



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Uniform Consistency for Nonparametric
Estimators in Null Recurrent Time Series**

Jiti Gao, Degui Li and Dag Tjøstheim

September 2011

Working Paper 13/11

Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series

JITI GAO^{†,1}, DEGUI LI[†] AND DAG TJØSTHEIM[‡]

[†]*Monash University and* [‡]*University of Bergen*

Abstract

This paper establishes a suite of uniform consistency results for nonparametric kernel density and regression estimators when the time series regressors concerned are nonstationary null-recurrent Markov chains. Under suitable conditions, certain rates of convergence are also obtained for the proposed estimators. Our results can be viewed as an extension of some well-known uniform consistency results for the stationary time series case to the nonstationary time series case.

JEL subject classifications: C13, C14, C22

Keywords: β -null recurrent Markov chain, nonparametric estimation, rate of convergence, uniform consistency

¹Jiti Gao is from Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Email address: jiti.gao@monash.edu.

1. Introduction

As shown in the literature, uniform consistency for nonparametric kernel density and regression estimators is not only important in estimation theory, but also useful in deriving results in specification testing theory. Existing studies by many authors mainly focus on the case where the observed time series data satisfy some stationarity conditions. Such studies include Liero (1989), Roussas (1990), Liebscher (1996), Masry (1996), Bosq (1998), Fan and Yao (2003), Ould-Saïd and Cai (2005) and others. Most of the existing results basically focus on uniform convergence on fixed compact sets. In a recent paper by Hansen (2008), the author makes significant progress towards establishing uniform convergence on unbounded sets for a general class of nonparametric functionals when the time series data are stationary and α -mixing. Kristensen (2009) extends Hansen's result to the heterogeneous dependent case under an α -mixing condition. By contrast, there is little work on uniform consistency of nonparametric kernel estimators for nonstationary time series without any mixing condition.

Phillips and Park (1998) are among the first to study nonparametric estimation in an autoregression model with integrated regressors and they develop a local-time approach for the establishment of their asymptotic theory. In the same period, Karlsen and Tjøstheim (1998, 2001) independently establish nonparametric kernel estimation in the nonstationary case where the time series regressors are nonstationary null-recurrent Markov chains. The authors establish various asymptotic results. For the recent development of nonparametric and semiparametric estimation in nonstationary time series and diffusion models, we refer to Karlsen, Myklebust and Tjøstheim (2007, 2010), Bandi and Moloche (2008), Cai, Li and Park (2009), Wang and Phillips (2009a, 2009b), Chen, Li and Zhang (2010), Chen, Gao and Li (2011) and the references therein. In the field of model specification testing, Gao *et al* (2009a, 2009b) establish asymptotically consistent tests in both autoregression and co-integration cases. In addition, the supplementary material for the papers by Gao *et al* (2009a, 2009b) briefly discuss weak uniform consistency for a nonparametric kernel density estimator for the case where the time series involved follow a random walk process.

This paper systematically studies the strong and weak uniform consistency results for a class of nonparametric kernel density and regression estimators for the case where the time series data involved are nonstationary null-recurrent Markov chains. In the weak uniform consistency result discussed in Section 3, we obtain a sharp rate of convergence

of the order $O_P\left(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\right)$ when the regressors are β -null recurrent Markov processes (see Section 3 for the definition of $L_s(\cdot)$). Such sharp rate of convergence is comparable to $O_P\left(\sqrt{\frac{\log n}{nh}}\right)$, which is a conventional result for the rate of convergence for nonparametric kernel estimators in the stationary time series case. In the strong uniform consistency result discussed in Section 3, we obtain a rate of convergence of the order $o\left(\sqrt{\frac{1}{n^{\beta-\varepsilon_0}h}}\right)$ for some small $0 < \varepsilon_0 < \beta$. Such a rate would be very close to the sharp rate of convergence $O\left(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\right)$ when ε_0 is close to zero. The uniform consistency results established in this paper not only strengthen existing point-wise consistency results given in Karlsen and Tjøstheim (2001), but also naturally extend some corresponding results in Hansen (2008) for the stationary time series case.

The rest of the paper is organized as follows. Some basic definitions and results for Markov chains are introduced in Section 2. The main results are stated in Section 3. Applications of the main results to the density estimation, the Nadaraya–Watson kernel and the local linear estimation methods are given in Section 4. The conclusions are given in Section 5. Some basic results in Markov theory are summarized in Appendix A. All the proofs are given in Appendix B.

