty. \tag{16}$$

Corollary 2. *Let the conditions of Theorem 3 be satisfied. For a fixed integer $h \geq 0$ let the column random vector $X_{t,h} = (X_{t-h}, \ldots, X_t)^T$, where ^{T} stands for transpose and $\Gamma(h) = E(X_0 X_{h,h})$. Then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} [X_t X_{t+h,h} - \Gamma(h)] \Rightarrow N(0, \Sigma_h) \tag{17}$$

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

In Section 3.1, we will illustrate how to apply Theorem 3 and Corollary 2 to the GARCH models. Notice that (14) and (16) allow some non-summable sequences ψ_k. Consider the Fractional ARIMA$(0, d, 0)$ model $(1 - B)^d X_n = a_n$, where B is the back-shift operator ($BX_n = X_{n-1}$) and $-1/2 < d < 1/2$. Then $X_n = (1 - B)^{-d} a_n = \sum_{i=0}^{\infty} \psi_i a_{n-i}$ and as $j \to \infty$, $\psi_j = \Gamma(j + d)/[\Gamma(j)\Gamma(d)] \sim j^{d-1}/\Gamma(d)$. If $1/4 > d > 0$, then (16) holds. Asymptotic distribution for sample correlations can be easily obtained from Corollary 2.

Remark 3. In Theorem 6.7 of HH (pp 188, 1980), a CLT for sample correlations is derived under the condition (16), $\{a_t\}_{t \in \mathbb{Z}}$ are martingale differences and

$$\mathbb{E}(a_t^2 | \mathcal{F}_{t-1}) = \text{a positive constant}. \tag{18}$$

Relation (18) generalizes the case in which $\{a_n\}$ are iid. Chung (2002) recently derives various limit theorems for martingale differences $\{a_t\}_{t \in \mathbb{Z}}$ satisfying (18). Unfortunately the latter assumption appears too restrictive and it excludes many important models. One of the most interesting case is the ARCH model. To see this, let $a_t = \epsilon_t \sqrt{\theta_1^2 + \theta_2^2 a_{t-1}^2}$ be the ARCH(1) model, where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ are iid with mean 0 and variance 1 and θ_1 and θ_2 are parameters. Then $\mathbb{E}(a_t^2 | \mathcal{F}_{t-1}) = \theta_1^2 + \theta_2^2 a_{t-1}^2$, which cannot be almost surely constant unless $\theta_2 = 0$. Thus limit theorems by Chung (2002) and HH (1980) can not be directly applied to linear processes with ARCH innovations. Our results avoids this severe limitation.

Remark 4. If a_n are martingale differences, then the $\mathcal{L}^p$ weakly dependence condition trivially holds if $a_n \in \mathcal{L}^p$. Since $\mathcal{P}_1 a_i a_j = 0$ for $i > j \geq 1$, condition (15) is reduced to

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

On the other hand, if (18) holds, then for $i \geq 2$, $E(a_i^2|\mathcal{F}_1) = E[E(a_i^2|\mathcal{F}_{i-1})|\mathcal{F}_1]$ is almost surely a constant and hence $\mathcal{P}_1 a_i^2 = 0$ almost surely.

3 Applications

The class of non-linear time series models that (1) represents is clearly huge. For example, it includes threshold autoregressive model (TAR, Tong, 1990), bilinear model (Meyn and Tweedie, 1994), generalized autoregressive conditional heteroscedastic model (GARCH, Bollerslev, 1986), random coefficient autoregressive model (RCA) (Nicholls and Quinn, 1982) etc. Here we argue that those widely used non-linear time series models are $\mathcal{L}^p$ weakly dependent, and moreover, $\|\mathcal{P}_1 a_n\|_p$ decays to 0 exponentially fast. Namely there exists $C > 0$ and $\rho \in (0,1)$ such that $\|\mathcal{P}_1 a_n\|_p \leq C\rho^n$ for all $n \geq 0$. In this section $C > 0$ and $\rho \in (0,1)$ stand for constants which may vary from line to line.

Let $\{\epsilon'_t\}_{t\in\mathbb{Z}}$ be an iid copy of $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and $a'_n = F(\ldots, \epsilon'_{-1}, \epsilon'_0, \epsilon_1, \ldots, \epsilon_n)$ be a coupled version of a_n. We say that $\{a_t\}$ is *geometrically moment contracting (GMC)* if there exists $\alpha > 0$ such that for all $n \geq 0$,

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

The GMC condition (19) implies that the process $\{a_t\}$ forgets the past $\mathcal{F}_0$ at an exponential rate by measuring the distance between a_n and its coupled version a'_n. Recently Hsing and Wu (2003) obtains asymptotic theory for U-statistics of processing satisfying (19).

