Uniform convergence rates for a class of martingales with application in non-linear co-integrating regression

Qiying Wang and Nigel Chan

School of Mathematics and Statistics
The University of Sydney

May 21, 2012

Abstract

For a class of martingales, this paper provides a framework on the uniform consistency with broad applicability. The main condition imposed is only related to the conditional variance of the martingale, which holds true for stationary mixing time series, stationary iterated random function, Harris recurrent Markov chain and $I(1)$ processes with innovations being a linear process. Using the established results, this paper investigates the uniform convergence of the Nadaraya-Watson estimator in a non-linear cointegrating regression model. Our results not only provide sharp convergence rate, but also the optimal range for the uniform convergence to be held. This paper also considers the uniform upper and lower bound estimates for a functional of Harris recurrent Markov chain, which are of independent interests.

Key words and phrases: Harris recurrent Markov chain, Martingale, non-linearity, non-parametric regression, non-stationarity, uniform convergence.
MSC 2010: 62G20, 62G08, 60F99

1 Introduction

Let (u_k, x_k) with $x_k = (x_{k1}, ..., x_{kd}), d \geq 1$, be a sequence of random vectors. A common functional of interests $S_n(x)$ of (u_k, x_k) is defined by

$$S_n(x) = \sum_{k=1}^{n} u_k f[(x_k + x)/h], \quad x \in R^d, \tag{1.1}$$

where $h = h_n \to 0$ is a certain sequence of positive constants and $f(x)$ is a real function on R^d. Such functionals arise in nonparametric estimation problems, where f 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.

The uniform convergence of $S_n(x)$ in the situation that the (u_k, x_k) satisfy certain stationary conditions was studied in many articles. Liero (1989), Peligard (1991) and Ango Nze and Doukhan (2004) considered the uniform convergence over a fixed compact set, while Masry (1995), Bosq (1998) and Fan and Yao (2003) gave uniform results over an unbounded set. These work mainly focus on random sequence x_t which satisfies different types of mixing conditions. Investigating a more general framework, Andrew (1995) gave result on kernel estimate when the data sequence is near-epoch dependent on another underlying mixing sequence. More recently, Hansen (2008) provided a set of general uniform consistency results, allowing for stationary strong mixing multivariate data with infinite support, kernels with unbounded support and general bandwidth sequences. Kristensen (2009) further extended Hansen's results to the heterogenous dependent case under α-mixing condition. Also see Wu, Huang and Huang (2010) for kernel estimation in general time series settings.

In comparison to the extensive results where the x_k comes from a stationary time series data, there is little investigation on the the uniform convergence of $S_n(x)$ for the x_k being a non-stationary time series. In this regard, Gao, Li and Tjøstheim (2011) derived strong and weak consistency results for the case where the x_k is a null-recurrent Markov chain. Wang and Wang (2011) worked with partial sum processes of the type $x_k = \sum_{j=1}^{k} \xi_j$ where ξ_j is a general linear process. While the rate of convergence in Gao, etc (2011) is sharp, they impose the independence between u_k and x_k. Using a quite different method, Wang and Wang (2011) allowed for the endogeneity between u_k and x_k, but their results hold only for the x being in a fixed compact set.

The aim of this paper is to present a general uniform consistency result for $S_n(x)$ with broad applicability. As a framework, our assumption on the x_t is only related to the conditional variance of the martingale, that is, $\sum_{t=1}^{n} f^2[(x_t+x)/h]$. See Assumption 2.3 in Section 2. This of course is a "high level" condition, but it in fact is quite natural which holds true for many interesting and important examples, including stationary mixing time series, stationary iterated random function and Harris recurrent Markov chain. See Sections 2.2 and 2.3 for the identification of Assumption 2.3. This condition also holds true for $I(1)$ processes with innovations being a linear process, but the identification is complicated and requires quite different techniques. We will report related work in a separate paper. By using the established result, we investigate the uniform convergence of the Nadaraya-Watson estimator in a non-linear cointegrating regression model. It confirms that the uniform asymptotics in Wang and Wang (2011) can be extended to a unbounded set and the independence between the u_t and x_t in Gao, et al. (2011) can be removed. More importantly, our result not only provides sharp convergence rate, but also the optimal range for the uniform convergence to be held. It should be mentioned that our work on the uniform upper and lower bound estimation for a functional of Harris recurrent Markov chain is of independent interests.

This paper is organized as follows. Our main results are presented in next section, which includes the establishment of a framework on the uniform convergence for a class of martingale and uniform upper and lower bound estimation for a functional of Harris recurrent Markov chain. An application of the main results in non-linear cointegrating regression is given in Section 3. All 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. We also use the notation $\|x\| = \max_{1\le i\le d} |x_i|$.

2 Main results

2.1 Uniform convergence for a class of martingales

We make use of the following assumptions in the development of uniform convergence for the $S_n(x)$ defined by (1.1). Recall $x_k = (x_{k1}, ..., x_{kd})$ where $d \ge 1$ is an integer.

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

$$\sup_{\|x\|\le b_n} \sum_{t=1}^n f^2[(x_t + x)/h] \;=\; O_P(c_n). \tag{2.1}$$

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 remark that Assumption 2.1 ensures that $\{S_n(x), \mathcal{F}_n\}_{n\geq 1}$ is a martingale for each fixed x and is quite weak. Clearly, Assumption 2.1 is satisfies if u_t is a sequence of iid random variables, which is independent of $x_1, ..., x_t$, with $Eu_1 = 0$, $Eu_1^2 = \sigma^2$ and $E|u_1|^{2p} < \infty$. The Lipschitz condition used in Assumption 2.2 is standard in the investigation of uniform consistency, where we do not require the $f(x)$ has finite compact support. Assumption 2.3 is a "high level" condition for the x_k. We use it here to provide a framework. In Sections 2.2 and 2.3, we will show that this condition is in fact quite natural which holds true by many interesting and important examples. Assumption 2.4 provides the connections among the moment condition required in Assumption 2.1, the condition (2.1) and the bandwidth h. In many applications, we have $c_n = n^\alpha\, h^d\, l(n)$, where $0 < \alpha \leq 1$ and $l(n)$ is a slowly varying function at infinite. See Section 2.3 and Examples 1-3 in Section 2.2. In the typical situation that $c_n = n^\alpha\, h^d\, l(n)$, if there exists a $0 < \epsilon_0 < \alpha$ such that $n^{\alpha-\epsilon_0} h^d \to \infty$, the p required in Assumption 2.1 can be specified to $p = [1/\epsilon_0] + 1$.

We have the following main result.

THEOREM 2.1. *Under Assumptions 2.1-2.4, we have*

$$\sup_{\|x\|\leq b_n} \Big| \sum_{t=1}^n u_t\, f[(x_t + x)/h] \Big| \;=\; O_P\big[(c_n \log n)^{1/2}\big]. \tag{2.2}$$

If (2.1) is replaced by

$$\sup_{\|x\|\leq b_n} \sum_{t=1}^n f^2[(x_t + x)/h] \;=\; O(c_n), \quad a.s., \tag{2.3}$$

the result (2.2) can be strengthened to

$$\sup_{\|x\|\leq b_n} \Big| \sum_{t=1}^n u_t\, f[(x_t + x)/h] \Big| \;=\; O\big[(c_n \log n)^{1/2}\big], \quad a.s. \tag{2.4}$$

Theorem 2.1 can be extended to uniform convergence for the $S_n(x) = \sum_{t=1}^n u_t\, f[(x_t + x)/h]$ over unrestricted space R^d. This requires additional condition on the x_k and the tail decay for the function $f(x)$.

THEOREM 2.2. *In addition to Assumptions 2.1-2.4, $n \sup_{\|x\|>b_n/2} |f(x/h)| = O\big[(c_n \log n)^{1/2}\big]$ and there exists a $k_0 > 0$ such that*

$$b_n^{-k_0} \sum_{t=1}^n E||x_t||^{k_0} \;=\; O\big[(c_n \log n)^{1/2}\big]. \tag{2.5}$$

Then,

$$\sup_{x \in R^d} \Big| \sum_{t=1}^n u_t \, f[(x_t + x)/h] \Big| \;\; = \;\; O_P\big[(c_n \log n)^{1/2}\big]. \tag{2.6}$$

Similarly, if (2.1) is replaced by (2.3) and (2.5) is replaced by

$$b_n^{-k_0} \sum_{t=1}^n ||x_t||^{k_0} \;\; = \;\; O\big[(c_n \log n)^{1/2}\big], \quad a.s., \tag{2.7}$$

then

$$\sup_{x \in R^d} \Big| \sum_{t=1}^n u_t \, f[(x_t + x)/h] \Big| \;\; = \;\; O\big[(c_n \log n)^{1/2}\big], \quad a.s. \tag{2.8}$$

Remark 2.1. Theorems 2.1-2.2 allow for the x_t to be a stationary or non-stationary time series. See Examples 1-3 and Section 2.3 below. More examples on non-stationary time series will be reported in a separate paper. The rates of convergence in both theorems are sharp. For instance, in the well-known stationary situation such as those appeared in Examples 1-3, the c_n can be chosen as $c_n = nh$. Hence, when there are enough moment conditions on the u_t (i.e., p is large enough), we obtain the optimal rate $n^{2/5} \log^{3/5} n$, by taking $h \sim (\log n/n)^{1/5}$. In non-stationary situation, the rate of convergence is different. In particular we have $c_n = \sqrt{n}h$ for the x_t to be a random walk given in Corollary 2.1. The reason behind this fact is that the amount of time spent by the random walk around any particular point is of order $\sqrt{n}$ rather than n for a stationary time series. More explanation in this regard, we refer to Wang and Phillips (2009a, b).

