WEAK CONVERGENCE TO STOCHASTIC INTEGRALS FOR ECONOMETRIC APPLICATIONS

By

Hanying Liang, Peter C. B. Phillips, Hanchao Wang, and Qiying Wang

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1971

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Weak Convergence to Stochastic Integrals for Econometric Applications

H. Liang, P. C. B. Phillips, H. Wang and Q. Wang

Tongji University, Yale University, Zhejiang University and The University of Sydney

October 7, 2014

Abstract

Limit theory involving stochastic integrals is now widespread in time series econometrics and relies on a few key results on function space weak convergence. In establishing weak convergence of sample covariances to stochastic integrals, the literature commonly uses martingale and semimartingale structures. While these structures have wide relevance, many applications in econometrics involve a cointegration framework where endogeneity and nonlinearity play a major role and lead to complications in the limit theory. This paper explores weak convergence limit theory to stochastic integral functionals in such settings. We use a novel decomposition of sample covariances of functions of $I(1)$ and $I(0)$ time series that simplifies the asymptotic development and we provide limit results for such covariances when linear process, long memory, and mixing variates are involved in the innovations. The limit results extend earlier findings in the literature, are relevant in many econometric applications, and involve simple conditions that facilitate implementation in practice. A nonlinear extension of FM regression is used to illustrate practical application of the methods.

Key words and phrases: Decomposition, FM regression, Linear process, Long memory, Stochastic integral, Semimartingale, $\alpha-$mixing.

JEL Classifications: C22, C65

1 Introduction

A dominant feature of nonstationary time series is that limit theory formulae typically reflect the effects of a full trajectory of observed data, rather than just a few moment characteristics as happens in the stationary case. The primary mechanisms producing this trajectory dependence are the functional central limit theory that operates on the partial sum components and the weak convergence results that provide limit theory for sample covariance and score components to a stochastic integral form rather than a normal or mixed normal form as commonly applies in simpler settings.

In developing a general theory it is convenient to use an array structure in which random arrays $\{x_{nk}, y_{nk}, 1 \leq k \leq n, n \geq 1\}$ are constructed from some underlying nonstationary time series by suitable standardization to ensure a non-trivial limit. In particular, we suppose that there exists a vector limit process $\{W(t), G(t), 0 \leq t \leq 1\}$ to which $\{x_{n,\lfloor nt \rfloor}, y_{n,\lfloor nt \rfloor}\}$ converges weakly in the Skorohod space $D_{R^2}[0,1]$, where the floor function $\lfloor a \rfloor$ denotes the integer part of a. A common functional of interest S_n of $\{x_{nk}, y_{nk}\}$ is defined by the sample quantity

$$S_n = \int_0^1 f(y_{n,\lfloor nt \rfloor}) dx_{n,\lfloor nt \rfloor} = \sum_{k=0}^{n-1} f(y_{nk})\, \epsilon_{n,k+1}, \tag{1.1}$$

where $\epsilon_{nk} = x_{n,k} - x_{n,k-1}$ and f is a real function on $\mathbb{R}$. The quantity S_n is a sample covariance between the elements $f(y_{nk})$ and $\epsilon_{n,k+1}$. As indicated, such functionals arise frequently in the study of nonstationary time series, unit root testing and nonlinear cointegration regressions. They also arise in mathematical finance and the study of stochastic differential equations. In the nonstationary time series context, the array components y_{nk} may be standardized forms of certain nonstationary regressors, the ϵ_{nk} standardized error processes, and $f(\cdot)$ a nonlinear regression function or its derivatives. The sample covariance S_n may then represent a score function or moment function arising from instrumental variable or moment method estimation. Many examples of such functionals have appeared in the literature since the work of Park and Phillips (1999, 2000, 2001) on nonlinear regression with integrated processes.

The asymptotics of functionals like S_n are therefore of considerable interest and a substantial literature has arisen. In certain cases it is well-known that S_n converges weakly to a simple Itō stochastic integral so that $S_n \to_D \int_0^1 f[G(t)]dW(t)$ where $W(t)$ is Brownian motion and the process $\int_0^r f[G(t)]dW(t)$ is a continuous martingale. Results of this form began to emerge in the 1980s in statistics, probability, and econometrics. Chan and Wei (1988), Phillips (1987, 1988a), and Strasser (1986), for example, gave results for martingale arrays, and Kurtz and Protter (1991), Duffie and Protter (1992) and Jakubowski (1996) provided some general results when $\{x_{nk}\}$ is a semimartingale and the limit process $W(t)$ is a semimartingale.

In many econometric applications such as a cointegration framework, endogeneity is expected and it is therefore realistic to assume that the regressors y_{nk} are correlated with the innovations ϵ_{nk} at some leads and/or lags. This correlation can complicate the limit theory and the econometric literature provided several results involving the convergence properties of S_n in such cases. When $f(x) = x$, Phillips (1988b) considered linear processes with iid innovations; Phillips (1987), Hansen (1992) and De Jong and Davidson (2000a, b) allowed for mixing sequences; and more recently Ibragimov and Phillips (2008) also allowed for summands involving a smooth function $f(x)$ in (1.1). De Jong (2002), Chang and Park (2011) and Lin and Wang (2010) provided some related results.

The present paper has a similar goal to this econometric work but offers results that are convenient to implement and have wider applicability. Our main theorems allow for the ϵ_{nk} in (1.1) to be replaced by a linear process array $u_{nk} = \sum_{j=0}^{\infty} \varphi_j \, \epsilon_{n,k-j}$, for $\Delta y_{nk} := y_{n,k} - y_{n,k-1}$ to comprise long memory innovations, and for $(\Delta y_{nk}, \epsilon_{n,k+1})$ to be an α-mixing random sequence. Since u_{nk} includes all stationary and invertible ARMA process and is serially dependent and cross correlated with y_{nk}, our results apply in much empirical work. Further, the method of derivation is simple and straightforward, so the technical development and results are also of pedagogical value for students of nonstationary time series limit theory. The core of the development is a novel decomposition result for partial sums of the form $\sum_{k=0}^{n-1} f(y_{nk}) \, u_{n,k+1}$ that is of some independent interest, extending to the nonlinear functional case the linear decomposition used in earlier work (Phillips, 1988b).

This paper is organized as follows. Our main results are given in the next section, which provides some general discussion and remarks clarifying the difference between the current paper and earlier work. The extension to α-mixing random sequences is considered in Section 3. Some examples, remarks on applications, and an illustration of nonlinear fully modified (FM) regression are given in Section 4. Section 5 concludes and proofs are provided in Section 6. Throughout the paper, we denote constants by $C, C_1, ...$ which may differ at each appearance. $D_{\mathbb{R}^d}[0,1]$ denotes the space of càdlàg functions from $[0,1]$ to $\mathbb{R}^d$. We mention that the convergence of càdlàg functions such as $(x_n(t), y_n(t))$ can be considered either on $D_{\mathbb{R}}[0,1] \times D_{\mathbb{R}}[0,1]$ or $D_{\mathbb{R}^2}[0,1]$ in the Skorohod topology. The latter convergence is stronger as we require only one sequence $0 \le \lambda_n(t) \le 1$ of time changes in the Skorohod metric such that $(x_n[\lambda_n(t)], y_n[\lambda_n(t)])$ converges uniformly to $(x(t), y(t))$ on $t \in [0,1]$. 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

Let $\{\mathcal{F}_k^n\}$ be a array filtration so that, for each n, $\{x_{nk}, y_{nk}\}$ is an $\{\mathcal{F}_k^n\}$-adapted process and $\{x_{nk}\}$ is an $\{\mathcal{F}_k^n\}$-semimartingale with decomposition:

$$x_{nk} = M_{n,k} + A_{n,k},$$

where $M_{n,k}$ is a martingale and $A_{n,k}$ is a finite variation process. In commonly occurring applications, the arrays $\{x_{nk}, y_{nk}\}$ arise as standardized versions of partial sums of sequences of innovations, as in (2.4) below. The following assumptions concerning these components are used throughout this section.

A1. $\{x_{n,\lfloor nt\rfloor}, y_{n,\lfloor nt\rfloor}\} \Rightarrow \{W(t), G(t)\}$ on $D_{\mathbb{R}^2}[0,1]$ in the Skorohod topology.

A2. $\sup_n \left(\mathbb{E}M_{n,n}^2 + \sum_{k=1}^{n-1} \mathbb{E}|A_{n,k+1} - A_{n,k}|\right) < \infty$.

Assumption **A1** is assured by standard functional limit theory holding under well-known primitive conditions. The condition implies the array $\{x_{n,\lfloor nt\rfloor}, y_{n,\lfloor nt\rfloor}\}$ is suitably standardized to ensure the time series trajectories have stochastic process limits in $D_{\mathbb{R}^2}[0,1]$. Assumption **A2** places a uniform moment condition on the martingale $M_{n,n}$ and the increments of the finite variation process $A_{n,k}$.

THEOREM 2.1. *Suppose* ***A1*** *and* ***A2*** *hold. Then $W(t)$ is a semimartingale with respect to a filtration to which $W(t)$ and $G(t)$ are adapted, and for any continuous functions $g_1(s)$ and $g_2(s)$,*

$$\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g_1(y_{nk}),\ \sum_{k=0}^{n-1} g_2(y_{nk})\,\epsilon_{n,k+1}\Big\}$$
$$\Rightarrow \Big\{W(t),\ G(t),\ \int_0^1 g_1[G(t)]dt,\ \int_0^1 g_2[G(t)]\,dW(t)\Big\}, \tag{2.1}$$

on $D_{\mathbb{R}^4}[0,1]$ in the Skorohod topology.

Theorem 2.1 is known in the existing literature (e.g., Kurtz and Protter, 1991) but is not sufficiently general to cover many econometric applications where endogeneity and more general innovation processes are present. Our goal is to extend the framework to accommodate these applications and to do so under conditions that facilitate implementation. The analysis follows earlier econometric work on weak convergence to stochastic integrals by using linear process innovations. Explicitly, we investigate the convergence of sample quantities to functionals of stochastic processes and stochastic integrals similar to (2.1) in which the ϵ_{nk} are replaced by

$$u_{nk} = \sum_{j=0}^{\infty} \varphi_j \, \epsilon_{n,k-j},$$

where $\varphi = \sum_{j=0}^{\infty} \varphi_j \neq 0$ and $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$. The array u_{nk} includes all stationary and invertible ARMA time series arrays and may be serially dependent and cross correlated with y_{nk}.

Our first result is as follows.

THEOREM 2.2. *In addition to **A1** and **A2**, suppose that* $\sum_{k=1}^{n} \mathbb{E}\epsilon_{nk}^2 = O(1)$, $\sup_{k\in\mathbb{Z}} \mathbb{E}\epsilon_{nk}^2 \to 0$ *and*

$$\sup_{i,j\geq 1} \frac{1}{j^2} \mathbb{E}|y_{n,i+j} - y_{n,i}|^2 = o(n^{-1}). \tag{2.2}$$

Then, for any function $f(s)$ satisfying a local Lipschitz condition[1] *and for any continuous function $g(s)$, we have*

$$\Big\{x_{n,\lfloor nt \rfloor},\ y_{n,\lfloor nt \rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(y_{nk}),\ \sum_{k=0}^{n-1} f(y_{nk})\, u_{n,k+1}\Big\}$$
$$\Rightarrow\ \Big\{W(t),\ G(t),\ \int_0^1 g[G(t)]dt,\ \varphi \int_0^1 f[G(t)]\, dW(t)\Big\}, \tag{2.3}$$

on $D_{\mathbb{R}^4}[0,1]$ in the Skorohod topology.

