Econometric Theory, **34**, 2018, 1132–1157.
doi:10.1017/S0266466617000408

WEAK CONVERGENCE TO STOCHASTIC INTEGRALS UNDER PRIMITIVE CONDITIONS IN NONLINEAR ECONOMETRIC MODELS

JIANGYAN PENG
University of Electronic Science and Technology of China

QIYING WANG
University of Sydney

Limit theory with 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 semi-martingale structures. Liang, Phillips, Wang, and Wang (2016) (see also Wang (2015), Chap. 4.5) currently extended weak convergence to stochastic integrals by allowing for a linear process or a α-mixing sequence in innovations. While these martingale, linear process and α-mixing structures have wide relevance, they are not sufficiently general to cover many econometric applications that have endogeneity and nonlinearity. 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 innovations, which have applications in a wide range of econometric areas such as TAR, bilinear, and other nonlinear models.

1. INTRODUCTION

In econometrics with nonstationary time series, it is usually necessary to rely on convergence to stochastic integrals. This result is particularly vital to unit root testing linear and nonlinear cointegrating regression. To illustrate, in Section 5, we investigate an application of the limit theorems involving stochastic integrals in nonlinear cointegrating regression. For more examples, we refer to Park and Phillips (2000, 2001), Chang, Park, and Phillips (2001), Chan and Wang (2015),

The authors thank Professor Peter Phillips, Professor Saikkonen and two anonymous referees for their very helpful comments on the original version. This work was completed when Jiangyan Peng visited the University of Sydney with financial support from the China Scholarship Council (CSC). Peng thanks the University of Sydney for providing a 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 the China Postdoctoral Science Foundation (grant number: 2015M572467). Wang acknowledges research support from the Australian Research Council. Address corresponding to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia; e-mail: qiying@maths.usyd.edu.au.

Wang and Phillips (2009a, 2009b, 2016), Wang (2015, Chap. 5) and the references 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 for 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*

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

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

Kurtz and Protter (1991) [also see Jacod and Shiryaev (2003)] actually established a result with y_{nk} as a semi-martingale rather than **A1**. Toward a general result beyond the semi-martingale, Liang et al. (2016) and Wang (2015, Chap. 4.5) investigated an extension to linear process innovations and provided a convergence result for sample quantities $\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$ to functionals of stochastic processes, 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 $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$, and v_k being defined as in **A1**. Liang et al. (2016) and Wang (2015, Chap. 4.5) further considered an extension to α-mixing innovations. For other related results, we refer to Ibragimov and Phillips (2008), De Joon (2004), Chang and Park (2011), and Lin and Wang (2015).

While these results are elegant, they are not sufficiently general to cover many econometric applications that have endogeneity and more general innovation processes. In particular, the linear structure in (1.2) or a α-mixing sequence considered in Liang et al. (2016) is well-known to be restrictive and fails to include many important practical models such as threshold, nonlinear autoregressions, and so on. The aim of this paper is to fill this gap by providing new general results for the convergence to stochastic integrals in which there are some advantages in econometric applications. Explicitly, our frameworks consider the convergence of $S_n := \sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1}$, where w_k has the form:

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

with z_k satisfying certain regularity conditions specified in the next section and v_k being defined as in **A1**. 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 a martingale. Martingale approximation has been widely investigated in the literature. For current developments, we refer to Borovskikh and Korolyuk (1997). As shown in Section 3, these existing results for 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 covers the case in which u_k is a long memory process, while Theorem 2.2 addresses the case where u_k is a short memory process. The section shows 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 this is not the case when u_k is a long memory process under minor natural conditions in z_t. Section 3 provides three corollaries to 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 consider nonlinear cointegrating regression in Section 5, where the focus is the impact of endogeneity and nonlinearity to asymptotic behaviors of the parametric estimators. We conclude in Section 6. Proofs of all theorems are given in Section 7.

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 $x = (x_1, \ldots, x_m)$, we use the notation $||x|| = \sum_{j=1}^{m} |x_j|$. For a sequence of increasing σ-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 $\langle Z\rangle_p = (E|Z|^p)^{1/p} < \infty$. For a real function $f(x)$ on R^d, $f(x)$ is said to satisfy a local Lipschitz condition if, for any $K > 0$, there exists a constant C_K such that, for all $||x|| + ||y|| < K$,

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

Where there is no confusion, we generally use the index notation $x_{nk}(y_{nk})$ for $x_{n,k}(y_{n,k})$. All other notation is standard.

2. MAIN FRAMEWORKS

In this section, we establish the frameworks for convergence to stochastic integrals. Except for those mentioned explicitly, the notation is the same as in 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* $g(s)$ *in* R^d, *and any function* $f(x)$ *in* R^d *satisfying a local Lipschitz condition, we have*

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

As Liang et al. (2016) note, the local Lipschitz condition is a minor requirement and holds 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) does not modify the limit behaviors under minor natural conditions of 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 σ, which generally holds for short memory processes u_t. In this case, as the following theorem shows, z_t has an essential impact on the limit distributions.

Let $\mathrm{D}f(x) = \big(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}\big)'$. The development of our theory requires the following additional assumptions:

A3 $\mathrm{D}f(x)$ *is continuous in* 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 $\delta > 0$;

(ii) $E z_k u_k \to A_0 = (A_{10}, \ldots, A_{d0})$, as $k \to \infty$;

Set $\lambda_k = z_k u_k - E z_k u_k$.

