Martingale limit theorems revisited and non-linear cointegrating regression

Qiying Wang
The University of Sydney

January 19, 2011

Abstract

For a certain class of martingales, the convergence to mixture normal distribution is established under the convergence in distribution for the conditional variance. This is less restrictive in comparison with the classical martingale limit theorem where one generally requires the convergence in probability. The extension removes a main barrier in the applications of the classical martingale limit theorem to non-parametric estimates and inferences with non-stationarity, and essentially enhances the effectiveness of the classical martingale limit theorem as one of the main tools in the investigation of asymptotics in statistics, econometrics and other fields. The main result is applied to the investigations of asymptotics for the conventional kernel estimator in a nonlinear cointegrating regression, which essentially improves the existing works in literature.

Key words and phrases: Martingale, convergence in distribution, convergence in probability, cointegration, nonlinear functionals, nonparametric regression, kernel estimates.
AMS Classification: 60F05, 62E20.

1 Introduction

Let $\{M_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n\}$ be a zero-mean martingale array with the difference y_{ni}, and let M^2 be an a.s. finite random variable, where the σ-fields are nested, that is, $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for $1 \le i \le k_n, n \ge 1$. Suppose that, for all $\epsilon > 0$,

$$\sum_i E[y_{ni}^2 I(|y_{ni}| \ge \epsilon) \mid \mathcal{F}_{n,i-1}] \quad \to_P \quad 0, \tag{1.1}$$

and the conditional variance

$$\sum_i E[y_{ni}^2 \mid \mathcal{F}_{n,i-1}] \quad \to_P \quad M^2. \tag{1.2}$$

The classical martingale limit theorem (MLT) shows that $M_{n,k_n} = \sum_i y_{ni} \to_D Z$, where the r.v. Z has characteristic function $Ee^{itZ} = Ee^{-M^2t^2/2}$, $t \in R$. If M^2 is a constant, the nested structure of the σ-fields $\mathcal{F}_{n,i}$ is not necessary.

As one of the main conventional tools, the classical MLT is widely used in statistics, econometrics and other fields. In many applications, however, the convergence in probability in (1.2) seems to be too restrictive. To illustrate, consider a functional of the form

$$M_{n,i} = (nh^2)^{-1/4} \sum_{k=1}^{i} g(x_k/h)\epsilon_{k+1}, \quad i = 1, 2, ..., n, \tag{1.3}$$

where g is a real integrable function on R, ϵ_k is a stationary process, x_t is an integrated process (i.e., $I(1)$ process) such as a random walk, and $h \equiv h_n$ is a certain sequence of positive constants satisfying that $h \to 0$ and $nh^2 \to \infty$. Write $\mathcal{F}_k = \sigma(x_1, ..., x_k; \epsilon_1, ..., \epsilon_k)$. If $E(\epsilon_{k+1} \mid \mathcal{F}_k) = 0$ and $E(\epsilon_{k+1}^2 \mid \mathcal{F}_k) = 1$, $\{M_{ni}, \mathcal{F}_i\}_{i=1}^n$ forms a zero-mean martingale array with the conditional variance $V_n^2 = \frac{1}{\sqrt{nh}} \sum_{k=1}^n g^2(x_k/h)$. Unfortunately, the asymptotics of $M_{n,n}$ can not be obtained by the classical MLT. Indeed, under certain conditions on x_t such that $x_{[ns]}/\sqrt{n} \Rightarrow W(s)$ on $D[0,1]$, where $W(s)$ is a standard Brownian motion, we may prove

$$V_n^2 \quad \to_D \quad \int_{-\infty}^{\infty} g^2(x)dx\, L_W(1,0), \tag{1.4}$$

where $L_W(t,s)$ is the local time of the process $W(s)$ defined as in next section, but it is difficult or not possible to replace the convergence in distribution in (1.4) by the convergence in probability. See, e.g., Wang and Phillips (2009a) for instance. It should be pointed out that, given (1.4), one may enlarge the probability space in which $x_1, ..., x_t$ are equipped so that under this new space

$$V_n^2 \quad \to_P \quad \int_{-\infty}^{\infty} g^2(x)dx\, L_W(1,0). \tag{1.5}$$

This fact can not be used to establish the asymptotics of $M_{n,n}$ as well since, without independence between ϵ_t and x_t, this enlarged probability space may destroy the nested σ-fields structure (i.e., $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$) required in the classical MLT.

Such functionals $M_{n,n}$ having a martingale structure commonly arise in non-linear regression with integrated time series [Park and Phillips (1999, 2001)], non-parametric estimation in relation to nonlinear cointegration models [Phillips and Park (1998), Karlsen and Tjostheim (2001), Karlsen, et al. (2007), and Wang and Phillips (2009a, 2009b, 2010)] and many other related research areas [Marmer (2008), Gao, et.al (2009a, 2009b), Choi and Saikkonen (2010), Kasparis (2008, 2010), Kasparis and Phillips (2009)]. In such cases, g may be a kernel function K or a squared kernel function K^2, and the sequence h is the bandwidth used in the nonparametric regression. As illustrated above, the classical MLT is limited to establish the asymptotics of $M_{n,n}$, which, as is well-known, plays a key rule in the development of non-parametric estimates and inferences with non-stationarity.

The main aim of this paper is to provide an extension of the classical MLT. It is shown that, for a certain class of martingales such those having a shape as in (1.3), the condition (1.2), that is, the convergence in probability for the conditional variance in the classical MLT, can be reduced to less restrictive:

$$\sum_i E[y_{ni}^2 \mid \mathcal{F}_{n,i-1}] \quad \to_D \quad M^2. \tag{1.6}$$

As discussed above, this kind of extensions removes a main barrier in the applications of the classical MLT to non-parametric estimates and inferences with non-stationarity. It is expected that our results will essentially enhance the effectiveness of the classical MLT as one of the main tools in the investigation of asymptotics in statistics, econometrics and other fields. We finally remark that there are little useful works in literature to establish a central limit theorem for martingale under less restrictive condition (1.6). Theorem 3.4 of Hall and Heyde (1980) provided a partial result in this direction, but the result can not be used to our specified class of martingales as defined in next section. Park and Phillips (2001) considered a class of martingales having the shape as in (1.3), but imposed some extra conditions. More currently a unpublished manuscript by Jeganathan (2006) discussed convergence in distribution of row sum process to mixture of additive processes. When specified to the martingales defined as in (2.1) below, the conditions used by Jeganathan (2006) are quite complicated and also hard to identify. Unlike these existing works, our conditions are neat and the identification on the conditions is quite straightforward.

This paper is organized as follows. In next section, we present our main results. Theorem 2.1 provides a framework in an extension to the classical MLT. It is shown that, for the specified class of martingales defined as in (2.1) below, a minor and easy identified additional condition, together with the convergence in distribution for the conditional variance, is sufficient to establish the convergence to mixture normal distribution. It is interesting to notice that, general speaking, it is not possible to replace the convergence in probability for the conditional variance by the convergence in distribution without any additional conditions. This kind of additional condition might be close to necessary to investigate the asymptotics for the specified martingale defined as in (2.1). Theorem 2.2 gives an important application of Theorem 2.1 to general linear process, where we investigate the asymptotics for the functionals having shape as in (1.3). Our result includes the x_k being a partial sum of ARMA process and fractionally integrated processes, which are most common used in practice. Using Theorem 2.2 as a main tool, Theorems 2.3 and 2.4 investigate the asymptotics for the conventional kernel estimators in a non-linear cointegrating regression model. These results significantly improve those in existing literature. This section also presents several remarks for possible further extensions of our results. All technical proofs are given in Section 3.

Throughout the paper, we denote by $C, C_1, ...$ the constants, which may changes at each appearance. We also use the following definitions and notation. If $\alpha_n^{(1)}$, $\alpha_n^{(2)}$,..., $\alpha_n^{(k)}$ $(1 \le n \le \infty)$ are random elements of $D[0,1]$, we will understand the condition

$$(\alpha_n^{(1)}, \alpha_n^{(2)}, ..., \alpha_n^{(k)}) \to_D (\alpha_\infty^{(1)}, \alpha_\infty^{(2)}, ..., \alpha_\infty^{(k)})$$

to mean that for all $\alpha_\infty^{(1)}$, $\alpha_\infty^{(2)}$,..., $\alpha_\infty^{(k)}$-continuity sets A_1, A_2,...,A_k

$$P\big(\alpha_n^{(1)} \in A_1, \alpha_n^{(2)} \in A_1, ..., \alpha_n^{(k)} \in A_k\big) \to P\big(\alpha_\infty^{(1)} \in A_1, \alpha_\infty^{(2)} \in A_2, ..., \alpha_\infty^{(k)} \in A_k\big).$$

[see Billingsley (1968, Theorem 3.1) or Hall (1977)]. $D[0,1]^k$ will be used to denote $D[0,1] \times ... \times D[0,1]$, the k-times coordinate product space of $D[0,1]$. We still use $\Rightarrow$ to denote weak convergence on $D[0,1]$.