The local Lipschitz condition on $f(x)$ is a minor requirement and holds for many continuous functions. The condition was used in the limit theory of Ibragimov and Phillips (2008, Remark 3.2). Recall that the components $\epsilon_{nk} = x_{n,k} - x_{n,k-1}$ are standardized differences and $x_{n,\lfloor nt \rfloor} \Rightarrow W(t)$ on $D[0,1]$. It is natural therefore to assume that $\sum_{k=1}^{n} \mathbb{E}\epsilon_{nk}^2 = O(1)$ and $\sup_{k\in\mathcal{Z}} \mathbb{E}\epsilon_{nk}^2 \to 0$. The additional condition (2.2) holds for standardized sums of a long memory process such as $y_{nk} = \sum_{j=1}^{k} \xi_j/d_n, 1 \leq k \leq n$, where $d_n^2 = var(\sum_{j=1}^{n} \xi_j) \sim C\,n^{\alpha}$ with $1 < \alpha \leq 2$. An example is given in Section 4. It is therefore particularly convenient in that context. Note that in this case the standardization is $d_n = O\left(n^{\alpha/2}\right)$, which exceeds the usual $\sqrt{n}$ standardization for $I\,(1)$ processes. Then, $\sup_{i,j\geq 1} \frac{1}{j^2}\mathbb{E}|y_{n,i+j} - y_{n,i}|^2 = \sup_{j\geq 1} \frac{Cj^{\alpha}}{j^2 n^{\alpha}} = o\left(n^{-1}\right)$, as in (2.2). Interestingly, however, (2.2) excludes partial sums of a short memory process and the

[1] The function $f(s)$ is said to satisfy local Lipschitz condition if, for every $K > 0$, there exists a constant C_K such that

$$|f(x) - f(y)| \leq C_K |x - y|,$$

for all $x, y \in \mathbb{R}$ with $\max\{|x|, |y|\} \leq K$.

condition does not hold even for partial sums of $iid\,(0,\sigma^2)$ innovations for which it is easily seen that $\sup_{i,j\geq 1}\frac{1}{j^2}\mathbb{E}|y_{n,i+j}-y_{n,i}|^2=\sup_{j\geq 1}\frac{Cj}{j^2n}=O\left(n^{-1}\right)$. Our next theorem removes this restriction but imposes greater smoothness on $f(x)$, thereby showing that the time series structure of u_{nk} and its interaction with the properties of the nonlinear function f can have a significant effect on limit behavior.

To facilitate the analysis and for notational convenience, we next assume that both x_{nk} and y_{nk} are simple normalized partial sum processes of the following integrated process form

$$x_{nk}=\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\epsilon_j,\quad y_{nk}=\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\eta_j,\quad k=1,2,...,n. \tag{2.4}$$

Let $\mathcal{F}_k^n=\mathcal{F}_k$ for all $n\geq 1$ where $\mathcal{F}_k=\sigma(\epsilon_j,\eta_j,j\leq k)$ and $u_k=\sum_{j=0}^{\infty}\varphi_j\,\epsilon_{k-j}$ where $\varphi=\sum_{j=0}^{\infty}\varphi_j\neq 0$ and $\sum_{j=0}^{\infty}j\,|\varphi_j|<\infty$. We add the following conditions.

A3. *$f'(x)$ is locally bounded and*

$$|f'(x)-f'(y)|\leq C_K\,|x-y|^{\beta},\quad \text{for some } 0<\beta\leq 1/3$$

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

A4. $\sup_{j\geq 1,i\in\mathcal{Z}}\frac{1}{j}\sum_{k=1}^{j}E|\epsilon_{k+i}|^3<\infty$, $\sup_{i,j\geq 1}\frac{1}{j}\sum_{k=1}^{j}E\left(\eta_{k+i}^2\epsilon_{i+1}^2\right)<\infty$ *and*

$$\sup_{i,j\geq 1}\frac{1}{j}\mathbb{E}|\sum_{k=1}^{j}\eta_{k+i}|^2<\infty.$$

A5. *There exists a constant $A_0>0$ such that*

$$\sup_{i\geq 0}\Big|\sum_{j=0}^{n-1}\varphi_j\sum_{k=1}^{j}\mathbb{E}\big(\eta_{k+i}\epsilon_{i+1}\mid\mathcal{F}_i\big)-A_0\Big|=o_P(1).$$

Condition **A3** is trivially satisfied when the second derivative of $f(x)$ exists on $\mathbb{R}$. Assumptions **A4** and **A5** typically hold for short memory processes satisfying certain moment and stationarity conditions. For instance, if $(\{\epsilon_k,\eta_k\},\mathcal{F}_k)$ forms a martingale difference sequence with

$$\mathbb{E}(\epsilon_k\,\eta_k\mid\mathcal{F}_k)=\tau,\ a.s.\quad \text{for all } k\geq 1,$$

and $\sup_k\left(\mathbb{E}|\epsilon_k|^4+\mathbb{E}|\eta_k|^4\right)<\infty$, then **A4** and **A5** hold with $A_0=\tau\,\varphi$. Other standard cases that arise in econometric work are given in Section 4.

Our second result covers time series satisfying the above conditions for which we again have weak convergence to limit functionals that involve a stochastic integral with a stochastic correction that embodies the effects of endogeneity.

THEOREM 2.3. *Under **A1** – **A5** and for any continuous function $g(\cdot)$, we have*

$$\begin{aligned} &\Big\{x_{n,\lfloor nt\rfloor},\ y_{n,\lfloor nt\rfloor},\ \frac{1}{n}\sum_{k=1}^{n} g(y_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(y_{nk})\,u_{k+1}\Big\} \\ \Rightarrow\ &\Big\{W(t),\ G(t),\ \int_0^1 g[G(s)]ds,\ \varphi\int_0^1 f[G(s)]\,dW(s)+A_0\int_0^1 f'[G(s)]ds\Big\}, \end{aligned} \tag{2.5}$$

on $D_{\mathbb{R}^4}[0,1]$ in the Skorohod topology.

Remark 1. Corresponding to (2.5) we have weak convergence of the partial sum covariance process

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{\lfloor nt\rfloor} f(y_{nk})\,u_{k+1} \Rightarrow \varphi\int_0^t f[G(s)]\,dW(s)+A_0\int_0^t f'[G(s)]ds, \tag{2.6}$$

where the limit involves the scaled stochastic integral $\varphi\int_0^t f[G(s)]\,dW(s)$ and stochastic drift function $D(t) = A_0\int_0^t f'[G(s)]ds$. The stochastic integrals in (2.5) and (2.6) are scaled by the long run moving average coefficient $\varphi = \sum_{j=0}^{\infty}\varphi_j$, as expected from the (Beveridge Nelson) decomposition of $u_k = \sum_{j=0}^{\infty}\varphi_j\,\epsilon_{k-j} = \varphi\,\epsilon_k + \widetilde{\epsilon}_{k-1} - \widetilde{\epsilon}_k$, where $\widetilde{\epsilon}_k = \sum_{j=0}^{\infty}\widetilde{\varphi}_j\,\epsilon_{k-j}$ with $\widetilde{\varphi}_j = \sum_{m=j+1}^{\infty}\varphi_m$ as in Phillips and Solo (1992). To explain the last term of (2.6), define $H(t) = f'[G(t)]/f[G(t)]$ and assume that $\int_0^1 H(s)^2\,ds < \infty$, *a.s.* Then, $F(t) = \frac{A_0}{\varphi}\int_0^t H(s)\,ds$ has finite variation and $F'(t) = \frac{A_0}{\varphi}H(t)$. Defining the semimartingale $V(t) = W(t) + F(t)$, we observe that

$$\begin{aligned} \varphi\int_0^t f[G(s)]\,dV(s) &= \varphi\int_0^t f[G(s)]\,dW(s) + \varphi\int_0^t f[G(s)]\,dF(s) \\ &= \varphi\int_0^t f[G(s)]\,dW(s) + A_0\int_0^t f'[G(s)]ds, \end{aligned} \tag{2.7}$$

which gives the limit process (2.6) a stochastic integral representation that involves the same integrand $f[G(s)]$ but where the integral in (2.7) is taken with respect to the semimartingale $V(s)$. The stochastic drift $D(t) = A_0\int_0^t f'[G(s)]ds$ is therefore induced by the finite variation process of the semimartingale $V(s)$.

Remark 2. Theorem 2.2 is new and Theorem 2.3 extends Theorem 3.1 of Ibragimov and Phillips (IP) (2008), where $\eta_k = u_k$ is imposed in their theorem and ϵ_k is assumed to be a sequence of iid random variables. Theorem 4.3 of IP (2008) eliminated the restriction $\eta_k = u_k$ by allowing (η_k, u_k) to be a joint linear process, but a detailed proof in that case was not provided.

The approach adopted in IP (2008) is to use general methods of weak convergence of discrete time semimartingales to continuous time semimartingales to establish limit theory for sample covariances such as $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(y_{nk})\, u_{k+1}$. The idea is conceptually elegant, offers considerable generality, unifies convergence results for stationary and unity root cases, and uses the semimartingale convergence methods and conditions developed in Jacod and Shiryaev (1987/2003) to establish the limit theory. According to this approach, discrete time sample covariances are embedded in semimartingales and asymptotics are delivered via semimartingale convergence. The conditions involved in justifying the limit theory by this method involve the asymptotic behavior of the triplet of predictable characteristics of the semimartingale process, combined with conditions that identify the limit process as a stochastic integral. These conditions can be difficult to verify and the proofs are often lengthy and involve some complex derivations, as is evident in IP (2008). The derivation of (2.5) given here has the advantage of a direct self-contained approach that proceeds under more readily verified conditions.

Remark 3. One feature of the proof of Theorem 3.1 in IP (2008) raises an interesting technical difficulty that has wider implications in time series econometrics and financial econometrics. The issue relates to limit theory involving weak convergence to normal mixtures, such as those that occur in asymptotics for cointegrating estimators (Phillips, 1989, 1991; Phillips and Ouliaris, 1990; Jeganathan, 1993) and in the limit theory for empirical quadratic variation (realized variance) processes in financial econometrics (e.g., Mykland and Zhang, 2006). In such cases, stable (Réyni) convergence can be used to facilitate random normalization that leads to feasible test statistics with pivotal limit distributions. In the present context, the techniques used in IP require verification of the convergence of a composite functional that arises in characterizing the limit behavior of the sample covariance as a semimartingale (Lemma E2 of IP, 2008). To fix ideas, suppose that $X_n(t)$ and $Y_n(t) \geq 0, t \geq 0$, are two continuous processes, having limit processes $X(t)$ and $Y(t)$, respectively. IP need to verify the weak convergence of the composite functional

$$X_n[Y_n(t)] \Rightarrow X[Y(t)], \quad t \geq 0, \tag{2.8}$$

see IP (2008, Lemma E2, p. 942[2]). IP argue that if $X_n(t) \Rightarrow X(t)$ and $Y_n(t) \to_p Y(t) \geq 0$, then (2.8) follows by the same method as that used in Billingsley (1968, eq'n (17.9), p.

[2]There is a typographical error in the statement of Lemma E.2: "$X(s) \geq 0$" should read "$Y(s) \geq 0$".