2.2 Identifications of Assumption 2.3

This section provides several stationary time series examples which satisfy Assumption 2.3. Examples 1 and 2 come from Wu, et al. (2010), where more general settings on the x_t are established. Example 3 discusses a strongly mixing time series. This example comes from Hansen (2008). By making use of other related works such as Peligard (1991), Ango Nze and Doukhan (2004), Masry (1995), Bosq (1998) and Andrews (1995), similar results can be established for other mixing time series like ρ-mixing and near-epoch-dependent time series. In these examples, we only consider the situation that $d = 1$. The extension to $d > 1$ is straightforward and hence the details are omitted. Throughout Examples 1-3, we use the notation $f_h^2(x) = h^{-1} f^2(x/h)$.

Example on Harris recurrent Markov chain, which allows for stationary (positive recurrent) or non-stationary (null recurrent), is given in Section 2.3. In the section, we also consider the uniform lower bound, which is of independent interests. More examples on $I(1)$ processes with innovations being a linear process will be reported in a separate paper.

Example 1. Let $\{x_t\}_{t\geq 0}$ be a linear process defined by

$$x_t = \sum_{k=0}^{\infty} \phi_k \, \epsilon_{t-k},$$

where $\{\epsilon_j\}_{j\in Z}$ is a sequence of iid random variables with $E\epsilon_0^2 < \infty$ and a density p_ϵ satisfying $\sup_x |p_\epsilon^{(r)}(x)| < \infty$ and

$$\int_R |p_\epsilon^{(r)}(x)|^2 dx < \infty, \quad r = 0,\ 1, 2,$$

where $p_\epsilon^{(r)}(x)$ denotes the r-order derivative of $p_\epsilon(x)$. Suppose that $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$, and in addition Assumption 2.2, $f(x)$ has a compact support. It follows from Section 4.1 of Wu, et al. (2010) that, for any $h \to 0$ and $nh \log^{-1} n \to \infty$,

$$\sup_{x\in R} \Big| \frac{1}{n} \sum_{t=1}^{n} \big[f_h^2(x_t + x) - E f_h^2(x_t + x) \big] \Big| = O\Big[\sqrt{\frac{\log n}{nh}} + n^{-1/2} l(n) \Big] \quad a.s. \tag{2.9}$$

where $l(n)$ is a slowly varying function. Note that x_t is stationary process with a bounded density $g(x)$ under the given conditions on ϵ_k. Simple calculations show that

$$\sup_{x\in R} \sum_{t=1}^{n} f^2[(x_t + x)/h] = O_P(nh), \tag{2.10}$$

that is, x_t satisfies Assumption 2.3.

Example 2. Consider the nonlinear time series of the following form

$$x_k = R(x_{k-1}, \epsilon_k)$$

where R is a bivariate measurable function and ϵ_k are iid innovations. This is the iterated random function framework that encompasses a lot of popular nonlinear time series models. For example, if $R(x, \epsilon) = a_1 \, x \, I(x < \tau) + a_2 \, x \, I(x \geq \tau) + \epsilon$, it is the threshold autoregressive (TAR) model, (see Tong (1990)). If $R(x, \epsilon) = \epsilon \sqrt{a_1^2 + a_2^2 \, x}$, then it is autoregressive model with conditional heteroscedasticity (ARCH) model. Other nonlinear time series models, including random coefficient model, bilinear autoregressive model and exponential autoregressive model can be fitted in this framework similarly. See Wu and Shao (2004) for details.

In order to identify Assumption 2.3, we need some regularity conditions on the initial distribution of x_0 and the function $R(x, \epsilon)$. Define

$$L_\epsilon = \sup_{x \neq x'} \frac{|R(x, \epsilon) - R(x', \epsilon)|}{|x - x'|}. \tag{2.11}$$

Denote by $g(x|x_0)$ the conditional density of x_1 at x given x_0. Further let $g'(y|x) = \partial g(y|x)/\partial y$ and

$$I(x) = \Big[\int_R \Big|\frac{\partial}{\partial x} g(y|x)\Big|^2 dy\Big]^{1/2} \quad \textit{and} \quad J(x) = \Big[\int_R \Big|\frac{\partial}{\partial x} g'(y|x)\Big|^2 dy\Big]^{1/2} \tag{2.12}$$

$I(x)$ and $J(x)$ can be interpreted as a prediction sensitivity measure. These quantities measure the change in 1-step predictive distribution of x_1 with respect to change in initial value x_0. Suppose that

(i) there exist α and z_0 such that

$$E\big(|L_{\epsilon_0}|^\alpha + |R(z_0, \epsilon_0)|^\alpha\big) < \infty, \quad E[\log(L_{\epsilon_0})] < 0 \quad \textit{and} \quad EL_{\epsilon_0}^2 < 1;$$

(ii) $\sup_x [I(x) + J(x)] < \infty$;

(iii) in addition to Assumption 2.2, $f(x)$ has a compact support.

It follows from Section 4.2 of Wu, et al. (2010) that, for any $h \to 0$ and $nh \log^{-1} n \to \infty$

$$\sup_{x \in R} \Big|\frac{1}{n} \sum_{t=1}^n \big[f_h^2(x_t + x) - Ef_h^2(x_t + x)\big]\Big| = O\Big[\sqrt{\frac{\log n}{nh}} + n^{-1/2} l(n)\Big] \quad a.s. \tag{2.13}$$

where $l(n)$ is a slowly varying function. Note that x_t has a unique and stationary distribution under the given condition (i) and (ii). See Diaconis and Freedman (1999) for instance. Simple calculations show that

$$\sup_{x \in R} \sum_{t=1}^n f^2[(x_t + x)/h] = O_P(nh), \tag{2.14}$$

that is, x_t satisfies Assumption 2.3.

Example 3. Let $\{x_k\}_{k \geq 0}$ be a strictly stationary time series with density $g(x)$. Suppose that

(i) x_t is strongly mixing with mixing coefficients $\alpha(m)$ that satisfy $\alpha(m) \leq A m^{-\beta}$ where $\beta > 2$ and $A < \infty$;

(ii) $\sup_x |x|^q g(x) < \infty$ for some $q \geq 1$ satisfying $\beta > 2 + 1/q$ and there is some $j^* < \infty$ such that for all $j \geq j^*$, $\sup_{x,y} g_j(x, y) < \infty$ where $g_j(x, y)$ is the joint density of $\{x_0, x_j\}$;

(iii) in addition to Assumption 2.2, $f(x)$ has a compact support.

It follows from Theorem 4 (with $Y_i = 1$) of Hansen (2008) that, for any $h \to 0$ and $n^\theta h \log^{-1} n \to \infty$ with $\theta = \beta - 2 - 1/q$,

$$\sup_{x \in R} \Big| \frac{1}{n} \sum_{t=1}^{n} \big[f_h^2(x_t + x) - E f_h^2(x_t + x) \big] \Big| = O_P\Big[\sqrt{\frac{\log n}{nh}} \Big]. \tag{2.15}$$

If in addition $E|x_0|^{2q} < \infty$, the result (2.15) can be strengthened to almost surely convergence. Simple calculations show that

$$\sup_{x \in R} \sum_{t=1}^{n} f^2[(x_t + x)/h] = O_P(nh), \tag{2.16}$$

that is, x_t satisfies Assumption 2.3.

