Uniform convergence for Nadaraya-Watson estimators with non-stationary data

Nigel Chan and Qiying Wang

The University of Sydney

April 11, 2012

Abstract

This paper investigates the uniform convergence for the Nadaraya-Watson estimators in a non-linear cointegrating regression. Our results provide a optimal convergence rate without the compact set restriction, allowing for martingale innovation structure and the situation that the data regressor sequence is a partial sum of general linear process including fractionally integrated time series. We also investigate the uniform convergence for functionals of general non-stationary time series, which is of interests in its own right.

Key words and phrases: Cointegration, uniform convergence, non-parametric regression, kernel estimation, nonstationarity, nonlinearity.
JEL Classification: C13, C14, C22.

1 Introduction

The past few decades has witnessed significant developments on cointegration analysis, particularly linear models have dominated empirical work in the applications of these methods. While it is convenient for practical implementation, linear structure of the traditional cointegration model is often too restrictive for modeling propose. Empirical examples in this regard can be found in Granger and Teräsvirta (1993). In such situations, given the prevalence of nonlinear relationships in economics, it is expected that nonlinear cointegration captures the features of many long-run relationships in a more realistic manner.

Typical non-linear cointegrating regression model has the form

$$y_t = f(x_t) + u_t, \qquad t = 1, 2, ..., n, \tag{1}$$

where $\{u_t\}$ is a zero mean equilibrium error, x_t is a non-stationary regressor and $f(\cdot)$ is an unknown real function on R. With given observations (x_t, y_t) which may include non-stationary components, the point-wise estimation and inference of the unknown $f(\cdot)$ have been becoming increasing interests in literature. Phillips and Park (1998) studied nonparametric autoregression in the context of a random walk. Karlsen and Tjøstheim (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. Karlsen, et al. (2007) address the function estimation problem for a possibly nonlinear cointegrating relation, providing an asymptotic theory of estimation and inference for nonparametric forms of cointegration. Under similar conditions and using related Markov chain methods, Schienle (2008) investigated additive nonlinear versions of (1) and obtained a limit theory for nonparametric regressions under smooth backfitting. More recently, Wang and Phillips (2009a, 2009b, 2011) 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 of the type $x_t = \sum_{j=1}^{t} \xi_j$ where ξ_j is a general linear process. For other related current works, we refer to Kasparis and Phillips (2009), Park and Phillips (1999, 2001), Bandi (2004), Gao, et al (2009a, b), Choi and Saikkonen (2004, 2010), Marmer (2008), Chen, et al (2010), Wang and Phillips (2012), and Wang (2011).

Different from these aforementioned point-wise asymptotic work, this paper is concerned with the uniform convergence for the Nadaraya-Watson estimator $\hat{f}(x)$ of $f(x)$ in the non-linear cointegrating regression model (1), defined by

$$\hat{f}(x) = \frac{\sum_{s=1}^{n} y_s\, K[(x_s - x)/h]}{\sum_{s=1}^{n} K[(x_s - x)/h]}, \tag{2}$$

where $K(x)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. In this regard, for a near $I(1)$ regressor x_t, Wang and Wang (2011) established uniform consistency for both the regression and the volatility functions under a compact set. Without the compact set restriction, Gao, et al. (2011) derived strong and weak consistency results for the case where the x_k is a null-recurrent Markov chain, but imposed the independence between u_k and x_k. More currently, Wang and Chan (2011) investigated the uniform convergence for a class of martingales. As a direct application of their main result, Wang and Chan (2011) removed the restriction on the independence between u_k and x_k as required in Gao, et al. (2011), but their result only works for the x_k to be a recurrent Markov chain.

Using some quite different techniques, in this paper, we establish the uniform consistency of $\hat{f}(x)$ defined by (2) with sharp (probably optimal) convergence rate without the compact set restriction, under very general settings on the regressor x_t and the error presses u_t. Explicitly, our model allows for the regressor x_t to be a partial sum of general linear process and for a martingale innovation structure. We also investigate the uniform convergence for functionals of general non-stationary time series. This result is of interests in its own right.

This paper is organized as follow. Section 2 presents our main results. The uniform convergence for functionals of general non-stationary time series is considered in Section 3. Technical proofs are postponed to Section 4. Throughout the paper, we denote constants by $C, C_1, C_2, ...$ which may be different at each appearance.

2 Main results

Let $\{\xi_j, j \geq 1\}$ be a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k \, \epsilon_{j-k}, \tag{3}$$

where $\{\epsilon_j, -\infty < j < \infty\}$ is a sequence of iid random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$, $\mathbb{E}|\epsilon_0|^k < \infty$ for some $k > 2$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfies $\int_{-\infty}^{\infty}(1 + |t|)|\varphi(t)|dt < \infty$. Throughout the paper, the coefficients ϕ_k, $k \geq 0$ 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 $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

We also make use of the following assumptions in the asymptotic development.

Assumption 2.1. $x_t = \sum_{j=1}^{t} \xi_j$, where ξ_j is defined as in (3).

Assumption 2.2. $\{u_t, \mathcal{F}_t\}_{t\geq 1}$ is a martingale difference, where $\mathcal{F}_t = \sigma(x_1, ..., x_{t+1}, u_1, ..., u_t)$, satisfying $E(u_t^2 \mid \mathcal{F}_{t-1}) \to_{a.s.} \sigma^2 < \infty$, as $t \to \infty$ and $\sup_{t\geq 1} E|u_t|^{2p} < \infty$, where $p \geq 1 + [1/\delta_0]$ for some $0 < \delta_0 < \alpha$, where

$$\alpha \;=\; \begin{cases} \mu - 1/2, & \text{under C1,} \\ 1/2, & \text{under C2.} \end{cases} \tag{4}$$

Assumption 2.3. The kernel K satisfies that $\sup_x K(x) < \infty$, $\int_{-\infty}^{\infty}(1+|s|)K(s)ds < \infty$, and

$$|K(x) - K(y)| \leq C\,|x - y|.$$

for any $x, y \in R$, whenever $|x - y|$ is sufficiently small.

Assumption 2.4. There exists a real positive function $g(x)$ such that

$$|f(y) - f(x)| \;\leq\; C\,|y - x|^{\beta} g(x),$$

uniformly for some $0 < \beta \leq 1$ and any $(x, y) \in \Omega_\epsilon$, where ϵ can be chosen sufficient small and $\Omega_\epsilon = \{(x, y) : |y - x| \leq \epsilon, x \in R\}$.

Assumption 1 allows for the regressor x_t to be a fractionally integrated process (under **C1**) and also a partial sum of short memory linear process (under **C2**). The condition **C2** is used in Wang and Wang (2011) as well as Wang and Chan (2011). The condition **C1** is new. Putting $d_n^2 = \mathbb{E}x_n^2$, we have

$$d_n^2 = \mathbb{E}x_n^2 \sim \begin{cases} c_\mu n^{3-2\mu}\rho^2(n), & \text{under C1,} \\ \phi^2 n, & \text{under C2.} \end{cases} \tag{5}$$

where $c_\mu = 1/((1 - \mu)(3 - 2\mu)) \int_0^\infty x^{-\mu}(x + 1)^{-\mu}dx$. See, e.g., Wang, Lin and Gulati (2003) for instance. This notation and fact will be repeatedly used later without further explanation.

Assumption 2.2 is standard as in the stationary situation in which we impose a martingale structure so that $cov(u_{t+1}, x_t) = E[x_t E(u_{t+1} \mid \mathcal{F}_t)] = 0$. This conditional orthogonality conditions is the same as in Wang and Wang (2011) as well as Wang and Chan (2011), but weaker than that in Gao et al (2011), which assumes x_t and u_t are independent.