145), a method that requires the joint weak convergence

$$(X_n(t), Y_n(t)) \Rightarrow (X(t), Y(t)) \tag{2.9}$$

to hold. IP justify (2.9) by using theorem 4.4 of Billingsley (1968, p. 27). However, Billingsley's theorem 4.4 assumes that $Y_n(t) \to_p Y$ with $Y = a$, a constant, and constancy of the limit plays a role in that proof. When $Y_n(t) \to_p Y$ with Y a random variable, then the result (2.9) may no longer hold whereas the composite function limit (2.8) may still apply. Example 1 below illustrates this phenomenon. On the other hand, if the stronger condition $X_n(t) \Rightarrow_{\text{stably}} X(t)$, requiring stable weak convergence (Réyni, 1963; Aldous and Eagleson, 1978; Hall and Heyde, 1980), in conjunction with $Y_n(t) \to_p Y(t)$ holds, then the joint convergence (2.9) is valid and (2.8) follows by the same argument as in Billingsley (1968, p. 145). The difference is that $X_n(t) \Rightarrow_{\text{stably}} X(t)$ ensures joint weak convergence $((X_n(t), Y(t)) \Rightarrow (X(t), Y(t)))$ for all $Y(t)$ adapted to the same probability space, thereby enabling (2.9).[3]

Example 1. Let $Y_n(t) = Y(t) = \xi \mathbf{1}_{\{\xi \geq 0\}}$ for all t and for all n, where $\xi \equiv N(0,1)$. Further, define $X_n(t) = -\xi$ for all t and for all n. Then, $Y_n(t) \to_p Y(t) = \xi \mathbf{1}_{\{\xi \geq 0\}} \geq 0$, and $X_n(t) \Rightarrow X(t) = \xi \equiv N(0,1)$ because of the symmetry of the random variable ξ. However, the joint weak convergence (2.9) fails. In particular,

$$(X_n(t), Y_n(t)) \equiv \left(-\xi, \xi \mathbf{1}_{\{\xi \geq 0\}}\right) \neq_D \left(\xi, \xi \mathbf{1}_{\{\xi \geq 0\}}\right) \equiv (X(t), Y(t))$$

since $-\xi + \xi \mathbf{1}_{\{\xi \geq 0\}} \neq_D \xi + \xi \mathbf{1}_{\{\xi \geq 0\}}$. For instance, the additive functional $X_n(t) + Y_n(t) := f(X_n(t), Y_n(t)) \not\Rightarrow f(X(t), Y(t))$ because

$$\begin{aligned} P(X_n(t) + Y_n(t) \leq x) &= P\left(-\xi + \xi \mathbf{1}_{\{\xi \geq 0\}} \leq x\right) = P\left(-\xi \mathbf{1}_{\{\xi < 0\}} \leq x\right) \\ &\neq P\left(\xi + \xi \mathbf{1}_{\{\xi \geq 0\}} \leq x\right) = P(X(t) + Y(t) \leq x). \end{aligned}$$

On the other hand, $X_n[Y_n(t)] = -\xi$ for all t and for all n, while $X[Y(t)] = N(0,1)$ for all t, so that the composite functional $X_n[Y_n(t)] \Rightarrow X[Y(t)]$ and (2.8) holds. It follows that (2.9) is not a necessary condition for (2.8).

Remark 4. The core component in the proofs of Theorems 2.2 and 2.3 is a decomposition result involving the sample covariance function $\sum_{k=1}^{n} f(\widetilde{y}_{nk})\, \widetilde{u}_{n,k+1}$. This decomposition can be used together with Theorem 2.1 to provide an extension of the limit theory

[3] A standard example that illustrates the difference between $X_n(t) \Rightarrow_{\text{stably}} X(t)$ and $X_n(t) \Rightarrow X(t)$ is as follows (e.g. see Cheng and Chow, 2002). Let $X_{2k} = X$ and $X_{2k+1} = X'$ where X and X' are independent and have identical distributions. Then $P(X_{2k} \leq a, X \leq b) \to P(X \leq a \wedge b)$ and $P(X_{2k+1} \leq a, X \leq b) \to P(X \leq a)\, P(X \leq b)$, so that $X_n \Rightarrow X$ but $X_n \not\Rightarrow_{\text{stably}} X$.

to more general classes of processes. The idea extends the decomposition used in Phillips (1988b) to establish convergence to a stochastic integral with drift by writing the sample covariance in terms of a martingale component and a correction term. In the present case, the nonlinear component in $\sum_{k=1}^{n} f(\widetilde{y}_{nk})\,\widetilde{u}_{n,k+1}$ requires additional treatment in delivering the decomposition. We present the following result involving two sequences of random arrays $\widetilde{y}_{nk}$ and $\widetilde{\epsilon}_{nk}$ and the linear process $\widetilde{u}_{nk} = \sum_{j=0}^{\infty} \varphi_j \widetilde{\epsilon}_{n,k-j}$ with coefficients φ_j satisfying $\varphi = \sum_{j=0}^{\infty} \varphi_j \neq 0$ and $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$.

PROPOSITION 2.1. *Suppose that* $\max_{1\le k\le n} |\widetilde{y}_{nk}| = O_P(1)$,

$$\sup_{j\ge 1, i\in Z} \frac{1}{j} \sum_{k=1}^{j} \mathbb{E}\widetilde{\epsilon}_{n,k+i}^{2} \to 0, \qquad \text{as } n \to \infty. \tag{2.10}$$

Then, for any locally bounded function $f(x)$, we have

$$\sum_{i=1}^{m} f(\widetilde{y}_{n,i-1})\widetilde{u}_{n,i} = \varphi \sum_{i=0}^{m} f(\widetilde{y}_{ni})\widetilde{\epsilon}_{n,i+1} + \sum_{j=0}^{m-1} \varphi_j \sum_{i=0}^{m} [f(\widetilde{y}_{n,i+j}) - f(\widetilde{y}_{n,i})]\widetilde{\epsilon}_{n,i+1} + R(m), \tag{2.11}$$

where $R(m) = o_P(1)$ for each $1 \le m \le n$. If in addition $\max_{1\le i<k\le n} \frac{1}{k-i} \sum_{j=i}^{k} |\widetilde{\epsilon}_{nj}| = o_P(1)$, *then* $\max_{1\le m\le n} |R(m)| = o_P(1)$.

Remark 5. If $f(x)$ is a bounded function on $\mathbb{R}$, the condition $\max_{1\le k\le n} |\widetilde{y}_{nk}| = O_P(1)$ is not necessary. In other words, Proposition 2.1 holds without any restriction on the random sequence $\widetilde{y}_{nk}$.

Remark 6. As in Phillips (1988b), instead of (2.11), $\sum_{i=1}^{m} f(\widetilde{y}_{n,i-1})\widetilde{u}_{n,i}$ can be decomposed as

$$\sum_{i=1}^{m} f(\widetilde{y}_{n,i-1})\widetilde{u}_{n,i} = \varphi \sum_{i=0}^{m} f(\widetilde{y}_{ni-1})\widetilde{\epsilon}_{n,i} + \sum_{i=1}^{m} \left(f(\widetilde{y}_{n,i}) - f(\widetilde{y}_{n,i-1})\right) \widetilde{\epsilon}_{n,i}^{*} + r_m, \tag{2.12}$$

where $r_m = f(\widetilde{y}_{n,m})\widetilde{\epsilon}_{n,m}^{*} - f(\widetilde{y}_{n,0})\widetilde{\epsilon}_{n,0}^{*}$, $\widetilde{\epsilon}_{n,i}^{*} = \sum_{j=0}^{\infty} \varphi_j^{*} \widetilde{\epsilon}_{n,i-j}$, and $\varphi_j^{*} = \sum_{s=j+1}^{\infty} \varphi_s$. The decomposition (2.12), which is proved in the Appendix, is particularly useful in the linear case, i.e. when $f(x) = x$. To illustrate, let $f(\widetilde{y}_{ni}) = \widetilde{y}_{ni} = \sum_{k=1}^{i} \eta_k/\sqrt{n}$, and $\widetilde{\epsilon}_{n,i} = \varepsilon_i/\sqrt{n}$. Then, for $m = n$ we have

$$\sum_{k=1}^{n} \left(f(\widetilde{y}_{n,k}) - f(\widetilde{y}_{n,k-1})\right) \epsilon_{n,k}^{*} = \frac{1}{n} \sum_{k=1}^{n} \eta_k \left\{ \sum_{s=0}^{\infty} \varphi_s^{*} \varepsilon_{k-s} \right\} = \frac{1}{n} \sum_{k=1}^{n} \eta_k \varepsilon_k^{*} \to_{a.s.} \mathbb{E}\eta_0 \varepsilon_0^{*},$$

if the components (η_k, ε_k) are stationary and ergodic. We may simplify this result further if $\mathbb{E}\{\eta_0\varepsilon_{-\ell+i}\} = 0$ for all $\ell < i$, as happens for instance when ε_k is a martingale difference sequence. Indeed, in this situation,

$$\mathbb{E}\left\{\eta_0 \sum_{\ell=0}^{\infty} \varphi_\ell \varepsilon_{-\ell+i}\right\} = \mathbb{E}\left\{\eta_0 \sum_{\ell=i}^{\infty} \varphi_\ell \varepsilon_{i-\ell}\right\} = \mathbb{E}\left\{\eta_0 \sum_{s=0}^{\infty} \varphi_{s+i} \varepsilon_{-s}\right\},$$

and it follows that

$$\begin{aligned}
\mathbb{E}\eta_0\varepsilon_0^* &= \sum_{s=0}^{\infty} \varphi_s^* \mathbb{E}(\eta_0\varepsilon_{-s}) = \sum_{s=0}^{\infty} \mathbb{E}\left\{\eta_0 \left(\sum_{i=1}^{\infty} \varphi_{s+i}\right) \varepsilon_{-s}\right\} \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left\{\eta_0 \sum_{s=0}^{\infty} \varphi_{s+i}\varepsilon_{-s}\right\} = \sum_{i=1}^{\infty} \mathbb{E}\left\{\eta_0 \sum_{\ell=i}^{\infty} \varphi_\ell \varepsilon_{-\ell+i}\right\} \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left\{\eta_0 \sum_{\ell=0}^{\infty} \varphi_\ell \varepsilon_{i-\ell}\right\} = \sum_{i=1}^{\infty} \mathbb{E}\{\eta_0 u_i\} = \lambda_{\eta u},
\end{aligned}$$

where $u_i = \sum_{\ell=0}^{\infty} \varphi_\ell \varepsilon_{i-\ell}$ and $\lambda_{\eta u} = \sum_{i=1}^{\infty} \mathbb{E}\{\eta_0 u_i\}$ is the one-sided long run covariance between the time series (η_k, u_k), as in the correction terms given in Phillips (1987, 1988a, 1988b). In this linear case, therefore, the decomposition (2.12) leads to a simple constant correction term in the limit theory that involves $\lambda_{\eta u}$.