2 Main results

Motivated by non-parametric estimates and inferences with non-stationarity, this section considers a class of statistics defined by

$$S_n = \sum_{k=1}^{n} x_{k,n} \epsilon_{n,k+1}, \tag{2.1}$$

where $\{\epsilon_{n,k+1}, \mathcal{F}_{n,k}\}_{1 \le k \le n}$ forms a martingale difference and $x_{k,n}$ is adapted to $\mathcal{F}_{n,k}$ for $1 \le k \le n, n \ge 1$. Let $f_n(...)$ be a real function of its components, and $\eta_{n,k}$ and ξ_{nj} be two sequences of random variables. We specify the $x_{k,n}$ to have a form:

$$x_{k,n} = f_n\big(\epsilon_{n,1}, ..., \epsilon_{n,k};\ \eta_{n,1}, ..., \eta_{n,k};\ \xi_{n1}, \xi_{n2}, ...\big),$$

and the $\mathcal{F}_{n,k}$ to be an increasing σ-fields for each $n \geq 1$. We do not require the nested structure of $\mathcal{F}_{n,k}$ in this paper. The increase of $\mathcal{F}_{n,k}$ on k and the structure of $x_{k,n}$ are quite natural in many applications.

In order to investigate the asymptotics of S_n, we make use of the following assumptions.

Assumption 1. $\{\eta_{n,k+1}, \epsilon_{n,k+1}, \mathcal{F}_{n,k}\}_{1\leq k\leq n}$ *forms a martingale difference, where* $\eta_{n,k}$ *and* $\epsilon_{n,k}$ *satisfy that, as* $n \to \infty$ *first and then* $m \to \infty$

$$\max_{m\leq k\leq n} |E(\eta_{n,k+1}^2 \mid \mathcal{F}_{n,k}) - 1| \to 0, \quad \max_{m\leq k\leq n} |E(\epsilon_{n,k+1}^2 \mid \mathcal{F}_{n,k}) - 1| \to 0, \qquad a.s.$$

and for some $\delta > 0$,

$$\max_{\substack{1\leq k\leq n\\ n\geq 1}} \left[E(|\eta_{n,k+1}|^{2+\delta} \mid \mathcal{F}_{n,k}) + E(|\epsilon_{n,k+1}|^{2+\delta} \mid \mathcal{F}_{n,k})\right] < \infty.$$

By virtue of Assumption 1, it is readily seen that, for each $0 < s \leq 1$ and all $\epsilon > 0$, $\frac{1}{n}\sum_{k=1}^{[ns]} E(\eta_{n,k+1}^2 \mid \mathcal{F}_{n,k}) \to_{a.s.} s$ and $\frac{1}{n}\sum_{k=1}^{n} E(\eta_{n,k+1}^2 I(|\eta_{n,k+1}| \geq \epsilon\sqrt{n}) \mid \mathcal{F}_{n,k}) \to_P 0$. The classical martingale invariance principle implies that

$$W_n(t) \ := \ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \eta_{n,j+1} \ \Rightarrow \ W(t), \tag{2.2}$$

on $D[0,1]$, where $W(t)$ is a standard Winner process. See, e.g., McLeish (1974) for instance. Our next assumption on $x_{k,n}$ depends on $\eta_{n,k}$ as well as $W(t)$.

Assumption 2. *(i)* $\xi_{nj}, j \geq 1$, *are* $\mathcal{F}_{n,1}$*-measurable for each* $n \geq 1$, *and there exist a sequence of constants* $0 < d_n \to \infty$ *and a Gaussian process* $G(t)$, *which is independent of* $W(t)$, *such that*

$$\Xi_n(t) := \frac{1}{d_n}\sum_{j=1}^{[nt]} \xi_{nj} \ \Rightarrow \ G(t), \qquad on\ D[0,\infty). \tag{2.3}$$

(ii) $\max_{1\leq k\leq n} |x_{k,n}| = o_P(1)$ *and*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} |x_{k,n}|\, |E(\eta_{n,k+1}\epsilon_{n,k+1} \mid \mathcal{F}_{n,k})| \ = \ o_P(1). \tag{2.4}$$

(iii) There exists a positive functional $g^2(W,G)$ *of* $W(s), 0 \leq s \leq 1$ *and* $G(s), 0 \leq s < \infty$, *such that*

$$G_n^2 := \sum_{k=1}^{n} x_{k,n}^2 \ \to_D \ g^2(W,G). \tag{2.5}$$

We have the following main result.

THEOREM 2.1. *Under Assumptions 1-2, we have*

$$\{S_n,\ G_n^2\} \ \to_D \ \{g(W,G)\,N,\ g^2(W,G)\}, \tag{2.6}$$

where N is a standard normal variate independent of $g(W,G)$. Consequently,

$$S_n/G_n \to_D N(0,1). \tag{2.7}$$

Remark 1. *If the condition (2.5) can be strengthened to convergence in probability without changing the probability space and the σ-fields $\mathcal{F}_{n,k}$ is nested, i.e., $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for $1 \le i \le n$, Assumption 2(i), (2.4) and the part related to $\eta_{n,k}$ in Assumption 1 are not necessary. In this situation, Theorem 2.1 can be easily established by the classical MLT as stated in Section 1. Unfortunately, it is generally difficult to replace (2.5) by $G_n^2 \to_P g^2(W,G)$, except that $g^2(W,G)$ is a constant. Our result hence provides a useful extension to the classical MLT.*

Remark 2. *If $g^2(W,G)$ does not depend on $G(s), 0 \le s < \infty$, Assumption 2(i) related to $\xi_{n,j}$ is not necessary. This point is important since, in many applications, we may have $x_{k,n} = f_n(M_n)$ with $M_n = \sum_{k=1}^n \eta_{n,k} + \xi_n$, where ξ_n is negligible comparing to M_n (that is, M_n itself is not a martingale, but may be approached by a martingale. In regard to the martingale approaches of stable random variables sum, we refer to Wu and Woodroofe (2004) for further details). Theorem 2.1 can be used to this situation without Assumption 2(i) if the $\eta_{n,k}$ and $x_{k,n}$ satisfy the related conditions.*

Remark 3. *If $\eta_{n,k}$ is independent of $\epsilon_{n,k}$, the condition (2.4) holds true automatically. A simple sufficient condition to give (2.4) is $\frac{1}{\sqrt{n}}\sum_{k=1}^n |x_{k,n}| = o_P(1)$, if only $\max_{\substack{1\le k\le n\\ n\ge 1}} \big[E(|\eta_{n,k+1}|^2 \mid \mathcal{F}_{n,k}) + E(|\epsilon_{n,k+1}|^2 \mid \mathcal{F}_{n,k})\big] < \infty$. This condition can not be removed except the condition (2.5) can be strengthened to convergence in probability without changing the probability space, as explained in Remark 1. A simple example is given as follows. By letting $\eta_k = \epsilon_k, k \ge 1$, be a sequence of iid $N(0,1)$, $x_{k,n} = n^{-3/2}(\sum_{j=1}^k \epsilon_j)^2$ and $\mathcal{F}_k = \sigma(\epsilon_1, ..., \epsilon_k)$, we have that $\sum_{k=1}^n x_{k,n}^2 \to_D \int_0^1 W^4(t)dt$ and*

$$S_n = \sum_{k=1}^n x_{k,n}\epsilon_{k+1} \to_D \int_0^1 W^2(t)dW(t),$$

where $W(t)$ is a standard Winner process. The result (2.5) in Theorem 2.1 is not true for this example, since $\int_0^1 W^2(t)dW(t)$ is not equal in distribution to $\big(\int_0^1 W^4(t)dt\big)^{1/2} N$. On the other hand, the condition (2.4) can not be satisfied as

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n |x_{k,n}|\,|E(\epsilon_{k+1}^2 \mid \mathcal{F}_k)| = \frac{1}{n^2}\sum_{k=1}^n (\sum_{j=1}^k \epsilon_j)^2 \to_D \int_0^1 W^2(t)dt.$$

In what follows we consider an application of Theorem 2.1 to general linear processes. Let $\{\xi_j, j \ge 1\}$ be a linear process defined by

$$\xi_j = \sum_{k=0}^\infty \phi_k\, \nu_{j-k}, \tag{2.8}$$

where $\{\nu_j, -\infty < j < \infty\}$ is a sequence of iid random variables with $E\nu_0 = 0$, $E\nu_0^2 = 1$, $E|\nu_0|^{2+\delta} < \infty$ for some $\delta > 0$ and characteristic function $\varphi(t)$ of ν_0 satisfying $\int_{-\infty}^\infty |\varphi(t)|dt < \infty$. Throughout the section, the coefficients $\phi_k, k \ge 0$, are assumed to satisfy one of the following conditions:

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

C2. $\sum_{k=0}^\infty |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^\infty \phi_k \neq 0$.