Proposition 2. *(i) If (19) holds with some $\alpha \geq 1$, then $\|E(a_n|\mathcal{F}_0)\|_\alpha \leq C\rho^n$ and hence $\|\mathcal{P}_1 a_n\|_\alpha \leq C\rho^n$. (ii) Assume that $\{a_n\}$ satisfies (19) for some $\alpha_0 > 0$ and $a_n \in \mathcal{L}^q$ for some $q > 1$. Then (19) holds for all $\alpha \in (0, q)$.*

Proposition 3. *Assume that $\{a_t\} \in \mathcal{L}^4$ and (19) holds with $\alpha = 4$. Then there exist $C > 0$ and $\rho \in (0,1)$ such that for all $t,\ k \geq 0$,*

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

Hence $\{a_t\}$ satisfies (15).

Propositions 2 and 3 make Theorems 1, 2, and Corollary 2 applicable by providing easily verifiable and mild conditions. 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

$$Z_n = G(Z_{n-1}, \varepsilon_n). \tag{21}$$

Diaconis and Freedman (1999) show that $\{Z_n\}$ has a unique stationary distribution if

$$E(\log L_\varepsilon) < 0, E(L_\varepsilon^\alpha) < \infty \text{ and } E[|z_0 - G(z_0, \varepsilon)|^\alpha] < \infty \tag{22}$$

hold for some $\alpha > 0$ and z_0. The same set of conditions actually also imply the GMC (19) (cf Lemma 3 in Wu and Woodroofe, 2000).

Example 2. Let the TAR(1) $a_n = \phi_1 \max(a_{n-1}, 0) + \phi_2 \max(-a_{n-1}, 0) + \epsilon_n$, the condition (22) is satisfied when $L_\epsilon = \max(|\phi_1|, |\phi_2|) < 1$ and $E(|\epsilon_0|^\alpha) < \infty$ for some $\alpha > 0$. For the bilinear model $a_n = (\alpha_1 + \beta_1 \epsilon_n) a_{n-1} + \epsilon_n$, where α_1 and β_1 are real parameters and $E(|\epsilon_0|^\alpha) < \infty$ for some $\alpha > 0$, we have a close form of the Lipschitz constant $L_\epsilon = |\alpha_1 + \beta_1 \epsilon|$ and (22) holds if $E(L_\epsilon^\alpha) < 1$. Similarly for the RCA model $a_n = (\phi_1 + \eta_n) a_{n-1} + \epsilon_n$, where η_n are iid, then the Lipschitz constant $L_\epsilon = |\phi_1 + \eta_n|$.

3.1 GARCH Models

Let $\epsilon_t,\ t \in \mathbb{Z}$ be iid random variables with mean 0 and variance 1; let

$$a_t = \sqrt{h_t} \epsilon_t \text{ and } h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_q a_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \tag{23}$$

be the generalized autoregressive conditional heteroscedastic model $GARCH(p, q)$, where $\alpha_0 > 0$, $\alpha_j \geq 0$ for $1 \leq j \leq q$ and $\beta_i \geq 0$ for $1 \leq i \leq p$. Chen and An (1998) show that if

$$\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1, \tag{24}$$

then a_t is strictly stationary. Notice that a_t form martingale differences. Hence $\mathcal{P}_1 a_t = 0$ for $t \geq 2$, $\mathcal{P}_1 a_1 = a_1$ and (4) holds for any $r \geq 0$ if $a_1 \in \mathcal{L}^p$. The existence of moments for GARCH models has been widely studied; see He and Teräsvirta (1999), Ling (1999), and Ling and McAleer (2002) and references therein.

Let $Y_t = (a_t^2, \ldots, a_{t-q+1}^2, h_t, \ldots, h_{t-p+1})^T$ and $b_t = (\alpha_0 \epsilon_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^T$. It is well known that GARCH models admits the following representation (see for example Theorem 3.2.14 in Taniguchi and Kakizawa (2000)):

$$Y_t = M_t Y_{t-1} + b_t, \text{ where } M_t = \begin{pmatrix} \alpha_1\epsilon_t^2 & \ldots & \alpha_{q-1}\epsilon_t^2 & \alpha_q\epsilon_t^2 & \beta_1\epsilon_t^2 & \ldots & \beta_{p-1}\epsilon_t^2 & \beta_p\epsilon_t^2 \\ 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 \\ \alpha_1 & \ldots & \alpha_{q-1} & \alpha_q & \beta_1 & \ldots & \beta_{p-1} & \beta_p \\ 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \tag{25}$$

For a square matrix M let $\rho(M)$ be its largest eigenvalue. Let $\otimes$ be the usual Kronecker product; let $|Y|$ be the Euclidean length of the vector Y. Assume $E(\epsilon_t^4) < \infty$. 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 (2002) argue that the condition $\rho[E(M_t^{\otimes 2})] < 1$ is also necessary for the finiteness of the fourth moment. Our Proposition 4 states that the same condition actually implies (19) as well.