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

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

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

Remark 1. Condition **A3** is similar to that in previous studies (see, e.g., Wang, 2015; Liang et al., 2016). We require the moment condition $\sup_{k\geq 1} E|z_k|^{2+\delta}<\infty$ for some $\delta>0$ in **A4** (i) to remove the effect of higher orders from z_k. In terms of the convergence in (2.2), $\sup_{k\geq 1} E|z_k|^2<\infty$ is essential. It is not clear at the moment if we can reduce δ in **A4** (i) 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., we can denote w_k as in the structure of (1.3) (see, e.g., Phillips and Solo, 1992). For this w_k, Theorem 4.9 in Wang (2015) [also see Liang et al., 2016] established a result 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 this is required for all $i\geq 1$, (2.3) is difficult to verify when u_k is a nonlinear stationary process, such as $u_k=F(\epsilon_k,\epsilon_{k-1},\ldots)$, even in the case that (ϵ_k,v_k) are independent and identically distributed (i.i.d.) random vectors. In comparison, we can easily apply **A4** (ii) and (iii) [or (iii)′] to stationary causal processes and mixing sequences, as 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 since we do not require the condition $\sup_{n\geq 1}\frac{1}{\sqrt{n}}\sum_{k=1}^{n} E|z_{k-1}-z_k|<\infty$. Consequently, Theorems 2.1–2.2 provide an essential extension for the convergence to stochastic integrals, rather than a simple corollary to previous works. For related results on convergence to stochastic integrals, we refer to Kurtz and Protter (1991), Phillips (1988a), Jacod and Shiryaev (2003), Ibragimov and Phillips (2008), and Lin and Wang (2015).

3. THREE USEFUL COROLLARIES

This section investigates direct applications of Theorems 2.1 and 2.2. Section 3.1 considers the case where u_k is a long memory process and w_k is a stationary causal

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

process. Section 3.2 contributes to convergence for both u_k and w_k as stationary causal processes. Finally, in Section 3.3, we investigate the impact of near-epoch dependence in convergence to stochastic integrals. A detailed verification of assumptions for more practical models such as GARCH and nonlinear autoregressive time series will be given 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 ∞. Let F be a measurable function such that

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

is a well-defined stationary random variable with $Ew_0 = 0$ and $0 < Ew_0^2 < \infty$. The w_k is a stationary causal process that is discussed extensively 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 $\mathcal{F}_k = \sigma(\epsilon_i, \eta_i, i \le k)$ and assume $\sum_{i=1}^{\infty} i\, \langle \mathcal{P}_0 w_i \rangle_2 < \infty$. 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) in the cited paper. All processes w_k, v_k, and z_k are stationary and satisfy 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 $\Omega \cdot t$ and B_t is 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

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

After these notations, a simple application of Theorem 2.1 yields the following result for the case where 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 $w_k^* = F(\ldots, \eta_{-1}^*, \eta_0^*, \eta_1, \ldots c, \eta_{k-1}, \eta_k)$, *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*

$$\left\{ x_{n,\lfloor nt \rfloor},\ y_{n,\lfloor nt \rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(x_{nk})\, w_{k+1} \right\}$$

$$\Rightarrow \left\{ B_{3/2-\mu}(t), B_{2t}, \int_0^1 g\left[B_{3/2-\mu}(t)\right] dt, \int_0^1 f\left[B_{3/2-\mu}(t)\right] dB_{2t} \right\}. \tag{3.3}$$

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

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

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

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

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

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

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

Write $\mathcal{F}_k = \sigma(\epsilon_i, \eta_i, i \le k)$,

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

We use the following assumption 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 $\widetilde{\lambda}_k = u_k z_k - E u_k z_k$.

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

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

As in Section 3.1, all $u_k, w_k, z_{1k}, z_k, v_{1k}$ and v_k are stationary with decompositions of:

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

Furthermore, **A5** (i) [(ii), respectively] implies that $E\left(v_{10}^2 + z_{10}^2\right) < \infty$ $[E\left(|v_0|^{2+\delta} + |z_0|^{2+\delta}\right) < \infty$, respectively]. Consequently, it follows that

$$E|u_k z_k| < \infty \quad \text{and} \quad A_0 := E u_0 z_0 = \sum_{i=1}^{\infty} E(u_0 w_i) < \infty.$$

Now let $\Omega = \begin{pmatrix} 1 & \sigma^{-1} E v_{10} v_0 \\ \sigma^{-1} E v_{10} v_0 & E v_0^2 \end{pmatrix}$, where $\sigma^2 = E v_{10}^2$, and (B_{1t}, B_{2t}) be a bivariate Brownian motion with covariance matrix $\Omega \cdot t$. An application of Theorem 2.2 gives the following result.