Put $y_k = \sum_{j=1}^k \xi_j$ and $y_{k,n} = y_k/d_n + c_n'$, where $c_n' \to 0$ and $d_n^2 = Ey_n^2$. It is well-known that, with $c_\mu = \frac{1}{(1-\mu)(3-2\mu)}\int_0^\infty x^{-\mu}(x+1)^{-\mu}dx$,

$$d_n^2 = Ey_n^2 \sim \begin{cases} c_\mu\, n^{3-2\mu}\,\rho^2(n), & \text{under } \mathbf{C1}, \\ \phi^2\, n, & \text{under } \mathbf{C2}. \end{cases} \tag{2.9}$$

See, e.g., Wang, Lin and Gulati (2003) for instance. We consider the limit behavior of sample functions of the form:

$$S_{1n} \;=\; \big(\frac{c_n}{n}\big)^{1/2}\sum_{k=1}^n f\big(c_n\, y_{k,n}\big)\epsilon_{k+1}, \tag{2.10}$$

when $c_n \to \infty$ and $c_n/n \to 0$, under the following conditions:

Assumption 3. *f is a real function on R satisfying $\int_{-\infty}^\infty \big[|f(x)| + |f(x)|^{4+\gamma}\big]dx < \infty$ for some $\gamma > 0$.*

Assumption 4. *(i) $\mathcal{F}_k$ is a sequence of increasing σ-fields such that $\nu_k \in \mathcal{F}_k$ and ν_{k+1} is independent of $\mathcal{F}_k$ for all $k \ge 1$, and $\nu_k \in \mathcal{F}_1$ for all $k \le 0$. (ii) $\{\epsilon_k, \mathcal{F}_k\}_{k\ge1}$ forms a martingale difference satisfying, as $m \to \infty$, $\max_{k\ge m}|E(\epsilon_{k+1}^2 \mid \mathcal{F}_k) - 1| \to 0, a.s.$ and for some $\delta > 0$, $\max_{k\ge1} E(|\epsilon_{k+1}|^{2+\delta} \mid \mathcal{F}_k) < \infty$.*

Assumption 4 is a specified version of Assumption 1 by making use of the independence between ν_k. Typically, in applications, we may choose $\mathcal{F}_k = \sigma\{\nu_k, \nu_{k-1}, ...\}$ or $\mathcal{F}_k = \sigma\{\epsilon_k, ..., \epsilon_1; \nu_k, \nu_{k-1}, ...\}$ together with the condition that ν_{k+1} is independent of $\mathcal{F}_k$ for all $k \geq 1$. To investigate the asymptotics of S_{1n}, we start with the following notation. A fractional Brownian motion with $0 < \beta < 1$ on $D[0,1]$ is defined by

$$W_\beta(t) = \frac{1}{A(\beta)} \int_{-\infty}^0 \Big[(t-s)^{\beta-1/2} - (-s)^{\beta-1/2}\Big] dW^*(-s) + \int_0^t (t-s)^{\beta-1/2} dW(s),$$

where

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

$W(s), 0 \leq s < \infty$ is a standard Brownian motion, and $W^*(u), 0 \leq u < \infty$ is an independent copy of $W(s), 0 \leq s < \infty$. It is readily seen that $W_{1/2}(t) = W(t)$ and $W_\beta(t)$ has a continuous local time $L_{W_\beta}(t,s)$ with regard to (t,s) in $[0,\infty) \times R$. See, e.g., Theorem 22.1 of Geman and Horowitz (1980). Here and below, the process $\{L_\zeta(t,s), t \geq 0, s \in R\}$ is said to be the local time of a measurable process $\{\zeta(t), t \geq 0\}$ if, for any locally integrable function $T(x)$,

$$\int_0^t T[\zeta(s)]ds = \int_{-\infty}^\infty T(s) L_\zeta(t,s) ds, \quad \text{all } t \in R,$$

with probability one.

Note that $L_{W_\beta}(1,0)$ is a functional of $W(s), 0 \leq s \leq 1$ and $W^*(u), 0 \leq u < \infty$. Write $G_{1n}^2 = \frac{c_n}{n} \sum_{k=1}^n f^2(c_n\, y_{k,n})$. As a direct consequence of Theorem 2.1, we have the following theorem.

THEOREM 2.2. *Under Assumptions 3 and 4, when $c_n \to \infty$ and $c_n/n \to 0$, we have*

$$\big\{S_{1n},\ G_{1n}^2\big\} \ \to_D\ \big\{g(W,W^*)\,N,\ g^2(W,W^*)\big\}, \tag{2.11}$$

where N is a standard normal variate independent of $g^2(W,W^*) = \int_{-\infty}^\infty f^2(x)dx\, \mathcal{L}_W$ with $\mathcal{L}_W$ being defined by

$$\mathcal{L}_W = \begin{cases} L_{W_{3/2-u}}(1,0), & \textit{under } \boldsymbol{C1}, \\ L_W(1,0), & \textit{under } \boldsymbol{C2}. \end{cases}$$

Consequently,

$$S_{1n}/G_{1n} \to_D N(0,1). \tag{2.12}$$

Remark 4. *The result (2.11) under **C1** is new, in which the y_n includes the fractionally integrated process as an example. Theorem 3.2 of Park and Phillips (2001) investigated the asymptotics of S_{1n} under **C2**, but only allows the c_n to be equal $Ey_n^2 = \phi n$*

*and imposes strong restrictions on f and ν_0. Using similar ideas as in Park and Phillips (2001), Wang and Phillips (2009a) considered more general situations under **C2** (the result is included in the proof of their Theorem 3.2), but still imposes some additional conditions. The result (2.11) under **C2** essentially improves these existing results. It should be mentioned that, unlike these existing results, our proof is quite neat and straightforward by our extended martingale limit theorem, given in Theorem 2.1.*

Remark 5. *Under the conditions of Theorem 2.2, one always has $cov(y_t, \epsilon_{t+1}) = E[y_t E(\epsilon_{t+1} \mid \mathcal{F}_t)] = 0$. Wang and Phillips (2009b) investigated the asymptotics of S_{1n} under **C2**, imposing the ϵ_{t+1} to be serially dependent and cross correlated with y_s for $|t-s| < m_0$, where m_0 is a given constant. As a consequence, in Wang and Phillips (2009b), one may have $cov(y_t, \epsilon_{t+1}) \neq 0$. This makes the functionals in Wang and Phillips (2009b) essentially different from the one investigated in current Theorem 2.2. However, Wang and Phillips (2009b) does not cover the result (2.11) under **C2** as a special case, since in (2.11) we allow ϵ_{k+1} is cross correlated with y_s for $1 \leq s \leq k$.*

Remark 6. *It is interesting to notice that the result would be quite different if $c_n = 1$. Indeed, under much weaker conditions on $f, y_{k,n}$ and ϵ_k which adds to*

$$\Big\{y_{[nt],n}, \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_{j+1}\Big\} \Rightarrow \{U(t), V(t)\}$$

on $D[0,1]^2$, one may prove $\frac{1}{\sqrt{n}} \sum_{k=1}^n f\big(y_{k,n}\big)\epsilon_{k+1} \to_D \int_0^1 f[U(s)]dV(s)$, a stochastic integral. There are many papers which investigate the convergence to stochastic integrals. We only cite Kurtz and Protter (1991) and Hansen (1992) for reference.