2.3 Uniform bounds for functionals of Harris recurrent Markov chain

Let $\{x_k\}_{k \geq 0}$ be a Harris recurrent Markov chain with state space $(E, \mathcal{E})$, transition probability $P(x, A)$ and invariant measure π. We denote P_μ for the Markovian probability with the initial distribution μ, E_μ for correspondent expectation and $P^k(x, A)$ for the k-step transition of $\{x_k\}_{k \geq 0}$. A subset D of E with $0 < \pi(D) < \infty$ is called D-set of $\{x_k\}_{k \geq 0}$ if for any $A \in \mathcal{E}^+$,

$$\sup_{x \in E} E_x \Big(\sum_{k=1}^{\tau_A} I_D(X_k) \Big) < \infty,$$

where $\mathcal{E}^+ = \{A \in \mathcal{E} : \pi(A) > 0\}$ and $\tau_A = \inf\{n \geq 1 : \ x_n \in A\}$. As is well-known, D-sets not only exist, but generate the entire sigma $\mathcal{E}$, and for any D-sets C, D and any probability measure ν, μ on $(E, \mathcal{E})$,

$$\lim_{n \to \infty} \sum_{k=1}^{n} \nu P^k(C) / \sum_{k=1}^{n} \mu P^k(D) \ = \ \frac{\pi(C)}{\pi(D)}, \tag{2.17}$$

where $\nu P^k(D) = \int_{-\infty}^{\infty} P^k(x, D) \nu(dx)$. See Nummelin (1984) for instance.

Let a D-set D and a probability measure ν on $(E, \mathcal{E})$ be fixed. Define

$$a(t) \ = \ \pi^{-1}(D) \sum_{k=1}^{[t]} \nu P^k(D), \quad t \geq 0.$$

By recurrence, $a(t) \to \infty$. By virtue of (2.17), the asymptotic order of $a(t)$ depends only on $\{x_k\}_{k\geq 0}$. As in Chen (2000), a Harris recurrent Markov chain $\{X_k\}_{k\geq 0}$ is called β-regular if

$$\lim_{\lambda\to\infty} a(\lambda t)/a(\lambda) \quad = \quad t^{\beta}, \quad \forall t > 0, \tag{2.18}$$

where $0 < \beta \leq 1$. It is interesting to notice that, under the condition (2.18), the function $a(t)$ is regularly varying at infinity, i.e., there exists a slowly varying function $l(x)$ such that $a(t) \sim t^{\beta}\, l(t)$. This implies that the definition of β-regular Harris recurrent Markov chain is similar to that of β-null recurrent given in Karlsen and Tjøstheim (2001) and Gao, et al. (2011), but it is more natural and simple.

The following theorem provides uniform upper and lower bounds for a functional of x_t. The upper bound implies that x_t satisfies Assumption 2.3, allowing for the x_t being stationary ($\beta = 1$, positive recurrent Markov chain) and non-stationary ($0 < \beta < 1$, null recurrent Markov chain). The lower bound plays a key role in the investigation of the uniform consistency for the kernel estimator in a non-linear co-integrating regression, and hence is of independent interests. See Section 3 for more details. Both upper and lower bounds are optimal, which is detailed in Remarks 2.2 and 2.3.

THEOREM 2.3. *Suppose that*

(i) $\{x_k\}_{k\geq 0}$ is a β-regular Harris recurrent Markov chain, where the invariant measure π has a bounded density function $p(s)$ on R;

(ii) in addition to Assumption 2.2, $\int_{-\infty}^{\infty} |f(x)|dx < \infty$.

Then, for any $h > 0$ satisfying $n^{-\epsilon_0} a(n)\, h \to \infty$ for some $\epsilon_0 > 0$, we have

$$\sup_{|x|\leq n^m} \sum_{k=1}^{n} f^2[(x_k + x)/h] = O_P[a(n)\, h]. \tag{2.19}$$

where m can be any finite integer.

For a given sequence of constants $b_n > 0$, if there exists a constant $C_0 > 0$ such that, uniformly for n large enough,

$$\inf_{|x|\leq b_n+1} \sum_{k=1}^{n} Ef^2[(x_k + x)/h] \geq a(n)\, h/C_0, \tag{2.20}$$

then, for any $h > 0$ satisfying $n^{-\epsilon_0} a(n)\, h \to \infty$ for some $\epsilon_0 > 0$, we have

$$\Big\{ \inf_{|x|\leq b_n} \sum_{k=1}^{n} f^2[(x_k + x)/h] \Big\}^{-1} = O_P\big\{[a(n)\, h]^{-1}\big\}. \tag{2.21}$$

Remark 2.2. The result (2.21) implies that, for any $0 < \eta < 1$, there exists a constant $C_\eta > 0$ such that

$$P\Big(\inf_{|x|\le b_n}\sum_{k=1}^n f^2[(x_k+x)/h] \ge a(n)\,h/C_\eta\Big) \ge 1-\eta. \tag{2.22}$$

This makes both bounds on (2.19) and (2.21) are optimal. On the other hand, since the result (2.22) implies that

$$E\inf_{|x|\le b_n}\sum_{k=1}^n f^2[(x_k+x)/h] \ge a(n)\,h\,(1-\eta)/C_\eta,$$

for any $0 < \eta < 1$, the condition (2.20) is close to minimal.

Note that random walk is a 1/2-regular Harris recurrent Markov chain. The following corollary on a random walk shows the range $|x| \le b_n$ can be taken to be optimal as well.

COROLLARY 2.1. *Let $\{\epsilon_j, 1 \le j \le n\}$ be a sequence of iid random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfying $\int_{-\infty}^{\infty}|\varphi(t)|dt < \infty$. Write $x_t = \sum_{j=1}^t \epsilon_j, t \ge 1$. If in addition to Assumption 2.2, $\int_{-\infty}^{\infty}|f(x)|dx < \infty$, then, for $h > 0$ and $n^{1/2-\epsilon_0}\,h \to \infty$ where $0 < \epsilon_0 < 1/2$, we have*

$$\sup_{|x|\le n^m}\sum_{k=1}^n f^2[(x_k+x)/h] \;=\; O_P(\sqrt{n}\,h), \tag{2.23}$$

for any integer $m > 0$, and

$$\Big\{\inf_{|x|\le M_0\sqrt{n}}\sum_{k=1}^n f^2[(x_k+x)/h]\Big\}^{-1} \;=\; O_P\big\{(\sqrt{n}\,h)^{-1}\big\}, \tag{2.24}$$

for a fixed $M_0 > 0$.

Remark 2.3. For a random walk x_t defined as in Corollary 2.1, it was shown in Wang and Phillips (2009a) that

$$\frac{1}{\sqrt{n}h}\sum_{t=1}^n f^2[(x_t+y_n)/h] \;\to_D\; \int f^2(s)ds\,L_W(1,y), \tag{2.25}$$

where $L_W(1,y)$ is a local time of a Brownian motion W_t, and $y = 0$ if $y_n/\sqrt{n} \to 0$ and $y = y_0$ if $y_n/\sqrt{n} \to y_0$. Since $L_W(1,y) \to 0$, in probability, as $y \to \infty$, it follows from (2.25) that the range $\inf_{|x|\le M_0\sqrt{n}}$ in (2.24) can not be extended to $\inf_{|x|\le b_n}$ where $b_n/\sqrt{n} \to \infty$.

Remark 2.4. As in Examples 1-3, we may obtain a better result if $\{x_t\}_{t\geq 0}$ is stationary (positive null recurrent) and satisfied certain other restrictive conditions. Indeed, Kristensen (2009) provided such a result.

Let $\{x_n\}_{n\geq 0}$ be a time-homogeneous, geometrically ergodic Markov Chain. Denote the 1-step transition probability by $p(y|x)$, such that $P(x_{i+1} \in A \,|\, x_i) = \int_A p(y|x)dy$. Also denote the i-step transition probability by $p_i(y|x)$, such that $p_i(y|x) = \int_R p(y|z)p_{i-1}(z|x)dz$. Since x_t is geometrically ergodic, it has a density $g(x)$. Further suppose that

(i) (strong Doeblin condition) there exists $s \geq 1$ and $\rho \in (0,1)$ such that for all $y \in R$,

$$p_s(y|x) \geq \rho\, g(y); \tag{2.26}$$

(ii) $\partial^r p(y|x)/\partial y^r$ exists and is uniformly continuous for all x, for some $r \geq 1$;

(iii) $\sup_y [\, g(y) + |y|^q g(y)] < \infty$ for some $q \geq 1$,

(iv) in addition to Assumption 2.2, $f(x)$ has a compact support.

It follows from Kristensen (2009) that, for any $h \to 0$ and $nh \to \infty$,

$$\sup_{x\in R} \Big| \frac{1}{nh} \sum_{t=1}^{n} f^2[(x_t + x)/h] - g(x) \int f^2(s)ds \Big| = O_P\Big[h^r + \sqrt{\frac{\log n}{nh}}\Big], \tag{2.27}$$

which yields (2.19) with $a(n) = n$ and (2.21) with $a(n) = n$ and $b_n = C_0$, where C_0 is a constant such that $\inf_{|x|\leq C_0} g(x) > 0$.

