# Establishing conditions for weak convergence to stochastic integrals

August 17, 2016


Jiangyan Peng

School of Mathematical Sciences
School of Management and Economics
University of Electronic Science and Technology of China, Chengdu, 611731, P. R. China

Qiying Wang[1]

School of Mathematics and Statistics
University of Sydney, NSW, 2006, Australia



**Abstract**

Limit theory involving stochastic integrals plays a major role in time series econometrics. In earlier contributions on weak convergence to stochastic integrals, the literature commonly uses martingale and semimartingale structures. Liang, et al (2015) (see also Wang (2015), Chapter 4.5) currently extended the weak convergence to stochastic integrals by allowing for the linear process in the innovations. While these martingale and linear processes structures have wild relevance, they are not sufficiently general to cover many econometric applications where endogeneity and nonlinearity are present. This paper provides new conditions for weak convergence to stochastic integrals. Our frameworks allow for long memory processes, causal processes and near-epoch dependence in the innovations, which can be applied to a wild range of areas in econometrics, such as GARCH, TAR, bilinear and other nonlinear models.



*Key words and phrases*: Stochastic integral, convergence, long memory process, near-epoch dependence, linear process, causal process, TAR model, bilinear model, GARCH model.

*JEL Classification*: C22, C65



[1]Corresponding author: School of Mathematics and Statistics, The University of Sydney, NSW, 2006, Australia. E-mail: qiying@maths.usyd.edu.au.

# 1 Introduction

In econometrics with nonstationary time series, it is usually necessary to rely on the convergence to stochastic integrals. The latter result is particularly vital to nonlinear cointegrating regression. See Wang and Phillips (2009a, 2009b, 2016) for instance. Also see Wang (2015, Chapter 5) and the reference therein.

Let $(u_j, v_j)_{j\geq 1}$ be a sequence of random vectors on $R^d \times R$ and $\mathcal{F}_k = \sigma(u_j, v_j, j \leq k)$. Write

$$x_{nk} = \frac{1}{d_n}\sum_{j=1}^{k} u_j, \quad y_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} v_j,$$

where $0 < d_n^2 \to \infty$. As a benchmark, the basic result on convergence to stochastic integrals is given as follows. See, e.g., Kurtz and Protter (1991).

**THEOREM 1.1.** *Suppose*

***A1*** $(v_k, \mathcal{F}_k)$ *forms a martingale difference with* $\sup_{k\geq 1} E v_k^2 < \infty$*;*

***A2*** $\{x_{n,\lfloor nt\rfloor}, y_{n,\lfloor nt\rfloor}\} \Rightarrow \{G_t, W_t\}$ *on* $D_{\mathbb{R}^{d+1}}[0,1]$ *in the Skorohod topology.*

*Then, for any continuous functions* $g(s)$ *and* $f(s)$ *on* $R^d$*, we have*

$$\begin{aligned} &\big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, v_{k+1}\big\} \\ &\Rightarrow\ \big\{G_t,\ W_t,\ \int_0^1 g(G_t)dt,\ \int_0^1 f(G_t)\, dW_t\big\}, \end{aligned} \tag{1.1}$$

*on* $D_{\mathbb{R}^{2d+2}}[0,1]$ *in the Skorohod topology.*

Kurtz and Protter (1991) [also see Jacod and Shiryaev (2003)] actually established the result with $y_{nk}$ being a semimartingale instead of **A1**. Toward a general result beyond the semimartingale, Liang, et al. (2015) and Wang (2015, Chapter 4.5) investigated the extension to linear process innovations, namely, they provided the convergence of sample quantities $\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$ to functionals of stochastic processes and stochastic integrals, where

$$w_k = \sum_{j=0}^{\infty} \varphi_j\, v_{k-j}, \tag{1.2}$$

with $\varphi = \sum_{j=0}^{\infty} \varphi_j \neq 0$ and $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$. Liang, et al. (2015) and Wang (2015, Chapter 4.5) further considered the extension to $\alpha$-mixing innovations.

While these results are elegant, they are not sufficiently general to cover many econometric applications where endogeneity and more general innovation processes are present. In particular, the linear structure in (1.2) is well-known restrictive, failing to include many practical important models such as GARCH, threshold, nonlinear autoregressions, etc. The aim of this paper is to fill in the gap, providing new general results on the convergence to stochastic integrals in which there are some advantages in econometrical applications. Explicitly, our frameworks consider the convergence of $S_n := \sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$, where the $w_k$ has the form:

$$w_k = v_k + z_{k-1} - z_k, \tag{1.3}$$

with $z_k$ satisfying certain regular conditions specified in next section. The $\{w_k\}_{k\geq 1}$ in (1.3) is usually not a martingale difference, but $\sum_{k=1}^{n} w_k = \sum_{k=1}^{n} v_k + z_0 - z_n$ provides an approximation to martingale. Martingale approximation has been widely investigated in the literature. For a current development, we refer to Borovskikh and Korolyuk (1997). As evidenced in Section 3, these existing results on martingale approximation provide important technical support for the purpose of this paper.

This paper is organized as follows. In Section 2, we establish two frameworks for the convergence of $S_n$. Theorem 2.1 includes the situation that $u_k$ is a long memory process, while Theorem 2.2 is for the $u_k$ to be a short memory process. It is shown that, for a short memory $u_k$, the additional term $z_k$ in (1.3) has an essential impact on the limit behaviors of $S_n$, but it is not the case when $u_k$ is a long memory process under minor natural conditions on the $z_t$. Section 3 provides three corollaries of our frameworks on long memory processes, causal processes and near-epoch dependence, which capture the most popular models in econometrics. More detailed examples including linear processes, nonlinear transformations of linear processes, nonlinear autoregressive time series and GARCH model are given in Section 4. We conclude in Section 5. Proofs of all theorems are postponed to Section 6.

Throughout the paper, we denote constants by $C, C_1, C_2, \ldots$, which may differ at each appearance. $D_{\mathbb{R}^d}[0,1]$ denotes the space of càdlàg functions from $[0,1]$ to $\mathbb{R}^d$. If $x = (x_1, ..., x_m)$, we make use of the notation $||x|| = \sum_{j=1}^{m} |x_j|$. For a sequence of increasing $\sigma$-fields $\mathcal{F}_k$, we write $\mathcal{P}_k Z = E(Z|\mathcal{F}_k) - E(Z|\mathcal{F}_{k-1})$ for any $E|Z| < \infty$, and $Z \in \mathcal{L}^p (p > 0)$ if $\langle Z\rangle_p = (E|Z|^p)^{1/p} < \infty$. When no confusion occurs, we generally use the index notation $x_{nk}(y_{nk})$ for $x_{n,k}(y_{n,k})$. Other notation is standard.

# 2 Main results

In this section, we establish frameworks on convergence to stochastic integrals. Except mentioned explicitly, the notation is the same as Section 1.

**THEOREM 2.1.** *In addition to* ***A1–A2****, suppose that* $\sup_{k\geq 1} E\big(||z_k\, u_k||\big) < \infty$ *and* $d_n^2/n \to \infty$*. Then, for any continuous function* $g(s)$ *on* $R^d$ *and any function* $f(x)$ *on* $R^d$ *satisfying a local Lipschitz condition*[2]*, we have*

$$\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}\Big\}$$
$$\Rightarrow\ \Big\{G_t,\ W_t,\ \int_0^1 g(G_t)dt,\ \int_0^1 f(G_s)\, dW_s\Big\}. \tag{2.1}$$

As noticed in Liang, et al. (2015), the local Lipschitz condition is a minor requirement and hold for many continuous functions. If $\sup_{k\geq 1} E\big(||u_k||^2 + |z_k|^2\big) < \infty$, it is natural to have $\sup_{k\geq 1} E\big(||z_k\, u_k||\big) < \infty$ by Hölder's inequality. Theorem 2.1 indicates that, when $d_n^2/n \to \infty$, the additional term $z_k$ in (1.3) do not modify the limit behaviors under minor natural conditions on $z_k$ and $f(x)$.