2. Some basic results for Markov chains

Let $\{X_t, t \geq 0\}$ be a ϕ -irreducible Markov chain with transition probability \mathbf{P} and state space $(\mathbb{E}, \mathcal{E})$. This means that there exists a nontrivial measure ϕ on \mathbb{E} such that each ϕ -positive set \mathbb{A} is communicating with the whole state space, that is,

$$\sum_{n=1}^{\infty} \mathbf{P}^n(x, \mathbb{A}) > 0, \text{ for all } x \in \mathbb{E} \text{ whenever } \phi(\mathbb{A}) > 0. \quad (2.1)$$

We assume that ϕ is maximal in the sense that if ϕ^* is another irreducible measure, then ϕ^* is absolutely continuous with respect to ϕ . In this paper, $\mathbb{E} \subset \mathbb{R}$. Denote the class of nonnegative measurable functions with ϕ -positive support by \mathcal{E}^+ . For a set $\mathbb{B} \in \mathcal{E}$, we write $\mathbb{B} \in \mathcal{E}^+$ if $1_{\mathbb{B}} \in \mathcal{E}^+$, where $1_{\mathbb{B}}$ is the indicator function of set \mathbb{B} . A function $\eta \in \mathcal{E}^+$ is said to be a small function if there exist a measure λ , a positive constant b and an integer $m \geq 1$, so that

$$\mathbf{P}^m \geq b\eta \otimes \lambda. \quad (2.2)$$

And if λ satisfies the above inequality for some $\eta \in \mathcal{E}^+$, $b > 0$ and $m \geq 1$, then λ is called a small measure. A set \mathbb{B} is small if $1_{\mathbb{B}}$ is a small function.

To make asymptotics for nonparametric estimation possible, we assume that the ϕ -irreducible Markov chain $\{X_t\}$ is Harris recurrent.

DEFINITION 2.1. *The chain $\{X_t\}$ is Harris recurrent if, given a neighborhood \mathbb{N}_v of v ($v \in \mathbb{E}$) with $\phi(\mathbb{N}_v) > 0$, $\{X_t\}$ returns to \mathbb{N}_v with probability one.*

It is well-known that for a Markov chain on a countable state space which has a point of recurrence, a sequence splitted by the regeneration times becomes independent and identically distributed (i.i.d.) by the Markov property (see, for example, Chung 1967). For a general Markov process which does not have an obvious point of recurrence, as in Nummelin (1984), the Harris recurrence allows one to construct a split chain which decomposes the partial sum of the Markov process $\{X_t\}$ into blocks of i.i.d. parts and the negligible remaining parts.

Let T_t only take the values 0 and 1, and $\{(X_t, T_t), t \geq 0\}$ be the split chain whose detailed construction will be provided in Appendix A. Define

$$\tau_k = \begin{cases} \inf\{t \geq 0 : T_t = 1\}, & k = 0, \\ \inf\{t > \tau_{k-1} : T_t = 1\}, & k \geq 1, \end{cases} \quad (2.3)$$

and denote the total number of regenerations in the time interval $[0, n]$ by $N(n)$, that is,

$$N(n) = \begin{cases} \max\{k : \tau_k \leq n\}, & \text{if } \tau_0 \leq n, \\ 0, & \text{otherwise.} \end{cases} \quad (2.4)$$

Let f be a real function defined in \mathbb{R} . We explain how to decompose the partial sum $S_n(f) = \sum_{t=0}^n f(X_t)$ into a sum of i.i.d. random variables with one main part and two asymptotically negligible minor parts. Define

$$Z_k = \begin{cases} \sum_{t=0}^{\tau_0} f(X_t), & k = 0, \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} f(X_t), & 1 \leq k \leq N(n), \\ \sum_{t=\tau_{N(n)}+1}^n f(X_t), & k = (n). \end{cases}$$

And it is easy to check that

$$S_n(f) = Z_0 + \sum_{k=1}^{N(n)} Z_k + Z_{(n)}. \quad (2.5)$$

From Nummelin (1984)'s result, we know that $\{Z_k, k \geq 1\}$ is a sequence of i.i.d. random variables. In the decomposition (2.5) of $S_n(f)$, $N(n)$ plays the role of the number of observations. It follows from Lemma 3.2 in Karlsen and Tjøstheim (2001) that Z_0 and $Z_{(n)}$ converge to zero almost surely when they are divided by $N(n)$.

Note that Harris recurrence only yields stochastic rates of convergence for the non-parametric estimation, where both the distribution and the number of regenerations $N(n)$ have no a priori known structure but fully depend on the underlying process. We next impose some restrictions on the tail behavior of the distribution of the recurrence time of the chain.

DEFINITION 2.2. *A Markov chain $\{X_t\}$ is β -null recurrent if there exist a small non-negative function $f(\cdot)$, an initial measure λ , a constant $\beta \in (0, 1)$ and a slowly varying function $L_f(\cdot)$ such that as $n \rightarrow \infty$*

$$\mathbf{E}_\lambda \left[\sum_{t=0}^n f(X_t) \right] \sim \frac{1}{\Gamma(1 + \beta)} n^\beta L_f(n), \quad (2.6)$$

where \mathbf{E}_λ stands for the expectation with initial distribution λ , $\Gamma(\cdot)$ is the usual Gamma function and $a_n \sim b_n$ means that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.