Remark 7. *The individual asymptotics of the functionals like G_{1n}^2 has been investigated in Wang and Phillips (2009a) [also see Wang and Phillips (2010) for zero energy situation] under weaker condition than Assumption 3. Explicitly, it follows from (2.5) of Wang and Phillips (2010) [1] that, for each fixed $x \in R$ and all h satisfying $h \to 0$ ($h^2 \log n \to 0$ under **C2**) and $d_n/nh \to 0$,*

$$\begin{aligned}
&\Big\{\Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{k=1}^n g\big[(y_k - x)/h\big],\ \frac{d_n}{nh} \sum_{k=1}^n f_1\big[(y_k - x)/h\big],\ \frac{d_n}{nh} \sum_{k=1}^n f_2\big[(y_k - x)/h\big]\Big\} \\
&\qquad \to_D \ \Big\{\tau_0\, N\, \mathcal{L}_W^{1/2},\ \tau_1\, \mathcal{L}_W,\ \tau_2\, \mathcal{L}_W\Big\}, \qquad (2.13)
\end{aligned}$$

where $d_n^2 = Ey_n^2$ is given as in (2.9), $\tau_0^2 = \int_{-\infty}^{\infty} g^2(x)dx$, $\tau_1 = \int f_1(x)dx$ and $\tau_2 = \int f_2(x)dx$, provided that $\int |g(t)|dt < \infty$, $\int |\hat{g}(t)|dt < \infty$ and $|\hat{g}(t)| \leq C \min\{|t|, 1\}$, where

[1] The result is only for $x = 0$, but it is not difficult to see that the result holds true for any fixed $x \in R$.

$\hat{g}(x) = \int e^{itx} g(t)dt$; $|f_j(x)|$ *and* $f_j^2(x)$, $j = 1, 2$, *are Lebesgue integrable functions on* R *with* $\tau_1 \neq 0$ *and* $\tau_2 \neq 0$. *It is readily from (2.13) that*

$$\frac{\left(\frac{d_n}{nh}\right)^{-1/2} \sum_{k=1}^{n} g[(y_k - x)/h]}{\sum_{k=1}^{n} f_2[(y_k - x)/h]} \quad \to_D \quad \frac{\tau_0}{\tau_2} N \, \mathcal{L}_W^{-1/2}, \tag{2.14}$$

$$\frac{\sum_{k=1}^{n} f_1[(y_k - x)/h]}{\sum_{k=1}^{n} f_2[(y_k - x)/h]} \quad \to_P \quad \frac{\tau_1}{\tau_2}. \tag{2.15}$$

These results will be used in the proofs of Theorems 2.2-2.4.

As stated in Introduction, the results given in Theorem 2.2 play a key rule in the investigation of non-stationary cointegration regression. To illustrate, consider the following nonlinear cointegrating regression model

$$z_t = m(y_t) + \epsilon_{t+1}, \quad t = 1, 2, ..., n, \tag{2.16}$$

where ϵ_{t+1} is a zero mean stationary equilibrium error and m is an unknown function to be estimated with the observed data $\{z_t, y_t\}_{t=1}^n$. The conventional kernel estimator of $m(x)$ in model (2.16) is given by

$$\hat{m}(x) \quad = \quad \frac{\sum_{t=1}^n z_t K_h(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)}, \tag{2.17}$$

where $K_h(s) = \frac{1}{h} K(s/h)$, $K(x)$ is a nonnegative real function, and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. Note that $\hat{m}(x)$ has the usual decomposition:

$$\hat{m}(x) - m(x) \quad = \quad \frac{\sum_{t=1}^n \epsilon_{t+1} K_h(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)} + \frac{\sum_{t=1}^n \left[m(y_t) - m(x)\right] K_h(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)}, \tag{2.18}$$

and it is the first term to determine the asymptotic distribution, depending on the asymptotics of

$$(S_{2n}, G_{2n}^2) := \Big\{ \big(\frac{c_n}{n}\big)^{1/2} \sum_{t=1}^{n} \epsilon_{t+1} K(c_n \, y_{t,n}), \quad \frac{c_n}{n} \sum_{t=1}^{n} K^2(c_n \, y_{t,n}) \Big\}, \tag{2.19}$$

where $y_{t,n} = (y_t - x)/d_n$, $d_n^2 = E y_n^2$ and $c_n = d_n/h$. An application of Theorem 2.2 to (S_{2n}, G_{2n}^2) essentially improves the existing results on the asymptotics of the kernel estimator $\hat{m}(x)$, which was currently investigated in Wang and Phillips (2009a, 2010). To do this, we require the following additional assumptions.

Assumption 5. $y_t = \sum_{j=1}^t \xi_j$, *where* ξ_j *is defined as in (2.8) with* ϕ_k *satisfying* ***C1*** *or* ***C2***.

Assumption 6. *The kernel* K *satisfies that* $\int_{-\infty}^{\infty} K(s)ds = 1$ *and* $\sup_s K(s) < \infty$.

Assumption 7. *For given x, there exists a real function $m_1(s,x)$ and an $0<\gamma\le 1$ such that, when h sufficiently small, $|m(hy+x)-m(x)|\le h^\gamma\, m_1(y,x)$ for all $y\in R$ and $\int_{-\infty}^{\infty} K(s)\,[m_1(s,x)+m_1^2(s,x)]ds<\ \infty$.*

THEOREM 2.3. *Under Assumptions 4-7, for any h satisfying $nh/d_n\to\infty$ and $nh^{1+2\gamma}/d_n\to 0$, we have*

$$\Big(\frac{d_n}{nh}\Big)^{-1/2}(\hat{m}(x)-m(x))\quad\to_D\quad \tau\, N\,\mathcal{L}_W^{-1/2} \tag{2.20}$$

where $d_n^2=Ey_n^2$, given as in (2.9), $\tau^2=\int_{-\infty}^{\infty}K^2(s)ds$ and N is a standard normal variate independent of $\mathcal{L}_W$, which is given as in Theorem 2.2. We also have

$$\Big(h\sum_{t=1}^{n}K_h(y_t-x)\Big)^{1/2}(\hat{m}(x)-m(x))\quad\to_D\quad\tau\, N. \tag{2.21}$$

If Assumptions 6 and 7 are strengthened to the following Assumptions 6^* and 7^*, an explicit bias term may be incorporated into the limit theory (2.20) and (2.21).

Assumption 6^*. *(i) $K(x)$ satisfies that $\int K(y)dy=1$ and for some $p\ge 2$,*

$$\int y^pK(y)dy\neq 0,\qquad \int y^iK(y)dy=0,\quad i=1,2,...,p-1.$$

(ii) $K(x)$ has a compact support and is twice continuous differentiable on R.

Assumption 7^*. *For given fixed x, $m(x)$ has a continuous $p+1$ derivative in a small neighborhood of x, where $p\ge 2$ is defined as in Assumption 6^*.*

THEOREM 2.4. *Under Assumptions 4-5 and 6^*-7^*, for any h satisfying $nh/d_n\to\infty$ and $nh^{1+2(p+1)}/d_n\to 0$, we have*

$$\Big(\frac{d_n}{nh}\Big)^{-1/2}\left[\hat{m}(x)-m\,(x)-\frac{h^pm^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^pK(y)dy\right]\ \to_D\ \tau\, N\,\mathcal{L}_W^{-1/2}, \tag{2.22}$$

and also,

$$\Big(h\sum_{t=1}^{n}K_h(y_t-x)\Big)^{1/2}\left[\hat{m}(x)-m\,(x)-\frac{h^pm^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^pK(y)dy\right]\ \to_D\ \tau\, N\,. \tag{2.23}$$

Remark 8. *In past decade there have been increasing interests in the investigation of asymptotics for the kernel estimator $\hat{m}(x)$ of $m(x)$ in the model (2.16), under different setting on the regressor y_t and the error process ϵ_t. Phillips and Park (1998) studied* *nonparametric autoregression in the context of a random walk. Karlsen and Thostheim (2001) and Guerre (2004) studied nonparametric estimation for certain nonstationary processes in the framework of recurrent Markov chains. Karlsen, et al. (2007) developed an asymptotic theory for nonparametric estimation of a time series regression equation involving stochastically nonstationary time series. Under similar conditions and using related Markov chain methods, Schienle (2008) investigated additive nonlinear versions of (2.16) and obtained a limit theory for nonparametric regressions under smooth backfitting.*

More recently, Wang and Phillips (2009a) and Cai, et al. (2009) considered an alternative treatment by making use of local time limit theory and, instead of recurrent Markov chains, worked with partial sum representations y_t of linear process, as given in Assumption 5. Unlike these cited results where the independence between the regressor y_t and the error process ϵ_t is essentially imposed, our results allow the endogenous regressor case, and also allow the regressor y_t to be fractionally integrated processes. Furthermore Theorem 2.4 investigates the bias analysis, which does not attend in the previous articles. In another papers, Wang and Phillips (2009b, 2010) considered the errors u_t to be serially dependent and cross correlated with the regressor x_t for small lags. As stated in Remark 5, these works have different error structure as used in current Theorems 2.3 and 2.4.