Assumptions 2.3 and 2.4 are standard conditions on the kernel $K(x)$ and the unknown functional $f(x)$, which are the same as in Wang and Wang (2011). Assumptions 2.3 does not restrict the kernel to have a finite compact, as required in most of the previous works. Assumption 2.4 hosts for a wide set of functionals, including that $f(x) = \theta_1 + \theta_2 x + ... + \theta_k x^{k-1}$; $f(x) = \theta_1 + \theta_2\, x^{\theta_3}$; $f(x) = x(1+\theta x)^{-1} I(x \geq 0)$; $f(x) = (\theta_1 + \theta_2\, e^x)/(1+e^x)$.

We have the following main result.

Theorem 2.1. *Under Assumptions 2.1–2.4, for any* $B_n \leq M_0\, d_n/\log^{\gamma_0} n$*, any* h *satisfying* $h \to 0$ *and* $n^{\alpha-\delta_0} h \to \infty$*, we have*

$$\sup_{|x|\leq B_n} |\hat{f}(x) - f(x)| \;=\; O_P\big[\big(\frac{d_n}{nh}\big)^{1/2} \log^{1/2} n + h^{\beta}\,\delta_n\big], \tag{6}$$

where $M_0 > 0$ *is a fixed constant,* $\delta_n = \sup_{|x|\leq B_n} g(x)$*, and*

$$\gamma_0 \;=\; \begin{cases} \frac{4(3-2\mu)}{2\mu-1}, & \text{under C1 and } 9/10 < \mu < 1\ , \\ \frac{(5-2\mu)(3-2\mu)}{(2\mu-1)^2}, & \text{under C1 and } 1/2 < \mu \leq 9/10, \\ 4, & \text{under C2}\ . \end{cases}$$

Remark 2.1. *A better result can be obtained if we are only interested in the pointwise asymptotics for* $\hat{f}(x)$*. Indeed, as in Wang and Phillips (2009a, b) with minor modification, we may show that, for each fixed point* x*,*

$$\hat{f}(x) - f(x) \;=\; O_P\big[\big(\frac{d_n}{nh}\big)^{1/2} + h^{\beta}\big]. \tag{7}$$

Furthermore $\hat{f}(x)$ *has an asymptotic distribution that is mixing normal, under minor additional conditions to Assumption* 2.2.

In comparison to the stationary regressor situation, where the sharp rate of convergence in (6) is $O_P[(\log n/nh)^{1/2}]$ *(see Hansen (2008), for instance), the convergence rate in (6) is sharp (probably optimal). There is an essential difference for the rate of convergence between stationary and non-stationary time series. The reason behind the difference is mainly because, in non-stationary case, the amount of time spent by the process around any specific point is of order* d_n *rather than* n*. More explanation can be found in Remark 3.3 of Wang and Phillips (2009a).*

Remark 2.2. *The bandwidth condition* $n^{\alpha-\delta_0}h \to \infty$ *is related to the moment condition of the martingale sequence* u_t. *If high moments of* u_t *exists, we can choose* δ_0 *to be sufficiently small. When* $|u_k| \le C$, *we can obtain better bandwidth condition* $n^{\alpha}h\log^{-\theta} n \to \infty$, *where* θ *is a real number depending only on* γ_0. *In (6),* $h^{\beta}\,\delta_n \to 0$ *is necessary to guarantee the uniform consistency of* $\hat{f}(x)$. *If the regression function* $f(x)$ *has thin tail, e.g.* $f(x) = (\alpha+\beta e^x)/(1+e^x)$, δ_n *will be bounded by a constant. In a pointwise situation, the term* h^{β} *can be improved if we put a bias term in the left hand of (7). See Wang and Phillips (2011). It is not clear at the moment whether there are similar properties for the uniform consistency (6). The issue seems to be difficult, since we have to consider the uniform convergence for zero energy functionals of non-stationary time series, which is not available at the moment. We leave this topic for future work.*

Remark 2.3. *The result (6) improves Theorem 3.2 of Wang and Wang (2012), even assuming* x *is in a finite compact set. As stated in Remark 2.1, our rate of convergence may be optimal for this kind of uniform convergence. In a specific situation that* $x_t = \sum_{j=1}^{t}\epsilon_j$, *that is,* x_t *is a random walk, Wang and Chan (2011) proved that the range* $|x| \le B_n$ *in Theorem 2.1 can be improved to* $|x| \le M_0\sqrt{n}$. *It is not clear at the moment whether the similar result still holds under our general regressors. It seems that a quite different techniques is required to attack this problem and hence leave for future work. See Remark 3.2 for more details in this regard.*

Remark 2.4. *To outline the essentials of the arguments in the proof of Theorem 2.1, we split the error of* $\hat{f}(x) - f(x)$ *as*

$$\begin{aligned} \hat{f}(x) - f(x) &= \frac{\sum_{t=1}^{n} u_t K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} + \frac{\sum_{t=1}^{n}\left[f(x_t) - f(x)\right] K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} \\ &:= \Theta_{1n}(x) + \Theta_{2n}(x). \end{aligned} \tag{8}$$

Under Assumptions 2.3 and 2.4, it is readily seen that, whenever n *is sufficiently large,*

$$\sup_{|x|\le B_n} |\Theta_{2n}(x)| \;\le\; C h^{\beta}\,\delta_n.$$

Under Assumption 2.2, $\{S_n(x), \mathcal{F}_t\}_{t\ge 1}$ *is a class of martingale, where* $S_n(x) = \sum_{t=1}^{n} u_t K[(x_t - x)/h]$. *In order to estimate the* Θ_{1n}, *by Theorem 2.1 of Wang and Chan (2011) (See*

Appendix for a restatement of the theorem), it suffices to show the following results:

$$\sup_{|x|\le M_0\, d_n/\log^{\gamma_0} n} \sum_{k=1}^{n} K^2[(x_k - x)/h] \;=\; O_P(nh/d_n), \tag{9}$$

$$\Big[\inf_{|x|\le M_0\, d_n/\log^{\gamma_0} n} \sum_{t=1}^{n} K[(x_t - x)/h]\Big]^{-1} \;=\; O_P[d_n/(nh)], \tag{10}$$

for any h, satisfying $h \to 0$ and $n^{\alpha} h \log^{-\theta} n \to \infty$, where θ is a real number depending only on γ_0. The results (9) and (10) is a direct consequence of Theorem 3.2, together with Remark 3.1, where we established the uniform consistency for functionals of non-stationary time series .

To end this section, we introduce the following examples which satisfy the condition C1 and C2 respectively.

Example 1. Let $\{Y_t\}$ be an ARMA(p,q) process, defined by

$$\phi(B)Y_t = \theta(B)\epsilon_t, \tag{11}$$

where ϕ and θ are the p^{th} and q^{th} degree polynomials

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \tag{12}$$

and

$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q \tag{13}$$

and B is the backshift operator, $\{\epsilon_j, -\infty < j < \infty\}$ is white noise with mean 0 and variance σ^2. Suppose that $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeroes. If $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \le 1$, Y_t can be represented as

$$Y_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \tag{14}$$

where $\psi(z) = \theta(z)/\phi(z)$, $|z| \le 1$ satisfying $\sum_{k=0}^{\infty} |\psi_k| < \infty$ and $\sum_{k=0}^{\infty} \psi_k \neq 0$. See, e.g. Brockwell and Davis (1987). That is, Y_t is a linear process satisfying the condition (C2).