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

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

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

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

Theorem 3.2 provides a rather general result for both u_t and w_t as causal processes. In related research using a quite complicated technique from Jacod and Shiryaev (2003), Lin and Wang (2015) considered the case in which $u_t = w_t$. In comparison, by using Theorem 2.2, our proof is quite simple, as shown in Section 6. Furthermore, our condition **A5** is easy to verify. The following corollary provides an illustration that investigates the case where 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* $\sum_{i=1}^{\infty} i|\varphi_i| < \infty$ *and* $\varphi = \sum_{j=0}^{\infty} \varphi_j \neq 0$. *Result* (3.5) *holds true with* $\sigma = \varphi$ *and* $A_0 = \sum_{j=0}^{\infty}\sum_{i=1}^{\infty} \varphi_j E\epsilon_{-j} w_i$, *if, in addition to A3* (*with* $d = 1$),

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

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

We require Condition (3.6) 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 the 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 in Section 6. Some examples for w_k satisfying (3.6), including nonlinear transformations of linear processes, nonlinear autoregressive time series, and GARCH 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 $\mathcal{F}_s^t = \sigma(\eta_s, \ldots, \eta_t)$ 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 dependent on $\{\eta_k\}_{k\in\mathbb{Z}}$ in $\mathcal{L}_P$-norm for $p > 0$ if

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

where d_t is a sequence of positive constants, and $\nu(m) \to 0$ as $m \to \infty$. In 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 in 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) $\eta_k = (\eta_{k1},\ldots,\eta_{km}), k \in \mathbb{Z}$, is α-mixing of size -6;[1]
(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 $E\left(||u_0||^4 + |w_0|^4\right) < \infty$.

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

$$\Omega_1 = Eu_0' u_0 + 2\sum_{i=1}^{\infty} Eu_0' u_i, \quad \Omega_2 = Ew_0^2 + 2\sum_{i=1}^{\infty} Ew_0 w_i,$$

$$\rho = Eu_0' w_0 + \sum_{i=1}^{\infty}(Eu_0' w_i + Eu_i' w_0).$$

In terms of (3.7) and **A6**, Corollary 29.19 in Davidson (1994, Page 494) yields, as $n \to \infty$,

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

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

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

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

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

Theorem 3.3, under less moment conditions, extends Liang et al.'s (2016) Theorem 3.1 [see also Theorem 4.11 in Wang (2015)] from a α-mixing sequence to near-epoch dependence. We mention that the NED approach also allows the use of our results in many important practical models, such as the bilinear, GARCH, threshold autoregressive models, and so on. For details, we refer to Davidson (2002).

4. EXAMPLES: VERIFICATIONS OF (3.2) AND (3.6)

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

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

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

In this section, we verify (3.2) and (3.6) for some important practical examples, including linear processes, nonlinear transformations of linear processes, nonlinear autoregressive time series, and GARCH model. We mention that, when w_t is generated by a GARCH model, asymptotic behaviors in Theorems 3.1 and 3.2 can also be considered by Theorem 1.1 with certain modifications since, in this situation, w_t forms a martingale difference.

The examples considered in this section partially come from Wu (2005) and Wu and Min (2005). For convenience and except where mentioned explicitly, we use the same notation as in Section 3; in particular, recall 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

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

We mention that due to the stationarity of w_k and the i.i.d. properties of η_k,

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

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

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

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

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

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

We can easily extend the result above to a nonlinear transformation of w_k. To illustrate, 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

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

It is readily apparent that (3.2) and (3.6) still hold true when replacing w_k with h_k using the following facts:

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

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

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

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 $n \in \mathbb{N}$,

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

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 in (4.4) is a nonlinear autoregressive time series and we can easily verify the condition (4.5) using many popular nonlinear models, such as the 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: $w_n = (\alpha_1 + \beta_1 \eta_{n-1}) w_{n-1} + \eta_n$, where α_1 and β_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 only if $E(L_{\eta_0}^q) < 1$.

4.3. GARCH Model

Let $\{w_t\}_{t \ge 1}$ 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 $E\eta_1^2 = 1$, $\alpha_0 > 0$, $\alpha_j \ge 0$ for $1 \le j \le m$, $\beta_i \ge 0$ for $1 \le i \le l$, and $h_0 = O_p(1)$. If $\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{l} \beta_j < 1$, then w_t is a stationary process with 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 ith 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 that 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 in (4.7) satisfies (3.2) and (3.6).

5. NONLINEAR COINTEGRATING REGRESSION

There are extensive applications in econometrics for the limit theorems involving stochastic integrals such as those given in Theorems 3.1 and 3.2 (or Corollary 3.1). They arise frequently in time series regressions with integrated and near integrated processes, unit root testing and nonlinear co-integration theory. See, for instance, Park and Phillips (2000, 2001), Chang et al. (2001), Chan and Wang (2015), Wang and Phillips (2009a, 2009b, 2016) and Wang (2015). As noticed in Liang et al. (2016), using the theorems given in that paper, previous results

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

may be extended to a wider class of generating mechanisms such as those involving nonlinear functions and long memory innovations. The following nonlinear cointegrating regression model illustrates the use of the methods discussed in this paper. As in Liang et al. (2016), we focus on the impact of endogeneity and nonlinearity to the asymptotics, rather than the generality of the model. More applications will be considered in subsequent researches.

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 $E\epsilon_0^2 = E\eta_0^2 = 1$. Let

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

where F is a measurable function such that w_k is a well-defined stationary causal process with $Ew_0 = 0$ and $0 < Ew_0^2 < \infty$. Let u_k be a linear process defined by $u_k = \sum_{j=0}^{\infty} \varphi_j \epsilon_{k-j}$, with coefficients φ_j, $j \geq 0$, satisfying one of the following conditions:

C1. $\varphi_j \sim j^{-\mu}\rho(j)$, where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞.