The β -null Harris recurrence restricts the tail behavior of the recurrence time of the process to be a regularly varying function (see, for example, Galambos and Seneta 1973). In particular, for a stationary or positive recurrent process, we have $\beta = 1$. We next provide two examples of $\frac{1}{2}$ -null recurrent Markov process.

EXAMPLE 2.1. Let a random walk process be defined as

$$X_t = X_{t-1} + u_t, \quad t = 1, 2, \dots, \quad X_0 = 0, \quad (2.7)$$

where $\{u_t\}$ is a sequence of i.i.d. random variables. Existing literature, such as Kallianpur and Robbins (1954), shows that $\{X_t\}$ defined by (2.7) is a $\frac{1}{2}$ -null recurrent Markov chain under weak conditions on the distribution of u_t .

EXAMPLE 2.2. Consider a parametric threshold autoregressive (TAR) model of the form

$$X_t = \alpha_1 X_{t-1} I_{\{X_{t-1} \in \mathbb{C}\}} + \alpha_2 X_{t-1} I_{\{X_{t-1} \in \mathbb{C}^c\}} + v_t, \quad (2.8)$$

where \mathbb{C} is a compact subset of \mathbb{R} , \mathbb{C}^c is the complement of \mathbb{C} , $\alpha_2 = 1$, $-\infty < \alpha_1 < \infty$, $\{v_t\}$ is assumed to be i.i.d. with $\mathbf{E}[v_1] = 0$, $0 < \mathbf{E}[v_1^2] < \infty$ and $\mathbf{E}[v_1^4] < \infty$, and the

distribution of $\{v_t\}$ is absolutely continuous with respect to Lebesgue measure with $f(\cdot)$ being the density function satisfying $\inf_{x \in \mathbb{S}} f(x) > 0$ for all compact sets \mathbb{S} . Recently, Gao, Tjøstheim and Yin (2010) have shown that $\{X_t\}$ generated by (2.8) is a $\frac{1}{2}$ -null recurrent Markov chain.

3. Main results

Let $\{e_t\}$ be a sequence of i.i.d. random variables and independent of $\{X_t\}$. Define a general nonparametric quantity of the form

$$\Phi_n(x) = \frac{1}{N(n)h} \sum_{t=0}^n L\left(\frac{X_t - x}{h}\right) e_t, \quad (3.1)$$

where $L(\cdot)$ is a kernel function satisfying Assumption A2(i) below and h is a bandwidth. To establish uniform consistency results for the nonparametric quantity $\Phi_n(x)$ defined by (3.1), we need the following assumptions.

ASSUMPTION A1 (i) The invariant measure of the β -null recurrent Markov chain $\{X_t\}$ has a uniformly continuous density function $p_s(\cdot)$ on \mathbb{R} with $\sup_{x \in \mathbb{R}} p_s(x) < \infty$.

(ii) Let $\{e_t\}$ be a sequence of i.i.d. random variables with $\mathbf{E}[e_t] = 0$ and $\mathbf{E}[e_t^2] < \infty$. In addition, $\{e_t\}$ is independent of $\{X_t\}$.

ASSUMPTION A2 (i) $L(\cdot)$ is nonnegative and has some compact support $C(L)$ and $\mathcal{N}_x(h) := \{y : \frac{1}{h}L(\frac{y-x}{h}) > 0\}$ is a small set. In addition, $L(\cdot)$ satisfies a Lipschitz-type condition of the form: $|L(x) - L(y)| \leq C_L |x - y|$ for all $x, y \in C(L)$ and some constant $C_L > 0$.

(ii) The bandwidth h satisfies

$$n^{\varepsilon_0} h \rightarrow 0 \quad \text{and} \quad n^{\beta - \varepsilon_0} h \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty. \quad (3.2)$$

for some $0 < \varepsilon_0 < \beta$.

REMARK 3.1. (i) Assumption A1(i) corresponds to the analogous conditions on the density function in the stationary time series case. Moreover, it can be verified when $\{X_t\}$ is generated by the random walk defined in Example 2.1. Nummelin (1984) shows in this case that the invariant density function $p_s(x) \equiv 1$. A1(ii) is imposed to make sure that the compound process $\{(X_t, e_t)\}$ is still β -null recurrent. The assumption that $\{e_t\}$ is independent of $\{X_t\}$ can be relaxed by allowing a heteroscedasticity structure of the form

$e_t = \sigma(X_t)\epsilon_t$, where $\{\epsilon_t\}$ is assumed to be independent of $\{X_t\}$ and $\sup_{x \in \mathbb{R}} |\sigma(x)| < \infty$. Another extension is that the i.i.d. condition on $\{\epsilon_t\}$ in A1(ii) might also be relaxed to accommodate the case where the error term is stationary and weakly dependent (such as α -mixing).