Remark 2.5. It is much more complicated if x_t is a null recurrent Markov chain, even in the simple situation that x_t is a random walk defined as in Corollary 2.1. In this regard, we have (2.25), but it is not clear at the moment if it is possible to establish a result like

$$\sup_{|x|\leq b_n} \Big| \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} f^2[(x_t + x)/h] - \int f^2(s)ds\, L_W(1,x) \Big| = O_P(c_n), \tag{2.28}$$

for some $b_n \to \infty$ and $c_n \to 0$. Note that (2.28) implies that

$$\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} f^2[(x_t + y)/h] \;\; \to_P \;\; \int f^2(s)ds\, L_W(1,0), \tag{2.29}$$

for any fixed y. This is a stronger convergence than that given in (2.25). Our experience show that it might not be possible to prove (2.28) without enlarging the probability space in which the x_t host.

3 Applications in non-linear cointegrating regression

Consider a non-linear cointegrating regression model:

$$y_t = m(x_t) + u_t, \quad t = 1, 2, ..., n, \tag{3.1}$$

where u_t is a stationary error process and x_t is a non-stationary regressor. Let $K(x)$ be a non-negative real function and set $K_h(s) = h^{-1}K(s/h)$ where $h \equiv h_n \to 0$. The conventional kernel estimate of $m(x)$ in model (3.1) is given by

$$\hat{m}(x) = \frac{\sum_{t=1}^n y_t K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)}. \tag{3.2}$$

The point-wise limit behavior of $\hat{m}(x)$ has currently been investigated by many authors. Among them, Karlsen, et al. (2007) discussed the situation where x_t is a recurrent Markov chain. Wang and Phillips (2009a, 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. In another paper, Wang and Phillips (2009b) considered the errors u_t to be serially dependent and cross correlated with the regressor x_t for small lags. Also, see Phillips and Park (1998), Karlsen and Tjøstheim (2001), Guerre (2004) and Bandi (2004) for related work on non-linear, non-stationary autoregressions.

This section provides a uniform convergence for the $\hat{m}(x)$ by making direct use of Theorems 2.1 and 2.3 in developing the asymptotics. For reading convenience, we list the assumptions as follows.

Assumption 3.1. (i) $\{x_k\}_{k\geq 0}$ is a β-regular Harris recurrent Markov chain defined as in Section 3, where the invariant measure π has a bounded density function $p(s)$ on R; (ii) $\{u_t, \mathcal{F}_t\}_{t\geq 1}$ is a martingale difference, where $\mathcal{F}_t = \sigma(x_1, ..., x_{t+1}, u_1, ..., u_t)$, satisfying $E(u_t^2 \mid \mathcal{F}_{t-1}) \to_{a.s.} \sigma^2 < \infty$ and $\sup_{t\geq 1} E(|u_t|^{2p} \mid \mathcal{F}_{t-1}) < \infty$, where $p \geq 1 + 1/\epsilon_0$ for some $0 < \epsilon_0 < \beta$.

Assumption 3.2. The kernel K satisfies that $\int_{-\infty}^{\infty} K(s)ds < \infty$, $\sup_x K(x) < \infty$ and for any $x, y \in R$,

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

Assumption 3.3. There exists a real positive function $g(x)$ such that

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

uniformly for some $0 < \alpha \leq 1$ 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 3.1 is similar to, but weaker than those appeared in Karlsen, et al. (2007), where the authors considered the point-wise convergence in distribution.

Assumption 3.2 is a standard condition on $K(x)$ as in the stationary situation. The Lipschitz condition on $K(x)$ is not necessary if we only investigate the point-wise asymptotics. See Remark 3.2 for further details.

Assumption 3.3 requires a Lipschitz-type condition in a small neighborhood of the targeted set for the functionals to be estimated. This condition is quite weak, which may host a wide set of functionals. Typical examples include that $m(x) = \theta_1 + \theta_2 x + ... + \theta_k x^{k-1}$; $m(x) = \alpha + \beta\, x^\gamma$; $m(x) = x(1 + \theta x)^{-1} I(x \geq 0)$; $m(x) = (\alpha + \beta\, e^x)/(1 + e^x)$.

We have the following asymptotic results.

THEOREM 3.1. *Suppose Assumptions 3.1-3.3 hold, $h \to 0$ and $n^{-\epsilon_0} a(n) h \to \infty$ where $0 < \epsilon_0 < \beta$ is given as in Assumption 3.1. It follows that*

$$\sup_{|x| \leq b_n'} |\hat{m}(x) - m(x)| = O_P \left\{ \left[a(n)h\right]^{-1/2} \log^{1/2} n + h^\alpha\, \delta_n \right\}, \tag{3.3}$$

where $b_n' \leq b_n$, $\delta_n = \sup_{|x| \leq b_n'} g(x)$ and b_n satisfies that

$$\inf_{|x| \leq b_n + 1} \sum_{k=1}^{n} EK[(x_k + x)/h] \geq a(n)\, h / C_0,$$

for some $C_0 > 0$ and all n sufficiently large. In particular, for the random walk x_t defined as in Corollary 2.1, we have

$$\sup_{|x| \leq b_n'} |\hat{m}(x) - m(x)| = O_P \left\{ \left(nh^2\right)^{-1/4} \log^{1/2} n + h^\alpha\, \delta_n \right\}, \tag{3.4}$$

where $b_n' \leq M_0 \sqrt{n}$ for a fixed $M_0 > 0$ and $\delta_n = \sup_{|x| \leq b_n'} g(x)$.

Remark 3.1. When a high moment exists on the error u_t, the ϵ_0 can be chosen sufficient small so that there are more bandwidth choices in practice. It is understandable that the results (3.3) and (3.4) are meaningful if only $h^\alpha \delta_n \to 0$, which depends on the tail of the unknown regression function $m(x)$, the bandwidth h and the range $|x| \leq b_n'$. When $m(x)$ has a light tail such as $m(x) = (\alpha + \beta\, e^x)/(1 + e^x)$, δ_n may be bounded by a constant. In this situation, the b_n' in (3.4) can be chosen to be $M_0 \sqrt{n}$ for some fixed $M_0 > 0$.

In contrast to Theorem 2.3 and Remark 2.3, this kind of range $|x| \le M_0\sqrt{n}$ might be optimal, that is, the b_n' cannot be extended to $b_n'/\sqrt{n} \to \infty$ to establish the same rate of convergence as in (3.4).

Remark 3.2. Both results (3.3) and (3.4) are sharp. However, a better result can be obtained if we are only interested in the point-wise asymptotics for $\hat{m}(x)$. For instance, as in Wang and Phillips (2009a, b) with minor modification, we may show that, for each x,

$$\hat{m}(x) - m(x) \;=\; O_P\left\{(nh^2)^{-1/4} + h^{\alpha}\right\}, \tag{3.5}$$

whenever x_t is a random walk defined as in Corollary 2.1. Furthermore $\hat{m}(x)$ has an asymptotic distribution that is mixing normal, under minor additional conditions. More details are referred to Wang and Phillips (2009a, b).

Remark 3.3. Wang and Wang (2011) established a similar result to (3.4) with the x_t being a partial sum of linear process, but only for the x being a compact support and imposing a bounded condition on u_t. The setting on the x_t in this paper is similar to that given in Gao, et al. (2011), but our result provides the optimal range for the uniform convergence holding true and removes the independence between the error u_t and x_t required by Gao, et al. (2011).

4 Proofs of main results

Proof of Theorem 2.1. We split the set $A_n = \{x : \|x\| \le b_n\}$ into m_n balls of the form

$$A_{nj} \;=\; \{x : \|x - y_j\| \le 1/m_n'\}$$

where $m_n' = [nh^{-1}/(c_n \log n)^{1/2}]$, $m_n = (b_n m_n')^d$ and y_j are chosen so that $A_n \subset \bigcup A_{nj}$. It follows that

$$\begin{aligned}
& \sup_{\|x\|\le b_n} \Big|\sum_{t=1}^n u_t\, f[(x_t+x)/h]\Big| \\
\le\;& \max_{0\le j\le m_n} \sup_{x\in A_{nj}} \sum_{t=1}^n |u_t| \,\big|f[(x_t+x)/h] - f[(x_t+y_j)/h]\big| \\
&\qquad + \max_{0\le j\le m_n} \Big|\sum_{t=1}^n u_t\, f[(x_t+y_j)/h]\Big| \\
:=\;& \lambda_{1n} + \lambda_{2n}.
\end{aligned} \tag{4.1}$$

Recalling the Assumption 2.2, it is readily seen that

$$\begin{aligned}\lambda_{1n} &\le \sum_{t=1}^{n}|u_t| \max_{0\le j\le m_n}\sup_{x\in A_{nj}} \big|f[(x_t+x)/h]-f[(x_t+y_j)/h]\big| \\ &\le C\,(hm_n')^{-1}\sum_{t=1}^{n}|u_t| \\ &\le C\,(c_n\log n)^{1/2}\,\frac{1}{n}\sum_{t=1}^{n}|u_t| = O[(c_n\log n)^{1/2}], \quad a.s., \end{aligned} \tag{4.2}$$

by the strong law of large number.