C2. $\sum_{i=1}^{\infty} i|\varphi_i| < \infty$ and $\varphi \equiv \sum_{j=0}^{\infty} \varphi_j \neq 0$.

We consider the nonlinear in variables cointegrating model:

$$y_t = \alpha + \beta h(x_t) + w_{t+1}, \tag{5.1}$$

where $x_t = \sum_{k=1}^{t} u_k$ and $h(x)$ is an asymptotically homogeneous real function, i.e., there exist real functions $\nu(\lambda) > 0$, $H(x)$ and $R(\lambda x)$ (negligible when $\lambda \to \infty$ or $\lambda x \to \infty$) so that

$$h(\lambda x) = \nu(\lambda) H(x) + R(\lambda x). \tag{5.2}$$

Under given conditions, there exist endogeneity and nonlinearity in the model (5.1) and the regressor x_t is driven by short (under C2) or long (under C1) memory innovations u_k. We may write the least square estimators (LSE) of α and β as

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{t=1}^{n} y_t h(x_t) - n^{-1}\sum_{t=1}^{n} h(x_t)\sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} h(x_t)^2 - n^{-1}[\sum_{t=1}^{n} h(x_t)]^2} \\ &= \beta + \frac{\sum_{t=1}^{n} w_t\big[h(x_t) - n^{-1}\sum_{t=1}^{n} h(x_t)\big]}{\sum_{t=1}^{n} h(x_t)^2 - n^{-1}[\sum_{t=1}^{n} h(x_t)]^2},\end{aligned} \tag{5.3}$$

$$\begin{aligned}\hat{\alpha} &= \frac{1}{n}\sum_{t=1}^{n} y_t - \frac{\hat{\beta}}{n}\sum_{t=1}^{n} h(x_t) \\ &= \alpha + \frac{1}{n}\sum_{t=1}^{n} w_t - \frac{\hat{\beta}-\beta}{n}\sum_{t=1}^{n} h(x_t).\end{aligned} \tag{5.4}$$

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

The asymptotics of $\hat{\alpha}$ and $\hat{\beta}$ clearly depend on the structure of w_t, endogeneity and nonlinearity in the model (5.1), which can be considered by using some modifications of Theorem 3.1 and Corollary 3.1. To show this, we need the conditions imposed on w_t as in Theorem 3.1 and Corollary 3.1, and further conditions on $h(x)$ having a representation of (5.2).

For convenience of reading, we use the notation as in Section 3.1, i.e., we write

$$v_k = \sum_{i=0}^{\infty} \mathcal{P}_k w_{i+k}, \quad z_k = \sum_{i=1}^{\infty} E(w_{i+k}|\mathcal{F}_k), \quad \Omega = \begin{pmatrix} 1 & \rho \\ \rho & Ev_0^2 \end{pmatrix},$$

where $\rho = E\epsilon_0 v_0 = \sum_{i=0}^{\infty} E\epsilon_0 w_i$. We further let (B_{1t}, B_{2t}) be a bivariate Brownian motion with covariance matrix $\mathbf{\Omega} \cdot t$ and define a fractional Brownian motion $B_H(t)$ as in Section 3.1. We always assume $\sum_{i=1}^{\infty} i \, \langle \mathcal{P}_0 w_i \rangle_2 < \infty$, indicating that $E(v_0^2 + z_0^2) < \infty$. We also use the following assumption on $h(x)$.

A7 (i) $H(x)$ satisfies a local Lipschitz condition;

(ii) $H'(x)$ is continuous on R and

$$|H'(x) - H'(y)| \le C_K |x-y|^{\beta},$$

for some $\beta > 0$ and all $\max\{|x|, |y|\} \le K$, where C_K is a constant depending only on K;

(iii) There exists an $a(\lambda)$ so that $|R(\lambda x)| \le a(\lambda)(1 + |x|^{\gamma})$ for some $\gamma > 0$ and

$$|R(\lambda x) - R(\lambda y)| \le a(\lambda)\, |x-y|,$$

whenever x and y are in a compact set;

(iv) As $\lambda \to \infty$, $\lim_{\lambda\to\infty} a(\lambda)/\nu(\lambda) = 0$.

The following results provide the asymptotics of $\hat{\alpha}$ and $\hat{\beta}$.