The condition $d_n^2/n \to \infty$ usually holds when the components of $u_t$ are long memory processes. See Section 3.1 for example. The situation becomes very different if $d_n^2/n \to \sigma^2 < \infty$ for a constant $\sigma$, which generally holds for short memory processes $u_t$. In this situation, as seen in the following theorem, $z_t$ has an essential impact on the limit distributions.

Let $\mathrm{D}f(x) = \big(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_d}\big)'$. The following additional assumptions are required for our theory development.

**A3.** $\mathrm{D}f(x)$ *is continuous on* $R^d$ *and for any* $K > 0$*,*

$$||\mathrm{D}f(x) - \mathrm{D}f(y)|| \leq C_K\, ||x-y||^{\beta}, \quad \text{for some } \beta > 0,$$

*for* $\max\{||x||, ||y||\} \leq K$*, where* $C_K$ *is a constant depending only on* $K$*.*

**A4.** (i) $\sup_{k\geq 1} E||u_k||^2 < \infty$ and $\sup_{k\geq 1} E|z_k|^{2+\delta} < \infty$ for some $\delta > 0$;

[2]That is, for any $K > 0$, there exists a constant $C_K$ such that, for all $||x|| + ||y|| < K$,

$$|f(x) - f(y)| \leq C_K \sum_{j=1}^{d} |x_j - y_j|.$$

(ii) $Ez_ku_k \to A_0 = (A_{10}, ..., A_{d0})$, as $k \to \infty$;

Set $\quad \lambda_k = z_ku_k - Ez_ku_k.$

(iii) $\sup_{k\geq 2m} ||E(\lambda_k \mid \mathcal{F}_{k-m})|| = o_P(1)$, as $m \to \infty$; or

(iii)′ $\sup_{k\geq 2m} E\,||E(\lambda_k \mid \mathcal{F}_{k-m})|| = o(1)$, as $m \to \infty$.

**THEOREM 2.2.** *Suppose $d_n^2/n \to \sigma^2$, where $\sigma^2 > 0$ is a constant. Suppose **A1**–**A4** hold. Then, for any continuous function $g(s)$ on $R^d$, we have*

$$\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\,w_{k+1}\Big\}$$
$$\Rightarrow \Big\{G_t,\ W_t,\ \int_0^1 g(G_t)dt,\ \int_0^1 f(G_s)\,dW_s + \sigma^{-1}\sum_{j=1}^{d} A_{j0}\int_0^1 \frac{\partial f}{\partial x_j}(G_s)\,ds\Big\}. \tag{2.2}$$

**Remark 1.** Condition **A3** is similar to that in previous work. See, e.g, Liang, et al. (2015) and Wang (2015). The moment condition $\sup_{k\geq 1} E|z_k|^{2+\delta} < \infty$ for some $\delta > 0$ in **A4** (i) is required to remove the effect of higher order from $z_k$. In terms of the convergence in (2.2), $\sup_{k\geq 1} E|z_k|^2 < \infty$ is essentially to be necessary. It is not clear at the moment if the $\delta$ in **A4** (i) can be reduced to zero.

**Remark 2.** If $w_k$ satisfies (1.2), we may write $w_k = \varphi v_k + z_{k-1} - z_k$, where $z_k = \sum_{j=0}^{\infty} \bar{\varphi}_j\, v_{k-j}$ with $\bar{\varphi}_j = \sum_{m=j+1}^{\infty} \varphi_m$, i.e., $w_k$ can be denoted as in the structure of (1.3). See, e.g., Phillips and Solo (1992). For this $w_k$, Theorem 4.9 of Wang (2015) [also see Liang, et al. 2015] established a result that is similar to (2.2) by assuming (among other conditions) that, for any $i \geq 1$,

$$\sum_{j=0}^{\infty} \bar{\varphi}_j E\big(u_{j+i}v_i \mid \mathcal{F}_{i-1}\big) = A_0, \quad a.s., \tag{2.3}$$

where $A_0$ is a constant. Since it is required to be held for all $i \geq 1$, (2.3) is difficult to be verified for the $u_k$ to be a nonlinear stationary process such as $u_k = F(\epsilon_k, \epsilon_{k-1}, ...)$, even in the situation that $(\epsilon_k, v_k)$ are independent and identically distributed (i.i.d.) random vectors. In comparison, **A4** (ii) and (iii) [or (iii)′] can be easily applied to stationary causal processes and mixing sequences, as seen in Section 3.

**Remark 3.** We have $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} w_k = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} v_k + \frac{1}{\sqrt{n}}(z_0 - z_n)$, indicating that $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} w_k$ provides an approximation to the martingale $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} v_k$, under given conditions. However,

$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} w_k$ is not a semi-martingale as considered in Kurtz and Protter (1991), since we do not require the condition $\sup_{n\geq 1}\frac{1}{\sqrt{n}}\sum_{k=1}^{n} E|z_{k-1}-z_k|<\infty$. As a consequence, Theorems 2.1–2.2 provide an essential extension for the convergence to stochastic integrals, rather than a simple corollary of the previous works.

# 3 Three useful corollaries

This section investigates the applications of Theorems 2.1 and 2.2. Section 3.1 considers the situation that $u_k$ is a long memory process and $w_k$ is a stationary causal process. Section 3.2 contributes to the convergence for both $u_k$ and $w_k$ being stationary causal processes. Finally, in Section 3.3, we investigate the impact of near-epoch dependence in convergence to stochastic integrals. The detailed verification of assumptions for more practical models such as GARCH and nonlinear autoregressive time series will be presented in Section 4.

## 3.1 Long memory process

Let $(\epsilon_i,\eta_i)_{i\in\mathbb{Z}}$ be i.i.d. random vectors with zero means and $E\epsilon_0^2=E\eta_0^2=1$. Define a long memory linear process $u_k$ by

$$u_k=\sum_{j=1}^{\infty}\psi_j\epsilon_{k-j},$$

where $\psi_j\sim j^{-\mu}h(j)$, $1/2<\mu<1$ and $h(k)$ is a function that is slowly varying at $\infty$. Let $F$ be a measurable function such that

$$w_k=F(...,\eta_{k-1},\eta_k),\quad k\in\mathbb{Z},$$

is a well-defined stationary random variable with $Ew_0=0$ and $Ew_0^2<\infty$. The $w_k$ is known as a stationary causal process that has been extensively discussed in Wu (2005, 2007) and Wu and Min (2005).

Define $x_{nk}=\frac{1}{d_n}\sum_{j=1}^{k}u_j$ and $y_{nk}=\frac{1}{\sqrt{n}}\sum_{j=1}^{k}w_j$, where $d_n^2=var(\sum_{j=1}^{n}u_j)$. To investigate the convergence of $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}f(x_{nk})\,w_{k+1}$, we first introduce the following notation.

Write $\mathcal{F}_k=\sigma(\epsilon_i,\eta_i,i\leq k)$ and assume $\sum_{i=1}^{\infty}i\,\langle\mathcal{P}_0w_i\rangle_2<\infty$. The latter condition implies that $E(v_k^2 + z_k^2) < \infty$, where

$$v_k = \sum_{i=0}^{\infty} \mathcal{P}_k w_{i+k}, \quad z_k = \sum_{i=1}^{\infty} E(w_{i+k} | \mathcal{F}_k).$$

See Lemma 7 of Wu and Min (2005), namely, (35) there. All processes $w_k, v_k$ and $z_k$ are stationary satisfying the decomposition:

$$w_k = v_k + z_{k-1} - z_k. \tag{3.1}$$

We next let $\rho = E\epsilon_0 v_0 = \sum_{i=0}^{\infty} E\epsilon_0 w_i$, $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & Ev_0^2 \end{pmatrix}$, $(B_{1t}, B_{2t})$ be a bivariate Brownian motion with covariance matrix $\Omega\, t$ and $B_t$ be a standard Brownian motion independent of $(B_{1t}, B_{2t})$. We further define a fractional Brownian motion $B_H(t)$ depending on $(B_t, B_{1t})$ by

$$B_H(t) = \frac{1}{A(d)} \int_{-\infty}^{0} \Big[(t-s)^d - (-s)^d\Big] dB_s + \int_0^t (t-s)^d dB_{1s},$$

where

$$A(d) = \Big(\frac{1}{2d+1} + \int_0^{\infty} \Big[(1+s)^d - s^d\Big]^2 ds\Big)^{1/2}.$$

After these notation, a simple application of Theorem 2.1 yields the following result in the situation that $u_k$ is a long memory process and $w_k$ is a stationary causal process.