Example 2. Let $\{Z_t\}$ be a fractionally integrated process $\{Z_t\}$ initialized at $Z_0 = 0$, defined by

$$(1 - B)^d Z_t = \epsilon_t, \tag{15}$$

where $0 \le d < 1/2$, B is a backshift operator, and $\{\epsilon_j, -\infty < j < \infty\}$ is a sequence of i.i.d. random variables with $\mathbb{E}\epsilon_0 = 0$, $\mathbb{E}\epsilon_0^2 = 1$, and characteristic function $\phi(t)$ of ϵ_0 satisfying $\int_{-\infty}^{\infty}(1+|t|)|\phi(t)|dt < \infty$. The fractional difference operator $(1-B)^\gamma$ is defined by its Maclaurin series (by its binomial expansion, if γ is an integer):

$$(1 - B)^\gamma = \sum_{j=0}^{\infty} \frac{\Gamma(-\gamma + j)}{\Gamma(-\gamma)\Gamma(j+1)} B^j \quad \textit{where} \quad \Gamma(z) = \begin{cases} \int_0^\infty s^{z-1} e^{-s} ds, & \text{if } z > 0 \\ \infty, & \text{if } z = 0. \end{cases} \tag{16}$$

If $z < 0$, $\Gamma(z)$ is defined by the recursion formula $z\Gamma(z) = \Gamma(z+1)$.

We may present Z_t as $Z_t = \sum_{k=0}^{\infty} a(k)\epsilon_{t-k}$ with

$$a(k) = \frac{\Gamma(k + d)}{\Gamma(k+1)\Gamma(d)} \sim \frac{1}{\Gamma(d)} k^{d-1}, \tag{17}$$

as $k \to \infty$. See Wang, et al. (2003) for instance. That is, Z_t is a linear process satisfying the condition (C1) with $\mu = 1 - d$.

3 Uniform bounds for functionals of non-stationary time series

Consider a triangular array $x_{k,n}, 1 \le k \le n, n \ge 1$ constructed from some underlying time series. In most practical situations, $x_{k,n}$ is equal to x_k/d_n, where x_k is a partial sum and $0 < d_n \to \infty$ in such a way that x_n/d_n has a limit distribution. The functional of interest S_n of $x_{k,n}$ is defined by the sample average

$$S_n(x) = \sum_{k=1}^{n} g[c_n (x_{k,n} + x)], \quad x \in R,$$

where c_n is a certain sequence of positive constants and g is a real function on R. Such functionals commonly arise in non-linear regression with integrated time series [Park and Phillips (1999, 2001)] and non-parametric estimation in relation to nonlinear cointegration models [Phillips and Park (1998), Karlsen and Tjostheim (2001), Wang and Phillips (2009a, 2009b, 2011)]. The limit behavior of $S_n(x)$ in the situation that $c_n \to \infty$ and $n/c_n \to \infty$ is particularly interesting and important for practical applications as it provides a setting that accommodates a sufficiently wide range of bandwidth choices to be relevant for non-parametric kernel estimation.

For a fixed x, the limit distribution of $S_n(x)$ has been established by Wang and Phillips (2009a, 2009b, 2011) under very general setting on $x_{k,n}$. The aim of this section is investigate the uniform (upper and lower) bound for $S_n(x)$ on a compact set or on R. As discussed in Section 2, these results will be useful in the investigation of uniform convergence for kernel estimates in a non-linear cointegrating regression. We make use of the following assumptions in the development of main results.

Assumption 3.1. $\sup_x |g(x)| < \infty$, $\int_{-\infty}^{\infty} |g(x)|dx < \infty$ *and* $|g(x)-g(y)| \le C|x-y|$ *whenever* $|x-y|$ *is sufficient small on* R.

Assumption 3.2. *There exists a stochastic process* $G(t)$ *having a continuous local time* $L_G(t,s)$ *such that* $x_{[nt],n} \Rightarrow G(t)$, *on* $D[0,1]$, *where weak convergence is understood w.r.t the Skorohod topology on the space* $D[0,1]$.

Assumption 3.3. *For all* $0 \le k < l \le n, n \ge 1$, *there exist a sequence of constants* $d_{l,k,n} \sim C_0[n/(l-k)]^{-d}$ *for some* $0 < d < 1$ *and a sequence of increasing* σ*-fields* $\mathcal{F}_{k,n}$ *(define* $\mathcal{F}_{0,n} = \sigma\{\phi, \Omega\}$, *the trivial* σ*-field) such that* $x_{k,n}$ *are adapted to* $\mathcal{F}_{k,n}$ *and, conditional on* $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ *has a density* $h_{l,k,n}(x)$ *satisfying that* $h_{l,k,n}(x)$ *is uniformly bounded by a constant* K *and uniformly for* $j-k$ *sufficiently large*

$$\sup_y |h_{l,k,n}(y+u) - h_{l,k,n}(y)| \le C \min\{|u|, 1\}. \tag{18}$$

Theorem 3.1. *Under Assumptions 3.1 and 3.3, we have*

$$\sup_{|x| \le n^{m_0}} |S_n(x)| = O[(n/c_n)\log n], \quad a.s., \tag{19}$$

for any $c_n \to \infty$ *and* $c_n/n \to 0$, *and any fixed constant* $0 < m_0 < \infty$. *If there exist positive constants* m *(allow to be sufficient large) and* k *(allow to be sufficient small) such that* $n \sup_{|x|>n^m/2} |g(c_n x)| = O[(n/c_n)\log n]$ *and* $n^{-mk} \sum_{t=1}^n |x_{t,n}|^k = O(1)$ *a.s., then*

$$\sup_{x \in R} |S_n(x)| = O[(n/c_n)\log n], \quad a.s. \tag{20}$$

We mention that the additional conditions to establish (20) are close to minimal. The bound can be improved if we are only concerned with convergence in probability, as stated in the following theorem.

Theorem 3.2. *Under Assumptions 3.1-3.3, we have*

$$\sup_{|x|\le M_0/\log^{\gamma_1} n} |S_n(x)| = O_P(n/c_n), \tag{21}$$

where

$$\gamma_1 \quad = \quad \begin{cases} 4\left(\frac{d}{1-d}\right), & if \quad 0 < d \le 3/5, \\ \left(\frac{1+d}{1-d}\right)\left(\frac{d}{1-d}\right), & if \quad 3/5 < d < 1. \end{cases} \tag{22}$$

for any fixed $M_0 > 0$, $c_n \to \infty$ and $(n/c_n)\log^{-\theta} n \to \infty$, where $\theta = (1-d)\gamma_1/d$. If in addition $\int_{-\infty}^{\infty} g(x)dx \neq 0$, then

$$\left[\inf_{|x|\le M_0/\log^{\gamma_1} n} |S_n(x)|\right]^{-1} = \sup_{|x|\le M_0/\log^{\gamma_1} n} |S_n^{-1}(x)| = O_P(c_n/n). \tag{23}$$

Remark 3.1. *The requirement on $d_{l,k,n}$ is mild. Indeed, in most practical situations, $x_{k,n} = \sum_{j=1}^{k} \eta_j/d_n$, where $d_n^2 = var(\sum_{j=1}^{k} \eta_j) \sim C_0 n^{2d}$ for some $0 < d < 1$, as stated in the following examples. It is hence natural to assume $d_{l,k,n} \sim C_0[n/(l-k)]^{-d}$. This condition can be generalized to $d_{l,k,n} \sim C_0[n\rho(n)/(l-k)\rho(l-k)]^{-d}$, where $\rho(n)$ is slowly varying function at infinity or more generally to those as in Assumption 2.3 (i) of Wang and Phillips (2009a) without essential difficulty. We omit this kind of generalization here for notation convenience.*