THEOREM 5.1. *Suppose that (3.2),* **A7***(i) and (iii) hold. If* $u_k = \sum_{j=0}^{\infty} \varphi_j \epsilon_{k-j}$, *with coefficients* φ_j, $j \ge 0$, *satisfying C1, then*

$$\nu(d_n)\sqrt{n}\big(\hat{\beta}_n - \beta\big) \to_D \frac{\int_0^1 H[B_{3/2-\mu}(t)]dB_{2t} - B_{2t}\int_0^1 H[B_{3/2-\mu}(t)]dt}{\int_0^1 H^2[B_{3/2-\mu}(t)]dt - \big[\int_0^1 H[B_{3/2-\mu}(t)]dt\big]^2}, \tag{5.5}$$

where $d_n^2 = c_\mu n^{3-2\mu}\rho^2(n)$ *for some* $c_\mu > 0$. *If, in addition* $\nu(\lambda) \to \infty$, *then*

$$\sqrt{n}\big(\hat{\alpha}_n - \alpha\big) \to_D N(0, Ev_0^2). \tag{5.6}$$

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

THEOREM 5.2. *Suppose that (3.6),* **A7***(ii) and (iii) hold. If* $u_k = \sum_{j=0}^{\infty} \varphi_j \epsilon_{k-j}$, *with coefficients* φ_j, $j \geq 0$, *satisfying C2, then*

$$\nu(\sqrt{n}\varphi)\sqrt{n}\big(\hat{\beta}_n - \beta\big) \to_D \frac{\int_0^1 H(B_{1t})dB_{2t} + A_0 \int_0^1 H'(B_{1t})dt - B_{2t}\int_0^1 H(B_{1t})dt}{\int_0^1 H^2(B_{1t})dt - \big[\int_0^1 H(B_{1t})dt\big]^2}, \quad \textbf{(5.7)}$$

where $A_0 = \sum_{j=0}^{\infty}\sum_{i=1}^{\infty} \varphi_j E\epsilon_{-j} w_i$. *If, in addition* $\nu(\lambda) \to \infty$, *then*

$$\sqrt{n}\big(\hat{\alpha}_n - \alpha\big) \to_D N(0, Ev_0^2). \quad \textbf{(5.8)}$$

By virtue of (5.2), the proofs of Theorems 5.1 and 5.2 follow easily from Theorem 3.1 and Corollary 3.1, respectively, together with the continuous mapping theorem and the following proposition. We omit the details.

PROPOSITION 5.1. *Suppose that* $R(x)$ *satisfying* **A7***(iii). Then, for* u_t *and* w_t *satisfying the conditions of Theorems 5.1 or 5.2, we have*

$$\frac{1}{a(\widetilde{d}_n)\sqrt{n}} \sum_{t=1}^{n} R(x_t) w_{t+1} = O_P(1), \quad \textbf{(5.9)}$$

where $\widetilde{d}_n = \begin{cases} d_n, & \text{under the conditions of Theorem 5.1,} \\ \sqrt{n}\varphi, & \text{under the conditions of Theorem 5.2.} \end{cases}$

6. CONCLUSION

On weak convergence to stochastic integrals, we show that it is possible to extend the common martingale and semi-martingale structures to include long memory processes, causal processes, and near-epoch dependence in innovations. Our frameworks apply to the TAR, bilinear, and other nonlinear models. As illustrated in Section 5, asymptotics with non-stationary time series in econometrics usually rely on convergence to stochastic integrals. The authors hope the results derived in this paper prove useful in related areas, particularly in nonlinear cointegrating regressions where endogeneity and nonlinearity play major roles.

7. PROOFS

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

Proof of Theorem 2.1. We may write

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})w_{k+1} &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})(v_{k+1}+z_k-z_{k+1})\\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1}+\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}\Big[f\big(x_{nk}\big)-f\big(x_{n,k-1}\big)\Big]z_k+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1}+R_n+o_P(1), \quad \text{say.}\end{aligned} \tag{7.1}$$

Let $\Omega_K = \{x_{ni} : \max_{1\le i\le n}||x_{ni}|| \le K\}$. 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 $R_n = o_P(1)$ because $P(\Omega_K) \to 1$, as $K \to \infty$. Theorem 2.1 follows from Theorem 1.1. ■

Proof of Theorem 2.2. We may write

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})w_{k+1} &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})(v_{k+1}+z_k-z_{k+1})\\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1}+\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}\Big[f\big(x_{nk}\big)-f\big(x_{n,k-1}\big)\Big]z_k+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1}+\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}(x_{nk}-x_{n,k-1})\mathrm{D}f(x_{n,k-1})z_k+R_1(n)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})v_{k+1}+\frac{1}{\sqrt{n}d_n}\sum_{k=1}^{n-1}E(z_k u_k)\mathrm{D}f(x_{n,k-1})+R_1(n)+R_2(n)+o_p(1),\end{aligned} \tag{7.2}$$

where the remainder terms are

$$R_1(n) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} z_k\,[f(x_{nk})-f(x_{n,k-1})-(x_{nk}-x_{n,k-1})\mathrm{D}f(x_{n,k-1})]$$

$$R_2(n) = \frac{1}{\sqrt{n}d_n}\sum_{k=1}^{n-1}[z_k u_k - E(z_k u_k)]\mathrm{D}f(x_{n,k-1}).$$

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

$$R_i(n) = o_P(1), \quad i = 1,\ 2. \tag{7.3}$$

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

To prove (7.3), let $\Omega_K = \{x_{ni} : \max_{1\le i\le n} ||x_{ni}|| \le K\}$. 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,

$$\begin{aligned} E|R_1(n)|I(\Omega_K) &\le \frac{C_K}{\sqrt{n}} \sum_{k=1}^{n} E\left(||x_{nk} - x_{n,k-1}||^{1+\beta'} \, |z_k|\right) \\ &\le C_K \, n^{-(1+\beta'/2)} \sum_{k=1}^{n} E(||u_k||^{1+\beta'} \, |z_k|) = O(n^{-\beta'/2}), \end{aligned} \tag{7.4}$$