**THEOREM 3.1.** *Suppose $\sum_{i=1}^{\infty} i \langle \mathcal{P}_0 w_i \rangle_2 < \infty$ and, for some $\epsilon > 0$,*

$$\sum_{i=1}^{\infty} i^{1+\epsilon}\, E|w_i - w_i^*|^2 < \infty, \tag{3.2}$$

*where $w_k^* = F(..., \eta_{-1}^*, \eta_0^*, \eta_1, ..., \eta_{k-1}, \eta_k)$ and $\{\eta_k^*\}_{k\in\mathbb{Z}}$ is an i.i.d. copy of $\{\eta_k\}_{k\in\mathbb{Z}}$ and independent of $(\epsilon_k, \eta_k)_{k\in\mathbb{Z}}$. Then, for any continuous function $g(s)$ and any function $f(x)$ satisfying a local Lipschitz condition, we have*

$$\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}\Big\}$$
$$\Rightarrow\ \Big\{B_{3/2-\mu}(t), B_{2t},\ \int_0^1 g\big[B_{3/2-\mu}(t)\big]\, dt,\ \int_0^1 f\big[B_{3/2-\mu}(t)\big]\, dB_{2t}\Big\}. \tag{3.3}$$

We remark that condition $\sum_{i=1}^{\infty} i\langle \mathcal{P}_0 w_i\rangle_2 < \infty$ is close to be necessary. As shown in the proof of Theorem 3.1 (see Section 6), condition (3.2) can be replaced by

$$E\Big[\sum_{i=0}^{\infty} \mathcal{P}_k (w_{i+k} - w_{i+k}^*)\Big]^2 \to 0, \quad \text{as } k \to \infty,$$

which is required to remove the correlation between $\epsilon_{-j}$ and $v_j$ for $j \geq 1$ so that a bivariate process $(B_H(t), B_{2t})$ depending on $(B_t, B_{1t}, B_{2t})$ can be defined on $D_{\mathbb{R}^2}[0,1]$. Without this condition or equivalent, the limit distribution in (3.3) may have a different structure. Condition (3.2) is quite weak, which is satisfied by most of the commonly used models. Examples including nonlinear transformations of linear processes, nonlinear autoregressive time series and GARCH model will be given in Section 4.

## 3.2 Causal processes

As in Section 3.1, suppose that $(\epsilon_i, \eta_i)_{i\in\mathbb{Z}}$ are i.i.d. random vectors with zero means and $E\epsilon_0^2 = E\eta_0^2 = 1$. In this section, we let

$$u_k = F_1(..., \epsilon_{k-1}, \epsilon_k); \quad w_k = F_2(..., \eta_{k-1}, \eta_k), \quad k \in \mathbb{Z},$$

where $F_1$ and $F_2$ are measurable functions such that both $u_k$ and $w_k$ are well-defined stationary random variables with $Eu_0 = Ew_0 = 0$ and $Eu_0^2 + Ew_0^2 < \infty$, namely, both $u_k$ and $w_k$ are stationary causal processes.

This section investigates the convergence of $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$, where $x_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} u_j$. To this end, let $\mathcal{F}_k = \sigma(\epsilon_i, \eta_i, i \leq k)$,

$$\begin{aligned} z_{1k} &= \sum_{i=1}^{\infty} E(u_{i+k}|\mathcal{F}_k), \quad z_{2k} = \sum_{i=1}^{\infty} E(w_{i+k}|\mathcal{F}_k) \\ v_{1k} &= \sum_{i=0}^{\infty} \mathcal{P}_k u_{i+k}, \quad v_{2k} = \sum_{i=0}^{\infty} \mathcal{P}_k w_{i+k}. \end{aligned}$$

The following assumption is used in this section.

**A5** (i) $\sum_{i=1}^{\infty} i\langle \mathcal{P}_0 u_i\rangle_2 < \infty$; (ii) $\sum_{i=1}^{\infty} i\langle \mathcal{P}_0 w_i\rangle_{2+\delta} < \infty$, for some $\delta > 0$;

Set $\widetilde{\lambda}_k = u_k z_{2k} - Eu_k z_{2k}$.

(iii) $\sup_{k\geq 2m} |E(\widetilde{\lambda}_k \mid \mathcal{F}_{k-m})| = o_P(1)$, as $m \to \infty$; or

(iii)′ $\sup_{k\geq 2m} E\,|E(\widetilde{\lambda}_k \mid \mathcal{F}_{k-m})| = o(1)$, as $m \to \infty$.

As noticed in Section 3.1, all $u_k, w_k$, $z_{ik}$ and $v_{ik}, i = 1, 2$, are stationary, having the decompositions:

$$u_k = v_{1k} + z_{1,k-1} - z_{1k}, \quad w_k = v_{2k} + z_{2,k-1} - z_{2k}. \tag{3.4}$$

Furthermore **A5** (i) [(ii), respectively] implies that $E\big(v_{10}^2 + z_{10}^2\big) < \infty$ $[E\big(|v_{20}|^{2+\delta} + |z_{20}|^{2+\delta}\big) < \infty$, respectively]. As a consequence, it follows that

$$E|u_k z_{2k}| < \infty \quad \text{and} \quad A_0 := Eu_0 z_{20} = \sum_{i=1}^{\infty} E(u_0 w_i) < \infty.$$

We further let $\Omega = \begin{pmatrix} Ev_{10}^2 & Ev_{10}v_{20} \\ Ev_{10}v_{20} & Ev_{20}^2 \end{pmatrix}$ and $(B_{1t}, B_{2t})$ be a bivariate Brownian motion with covariance matrix $\Omega\, t$. We have the following result by making an application of Theorem 2.2.

**THEOREM 3.2.** *Suppose that **A3** (with $d = 1$) and **A5** hold. Then, for any continuous function $g(s)$, we have*

$$\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}\Big\}$$
$$\Rightarrow \Big\{B_{1t},\ B_{2t},\ \int_0^1 g(B_{1s})ds,\ \int_0^1 f(B_{1s})\, dB_{2s} + A_0 \int_0^1 f'[B_{1s}]ds\Big\}, \qquad (3.5)$$

*where $y_{nk} = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} w_j$.*

Theorem 3.2 provides a quite general result for both $u_t$ and $w_t$ are causal processes. In a related research, using a quite complicated technique originated from Jacod and Shiryaev (2003), Lin and Wang (2015) considered the specified situation that $u_t = w_t$. In comparison, by using Theorem 2.2, our proof is quite simple, as seen in Section 6. Furthermore our condition **A5** is easy to verify. An illustration is given in the following corollary, investigating the case that $u_k$ is a short memory linear process and $w_k$ is a general stationary causal process.

**COROLLARY 3.1.** *Suppose that $u_t = \sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}$, where $\sum_{i=1}^{\infty} i|\varphi_i| < \infty$. Result (3.5) holds true, if, in addition to A3 (with $d = 1$),*

$$\sum_{k=1}^{\infty} k\langle w_k - w_k^{'}\rangle_{2+\delta} < \infty, \quad \textit{for some } \delta > 0, \qquad (3.6)$$

*where $w_k^{'} = F_2(..., \eta_{-1}, \eta_0^*, \eta_1, ..., \eta_k)$ and $\{\eta_k^*\}_{k\in\mathbb{Z}}$ is an i.i.d. copy of $\{\eta_k\}_{k\in\mathbb{Z}}$ and independent of $(\epsilon_k, \eta_k)_{k\in\mathbb{Z}}$.*

Condition (3.6) is required to establish **A5** (ii). When $u_t = \sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}$ with $\sum_{i=1}^{\infty} i|\varphi_i| < \infty$, **A5** (iii) can be established under less restrictive condition: $\sum_{k=1}^{\infty} k\langle w_k - w_k^{'}\rangle_2 < \infty$ as seen in the proof of Corollary 3.1 given in Section 6. Some examples for $w_k$ satisfying (3.6), including nonlinear transformations of linear processes, nonlinear autoregressive time series and GARCH model are discussed in Section 4.