(ii) As discussed in condition B_2 in Section 5 of Karlsen and Tjøstheim (2001), A2(i) is needed in this kind of kernel estimation of null-recurrent time series. The small set requirement is a weak condition when combined with the compact support condition. For example, if $\{X_t\}$ is autoregressive given by $X_t = g(X_{t-1}) + x_t$, a sufficient condition for the smallness of $\mathcal{N}_x(h)$ is that $g(\cdot)$ is bounded on compact sets and that $\{x_t\}$ has density with respect to the Lebesgue measure and this density function is strictly positive on any compact set (see, for example, Doukhan and Ghindés 1980; Tjøstheim 1990). There are many other sufficient conditions that can be seen from Chapter 2.3 of Nummelin (1984). A2(ii) also imposes some mild conditions on the bandwidth parameter h for the null recurrent time series (cf. Karlsen, Myklebust and Tjøstheim 2007). When $\beta = 1$, our condition in (3.2) is slightly stronger than $h \rightarrow 0$ and $nh/\log n \rightarrow \infty$, which is commonly used for the stationary time series case.

In the stationary case, Hansen (2008) studies the uniform consistency results for a nonparametric estimate of the form

$$\Psi_*(\mathbf{x}) = \frac{1}{nh^d} \sum_{t=1}^n L_* \left(\frac{\mathbf{X}_t - \mathbf{x}}{h} \right) Y_t,$$

where $\{(\mathbf{X}_t, Y_t) : t \geq 1\}$ is a $(d+1)$ -dimensional vector of random variables and $L_*(\cdot)$ is a multivariate kernel function. Both weak and strong uniform consistency results are established in Theorems 2 and 3 of Hansen (2008). In Theorems 3.1 below, we establish a weak uniform consistency result for the nonparametric quantity defined by (3.1).

THEOREM 3.1. *Suppose that A1 and A2 hold. Let*

$$E[|e_1|^{4p_0}] < \infty \text{ for } p_0 = \left\lceil \frac{1+\beta}{\varepsilon_0} \right\rceil,$$

where $[x] \leq x$ is the largest integer part of x . Then,

$$\sup_{|x| \leq T_n} |\Phi_n(x)| = O_P \left(\sqrt{\frac{\log n}{n^\beta L_s(n) h}} \right), \quad (3.3)$$

where $T_n = M_0 n^{\beta-\varepsilon_0} L_s(n)$, in which $0 < \varepsilon_0 < \beta$, M_0 is any given positive constant, $L_s(\cdot)$ is chosen such that, for all small functions f , the asymptotic relation (2.6) holds with $L_f = \pi_s(f)L_s$, in which π_s will be defined in (A.2) in Appendix A.

REMARK 3.2. (i) Theorem 3.1 can be seen as an extension of the corresponding results in the stationary time series case to the nonstationary null recurrent time series case. When $\beta = 1$ and $L_s(\cdot)$ equals a non-zero constant, equation (3.3) reduces to some well-known results in the stationary case (see, for example, Theorem 2 in Hansen 2008). Taking $h \sim \left(\frac{\log n}{(n^\beta L_s(n))}\right)^{1/5}$, the right-hand side of (3.3) becomes $\left(\frac{\log n}{(n^\beta L_s(n))}\right)^{2/5}$, which reduces to an optimal rate in the stationary time series case when $\beta = 1$ and $L_s(\cdot)$ becomes a non-zero constant (see, for example, Stone 1980). The moment condition on $\{e_t\}$ implies that there exists a trade-off between the bandwidth condition and the moment condition on $\{e_t\}$. As ε_0 decreases and then the bandwidth condition becomes weaker, we need a stronger moment condition on $\{e_t\}$.

(ii) In particular, when $\beta = \frac{1}{2}$ with $L_s(\cdot)$ being a non-zero constant and $\varepsilon_0 \rightarrow 0$, the rate of convergence in (3.3) is close to $O_P\left(\sqrt{\frac{\log n}{\sqrt{n} h}}\right)$, which corresponds to $O_P\left(\sqrt{\frac{\log n}{n h}}\right)$ for the stationary time series case. This is mainly because in the $\frac{1}{2}$ -null recurrent case, the amount of time spent by the time series around any particular point is of order \sqrt{n} (see, for example, the random walk process defined in Example 2.1) rather than n for the stationary time series case.

In Theorem 3.2 below, we further establish a strong uniform rate of convergence under a slightly stronger condition on the moments of $\{e_t\}$.