Remark 9. *Many interesting facts are raised in the nonlinear cointegrating regression. For instance, a possible "optimal" bandwidth h which yields the best rate $\hat{m}(x) - m(x)$ or the minimal $E(\hat{m}(x) - m(x))^2$ is different from non-parametric regression with a stationary regressor; the use of augmented regression, as is common in linear cointegration modeling to address endogeneity, does not lead to bias deduction in nonparametric regression, except there is an asymptotic gain in variance deduction; the particular advantage that the local linear nonparametric estimator has bias reducing properties in comparison with the Nadaraya-Watson estimator $\hat{m}(x)$ defined as in (2.17) is lost when the regressor y_t is nonstationary. For more detailed discussions of these interesting facts, we refer to Wang and Phillips (2009a, 2009b, 2010).*

3 Proofs of main results

3.1. Proof of Theorem 2.1. For convenience of notation, we omit the index n on $\eta_{n,k}$, $\epsilon_{n,k}$ and $\xi_{n,k}$ in the following proof. It suffices to show that, for any $\alpha, \beta \in R$

$$\lim_{n\to\infty} E e^{i\alpha S_n + i\beta G_n^2} = E e^{(-\frac{\alpha^2}{2}+i\beta)\sigma^2}, \tag{3.1}$$

where $\sigma^2 = g^2(W,G)$. Let $\lambda > 0$ be a continuous point of σ^2. Write

$$x^*_{k,n} = x_{k,n} I(\sum_{j=1}^{k} x^2_{j,n} \le \lambda), \quad S^*_n = \sum_{k=1}^{n} x^*_{k,n}\epsilon_{k+1}$$

and $\sigma^2_\lambda = \sigma^2 I(\sigma^2 \le \lambda) + \lambda I(\sigma^2 > \lambda)$. It is readily seen that

$$\max_{1\le k\le n} |x^*_{k,n}| \le \max_{1\le k\le n} |x_{k,n}| = o_P(1) \tag{3.2}$$

and because of (2.4),

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x^*_{k,n}|\, |E(\eta_{n,k+1}\epsilon_{n,k+1} \mid \mathcal{F}_{n,k})| \\ \le \; & \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{k,n}|\, |E(\eta_{n,k+1}\epsilon_{n,k+1} \mid \mathcal{F}_{n,k})| = o_P(1). \end{aligned} \tag{3.3}$$

Since λ is a continuous point of σ^2, the continuous mapping theorem shows that

$$\widetilde{G}^2_n \;:=\; G^2_n I(G^2_n \le \lambda) + \lambda\, I(G^2_n > \lambda) \to_D \sigma^2_\lambda.$$

Now, by noting $\widetilde{G}^2_n - \max_{1\le k\le n} x^2_{k,n} \le \sum_{k=1}^n x^{*2}_{k,n} \le \widetilde{G}^2_n$, we have

$$G^{*2}_n := \sum_{k=1}^{n} x^{*2}_{k,n} \quad \to_D \quad \sigma^2_\lambda. \tag{3.4}$$

Combining all above facts, the result (3.1) will follow if we prove, for all $\alpha, \beta \in R$,

$$\lim_{n\to\infty} Ee^{i\alpha\, S^*_n + i\beta G^{*2}_n} \quad = \quad Ee^{(-\frac{\alpha^2}{2} + i\beta)\sigma^2_\lambda}, \tag{3.5}$$

Indeed, this claim follows from the facts that $S^*_n = S_n$ and $G^{*2}_n = G^2_n$ on the set $G^2_n \le \lambda$, and $\lim_{\lambda\to\infty} P(G^2_n > \lambda) = 0$.

Recall that $ES^{*2}_n \le C\, EG^{*2}_n \le 2C\,\lambda$. $\{S^*_n, G^{*2}_n\}_{n\ge 1}$ is tight on $D[0,1]^2$. Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

$$\{S^*_{n''}, G^{*2}_{n''}\} \quad \to_D \quad (S, \sigma^2_\lambda), \tag{3.6}$$

where S is a limit random variable of $S^*_{n''}$. For $\beta_{1k}, \beta_{2k} \in R$, $0 = u_{10} < u_{11} < ... < u_{1N_1} = 1$ and $0 = u_{20} < u_{21} < ... < u_{2N_2} < \infty$, define

$$V_1 \quad = \quad \sum_{k=1}^{N_1} \beta_{1k}[W(u_{1k}) - W(u_{1,k-1})],$$

and

$$V_2 \;=\; \sum_{k=1}^{N_2} \beta_{2k}[G(u_{2k}) - G(u_{2,k-1})].$$

By virtue of (3.6) and recalling that σ_λ^2 is $\mathcal{F} = \sigma\{W(s), 0 \le s \le 1; G(t), 0 \le t < \infty\}$-measurable, to establish (3.5), it suffices to show that, for all $\alpha, \beta, \beta_{1k}, \beta_{2k} \in R$, $0 = u_{10} < u_{11} < ... < u_{1N_1} = 1$ and $0 = u_{20} < u_{21} < ... < u_{2N_2} < \infty$,

$$Ee^{iV_1 + iV_2 + i\alpha S + \frac{1}{2}\alpha^2\sigma_\lambda^2} \;=\; Ee^{iV_1 + iV_2}. \tag{3.7}$$

Indeed, the result (3.7) implies that

$$E(e^{i\alpha S + i\beta\sigma_\lambda^2} \mid \mathcal{F}) \;=\; e^{(-\frac{1}{2}\alpha^2 + i\beta)\sigma_\lambda^2}.$$

This, together with (3.6), yields that, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

$$\lim_{n'' \to \infty} Ee^{i\alpha\, S^*_{n''} + i\beta G^{*2}_{n''}} \;=\; Ee^{i\alpha S + i\beta\sigma_\lambda^2} = Ee^{(-\frac{\alpha^2}{2} + i\beta)\sigma_\lambda^2}.$$

The result (3.1) hence follows because the limitation does not depend on the choice of the subsequence.

We next prove (3.7). To this end, write

$$V_{1n} \;=\; \sum_{k=1}^{N_1} \beta_{1k}[W_n(u_{1k}) - W_n(u_{1,k-1})] \;:=\; \sum_{k=1}^{n} \beta^*_{1k}\, \eta_{k+1},$$

where $\beta^*_{1k} = \beta_{1j}/\sqrt{n}$ when $[nu_{1,j-1}] < k \le [nu_{1j}]$, for $j = 1, ..., N_1$;

$$V_{2n} \;=\; \sum_{k=1}^{N_2} \beta_{2k}[\Xi_n(u_{2k}) - \Xi_n(u_{2,k-1})] \;:=\; \sum_{k=1}^{n} \beta^*_{2k}\, \xi_k,$$

where $\beta^*_{2k} = \beta_{2j}/d_n$ when $[nu_{2,j-1}] < k \le [nu_{2j}]$, for $j = 1, ..., N_2$; and

$$\Gamma_n \;=\; \sum_{k=1}^{n} E\big\{(e^{i\beta^*_{1k}\,\eta_{k+1} + i\alpha\, x^*_{k,n}\,\epsilon_{k+1}} - 1) \mid \mathcal{F}_{n,k}\big\}$$

We need the following two propositions. Their proofs will be given in Sections 3.5 and 3.6.

PROPOSITION 3.1. *For any $\alpha, \beta_{1k} \in R$ and $0 = u_{10} < u_{11} < ... < u_{1N_1} = 1$, we have*

$$\Gamma_n \;\to_D\; \Gamma := -\frac{1}{2}\sum_{k=1}^{N_1} \beta_{1k}^2(u_{1k} - u_{1,k-1}) - \frac{1}{2}\alpha^2\sigma_\lambda^2, \tag{3.8}$$

and $e^{|\Gamma_n|}$ is uniformly integrable.