3 Extension to α-mixing sequences

Let $\{u_i, v_i\}_{i\geq 1}$ be a sequence of stationary α-mixing random variables[4] with mean zero and coefficients $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 6$, and $\mathbb{E}|u_1|^6 + \mathbb{E}|v_1|^6 < \infty$. Write

$$U_{nk} = \frac{1}{\sqrt{n}\sigma_u} \sum_{i=1}^{k} u_i, \quad V_{nk} = \frac{1}{\sqrt{n}\sigma_v} \sum_{i=1}^{k} v_i, \quad 1 \leq k \leq n,$$

where $\sigma_u^2 = \mathbb{E}u_1^2 + 2\sum_{i=1}^{\infty} \mathbb{E}u_1u_{1+i}$ and $\sigma_v^2 = \mathbb{E}v_1^2 + 2\sum_{i=1}^{\infty} \mathbb{E}v_1v_{1+i}$ are the long run variances of u_i and v_i. According to standard functional limit theory and for any continuous function $g(x)$

$$\big\{U_{n,\lfloor nt\rfloor}, V_{n,\lfloor nt\rfloor}, \frac{1}{n}\sum_{k=1}^{n} g(U_{nk})\big\} \Rightarrow \big\{U(t), V(t), \int_0^1 g[U(t)]dt\big\}, \tag{3.1}$$

[4] A sequence $\{\zeta_k, k \geq 1\}$ is said to be α-mixing if

$$\alpha(n) := \sup_{k\geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m = \sigma\{\zeta_l, \zeta_{l+1}, \ldots, \zeta_m\}$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l < m$.

on $D_{\mathbb{R}^3}[0,1]$, where $(U(t), V(t))$ is bivariate Brownian motion with covariance matrix:

$$\Omega = \begin{pmatrix} 1 & \sigma_{uv}/\sigma_u\sigma_v \\ \sigma_{uv}/\sigma_u\sigma_v & 1 \end{pmatrix},$$

where $\sigma_{uv} = \mathbb{E}u_1v_1 + \sum_{i=1}^{\infty}(\mathbb{E}u_1v_{1+i} + \mathbb{E}v_1u_{1+i})$ is the long run covariance of (u_i, v_i). See, De Jong and Davidson (2000a, b), for instance.

Write $\Lambda_{vu} = \sum_{k=1}^{\infty} \mathbb{E}(u_1v_{k+1})$ and $\Delta_{vu} = \sum_{k=0}^{\infty} \mathbb{E}(u_1v_{k+1})$. Regarding weak convergence of the sample covariance functional $\frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})v_{k+1}$, we have the following result.

THEOREM 3.1. *For any function $f(x)$ satisfying **A3** and for any continuous function $g(s)$, we have*

$$\begin{aligned} &\Big\{U_{n,\lfloor nt\rfloor}, V_{n,\lfloor nt\rfloor}, \frac{1}{n}\sum_{k=1}^{n} g(U_{nk}),\ \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})v_{k+1}\Big\} \\ \Rightarrow\ &\Big\{U(t), V(t), \int_0^1 g[U(t)]dt,\ \int_0^1 f[U(t)]dV(t) + \widetilde{\Lambda}_{vu}\int_0^1 f'[U(t)]dt\Big\}, \end{aligned} \tag{3.2}$$

where $\widetilde{\Lambda}_{vu} = \frac{1}{\sigma_u\sigma_v}\Lambda_{vu}$. We also have

$$\begin{aligned} &\Big\{U_{n,\lfloor nt\rfloor}, V_{n,\lfloor nt\rfloor}, \frac{1}{n}\sum_{k=1}^{n} g(U_{nk}),\ \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n} f(U_{nk})v_{k}\Big\} \\ \Rightarrow\ &\Big\{U(t), V(t), \int_0^1 g[U(t)]dt,\ \int_0^1 f[U(t)]dV(t) + \widetilde{\Delta}_{vu}\int_0^1 f'[U(t)]dt\Big\}, \end{aligned} \tag{3.3}$$

where $\widetilde{\Delta}_{vu} = \frac{1}{\sigma_u\sigma_v}\Delta_{vu}$.

The quantities $\widetilde{\Lambda}_{vu} = \frac{1}{\sigma_u\sigma_v}\Lambda_{vu}$ and $\widetilde{\Delta}_{vu} = \frac{1}{\sigma_u\sigma_v}\Delta_{vu}$ in (3.2) and (3.3) are standardized versions of the one-sided long run covariances $\Lambda_{vu} = \sum_{k=1}^{\infty}\mathbb{E}(u_1v_{k+1})$ and $\Delta_{vu} = \sum_{k=0}^{\infty}\mathbb{E}(u_1v_{k+1})$. These quantities embody temporal correlation effects between the stationary inputs (u_i, v_i) and they commonly arise in sample covariance limits between $I(1)$ and $I(0)$ time series in linear models, as detailed in early work (Phillips, 1987, 1988a, 1988b; Park and Phillips, 1988, 1989) on nonstationary time series regression.

Convergence to stochastic integrals for mixing sequence was first considered in Hansen (1992) and later by De Jong and Davidson (2000a, b) with $f(x) = x$. The first extension to general $f(x)$ was investigated in an unpublished paper de Jong (2002). The technique used in that work requires $\sup_{0\le t\le 1}(|U_{n,\lfloor nt\rfloor} - U(t)| + |V_{n,\lfloor nt\rfloor} - V(t)|) \to_{a.s.} 0$ and $D[0,1]^2$ is equipped with uniform metric. This uniform strong convergence condition is quite stringent. The conditions of Theorem 3.1 are simple and only require that $\{u_i, v_i\}_{i\ge 1}$ is stationary and α-mixing with a power law decay rate and corresponding moment condition. These conditions are widely applicable and verification is straightforward under simple primitive conditions. The sixth moment condition on the components (u_i, v_i) appears more restrictive than usual and is made for technical reasons to simplify proofs. The authors conjecture that the condition may be relaxed.

4 Econometric applications

Let $\{\epsilon_i, \eta_i\}_{i\in\mathcal{Z}}$ be an *iid* sequence with zero means, unit variances and covariance $\rho = \mathbb{E}\epsilon_0\eta_0$. According to standard functional limit theory we have the weak convergence

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\eta_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\eta_{-i}) \Rightarrow \big(W(t),\ W_1(t),\ W_2(t)\big)$$

on $D_{\mathbb{R}^3}[0,1]$ in the Skorohod topology, where $W_2(t)$ is a standard Brownian motion independent of $(W(t), W_1(t))$, which is bivariate Brownian motion with covariance matrix:

$$\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Define the linear process $u_k = \sum_{j=0}^{\infty}\varphi_j\,\epsilon_{k-j}$ with $\varphi = \sum_{j=0}^{\infty}\varphi_j \neq 0$ and $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$, and the standardized array $z_{nk} = \frac{1}{d_n}\sum_{j=1}^{k} z_j$, where z_j is a functional of $\eta_j, \eta_{j-1}, ...$ satisfying $\mathbb{E}z_j = 0$ and $d_n^2 = var(\sum_{j=1}^{n} z_j)$. Theorems 2.2 and 2.3 can be used to establish the asymptotic distribution of the sample covariance functional

$$S_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(z_{nk})\,u_{k+1},$$

for many arrays z_{nk} that arise in regression applications in econometrics. The following are two examples involving partial sums of long and short memory linear processes.

Example 2. (Long memory linear process). Let $z_j = \sum_{k=0}^{\infty}\psi_k\,\eta_{j-k}$, where $\psi_k \sim k^{-\mu}\,h(k)$, where $1/2 < \mu < 1$ and $h(k)$ is a function that is slowly varying at ∞. Then, for any function $f(s)$ satisfying a local Lipschitz condition and for any continuous function $g(s)$, we have by Theorem 2.2, as verified in the Section 6,

$$\Big\{\frac{1}{n}\sum_{k=1}^{n} g(z_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(z_{nk})\,u_{k+1}\Big\} \Rightarrow \Big\{\int_0^1 g[G(t)]dt,\ \varphi\int_0^1 f[G(t)]\,dW(t)\Big\}, \tag{4.1}$$

where $G(t) = W_{3/2-\mu}(t)$ and $W_d(t)$ is a fractional Brownian motion defined by

$$W_d(t) = \frac{1}{A(d)} \int_{-\infty}^{0} \Big[(t-s)^d - (-s)^d\Big] dW_2(s) + \int_0^t (t-s)^d dW_1(s),$$

with

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

Example 3. (Short memory linear process). Let $z_j = \sum_{k=0}^\infty \psi_k \, \eta_{j-k}$, where $\sum_{k=0}^\infty |\psi_k| < \infty$. Suppose that $\mathbb{E}|\epsilon_0|^4 + \mathbb{E}|\eta_0|^4 < \infty$. Then, for any function $f(s)$ satisfying **A3** and for any continuous function $g(s)$, we have by Theorem 2.3

$$\Big\{ \frac{1}{n}\sum_{k=1}^n g(z_{nk}), \ \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(z_{nk})\, u_{k+1}\Big\} \Rightarrow \Big\{ \int_0^1 g[W_1(t)]dt, \ \varphi \int_0^1 f[W_1(t)]\, dW(t) + A_0 \int_0^1 f'[W_1(t)]dt\Big\}, \tag{4.2}$$

where $A_0 = \rho \sum_{j=1}^\infty \varphi_j \sum_{k=0}^j \psi_k$, as verified in Section 6.

Limit theorems involving stochastic integrals such as those given in (3.2), (4.1) and (4.2) have many applications in econometrics. They arise frequently in time series regressions with integrated and near integrated processes, unit root testing and nonlinear co-integration theory. Examples can be found in Park and Phillips (2000, 2001), Chang, et al. (2001), Wang and Phillips (2009a, b, 2011), Chang and Park (2011), Chan and Wang (2014) and Wang (2014). Using the theorems given here, previous results such as these may be extended to to a wider class of generating mechanisms such as those involving nonlinear functions and long memory innovations, thereby justifying the use of these asymptotic results for estimation and inference in empirical work under broadly applicable conditions. The following nonlinear cointegrating regression model illustrates the use of the methods.

Example 4. (Nonlinear cointegrating regression)

We consider the nonlinear in variables cointegrating model

$$y_t = \alpha + \beta x_t^2 + v_t, \quad t \ge 1, \tag{4.3}$$

where $x_t = \sum_{j=1}^t u_j$ and $\{u_i, v_i\}_{i\ge 1}$ is stationary α-mixing time series with zero mean. The least squares estimates of α and β are

$$\hat\alpha = \frac{1}{n}\sum_{t=1}^n y_t - \frac{\hat\beta}{n}\sum_{t=1}^n x_t^2, \qquad \hat\beta = \frac{\sum_{t=1}^n y_t x_t^2 - n^{-1}\sum_{t=1}^n x_t^2 \sum_{t=1}^n y_t}{\sum_{t=1}^n x_t^4 - n^{-1}(\sum_{t=1}^n x_t^2)^2}.$$

In the analysis that follows it is convenient to use the same notation for the components $\sigma_u, \sigma_v, \widetilde{\Delta}_{vu}, U_{nk}, V_{nk}, U(t)$ and $V(t)$ given earlier in Section 3. Accordingly, we can write the estimation errors for $\hat{\beta}$ and $\hat{\alpha}$ as

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^n v_t(x_t^2 - n^{-1}\sum_{t=1}^n x_t^2)}{\sum_{t=1}^n x_t^4 - n^{-1}(\sum_{t=1}^n x_t^2)^2} \tag{4.4}$$

$$= n^{-3/2}\sigma_u^{-2}\sigma_v \frac{\frac{1}{\sqrt{n}\sigma_v}\sum_{t=1}^n U_{n,t}^2 v_t - \frac{1}{n}\sum_{t=1}^n U_{n,t}^2 \frac{1}{\sqrt{n}\sigma_v}\sum_{t=1}^n v_t}{\frac{1}{n}\sum_{t=1}^n U_{n,t}^4 - (\frac{1}{n}\sum_{t=1}^n U_{n,t}^2)^2}, \tag{4.5}$$

$$\hat{\alpha} - \alpha = \frac{1}{n}\sum_{t=1}^n v_t - \frac{\hat{\beta} - \beta}{n}\sum_{t=1}^n x_t^2$$

$$= n^{-1/2}\sigma_v\big[V_{n,n} - n^{3/2}\sigma_u^2\sigma_v^{-1}(\hat{\beta} - \beta)\frac{1}{n}\sum_{t=1}^n U_{n,t}^2\big].$$