In order to investigate λ_{2n}, write $u_t' = u_t I[|u_t|\le (c_n/\log n)^{1/2}]$ and $u_t^* = u_t' - E(u_t' \mid \mathcal{F}_{t-1})$. Recalling $E(u_t\mid\mathcal{F}_{t-1})=0$ and $\sup_x|f(x)|<\infty$, we have

$$\begin{aligned}\lambda_{2n} &\le \max_{0\le j\le m_n}\big|\sum_{t=1}^{n} u_t^*\, f[(x_t+y_j)/h]\big| \\ &\quad + \max_{0\le j\le m_n}\big|\sum_{t=1}^{n}\big[|u_t-u_t'|+E(|u_t-u_t'|\mid\mathcal{F}_{t-1})\big]\,f[(x_t+y_j)/h]\big| \\ &\le \max_{0\le j\le m_n}\big|\sum_{t=1}^{n} u_t^*\, f[(x_t+y_j)/h]\big| + C\sum_{t=1}^{n}\big[|u_t-u_t'|+E(|u_t-u_t'|\mid\mathcal{F}_{t-1})\big] \\ &:= \lambda_{3n}+\lambda_{4n}. \end{aligned} \tag{4.3}$$

Routine calculations show that, under $\sup_{t\ge1}E(|u_t|^{2p}\mid\mathcal{F}_{t-1})<\infty$ and $n\,c_n^{-p}\log^{p-1}n = O(1)$,

$$\begin{aligned}\lambda_{4n} &\le \sum_{t=1}^{n}\Big[|u_t|I\{|u_t|>(c_n/\log n)^{1/2}\} + E\big(|u_t|I\{|u_t|>(c_n/\log n)^{1/2}\}|\mathcal{F}_{t-1}\big)\Big] \\ &\le C\,\big(\frac{c_n}{\log n}\big)^{(1-2p)/2}\sum_{t=1}^{n}\big[|u_t|^{2p}+E(|u_t|^{2p}\mid\mathcal{F}_{t-1})\big] \\ &\le C\,(c_n\log n)^{1/2}\,\frac{1}{n}\sum_{t=1}^{n}\big[|u_t|^{2p}+E(|u_t|^{2p}\mid\mathcal{F}_{t-1})\big] \\ &= O\big[(c_n\log n)^{1/2}\big], \qquad a.s., \end{aligned} \tag{4.4}$$

by the strong law of large number again.

We next consider λ_{3n}. Assumptions 2.1 and 2.3 imply that

$$\max_{0\le j\le m_n}\sum_{t=1}^{n} f^2[(x_t+y_j)/h]E[(u_t^*)^2\mid\mathcal{F}_{t-1}] = O_P(c_n).$$

For any $\eta>0$, there exists a $M_0>0$ such that

$$P\big(\max_{0\le j\le m_n}\sum_{t=1}^{n}\sigma_t^2 \ge M_0 c_n\big)\le \eta.$$

where $\sigma_t^2 = f^2[(x_t + y_j)/h]E[(u_t^*)^2 \mid \mathcal{F}_{t-1}]$, whenever n is sufficiently large. This, together with $|u_t^*| \le 2(c_n/\log n)^{1/2}$ and the well-known martingale exponential inequality (see, e.g., de la Pana (1999)), implies that, for any $\eta > 0$, there exists a $M_0 \ge 6d(k+3)$ (k is as in Assumption 2.3) such that, whenever n is sufficiently large,

$$\begin{aligned} & P[\lambda_{3n} \ge M_0(c_n \log n)^{1/2}] \\ \le\ & P\Big[\lambda_{3n} \ge M_0(c_n \log n)^{1/2},\ \max_{0\le j\le m_n} \sum_{t=1}^n \sigma_t^2 \le M_0 c_n\Big] + \eta \\ \le\ & \sum_{j=0}^{m_n} P\Big[\sum_{t=1}^n u_t^* f[(x_k + y_j)/h] \ge M_0(c_n \log n)^{1/2},\ \sum_{t=1}^n \sigma_t^2 \le M_0\, c_n\Big] + \eta \\ \le\ & m_n \exp\big\{-\frac{M_0^2\, c_n \log n}{6\, M_0 c_n}\big\} + \eta \le m_n n^{-M_0/6} + \eta \le 2\eta, \end{aligned} \tag{4.5}$$

where we have used the following fact:

$$m_n \ \le\ C[n^{k+1}\, h^{-1}/(c_n \log n)^{1/2}]^d \le C_1\, n^{(k+2)d},$$

as $c_n \to \infty$ and $nh \to \infty$. This yields $\lambda_{3n} = O_P\big[(c_n \log n)^{1/2}\big]$. Combining (4.1)-(4.5), we establish (2.2).

To prove (2.4), by checking (4.1)-(4.4), it suffices to show that

$$\lambda_{3n} \ =\ O\big[(c_n \log n)^{1/2}\big], \qquad a.s. \tag{4.6}$$

under the alternative condition (2.3). In fact, by virtue of (2.3), it follows that

$$\max_{0\le j\le m_n} \sum_{t=1}^n f^2[(x_t + y_j)/h]E[(u_t^*)^2 \mid \mathcal{F}_{t-1}] = O(c_n), \quad a.s.$$

Similarly to proof of (4.5), we have for sufficiently large M_0 ($M_0 \ge 6d(k+4)$, say),

$$\begin{aligned} & P[\lambda_{3n} \ge M_0(c_n \log n)^{1/2},\ i.o.] \\ =\ & P\Big[\lambda_{3n} \ge M_0(c_n \log n)^{1/2},\ \max_{0\le j\le m_n} \sum_{k=1}^n \sigma_k^2 \le M_0 c_n,\ i.o.\Big] \\ \le\ & \lim_{s\to\infty} \sum_{n=s}^\infty P\Big[\lambda_{3n} \ge M_0(c_n \log n)^{1/2},\ \max_{0\le j\le m_n} \sum_{k=1}^n \sigma_k^2 \le M_0 c_n\Big] \\ \le\ & \lim_{s\to\infty} \sum_{n=s}^\infty m_n \exp\big\{-\frac{M_0^2\, c_n \log n}{6\, M_0 c_n}\big\} \\ \le\ & C \lim_{s\to\infty} \sum_{n=s}^\infty n^{(k+2)d}\, n^{-M_0/6} = 0, \end{aligned} \tag{4.7}$$

which yields (4.6). The proof of Theorem 2.1 is now complete. □

Proof of Theorem 2.2. We only prove (2.6). It is similar to prove (2.8) and hence the details are omitted. We may write

$$\begin{aligned}\sum_{t=1}^{n} u_t\, f[(x_t+x)/h] &= \sum_{t=1}^{n} u_t\, f[(x_t+x)/h] I(\|x_t\| \le b_n/2) \\ &\quad + \sum_{t=1}^{n} u_t\, f[(x_t+x)/h] I(\|x_t\| > b_n/2) \\ &= \lambda_{5n}(x) + \lambda_{6n}(x), \quad \text{say.} \end{aligned} \tag{4.8}$$

It is readily seen from (2.2) and $n \sup_{\|x\|>b_n/2} |f(x/h)| = O\big[(c_n \log n)^{1/2}\big]$ that

$$\begin{aligned}\sup_{x\in R^d} |\lambda_{5n}(x)| &\le \sup_{\|x\|\le b_n} |\lambda_{5n}(x)| + \sup_{\|x\|>b_n} |\lambda_{5n}(x)| \\ &\le O_P\big[(c_n \log n)^{1/2}\big] + \sup_{\|x\|>b_n/2} |f(x/h)| \sum_{t=1}^{n} |u_t| \\ &\le O_P\big[(c_n \log n)^{1/2}\big],\end{aligned}$$

as $\frac{1}{n}\sum_{t=1}^{n} |u_t| = O(1), a.s.$ by the strong law. As for $\lambda_{6n}(x)$, we have

$$\begin{aligned} E\sup_{x\in R^d} |\lambda_{6n}(x)| &\le C \sum_{t=1}^{n} E\big[|u_t|\, I(\|x_t\| > b_n/2)\big] \\ &\le C \sum_{t=1}^{n} P(\|x_t\| > b_n/2) \le C\, b_n^{-k_0} \sum_{t=1}^{n} E\|x_t\|^{k_0} \\ &= O\big[(c_n \log n)^{1/2}\big],\end{aligned}$$

which yield $\sup_{x\in R^d} |\lambda_{6n}(x)| = O_P\big[(c_n \log n)^{1/2}\big]$. Taking these estimates into (4.8), we obtain (2.6). The proof of Theorem 2.2 is complete. □

Proof of Theorem 2.3. First assume there exists a $C \in \mathcal{E}^+$ such that

$$P(x, A) \ge bI_C(x)\,\nu(A), \quad x \in E, \quad A \in \mathcal{E}, \tag{4.9}$$

for some $b > 0$ and probability measure ν on $(E, \mathcal{E})$ with $\nu(C) > 0$. Under this addition assumption, Theorem 2.3 can be established by using the so-called split chain technique.