### 3.3 Near-epoch dependence

Let $\{A_k\}_{k\geq 1}$ be a sequence of random vectors whose coordinates are measurable functions of another random vector process $\{\eta_k\}_{k\in\mathbb{Z}}$. Define $\mathcal{F}_s^t = \sigma(\eta_s, ..., \eta_t)$ for $s \leq t$ and denote by $\mathcal{F}_t$ for $\mathcal{F}_{-\infty}^t$. As in Davidson (1994), $\{A_k\}_{k\geq 1}$ is said to be near-epoch dependence on $\{\eta_k\}_{k\in\mathbb{Z}}$ in $\mathcal{L}_P$-norm for $p > 0$ if

$$\langle A_t - E\big(A_t \mid \mathcal{F}_{t-m}^{t+m}\big)\rangle_p \leq d_t\, \nu(m),$$

where $d_t$ is a sequence of positive constants, and $\nu(m) \to 0$ as $m \to \infty$. For short, $\{A_k\}_{k\geq 1}$ is said to be $\mathcal{L}_P$-NED of size $-\mu$ if $d_t \leq \langle A_t\rangle_p$ and $\nu(m) = O(m^{-\mu-\epsilon})$ for some $\epsilon > 0$.

For $k \geq 1$, let $x_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} u_j$ and $y_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} w_j$, where $(u_k, w_k)_{k\geq 1}$ defined on $R^{d+1}$ is a stationary process. This section investigates the convergence of $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$ in the following conditions:

**A6** (i) $\eta_k = (\eta_{k1}, ..., \eta_{km}), k \in \mathbb{Z}$, is $\alpha$-mixing of size $-6$ [3];

(ii) $(u_k)_{k\geq 1}$ is $\mathcal{L}_2$-NED of size $-1$ and $u_k$ is adapted to $\mathcal{F}_k$;

(iii) $(w_k)_{k\geq 1}$ is $\mathcal{L}_{2+\delta}$-NED of size $-1$ for some $\delta > 0$;

(iv) $E(u_0, w_0) = 0$ and $E\big(||u_0||^4 + |w_0|^4\big) < \infty$.

Due to the stationarity of $(u_k, w_k)_{k\geq 1}$, it follows easily from **A6** that

$$\Omega \;:=\; \lim_{n\to\infty} \frac{1}{n}\sum_{i,j=1}^{n} E(M_i' M_j) = \begin{pmatrix} \Omega_1 & \rho \\ \rho' & \Omega_2 \end{pmatrix}, \tag{3.7}$$

where $M_k = (u_k, w_k)$ and

$$\Omega_1 = Eu_0^{'}u_0 + 2\sum_{i=1}^{\infty} Eu_0^{'}u_i, \quad \Omega_2 = Ew_0^2 + 2\sum_{i=1}^{\infty} Ew_0 w_i,$$
$$\rho = Eu_0^{'}w_0 + \sum_{i=1}^{\infty}(Eu_0^{'}w_i + Eu_i^{'}w_0).$$

[3]For a definitions of $\alpha$-mixing, we refer to Davidson (1994).

For a proof of (3.7), see Section 6. In terms of (3.7) and **A6**, Corollary 29.19 of Davidson (1994, Page 494) yields that, as $n \to \infty$,

$$\big(x_{n,[nt]}, y_{n,[nt]}\big) \Rightarrow (B_{1t}, B_{2t}), \tag{3.8}$$

where $(B_{1t}, B_{2t})$ is a $d+1$-dimensional Brownian motion with covariance matrix $\Omega\, t$. Now, by using Theorem 2.2, we have the following theorem.

**THEOREM 3.3.** *Suppose* ***A3*** *and* ***A6*** *hold. For any continuous function $g(s)$ on $R^d$, we have*

$$\begin{aligned}\big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}\big\}\\ \Rightarrow\ \big\{B_{1t},\ B_{2t},\ \int_0^1 g(B_{1s})ds,\ \int_0^1 f(B_{1s})\, dB_{2s} + \int_0^1 A_0\, \mathrm{D}f[B_{1s}]ds\big\},\end{aligned} \tag{3.9}$$

*where $A_0 = \sum_{i=1}^{\infty} E(u_0 w_i)$.*

Theorem 3.3, under less moment conditions, provides an extension of Theorem 3.1 in Liang, et al. (2005) [see also Theorem 4.11 of Wang (2005)] from $\alpha-$mixing sequence to near-epoch dependence. We mentioned that NED approach also allows for our results to be used in many practical important models such as bilinear, GARCH, threshold autoregressive models, etc. For the details, we refer to Davidson (2002).

# 4 Examples: verifications of (3.2) and (3.6)

As in Section 3.1 and 3.2, define a stationary causal process by

$$w_k = F(..., \eta_{k-1}, \eta_k), \quad k \in \mathbb{Z},$$

where $\eta_i, i \in \mathbb{Z}$, are i.i.d. random variables with mean zero and $E\eta_0^2 = 1$ and $F$ is a measurable function such that $Ew_0 = 0$ and $Ew_0^2 < \infty$.

In this section, we verify (3.2) and (3.6) for some practical important examples, including linear processes, nonlinear transformations of linear processes, nonlinear autoregressive time series and GARCH model. These examples partially come from Wu (2005) and Wu and Min (2005). For the convenience of reading, except mentioned explicitly, we use the notation as in Section 3, in particular, we recall the notation that $\{\eta_k^*\}_{k\in\mathbb{Z}}$ is an i.i.d. copy of $\{\eta_k\}_{k\in\mathbb{Z}}$ and independent of $(\epsilon_k, \eta_k)_{k\in\mathbb{Z}}$, and

$$w_k^* = F(..., \eta_{-1}^*, \eta_0^*, \eta_1, ..., \eta_{k-1}, \eta_k) \quad \text{and} \quad w_k' = F(..., \eta_{-1}, \eta_0^*, \eta_1, ..., \eta_{k-1}, \eta_k).$$

We mention that, due to the stationarity of $w_k$ and i.i.d. properties of $\eta_k$,

$$\begin{aligned} E|\mathcal{P}_0 w_n|^p &\leq E|w_n - w_n'|^p \\ &\leq C_p \big(E|w_n - w_n^*|^p + E|w_{n+1} - w_{n+1}^*|^p\big), \end{aligned} \tag{4.1}$$

for any $p \geq 1$, where $C_p$ is a constant depending only on $p$. As a consequence, both (3.2) and (3.6) hold if we can prove

$$E|w_n - w_n^*|^{2+\delta} \leq C\, n^{-4-3\delta}, \tag{4.2}$$

for some $\delta > 0$ and all $n$ sufficiently large.

## 4.1 Linear process and its nonlinear transformation

Consider a linear process $w_k$ defined by $w_k = \sum_{j=0}^{\infty} \theta_j \eta_{k-j}$ with $E\eta_0 = 0$. Routine calculation show that $w_k - w_k' = \theta_k(\eta_0 - \eta_0^*)$ and $w_k - w_k^* = \sum_{j=0}^{\infty} \theta_{j+k}(\eta_{-j} - \eta_{-j}^*)$. Hence,

- if $\sum_{j=1}^{\infty} j|\theta_j| < \infty$, $\sum_{j=1}^{\infty} j^{2+\delta}\theta_j^2 < \infty$ and $E|\eta_0|^{2+\delta} < \infty$ for some $\delta > 0$, then (3.2) and (3.6) hold true.