PROPOSITION 3.2. *For any* $\alpha, \beta_{1k}, \beta_{2k} \in R$, $0 = u_{10} < u_{11} < ... < u_{1N_1} = 1$ *and* $0 = u_{20} < u_{21} < ... < u_{2N_2} < \infty$, *we have*

$$I_n \;:=\; E\Big|E\big(e^{i\,\alpha S_n^* + iV_{1n} - \Gamma_n} \mid \mathcal{F}_{n,1}\big) - 1\Big| = o(1). \tag{3.9}$$

Furthermore, as $n \to \infty$, *we have*

$$Ee^{i\,\alpha S_n^* + iV_{1n} + iV_{2n} - \Gamma_n} \;\to\; Ee^{iV_2}. \tag{3.10}$$

We are now ready to prove (3.7). First recall that $\{S_n^*\}_{n\geq 1}$ is tight, and

$$V_{1n} \to_D V_1, \quad V_{2n} \to_D V_2, \tag{3.11}$$

by virtue of (2.2) and (2.3). These facts, together with (3.8), yield that

$$\{S_n^*,\ V_{1n},\ V_{2n},\ \Gamma_n\}_{n\geq 1}$$

is tight on $D[0,1]^4$. Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

$$\{S_{n''}^*,\ V_{1n''},\ V_{2n''},\ \Gamma_{n''}\} \to_D \big\{S,\ V_1,\ V_2,\ \Gamma\big\}. \tag{3.12}$$

on $D[0,1]^4$. Consequently it follows from the uniformity of $Ee^{iS_{n''}^* + iV_{1n''} + iV_{2n''} - \Gamma_{n''}}$ because of Proposition 3.1 that, as $n'' \to \infty$,

$$Ee^{iS_{n''}^* + iV_{1n''} + iV_{2n''} - \Gamma_{n''}} \;\to\; Ee^{iS + iV_1 + iV_2 - \Gamma}. \tag{3.13}$$

This, together with (3.10), yields that

$$Ee^{iS + iV_1 + iV_2 - \Gamma} \;=\; Ee^{iV_2}, \tag{3.14}$$

and hence

$$Ee^{iV_1 + iV_2 + iS + \frac{1}{2}\alpha^2\sigma_\lambda^2} \;=\; e^{-\frac{1}{2}\sum_{k=1}^{N_1}\beta_{1k}^2(u_{1k} - u_{1,k-1})}\, Ee^{iV_2} = Ee^{iV_1 + iV_2},$$

where we have used the independence between $W(t)$ and $G(t)$ and the fact that

$$Ee^{iV_1} \;=\; e^{-\frac{1}{2}\sum_{k=1}^{N_1}\beta_{1k}^2(u_{1k} - u_{1,k-1})}.$$

This proves (3.7) and also completes the proof of Theorem 1. □

3.2. Proof of Theorem 2.2. We may write $S_{1n} = \sum_{k=1}^{n} x_{k,n}\epsilon_{k+1}$ with

$$x_{k,n} \;=\; \big(\frac{c_n}{n}\big)^{1/2} f(c_n y_{k,n}) = f_n(\epsilon_1, ..., \epsilon_k; \eta_1, ..., \eta_k : \xi_1, \xi_2, ...),$$

where $\eta_j = \nu_j, 1 \le j \le n$ and $\xi_j = \nu_{-j+1}, j \ge 1$. It suffices to check that η_j, ϵ_j and $x_{j,n}$ satisfy Assumptions 1 and 2. In fact, by Assumption 4, it is readily seen that $\{\eta_{j+1}, \epsilon_{j+1}, \mathcal{F}_j\}_{j\ge 1}$ forms a martingale difference satisfying Assumption 1 and $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_j \Rightarrow W(t)$ on $D[0,1]$, where $W(t)$ is a standard Winner process. By the definitions of $\mathcal{F}_k$ and ξ_k, $\xi_k, k \ge 1$, is $\mathcal{F}_1$ measurable and obviously $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\xi_j \Rightarrow W^*(t)$, on $D[0,\infty)$, where $W^*(t)$ is a standard Winner process independent of $W(s)$. This gives Assumption 2(i). On the other hand, it follows from Corollary 2.2 and Remark 2.1 of Wang and Phillips (2009a) that, for any $0 \le \gamma' \le 1+\gamma$,

$$\frac{c_n}{n}\sum_{k=1}^n |f(c_n y_{k,n})|^{1+\gamma'} \quad \to_D \quad \int_{-\infty}^{\infty} |f(x)|^{1+\gamma'}dx \begin{cases} L_{W_{3/2-u}}(1,0), & \text{under } \mathbf{C1}, \\ L_W(1,0), & \text{under } \mathbf{C2}, \end{cases} \tag{3.15}$$

whenever $c_n \to \infty$ and $c_n/n \to 0$. By virtue of (3.15), simple calculations show that

$$\begin{aligned} \max_{1\le k\le n}|x_{k,n}| &\le \big(\frac{c_n}{n}\big)^{1+\gamma/2}\sum_{k=1}^n |f(c_n y_{k,n})|^{2+\gamma} = o_P(1), \\ \frac{1}{\sqrt{n}}\sum_{k=1}^n |x_{k,n}| &\le c_n^{-1/2}\frac{c_n}{n}\sum_{k=1}^n |f(c_n y_{k,n})| = o_P(1), \end{aligned}$$

and

$$\sum_{k=1}^n x_{k,n}^2 = \frac{c_n}{n}\sum_{k=1}^n f^2(c_n y_{k,n}) \quad \to_D \quad g^2(W, W^*).$$

This gives Assumption 2 (ii) and (iii). Combining all these fact, we prove (2.11) by Theorem 2.1. □

3.3. Proof of Theorem 2.3. It follows from Theorem 2.2 that

$$\{S_{2n}, G_{2n}^2\} \quad \to_D \quad \{\tau\,\mathcal{L}_W^{1/2}\,N\,,\tau^2\,\mathcal{L}_W\}, \tag{3.16}$$

whenever $c_n/n \to 0$ and $c_n \to \infty$, where $c_n = d_n/h$. This, together with (2.15) with $f_1(y) = K^2(y)$ and $f_2(y) = K(y)$, implies that

$$\big(\frac{d_n}{nh}\big)^{-1/2}\frac{\sum_{t=1}^n \epsilon_{t+1}K_h(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)} \quad = \quad \frac{S_{2n}}{G_{2,n}^2}\,\frac{\sum_{t=1}^n K_h^2(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)} \to_D \tau\,N\,\mathcal{L}_W^{-1/2}. \tag{3.17}$$

On the other hand, it is readily from Assumption 7 and (2.15) with $f_1(y) = K(y)\,m_1(y,x)$ and $f_2(y) = K(y)$ that

$$\begin{aligned} &\big(\frac{d_n}{nh}\big)^{-1/2}\frac{\sum_{t=1}^n \big|m(y_t) - m(x)\big|\,K_h(y_t - x)}{\sum_{t=1}^n K_h(y_t - x)} \\ \le\;& \big(\frac{d_n}{nh}\big)^{-1/2}h^\gamma\,\frac{\sum_{t=1}^n |m_1[(y_t - x)/h, x]|\,K[(y_t - x)/h]}{\sum_{t=1}^n K[(y_t - x)/h]} = o_P(1), \end{aligned}$$

since $nh^{2\gamma+1}/d_n \to \infty$. Taking these estimates into (2.18), we proves (2.20). The proof of (2.21) is similar, and hence the details are omitted. $\Box$

3.4. Proof of Theorem 2.4. By virtue of (2.18) and (3.17), to establish (2.22), it suffices to show that, under Assumptions 6^* and 7^*, for any h satisfying $nh^{1+2(p+1)}/d_n \to 0$,

$$\begin{aligned}\Lambda_{1n} &:= \frac{\sum_{t=1}^{n}\big[m(y_t)-m(x)\big]\,K_h(y_t-x)}{\sum_{t=1}^{n}K_h(y_t-x)} \\ &= \frac{h^p m^{(p)}(x)}{p!}\int_{-\infty}^{\infty} y^p K(y)dy + o_P\big[(\frac{d_n}{nh})^{1/2}\big]. \qquad (3.18)\end{aligned}$$