To this end, define new random variables $Y_0, Y_1, ...$ and $\bar{x}_0, \bar{x}_1, ...$ by

$$\begin{aligned} P(\bar{x}_0 \in A) &= \nu(A) \\ P(Y_n = 1 \mid \bar{x}_n = x) &= h(x) \\ P(Y_n = 0 \mid \bar{x}_n = x) &= 1 - h(x) \\ P(\bar{x}_{n+1} \in A \mid \bar{x}_n = x, Y_n = 1) &= \nu(A) \\ P(\bar{x}_{n+1} \in A \mid \bar{x}_n = x, Y_n = 0) &= \frac{P(x, A) - h(x)\nu(A)}{1 - h(x)}, \end{aligned}$$

where $h(x) = bI_C(x)$. As easily seen, $\{\bar{x}_n, Y_n\}_{n=0}^{\infty}$ is a Harris recurrent Markov chain with state space $E \times \{0, 1\}$ and $\{\bar{x}_n\}_{n=0}^{\infty}$ has the same transition probability $P(x, A)$ as those of $\{x_n\}_{n=0}^{\infty}$. Since our result is free of the initial distribution, $\{x_n\}_{n=0}^{\infty}$ can be assumed to be identical with $\{\bar{x}_n\}_{n=0}^{\infty}$, i.e., x_0 has the distribution ν.

Further define $\rho_0 = -1$,

$$\rho_k = \min\{i : i \geq \rho_{k-1}, Y_i = 1\}, \quad k = 1, 2, ...$$

$N(n) = \max\{k : \rho_k \leq n\}$, and

$$Z_j(x) = \sum_{k=\rho_{j-1}+1}^{\rho_j} f^2[(x_k + x)/h], \quad Z_{jn}(x) = \sum_{k=\rho_{j-1}\wedge n+1}^{\rho_j \wedge n} f^2[(x_k + x)/h],$$

for $j = 1,\ 2, ...$ It is well-known that the blocks

$$(x_{\rho_i+1}, ..., x_{\rho_{i+1}}), \quad i = 0, 1, 2, ...$$

are iid blocks, x_{ρ_i+1} having the distribution ν. Hence, for each h and x, $\{Z_j^*(x), \rho_j - \rho_{j-1}\}_{j=1}^{\infty}$, where $Z_j^*(x) = Z_j(x)$ or $Z_{jn}(x)$ is a sequence of iid random vectors. Furthermore, by recalling that π has a bounded density function $p(s)$, $\int_{-\infty}^{\infty} |f(x)|dx < \infty$ and $\sup_s |f(s)| < \infty$, we have

$$\begin{aligned} EZ_1(x) &= b \int_{-\infty}^{\infty} f^2[(s + x)/h]\pi(ds) \\ &= b\, h \int_{-\infty}^{\infty} f^2(s)\, p(-x + sh)ds \leq C^*\, h, \end{aligned} \tag{4.10}$$

for any $x \in R$ and

$$\sup_{x \in R} E|Z_1(x)|^{2k} \leq C\, h, \tag{4.11}$$

for any integer k. See Lemma 5.2 of Karlsen and Tjostheim (2001) or Lemma B.1 of Gao, et al (2011). We also have the following lemma.

LEMMA 4.1. *Suppose that $d_n \sim C_0 a(n)$, where $C_0 > 0$ is a constant, and all $y_j, j = 0, 1, ..., m_n$ are different, where $|y_j| \le n^{m_0}$ and $m_n \le n^{m_1}$ for some $m_0, m_1 > 0$. Then,*

$$R_n := \max_{0\le j\le m_n} \Big|\sum_{k=0}^{d_n} \big[Z_k^*(y_j) - EZ_k^*(y_j)\big]\Big| = O_P[n^{-\epsilon_0/4}\, a(n)\, h], \tag{4.12}$$

$$\Delta_n := \max_{0\le j\le m_n} E\Big|\sum_{k=0}^{d_n} \big[Z_k^*(y_j) - EZ_k^*(y_j)\big]\Big| = O[n^{-\epsilon_0/4}\, a(n)\, h], \tag{4.13}$$

where ϵ_0 is a constant such that $n^{-\epsilon_0} a(n)h \to \infty$.

Proof. Only consider $Z_k^*(x) = Z_k(x)$, as the situation that $Z_k^*(x) = Z_{kn}(x)$ is similar. To this end, write $\widetilde{Z}_i(y_j) = Z_i(y_j) I(|Z_i(y_j)| \le n^{-\epsilon_0/2}\, a(n)h)$ and $\widehat{Z}_i(y_j) = Z_i(y_j) I(|Z_i(y_j)| > n^{-\epsilon_0/2}\, a(n)\, h)$. We have

$$\begin{aligned} R_n &\le \max_{0\le j\le m_n} \big|\sum_{i=1}^{d_n} \big[\widetilde{Z}_i(y_j) - E\widetilde{Z}_i(y_j)\big]\big| + \max_{0\le j\le m_n} \sum_{i=1}^{d_n} \big[\widehat{Z}_i(y_j) + E\widehat{Z}_i(y_j)\big] \\ &:= R_{1n} + R_{2n}. \end{aligned} \tag{4.14}$$

By taking $k \ge (m_1+2)/\epsilon_0$ in (4.11) and noting $n^{-\epsilon_0} a(n)h \to \infty$, simple calculations show that

$$\begin{aligned} E\, R_{2n} &\le C\, m_n\, a(n) \max_{0\le j\le m_n} EZ_1(y_j) I(|Z_i(y_j)| > n^{-\epsilon_0/2} a(n)h) \\ &\le C_1\, a(n)\, h\, (n^{m_1+1-k\epsilon_0} h^{-1}) \le C_1\, a(n)\, h\, (nh)^{-1} \\ &\le C_2\, n^{-\epsilon_0/2} a(n)\, h, \end{aligned} \tag{4.15}$$

which yields $R_{2n} = O_P[n^{-\epsilon_0/2}\, a(n)\, h]$. As for R_{1n}, by using (4.11) with $k = 2$ and noting

$$Ee^{t(\widetilde{Z}_i(y_j) - E\widetilde{Z}_i(y_j))} \le 1 + \frac{t^2}{2} EZ_1^2(y_j) e^{2tn^{-\epsilon_0/2} a(n)h} \le e^{C_0\, t^2\, h},$$

for any $t \le (n^{-\epsilon_0/2} a(n)h)^{-1}$ and some $C_0 > 0$, the standard Markov inequality implies that

$$\begin{aligned} & P(R_{1n} \ge M\, n^{-\epsilon_0/4} a(n)h) \\ \le\ & C\, m_n \max_{0\le j\le m_n} P\Big(\big|\sum_{i=1}^{C_\epsilon\, a(n)} \big[\widetilde{Z}_i(y_j) - E\widetilde{Z}_i(y_j)\big]\big| \ge M n^{-\epsilon_0/4} a(n)h\Big) \\ \le\ & C\, m_n\, \exp(-M\, t n^{-\epsilon_0/4} a(n)h + C_\epsilon a(n) t^2 h) \\ \le\ & C\, m_n\, \exp(-M\, n^{\epsilon_0/4}/4) \to 0, \end{aligned} \tag{4.16}$$

as $n \to \infty$. Hence, $R_{1n} = O_P[n^{-\epsilon_0/4}\, a(n)\, h]$. Combining (4.14) – (4.16), we prove (4.12).