Proposition 4. *Assume that ϵ_t are iid with mean 0 and variance 1, $E(\epsilon_t^4) < \infty$ and $\rho[E(M_t^{\otimes 2})] < 1$. Then $E(|a_n - a'_n|^4) \le C\rho^n$ for some $C < \infty$ and $\rho \in (0, 1)$. Namely (19) holds.*

Under the conditions of Proposition 4, it is clear that Proposition 3 and hence Theorem 3 are applicable since $\|\mathcal{P} a_i a_j\|$ and $\|\mathcal{P} a_i\|$ decays to zero exponentially fast. To derive asymptotic distributions related to stationary processes, traditional approaches normally require strong mixing conditions. However, it is a challenging problem to obtain strong mixing conditions of GARCH models; see Carrasco and Chen (2002) for an recent attempt. Our approach, however, provides a framework that completely avoids strong mixing conditions.

4 Proofs

We assume $E(a_n) = 0$ throughout. 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 2. *Let $\{D_i,\ i \in \mathbb{N}\}$ be a martingale difference sequence in $\mathcal{L}^p$ for some $p \geq 2$. Then $\| \sum_{i=1}^{\infty} D_i \|_p \leq C_p [\sum_{i=1}^{\infty} \|D_i\|_p^2]^{1/2}$, where $C_p = 18p^{3/2}/(p-1)^{1/2}$.*

Proof of Lemma 2. It is a straightforward consequence of Burkholder's inequality (cf Theorem 11.2.1 in Chow and Teicher (1978)) and Minkowski's inequality

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

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

Part (i) of Lemma 3 is also used in WLG (2002). For the sake of completeness, we provide a proof here. Part (ii) is well-known and it is an easy consequence of Karamata's theorem.

Lemma 3. *(i) Let $\psi_k = \ell(k)/k, k \geq 1$ 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 $\lim_{n\to\infty} (1/\sigma_n^2) \sum_{j=n}^{\infty} (\Psi_j - \Psi_{j-n})^2 = 0$. (ii) Let $\psi_k = \ell(k)/k^\beta, k \geq 1$ and $1/2 < \beta < 1$. Then $\Psi_n \sim n^{1-\beta}\ell(n)/(1-\beta)$ and $\sigma_n \sim n^{3/2-\beta}\ell(n)c_\beta$, where $c_\beta^2 = \int_0^\infty [x^{1-\beta} - \max(x-1,0)^{1-\beta}]^2 dx/(1-\beta)^2$.*

Proof of Lemma 3. (i) Since ℓ is slowly varying, there exists $N_0 \in \mathbb{N}$ such that either $\ell(n) > 0$ for all $n \geq N_0$ or $\ell(n) < 0$ for all $n \geq N_0$. Without loss of generality we assume the former. So $\sum_{k=1}^{\infty} |\psi_k| = \infty$ implies that $\lim_{n\to\infty} |\Psi_n| / \sum_{k=0}^{n} |\psi_k| = 1$. For any $0 < \delta < 1$ and $G > 1$,

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

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

By Karamata's theorem, $\sigma_n^2 \geq \sum_{j=0}^{n-1} \Psi_n^2 \sim n\Psi_n^2$ since Ψ_n is slowly varying. 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

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n\Psi_n^2}\sum_{j=n}^{\infty}(\Psi_j - \Psi_{j-n})^2 &= \limsup_{n\to\infty} \frac{1}{n\Psi_n^2}\sum_{j=n}^{(1+\delta)n}(\Psi_j - \Psi_{j-n})^2 \\ &\leq \limsup_{n\to\infty} \frac{1}{n\Psi_n^2}\sum_{j=n}^{(1+\delta)n} 2(\Psi_j^2 + \Psi_{j-n}^2) \leq 4\delta\end{aligned}$$

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

Lemma 4. *Let $\gamma > 0$ and $\ell(n)$ be a slowly varying function. Then*

$$\lim_{n\to\infty} \max_{1\leq k\leq n} (k/n)^{\gamma}|\ell(k)/\ell(n) - 1| = 0. \tag{26}$$

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

Proof of Lemma 4. Let $\ell(m)$ have the representation $\ell(m) = c_m e^{\int_1^m \eta(u)/u du}$ (Bingham, Goldie and Teugels, 1987) with $c_m \to c > 0$. Choose $K_0 \in \mathbb{N}$ be sufficiently large such that $i^{-\gamma/4} \leq \ell(i) \leq i^{\gamma/4}$ holds for all $i \geq K_0$. Then

$$\limsup_{n\to\infty} \max_{1\leq k\leq K_0} (k/n)^{\gamma}|\ell(k)/\ell(n) - 1| = 0, \tag{27}$$

$$\limsup_{n\to\infty} \max_{K_0\leq k\leq\sqrt{n}} (\frac{k}{n})^{\gamma}|\frac{\ell(k)}{\ell(n)} - 1| = \limsup_{n\to\infty} \max_{K_0\leq k\leq\sqrt{n}} (\frac{k}{n})^{\gamma} k^{\gamma/4} n^{\gamma/4} = 0, \tag{28}$$