Indeed (3.6) follows from $\sum_{k=1}^{\infty} k\langle w_k - w_k'\rangle_{2+\delta} \leq \sum_{k=1}^{\infty} k \cdot |\theta_k| \cdot \langle \eta_0 - \eta_0^*\rangle_{2+\delta} < \infty$; and (3.2) from

$$\begin{aligned} \sum_{i=1}^{\infty} i^{1+\delta} \langle w_i - w_i^*\rangle_2^2 &= \sum_{i=1}^{\infty} i^{1+\delta} E[\sum_{j=i}^{\infty} \theta_j(\eta_{i-j} - \eta_{i-j}^*)]^2 \\ &\leq \sum_{i=1}^{\infty} i^{1+\delta} \sum_{j=i}^{\infty} \theta_j^2 E[(\eta_0 - \eta_0^*)]^2 \leq C \sum_{j=1}^{\infty} j^{2+\delta}\theta_j^2 < \infty. \end{aligned}$$

The result above can be easily extended to a nonlinear transformation of $w_k$. To see the claim, let

$$h_k = G(w_k) - EG(w_k),$$

where $G$ is a Lipschitz continuous function, i.e., there exists a constant $C < \infty$ such that

$$|G(x) - G(y)| \leq C|x - y|, \qquad \text{for all } x, y \in \mathbb{R}. \tag{4.3}$$

It is readily seen that (3.2) and (3.6) still hold true with the $w_k$ being replaced by $h_k$ by using the following facts:

$$|h_k - h_k'| \leq C|w_k - w_k'| \quad \text{and} \quad |h_k - h_k^*| \leq C|w_k - w_k^*|.$$

□

## 4.2 Nonlinear autoregressive time series

Let $w_n$ be generated recursively by

$$w_n = R(w_{n-1}, \eta_n), \; n \in \mathbb{Z}, \tag{4.4}$$

where $R$ is a measurable function of its components. Let

$$L_{\eta_0} = \sup_{x \neq x'} \frac{|R(x, \eta_0) - R(x', \eta_0)|}{|x - x'|}$$

be the Lipschitz coefficient. Suppose that, for some $q > 2$ and $x_0$,

$$E(\log L_{\eta_0}) < 0 \text{ and } E(L_{\eta_0}^q + |x_0 - R(x_0, \eta_0)|^q) < \infty. \tag{4.5}$$

Lemma 2 (i) of Wu and Min (2005) proved that there exist $C = C(q) > 0$ and $r_q \in (0, 1)$ such that, for all $n \in \mathbb{N}$,

$$E|w_n - w_n^*|^q \leq C r_q^n. \tag{4.6}$$

Since (4.6) implies (4.2), the $w_n$ defined by (4.4) satisfies (3.2) and (3.6).

We mention that the $w_n$ defined by (4.4) is a nonlinear autoregressive time series and the condition (4.5) can be easily verified by many popular nonlinear models such as threshold autoregressive (TAR), bilinear autoregressive, ARCH and exponential autoregressive (EAR) models. The following illustrations come from Examples 3-4 in Wu and Min (2005).

**TAR model**: $w_n = \phi_1 \max(w_{n-1}, 0) + \phi_2 \max(-w_{n-1}, 0) + \eta_n$. Simple calculation implies that if $L_{\eta_0} = \max(|\phi_1|, |\phi_2|) < 1$ and $E(|\eta_0|^q) < \infty$ for some $q > 0$, then (4.5) is satisfied.

**Bilinear model**: $w_n = (\alpha_1 + \beta_1 \eta_n) w_n + \eta_n$, where $\alpha_1$ and $\beta_1$ are real parameters and $E(|\eta_0|^q) < \infty$ for some $q > 0$. Note that $L_{\eta_0} = |\alpha_1 + \beta_1 \eta_0|$. (4.5) holds if only $E(L_{\eta_0}^q) < 1$. □

## 4.3 GARCH model

Let $\{w_t\}_{t \geq 1}$ be a GARCH$(l, m)$ model defined by

$$w_t = \sqrt{h_t} \eta_t \text{ and } h_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i w_{t-i}^2 + \sum_{j=1}^{l} \beta_j h_{t-j}, \tag{4.7}$$

where $\eta_t \sim i.i.d.$ with $E\eta_1 = 0$ and $E\eta_1^2 = 1$, $\alpha_0 > 0$, $\alpha_j \geq 0$ for $1 \leq j \leq m$, $\beta_i \geq 0$ for $1 \leq i \leq l$, and $h_0 = O_p(1)$. It is well-known that, if $\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{l} \beta_j < 1$, then $w_t$

is a stationary process having the following representation (see, e.g., Theorem 3.2.14 in Taniguchi and Kakizawa (2000)):

$$Y_t = M_t Y_{t-1} + b_t \quad \text{with } M_t = (\theta\eta_t^2, e_1, \ldots, e_{m-1}, \theta, e_{m+1}, \ldots, e_{l+m-1})^T,$$

where $Y_t = (w_t^2, \ldots, w_{t-m+1}^2, h_t, \ldots, h_{t-l+1})^T$ and $b_t = (\alpha_0\eta_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^T$ and $\theta = (\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_l)^T$; $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is the unit column vector with $i$th element being 1, $1 \le i \le l + m$.

Suppose that $E|\eta_0|^4 < \infty$ and $\rho[E(M_t^{\bigotimes 2})] < 1$, where $\rho(M)$ is the largest eigenvalue of the square matrix $M$ and $\bigotimes$ is the usual Kronecker product. Proposition 3 in Wu and Min (2005) implies for some $C < \infty$ and $r \in (0, 1)$,

$$E(|w_n - w_n^*|^4) \le C r^n. \tag{4.8}$$

Since (4.8) implies (4.2), the $w_n$ defined by (4.7) satisfies (3.2) and (3.6). □

# 5 Conclusion

On weak convergence to stochastic integrals, we have shown that the commonly used martingale and semimartingale structures can be extended to include the long memory processes, the causal processes and the near-epoch dependence in the innovations. Our frameworks can be applied to GARCH, TAR, bilinear and other nonlinear models. In econometrics with non-stationary time series, it is usually necessary to rely on the convergence to stochastic integrals. The authors hope these results derived in this paper prove useful in the related areas, particularly, in nonlinear cointegrating regression where endogeneity and nonlinearity play major roles.

# 6 Proofs

This section provides the proofs of all theorems. Except mentioned explicitly, the notation used in this section is the same as in previous sections.

*Proof of Theorem 2.1.* We may write

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})w_{k+1} =& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})(v_{k+1}+z_k-z_{k+1}) \\
=& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1} + \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}\Big[f\big(x_{nk}\big)-f\big(x_{n,k-1}\big)\Big]\,z_k + o_p(1) \\
=& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1} + R_n + o_P(1), \quad \text{say.} \qquad (6.1)
\end{aligned}$$

Write $\Omega_K = \{x_{ni} : \max_{1\le i\le n} ||x_{ni}|| \le K\}$. Since $f$ satisfies the local Lipschitz condition, it is readily seen from $\sup_k E||z_k u_k|| < \infty$ that, as $n\to\infty$,

$$E|R_n|I(\Omega_K) \ \le \ C_K\,\frac{1}{\sqrt{n}d_n}\sum_{k=1}^{n} E||z_k u_k|| \le C_K\,(n/d_n^2)^{1/2} \to 0.$$

This implies that $R_n = o_P(1)$ due to $P(\Omega_K)\to 1$, as $K\to\infty$. Theorem 2.1 follows from Theorem 1.1. □

*Proof of Theorem 2.2.* We may write

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})w_{k+1} =& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})(v_{k+1}+z_k-z_{k+1}) \\
=& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1} + \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}\Big[f\big(x_{nk}\big)-f\big(x_{n,k-1}\big)\Big]z_k + o_p(1) \\
=& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1} + \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}(x_{nk}-x_{n,k-1})\mathrm{D}f(x_{n,k-1})z_k + R_1(n) + o_p(1) \\
=& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1} + \frac{1}{\sqrt{n}d_n}\sum_{k=1}^{n-1} E(z_k u_k)\mathrm{D}f(x_{n,k-1}) + R_1(n) + R_2(n) + o_p(1), (6.2)
\end{aligned}$$

where the remainder terms are

$$\begin{aligned}
R_1(n) =& \ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} z_k\left[f(x_{nk})-f(x_{n,k-1})-(x_{nk}-x_{n,k-1})\mathrm{D}f(x_{n,k-1})\right] \\
R_2(n) =& \ \frac{1}{\sqrt{n}d_n}\sum_{k=1}^{n-1}[z_k u_k - E(z_k u_k)]\mathrm{D}f(x_{n,k-1}).
\end{aligned}$$