THEOREM 3.2. *Let A1 and A2 hold. If, in addition,*

$$E[|e_1|^{2m_0}] < \infty \quad \text{with} \quad m_0 = 2 \left(\left\lceil \frac{\beta + 1}{\varepsilon_0} \right\rceil + 1 \right), \quad (3.4)$$

then,

$$\sup_{|x| \leq T_n} |\Phi_n(x)| = o\left(\frac{1}{\sqrt{n^{\beta - \varepsilon_0} h}}\right) \quad \text{a.s.}, \quad (3.5)$$

where ε_0 is defined as in A2(ii) and $T_n = M_0 n^{\beta - \varepsilon_0} L_s(n)$.

REMARK 3.3. Equation (3.5) can be viewed as a result corresponding to some existing results in the stationary time series case (see, for example, Theorem 3 of Hansen 2008). We can see that the rate of convergence in (3.5) is very close to the sharp rate obtained in Theorem 3.1 when ε_0 is close to zero. In this case, the moment condition (3.4) becomes stronger when ε_0 becomes smaller.

4. Applications in density and regression estimation

Define the kernel density estimator of the invariant density function $p_s(x)$ by

$$\widehat{p}_n(x) = \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right), \quad (4.1)$$

where $K(\cdot)$ is a kernel function. We next establish weak and strong uniform convergence rates for $\widehat{p}_n(x)$.

THEOREM 4.1. *Suppose that A1 and A2(ii) hold. Let $p_s(x)$ be thrice continuously differentiable with $\sup_{x \in \mathbb{R}} (|p'_s(x)| + |p''_s(x)| + |p'''_s(x)|) \leq C_p < \infty$. Suppose that $K(\cdot)$ has some compact support $\mathbb{C}(K)$ and satisfies the Lipschitz-type condition: $|K(x) - K(y)| \leq C_K |x - y|$ for all $x, y \in \mathbb{C}(K)$ and some constant $C_K > 0$. In addition, $K(\cdot)$ is a symmetric probability density function. Then, we have for $T_n = M_0 n^{\beta - \varepsilon_0} L_s(n)$ and n large enough*

$$\sup_{|x| \leq T_n} |\widehat{p}_n(x) - p_s(x)| = O_P(h^2) + O_P\left(\sqrt{\frac{\log n}{n^\beta L_s(n) h}}\right) \quad (4.2)$$

and

$$\sup_{|x| \leq T_n} |\widehat{p}_n(x) - p_s(x)| = O(h^2) + o\left(\frac{1}{\sqrt{n^{\beta - \varepsilon_0} h}}\right) \quad a.s. \quad (4.3)$$

REMARK 4.1. (i) The above theorem can be seen as an extension of Theorem 5.3 in Fan and Yao (2003) and Theorems 6 and 7 in Hansen (2008) from the stationary time series case to the nonstationary time series case. Karlsen and Tjøstheim (2001) obtain the point-wise consistency of $\widehat{p}_n(x)$ in the null recurrent time series case where

$$n^{\varepsilon_0} h \rightarrow 0 \quad \text{and} \quad n^{\frac{\beta}{2} - \varepsilon_0} h \rightarrow \infty \quad \text{for } 0 < \varepsilon_0 < \frac{\beta}{2}.$$

Theorem 4.1 not only weakens their bandwidth condition but also extends their point-wise consistency result to the uniform consistency result with possible rates.

(ii) The uniform consistency results in Theorem 4.1 may be thought to be of a somewhat academic character as $N(n)$ in the definition of (4.1) is not observable. However, it can be used in practice when $N(n)$ is linked with a directly observable hitting time. Indeed, if $\mathbb{C}_* \in \mathcal{E}^+$, the number of times that the process is visiting \mathbb{C}_* up to the time n is defined by $N_{\mathbb{C}_*}(n) = \sum_{t=0}^n I_{\mathbb{C}_*}(X_t)$. By Lemma 3.2 in Karlsen and Tjøstheim (2001), we have

$$\frac{N_{\mathbb{C}_*}(n)}{N(n)} \rightarrow \pi_s I_{\mathbb{C}_*} \quad a.s., \quad (4.4)$$

where π_s will be defined in (A.2) in Appendix A. Define

$$\widehat{p}_n^{\mathbb{C}_*}(x) = \frac{1}{N_{\mathbb{C}_*}(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right).$$

Note that

$$\widehat{p}_n^{\mathbb{C}_*}(x) = \frac{N(n)}{N_{\mathbb{C}_*}(n)} \left(\frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) \right) = \frac{N(n)}{N_{\mathbb{C}_*}(n)} \widehat{p}_n(x). \quad (4.5)$$