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

$$\limsup_{n\to\infty} \max_{n\delta\leq k\leq n} (\frac{k}{n})^{\gamma}|\frac{\ell(k)}{\ell(n)} - 1| \leq \limsup_{n\to\infty} \max_{n\delta\leq k\leq n} |\frac{\ell(k)}{\ell(n)} - 1| = 0. \tag{29}$$

By (27), (28) and (29), for sufficiently large n,

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

Thus the lemma follows since $\max_{\sqrt{n}\leq u\leq \delta n}\eta(u) \leq \gamma/2$ for sufficiently large n and δ is arbitrarily chosen. $\diamondsuit$

Lemma 5. *Let* $\{\eta_k, k \in \mathbb{Z}\}$ *be a stationary and ergodic process with mean* 0 *and define* $\Delta_n = \sum_{j=0}^{\infty}(1/\sigma_n^2)(\Psi_j - \Psi_{j-n})^2\eta_j$*, where* $\{\psi_k,\ k \geq 0\}$ *satisfies conditions in (ii) of Theorem 2. Then* $\lim_{n\to\infty} E|\Delta_n| = 0$.

Proof of Lemma 5. Let $T_n = \sum_{k=0}^{n-1}\eta_k$ and $\delta_k = \sup_{n\geq k} E|T_n|/n$. By the ergodic theorem, $\delta_k \downarrow 0$. We first consider the case $\frac{1}{2} < \beta < 1$. For any fixed $M > 1$ write

$$\Delta_n = [\sum_{j=0}^{n-1} + \sum_{j=n}^{(1+M)n} + \sum_{j=(1+M)n+1}^{\infty}]\frac{1}{\sigma_n^2}(\Psi_j - \Psi_{j-n})^2\eta_j =: \Delta_{n,1} + \Omega_{n,M} + \Xi_{n,M}. \quad (30)$$

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, $\Delta_{n,1} = \Psi_{n-1}^2 T_n/\sigma_n^2 + \sum_{j=1}^{n-1}(1/\sigma_n^2)(\Psi_j^2 - \Psi_{j-1}^2)T_j$. Thus

$$\limsup_{n\to\infty} E|\Delta_{n,1}| \leq \limsup_{n\to\infty}\frac{\Psi_{n-1}^2 E|T_n|}{\sigma_n^2} + \limsup_{n\to\infty}\sum_{j=1}^{n-1}\frac{|\Psi_j^2 - \Psi_{j-1}^2|E|T_j|}{\sigma_n^2} \quad (31)$$

Since $E|T_n| = o(n)$ and $n\Psi_{n-1}^2/\sigma_n^2 = \mathcal{O}(1)$ by Lemma 3, the first term in the proceeding display vanishes. For the second one, let $K \geq 1$ be a fixed integer. Then

$$\limsup_{n\to\infty} E|\Delta_{n,1}| \leq \limsup_{n\to\infty}\sum_{j=1}^{K-1}\frac{|\Psi_j^2 - \Psi_{j-1}^2|E|T_j|}{\sigma_n^2} + \delta_K \limsup_{n\to\infty}\sum_{j=K}^{n-1}\frac{j|\Psi_j^2 - \Psi_{j-1}^2|}{\sigma_n^2}. \quad (32)$$

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

$$\limsup_{n\to\infty}\sum_{j=1}^{n-1}\frac{j|\Psi_j^2 - \Psi_{j-1}^2|}{\sigma_n^2} = \limsup_{n\to\infty}\sum_{j=1}^{n-1}\frac{\mathcal{O}[j^{2-2\beta}\ell^2(j)]}{\sigma_n^2} = \limsup_{n\to\infty}\frac{\mathcal{O}[n^{3-2\beta}\ell^2(n)]}{\sigma_n^2} = \mathcal{O}(1) \quad (33)$$

by Karamata's theorem. So (32) implies that $\limsup_{n\to\infty} E|\Delta_{n,1}| = 0$ since K is arbitrarily chosen and $\delta_K \downarrow 0$ as $K \to \infty$. The claim $\lim_{n\to\infty} E|\Omega_{n,M}| = 0$ can be similarly proved. Actually, by the same arguments in (31) and (32), it suffices to show the analogy of (33):

$$\limsup_{n\to\infty}\frac{1}{\sigma_n^2}\sum_{j=1}^{nM} j|(\Psi_{n+j-1} - \Psi_{j-1})^2 - (\Psi_{n+j} - \Psi_j)^2| < \infty. \quad (34)$$

Simple algebra gives

$$|(\Psi_{n+j-1} - \Psi_{j-1})^2 - (\Psi_{n+j} - \Psi_j)^2| \leq 2|\psi_{n+j} - \psi_j||\Psi_{n+j-1} - \Psi_{j-1}| + |\psi_{n+j} - \psi_j|^2. \quad (35)$$