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

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

This implies that $R_1(n) = O_P(n^{-\beta'/2})$ because $P(\Omega_K) \to 1$ 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

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

As in the proof of (7.4), it is clear from **A3** that

$$E|R_{21}(n)|I(\Omega_K) \le C_K m n^{-1} \sup_{k\ge 1} E||\lambda_k|| \le C_K n^{-1} \log n,$$

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

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

$$IR_1(n) = \frac{1}{n\sigma} \sum_{k=2m}^{n-1} \left[\lambda_k - E(\lambda_k \mid \mathcal{F}_{k-m-1})\right] x_k^*,$$

$$IR_2(n) = \frac{1}{n\sigma} \sum_{k=2m}^{n-1} E(\lambda_k \mid \mathcal{F}_{k-m-1}) x_k^*,$$

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

where $x_k^* = \mathrm{D}f(x_{n,k-m-1})I(\max_{1\le j\le k-m-1}||x_{nj}||\le K)$. Since **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 $\sup_{k\ge 1}E||\lambda_{1k}(j)||^{1+\delta} \le C\sup_{k\ge 1}E||\lambda_k||^{1+\delta} < \infty$ for some $\delta > 0$. The classical result from the strong law for martingales (see, e.g., Hall and Heyde (1980, Thm. 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 $R_{22}(n) = o_P(1)$ because $P(\Omega_K)\to 1$ as $K\to\infty$, and the fact that, on Ω_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 for (2.2). ■

Proof of Theorem 3.1. First, note that

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

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

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 verify **A2**, i.e., on $D_{\mathbb{R}^2}[0,1]$,

$$\left(\frac{1}{d_n}\sum_{j=1}^{[nt]} u_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} w_j\right) \Rightarrow \left(B_{3/2-\mu}(t), B_{2t}\right). \tag{7.5}$$

We next prove (7.5). Since $\{(\epsilon_k, v_k), \mathcal{F}_k\}_{k\geq 1}$ forms a stationary martingale difference with covariance matrix Ω. Applying the classic martingale limit theorem [see, e.g., Theorem 3.9 in Wang (2015)], yields

$$\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} v_j\right) \Rightarrow (B_{1t}, B_{2t}), \tag{7.6}$$

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

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

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 ϵ_{-i} is independent of (ϵ_i, v_i^*) for $i \geq 1$. If we have the condition:

$$\frac{1}{\sqrt{n}}\max_{1\leq k\leq n}\left|\sum_{j=1}^{k}(v_j - v_j^*)\right| = o_P(1), \tag{7.7}$$

it follows from (7.6) that

$$\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_{-j},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} v_j\right) \Rightarrow \left(B_t, B_{1t}, B_{2t}\right), \tag{7.8}$$

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

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

Result (7.8) implies that

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

on $D_{\mathbb{R}^3}[0,1]$. Thus, (7.5) follows from the continuous mapping theorem and similar arguments to those in Wang et al. (2003).

It remains to show that (3.2) implies (7.7). In fact, by noting $\{v_k - v_k^*, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference, the martingale maximum inequality clearly shows that, for any $\epsilon > 0$,

$$P\left(\max_{1\leq k\leq n}\left|\sum_{j=1}^{k}(v_j - v_j^*)\right| \geq \epsilon\sqrt{n}\right) \leq \frac{2}{n\epsilon^2}\sum_{j=1}^{n} E(v_j - v_j^*)^2$$
$$\leq \frac{2}{n\epsilon^2}\sum_{k=1}^{n} E\left[\sum_{i=0}^{\infty}\mathcal{P}_k(w_{i+k} - w_{i+k}^*)\right]^2. \tag{7.9}$$

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

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

as $k \to \infty$. Taking this estimate into (7.9), we have (7.7), and complete the proof of Theorem 3.1 in addition. ■

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

$$\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} u_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} w_k\right) \Rightarrow (B_{1t}, B_{2t}). \tag{7.10}$$

In fact, by noting that $\left\{(v_{1k}, v_k), \mathcal{F}_k\right\}_{k\geq 1}$ forms a stationary martingale difference with $E\left(v_{10}^2 + v_0^2\right) < \infty$, the classical martingale limit theorem [see, e.g., Theorem 3.9 in Wang (2015)] yields

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

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

$$\Omega_t = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{[nt]} cov\left[\begin{pmatrix}\sigma^{-1}v_{1k}\\ v_k\end{pmatrix}\left(\sigma^{-1}v_{1k}, v_k\right)\right] = \Omega t.$$