By (4.2)–(4.5) and noting that $\pi_s I_{\mathbb{C}_*} > 0$, we have

$$\sup_{|x| \leq T_n} |\widehat{p}_n^{\mathbb{C}_*}(x) - p_s(x)/(\pi_s I_{\mathbb{C}_*})| = O_P(h^2) + O_P\left(\frac{\sqrt{\log(n)}}{\sqrt{L_s(n)}} \cdot \frac{1}{\sqrt{n^\beta h}}\right) \quad (4.6)$$

and

$$\sup_{|x| \leq T_n} |\widehat{p}_n^{\mathbb{C}_*}(x) - p_s(x)/(\pi_s I_{\mathbb{C}_*})| = O(h^2) + o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s. \quad (4.7)$$

We now consider a nonlinear nonstationary regression model of the form

$$Y_t = m(X_t) + e_t, \quad 0 \leq t \leq n, \quad (4.8)$$

where $\{X_t\}$ is a β -null recurrent Markov chain, $\{e_t\}$ is a sequence of i.i.d. errors with $E[e_1] = 0$ and $0 < E[e_1^2] < \infty$, $m(\cdot)$ is an unknown function, and $\{e_t\}$ is independent of $\{X_t\}$. Such nonlinear cointegration models have been studied by several authors. For example, Karlsen, Myklebust and Tjøstheim (2007), and Wang and Phillips (2009a) consider estimating the regression function by the Nadaraya–Watson (NW) estimator of the form

$$\widehat{m}_n(x) = \sum_{t=0}^n w_{n,t}(x) Y_t \quad \text{with} \quad w_{n,t} = \frac{K\left(\frac{X_t - x}{h}\right)}{\sum_{s=0}^n K\left(\frac{X_s - x}{h}\right)}. \quad (4.9)$$

They then establish asymptotic distributions for $\widehat{m}_n(x)$ using different methods. As an application of Theorems 3.1 and 3.2, we establish rates for both the weak and strong uniform consistency results for the NW estimator $\widehat{m}_n(x)$ in Theorem 4.2 below.

THEOREM 4.2. *Assume that the conditions of Theorem 4.1 are satisfied. In addition, let $m(x)$ be twice continuously differentiable,*

$$\delta_n^2 n^{\beta-\varepsilon_0} h \rightarrow \infty, \quad h^2 \delta_n^{-1} \rightarrow 0, \quad \delta_{in}^* h^i \rightarrow 0 \quad \text{for } i = 1, 2, \quad (4.10)$$

where $\delta_n = \inf_{|x| \leq T_n} p_s(x) > 0$ and $\delta_{in}^* = \sup_{|x| \leq T_n} |m^{(i)}(x)| / \delta_n$ for $i = 1, 2$ and $T_n = M_0 n^{\beta-\varepsilon_0} L_s(n)$.

(i) *If, in addition, the moment condition on $\{e_t\}$ in Theorem 3.1 is satisfied, then we have*

$$\sup_{|x| \leq T_n} |\widehat{m}_n(x) - m(x)| = O_P\left(\frac{\sqrt{\log n}}{\delta_n \sqrt{n^\beta L_s(n)} h} + \delta_{2n}^* h^2\right) + o_P(\delta_{1n}^* h). \quad (4.11)$$

(ii) If, in addition, the moment condition on $\{e_t\}$ in Theorem 3.2 is satisfied, then we have

$$\sup_{|x| \leq T_n} |\widehat{m}_n(x) - m(x)| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0} h}} + \delta_{1n}^* h\right) + O(\delta_{2n}^* h^2) \quad a.s. \quad (4.12)$$

REMARK 4.2. (i) The conditions imposed for the establishment of Theorem 4.2 are reasonable and justifiable. We can show that the conditions in (4.10) can be easily verified in the case where the regressor $\{X_t\}$ is defined as in either Example 2.1 or Example 2.2. $p_s(x) \equiv 1$ in the first example and $p_s(x) \rightarrow 1$ as $|x| \rightarrow \infty$ in the second example, and thus the first two parts of (4.10) can be derived from (3.2) in Assumption A2 (ii). The last part of (4.10) imposes certain restrictions on the functional form of $m(\cdot)$. Several classes of functional forms of $m(\cdot)$ are included as long as $m(x)$ is of the form $m(x) = O(|x|^{1+\zeta})$ for some $0 < \zeta < 1$ when x is large enough. Particularly when $m(x) = a + bx$ and $\{X_t\}$ is generated by either Example 2.1 or Example 2.2, the last part of (4.10) is satisfied trivially.

(ii) Theorem 4.2 can be viewed as an extension of Theorem 3.3 in Bosq (1998) and Theorems 8 and 9 in Hansen (2008) from the stationary regression time series case to the nonstationary time series case. When $\{X_t\}$ is the random walk defined by Example 2.1, it is easy to check that (4.11) and (4.12) hold with $\delta_n = 1$, $\beta = \frac{1}{2}$ and $L_s(\cdot)$ being a positive constant.