The proof of (4.13) is similar except more simpler. Indeed, by independence of $\widetilde{Z}_i(x)$, we obtain

$$\begin{aligned}\Delta_n &\le \max_{0\le j\le m_n} E\big|\sum_{i=1}^{d_n}\big[\widetilde{Z}_i(y_j)-E\widetilde{Z}_i(y_j)\big]\big| + 2\max_{0\le j\le m_n}\sum_{i=1}^{d_n} E\widehat{Z}_i(y_j)\\ &\le 2\max_{0\le j\le m_n} d_n^{1/2}\,[E\widetilde{Z}_1^2(y_j)]^{1/2} + C\,n^{-\epsilon_0/2}a(n)\,h\\ &\le C\,n^{-\epsilon_0/4}a(n)\,h,\end{aligned}$$

due to the fact:

$$E\widetilde{Z}_1^2(y_j) \;\le\; n^{-\epsilon_0/2}\,a(n)hEZ_1(y_j) \le C\,n^{-\epsilon_0/2}\,a(n)h^2.$$

The proof of Lemma 4.1 is complete. □

We are now ready to prove (2.19) and (2.21) under the additional condition (4.9).

(2.19) first. As in proof of (4.1) and (4.2), but letting $y_j = -[n^m]-1+j\,/m_n', j = 0,1,2,...,m_n$, where $m_n' = [nh^{-2}/a(n)]$ and $m_n = 2([n^m]+1)m_n'$, we have

$$\sup_{|x|\le n^m}\sum_{k=0}^{n} f^2[(x_k+x)/h] \;\le\; \max_{0\le j\le m_n}\sum_{k=0}^{n} f^2[(x_k+y_j)/h] + C\,a(n)\,h. \tag{4.17}$$

Note that

$$\sum_{k=0}^{n} f^2[(x_k+x)/h] \;\le\; \sum_{k=0}^{\rho_{N(n+1)}} f^2[(x_k+x)/h] = \sum_{i=1}^{N(n+1)} Z_i(x),$$

and $\big\{N(n)/a(n)\big\}_{n\ge1}$ is bounded in probability. See, e.g., Chen (2000). For each $\epsilon>0$, there exist $0<C_\epsilon, C_{1\epsilon}<\infty$ such that

$$P(C_{1\epsilon}a(n)\le N(n)\le C_\epsilon a(n)) \ge 1-\epsilon, \tag{4.18}$$

whenever n is sufficiently large. Consequently, for each $a>0, \epsilon>0$ and n large enough,

$$P\big(\max_{0\le j\le m_n}\sum_{k=0}^{n} f^2[(x_k+y_j)/h]\ge a\big) \;\le\; P\big(\max_{0\le j\le m_n}\sum_{i=1}^{C_\epsilon\,a(n)} Z_i(y_j)\ge a\big)+\epsilon.$$

This, together with (4.12) with $Z_k^*(x)=Z_k(x)$, implies (2.19) under (4.9), since

$$\max_{0\le j\le m_n}\sum_{i=1}^{C_\epsilon\,a(n)} Z_i(y_j) \;\le\; C_\epsilon\,a(n)\max_{0\le j\le m_n} EZ_1(y_j)+R_n = O_P[a(n)\,h].$$

We next consider (2.21) under (4.9). To this regard, let $y_j = -[b_n]-1+j\,/m_n', \quad j = 0,1,2,...,m_n$, where $m_n' = [n^{1+\epsilon_0/2}h^{-2}/a(n)]$ and $m_n = 2([b_n]+1)m_n'$. Since

$$\begin{aligned}&\max_{0\le j\le m_n-1}\sup_{x\in[y_j,y_{j+1}]}\sum_{t=1}^{n}\big|f^2[(x_t+x)/h]-f^2[(x_t+y_j)/h]\big|\\ &\qquad\le\; C\,n\,h^{-1}\max_{0\le j\le m_n-1}|y_{j+1}-y_j| \le Cn^{-\epsilon_0/2}\,a(n)h,\end{aligned}$$

it is readily seen that

$$\inf_{|x|\le b_n} \sum_{t=1}^{n} f^2[(x_t+x)/h] \ \ \geq \ \ \Delta_{1n} - O_P[n^{-\epsilon_0/2}\, a(n)h], \tag{4.19}$$

where $\Delta_{1n} = \inf_{1\le j\le m_n} \sum_{t=1}^n f^2[(x_t+y_j)/h]$. Furthermore, by recalling (4.18) and noting that

$$\sum_{k=0}^{n} f^2[(x_k+x)/h] \ \ \geq \ \ \sum_{k=0}^{\rho_{N(n)}\wedge n} f^2[(x_k+x)/h] = \sum_{i=1}^{N(n)} Z_{in}(x),$$

we have, for each $a>0, \epsilon>0$ and n large enough,

$$P\big(\Delta_{1n}\ge a\big) \ \ \geq \ \ P\Big(\inf_{0\le j\le m_n} \sum_{i=1}^{C_{1\epsilon}\, a(n)} Z_{in}(y_j)\ge a\Big) - \epsilon. \tag{4.20}$$

On the other hand, it follows from (4.12) with $Z_k^*(x) = Z_{kn}(x)$ that

$$\begin{aligned}\inf_{0\le j\le m_n} \sum_{i=1}^{C_{1\epsilon}\, a(n)} Z_{in}(y_j) \ \ &\geq \ \ C_{1\epsilon}\, a(n) \inf_{0\le j\le m_n} EZ_{1n}(y_j) - R_n \\ &\geq \ \ C_{1\epsilon}\, a(n) \inf_{0\le j\le m_n} EZ_{1n}(y_j) - O_P[n^{-\epsilon_0/4}\, a(n)h]. \end{aligned}\tag{4.21}$$

Combining (4.19)-(4.21), the result (2.21) under (4.9) will follow if we prove: there exists a $b_0>0$ such that

$$\inf_{0\le j\le m_n} EZ_{1n}(y_j) \ge b_0\, h, \tag{4.22}$$

for all n sufficiently large. To prove (4.22), first note that there exists a $b_1>0$ such that $EN^2(n)/a^2(n)\le b_1$. See Lemma 3.3 of Karlsen and Tjøstheim (2001) for instance. Therefore, by taking $d_n = [b_2 a(n)]+1$, where $b_2>b_1$ is chosen later, we have for some $b_0>0$,

$$\begin{aligned}\inf_{0\le j\le m_n} EZ_{1n}(y_j) \ \ &= \ \ \frac{1}{d_n}\inf_{0\le j\le m_n} E\sum_{i=1}^{d_n} Z_{in}(y_j) = \frac{1}{d_n}\inf_{0\le j\le m_n} E\sum_{t=1}^{\rho_{d_n}\wedge n} f^2[(x_t+y_j)/h] \\ &\geq \ \ \frac{1}{d_n}\inf_{0\le j\le m_n} E\Big(\sum_{t=1}^{n} f^2[(x_t+y_j)/h] - I(\rho_{d_n}\le n)\sum_{t=1}^{\rho_{d_n}} f^2[(x_t+y_j)/h]\Big) \\ &\geq \ \ \frac{1}{d_n}\Big(\inf_{|x|\le b_n+1} E\sum_{k=1}^{n} f^2[(x_t+x)/h] - M_n\Big) \\ &\geq \ \ \frac{1}{d_n}\big[a(n)h/C_0 - M_n\big] \\ &\geq \ \ b_0\, h\,, \end{aligned}$$

whenever n is sufficiently large, where we have used the condition (2.20) and the fact: it follows from (4.10), (4.13) and $\rho_{d_n} \le n$ if and only if $N(n) > d_n$ that

$$\begin{aligned} M_n &:= \max_{0\le j\le m_n} E\Big[I(\rho_{d_n}\le n)\sum_{t=1}^{\rho_{d_n}} f^2[(x_t+y_j)/h]\Big] \\ &= \max_{0\le j\le m_n} E\Big[I(\rho_{d_n}\le n)\sum_{i=1}^{d_n} Z_i(y_j)\Big] \\ &\le d_n \max_{0\le j\le m_n} EZ_1(y_j)\,P(N(n)\ge d_n) + \max_{0\le j\le m_n} E\Big|\sum_{i=1}^{d_n}\big[Z_i(y_j)-EZ_i(y_j)\big]\Big| \\ &\le b_2^{-1}C^* h\, a^{-1}(n)\,EN^2(n) + O[n^{-\epsilon_0/4}a(n)h] \\ &\le C_0^{-1}\,a(n)h/2 \end{aligned}$$

by choosing $b_2 = 3\,C_0\,b_1\,C^*$ and n sufficiently large. This proves (4.22) and also completes the proof of (2.21) under (4.9).