Since $|\psi_{n+j}-\psi_j| = \mathcal{O}[j^{-\beta}\ell(j)]$ and $|\Psi_{n+j-1}-\Psi_{j-1}| = \mathcal{O}[n^{1-\beta}\ell(n)]$ for $1 \le j \le nM$, (34) follows from $\sum_{j=1}^{nM} jj^{-\beta}\ell(j) = \mathcal{O}[n^{2-2\beta}\ell^2(n)] = \mathcal{O}(\sigma_n^2/n)$ and

$$\limsup_{n\to\infty}\sum_{j=1}^{nM}\frac{j}{\sigma_n^2}|\psi_{n+j}-\psi_j||\Psi_{n+j-1}-\Psi_{j-1}| = \sum_{j=1}^{nM}\frac{\mathcal{O}[j^{1-\beta}\ell(j)n^{1-\beta}\ell(n)]}{\sigma_n^2} = \frac{\mathcal{O}[n^{3-2\beta}\ell^2(n)]}{\sigma_n^2}$$

by Karamata's theorem. Let $M \ge 1$. Since ℓ is slowly varying, there exists a constant $c_* > 0$ such that for all sufficiently large n, $|\Psi_j - \Psi_{j-n}| \le c_* n j^{-\beta}\ell(j)$ holds for all $j \ge j_n = (1+M)n+1$. Therefore by (30) and Karamata's theorem,

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

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

For the case $\beta = 1$, by (ii) of Lemma 3 and (30), it follows as in the the proof of the case $1/2 < \beta < 1$ from

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

since $\ell(j)\Psi_j$ is also a slowly varying function and $n\ell(n)\Psi_n/\sigma_n^2 = 0$ by Lemma 3. ♢

Lemma 6. *Assume that $\{a_n\}$ is $\mathcal{L}^p$ ($p \ge 2$) 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 $n \in \mathbb{N}$,*

$$\|S_n\|_p \le C\sigma_n. \tag{36}$$

Proof of Lemma 6. Observe that $E(a_t|\mathcal{F}_0) = \sum_{k=-\infty}^{0}\mathcal{P}_k a_t$. Then

$$\sum_{t=0}^{\infty}\|E(a_t|\mathcal{F}_0)\|_p \le \sum_{t=0}^{\infty}\sum_{k=-\infty}^{0}\|\mathcal{P}_k a_t\|_p = \sum_{t=0}^{\infty}\sum_{j=t+1}^{\infty}\|\mathcal{P}_1 a_j\|_p = \sum_{t=1}^{\infty} t\|\mathcal{P}_1 a_t\|_p < \infty. \tag{37}$$

Hence $b_k := \sum_{t=k}^{\infty} E(a_t|\mathcal{F}_k) \in \mathcal{L}^p$ and the Poisson equation $b_k = a_k + E(b_{k+1}|\mathcal{F}_k)$ holds. Let $d_k = b_k - E(b_k|\mathcal{F}_{k-1})$, 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

$$S_n^* = \sum_{j=0}^{\infty}(\Psi_j - \Psi_{j-n})d_{n-j} \text{ and } S_n^{\#} = \sum_{j=0}^{\infty}\psi_j[E(b_{1-j}|\mathcal{F}_{-j}) - E(b_{n+1-j}|\mathcal{F}_{n-j})]. \tag{38}$$

The essence of our approach is to approximate S_n by S_n^*, which admits martingale structures.

By Lemma 2, $\|S_n^*\|_p \le C_p \sigma_n \|d_0\|_p$. Inequality (37) implies that $\|d_0\|_p \le 2\|b_0\|_p < \infty$. To establish (36), it then remains to verify that $\sum_{j=0}^{\infty} \psi_j E(b_{1-j}|\mathcal{F}_{-j}) \in \mathcal{L}^p$, which entails $\|S_n^{\#}\|_p = \mathcal{O}(1)$. To this end, for $k \ge 0$ let $V_k = \sum_{i=-\infty}^{0} \psi_{-i} \mathcal{P}_{i-k} b_{1+i}$. Then $\sum_{j=0}^{\infty} \psi_j E(b_{1-j}|\mathcal{F}_{-j}) = \sum_{k=0}^{\infty} V_k$. Since $\mathcal{P}_{i-k} b_{1+i}$ forms martingale differences in i, by Lemma 2,

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

Therefore,

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

which is finite in view of

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

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

Proof of Corollary 1. It follows from Lemma 1 in view of

$$\sum_{t=1}^{\infty} \|\mathcal{P}_1 X_t\| \le \sum_{t=1}^{\infty} \sum_{i=0}^{\infty} |\psi_i| \|\mathcal{P}_1 a_{t-i}\| = \sum_{i=0}^{\infty} \sum_{t=i+1}^{\infty} |\psi_i| \|\mathcal{P}_1 a_{t-i}\| = \sum_{i=0}^{\infty} |\psi_i| \sum_{k=1}^{\infty} \|\mathcal{P}_1 a_k\| < \infty.$$

since $\mathcal{P}_1 a_k = 0$ for $k \le 0$.