This is exactly same as in the proof of Theorem 2.2 of Wang and Phillips (2010). For the sake of completeness, we rewrite the proof as follows since it is not very complex. The numerator of Λ_{1n} involves

$$\sum_{t=1}^{n}\{m(y_t)-m(x)\}\,K\left(\frac{y_t-x}{h}\right) = \sum_{j=1}^{p+1} I_j, \qquad (3.19)$$

where

$$\begin{aligned} I_j &= \frac{m^{(j)}(x)}{j!}\sum_{t=1}^{n}(y_t-x)^j\ K\left(\frac{y_t-x}{h}\right), \quad j=1,2,...,p,\\ I_{p+1} &= \sum_{t=1}^{n}\Big\{m(y_t)-\sum_{j=0}^{p}\frac{m^{(j)}(x)}{j!}(y_t-x)^j\Big\}K\left(\frac{y_t-x}{h}\right).\end{aligned}$$

Write $H_j(x)=x^jK(x), j=1,2...,p$. Recall that $K(x)$ has a compact support with twice continuous differentials. For $j=1,...,p-1$, $H_j(x)$ are twice continuous differentiable satisfying $\int_{-\infty}^{\infty}\big[|H_j'(x)|+|H_j''(x)|\big]dx<\infty$. Hence $\widehat{H}_j(t), j=1,...,p-1$, is integrable, where $\widehat{H}_j(t)=\int_{-\infty}^{\infty}e^{itx}H_j(x)dx$. See Proposition 18.1.2 of Gasquet and Witomski (1998) for instance. Furthermore, for $j=1,...,p-1$, $|\widehat{H}_j(x)|\le C\min\{|t|,1\}$ as $\int H_j(x)dx=0$. By virtue of these facts, together with $\int K(x)dx=1$ and $\int H_p(x)dx\neq 0$, it is readily seen from (2.13) that

$$\frac{h^{-j}\left(\frac{d_n}{nh}\right)^{-1/2}I_j}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)} \to_D \ \sigma_j\,N\,\mathcal{L}_W^{-1/2}, \quad j=1,2,...,p-1, \qquad (3.20)$$

where $\sigma_j^2=\big[\frac{m^{(j)}(x)}{j!}\big]^2\int H_j^2(x)dx$, and

$$\frac{h^{-p}\,I_p}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)} \to_P \ \frac{m^{(p)}(x)}{p!}\int H_p(x)dx. \qquad (3.21)$$

On the other hand, by noting $\lim_{h\to 0}\sup_{y\in\Omega}|m^{(p+1)}(yh+x)| \le C$ by Assumption 7*, Taylor expansion yields

$$|I_{p+1}| \;\le\; C\sum_{t=1}^{n}|y_t-x|^{p+1}K\left(\frac{y_t-x}{h}\right),$$

and hence

$$\frac{h^{-(p+1)}|I_{p+1}|}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)} \;\le\; C\frac{\sum_{t=1}^{n}H_{p+1}\left(\frac{y_t-x}{h}\right)}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)} \;\to_P\; C\int H_{p+1}(x)dx, \tag{3.22}$$

where $H_{p+1}(x)=|x|^{p+1}K(x)$. Combining (3.19)-(3.22), simple calculation show that

$$\begin{aligned}
&\Big(\frac{d_n}{nh}\Big)^{-1/2}\Big[\Lambda_{1n}-\frac{h^p m^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^pK(y)dy\Big]\\
&\le\; \frac{\left(\frac{d_n}{nh}\right)^{-1/2}}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)}\sum_{j=1}^{p-1}|I_j| + \frac{\left(\frac{d_n}{nh}\right)^{-1/2}|I_{p+1}|}{\sum_{t=1}^{n}K\left(\frac{y_t-x}{h}\right)}\\
&=\; O_P\big[h+\Big(\frac{d_n}{nh}\Big)^{-1/2}h^{p+1}\big]=o_P(1),
\end{aligned}$$

whenever $nh^{1+2(p+1)}/d_n\to 0$. This proves (3.18). The proof of (2.23) is similar and hence the details are omitted. The proof of Theorem 2.3 is now complete. □

3.5. Proof of Proposition 3.1. Write $Y_{nm}=i\,\beta^*_{1m}\eta_{m+1}+i\,\alpha\,x^*_{m,n}\,\epsilon_{m+1}$. By Assumption 1, simple calculations show that

$$E\big\{(e^{Y_{nm}}-1)\mid\mathcal{F}_{n,m}\big\} \;=\; -\frac{1}{2}E\big\{[\beta^*_{1m}\eta_{m+1}+\alpha x^*_{m,n}\epsilon_{m+1}]^2\mid\mathcal{F}_{n,m}\big\}+R_{nm},$$

where, for the $\delta>0$ defined as in Assumption 1,

$$|R_{nm}| \;\le\; E\big(|Y_{nm}|^{2+\delta}\mid\mathcal{F}_{n,m}\big)\le C\big(n^{-(2+\delta)/2}+|x^*_{m,n}|^{2+\delta}\big).$$

Recall (3.2) and (3.4). It is readily seen that

$$\sum_{m=1}^{n}|R_{nm}| \;\le\; cn^{-\delta/2}+C\max_{1\le k\le n}|x^*_{nk}|^{\delta}\sum_{m=1}^{n}x^{*2}_{nm}=o_P(1).$$

It now follows from Assumption 1 again that

$$\begin{aligned}
\Gamma_n &= \sum_{m=1}^n E\big\{(e^{Y_{nm}}-1) \mid \mathcal{F}_{n,m}\big\} \\
&= -\frac{1}{2}\sum_{m=1}^n E\big\{[\beta_{1m}^*\eta_{m+1}+\alpha x_{m,n}^*\epsilon_{m+1}]^2 \mid \mathcal{F}_{n,m}\big\} + o_P(1) \\
&= -\frac{1}{2}\sum_{m=1}^n \beta_{1m}^{*2} E\big(\eta_{m+1}^2 \mid \mathcal{F}_{n,m}\big) - \frac{\alpha^2}{2}\sum_{m=1}^n x_{m,n}^{*2} E(\epsilon_{m+1}^2 \mid \mathcal{F}_{n,m}) \\
&\quad - \sum_{m=1}^n \alpha\, \beta_{1m}^*\, x_{m,n}^*\, E\big\{(\eta_{m+1}\epsilon_{m+1}) \mid \mathcal{F}_{n,m}\big\} + o_P(1) \\
&= -\frac{1}{2n}\sum_{m=1}^{N_1} \beta_{1m}^2([nu_{1m}]-[nu_{1,m-1}]) - \frac{\alpha^2}{2}\sum_{m=1}^n x_{m,n}^{*2} + o_P(1) \\
&\to_D -\frac{1}{2}\sum_{m=1}^{N_1} \beta_{1m}^2(u_{1m}-u_{1,m-1}) - \frac{\alpha^2}{2}\sigma_\lambda^2,
\end{aligned}$$

where we used the fact that, by (3.3),

$$\begin{aligned}
&|\sum_{m=1}^n \alpha\, \beta_{1m}^*\, x_{m,n}^*\, E\big\{(\eta_{m+1}\epsilon_{m+1}) \mid \mathcal{F}_{n,m}\big\}| \\
&\leq C\sum_{m=1}^n |x_{m,n}^*|\, |E\big\{(\eta_{m+1}\epsilon_{m+1}) \mid \mathcal{F}_{n,m}\big\}| = o_P(1).
\end{aligned}$$

This proves (3.8). Recall that $E(\epsilon_{k+1}^2 \mid \mathcal{F}_{n,k}) + E(\eta_{k+1}^2 \mid \mathcal{F}_{n,k}) \leq C$ and note that

$$\big|E\big\{(e^{Y_{nm}}-1) \mid \mathcal{F}_{n,m}\big\}\big| \leq \frac{1}{2}E(Y_{nm}^2 \mid \mathcal{F}_{n,m}) \leq C\,(n^{-1}+x_{m,n}^{*2}). \tag{3.23}$$

We have

$$|\Gamma_n| \leq C(1+\sum_{m=1}^n x_{m,n}^{*2}) \leq C(1+2\lambda).$$

The uniformly integrality of $e^{|\Gamma_n|}$ is obvious. The proof of Proposition 3.1 is now complete. □