By virtue of Theorem 1.1, to prove (2.2), it suffices to show that

$$R_i(n) = o_P(1), \quad i=1,\ 2. \qquad (6.3)$$

To prove (6.3), write $\Omega_K = \{x_{ni} : \max_{1\le i\le n} ||x_{ni}|| \le K\}$. Note that **A3** implies that, for any $K > 0$ and $\max\{||x||, ||y||\} \le K$, $||\mathrm{D}f(x)|| \le C_K$ and

$$|f(x) - f(y) - (x-y)\,\mathrm{D}f(x)| \le C_K\, ||x-y||^{1+\beta'},$$

where $\beta' = \min\{\delta/(2+\delta), \beta\}$ for $\delta > 0$ given in **A4**(i). Then,

$$\begin{aligned} E|R_1(n)|I\big(\Omega_K\big) &\le \frac{C_K}{\sqrt{n}} \sum_{k=1}^n E\big(||x_{nk} - x_{n,k-1}||^{1+\beta'}\,|z_k|\big) \\ &\le C_K\, n^{-(1+\beta'/2)} \sum_{k=1}^n E(||u_k||^{1+\beta'}\,|z_k|) = O(n^{-\beta'/2}), \end{aligned} \tag{6.4}$$

where we have used the fact that, due to **A4**(i),

$$\sup_{k\ge1} E(||u_k||^{1+\beta'}|z_k|) \le \sup_{k\ge1}\big(E||u_k||^2\big)^{(1+\beta')/2} \sup_{k\ge1}\big(E|z_k|^{2+\delta}\big)^{1/(2+\delta)} < \infty.$$

This implies that $R_1(n) = O_P(n^{-\beta'/2})$ due to $P(\Omega_K) \to 1$ as $K \to \infty$.

It remains to show $R_2(n) = o_P(1)$. To this end, let $m = m_n \to \infty$ and $m_n \le \log n$. By recalling $\lambda_k = z_k u_k - E(z_k u_k)$, we have

$$\begin{aligned} R_2(n) &= \frac{1}{n\sigma}\sum_{k=1}^{2m} \lambda_k \mathrm{D}f(x_{n,k-1}) + \frac{1}{n\sigma}\sum_{k=2m}^{n-1} \lambda_k \mathrm{D}f(x_{n,k-m-1}) \\ &\quad + \frac{1}{n\sigma}\sum_{k=2m}^{n-1} \lambda_k\big[\mathrm{D}f(x_{n,k-1}) - \mathrm{D}f(x_{n,k-m-1})\big] = R_{21}(n) + R_{22}(n) + R_{23}(n). \end{aligned}$$

As in the proof of (6.4), it is readily seen from **A3** that

$$\begin{aligned} E|R_{21}(n)|I\big(\Omega_K\big) &\le C_K m n^{-1} \sup_{k\ge1} E||\lambda_k|| \le C_K n^{-1}\log n, \\ E|R_{23}(n)|I\big(\Omega_K\big) &\le C_K n^{-1} \sum_{k=1}^n E\big(||x_{n,k-1} - x_{n,k-m-1}||^{\beta'}\,||\lambda_k||\big) \\ &\le C_K n^{-1-\beta'/2} \sum_{k=1}^n \sum_{j=k-m}^{k-1} E(||u_j||^{\beta'}||\lambda_k||) \le C_K n^{-\beta'/2}\log n, \end{aligned}$$

where $\beta' = \min\{\delta/(2+\delta), \beta\}$. Hence $R_{21}(n) + R_{23}(n) = o_P(1)$ due to $P(\Omega_K) \to 1$ as $K \to \infty$. To estimate $R_{22}(n)$, write

$$\begin{aligned} IR_1(n) &= \frac{1}{n\sigma}\sum_{k=2m}^{n-1} \big[\lambda_k - E(\lambda_k \mid \mathcal{F}_{k-m-1})\big] x_k^*, \\ IR_2(n) &= \frac{1}{n\sigma}\sum_{k=2m}^{n-1} E(\lambda_k \mid \mathcal{F}_{k-m-1}) x_k^*, \end{aligned}$$

where $x_k^* = \mathrm{D}f(x_{n,k-m-1})I(\max_{1\le j\le k-m-1}||x_{nj}||\le K)$. Due to **A4** (iii) and **A3**,

$$|IR_2(n)| \ \le \ \frac{C_K}{n}\sum_{k=1}^{n}||E(\lambda_k \mid \mathcal{F}_{k-m-1})|| \le \sup_{k\ge 2m}||E(\lambda_k \mid \mathcal{F}_{k-m-1})|| = o_P(1).$$

Similarly, if **A4** (iii)$'$ and **A3** hold, then

$$E|IR_2(n)| \ \le \ \frac{C_K}{n}\sum_{k=1}^{n}E||E(\lambda_k \mid \mathcal{F}_{k-m-1})|| \le \sup_{k\ge 2m}E\,||E(\lambda_k \mid \mathcal{F}_{k-m-1})|| = o(1),$$

which yields $|IR_2(n)| = o_P(1)$. On the other hand, we have

$$IR_1(n) = \sum_{j=0}^{m} IR_{1j}(n),$$

where

$$IR_{1j}(n) = \frac{1}{n\sigma}\sum_{k=2m}^{n-1}\big[E(\lambda_k \mid \mathcal{F}_{k-j}) - E(\lambda_k \mid \mathcal{F}_{k-j-1})\big]x_k^*.$$

Let $\lambda_{1k}(j) = \big[E(\lambda_k \mid \mathcal{F}_{k-j}) - E(\lambda_k \mid \mathcal{F}_{k-j-1})\big]x_k^*$. Note that, for each $j\ge 0$,

$$IR_{1j}(n) = \frac{1}{n\sigma}\sum_{k=2m}^{n-1}\lambda_{1k}(j)$$

is a martingale with $\sup_{k\ge1}E||\lambda_{1k}(j)||^{1+\delta}\le C\sup_{k\ge1}E||\lambda_k||^{1+\delta}<\infty$ for some $\delta>0$. The classical result on strong law for martingale (see, e.g., Hall and Heyde (1980, Theorem 2.21, Page 41)) yields

$$IR_{1j}(n) = o_{a.s}(\log^{-2} n),$$

for each $0\le j\le m\le \log n$, implying $IR_1(n) = \sum_{j=0}^{m} IR_{1j}(n) = o_P(1)$.

We now have $R_{22}(n) = o_P(1)$ due to $P(\Omega_K)\to 1$ as $K\to\infty$, and the fact that, on $\Omega_k$,

$$R_{22}(n) = \frac{1}{n\sigma}\sum_{k=2m}^{n-1}\lambda_k x_k^* = IR_1(n) + IR_2(n) = o_P(1).$$

Combining these results, we prove $R_2(n) = o_P(1)$ and also complete the proof of (2.2). □

*Proof of Theorem 3.1.* Except mentioned explicitly, notation used in this section is the same as in Section 3.1. First note that

$$d_n^2 = var(\sum_{j=1}^{n}u_j) \sim c_\mu\, n^{3-2\mu}h^2(n), \ \text{with}\ c_\mu = \frac{1}{(1-\mu)(3-2\mu)}\int_0^\infty x^{-\mu}(x+1)^{-\mu}dx,$$

i.e., $d_n^2/n \to \infty$. See, e.g., Wang, Lin and Gullati (2003). By recalling (3.1) and using Theorem 2.1, Theorem 3.1 will follow if we may verify **A2**, i.e., on $D_{\mathbb{R}^2}[0,1]$,

$$\Big(\frac{1}{d_n}\sum_{j=1}^{[nt]} u_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} w_j\Big) \quad\Rightarrow\quad \big(B_{3/2-\mu}(t), B_{2t}\big). \tag{6.5}$$