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

To show the finite dimensional convergence, let $b_k = \sum_{t=k}^{\infty} E(a_t|\mathcal{F}_k) \in \mathcal{L}^2$ and $d_k = b_k - E(b_k|\mathcal{F}_{k-1})$, which are in $\mathcal{L}^2$ from the proof of Lemma 6. Recall (38) for $S_n^{\#}$ and S_n^*. Then (6) implies that $\|E(S_n^*|\mathcal{F}_0)\| = o(\|S_n^*\|)$ and $\|S_n\| \sim B_n \|d_1\| \to \infty$. By Theorem 1 in

Wu and Woodroofe (2003), $\ell_1(n) = \|S_n^*\|/\sqrt{n}$ 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

$$\max_{1\le k\le n} \|S_k^* - \sum_{i=1}^{k} H_{ni}\| = o(\|S_n^*\|). \tag{39}$$

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

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

Therefore, the finite dimensional convergence trivially follows from (39) since for $0 < t < 1$,

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

For the tightness, by Theorem 12.3 in Billingsley (1968), we need to show that there exists a constant $C < \infty$ and $\tau > 1$ such that for all $1 \le k \le n$,

$$E[|W_n(k/n)|^\alpha] \le C(k/n)^\tau. \tag{40}$$

We claim that (40) holds for $\tau = (2+\alpha)/4 > 1$. By Lemma 6, (40) is reduced to

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

By Lemma 4, the limit in the proceeding 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.

Since $\{a_n\}$ are $\mathcal{L}^2$ weakly dependent with order 1, from the proof of Lemma 6, we can define $b_k = \sum_{t=k}^\infty E(a_t|\mathcal{F}_k) \in \mathcal{L}^2$ and $d_k = b_k - E(b_k|\mathcal{F}_{k-1})$. Recall (38) for $S_n^{\#}$ and S_n^*. Note that $\|S_n^{\#}\| = \mathcal{O}(1)$, 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. The Lindeberg condition is obvious in view of

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

Let $\eta_{n,i} = E(d_{n-i}^2|\mathcal{F}_{i-1}) - E(d_{n-i}^2)$. Then the convergence of conditional variance easily follows from Lemma 6, which asserts that $\lim_{n\to\infty} E|\sum_{i=0}^{\infty} \zeta_{n,i}^2 \eta_{n,i}| = 0$.

For the tightness, we shall show that (40) holds with $\alpha = 2$ and $\tau = H + 1/2$. By (ii) of Lemma 3, $\ell^*(n) = \sigma_n/n^H$ is slowly varying. Then (40) is equivalent to

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

which is an easy consequence of Lemma 4. ♢

Proof of Theorem 3. Let $\overline{z}_{t,1} = \sum_{i=0}^{t-1} \psi_i a_{t-i}$ and $\underline{z}_{t,0} = \sum_{i=t}^{\infty} \psi_i a_{t-i}$. Then $X_t = \overline{z}_{t,1} + \underline{z}_{t,0}$. Since $\underline{z}_{t,0}$ is measurable with respect to $\mathcal{F}_0$, $\mathcal{P}_1 \underline{z}_{t,0}^2 = 0$, and by the triangle inequality,

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

We will show $\sum_{t=1}^{\infty} \|\mathcal{P}_1\overline{z}_{t,1}^2\| < \infty$ and $\sum_{t=1}^{\infty} \|\underline{z}_{t,0}\|_4.\|\mathcal{P}_1\overline{z}_{t,1}\|_4 < \infty$. For the former,

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

where the last inequality is due to $\sum_{i=0}^{\infty} |\psi_i\psi_j| \le \sum_{i=0}^{\infty} \psi_i^2$. Hence (15) entails the former inequality. By (14), the latter one holds if we can show that $\|\underline{z}_{t,0}\|_4 \le C\sqrt{A_t}$ for some constant $C > 0$ in view of

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

To prove $\|\underline{z}_{t,0}\|_4 \le C\sqrt{A_t}$, define $U_k = \sum_{j=t}^{\infty} \psi_j E(a_{t-j}|\mathcal{F}_{t-j-k})$, $k \ge 0$. 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 $\psi_{-i}\mathcal{P}_{t+i-k}a_{t+i}$ of $U_k - U_{k+1}$ form martingale differences in i. By Lemma 2,

$$\|U_k - U_{k+1}\|_4 \le C_4 \left\{ \sum_{i=-\infty}^{-t} \|\psi_{-i}\mathcal{P}_{t+i-k}a_{t+i}\|_4^2 \right\}^{1/2} = C_4 A_t^{1/2}\|\mathcal{P}_0 a_k\|_4,$$

which implies that

$$\|z_{t,0}\|_4 = \|\sum_{k=0}^{\infty}(U_k - U_{k+1})\|_4 \le \sum_{k=0}^{\infty}\|U_k - U_{k+1}\|_4 = \mathcal{O}(A_t^{1/2})$$

since $\{a_k\}$ are $\mathcal{L}^4$ weakly dependent, namely $\sum_{k=1}^{\infty}\|\mathcal{P}_0 a_k\|_4 < \infty$. ◇