Consequently, we have

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

by recalling $E(|z_{10}|^2 + |z_0|^2) < \infty$,

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

This yields (7.10), and also completes the proof of Theorem 3.2. ∎

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

Because (4.1), **A5** (ii) is implied by (3.6). It remains to show that A5 (iii) holds 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

$$\begin{aligned}
E\left(\widetilde{\lambda}_k \mid \mathcal{F}_{k-m}\right) &= \sum_{j=-\infty}^{k-m} \mathcal{P}_j(u_k z_k) = \sum_{i=0}^{\infty} \varphi_i \sum_{j=m}^{\infty} \mathcal{P}_{k-j}(\epsilon_{k-i} z_k) \\
&= \sum_{i=0}^{\infty} \varphi_i \left(\sum_{j=m}^{\max\{m,i\}} + \sum_{j=\max\{m,i\}+1}^{\infty} \right) \mathcal{P}_0(\epsilon_{j-i} z_j) \\
&= \sum_{i=0}^{\infty} \varphi_i \sum_{j=m}^{\max\{m,i\}} \mathcal{P}_0(\epsilon_{j-i} z_j) + \sum_{i=0}^{\infty} \varphi_i \sum_{j=\max\{m,i\}+1}^{\infty} \sum_{t=1}^{\infty} \mathcal{P}_0(\epsilon_{j-i} w_{t+j}) \\
&:= A_{1m} + A_{2m}.
\end{aligned} \tag{7.11}$$

It is readily seen from $E|z_k|^2 = E|z_0|^2 < \infty$ that

$$E|A_{1m}| \le 2 \sum_{i=m}^{\infty} i\, |\varphi_i| (E\epsilon_0^2)^{1/2} (E z_0^2)^{1/2} \to 0,$$

as $m \to \infty$. As for A_{2m}, by noting $\mathcal{P}_0(\epsilon_{j-i} w_{t+j}) = E\left[\epsilon_{j-i}(w_{t+j} - w_{t+j}^{'}) \mid \mathcal{F}_0\right]$ when $j > i$, 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 (7.11), we obtain

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

implying A5 (iii). ∎

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

$$\langle E(w_k \mid \mathcal{F}_{k-m}) \rangle_{2+\delta} \le C \langle w_1 \rangle_4 m^{-\gamma}, \tag{7.12}$$

$$\langle w_k - E(w_k \mid \mathcal{F}_{k+m}) \rangle_{2+\delta} \le C \langle w_1 \rangle_4 m^{-\gamma}, \tag{7.13}$$

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

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

where, as in Section 3.2,

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

We clearly see that both v_k and z_k are stationary and $(v_k, \mathcal{F}_k)_{k \ge 1}$ forms a martingale difference with $Ev_1^2 \le 2Ew_1^2 + 4Ez_1^2 < \infty$, since, by (7.12), the following result holds (implying $Ez_1^2 < \infty$):

$$\langle z_{k,j} \rangle_{2+\delta} \le \sum_{i=j+1}^{\infty} \langle E(w_i | \mathcal{F}_0) \rangle_{2+\delta} \le C \langle w_1 \rangle_4 \sum_{i=j+1}^{\infty} i^{-\gamma} < \infty, \tag{7.14}$$

for any $j \ge 0$, where $z_{k,j} = \sum_{i=j+1}^{\infty} E(w_{i+k} | \mathcal{F}_k)$. By (7.12) and (7.13), for any $k \ge 1$, we also have

$$\begin{aligned} |E(w_1 w_k)| &\le E\big(|w_1 - w_1^*||w_k|\big) + E\big[|w_1^*||E(w_k \mid \mathcal{F}_{k/2})|\big] \\ &\le \langle w_1 \rangle_2 \big\{ \langle w_1 - w_1^* \rangle_2 + \langle E(w_k \mid \mathcal{F}_{k/2}) \rangle_2 \big\} \\ &\le C \langle w_1 \rangle_2 \langle w_1 \rangle_4 k^{-\gamma}, \end{aligned} \tag{7.15}$$

where $w_1^* = E(w_1 \mid \mathcal{F}_{k/2})$. We use the result (7.15) later.

Note that w_k has structure (1.3) with the v_k satisfying **A1**; (3.8) implies **A2**; and **A6** (iii) and (7.14) with $j = 0$ imply **A4** (i). Using Theorem 2.2, Theorem 3.3 will follow if we prove (3.7) and

$$\sup_{k \ge 2m} E||E\big(\lambda_k \mid \mathcal{F}_{k-m}\big)|| \to 0, \tag{7.16}$$

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

By recalling the stationarity of $(u_k, w_k)_{k \ge 1}$, to prove (3.7), it suffices to show that Ω_1, Ω_2 and ρ are finite. In fact, (7.15) implies that $|\Omega_2| \le E w_0^2 + C \sum_{j=1}^{\infty} j^{-\gamma} < \infty$. Similarly, we may prove that $(u_k)_{k \ge 1}$ is a stationary $\mathcal{L}_2$-mixingale of size -1 with constant $\langle u_0 \rangle_4$. Thus, the same argument yields $|\Omega_1| < \infty$ and $|\rho| < \infty$.