Remark 3.2. *It is readily seen that $d_{l,k,n} \sim C_0[n/(l-k)]^{-d}$ for some $0 < d < 1$ satisfies Assumption 2.3 (i) of Wang and Phillips (2009a). If in addition to Assumptions 3.1-3.3, $\int_{\infty}^{\infty} g(x)dx \neq 0$. Theorem 2.1 and Remark 2.1 of Wang and Phillips (2009a) yield that*

$$\frac{c_n}{n}\sum_{t=1}^{n} g[c_n\,(x_{t,n} + y_n)] \to_D \int_{-\infty}^{\infty} g(x)dx\, L_G(1, y), \tag{24}$$

whenever $c_n \to \infty$, $n/c_n \to \infty$ and $y_n \to y$, where $L_G(1, 0)$ is the local time process defined by

$$L_G(t, s) = \lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_0^t I\big\{|G(r) - s| \le \epsilon\big\}dr.$$

Note that $P(0 < L_G(1,x) < \infty) = 1$, for any fixed $x \in R$ and $L_G(1,x) \to 0$, as $x \to \infty$. It might be possible to improve the range $|x| \le M_0/\log^{\gamma_1} n$ in (21) and (23) into $x \in R$ and $|x| \le M_0$ (not possible for $|x| \le b_n$ in (23) where $b_n \to \infty$), respectively. However, to do this, we require a quite different technique, and hence leave it for future work.

The essential behind the proof of Theorem 3.2 is a fact that $S_n(x)$ can be approximated by $S_n(0)$ under a reasonable rate. Explicitly we have the following theorem.

Theorem 3.3. *Let $\gamma \ge 0$. Under Assumptions 3.1-3.3, we have*

$$\sup_{|x| \le M_0/\log^{\gamma} n} |S_n(x) - S_n(0)| = O_P[(n/c_n)\log^{1-\lambda} n], \tag{25}$$

for any fixed $M_0 > 0$, $c_n \to \infty$ and $(n/c_n)\log^{-\theta} n \to \infty$, where $\theta = \max\{\lambda+1, \eta\}$,

$$\lambda = \begin{cases} \frac{(1-d)^2\gamma/d-(2d-1)}{2-d}, & \text{if } 1/2 < d < 1 \text{ and } \gamma \ge d(5d-1)/(2d-1), \\ \frac{(1+\gamma)(1-d)-2d}{1+d}, & \text{otherwise}, \end{cases} \tag{26}$$

and

$$\eta = \begin{cases} \gamma + (\lambda+1)(1-2d)/(1-d), & if \quad 0 < d < 1/2, \\ \gamma - 1, & if \quad d = 1/2, \\ (1-d)\gamma/d, & if \quad 1/2 < d < 1. \end{cases} \tag{27}$$

To end this section, we introduce the following examples on $x_{k,n}$, which satisfy Assumptions 3.2 and 3.3.

Example 3.1. Let $\{\xi_j, j \ge 1\}$ be a stationary sequence of Gaussian random variables with $\mathbb{E}\xi_1 = 0$ and covariances $\gamma(j-i) = \mathbb{E}\xi_j\xi_i$ satisfying the following condition for some $0 < \beta < 2$ and $\lambda > 1$,

$$d_n^2 \equiv \sum_{1 \le i,j \le n} \gamma(j-i) \sim n^{\beta} \quad and \quad |\tilde{\gamma}_{l,k}| \le \lambda d_k d_{l-k}, \tag{28}$$

as $\min\{k, l-k\} \to \infty$, where

$$\tilde{\gamma}_{l,k} = \sum_{i=1}^{k} \sum_{j=k+1}^{l} \gamma(j-i) \tag{29}$$

Let $x_{k,n} = \sum_{j=1}^{k} \xi_j / d_n$, $1 \le k \le n$. Then $x_{k,n}$ satisfies Assumptions 3.2 and 3.3 with $G(t) = W_{\beta/2}(t)$. See Corollary 2.1 of Wang and Phillips (2009a).

Here and below, $W_\beta(t)$ denotes fractional Brownian motion with $0 < \beta < 1$ on $D[0,1]$, defined as follows:

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

where $W(s)$ is a standard Brownian motion and

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

Example 3.2. Let $x_{k,n} = x_k/d_n$, where x_k is defined as in Assumption 2.1 and d_n is defined as in (5). Then $x_{k,n}$ satisfies Assumptions 3.2 and 3.3 with

$$G(t) \quad = \quad \begin{cases} W_{\mu-3/2}(t), & \text{under C1,} \\ W(t), & \text{under C2.} \end{cases} \tag{30}$$

The verification of $x_{k,n}$ satisfying Assumption 3.2 and 3.3 are largely similar to that the proof of Corollary 2.2 of Wang and Phillips (2009a). The only additional work is to check (18), which is given in Appendix.

4 Proofs of main results

This section provides proofs of the main results. we start with a lemma, which will be heavily used in the proof of main results. Throughout this section, we denote constants by $C, C_1, C_2, ...$, which may be different at each appearance.

Lemma 4.1. *For any real function $l(x)$ satisfying $\sup_x |l(x)| < \infty$ and $\int_{-\infty}^{\infty} |l(x)|dx < \infty$, there exist a constant H_0 not depending on t_1, t_2, t_3 and m such that*

$$\begin{aligned} & \sup_x E\big(|\sum_{k=t_2}^{t_3} l[c_n (x_{k,n} + x)]|^m \mid \mathcal{F}_{n,t_1}\big) \\ \le \quad & H_0^m (m+1)!\, n^d\, c_n^{-1} (t_3 - t_1)^{1-d} \big[1 + \big\{(t_3 - t_2)^{1-d} n^d\, c_n^{-1}\big\}^{m-1}\big]. \end{aligned} \tag{31}$$

for all $0 \le t_1 < t_2 < t_3 \le n$ and integer $m \ge 1$. In particular, by letting $t_1 = 0, t_2 = 1$

and $t_3 = n$, *we have*

$$\sup_x E\Big|\sum_{k=1}^{n} l[c_n\,(x_{k,n}+x)]\Big|^m \quad\le\quad H_0^m\,(m+1)!\,(n/c_n)^m. \tag{32}$$

Proof. First recall that, given on $\mathcal{F}_{s,n}$, $(x_{t,n}-x_{s,n})/d_{t,s,n}$ has a density $h_{t,s,n}(x)$ which is uniformly bounded by a constant K. Simple calculations show that, for $1\le s<t\le n$,

$$\begin{aligned}\mathbb{E}\big\{|l[c_n(x_{t,n}+x)]|\mid\mathcal{F}_{s,n}\big\} &= \int_{-\infty}^{\infty}|l[c_nd_{t,s,n}\,y+c_n(x_{s,n}+x)]|h_{t,s,n}(y)dy\\ &\le \frac{K}{c_nd_{t,s,n}}\int_{-\infty}^{\infty}|l[y+c_n(x_{s,n}+x)]|dy\\ &\le K\,l_1/(c_nd_{t,s,n}),\end{aligned}\tag{33}$$

where $l_1=\int_{-\infty}^{\infty}|l(x)|dx$. By virtue of this estimate, it follows from conditional arguments repeatedly that, for any $t_2\le k_1<k_2<...<k_m\le t_3$,

$$\begin{aligned}&\mathbb{E}\Big(\big|l[c_n(x_{k_1,n}+x)]\,...\,l[c_n(x_{k_m,n}+x)]\big|\mid\mathcal{F}_{n,t_1}\Big)\\ \le\;&\mathbb{E}\Big(\big|l[c_n(x_{k_1,n}+x)]\,...\,l[c_n(x_{k_{m-1},n}+x)]\big|\,\mathbb{E}\big(|l[c_n(x_{k_m,n}+x)]|\mid\mathcal{F}_{n,k_{m-1}}\big)\mid\mathcal{F}_{n,t_1}\Big)\\ \le\;&K\,l_1\,c_n^{-1}\,d_{k_m,k_{m-1},n}^{-1}\,\mathbb{E}\Big(\big|l[c_n(x_{k_1,n}+x)]\,...\,l[c_n(x_{k_{m-1},n}+x)]\big|\;\mid\mathcal{F}_{n,t_1}\Big)\\ \le\;&......\\ \le\;&(K\,l_1)^m\,c_n^{-m}\,d_{k_1,t_1,n}^{-1}\,d_{k_2,k_1,n}^{-1}\cdots d_{k_m,k_{m-1},n}^{-1}.\end{aligned}$$