Direct application of Theorem 3.1 and the continuous mapping theorem yields the following limit theory under the assumptions that the α-mixing decay rate is $\alpha(m) = O(m^{-\gamma})$ for some $\gamma > 6$ and the moment condition $\mathbb{E}|u_1|^6 + \mathbb{E}|v_1|^6 < \infty$ holds. Specifically, we have

$$n^{3/2}\sigma_u^2\sigma_v^{-1}(\hat{\beta} - \beta) \to_D \quad Y, \tag{4.6}$$

$$n^{1/2}\sigma_v^{-1}(\hat{\alpha} - \alpha) \to_D \quad V(1) - Y\int_0^1 U^2(t)dt, \tag{4.7}$$

where

$$Y = \frac{\int_0^1 U^2(t)dV(t) + 2\widetilde{\Delta}_{vu}\int_0^1 U(t)dt - V(1)\int_0^1 U^2(t)dt}{\int_0^1 U^4(t)dt - (\int_0^1 U^2(t)dt)^2} \tag{4.8}$$

$$= \frac{\int_0^1 \widetilde{U^2}(t)dV(t) + 2\widetilde{\Delta}_{vu}\int_0^1 U(t)dt}{\int_0^1 \big[\widetilde{U^2}(t)\big]^2 dt},$$

where $\widetilde{U^2}(t) := U^2(t) - \int_0^1 U^2(t)dt$ is a demeaned version of $U^2(t)$. The limit (4.8) follows from the joint weak convergence (3.3) of Theorem 3.1. In particular for the sample covariance term in the numerator of (4.5) we have

$$\frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^n U_{nk}^2 v_k \Rightarrow \int_0^1 U(t)^2 dV(t) + 2\widetilde{\Delta}_{vu}\int_0^1 U(t)dt.$$

The convergence rate for the intercept $\hat{\alpha}$ is $\sqrt{n}$, as usual, but the limit distribution is not normal. So the intercept asymptotics bear the effect of the slope coefficient limit distribution. That distribution is non-normal and is delivered by joint weak convergence of the sample covariance in the numerator of (4.4) in conjunction with the quadratic functional of x_t^2 in the denominator. The slope coefficient $\hat{\beta}$ has an $n^{3/2}$ convergence rate, reflecting the stronger signal $\sum_{t=1}^{n} x_t^4$ from the squared $I(1)$ regressor x_t^2.

Example 5. (Nonlinear FM regression)

In view of the nuisance parameters involved in Y in (4.8) the limit theory in (4.6) and (4.7) is not immediately amenable to inference. As usual, corrections to least squares regression are required to achieve feasible inference by removing the nuisance parameters to produce estimates with a limiting mixed normal distribution and asymptotically pivotal statistics for testing. A simple mechanism to achieve these corrections in the linear cointegrating case is fully modified (FM) least squares (Phillips and Hansen, 1990). That approach extends to the present case, as we now demonstrate.

The details follow Phillips and Hansen (1990) in broad outline with modifications that account for the nonlinearity. Note first that, just as in Theorem 3.1 and (3.3), we have the joint convergence

$$\begin{bmatrix} \frac{1}{\sqrt{n}\sigma_v} \sum_{k=1}^{n} f(U_{nk}) v_k \\ \frac{1}{\sqrt{n}\sigma_u} \sum_{k=1}^{n} f(U_{nk}) u_k \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^1 f[U(t)]dV(t) + \widetilde{\Delta}_{vu} \int_0^1 f'[U(t)]dt \\ \int_0^1 f[U(t)]dU(t) + \widetilde{\Delta}_{uu} \int_0^1 f'[U(t)]dt \end{bmatrix},$$

where

$$\widetilde{\Delta}_{uu} = \frac{1}{\sigma_u^2}\Delta_{uu} = \frac{1}{\sigma_u^2}\sum_{k=0}^{\infty}\mathbb{E}(u_1 u_{k+1}) \text{ and } \widetilde{\Delta}_{vu} = \frac{1}{\sigma_v\sigma_u}\Delta_{vu} = \frac{1}{\sigma_v\sigma_u}\sum_{k=0}^{\infty}\mathbb{E}(u_1 v_{k+1}).$$

Next, observe that least squares estimates of (4.3) may be used to construct conventional (lag kernel based) consistent estimates of the long run variance and covariance parameters $\sigma_u^2, \sigma_v^2, \sigma_{uv}$, which we denote by $\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\sigma}_{uv}$ (e.g., Park and Phillips, 1988). To develop the FM regression estimates of (4.3), we define the augmented regression equation

$$y_t = \alpha + \beta x_t^2 + \frac{\sigma_{vu}}{\sigma_u^2}\Delta x_t + w_{v.u,t}, \quad w_{v.u,t} = v_t - \frac{\sigma_{vu}}{\sigma_u^2}u_t, \tag{4.9}$$

where $\sigma_{vu} = \rho_{vu}\sigma_v\sigma_u$, and ρ_{uv} is the long run correlation coefficient between u_i and v_i. The control variable $\frac{\sigma_{vu}}{\sigma_u^2}\Delta x_t$ in (4.9) captures the (long run) endogeneity effect in the regression equation. The corresponding endogeneity-corrected dependent variable is $y_t^+ := y_t - \frac{\sigma_{vu}}{\sigma_u^2}\Delta x_t$, which is estimated by $\hat{y}_t^+ = y_t - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2}\Delta x_t$. The equation error in (4.9) is $w_{v.u,t}$ which is stationary with zero mean and long run variance $\sigma_v^2 - \frac{\sigma_{vu}^2}{\sigma_u^2} = \sigma_v^2\left(1-\rho_{uv}^2\right)$. Next, define the serial correlation correction $\hat{\Delta}_{v.u} = \hat{\Delta}_{vu} - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2}\hat{\Delta}_{uu}$ constructed in the usual way (Phillips and Hansen, 1990) as a consistent estimate of the one-sided long run covariance

$$\Delta_{v.u} = \sum_{k=0}^{\infty}\mathbb{E}(u_1 w_{v.u,k+1}) = \Delta_{vu} - \frac{\sigma_{vu}}{\sigma_u^2}\Delta_{uu},$$

where

$$\Delta_{uu} = \sum_{k=0}^{\infty} \mathbb{E}(u_1 u_{k+1}) \text{ and } \Delta_{vu} = \sum_{k=0}^{\infty} \mathbb{E}(u_1 v_{k+1}).$$

Define the demeaned regressor as $\widetilde{x}_t^2 := x_t^2 - n^{-1} \sum_{t=1}^n x_t^2$. Then, the FM regression estimator of the slope coefficient β in (4.3) is constructed as

$$\hat{\beta}^+ = \frac{\sum_{t=1}^n \left\{ \hat{y}_t^+ x_t^2 - 2\sqrt{n}\hat{\Delta}_{v.u} x_t \right\} - n^{-1} \sum_{t=1}^n x_t^2 \sum_{t=1}^n \hat{y}_t^+}{\sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2} = \frac{\sum_{t=1}^n \left\{ \hat{y}_t^+ \widetilde{x}_t^2 - 2\sqrt{n}\hat{\Delta}_{v.u} x_t \right\}}{\sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2},$$

which embodies the endogeneity correction in $\hat{y}_t^+$ and the temporal correlation correction $\hat{\Delta}_{v.u}$. Noting that $\sum_{t=1}^n \widetilde{x}_t^2 = 0$, $\sum_{t=1}^n x_t^2 \widetilde{x}_t^2 = \sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2$ and

$$\begin{aligned} \hat{y}_t^+ &= y_t - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2} \Delta x_t = \alpha + \beta x_t^2 + v_t - \frac{\sigma_{vu}}{\sigma_u^2} \Delta x_t + \left(\frac{\sigma_{vu}}{\sigma_u^2} - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2} \right) \Delta x_t \\ &= \alpha + \beta x_t^2 + w_{v.u,t} + \left(\frac{\sigma_{vu}}{\sigma_u^2} - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2} \right) u_t, \end{aligned}$$

we may write the estimation error of $\hat{\beta}^+$ as

$$\begin{aligned} \hat{\beta}^+ - \beta &= \frac{\sum_{t=1}^n \left\{ \widetilde{x}_t^2 w_{v.u,t} - 2\sqrt{n}\Delta_{v.u} x_t + 2\sqrt{n} \left(\Delta_{v.u} - \hat{\Delta}_{v.u} \right) x_t + \left(\frac{\sigma_{vu}}{\sigma_u^2} - \frac{\hat{\sigma}_{vu}}{\hat{\sigma}_u^2} \right) u_t \widetilde{x}_t^2 \right\}}{\sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2} \\ &= \frac{\sigma_v}{n^{3/2}\sigma_u^2} \frac{\frac{1}{\sqrt{n}\sigma_v} \sum_{t=1}^n \widetilde{U}_{n,t}^2 w_{v.u,t} - 2\widetilde{\Delta}_{v.u} \frac{1}{n} \sum_{t=1}^n U_{n,t} + o_p(1)}{\frac{1}{n} \sum_{t=1}^n \left(\widetilde{U}_{n,t}^2 \right)^2}, \end{aligned}$$

where $\widetilde{U}_{n,t}^2 = U_{n,t}^2 - n^{-1} \sum_{t=1}^n U_{n,t}^2$ and $\widetilde{\Delta}_{v.u} = \frac{\Delta_{v.u}}{\sigma_u \sigma_v}$. Then, defining

$$V_{v.u}(t) := V(t) - \rho_{vu} U(t) = BM\left(1 - \rho_{vu}^2\right)$$

and noting $U(t)$ is independent of $V_{v.u}(t)$, we have

$$\begin{aligned} n^{3/2}\left(\hat{\beta}^+ - \beta \right) &\to_D \frac{\sigma_v}{\sigma_u^2} \frac{\int_0^1 \widetilde{U}^2(t) dV_{v.u}(t) + 2\widetilde{\Delta}_{v.u} \int_0^1 U(t)dt - 2\bar{\Delta}_{v.u} \int_0^1 U(t)dt}{\int_0^1 \left[\widetilde{U}^2(t)\right]^2 dt} \\ &= \frac{\sigma_v}{\sigma_u^2} \frac{\int_0^1 \widetilde{U}^2(t) dV_{v.u}(t)}{\int_0^1 \left[\widetilde{U}^2(t)\right]^2 dt} \equiv MN\left(0, \frac{\sigma_v^2 \left(1 - \rho_{vu}^2\right)}{\int_0^1 \left(\sigma_u^2 \widetilde{U}^2(t) \right)^2 dt} \right), \end{aligned}$$

giving a mixed normal (MN) limit distribution that is centred on the origin.