Proof of Proposition 1. By the Cauchy inequality, the proposition follows from

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

in view of $\sum_{t=1}^{i-1}t^{-1/2} \le 2\sqrt{i}$ for $i \ge 2$. ◇

Proof of Corollary 2. We first consider $h = 1$. Let $\psi_0' = \psi_0$, $\psi_k' = \psi_k + \psi_{k-1}$ for $k \ge 1$ and $A_k' = \sum_{i=k}^{\infty}|\psi_i'|^2$. Since ψ_k are square summable, it is easily seen that (14) implies $\sum_{i=0}^{\infty}|\psi_i'|\sqrt{A_{i+1}'} < \infty$, which by Theorem 3 entails $\sum_{t=1}^{\infty}\|\mathcal{P}_1(X_t + X_{t+1})^2\| < \infty$ in view of $X_t + X_{t+1} = \sum_{k=0}^{\infty}\psi_k' a_{t+1-k}$. Similarly, $\sum_{t=1}^{\infty}\|\mathcal{P}_1(X_t - X_{t+1})^2\| < \infty$. Hence by the elementary identity $X_t X_{t+1} = [(X_t + X_{t+1})^2 - (X_t - X_{t+1})^2]/4$, we have $\sum_{t=1}^{\infty}\|\mathcal{P}_1(X_t X_{t+1})\| < \infty$ by the triangle inequality. Therefore, $\sum_{t=1}^{\infty}\|\mathcal{P}_1(X_t X_{t+h,h})\| < \infty$ easily follows and (17) holds by the Cramer-Wold device. ◇

Proof of Proposition 3. We shall show that $\|\mathcal{P}_1 a_t a_{t+k}\| \le C\rho^k$ and $\|\mathcal{P}_1 a_t a_{t+k}\| \le C\rho^t$ hold respectively, which imply that $\|\mathcal{P}_1 a_t a_{t+k}\| \le C\min(\rho^k, \rho^t) \le C\rho^{(k+t)/2}$. In this proof constants $C > 0$ and $\rho \in (0,1)$ may change from line to line. For the former, for $t,\ k \ge 0$, by Cauchy's inequality,

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

Thus $\|\mathcal{P}_1 a_t a_{t+k}\| \le \|E(a_t a_{t+k}|\mathcal{F}_1)\| + \|E(a_t a_{t+k}|\mathcal{F}_0)\| \le C\rho^k$. For the latter, observe that $\gamma_k = E(a_t a_{t+k}) = E(a_t' a_{t+k}'|\mathcal{F}_0)$,

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

which, combined with a similar inequality $\|E(a_t a_{t+k}|\mathcal{F}_1) - \gamma_k\| \leq C\rho^t$, yields $\|\mathcal{P}_1 a_t a_{t+k}\| \leq C\rho^t$ via triangle inequality. $\diamondsuit$

Proof of Proposition 2. (i) Since $E(a'_n|\mathcal{F}_0) = 0$, it easily follows from $\|E(a_n - a'_n|\mathcal{F}_0)\|_\alpha \leq \|a_n - a'_n\|_\alpha$. (ii) Let $\lambda_n = \rho^{n/(2\alpha)}$. By Markov's inequality, (19) implies $P(|a_n - a'_n| \geq \lambda_n) \leq C\lambda_n^\alpha$. By Hölder's inequality,

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

which entails $E(|a_n - a'_n|^p) \leq C\rho^n$ since $a_n \in \mathcal{L}^q$. Observe that $E(a'_n|\mathcal{F}_0) = 0$. Again by Hölder's inequality, $\|E(a_n|\mathcal{F}_0)\|_p = \|E(a_n - a'_n|\mathcal{F}_0)\|_p \leq \|a_n - a'_n\|_p$. Finally, $\|E(a_n|\mathcal{F}_1) - E(a_n|\mathcal{F}_0)\|_p \leq \|E(a_{n-1}|\mathcal{F}_0)\|_p + \|E(a_n|\mathcal{F}_0)\|_p$ implies $\|\mathcal{P}_1 a_n\|_p \leq C\rho^n$. $\diamondsuit$

Proof of Proposition 4. Let Y'_0, independent of $\{\epsilon_t,\ t \in \mathbb{Z}\}$ be an iid copy of Y_0 and define recursively $Y'_t = M_t Y'_{t-1} + b_t$, $t \geq 1$; let $Y^*_n = Y_n - Y'_n$. Then $Y^*_n = M_n Y^*_{n-1}$. 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 $\rho[E(M_t^{\otimes 2})] < 1$. $\diamondsuit$

REFERENCES

Baillie, R.T., C.F. Chung & M.A. Tieslau (1996) Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics* 11, 23–40.

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

Bingham, N.H., C.M. Goldie & J.L. Teugels (1987) *Regular Variation.* Cambridge, UK: Cambridge University Press.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Box, G.E.P. & G.M. Jenkins (1976) *Time series analysis, forecasting and control.* Holden-Day, San Francisco.