3.6. Proof of Proposition 3.2. Write $Y_{nm} = i\,\beta_{1m}^*\eta_{m+1} + i\,\alpha\, x_{m,n}^*\,\epsilon_{m+1}$ as in the proof of Proposition 3.1, and let $E_1X = E(X \mid \mathcal{F}_{n,1})$. We have that

$$\begin{aligned}
I_n &= E\Big|E_1\exp\Big\{\sum_{k=1}^n Y_{nk} - \sum_{k=1}^n E\big[\big(e^{Y_{nk}}-1\big) \mid \mathcal{F}_{n,k}\big]\Big\} - 1\Big| \\
&\leq \sum_{m=2}^n E\Big|E_1\Big[\exp\Big\{\sum_{k=1}^{m-1} Y_{nk} - \sum_{k=1}^m E\big[\big(e^{Y_{nk}}-1\big) \mid \mathcal{F}_{n,k}\big]\Big\}\big\{e^{Y_{nm}} - e^{E\big[\big(e^{Y_{nm}}-1\big)\mid\mathcal{F}_{n,m}\big]}\big\}\Big]\Big| \\
&\qquad + E\Big|E_1\exp\Big\{Y_{n1} - E\big[\big(e^{Y_{n1}}-1\big) \mid \mathcal{F}_{n,1}\big]\Big\} - 1\Big| \\
&:= I_{1n} + I_{2n}.
\end{aligned} \tag{3.24}$$

Recall that (3.23). It follows from $\max_{1\le m\le n} x_{nm}^{*2} = o_P(1)$ and $\max_{1\le m\le n} x_{nm}^{*2} \le \lambda$ that

$$\begin{aligned}
\Delta_{1n} &:= \Big|E_1 \exp\Big\{Y_{n1} - E\big[\big(e^{Y_{n1}}-1\big) \mid \mathcal{F}_{n,1}\big]\Big\} - 1\Big| \\
&\le \Big|e^{-E\left[\left(e^{Y_{n1}}-1\right)|\mathcal{F}_{n,1}\right]} - 1\Big| + \Big|E\big[\big(e^{Y_{n1}}-1\big) \mid \mathcal{F}_{n,1}\big]\Big| \\
&\le C(n^{-1}+x_{n1}^{*2})e^{C(n^{-1}+x_{n1}^{*2})} = o_P(1),
\end{aligned}$$

and Δ_{1n} is uniformly integrable as $\Delta_{1n} \le C(1+\lambda)e^{\lambda}$. This implies that $I_{2n} = E\Delta_{1n} \to 0$, as $n \to \infty$.

To consider I_{1n}, write

$$\begin{aligned}
u_{n,m} &= e^{\sum_{k=1}^{m-1} Y_{nk} - \sum_{k=1}^{m} E\left[(e^{Y_{nk}}-1)|\mathcal{F}_{n,k}\right]}, \\
v_{n,m} &= e^{Y_{nm}} - e^{E\left[(e^{Y_{nm}}-1)|\mathcal{F}_{n,m}\right]}.
\end{aligned}$$

Using (3.23) and $\sum_{k=1}^{n} x_{kn}^{*2} \le 2\lambda$ again, it follows that

$$|u_{n,m}| \le e^{\sum_{k=1}^{m}|E\left[(e^{Y_{nk}}-1)|\mathcal{F}_{n,k}\right]|} \le e^{C\sum_{k=1}^{m}(n^{-1}+x_{nk}^{*2})} \le e^{C(1+2\lambda)}. \tag{3.25}$$

As for $E(v_{n,m} \mid \mathcal{F}_{n,m})$, by using $|e^x - 1 - x| \le |x|^{(2+\delta)/2} e^{|x|}$, we have

$$\begin{aligned}
|E(v_{n,m} \mid \mathcal{F}_{n,m})| &= \big|E\big\{(e^{Y_{nm}}-1) \mid \mathcal{F}_{n,m}\big\} + 1 - e^{E\left[(e^{Y_{nm}}-1)|\mathcal{F}_{n,m}\right]}\big| \\
&\le \big|E\big\{(e^{Y_{nm}}-1) \mid \mathcal{F}_{n,m}\big\}\big|^{(2+\delta)/2} e^{|E\left[(e^{Y_{nm}}-1)|\mathcal{F}_{n,m}\right]|} \\
&\le C\, e^{C(1+2\lambda)} (n^{-1}+x_{nm}^{*2})^{(2+\delta)/2}.
\end{aligned}$$

Now, by recalling $\mathcal{F}_{n,m} \subseteq \mathcal{F}_{n,m+1}$ for any $n \ge m \ge 1$ and $n \ge 1$, it is readily seen that

$$\begin{aligned}
I_{1n} &\le \sum_{m=2}^{n} E\big[|u_{n,m}|\,|E(v_{n,m} \mid \mathcal{F}_{n,m})|\big] \\
&\le C\, e^{2C(1+2\lambda)}\, E\sum_{m=2}^{n}(n^{-1}+x_{nm}^{*2})^{(2+\delta)/2} \to 0,
\end{aligned}$$

since

$$\begin{aligned}
\Delta_{2n} &:= \sum_{m=2}^{n}(n^{-1}+x_{nm}^{*2})^{(2+\delta)/2} \\
&\le (n^{-\delta/2} + \max_{1\le m\le n}|x_{nk}^{*}|^{\delta})(1+\sum_{m=1}^{n} x_{nk}^{*2}) = o_P(1),
\end{aligned}$$

and Δ_{2n} is uniformly integrable by noting $\Delta_{2n} \le C(1+\lambda^{1+\delta/2})$. Taking these facts into (3.24), we proves (3.9).

The proof of (3.10) is simple. Indeed it follows from (3.9), V_{2n} is $\mathcal{F}_{n,1}$-measurable and and $Ee^{iV_{1n}} \to Ee^{iV_1}$ due to (2.3) that

$$\begin{aligned} & |Ee^{i\,\alpha S_n^* + iV_{1n} + iV_{2n} - \Gamma_n} - Ee^{iV_2}| \\ \le \;& E\big|E\big(e^{i\,\alpha S_n^* + iV_{1n} - \Gamma_n} \mid \mathcal{F}_{n,1}\big) - 1\big| + |Ee^{iV_{2n}} - Ee^{iV_1}| \to 0. \end{aligned}$$

The proof of Proposition 3.2 is now complete. □

REFERENCES

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

Cai, Z., Li, Q. and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *J. Econometrics* **148**, 101–113.

Choi, I. and Pentti, S. (2010). Tests for nonlinear cointegration. Forthcoming in Econometric Theory

Gao, J., K., Maxwell, Lu, Z., Tjøstheim, D. (2009). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, **25**, 1869–1892.

Gao, J., K., Maxwell, Lu, Z., Tjøstheim, D. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *Ann. Statist.* **37**, 3893–3928.

Gaquet, C. and Witomski, P. (1999). *Fourier Analysis and Applications.* Springer, New York.

Geman, D, J. Horowitz (1980). Occupation densities. *Ann. Probab.*, **8** , 1–67.

Guerre, E. (2004). Design-Adaptive pointwise nonparametric regression estimation for recurrent Markov time series. *Unpublished pamnuscript.*

Hall, P. (1977). Martingale invariance principles. *Annals of Probability,* **5**, 875-887.

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

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8**, 489500.

Hong, S. H. and Phillips, P. C. B. (2010) Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. *Journal of Business and Economic Statistics*, forthcoming.

Jeganathan, P. (2006). Convergence in distribution of row sum processes to mixtures of addive processes. *unpublished manuscript.*

Karlsen, H. A.; Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29**, 372–416.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model, *Ann. Statist.*, **35**, 252-299.

Kasparis, I. (2008). Detection on functional form misspecification in cointegrating relations, *Econometric Theory*, **24**, 1373-1403.

Kasparis, I. (2011). Functional form misspecification in regressions with a unit root, *Econometric Theory*, **27**, forthcoming.

Kasparis, I. and Phillips, P. C. B. (2009). Dynamic Misspecification in Nonparametric Cointegrating Regression. *Yale University Working Paper.*

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19**, 10351070.

Marmer, V. (2008). Nonlinearity, nonstationarity, and spurious forecasts. *J. Econometrics* **142**, 1–27.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 620628.

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

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

Phillips, P. C. B. and J. Y. Park (1998). Nonstationary Density Estimation and Kernel Autoregression. *Cowles Foundation discuss paper No. 1181.*

Schienle, M. (2008). Nonparametric Nonstationary Regression, Unpublished Ph.D. Thesis, University of Mannheim.

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 P. C. B. Phillips (2009a). Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression, *Econometric Theory*, **25**, 710-738.

Wang, Q. and P. C. B. Phillips (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **77**, 1901-1948.

Wang, Q. and P. C. B. Phillips (2010). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, forthcoming.

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

The authors acknowledge partial 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.