This limit theory for $n^{3/2}\left(\hat{\beta}^+ - \beta\right)$ leads naturally to pivotal statistical inference just as in the linear case. In particular, the (semiparametric) cointegrating t ratio for β is

$$t_\beta = \frac{\hat{\beta}^+ - \beta}{s_\beta^+} \to_D N(0,1),$$

where the standardization has the usual form $s_\beta^+ = \left(\hat{\sigma}_{v.u}^2 / \sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2\right)^{1/2}$, which employs the long run error variance estimate $\hat{\sigma}_{v.u}^2 = \hat{\sigma}_v^2 - \hat{\sigma}_{vu}^2/\hat{\sigma}_u^2$. Then

$$
\begin{aligned}
t_\beta &= \frac{\hat{\beta}^+ - \beta}{s_\beta^+} = \frac{n^{3/2}\left(\hat{\beta}^+ - \beta\right)}{\left(\hat{\sigma}_{v.u}^2 / \frac{1}{n^3}\sum_{t=1}^n \left(\widetilde{x}_t^2\right)^2\right)^{1/2}} \\
&= \frac{\sigma_v}{\hat{\sigma}_{v.u}} \frac{\frac{1}{\sqrt{n}\sigma_v}\sum_{t=1}^n \widetilde{U}_{n,t}^2 w_{v.u,t} - 2\widetilde{\Delta}_{v.u}\frac{1}{n}\sum_{t=1}^n U_{n,t} + o_p(1)}{\left(\frac{1}{n}\sum_{t=1}^n \left(\widetilde{U}_{n,t}^2\right)^2\right)^{1/2}} \\
&\to_D \frac{\sigma_v}{\sigma_{v.u}} \frac{\int_0^1 \widetilde{U}^2(t) dV_{v.u}(t)}{\left(\int_0^1 \left[\widetilde{U}^2(t)\right]^2 dt\right)^{1/2}} \equiv N(0,1),
\end{aligned}
$$

since $\left(\int_0^1 \left[\widetilde{U}^2(t)\right]^2 dt\right)^{-1/2} \int_0^1 \widetilde{U}^2(t) dV_{v.u}(t) = N\left(0, \left(1-\rho_{vu}^2\right)\right)$ and $\sigma_v^2\left(1-\rho_{vu}^2\right)/\sigma_{v.u}^2 = 1$.

5 Conclusion

Many applications in time series econometrics involve cointegrating links where nonlinearities, endogeneity, and long memory effects complicate the usual limit theory for linear cointegrated systems. The weak convergence limit theory given here provides simple conditions under which that limit theory is extended to such cases, including sample covariances involving nonlinear functions with limiting forms as stochastic integrals with stochastic drift functionals. The results obtained complement earlier limit theory and show how regression methods like FM regression may be extended to a nonlinear framework. The authors hope the results are accessible and prove useful in econometric applications of time series regression with nonstationary, nonlinear, and long memory components.

6 Proofs

Proof of Proposition 2.1. For notational convenience, we remove the tilde affix on $\widetilde{y}_{nk}$, $\widetilde{\epsilon}_{nk}$ and $\widetilde{u}_{nk}$ in what follows. Simple calculations show that

$$\begin{aligned}
\sum_{i=1}^{m} f(y_{n,i-1})u_{ni} &= \sum_{i=1}^{m} f(y_{n,i-1})(\sum_{j=0}^{i-1}+\sum_{j=i}^{\infty})\varphi_j\epsilon_{n,i-j} \\
&= \sum_{j=0}^{m-1}\varphi_j \sum_{i=1+j}^{m} f(y_{n,i-1})\epsilon_{n,i-j} + \sum_{i=1}^{m}\sum_{j=0}^{\infty}\varphi_{j+i} f(y_{n,i-1})\epsilon_{n,-j} \\
&= \sum_{j=0}^{m-1}\varphi_j \sum_{i=0}^{m-j-1} f(y_{n,i+j})\epsilon_{n,i+1} + \sum_{j=0}^{\infty}\epsilon_{n,-j}\sum_{i=1}^{m}\varphi_{j+i} f(y_{n,i-1}) \\
&= \sum_{j=0}^{m-1}\varphi_j \sum_{i=0}^{m} f(y_{n,i+j})\epsilon_{n,i+1} - \sum_{j=0}^{m-1}\varphi_j \sum_{i=m-j}^{m} f(y_{n,i+j})\epsilon_{n,i+1} \\
&\qquad + \sum_{j=0}^{\infty}\epsilon_{n,-j}\sum_{i=1}^{m}\varphi_{j+i} f(y_{n,i-1}) \\
&= \varphi\sum_{i=0}^{m} f(y_{ni})\epsilon_{n,i+1} + \sum_{j=0}^{m-1}\varphi_j\sum_{i=0}^{m}[f(y_{n,i+j})-f(y_{n,i})]\epsilon_{n,i+1} \\
&\qquad -R_1(m)-R_2(m)+R_3(m),
\end{aligned}$$

where $R_1(m)=\sum_{j=m}^{\infty}\varphi_j\sum_{i=0}^{m} f(y_{ni})\epsilon_{n,i+1}$,

$$R_2(m)=\sum_{j=0}^{m-1}\varphi_j\sum_{i=m-j}^{m} f(y_{n,i+j})\epsilon_{n,i+1}, \quad R_3(m)=\sum_{j=0}^{\infty}\epsilon_{n,-j}\sum_{i=1}^{m}\varphi_{j+i} f(y_{n,i-1}).$$

It suffices to show that, for each $1 \le m \le n$,

$$|R_j(m)| = o_P(1), \quad j=1,2,3, \tag{6.1}$$

and under the additional condition $\max_{1\le i<k\le n}\frac{1}{k-i}\sum_{j=i}^{k}|\epsilon_{nj}| = o_P(1)$

$$\max_{1\le m\le n}|R_j(m)| = o_P(1), \quad j=1,2,3. \tag{6.2}$$

To this end, write $\Omega_K = \{y_{ni} : \max_{1\le i\le n}|y_{ni}| \le K\}$. As $f(x)$ is a locally bounded function, we have $\max_{1\le k\le n}|f(y_{nk})| \le A_K$, on Ω_K, for some $A_K > 0$. Also note that, under (2.10),

$$\sup_{-\infty<i<j<\infty}\frac{1}{j-i}\sum_{k=i+1}^{j}\mathbb{E}|\epsilon_{nk}| \le \sup_{j\ge1,i\in Z} j^{-1/2}\big(\sum_{k=1}^{j}\mathbb{E}\epsilon_{n,k+i}^2\big)^{1/2} \to 0,$$

as $n \to \infty$, due to Hölder's inequality. Combining these facts and $\sum_{j=0}^{\infty} j\,|\varphi_j| < \infty$, we have

$$\begin{aligned}
\mathbb{E}|R_1(m)|I(\Omega_K) &\leq A_K \sum_{j=m}^{\infty} |\varphi_j| \sum_{i=0}^{m} \mathbb{E}|\epsilon_{n,i+1}| = o(1), &(6.3)\\
\mathbb{E}|R_2(m)|I(\Omega_K) &\leq A_K \sum_{j=0}^{m-1} |\varphi_j| \sum_{i=m-j}^{m} \mathbb{E}|\epsilon_{n,i+1}| = o(1), &(6.4)\\
\mathbb{E}\max_{1\leq m\leq n} |R_3(m)|I(\Omega_K) &\leq A_K \sum_{j=0}^{\infty} \mathbb{E}|\epsilon_{n,-j}| \sum_{i=j}^{\infty} |\varphi_i| \leq C \sum_{i=0}^{\infty} |\varphi_i| \sum_{k=0}^{i} \mathbb{E}|\epsilon_{n,-k}| \\
&= o(1). &(6.5)
\end{aligned}$$

Hence $(|R_1(m)| + |R_2(m)| + |R_3(m)|)I(\Omega_K) = o_P(1)$ for each m. This proves (6.1) as $P(\max_{1\leq i\leq n} |y_{ni}| > K) \to 0$ as $K \to \infty$.

We next prove (6.2). In fact, as in (6.3), we have

$$\begin{aligned}
\max_{1\leq m\leq n} |R_1(m)|I(\Omega_K) &\leq A_K \max_{1\leq m\leq n} \sum_{j=m}^{\infty} j\,|\varphi_j| \frac{1}{m} \sum_{i=0}^{m} |\epsilon_{n,i+1}| \\
&\leq C\,A_K \max_{1\leq m\leq n} \frac{1}{m} \sum_{i=0}^{m} |\epsilon_{n,i+1}| = o_P(1),
\end{aligned}$$

due to the additional condition $\max_{1\leq i<k\leq n} \frac{1}{k-i} \sum_{j=i}^{k} |\epsilon_{nj}| = o_P(1)$. This yields $\max_{1\leq m\leq n} |R_1(m)| = o_P(1)$ due to $P(\Omega_K) \to 1$ as $K \to \infty$. Similarly, we have (6.2) with $j = 2$. The result $\max_{1\leq m\leq n} |R_3(m)| = O_P(1)$ follows from (6.5) and $P(\Omega_K) \to 1$ as $K \to \infty$. The proof of Proposition 2.1 is now complete. □

*Proof of Expression (2.12)***.** Removing the tilde affix again, applying the BN decomposition (Phillips and Solo, 1992), using summation by parts, and setting $\epsilon^*_{n,i} = \sum_{j=0}^{\infty} \varphi^*_j \epsilon_{n,i-j}$ with $\varphi^*_j = \sum_{s=j+1}^{\infty} \varphi_s$,we have

$$\begin{aligned}
&\sum_{i=1}^{m} f(y_{n,i-1})u_{ni} = \sum_{i=1}^{m} f(y_{n,i-1}) \left(\sum_{j=0}^{\infty} \varphi_j \epsilon_{n,i-j}\right) = \sum_{i=1}^{m} f(y_{n,i-1}) \left\{ \left(\sum_{j=0}^{\infty} \varphi_j\right) \epsilon_{n,i} + \epsilon^*_{n,i-1} - \epsilon^*_{n,i} \right\} \\
&= \varphi \sum_{i=1}^{m} f(y_{n,i-1})\epsilon_{n,i} - \sum_{i=1}^{m} f(y_{n,i-1}) \left(\epsilon^*_{n,i} - \epsilon^*_{n,i-1}\right) \\
&= \varphi \sum_{i=1}^{m} f(y_{n,i-1})\epsilon_{n,i} - \left[\left\{ f(y_{n,m})\epsilon^*_{n,m} - f(y_{n,0})\epsilon^*_{n,0} \right\} - \sum_{i=1}^{m} \left\{ f(y_{n,i}) - f(y_{n,i-1}) \right\} \epsilon^*_{n,i} \right] \\
&= \varphi \sum_{i=1}^{m} f(y_{n,i-1})\epsilon_{n,i} + \sum_{i=1}^{m} \left\{ f(y_{n,i}) - f(y_{n,i-1}) \right\} \epsilon^*_{n,i} + r_m
\end{aligned}$$

where $r_m = f(y_{n,m})\epsilon^*_{n,m} - f(y_{n,0})\epsilon^*_{n,0}$.