We now consider general situation. Let $0 < t < 1$ be fixed. Define a transition probability $P_t(x,A)$ on $(E,\mathcal{E})$ by

$$P_t(x,A) = (1-t)\sum_{k=1}^{\infty} t^{k-1}\,P^k(x,A), \quad x\in E, \quad A\in\mathcal{E}.$$

Let $\{\beta_n\}_{n\ge1}$ be an iid Bernoulli random variables with the common law

$$P(\beta_1 = 0) = t \quad \text{and} \quad P(\beta_1 = 1) = 1-t$$

and assume $\{\beta_n\}_{n\ge1}$ and $\{X_n\}_{n\ge0}$ are independent. Define a renewal sequence $\{\sigma(k)\}_{k\ge0}$ by

$$\sigma(0)=0 \quad \text{and} \quad \sigma(k)=\inf\{n : n\ge\sigma(k-1); \beta_n = 1\}, \quad k\ge1.$$

With these notations, $\{x_{\sigma(n)}\}_{n\ge0}$ is a Harris recurrent Markov chain with the invariant measure π. The transition probability $P_t(x,A)$ of $\{x_{\sigma(n)}\}_{n\ge0}$ satisfies the additional condition (4.9) and

$$a_t(n) := \pi(D)^{-1}\sum_{k=1}^{n}\nu P_t^k(D) \sim (1-t)^{1-\gamma}\,a(n).$$

See Chen (2000) for instance. By virtue of these facts, it follows from the first part proof of (2.19) that, for any fixed $m>0$ and $h>0$,

$$\sup_{|x|\le n^m}\sum_{k=1}^{\sigma(n)}\beta_k f^2[(x_k+x)/h] = \sup_{|x|\le n^m}\sum_{k=1}^{n} f^2[(x_{\sigma(k)}+x)/h] = O_P[a_t(n)\,h].$$

Now by noting $\sigma([\lambda n])/n \to_{a.s.} \lambda/(1-t)$ by the strong law and taking λ such that $\lambda/(1-t) \geq 1$, simple calculations show that

$$\sup_{|x|\leq n^m} \sum_{k=1}^{n} \beta_k f^2[(x_k+x)/h] \;\leq\; \sup_{|x|\leq n^m} \sum_{k=1}^{\sigma([\lambda n])} f^2[(x_{\sigma(k)}+x)/h] = O_P[a(n)\,h].$$

Similarly,

$$\sup_{|x|\leq n^m} \sum_{k=1}^{n} (1-\beta_k) f^2[(x_k+x)/h] = O_P[a(n)\,h],$$

and hence the result (2.19) under general situation follows.

The proof of (2.21) under general situation is similar and hence the details are omitted. □

Proof of Corollary 2.1. We first notice that

F. $x_k = \sum_{j=1}^{k} \epsilon_j$ is a Harris null recurrent Markov chain, satisfying (4.9), $a(t) = \sqrt{t}$ and the invariant measure π is the Lebesgue measure.

Due to the fact **F**, (2.23) follows immediately from Theorem 2.3.

To prove (2.24), by Theorem 2.3, it suffices to show that (2.20) holds true with $b_n = M_0\sqrt{n}$ and $a(n) = \sqrt{n}$. In fact, under the conditions of Corollary 2.1, $x_k/\sqrt{k}$ has a density $p_k(x)$, satisfying $\sup_x |p_k(x) - \phi(x)| \to 0$, as $k \to \infty$, where $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$, due to the central limit theorem. This implies that

$$\inf_{|x|\leq 3M_0} p_k(x) \geq \inf_{|x|\leq 3M_0} \phi(x) - \sup_x |p_k(x) - \phi(x)| \geq A_0 > 0,$$

for some $A_0 > 0$ and all sufficiently large k. Hence, for $n/2 < k \leq n$ and n sufficiently large, we have

$$\begin{aligned}
\inf_{|x|\leq M_0\sqrt{n}+1} Ef^2[(x_k+x)/h] &= \inf_{|x|\leq M_0\sqrt{n}+1} \int_{-\infty}^{\infty} f^2[(\sqrt{k}y+x)/h]\, p_k(y)dy \\
&\geq \frac{h}{\sqrt{k}} \inf_{|x|\leq M_0\sqrt{n}+1} \int_{-\infty}^{\infty} f^2(y)\, p_k[(yh-x)/\sqrt{k}]dy \\
&\geq \frac{h}{\sqrt{k}} \inf_{|x|\leq 3M_0} p_k(x) \int_{|y|\leq M_1} f^2(y)dy \\
&\geq \frac{A_0 h}{2\sqrt{n}} \int_{|y|\leq M_1} f^2(y)dy,
\end{aligned}$$

where M_1 is chosen such that $\int_{|y|\leq M_1} f^2(y)dy > 0$. Consequently, there exists a constant $C_0 > 0$ such that

$$\inf_{|x|\leq M_0\sqrt{n}+1} \sum_{k=1}^{n} Ef^2[(x_k+x)/h] \geq \inf_{|x|\leq M_0\sqrt{n}+1} \sum_{k=n/2}^{n} Ef^2[(x_k+x)/h] \geq \sqrt{n}\,h/C_0,$$

as required. The proof of Corollary 2.1 is now complete. □

Proof of Theorem 3.1. We may write $\hat{m}(x) - m(x)$ as

$$\begin{aligned} \hat{m}(x) - m(x) &= \frac{\sum_{t=1}^n u_t K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)} + \frac{\sum_{t=1}^n \left[m(x_t) - m(x)\right] K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)} \\ &:= \Theta_{1n}(x) + \Theta_{2n}(x). \end{aligned} \tag{4.23}$$

Note that, for any $|x| \le b_n'$, there exists a $C_0 > 0$ such that $K[(x_t - x)/h] = 0$ if $|x_t - x| \ge hC_0$. It follows from Assumption 3.3 that, whenever n is sufficiently large,

$$\sup_{|x|\le b_n'} |\Theta_{2n}(x)| \;\le\; C_1\,\delta_n \sup_{|x|\le b_n'} \frac{\sum_{t=1}^n |x_t - x|^\alpha\, K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} \le Ch^\alpha\,\delta_n.$$

This, together with (2.21) [taking $f^2(s) = K(s)$] in Theorem 2.3, implies that (3.3) will follow if we prove

$$\sup_{|x|\le b_n} \sum_{t=1}^n u_t K[(x_t - x)/h] \;=\; O_P\big[a(n)h\big]^{1/2} \log^{1/2} n\big]. \tag{4.24}$$

In fact, with $p \ge 1 + 1/\epsilon_0$ and $c_n = a(n)h \to \infty$, we have

$$n\, c_n^{-p} \log^{p-1} n \le (n^{-\epsilon_0} a(n)h)^{-1-1/\epsilon_0} n^{-\epsilon_0} \log^{p-1} n \to 0,$$

since $n^{-\epsilon_0}a(n)h \to \infty$. Now, by recalling (2.19), it is readily seen that the conditions of Theorem 2.1 hold for $f(x) = K(x)$ and $c_n = a(n)h$. The result (4.24) follows from (2.2) in Theorem 2.1. □

Acknowledgment: The authors thank Associate Editor, two referees and Professor Jiti Gao for helpful comments on the original version. Wang acknowledges the 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.wang@sydney.edu.au.

5 Reference

Andrew, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11**, 560–596.

Ango Nze, P. and Doukhan P. (2004). Weak dependence: Models and applications to econometrics *Econometric Theory* **20**, 995–1045.

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

Berman, S. M. (1969). Local times and sample function properties of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **137**, 277–299.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, 2nd ed. Lecture Notes in Statistics 110*, 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, X. (2000). On the limit laws of the second order for additive functionals of Harris recurrent Markov chains. *Probability Theory And Related Fields* **116**, 89–123.

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 41–76.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.*. Springer-Verlag.

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.

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. *Annal of Statistics* **29**, no. 2, 372–416.

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

Kristensen, D. (2009). Uniform convergence rates for kernel estimators with heterogeneous dependent data. *Econometric Theory* **25**, 1433–1445.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications.* **65**, 69–80.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates *Journal of Time Series Analysis.* **17**, 571–599.

Nummelin, E. (1984). General Irreducible Markov Chains and Non-negative Operaors. *Cambridge University Press,* Cambridge, England.

Peligrad, M. (1991). Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence conditions. *Stochastics and Stochastic Reports* **40**, 147–168.

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

Tong, H. (1990). Non-linear Time Series: A Dynamical System Approach. *Oxford University Press*

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

Wu, W. B., Huang, Y. and Huang, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications* **120**, 2412–2431.

Wu, W. B., Shao, X. (2004). Limit theorems for iterated random functions *Journal of Applied Probability* **41**, 425–436.