We finally apply the local linear method for the estimation of $m(\cdot)$, and establish both the weak and strong uniform consistency results for the proposed local linear estimator. As in Fan and Gijbels (1996), the local linear estimator of $m(x)$ is defined by

$$\widetilde{m}_n(x) = \sum_{t=0}^n \widetilde{w}_{n,t}(x) Y_t, \quad \text{where } \widetilde{w}_{n,t}(x) = \frac{\widetilde{K}_{x,h}(X_t)}{\sum_{s=0}^n \widetilde{K}_{x,h}(X_s)}, \quad (4.13)$$

in which $\widetilde{K}_{x,h}(X_t) = \frac{1}{h} \widetilde{K}_n\left(\frac{X_t-x}{h}\right)$, $\widetilde{K}_n\left(\frac{X_t-x}{h}\right) = K\left(\frac{X_t-x}{h}\right) (S_{n,2}(x) - \left(\frac{X_t-x}{h}\right) S_{n,1}(x))$ with $S_{n,j}(x) = \frac{1}{N(n)h} \sum_{s=0}^n K\left(\frac{X_s-x}{h}\right) \left(\frac{X_s-x}{h}\right)^j$ for $j = 1, 2$.

The following theorem can be seen as an extension of Theorems 10 and 11 in Hansen (2008) from the stationary time series case to the nonstationary time series case.

THEOREM 4.3. *Let the conditions of Theorem 4.2 hold and $T_n = M_0 n^{\beta-\varepsilon_0} L_s(n)$.*

(i) *If the moment condition on $\{e_t\}$ in Theorem 3.1 is satisfied, then we have*

$$\sup_{|x| \leq T_n} |\widetilde{m}_n(x) - m(x)| = O_P\left(\frac{\sqrt{\log n}}{\delta_n \sqrt{n^\beta L_s(n) h}}\right) + O_P(\delta_{2n}^* h^2). \quad (4.14)$$

(ii) If the moment condition on $\{e_t\}$ in Theorem 3.2 is satisfied, then we have

$$\sup_{|x| \leq T_n} |\tilde{m}_n(x) - m(x)| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0} h}}\right) + O(\delta_{2n}^* h^2) \quad a.s. \quad (4.15)$$

REMARK 4.3. (i) Note that the first-order bias term involved in (4.14) and (4.15) is eliminated when the local-linear estimation method is employed. As a consequence, the class of functional forms for $m(x)$ is enlarged to include the case where $m(x) = O(|x|^{2+\zeta})$ for some $0 < \zeta < 1$ when x is large enough.

(ii) Note that the proofs of Theorems 4.2 and 4.3 show that the conclusions of Theorems 4.2 and 4.3 remain true in the case where $X_t = Y_{t-1}$, $\{Y_s\}$ and $\{e_t\}$ are independent for all $s < t$, and the functional form of $m(\cdot)$ is chosen such that $\{Y_t\}$ is a β -null recurrent Markov chain.

5. Conclusions

We have established several results for both the weak and strong uniform convergence with rates for some commonly-used nonparametric estimators in the case where the regressors are nonstationary null recurrent time series. Our main results have extended some existing uniform consistency results from the stationary time series case to the nonstationary time series case. In particular, we have obtained a sharp rate of convergence in the weak uniform consistency case. The established results are expected to be useful in deriving asymptotic theory for semiparametric estimation and specification testing for nonstationary null recurrent time series.

Note that, in this paper, we have only considered the case where $\{X_t\}$ is univariate. The main reason is that $\{X_t\}$ is not necessarily Harris recurrent when it is multivariate. When $\{X_t\}$ is a multivariate random walk process, for example, it is transient. For the case where a vector of independent univariate random walk regressors is involved, we refer to Schienle (2008), and Cai, Li and Park (2009).

6. Acknowledgments

The main ideas of this paper were based on the discussion of the authors during the visit by the third author to Australia in February 2009. The work of the authors was supported financially by two Australian Research Council Discovery Grants under Grant Numbers: DP0558602 and DP0879088.