We next prove (6.5). Since $\{(\epsilon_k, v_k), \mathcal{F}_k\}_{k\geq 1}$ forms a stationary martingale difference with covariance matrix $\Omega$, an application of the classical martingale limit theorem [see, e.g., Theorem 3.9 of Wang (2015)] yields that

$$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} v_j\Big) \Rightarrow \big(B_{1t}, B_{2t}\big), \tag{6.6}$$

on $D_{\mathbb{R}^2}[0,1]$. Recall that, for $k \geq 1$,

$$w_k^* = F(..., \eta_{-1}^*, \eta_0^*, \eta_1, ..., \eta_{k-1}, \eta_k),$$

where $\{\eta_k^*\}_{k\in\mathbb{Z}}$ is an i.i.d. copy of $\{\eta_k\}_{k\in\mathbb{Z}}$ and independent of $(\epsilon_k, \eta_k)_{k\in\mathbb{Z}}$. Let $v_k^* = \sum_{i=0}^{\infty} \mathcal{P}_k w_{i+k}^*$. Note that $\epsilon_{-i}$ is independent of $(\epsilon_i, v_i^*)$ for $i \geq 1$. If we have the condition:

$$\frac{1}{\sqrt{n}}\max_{1\leq k\leq n}\Big|\sum_{j=1}^{k}(v_j - v_j^*)\Big| = o_P(1), \tag{6.7}$$

it follows from (6.6) that

$$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_{-j},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} v_j\Big) \Rightarrow \big(B_t, B_{1t}, B_{2t}\big), \tag{6.8}$$

on $D_{\mathbb{R}^3}[0,1]$, where $B_t$ is a standard Brownian motion independent of $\big(B_{1t}, B_{2t}\big)$. Note that

$$\max_{1\leq k\leq n}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{k} w_j - \frac{1}{\sqrt{n}}\sum_{j=1}^{k} v_j\Big| \leq \max_{1\leq k\leq n}|z_k|/\sqrt{n} = o_P(1).$$

Result (6.8) implies that

$$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_{-j},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} w_j\Big) \Rightarrow \big(B_t, B_{1t}, B_{2t}\big),$$

on $D_{\mathbb{R}^3}[0,1]$. As a consequence, (6.5) follows from the continuous mapping theorem and similar arguments to those in Wang, Lin and Gullati (2003).

It remains to show that (3.2) implies (6.7). In fact, by noting $\{v_k - v_k^*, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference, it is readily seen from martingale maximum inequality that, for any $\epsilon > 0$,

$$P\Big(\max_{1\leq k\leq n}\Big|\sum_{j=1}^k (v_j - v_j^*)\Big| \geq \epsilon\sqrt{n}\Big) \leq \frac{2}{n\epsilon^2}\sum_{j=1}^n E(v_j - v_j^*)^2$$
$$\leq \frac{2}{n\epsilon^2}\sum_{k=1}^n E\Big[\sum_{i=0}^\infty \mathcal{P}_k(w_{i+k} - w_{i+k}^*)\Big]^2. \tag{6.9}$$

By Hölder's inequality and (3.2), we have

$$E\Big[\sum_{i=0}^\infty \mathcal{P}_k(w_{i+k} - w_{i+k}^*)\Big]^2 \leq \sum_{i=0}^\infty (i+k)^{-1-\epsilon}\sum_{i=0}^\infty (i+k)^{1+\epsilon} E\Big[\mathcal{P}_k(w_{i+k} - w_{i+k}^*)\Big]^2$$
$$\leq C\sum_{i=k}^\infty i^{1+\epsilon}\, E(w_i - w_i^*)^2 \to 0,$$

as $k\to\infty$. Taking this estimate into (6.9), we yield (6.7) and also complete the proof of Theorem 3.1. □

*Proof of Theorem 3.2.* As in the proof of Theorem 3.1, by recalling (3.4) and using Theorem 2.2, we only need to verify **A2**, i.e., on $D_{\mathbb{R}^2}[0,1]$,

$$\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} u_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} w_k\Big) \ \Rightarrow\ (B_{1t}, B_{2t}). \tag{6.10}$$

In fact, by noting that $\big\{(v_{1k}, v_{2k}), \mathcal{F}_k\big\}_{k\geq 1}$ forms a stationary martingale difference with $E\big(v_{10}^2 + v_{20}^2\big) < \infty$, the classical martingale limit theorem [see, e.g., Theorem 3.9 of Wang (2015)] yields that

$$\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} v_{1k}, \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} v_{2k}\Big) \Rightarrow (B_{1t}, B_{2t}),$$

on $D_{\mathbb{R}^2}[0,1]$, where $(B_{1t}, B_{2t})_{t\geq 0}$ is a 2-dimensional Gaussian process with zero means, stationary and independent increments, and covariance matrix:

$$\Omega_t = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{[nt]} cov\Big[\begin{pmatrix} v_{1k}\\ v_{2k}\end{pmatrix}\big(v_{1k}, v_{2k}\big)\Big] = \Omega\, t.$$

As a consequence, we have

$$\big(x_{n,[nt]}, y_{n,[nt]}\big) = \Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} v_{1k}, \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} v_{2k}\Big) + R_{n,t}$$
$$\Rightarrow (B_{1t}, B_{2t}),$$

due to the fact that, by recalling $E(|z_{10}|^2+|z_{20}|^2)<\infty$,

$$\sup_{0\le t\le 1}||R_{n,t}||\le \max_{1\le k\le n}(|z_{1k}|+|z_{2k}|)/\sqrt{n}=o_P(1).$$

This yields (6.10), and also completes the proof of Theorem 3.2. □

*Proof of Corollary 3.1.* We only need to verify **A5**. First of all, simple calculation shows that $\mathcal{P}_k u_{i+k}=\varphi_i\epsilon_k$. As a consequence, $\sum_{i=1}^{\infty} i\,\langle \mathcal{P}_0 u_i\rangle_2<\infty$, that is, A5 (i) holds.

Due to (4.1), **A5** (ii) is implied by (3.6). It remains to show that A5 (iii) holds true if $\sum_{t=1}^{\infty} t\,\langle w_t-w_t^{'}\rangle_2<\infty$, as the latter is a consequence of (3.6). In fact, by letting $\sum_{i=k}^{j}=0$ if $j<k$, we may write

$$\begin{aligned}
E\big(\widetilde{\lambda}_k\mid \mathcal{F}_{k-m}\big) &= \sum_{j=-\infty}^{k-m}\mathcal{P}_j(u_k z_{2,k})=\sum_{i=0}^{\infty}\varphi_i\sum_{j=m}^{\infty}\mathcal{P}_{k-j}(\epsilon_{k-i}z_{2,k})\\
&= \sum_{i=0}^{\infty}\varphi_i\Big(\sum_{j=m}^{\max\{m,i\}}+\sum_{j=\max\{m,i\}+1}^{\infty}\Big)\mathcal{P}_0(\epsilon_{j-i}z_{2,j})\\
&= \sum_{i=0}^{\infty}\varphi_i\sum_{j=m}^{\max\{m,i\}}\mathcal{P}_0(\epsilon_{j-i}z_{2,j})+\sum_{i=0}^{\infty}\varphi_i\sum_{j=\max\{m,i\}+1}^{\infty}\sum_{t=1}^{\infty}\mathcal{P}_0(\epsilon_{j-i}w_{t+j})\\
&:= A_{1m}+A_{2m}. \qquad (6.11)
\end{aligned}$$

It is readily seen from $E|z_{2k}|^2=E|z_{20}|^2<\infty$ that

$$E|A_{1m}|\ \le\ 2\sum_{i=m}^{\infty} i\,|\varphi_i|(E\epsilon_0^2)^{1/2}\,(Ez_{20}^2)^{1/2}\to 0,$$

as $m\to\infty$. As for $A_{2m}$, by noting $\mathcal{P}_0(\epsilon_{j-i}w_{t+j})=E\big[\epsilon_{j-i}(w_{t+j}-w_{t+j}^{'})\mid \mathcal{F}_0\big]$ whenever $j>i$, we have

$$\begin{aligned}
E|A_{2m}| &\le \sum_{i=0}^{\infty}|\varphi_i|\sum_{j=m+1}^{\infty}\sum_{t=1}^{\infty}E|\epsilon_{j-i}(w_{t+j}-w_{t+j}^{'})|\\
&\le C\sum_{j=m+1}^{\infty}\sum_{t=1+j}^{\infty}\langle w_t-w_t^{'}\rangle_2\\
&\le C\sum_{t=m}^{\infty} t\,\langle w_t-w_t^{'}\rangle_2\to 0,
\end{aligned}$$

as $m\to\infty$. Taking these estimates into (6.11), we obtain

$$E\big[\sup_{k\ge 2m}|E\big(\widetilde{\lambda}_k\mid \mathcal{F}_{k-m}\big)|\big]\ \le\ E|A_{1m}|+E|A_{2m}|\to 0,$$