To prove (7.16), 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,α_m} is given as in (7.14). Because (7.14), we have

$$E||E(\lambda_{k,2} \mid \mathcal{F}_{k-m})|| \leq E||\lambda_{k,2}|| \leq 2\langle z_{k,\alpha_m}\rangle_2 \langle u_0\rangle_2 \to 0, \tag{7.17}$$

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 in Davidson (1994) implies that A_k is $\mathcal{L}_1$-NED of size -1. Consequently, as in the proof of (7.12), there exists a sequence of υ_m such that $\upsilon_m \to 0$ and

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

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

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

as $m \to \infty$, by taking α_m as an integer sequence in which $\alpha_m \to \infty$ and $\alpha_m \upsilon_m \to 0$. This, together with (7.17), yields

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

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

Proof of Proposition 5.1. Recalling $w_k = \upsilon_k + z_{k-1} - z_k$, we may write

$$\begin{aligned}\sum_{k=1}^{n} R(x_k) w_{k+1} &= \sum_{k=1}^{n} R(x_k)(\upsilon_{k+1} + z_k - z_{k+1}) \\ &= \sum_{k=1}^{n} R(x_k)\upsilon_{k+1} + \sum_{k=1}^{n} \Big[R(x_k) - R(x_{k-1})\Big] z_k + R(x_n) z_{n+1} - R(x_0) z_1 \\ &= \Delta_{1n} + \Delta_{2n} + \Delta_{3n}, \quad \text{say.}\end{aligned} \tag{7.18}$$

For some $K > 0$, let $\tilde{x}_k = x_k I(|x_k|/\widetilde{d}_n \leq K)$,

$$\widetilde{\Delta}_{1n} = \sum_{k=1}^{n} R(\tilde{x}_k)\upsilon_{k+1} \quad \text{and} \quad \widetilde{\Delta}_{2n} = \Delta_{2n} I(\max_{1\leq i\leq n} |x_i|/\widetilde{d}_n \leq K).$$

https://doi.org/10.1017/S0266466617000408 Published online by Cambridge University Press

Under **A7** (iii), it is readily seen from $\sup_k E|z_k u_k| < \infty$ that, for any $K > 0$,

$$E|\widetilde{\Delta}_{2n}| \leq \frac{a(\widetilde{d}_n)}{\widetilde{d}_n} \sum_{k=1}^{n} E||z_k u_k|| \leq C \frac{na(\widetilde{d}_n)}{\widetilde{d}_n},$$

implying $\Delta_{2n} = O_P\big[\sqrt{n}\, a(\widetilde{d}_n)\big]$ because, again as $K \to \infty$,

$$P(\Delta_{2n} \neq \widetilde{\Delta}_{2n}) \leq P(\max_{1\leq i\leq n} |x_i|/\widetilde{d}_n > K) \to 0.$$

Similarly, $\Delta_{2n} = O_P\big[\sqrt{n}\, a(\widetilde{d}_n)\big]$. As for Δ_{1n}, by noting that Δ_{1n} forms a martingale, it follows from **A7** (iii) again that

$$E\widetilde{\Delta}_{1n}^2 \leq \sum_{k=1}^{n} E R^2(\tilde{x}_k) \leq na^2(\widetilde{d}_n)(1 + C_K^{\gamma}),$$

indicating that $\Delta_{1n} = O_P\big[\sqrt{n}\, a(\widetilde{d}_n)\big]$ because, as $K \to \infty$,

$$P(\Delta_{1n} \neq \widetilde{\Delta}_{1n}) \leq P(\max_{1\leq i\leq n} |x_i|/\widetilde{d}_n > K) \to 0.$$

Combining all these facts, we prove $\sum_{k=1}^{n} R(x_k) \omega_{k+1} = O_P\big[\sqrt{n}\, a(\widetilde{d}_n)\big]$, as required. ■

NOTE

1. For a definition of α-mixing, we refer to Davidson (1994, Chap. 14).

REFERENCES

Borovskikh, Y.V. & V.S. Korolyuk (1997) *Martingale Approximation*. VSP

Chang, Y., J.Y. Park, & P.C.B. Phillips (2001) Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal* 4, 1–36.

Chang, Y. & J.Y. Park (2011) Endogeneity in nonlinear regressions with integrated time series. *Econometric Reviews* 30, 51–87.

Chan, N. & Q. Wang (2015) Nonlinear regression with nonstationary time series. *Journal of Econometrics* 185, 182–195.

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.

De Joon, R. (2004) Nonlinear estimators with integrated regressors but without exogeneity. Econometric Society North American Winter Meetings, no 324, Econometric Society 2004.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and its Application*. Probability and Mathematics Statistics. Academic Press, Inc.

Ibragimov, R. & P.C.B. Phillips (2008) Regression asymptotics using martingale convergence methods. *Econometric Theory* 24, 888–947.

Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes*, 2nd ed. Springer-Verlag.

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

Liang, H.Y., P.C.B. Phillips, H. Wang, & Q. Wang (2016) Weak convergence to stochastic integrals for econometric applications. *Econometric Theory* 32, 1349–1375.

Lin, Z. & H. Wang (2016) On convergence to stochastic integrals. *Journal of Theoretical Probability* 29(3), 717–736.

Park, J.Y. & P.C.B. Phillips (2000) Nonstationary binary choice. *Econometrica* 68, 1249–1280.

Park, J.Y. & P.C.B. Phillips (2001) Nonstationary regressions with integrated time series. *Econometrica* 69, 117–161.

Phillips, P.C.B. (1988a) Weak convergence to sample covariance matrices to stochastic integrals via martingale approximation. *Econometric Theory* 4, 528–533.

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

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

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

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

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

Wang, Q. (2015) *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.

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

Wu, W. & M. Rosenblatt (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 101, 14150–14154.

Wu, W. & W. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115(6), 939–958.

Wu, W. (2007) Strong invariance principles for dependent random variables. *Annals of Probability* 35(6), 2294–2320.