Therefore, by recalling $d_{t,s,n} \sim C_0\,[(n/(t-s)]^{-d}$ for some $0 < d < 1$ and letting $H_0 = \max\{1, K\, l_0\, l_1\, C_0\}$ where $l_0 = \sup_x |l(x)|$, we have

$$
\begin{aligned}
& \mathbb{E}\Big(|\sum_{k=t_2}^{t_3} l[c_n\,(x_{k,n}+x)]|^m \mid \mathcal{F}_{n,t_1}\Big) \\
\leq\ & \max\{1, l_0^{m-1}\}\Big[\sum_{k_1=t_2}^{t_3} \mathbb{E}\Big(\big|l\big[c_n(x_{t,n}+x)\big]\big| \mid \mathcal{F}_{n,t_1}\Big) \\
& +2 \sum_{t_2\leq k_1<k_2\leq t_3} \mathbb{E}\Big(\big|l\big[c_n(x_{k_1,n}+x)\big]\,l\big[c_n(x_{k_2,n}+x)\big]\big| \mid \mathcal{F}_{n,t_1}\Big) + ... \\
& +m! \sum_{t_2\leq k_1<...<k_m\leq t_3} \mathbb{E}\Big(\big|l\big[c_n(x_{k_1,n}+x)\big]...\,l\big[c_n(x_{k_m,n}+x)\big]\big| \mid \mathcal{F}_{n,t_1}\Big)\Big] \\
\leq\ & H_0^m \Big[n^d\, c_n^{-1} \sum_{k_1=t_2}^{t_3} (k_1-t_1)^{-d} \\
& \qquad +2\, n^{2d}\, c_n^{-2} \sum_{t_2\leq k_1<k_2\leq t_3} (k_1-t_1)^{-d}(k_2-k_1)^{-d} + ... \\
& \qquad +m!\, n^{md}\, c_n^{-m} \sum_{t_2\leq k_1<...<k_m\leq t_3} (k_1-t_1)^{-d}(k_2-k_1)^{-d}...(k_m-k_{m-1})^{-d}\Big] \\
\leq\ & H_0^m\,(m+1)!\, n^d\, c_n^{-1}(t_3-t_1)^{1-d}\big[1+\big\{(t_3-t_2)^{1-d} n^d\, c_n^{-1}\big\}^{m-1}\big].
\end{aligned}
$$

This proves Lemma 4.1. □

We are now ready to prove the main results.

Proof of Theorem 3.1. Let

$$y_j = -[n^{m_0}] - 1 + j\,/m_n', \quad j = 0, 1, 2, ..., m_n, \tag{34}$$

where $m_n' = [(n/c_n)^{1/2} c_n^2]$ and $m_n = 2([n^{m_0}] + 1)m_n'$. It follows that

$$
\begin{aligned}
\sup_{|x|\leq n^{m_0}} \big|S_n(x)\big| \ \leq\ & \max_{0\leq j\leq m_n-1}\ \sup_{x\in[y_j,y_{j+1}]} \sum_{t=1}^{n} \big|g[c_n(x_{t,n}+x)] - g[c_n(x_{t,n}+y_j)]\big| \\
& + \max_{0\leq j\leq m_n} \big|\sum_{t=1}^{n} g[c_n(x_{t,n}+y_j)]\big| \\
:=\ & \lambda_{1n} + \lambda_{2n}.
\end{aligned} \tag{35}
$$

It follows from Assumption 3.1 that

$$\lambda_{1n} \;\; \le \;\; C\, n\, c_n \max_{0\le j\le m_n-1} |y_{j+1}-y_j| = O\big[(n/c_n)^{1/2}\big], \tag{36}$$

which yields $\lambda_{1n} = O(n/c_n)$, a.s., as $n/c_n \to \infty$.

We use Lemma 4.1 to estimate λ_{2n}. To this end, let $m = \log n$ in (32). It follows from Markov inequality and the Stirling approximation of $(m+1)!$ that, for any $M_0 \ge e^{m_0+3} H_0$ where H_0 is given as in Lemma 4.1,

$$\begin{aligned}
& P\Big(\max_{1\le j\le m_n}\Big|\sum_{t=1}^n g\big[c_n(x_t+y_j)\big]\Big| \ge M_0\,(n/c_n)\,\log n,\ i.o.\Big) \\
& \le \lim_{s\to\infty}\sum_{n=s}^\infty\sum_{j=1}^{m_n} P\Big(\Big|\sum_{t=1}^n g\big[c_n(x_{t,n}+y_j)\big]\Big|^m \ge \big[M_0\,(n/c_n)\,\log n\big]^m\Big) \\
& \le \lim_{s\to\infty}\sum_{n=s}^\infty \frac{m_n}{\big[M_0\,(n/c_n)\,\log n\big]^m}\max_{1\le j\le m_n}\mathbb{E}\Big|\sum_{t=1}^n g\big[c_n(x_t+y_j)\big]\Big|^m \\
& \le \lim_{s\to\infty}\sum_{n=s}^\infty \frac{m_n\, H_0^m (m+1)!}{\big[M_0\,\log n\big]^m} \\
& \le C\,\lim_{s\to\infty}\sum_{n=s}^\infty \frac{n^{m_0+1}\, H_0^m}{\big[M_0\,\log n\big]^m}\sqrt{2\pi\,(m+1)}\,\Big(\frac{m+1}{e}\Big)^{m+1} \\
& \le C\,\lim_{s\to\infty}\sum_{n=s}^\infty e^{-(m_0+3)\log n}\, n^{m_0+1}\,\log n \\
& \le C_1 \lim_{s\to\infty}\sum_{n=s}^\infty n^{-2} = 0.
\end{aligned} \tag{37}$$

This proves $\lambda_{2n} = O[(n/c_n)\log n]$, a.s. Taking the estimates of λ_{1n} and λ_{2n} into (35), we obtain the required (19).