implying A5 (iii). □

*Proof of Theorem 3.3.* First note that, under **A6**, it follows from Theorem 17.5 of Davidson (1994) that $w_k, k \in \mathbb{Z}$, is a stationary $\mathcal{L}_{2+\delta}$-mixingale of size $-1$ with constant $\langle w_0 \rangle_4$,

$$\begin{aligned} \langle E(w_k \mid \mathcal{F}_{k-m}) \rangle_{2+\delta} &\leq C \, \langle w_1 \rangle_4 \, m^{-\gamma}, \quad (6.12) \\ \langle w_k - E(w_k \mid \mathcal{F}_{k+m}) \rangle_{2+\delta} &\leq C \, \langle w_1 \rangle_4 \, m^{-\gamma}, \quad (6.13) \end{aligned}$$

hold for all $k$, $m \geq 1$ and some $\gamma > 1$. Furthermore, by Theorem 16.6 of Davidson (1994), we may write

$$w_k = v_k + z_{k-1} - z_k,$$

where, as in Section 3.2,

$$v_k = \sum_{i=0}^{\infty} \mathcal{P}_k w_{i+k}, \quad z_k = \sum_{i=1}^{\infty} E(w_{i+k} | \mathcal{F}_k).$$

It is readily seen that both $v_k$ and $z_k$ are stationary and $(v_k, \mathcal{F}_k)_{k \geq 1}$ forms a martingale difference with $Ev_1^2 \leq 2Ew_1^2 + 4Ez_1^2 < \infty$, since, by (6.12), the following result holds (implying $Ez_1^2 < \infty$):

$$\langle z_{k,j} \rangle_{2+\delta} \leq \sum_{i=j+1}^{\infty} \langle E(w_i | \mathcal{F}_0) \rangle_{2+\delta} \leq C \langle w_1 \rangle_4 \sum_{i=j+1}^{\infty} i^{-\gamma} < \infty, \quad (6.14)$$

for any $j \geq 0$, where $z_{k,j} = \sum_{i=j+1}^{\infty} E(w_{i+k} | \mathcal{F}_k)$. By (6.12) and (6.13), for any $k \geq 1$, we also have

$$\begin{aligned} |E(w_1 w_k)| &\leq E\big(|w_1 - w_1^*| \, |w_k|\big) + E\big[|w_1^*| \, |E(w_k \mid \mathcal{F}_{k/2})|\big] \\ &\leq \langle w_1 \rangle_2 \big\{ \langle w_1 - w_1^* \rangle_2 + \langle E(w_k \mid \mathcal{F}_{k/2}) \rangle_2 \big\} \\ &\leq C \, \langle w_1 \rangle_2 \langle w_1 \rangle_4 \, k^{-\gamma}, \quad (6.15) \end{aligned}$$

where $w_1^* = E(w_1 \mid \mathcal{F}_{k/2})$. The result (6.15) will be used later.

Since $w_k$ has structure (1.3) with the $v_k$ satisfying **A1**, (3.8) implies **A2** and **A6** (iii) and (6.14) with $j = 0$ imply **A4** (i), by using Theorem 2.2, Theorem 3.3 will follow if we prove (3.7) and

$$\sup_{k \geq 2m} E||E\big(\lambda_k \mid \mathcal{F}_{k-m}\big)|| \to 0, \quad (6.16)$$

where $\lambda_k = z_k u_k - E z_k u_k$, as $m \to \infty$.

By recalling the stationarity of $(u_k, w_k)_{k\geq 1}$, to prove (3.7), it suffices to show that $\Omega_1, \Omega_2$ and $\rho$ are finite. In fact (6.15) implies that $|\Omega_2| \leq E w_0^2 + C \sum_{j=1}^{\infty} j^{-\gamma} < \infty$. Similarly, we may prove that $(u_k)_{k\geq 1}$ is a stationary $\mathcal{L}_2$-mixingale of size $-1$ with constant $\langle u_0 \rangle_4$. As a consequence, the same argument yields $|\Omega_1| < \infty$ and $|\rho| < \infty$.

In order to prove (6.16), let $z_k^* = z_k - z_{k,\alpha_m} = \sum_{i=1}^{\alpha_m} E(w_{i+k}|\mathcal{F}_k)$,

$$\lambda_{k,1} = z_k^* u_k - E z_k^* u_k, \qquad \lambda_{k,2} = z_{k,\alpha_m} u_k - E z_{k,\alpha_m} u_k,$$

where $\alpha_m \to \infty$ and $z_{k,\alpha_m}$ is given as in (6.14). Due to (6.14), we have

$$E||E(\lambda_{k,2} \mid \mathcal{F}_{k-m})|| \leq E||\lambda_{k,2}|| \leq 2 \langle z_{k,\alpha_m} \rangle_2 \langle u_0 \rangle_2 \to 0, \tag{6.17}$$

as $m \to \infty$, uniformly for any $k \geq 2m$ and any integer sequence $\alpha_m \to \infty$. By recalling that $u_k$ is adapted to $\mathcal{F}_k$ and $\mathcal{F}_{k-m} \subset \mathcal{F}_k$, we may write

$$E||E(\lambda_{k,1} \mid \mathcal{F}_{k-m})|| \leq \sum_{i=1}^{\alpha_m} E||E(A_k \mid \mathcal{F}_{k-m})||,$$

where $A_k = u_k w_{i+k} - E u_k w_{i+k}$. Since both $u_k$ and $w_k$ are $\mathcal{L}_2$-NED of size $-1$, Corollary 17.11 of Davidson (1994) implies that $A_k$ is $\mathcal{L}_1$-NED of size $-1$. As a consequence, as in the proof of (6.12), there exist a sequence of $v_m$ such that $v_m \to 0$ and

$$E||E(A_k \mid \mathcal{F}_{k-m})|| \leq C\, v_m.$$

Hence, uniformly for $k \geq 2m$,

$$E||E(\lambda_{k,1} \mid \mathcal{F}_{k-m})|| \leq C\, \alpha_m v_m \to 0,$$

as $m \to \infty$, by taking $\alpha_m$ to be such an integer sequence that $\alpha_m \to \infty$ and $\alpha_m v_m \to 0$. This, together with (6.17), yields

$$\sup_{k\geq 2m} E||E(\lambda_k \mid \mathcal{F}_{k-m})|| \leq C\, (\alpha_m v_m + 2 \langle z_{k,\alpha_m} \rangle_2 \langle u_0 \rangle_2) \to 0,$$

as $m \to \infty$, as required. The proof of Theorem 3.3 is now complete. □

**Acknowledgements:** This work is completed when the first author Jiangyan Peng visited the University of Sydney under the financial support of China Scholarship Council (CSC) from Chinese government. Peng thanks the University of Sydney for providing friendly research environment. Peng also acknowledges research support from the

National Natural Science Foundation of China (project no: 71501025), Applied Basic Project of Sichuan Province (grant number: 2016JY0257) and China Postdoctoral Science Foundation (grant number: 2015M572467). Wang acknowledges research support from Australian Research Council.

## REFERENCES

Borovski, Y. and Korolyuk, V. (1997). Martingale approximation. *VSP*, Utrecht.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press.

Davidson J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics*, **106**, 243-269.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press.

Jacod, J. and A. N. Shiryaev (1987/2003). *Limit Theorems for Stochastic Processes.* New York: Springer–Verlag.

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, **19**, 1035–1070.

Liang, H. Y., Phillips, P. C. B., Wang, H. C. and Wang Q. (2015). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory,* **0**, 1-27.

Lin, Z. and Wang, H. (2015). On convergence to stochastic integrals. *Journal of theoretical probability*, 1-20.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for Linear Processes, *Ann. Statist.* **20**, 971-1001.

Taniguchi, M and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series,* New York: Springer.

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

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, **25**, 710–738.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, **77**, 1901–1948.

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression.* World Scientific.

Wang, Q. and Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, **32**, 359-401.

Wu, W. (2005). Nonlinear system theory: another look at dependence. *Proc Natl Acad Sci*, **101**, 14150-14154.

Wu, W. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic processes and their applications*, **115**, 939-958.

Wu, W. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.*, **35**, 2294-2320.