Proof of Theorem 2.2. It is readily seen that

$$\sup_{j\geq 1, i\in Z} \frac{1}{j}\sum_{k=1}^{j} \mathbb{E}\epsilon_{n,k+i}^2 \leq \sup_{k\in Z} \mathbb{E}\epsilon_{nk}^2 \to 0,$$

and $\max_{1\leq k\leq n}|y_{nk}| \to_D \sup_{0\leq t\leq 1}|G(t)| = O_P(1)$ by the condition **A1**. By Theorem 2.1 and Proposition 2.1 with $\widetilde{y}_{nk} = y_{nk}$ and $\widetilde{\epsilon}_{nk} = \epsilon_{nk}$, Theorem 2.2 will follow if we prove, for all $K > 0$,

$$\Delta_n := I(\max_{1\leq k\leq 2n}|y_{nk}| \leq K)\sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n}[f(y_{n,i+j}) - f(y_{n,i})]\epsilon_{n,i+1} = o_P(1). \tag{6.6}$$

In fact, by Hölder's inequality, (2.2) and the fact that $f(s)$ satisfies the local Lipschitz condition, we have

$$\begin{aligned}
\mathbb{E}|\Delta_n| &\leq C\sum_{j=0}^{n-1}|\varphi_j|\sum_{i=0}^{n}\mathbb{E}\big(|y_{n,i+j} - y_{n,i}|\,|\epsilon_{n,i+1}|\big) \\
&\leq C\sum_{j=0}^{n-1}|\varphi_j|\big(\sum_{i=0}^{n-1}\mathbb{E}|y_{n,i+j} - y_{n,i}|^2\big)^{1/2}\big(\sum_{i=0}^{n-1}\mathbb{E}|\epsilon_{n,i+1}|^2\big)^{1/2} \\
&\leq C\big(\sup_{i,j\geq 1}\frac{n}{j^2}\mathbb{E}|y_{n,i+j} - y_{n,i}|^2\big)^{1/2}\sum_{j=0}^{n-1}j|\varphi_j|\big(\sum_{i=0}^{n-1}\mathbb{E}|\epsilon_{n,i+1}|^2\big)^{1/2} \\
&= o(1),
\end{aligned} \tag{6.7}$$

due to $\sum_{i=0}^{n-1}\mathbb{E}|\epsilon_{n,i+1}|^2 = O(1)$. This proves (6.6) and completes the proof of Theorem 2.2. □

Proof of Theorem 2.3. Let $\widetilde{y}_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k}\eta_j$ and $\widetilde{\epsilon}_{nk} = \frac{1}{\sqrt{n}}\epsilon_j$. Due to **A1** and $\sup_{j\geq 1, i\in\mathcal{Z}}\frac{1}{j}\sum_{k=1}^{j}\mathbb{E}|\epsilon_{k+i}|^3 < \infty$, we have $\max_{1\leq k\leq n}|\widetilde{y}_{nk}| = O_P(1)$ and (2.10), respectively. Proposition 2.1 with $\widetilde{y}_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k}\eta_j$ and $\widetilde{\epsilon}_{nk} = \frac{1}{\sqrt{n}}\epsilon_j$ yields

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f(y_{n,i-1})u_i &= \frac{\varphi}{\sqrt{n}}\sum_{i=0}^{n}f(y_{ni})\epsilon_{i+1} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n}[f(y_{n,i+j}) - f(y_{ni})]\epsilon_{i+1} + o_P(1).
\end{aligned} \tag{6.8}$$

Noting $f(y_{n,i+j}) - f(y_{ni}) = f'(y_{ni})(y_{n,i+j} - y_{n,i}) + \int_{y_{ni}}^{y_{n,i+j}} \big[f'(x) - f'(y_{ni})\big]dx$, we have

$$\begin{aligned}
&\sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n}[f(y_{n,i+j}) - f(y_{ni})]\epsilon_{i+1} \\
=& \sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n} f'(y_{ni})\,\delta_{n,ij} + R_1(n) \\
=& \sum_{i=0}^{n} f'(y_{ni})\,\mathbb{E}\big(Z_{ni}\mid\mathcal{F}_i\big) + R_1(n) + R_2(n), \qquad (6.9)
\end{aligned}$$

where $\delta_{n,ij} = (y_{n,i+j} - y_{n,i})\epsilon_{i+1} = \frac{1}{\sqrt{n}}\sum_{k=1}^{j}\eta_{k+i}\epsilon_{i+1}$, $Z_{ni} = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\varphi_j\sum_{k=1}^{j}\eta_{k+i}\epsilon_{i+1}$,

$$\begin{aligned}
|R_1(n)| &\le \sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n}|\epsilon_{i+1}|\,\Big|\int_{y_{ni}}^{y_{n,i+j}}\big[f'(x) - f'(y_{ni})\big]dx\,\Big| \\
&\le \sum_{j=0}^{n-1}\varphi_j\sum_{i=0}^{n}|\epsilon_{i+1}|\int_0^{|y_{n,i+j}-y_{ni}|}\big|f'(x+y_{ni}) - f'(y_{ni})\big|dx\,, \\
R_2(n) &= \sum_{i=0}^{n} f'(y_{ni})\,\big[Z_{ni} - \mathbb{E}\big(Z_{ni}\mid\mathcal{F}_i\big)\big].
\end{aligned}$$

Write $\Omega_K = \{y_{ni} : \max_{1\le i\le 2n}|y_{ni}| \le K/3\}$. Note that $|x + y_{ni}| \le K$ whenever $0 \le x \le |y_{n,i+j} - y_{ni}|$. It follows from **A3** that

$$\begin{aligned}
\mathbb{E}|R_1(n)|I\big(\Omega_K\big) &\le C_K\sum_{j=0}^{n-1}|\varphi_j|\sum_{i=0}^{n}\mathbb{E}\big(|y_{n,i+j} - y_{n,i}|^{1+\beta}\,|\epsilon_{i+1}|\big) \\
&\le C_K\,n^{-(1+\beta)/2}\sum_{j=0}^{n-1}|\varphi_j|\sum_{i=0}^{n}\mathbb{E}\big(|\sum_{k=1}^{j}\eta_{k+i}|^{1+\beta}\,|\epsilon_{i+1}|\big) \\
&\le C_K\,n^{-(1+\beta)/2}\,\big(\sup_{i,j\ge1}\frac{n}{j}\,\mathbb{E}|\sum_{k=1}^{j}\eta_{k+i}|^2\big)^{(1+\beta)/2}\sum_{j=0}^{n-1}j|\varphi_j|\big(\sum_{i=0}^{n}\mathbb{E}|\epsilon_{i+1}|^{2/(1-\beta)}\big)^{(1-\beta)/2} \\
&= O(n^{-(\beta-1)/2}), \qquad (6.10)
\end{aligned}$$

due to $\beta \le 1/3$ and the condition **A4**. This implies that $R_1(n) = O_P(n^{(1-\beta)/2})$, as $P(\Omega_K) \to 1$ as $K \to \infty$.

To discuss $R_2(n)$, let $R_2(n)^* = \sum_{i=0}^{n} f'(y_{ni})I(\max_{1\le k\le i}|y_{nk}| \le K)\,\big[Z_{ni} - \mathbb{E}\big(Z_{ni}\mid\mathcal{F}_i\big)\big]$.

Recalling that y_{ni} is adapted to $\mathcal{F}_i$ and $f'(x)$ is locally bounded, we have

$$\begin{aligned}
\mathbb{E}|R_2(n)^*|^2 &= \mathbb{E}\Big|\sum_{i=0}^{n} f'(y_{ni}) I(\max_{1\le k\le i}|y_{nk}|\le K)\,\big[Z_{ni}-\mathbb{E}\big(Z_{ni}\mid \mathcal{F}_i\big)\big]\Big|^2 \\
&\le \sum_{i=0}^{n}\mathbb{E}\Big\{|f'(y_{ni})|^2 I(\max_{1\le k\le i}|y_{nk}|\le K)\,\mathbb{E}\big(\big[Z_{ni}-\mathbb{E}\big(Z_{ni}\mid\mathcal{F}_i\big)\big]^2\mid\mathcal{F}_i\big)\Big\} \\
&\le C\sum_{i=0}^{n}\mathbb{E}\big[Z_{ni}-\mathbb{E}\big(Z_{ni}\mid\mathcal{F}_i\big)\big]^2 \\
&\le \frac{2C}{n}\sum_{i=0}^{n}\sum_{j=0}^{n-1}(j+1)|\varphi_j|\sum_{j=0}^{n-1}\frac{|\varphi_j|}{j+1}\mathbb{E}\big(\sum_{k=1}^{j}\eta_{k+i}\epsilon_{i+1}\big)^2 \\
&\le \frac{C_1}{n}\sum_{i=0}^{n}\sum_{j=0}^{n-1}|\varphi_j|\sum_{k=1}^{j}\mathbb{E}\big(\eta_{k+i}^2\epsilon_{i+1}^2\big) = O(1),
\end{aligned}$$

whenever **A4** holds. Now, by noting $R_2(n)=R_2(n)^*$ on $\Omega_K=\{y_{ni}:\max_{1\le i\le n}|y_{ni}|\le K\}$, it is readily seen that $R_2(n)=O_P(1)$ due to $P(\Omega_K)\to 1$ as $K\to\infty$.

Combining all these facts, we obtain

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(y_{n,i-1})u_i &= \frac{\varphi}{\sqrt{n}}\sum_{i=0}^{n} f(y_{ni})\epsilon_{i+1} \\
&\quad + \frac{1}{n}\sum_{i=0}^{n} f'(y_{ni})\sum_{j=0}^{n-1}\varphi_j\sum_{k=1}^{j}\mathbb{E}\big(\eta_{k+i}\epsilon_{i+1}\mid\mathcal{F}_i\big)+o_P(1),
\end{aligned}$$

which yields (2.5) due to **A5** and Theorem 2.1. □

Proof of Theorem 3.1. We start with some preliminaries. Let $\mathcal{F}_t=\sigma(u_i,v_i,1<i<t)$, and $\mathcal{F}_s=\sigma(\phi,\Omega)$ be the trivial σ-field for $s<0$. Put $z_i=\sum_{k=1}^{\infty}\mathbb{E}(v_{i+k}|\mathcal{F}_i)$ and $\epsilon_i=\sum_{k=0}^{\infty}[\mathbb{E}(v_{i+k}|\mathcal{F}_i)-\mathbb{E}(v_{i+k}|\mathcal{F}_{i-1})]$. Recalling $\alpha(n)=O(n^{-\gamma})$ for some $\gamma>6$, $\mathbb{E}u_1=\mathbb{E}v_1=0$ and $\mathbb{E}|u_1|^6+\mathbb{E}|v_1|^6<\infty$, standard arguments (see, McLeish (1975), for instance) show that $||\mathbb{E}(v_{i+k}|\mathcal{F}_i)||_3\le C\alpha(k)^{1/6}\,||v_1||_6$ and

$$||z_i||_3\le\sum_{k=1}^{\infty}||\mathbb{E}(v_{i+k}|\mathcal{F}_i)||_3\le C\,||v_1||_6\sum_{k=1}^{\infty}k^{-\gamma/6}<\infty, \tag{6.11}$$

where $||X||_p=(\mathbb{E}|X|^p)^{1/p}$. We further have $\sup_{i\ge1}\mathbb{E}\epsilon_i^2<\infty$,

$$\sup_{i\ge1}\mathbb{E}(|u_i|^{r_1}|z_i|^{r_2})\le(\mathbb{E}u_1^6)^{r_1/6}(\mathbb{E}|z_1|^3)^{r_2/3}<\infty,\quad \text{for any } 1\le r_1,r_2\le 2. \tag{6.12}$$

Consequently, by letting $\lambda_k=u_kz_k-\mathbb{E}(u_kz_k)$, it follows that

$$\sup_{k\ge1}\mathbb{E}|\mathbb{E}(\lambda_k\mid\mathcal{F}_{k-m})|\le 6\alpha^{1/2}(m)\sup_{k\ge1}||\lambda_k||_2\to 0, \tag{6.13}$$

as $m \to \infty$.