To prove (20), we first write

$$\begin{aligned}
\sum_{k=1}^n g[c_n\,(x_{k,n}+x)] &= \sum_{k=1}^n g[c_n\,(x_{k,n}+x)] I(|x_{t,n}| \le n^m/2) \\
&\quad + \sum_{k=1}^n g[c_n\,(x_{k,n}+x)] I(|x_{t,n}| > n^m/2) \\
&:= \lambda_{1n}(x) + \lambda_{2n}(x)
\end{aligned} \tag{38}$$

It follows from (19) and $n \sup_{|x|>n^m} |g(c_n\, x)| = O[(n/c_n)\log n]$ that

$$\begin{aligned}\sup_{x\in R}|\lambda_{1n}(x)| &\le \sup_{|x|\le n^m}|\lambda_{1n}(x)| + \sup_{|x|>n^m}|\lambda_{1n}(x)| \\ &\le O[(n/c_n)\log n] + n \sup_{|x|>n^m/2}|g(c_n\, x)| \quad a.s. \\ &\le O[(n/c_n)\log n] \quad a.s., \end{aligned} \tag{39}$$

As for $\lambda_{2n}(x)$, we have

$$\begin{aligned}\sup_{x\in R}|\lambda_{2n}(x)| &\le C\sum_{t=1}^{n} I(|x_{t,n}| > n^m/2) \le Cn^{-mk/2}\sum_{t=1}^{n}|x_{t,n}|^k \\ &= O(1) \quad a.s., \end{aligned} \tag{40}$$

Combining (38)-(40), we prove (20). The proof of Theorem 3.1 is now complete. □

Proof of Theorem 3.2. It follows immediately from Theorem 3.3 and Remark 3.2. We omit the detalis.

Proof of Theorem 3.3. Let $\eta_n = (n/c_n)\log^{-\lambda} n$, $b_n = [n\log^{-\nu} n]$, where $\nu = (\lambda+1)/(1-d)$, and let T_n be the largest integer s such that $sb_n \le n$. Also write $y_j = -M_0/\log^\gamma n + j/m_n', j = 0, 1, 2, ..., m_n$, where $m_n' = [(n/c_n)^{1/2}c_n^2]$ and $m_n = [M_0 m_n'/\log^\gamma n] + 1$. It is readily seen that

$$n/b_n \sim \log^\nu n, \quad T_n b_n \le n, \quad m_n \le Cn^2, \tag{41}$$

due to $c_n \to \infty$ and $c_n/n \to 0$. Furthermore, by recalling the definitions of λ and η, tedious but elementary calculations show that, whenever $\gamma \ge 0$,

$$d\nu - \eta \le 1 - 2\lambda, \quad (d-1)\nu + \lambda \le -1, \quad 2d\nu - \gamma \le 1 - \lambda. \tag{42}$$

We now return to the proof of Theorem 3.3. Using the similar arguments as in the proof of (35), we have

$$
\begin{aligned}
& \sup_{|x|\le M_0/\log^{\gamma} n} |S_n(x) - S_n(0)| \\
\le & \sup_{|x|\le M_0/\log^{\gamma} n} |S_n(y_j) - S_n(0)| + O_{a.s.}[(n/c_n)^{1/2}] \\
\le & \max_{1\le j\le m_n} |\sum_{s=2}^{T_n-1} \Delta_{ns}(y_j)| + \max_{1\le j\le m_n} \Delta_n(y_j) + O_{a.s.}[(n/c_n)^{1/2}], \qquad (43)
\end{aligned}
$$

where, for $s = 1, ..., T_n$,

$$
\begin{aligned}
\Delta_{ns}(x) &= \sum_{t=sb_n+1}^{(s+1)b_n} \big(g[c_n(x_{n,t}+x)] - g(c_n\, x_{n,t})\big), \\
\Delta_n(x) &\le \Big(\sum_{t=1}^{2b_n} + \sum_{t=T_n b_n}^{n}\Big) \big|g[c_n(x_{n,t}+x)] - g(c_n\, x_{n,t})\big|.
\end{aligned}
$$

Recall $\eta_n = (n/c_n)\log^{-\lambda} n$. Using Theorem 3.1, it is readily seen that

$$
\begin{aligned}
\max_{1\le j\le m_n} \Delta_n(y_j) &\le C\big[(b_n + |n - T_n b_n|)/c_n\big]\log n \\
&\le C\,(n/c_n)\,\log^{1-\nu} n \le C\,\eta_n\,\log n, \quad a.s.
\end{aligned}
$$

This, together with (43), implies that (23) will follow if we prove

$$
\max_{1\le j\le m_n} \Big(|\sum_{\substack{s=2\\ s\in even}}^{T_n} \Delta_{ns}(y_j)| + |\sum_{\substack{s=2\\ s\in odd}}^{T_n} \Delta_{ns}(y_j)|\Big) = O_P(\eta_n\,\log n). \qquad (44)
$$

We only prove (44) for $s \in even$. The other is similar and hence the details are omitted. To this end, let $\mathcal{F}^*_{n,v} = \mathcal{F}_{n,(2v+1)b_n}$, $v \ge 0$, and $M_1 > 0$ is chosen later,

$$
\begin{aligned}
\Delta'_{ns}(x) &= \Delta_{n,2s}(x) I(|\Delta_{n,2s}(x)| \le M_1\,\eta_n), \\
\Delta^*_{ns}(x) &= \Delta'_{n,s}(x) - \mathbb{E}\big(\Delta'_{n,s}(x) \mid \mathcal{F}^*_{n,s-1}\big).
\end{aligned}
$$

Under these notation, to prove (44) for $s \in even$, it suffices to show

$$\begin{aligned}
\lambda_{1n} &:= \max_{1\le j\le m_n} |\sum_{s=1}^{T_n/2} \Delta^*_{ns}(y_j)| = O_P(\eta_n \log n), \qquad (45)\\
\lambda_{2n} &:= \max_{1\le j\le m_n} |\sum_{s=1}^{T_n/2} \mathbb{E}\big(\Delta_{n,2s}(y_j) \mid \mathcal{F}^*_{n,s-1}\big)| = O_P(\eta_n \log n), \qquad (46)\\
\lambda_{3n} &:= \max_{1\le j\le m_n} |\sum_{s=1}^{T_n/2} \Big(\Delta_{n,2s}(y_j) I(|\Delta_{n,2s}(y_j)| > M_1\,\eta_n)\\
&\qquad + \mathbb{E}\Big[\Delta_{n,2s}(y_j) I(|\Delta_{n,2s}(y_j)| > M_1\,\eta_n) \mid \mathcal{F}^*_{n,s-1}\Big]\Big)\\
&= O_P(\eta_n \log n). \qquad (47)
\end{aligned}$$

We start with (46). Note that, for any $2sb_n < t \le (2s+1)b_n$ and $|x| \le M_0/\log^{\gamma} n$ (letting $s_n = (2s-1)b_n$),

$$\begin{aligned}
&\Big|E\big[g[c_n(x_{t,n}+x)] - g(c_n\,x_{t,n})\Big|\mathcal{F}^*_{n,s-1}\big]\Big| = \Big|E\big[g[c_n(x_{t,n}+x)] - g(c_n\,x_{t,n})\Big|\mathcal{F}_{n,s_n}\big]\Big|\\
&= \Big|\int_{-\infty}^{\infty}\Big(g[c_n(x_{s_n,n}+d_{t,s_n,n}y+x)] - g[c_n\,(x_{s_n,n}+d_{t,s_n,n}y)]\Big)\,h_{t,s_n,n}(y)dy\Big|\\
&\le d^{-1}_{t,s_n,n}\int_{-\infty}^{\infty} g[c_n(y+x_{s_n,n})]\big|h_{t,s_n,n}[(y-x)/d_{t,s_n,n}] - h_{t,s_n,n}(y/d_{t,s_n,n})\big|\,dy\\
&\le C\,c_n^{-1} d^{-1}_{t,s_n,n}\min\{|x|d^{-1}_{t,s_n,n}, 1\} \le C\,|x|\,c_n^{-1}(n/b_n)^{2d}\\
&\le C\,c_n^{-1}\log^{2d\nu-\gamma} n, \qquad (48)
\end{aligned}$$

due to Assumption 3.3 and $d_{t,s,n} \sim C_0[n/(t-s))]^{-d}$. It is readily seen that

$$\begin{aligned}
\lambda_{2n} &\le \sum_{s=1}^{T_n/2}\max_{1\le j\le m_n} |\mathbb{E}\big(\Delta_{n,2s}(y_j) \mid \mathcal{F}^*_{n,s-1}\big)|\\
&\le C\,(n/c_n)\,\log^{2d\nu-\gamma} n = O_P(\eta_n \log n), \qquad (49)
\end{aligned}$$

due to (42), which yields (46).