Bradley, R. (1986) Basic properties of strong mixing conditions. In E. Eberlein and M. Taqqu, editor, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 165–192. Birkhauser, Boston, MA.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods.* New York: Springer-Verlag.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chen, M.& H. An (1998) A note on the stationarity and existence of moments of the GARCH models. *Statistica Sinica* 8, 505–510.

Chung, C. F. (2002) Sample Means, Sample Autocovariances, and Linear Regression of Stationary Multivariate Long Memory Processes. *Econometric Theory* 18, 51–78.

Chow, Y.S. & H. Teicher (1988) *Probability Theory*, 2nd ed. New York: Springer-Verlag.

Davydov, Yu.A (1970) The invariance principle for stationary process. *Theory of Probability and Its Applications* 15, 487–498.

Diaconis, P. & D. Freedman (1999) Iterated random functions. *SIAM Review* 41, 41–76.

Diananda, P.H. (1953) Some probability limit theorems with statistical applications. *Proceedings of Cambridge Philosophy Society*, 239–246.

Giraitis, L. & D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probability Theory and Related Fields* 86, 87–104.

Gorodetskii, V.V. (1977) Convergence to Semi-Stable Gaussian Processes. *Theory of Probability and Its Applications* 22, 498–508.

Haggan, V. & T., Ozaki (1981) Modeling Nonlinear Vibrations Using an Amplitude-Dependent Autoregressive Time Series Models. *Biometrika* 68, 189–196.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theorem and Its Application*, New York, Academic Press.

Hannan, E.J. (1976) The asymptotic distribution of serial covariances. *Annals of Statistics* 4, 396–399.

Hannan, E.J. & C.C. Heyde (1972) On limit theorems for quadratic functions of discrete time series. *Annals of Mathematical Statistics* 43, 2058–2066.

Hauser, M.A. & R.M. Kunst (2001) Forecasting high-frequency financial data with the ARFIMA-ARCH model. *Journal of Forecasting* 20, 501–518.

He, C., T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.

Hosking, J.R.M (1981) Fractional Differencing. *Biometrika* 73, 217–221.

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

Hsing, T., Wu, W. B. On weighted U statistics for stationary processes. To appear, Ann. Probability, 2003

Lien, D., Y. K. Tse (1999) Forecasting the Nikkei spot index with fractional cointegration. *Journal of Forecasting* 18, 259–273.

Lien, D., Y. K. Tse (1999) Fractional cointegration and futures hedging. *Journal of Futures Markets* 19, 457–474.

Ling, S., W. K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.

Ling, S., M. McAleer (2002) Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH$(r,\ s)$ models. *Econometric Theory* 18, 722–729.

Mandelbrot, B., Van Ness, W.J. (1969) Fractional Brownian motions, fractional noises and applications, *SIAM Review*, 10, 422-437.

McLeish, D.L. (1975) Invariance Principles for dependent variables. *Z. Wahrsch. Verw. Gebiete* 32, 165–178

McLeish, D.L. (1977) On the invariance principle for non stationary mixingales. *Annals of Probability* 5, 616–621.

Meyn, S.P., R.L. Tweedie (1994) *Markov Chains and Stochastic Stability* , Springer-Verlag.

Nicholls, D.F., Quinn, B.G.(1982) *Random coefficient autoregressive models: an introduction.* New York: Springer.

Phillips, P.C.B., V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001

Peligrad, M. (1986) Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (A survey), In E. Eberlein and M. Taqqu, editor, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 193–223. Birkhauser, Boston, MA.

Romano, P.J. & L.A. Thombs (1996) Inference For Autocorrelations Under Weak Assumption. *Journal of the American Statistical Association* 91, 590–600

Taniguchi, M., Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time*

Series, New York: Springer.

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

Wang, Q.Y., Y.X. Lin, C.M. Gulati (2002) The invariance principle for linear processes with applications. *Econometric Theory* 18, 119–139.

Woodroofe, M. (1992) A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Processes and their Applications* 41, 33–44.

Wu, W. B. (2002) Central limit theorems for functionals of linear processes and their applications. *Statistica Sinica* 12, 635-649.

Wu, W. B. (2003a) Empirical Processes of Long-memory Sequences. 9, 809–831, *Bernoulli.*

Wu, W. B. (2003b). Additive Functionals of Infinite-Variance Moving Averages. *Statistica Sinica* 13, 1259–1267

Wu, W. B., M. Woodroofe (2000) A central limit theorem for iterated random functions. *Journal of Applied Probability.* 37, 748–755

Wu, W. B., M. Woodroofe (2004) Martingale approximations for sums of stationary processes. *Ann. Probability* 32, 1674–1690

Wu, W. B., J. Mielniczuk (2002) Kernel density estimation for linear processes. *Annals of Statistics* 30, 1441–1459.

Yokoyama, R. (1995) On the central limit theorem and law of the iterated logarithm for stationary processes with applications to linear processes. *Stochastic Processes and their Applications* 59, 343–351.