We are now ready to prove Theorem 3.1. It is readily seen that $v_i = \epsilon_i + z_{i-1} - z_i$, $\{\epsilon_i, \mathcal{F}_i, i \geq 1\}$ forms a sequence of martingale differences, and

$$\begin{aligned}
\frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})v_{k+1} &= \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})(\epsilon_{k+1} + z_k - z_{k+1}) \\
&= \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})\epsilon_{k+1} + \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1}[f(U_{nk}) - f(U_{n,k-1})]z_k \\
&= \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{n-1} f(U_{nk})\epsilon_{k+1} + \frac{\widetilde{\Lambda}}{n}\sum_{k=1}^{n-1} f'(U_{n,k-1}) + R_1(n) + R_2(n), \qquad (6.14)
\end{aligned}$$

where $\widetilde{\Lambda} = \frac{\mathbb{E}(u_1 z_1)}{\sigma_u\sigma_v} = \frac{\sum_{k=1}^{\infty}\mathbb{E}(u_1 v_{k+1})}{\sigma_u\sigma_v}$, and the remainder terms are

$$\begin{aligned}
R_1(n) &= \frac{1}{\sqrt{n}\sigma_u\sigma_v}\sum_{k=1}^{n-1} z_k \int_{U_{n,k-1}}^{U_{nk}} \big[f'(x) - f'(U_{n,k-1})\big]dx\,, \\
R_2(n) &= \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^{n-1} f'(U_{n,k-1})[u_k z_k - \mathbb{E}(u_k z_k)].
\end{aligned}$$

Write $Y_{n,\lfloor nt\rfloor} = \frac{1}{\sqrt{n}\sigma_v}\sum_{k=1}^{\lfloor nt\rfloor}\epsilon_k$. By virtue of Theorem 2.1 with $\epsilon_{nk} = v_k/(\sqrt{n}\sigma_v)$ and $y_{nk} = U_{nk}$, to prove (3.2), it suffices to show that

$$\big\{U_{n,\lfloor nt\rfloor}, Y_{n,\lfloor nt\rfloor}\big\} \Rightarrow \big\{U(t), V(t)\big\}, \qquad (6.15)$$

and

$$R_i(n) = o_P(1), \quad i = 1,\ 2. \qquad (6.16)$$

The proof of (6.15) is simple. Indeed, by observing that

$$\sup_{0\leq t\leq 1}|Y_{n,\lfloor nt\rfloor} - V_{n,\lfloor nt\rfloor}| = \frac{1}{\sqrt{n}\sigma_v}\sup_{0\leq t\leq 1}\Big|\sum_{k=1}^{\lfloor nt\rfloor}(\epsilon_k - v_k)\Big| \leq \frac{1}{\sqrt{n}\sigma_v}\max_{1\leq k\leq n}|z_k|,$$

(6.15) follows from (3.1) and the fact that, for any $\eta > 0$ and $0 < \delta \leq 1$,

$$P(\max_{1<i<n}|z_i| > \eta\sqrt{n}) < \sum_{i=1}^{n} P(|z_i| > \eta\sqrt{n}) < Cn^{-1-\delta/2}\sum_{i=1}^{n}\mathbb{E}|z_i|^{2+\delta} \to 0,$$

due to (6.11).

To prove (6.16), write $\Omega_K = \{U_{ni} : \max_{1\leq i\leq n}|U_{ni}| \leq K\}$. As in the proof of (6.10), it follows from **A3** and (6.12) that

$$\begin{aligned}
\mathbb{E}|R_1(n)|I\big(\Omega_K\big) &\leq \frac{C_K}{\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\big(|U_{nk} - U_{n,k-1}|^{1+\beta}\,|z_k|\big) \\
&\leq C_K\, n^{-(1+\beta/2)}\sum_{k=1}^{n}\mathbb{E}(|u_k|^{1+\beta}\,|z_k|) = O(n^{-\beta/2}). \qquad (6.17)
\end{aligned}$$

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 $m = \lfloor \log n \rfloor$ and recall $\lambda_k = u_k z_k - \mathbb{E}(u_k z_k)$. We have

$$\begin{aligned} R_2(n) &= \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^n f'(U_{n,k-m-1})\,\lambda_k + \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^n \left[f'(U_{n,k-1}) - f'(U_{n,k-m-1})\right]\lambda_k \\ &= R_{21}(n) + R_{22}(n), \quad \text{say.} \end{aligned} \tag{6.18}$$

As in the proof of (6.17), it is readily seen that

$$\begin{aligned} \mathbb{E}|R_{22}(n)|I\big(\Omega_K\big) &\leq C_K n^{-1}\sum_{k=1}^n \mathbb{E}\big(\big|U_{n,k-1} - U_{n,k-m-1}\big|^\beta\,|\lambda_k|\big) \\ &\leq C_K n^{-1-\beta/2}\sum_{k=1}^n \sum_{j=k-m}^{k-1} \mathbb{E}(|u_j|^\beta|\lambda_k|) \leq C n^{-\beta/2}\log n, \end{aligned}$$

as $0 < \beta \leq 1/3$. Hence $R_{22}(n) = o_P(1)$ due to $P(\Omega_K) \to 1$ as $K \to \infty$. To estimate $R_{21}(n)$, write

$$\begin{aligned} IR_1(n) &= \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^n U_k^*\left[\lambda_k - \mathbb{E}(\lambda_k \mid \mathcal{F}_{k-m-1})\right], \\ IR_2(n) &= \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^n U_k^*\,\mathbb{E}(\lambda_k \mid \mathcal{F}_{k-m-1}), \end{aligned}$$

where $U_k^* = f'(U_{n,k-m-1})I(\max_{1\leq j\leq k-m-1}|U_{n,j}| \leq K)$. It is readily seen from (6.12) and (6.13) that

$$\begin{aligned} \mathbb{E}IR_1^2(n) &\leq \frac{C}{n^2}\sum_{k=1}^n \mathbb{E}\big[\lambda_k - \mathbb{E}(\lambda_k \mid \mathcal{F}_{k-m-1})\big]^2 = O(n^{-1}), \\ \mathbb{E}|IR_2(n)| &\leq \frac{C}{n}\sum_{k=1}^n |\mathbb{E}(\lambda_k \mid \mathcal{F}_{k-m-1})| = o(1), \end{aligned}$$

which yields $IR_1(n) + IR_2(n) = o_P(1)$. We now have $R_{21}(n) = o_P(1)$ due to $P(\Omega_K) \to 1$ as $K \to \infty$, and the fact that, on Ω_k,

$$R_{21}(n) = \frac{1}{n\sigma_u\sigma_v}\sum_{k=1}^n U_k^*\,\lambda_k = IR_1(n) + IR_2(n) = o_P(1).$$

Combining these results proves $R_2(n) = o_P(1)$ and also completes the proof of (3.2). The proof of (3.3) is essentially the same and the details are omitted. □

Proof of (4.1). It suffices to identify the conditions of Theorem 2.2 with $x_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k}\epsilon_j$ and $y_{nk} = z_{nk}$. In fact, it is trivial to have **A2**. By the continuous mapping theorem and similar arguments to those in Wang, Lin and Gullati (2003), we have

$$\big(x_{n,\lfloor nt\rfloor}, y_{n,\lfloor nt\rfloor}\big) = \big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ z_{n,\lfloor nt\rfloor}\big) \Rightarrow \big(W(t), G(t)\big),$$

on $D_{\mathbb{R}^2}[0,1]$ in the Skorohod topology, which yields **A1**. Finally, due to the stationarity of z_j and $d_n^2 = \mathbb{E}|\sum_{k=1}^{n} z_k|^2 \sim c_\mu\, n^{3-2\mu}\, h^2(n)$ with $c_\mu = \frac{1}{(1-\mu)(3-2\mu)}\int_0^\infty x^{-\mu}(x+1)^{-\mu}dx$ [see, e.g., Wang, Lin and Gullati (2003)], we have

$$\sup_{i,j\geq 1}\frac{1}{j^2}\mathbb{E}|z_{n,i+j} - z_{n,i}|^2 = \frac{1}{d_n^2}\sup_{j\geq 1}\frac{1}{j^2}\mathbb{E}|\sum_{k=1}^{j} z_k|^2 = o(n^{-1}),$$

which yields (2.2). □

Proof of (4.2). It suffices to identify the conditions of Theorem 2.3 with $\epsilon_{nk} = \epsilon_k$ and $\eta_{n,k} = z_k$, $k = 1, ..., n$. This is straightforward and the details are omitted. □

Acknowledgements: Liang acknowledges research support from the National Natural Science Foundation of China (11271286) and the Specialized Research Fund for the Doctor Program of Higher Education (20120072110007). Phillips acknowledges support from the NSF under Grant No. SES 12-58258. Wang acknowledges research support from the Australian Research Council. Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia; e-mail:qiying@maths.usyd.edu.au.

REFERENCES

Aldous, D.J. and Eagleson, G.K.(1978). On mixing and stability of limit theorems, *Ann. Probab.* **2,** p325-331.

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

Chan, N. and Wang, Q. (2014). Nonlinear regression with nonstationary time series. *Journal of Econometrics*, forthcoming.

Chan, N. H. and Wei, C. Z. (1988). Limit distributions of least-squares estimates of unstable auto-regressive processes. *Annals of Statistics*, **16**, 367–401.

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

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

Cheng, T.-L. and Chow, Y.-S. (2002). On stable convergence in the central limit theorem. *Statistics and Probability Letters*, **57**, 307-313.

De Jong R, Davidson J. (2000 a). The functional central limit theorem and weak convergence to stochastic integral I: weak dependent processes. *Econometric Theory*, **16**, 621–642.

De Jong R, Davidson J. (2000 b). The functional central limit theorem and weak convergence to stochastic integral II: fractionally integrated processes. *Econometric Theory*, **16**, 643–666.

Duffie, D. and Protter, P. (1992). From discrete to continuous time finance: weak convergence of the financial gain process. *Mathematical finance*, **2**, 1–15.

de Jong, R. (2002). Nonlinear Regression with Integrated Regressors but Without Exogeneity, Mimeograph, Department of Economics, Michigan State University.

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

Hansen, B. E. (1992). Covergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, **8**, 489–500.

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

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

Jakubowski A. (1996). Convergence in various topologies for stochastic integrals driven by semimartingales. *Annals of Probability*, **24**, 2141–2153.

Jeganathan, P. (1995). "Some aspects of asymptotic theory with applications to time series models," *Econometric Theory* 11, 818-867.

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

Lin Z and Wang H. (2010). On Convergence to Stochastic Integrals. arXiv:1006.4693 [math.PR]

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of Probability,* **3**, 829-839.

Mykland, P. A. and Zhang L. (2006). ANOVA for diffusions and Ito processes". *Annals of Statistics*, 34, 1931–196

Park, J. Y., Phillips, P. C. B. (1988). "Statistical Inference in Regressions With Integrated Processes: Part 1," *Econometric Theory* 4, 468–497.

Park, J. Y. and Phillips, P. C. B. (1989). "Statistical Inference in Regressions With Integrated Processes: Part 2," *Econometric Theory* 5, 95-131.

Park, J.Y. and Phillips, P. C. B. (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, **15**, 269-298.

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

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

Phillips, P. C. B. (1987). "Time Series Regression with a Unit Root," *Econometrica,* 55, 277–302.

Phillips, P. C. B. (1988a). Multiple regression with integrated processes. In N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.

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

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (1991). "Optimal Inference in Cointegrated Systems," *Econometrica* 59, 283–306.

Phillips, P. C. B. and Hansen B. E. (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* 57, 99–125.

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

Phillips, PCB and Outliaris, S. (1990). Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, **58**, 165-193.

Rényi, A. (1963). On stable sequences of events, *Sankhya Ser. A,* **25**, 293-302.

Strasser, H. (1986). Martingale difference arrays and stochastic integrals. *Probabity Theory and Related Fields*, **72**, 83–89.

Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression, *Econometric Theory*, **30** , 509–535.

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. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, **27**, 235-259.

Wang, Q. and Phillips, P. C. B. (2012). A Specification Test for Nonlinear Nonstationary Models. *Annals of Statistics*, **40**, 727-758.