Next for (47). Using Lemma 4.1 with $t_1 = 0, t_2 = 2sb_n + 1$ and $t_3 = (2s+1)b_n$, for any integer $m \ge 1$,

$$\begin{aligned}
\sup_x \mathbb{E}|\Delta_{n,2s}(x)|^m &\le H_0^m (m+1)!\,(n/c_n)\,\big\{1 + \big[(n/c_n)(n/b_n)^{d-1}\big]^{m-1}\big\}\\
&\le 2H_0^m(m+1)!(n/c_n)^m(n/b_n)^{(d-1)(m-1)},
\end{aligned}$$

whenever $(n/c_n)\log^{-(\lambda+1)} n \to \infty$. By virtue of this fact, we have

$$\begin{aligned} E\lambda_{3n} &\le 2\sum_{j=1}^{m_n}\sum_{s=1}^{T_n/2} \mathbb{E}\Delta_{n,2s}(y_j) I(|\Delta_{n,2s}(y_j)| > M_1\,\eta_n) \\ &\le 2\,m_n T_n\; H_0^m (m+1)!(n/c_n)\Big[\frac{(n/c_n)(n/b_n)^{d-1}}{M_1\,\eta_n}\Big]^{m-1} \\ &\le C\,n^4 (H_0/M_1)^m\,(m+1)!\log^{-(m-1)} n, \end{aligned}$$

due to (41) and $(d-1)\nu + \lambda \le -1$ by (42). Taking $m = \log n$ and letting $M_1 \ge 5H_0$, it follows from the Stirling approximation of $(m+1)!$ that

$$E\lambda_{3n} \;\le\; Cn^4 \log^5 n \exp\{-(M_1/H_0)\,\log n\} \le Cn^{-1}\log^5 n \to 0, \qquad (50)$$

which implies that $\lambda_{3n} = o_P(1)$. Hence (47) follows.

We finally consider (45). First note that, similarly to the proof of (48),

$$\begin{aligned} I_{k,j} &:= \Big|\mathbb{E}\Big(\{g[c_n(x_{n,j}+x)] - g[c_n x_{n,j}]\} \mid \mathcal{F}_{n,k}\Big)\Big| \\ &\le d_{j,k,n}^{-1}\int_{-\infty}^{\infty} g[c_n(x_{n,k}+y)]\big|h_{j,k,n}[(y-x)/d_{j,k,n}] - h_{j,k,n}(y/d_{j,k,n})\big|\,dy \\ &\le C\,c_n^{-1}\,d_{j,k,n}^{-1}\min\{|x| d_{j,k,n}^{-1}, 1\} \\ &\le C\,c_n^{-1}\,[n/(j-k)]^d \min\{|x|[n/(j-k)]^d, 1\}, \end{aligned}$$

for any $k < j$. This, together with (33), implies that, for any $|x| \le M_0/\log^\gamma n$,

$$
\begin{aligned}
& E[\Delta_{ns}^{*2}(x)|\mathcal{F}^*_{n,s-1}] \le 2E[\Delta^2_{n,2s}(x)|\mathcal{F}_{n,(2s-1)b_n}] \\
\le & \sum_{k=2sb_n+1}^{(2s+1)b_n} E\Big(\{\, g[c_n(x_{n,t}+x)] - g[c_n x_{n,t}]\,\}^2 \,|\, \mathcal{F}_{n,(2s-1)b_n}\Big) \\
& + 2 \sum_{2sb_n+1\le k<j\le(2s+1)b_n} \Big| E\Big(\{g[c_n(x_{n,k}+x)] - g[c_n x_{n,k}]\} \\
& \qquad \{g[c_n(x_{n,j}+x)] - g[c_n x_{n,j}]\} \Big| \mathcal{F}_{n,(2s-1)b_n}\Big)\Big| \\
\le & \, C\,(n/c_n)(n/b_n)^{d-1} + 2 \sum_{2sb_n+1\le k<j\le(2s+1)b_n} E\Big(|g[c_n(x_{n,k}+x)] - g[c_n x_{n,k}]|\,|I_{k,j}|\,\Big|\,\mathcal{F}_{n,(2s-1)b_n}\Big) \\
\le & \, C\,(n/c_n)(n/b_n)^{d-1} + C\, n^{2d}\, c_n^{-2}\, b_n^{-d} \sum_{2sb_n+1\le k<j\le(2s+1)b_n} (j-k)^{-d} \min\{n^d \log^{-\gamma} n (j-k)^{-d}, 1\} \\
\le & \, C\,(n/c_n)(n/b_n)^{d-1} + C\, n^{2d}\, c_n^{-2}\, b_n^{1-d} \sum_{k=1}^{b_n} k^{-d} \min\{(n/k)^d \log^{-\gamma} n, 1\} \\
\le & \, C\,(n/c_n)(n/b_n)^{d-1}\big[1 + (n/c_n)\log^{-\eta} n\big], \qquad (51)
\end{aligned}
$$

where we have used the fact: for $0 < d < 1$, letting $\zeta = \gamma/d$,

$$
\begin{aligned}
& \sum_{k=1}^{b_n} k^{-d} \min\{(n/k)^d \log^{-\gamma} n, 1\} \\
\le & \sum_{k=1}^{n/\log^\zeta n} k^{-d} + n^d \log^{-\gamma} n \sum_{k=n/\log^\zeta n+1}^{b_n} k^{-2d} \\
\le & \; C\, n^{1-d} \log^{-\eta} n.
\end{aligned}
$$

and η is given in (27). It follows from this estimate that

$$
\begin{aligned}
& \max_{0\le j\le m_n} \sum_{s=1}^{T_n/2} \mathbb{E}[\Delta^{*2}_{ns}(y_j) \mid \mathcal{F}^*_{n,s-1}] \\
\le & \; C\,(n/c_n)(n/b_n)^d \big[1 + (n/c_n)\log^{-\eta} n\big] \\
\le & \; C(n/c_n)^2 \log^{d\nu-\eta} n \le C\,\eta_n^2 \log n,
\end{aligned}
$$

due to (42) and $(n/c_n)\log^{-\eta} n \to \infty$. This, together with the facts that $|\Delta^*_{ns}(y_j)| \le \eta_n$ and for each j, $\{\Delta^*_{ns}(y_j), \mathcal{F}^*_{n,s}\}$ forms a martingale difference, it follows from the wellknown martingale exponential inequality (see, e.g., de la Pana (1999)) that, there exists a $M_0 \geq 3$ such that, as $n \to \infty$,

$$\begin{aligned}
& P[\lambda_{1n} \geq M_0 \eta_n \log n] \\
\leq\ & P\Big[\lambda_{1n} \geq M_0 \eta_n \log n, \ \max_{0 \leq j \leq m_n} \sum_{s=1}^{T_n/2} \mathbb{E}[\Delta_{ns}^{*2}(y_j) \mid \mathcal{F}_{n,s-1}^*] \leq C\, \eta_n^2 \log n\Big] + o(1) \\
\leq\ & \sum_{j=0}^{m_n} P\Big[\sum_{s=1}^{T_n/2} \Delta_{ns}^*(y_j) \geq M_0 \eta_n \log n, \ \sum_{s=1}^{T_n/2} \mathbb{E}[\Delta_{ns}^{*2}(y_j) \mid \mathcal{F}_{n,s-1}^*] \leq C\, \eta_n^2 \log n\Big] + o(1) \\
\leq\ & m_n \exp\Big\{ - \frac{M_0^2 \log^2 n}{2C \log n + 2M_0 \log n}\Big\} + o(1) \\
\leq\ & m_n \exp\{-M_0 \log n\} + o(1) \to 0, \qquad (52)
\end{aligned}$$

where the last inequality follows from (41). This yields $\lambda_{1n} = O_P\big(\eta_n \log n\big)$. Combining (49)-(52), we establish (44). □

REFERENCES

Bandi, F. (2004). On Persistence and nonparametric estimation (with an application to stock return predictability). *Unpublished manuscript.*

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* New York: Springer Verlag

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

Chen, J., Gao, J and Li, D. (2010). Semiparametric Regression Estimation in Null Recurrent Nonlinear Time Series. *Unpublished manuscript.*

Choi, I. and Saikkonen, P. (2004). Testing linearity in cointegrating smooth transition regressions. *The Econometrics Journal,* **7**, 341–365.

Choi, I. and Saikkonen, P. (2009). Tests for nonlinear cointegration. *Econometric Theory*, **26**, 682–709.

Gao, J., Li, D. and Tjøstheim, D. (2011). Uniform consistency for nonparametric estimates in null recurrent time series. *Working paper series No. 0085,* , The University of Adelaide, School of Economics.

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

Gao, J., K., Maxwell, Lu, Z., Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics*, **37**, 3893–3928.

Granger, C. W. J. and Teräsvirta, T. (1993), *Modelling nonlinear economic relationships*, Oxford University Press, New York.

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

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**, 726–748.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *The Annals of Statistics*, **29**, 372–416.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model, *The Annals of Statistics*, **35**, 252–299.

Kasparis, I. and Phillips, P.C.B. (2009). Dynamic misspecification in nonparametric cointegrating regression. Cowles foundation discussion paper No. 1700.

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

De La Pena, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *The Annals of Probability,* **27**, 537–564.

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics 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. (2011). Martingale limit theorems revisited and non-linear cointegrating regression. *Working Paper.*

Wang, Q. and Chan, N. (2011). Uniform convergence rates for a class of martingales with application in non-linear co-integrating regression. *Working paper.*

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

Wang, Q. and P. C. B. Phillips (2012). A Specification Test for Nonlinear Nonstationary Models. *The Annals of Statistics*, *forthcoming*.

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

Wang, Q. and Wang, R. (2011). Non-parametric cointegrating regression with NNH errors. *Econometric Theory*, *forthcoming*.

Appendix 1. Example 3.2: Verification of Assumption 3.3. Write

$$\begin{aligned} x_l &= \sum_{j=1}^{l} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} \\ &= x_k + \sum_{j=s+1}^{t} \sum_{i=-\infty}^{s} \epsilon_i \phi_{j-i} + \sum_{j=k+1}^{l} \sum_{i=s+1}^{j} \epsilon_i \phi_{j-i} \\ &:= x_{k,l}^* + x_{k,l}', \end{aligned} \tag{53}$$

The similar arguments as in the proof of Corollary 2.2 in Wang and Phillips (2009a) yields that $x_{k,l}'/d_{l-k}$, where d_n is defined as in (5), has a density $h_{l,k}(x)$ and $\int_{-\infty}^{\infty}(1+|t|)|\varphi_{l,k}(t)|dt < \infty$ uniformly for $0 \le k < l \le n$, where $\varphi_{l,k}(t) = Ee^{itx_{l,k}'/d_{l-k}}$, due to $\int(1+|t|)|Ee^{it\epsilon_0}|dt < \infty$. Hence, conditional on $\mathcal{F}_{k,n} = \sigma(\epsilon_j, -\infty < j \le k)$,

$$(x_{l,n} - x_{k,n})/d_{l,k,n} \text{ has a density } h_{l,k}(x - x_{k,l}^*/d_{l-k}) \tag{54}$$

where $x_{t,n} = x_t/d_n$ and $d_{l,k,n} = d_{l-k}/d_n$. Furthermore, for any $u \in R$, we have

$$\begin{aligned} & \sup_x \left| h_{l,k}(x - x_{k,l}^*/d_{l-k} + u) - h_{l,k}(x - x_{k,l}^*/d_{l-k}) \right| \\ \le\ & \sup_x |h_{l,k}(x+u) - h_{l,k}(x)| \\ \le\ & C \Big| \int_{-\infty}^{\infty} \big(e^{-it(x+u)} - e^{-itx}\big) \varphi_{l,k}(t) dt \Big| \\ \le\ & C \min\{|u|, 1\} \int_{-\infty}^{\infty} (1+|t|)\, |\varphi_{l,k}(t)| dt \le C_1 \min\{|u|, 1\}. \end{aligned}$$

That is, Assumption 3.3 is satisfied for $x_{t,n} = x_t/d_n$, where x_t is given as in Assumption 2.1. □

Appendix 2. Uniform convergence for a class of martingales. This result comes from Theorem 2.1 of Wang and Chan (2011). Let (u_k, x_k) with $x_k = (x_{k1}, ..., x_{kd}), d \ge 1$, be a sequence of random vectors.

Assumption 2.1. $\{u_t, \mathcal{F}_t\}_{t\ge1}$ is a martingale difference, where $\mathcal{F}_t = \sigma(x_1, ..., x_{t+1}, u_1, ..., u_t)$, satisfying $E(u_t^2 \mid \mathcal{F}_{t-1}) \to_{a.s.} \sigma^2 < \infty$ and $\sup_{t\ge1} E|u_t|^{2p} < \infty$ for some $p \ge 1$ specified in Assumption 2.4 below.

Assumption 2.2. $f(x)$ is a real function on R^d satisfying $\sup_{x\in R^d} |f(x)| < \infty$ and $|f(x) - f(y)| \le C\,\|x-y\|$ for all $x, y \in R^d$ and some constant $C > 0$.

Assumption 2.3. There exist positive constant sequences $c_n \uparrow \infty$ and b_n with

$b_n = O(n^k)$ for some $k > 0$ such that

$$\sup_{\|x\| \le b_n} \sum_{t=1}^{n} f^2[(x_t + x)/h] \;=\; O_P(c_n). \tag{55}$$

Assumption 2.4. $h \to 0, nh \to \infty$ and $n\, c_n^{-p} \log^{p-1} n = O(1)$, where c_n is defined as in Assumption 2.3 and p is defined as in Assumption 2.1.

We have the following main result.

Theorem 4.1. *Under Assumptions 2.1-2.4, we have*

$$\sup_{\|x\| \le b_n} \Big| \sum_{t=1}^{n} u_t\, f[(x_t + x)/h] \Big| \;=\; O_P\big[(c_n \log n)^{1/2}\big]. \tag{56}$$

If (55) is replaced by

$$\sup_{\|x\| \le b_n} \sum_{t=1}^{n} f^2[(x_t + x)/h] \;=\; O(c_n), \quad a.s., \tag{57}$$

the result (56) can be strengthened to

$$\sup_{\|x\| \le b_n} \Big| \sum_{t=1}^{n} u_t\, f[(x_t + x)/h] \Big| \;=\; O\big[(c_n \log n)^{1/2}\big], \quad a.s. \tag{58}$$

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.