REFERENCES

- Bandi, F. & G. Moloche (2008) On the functional estimation of multivariate diffusion processes. Working paper at the University of Chicago.
- Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed. Lecture Notes in Statistics 110. Springer–Verlag.
- Cai, Z., Q. Li & J. Park (2009) Functional–coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101–113.
- Chen, J., J. Gao & D. Li (2011) Estimation in semiparametric regression with nonstationary regressors. To appear in *Bernoulli*.
- Chen, J., D. Li & L. Zhang (2010) Robust estimation in a nonlinear cointegration model. *Journal of Multivariate Analysis* 101, 706–717.
- Chung, K. L. (1967) *Markov Chains with Stationary Transition Probabilities*. Springer–Verlag, 2nd Edition.
- Doukhan, P. & M. Ghindés (1980) Estimations dans le processus: “ $X_{n+1} = f(X_n) + \varepsilon_n$ ”. *C. R. Acad. Sci. Paris Sér. A–B* 291, A61–A64.
- Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.
- Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- Galambos, J. & E. Seneta (1973) Regularly varying sequences. *Proceedings of the American Mathematical Society* 41, 110–116.
- Gao, J. (2007) *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, London.
- Gao, J., M. L. King, Z. Lu & D. Tjøstheim (2009a) Specification testing in nonstationary time series autoregression. *Annals of Statistics* 7, 3893–3928.
- Gao, J., M. L. King, Z. Lu & D. Tjøstheim (2009b) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.
- Gao, J., D. Tjøstheim & J. Yin (2010) Estimation in threshold autoregressive models with a stationary and unit root regime. Available at www.adelaide.edu.au/directory/jiti.gao.
- Hansen, B. E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.
- Kallianpur, G. & H. Robbins (1954) The sequence of sums of independent random variables. *Duke Mathematical Journal* 21, 285–307.
- Karlsen, H. A. & D. Tjøstheim (1998) Nonparametric estimation in null recurrent time series. Discussion paper available at Sonderforschungsbereich 373 50, Humboldt University.

- Karlsen, H. A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
- Karlsen, H. A., T. Myklebust & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35, 252–299.
- Karlsen, H. A., T. Myklebust & D. Tjøstheim (2010) Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference* 140, 3619–3626.
- Kasahara, Y. (1984) Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *Journal of Mathematics of Kyoto University* 24, 521–538.
- Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogenous 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.
- Liero, H. (1989) Strong uniform consistency of nonparametric regression function estimates. *Probability Theory and Related Fields* 82, 587–614.
- Lin, G. (1998) On the Mittag–Leffler distributions. *Journal of Statistical Planning and Inference* 74, 1–9.
- 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 Operators*. Cambridge University Press.
- Ould–Saïd, E. & Z. Cai (2005) Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *Journal of Nonparametric Statistics* 17, 797–806.
- Phillips, P. C. B. & J. Park (1998) Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper 1181.
- Roussas, G. G. (1990) Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications* 36, 107–116.
- Schienle, M. (2008) Nonparametric nonstationary regression. Working paper available from the Department of Economics, University of Mannheim, Germany.
- Stone, C. J. (1980) Optimal rates of convergence for nonparametric estimators. *Annals of Statistics* 8, 1348–1360.
- Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
- van der Vaart, A. W. & J. Wellner (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer.

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

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

Appendix A: Useful results in Markov theory

To make this paper more self-contained, we summarize some useful terms and facts in Markov theory in this appendix. We adopt the same notation as used in Nummelin (1984) and Karlsen and Tjøstheim (2001). Let $\{X_t, t \geq 0\}$ be a Markov chain with transition probability \mathbb{P} and state space $(\mathbb{E}, \mathcal{E})$, and ϕ be a measure on $(\mathbb{E}, \mathcal{E})$.

Let η be a nonnegative measurable function and λ be a measure. We define the kernel $\eta \otimes \lambda$ by

$$\eta \otimes \lambda(x, \mathbb{A}) = \eta(x)\lambda(\mathbb{A}), \quad (x, \mathbb{A}) \in (\mathbb{E}, \mathcal{E}).$$

If K is a kernel, we define the function $K\eta$, the measure λK and the number $\lambda\eta$ by

$$K\eta(x) = \int K(x, dy)\eta(y), \quad \lambda K(\mathbb{A}) = \int \lambda(dx)K(x, \mathbb{A}), \quad \lambda\eta = \int \lambda(dx)\eta(x).$$

The convolution of two kernels K_1 and K_2 is defined by

$$K_1 K_2(v, \mathbb{A}) = \int K_1(v, dy)K_2(y, \mathbb{A}).$$

Let (2.2) hold. By Theorem 2.1 and Proposition 2.6 in Nummelin (1984), we know that for a ϕ -irreducible Markov chain, there exists a minorization inequality: there are a small function s , a probability measure ν and an integer $m_0 \geq 1$ such that $\mathbb{P}^{m_0} \geq s \otimes \nu$. As pointed out by Karlsen and Tjøstheim (2001), it causes some technical difficulties to have $m_0 > 1$ and it is not a severe restriction to assume $m_0 = 1$. So in this appendix, we always assume that the minorization inequality

$$\mathbb{P} \geq s \otimes \nu \tag{A.1}$$

holds with $\nu(\mathbb{E}) = 1$, $0 \leq s(x) \leq 1$, $x \in \mathbb{E}$.

As mentioned in Section 2, we will apply the so-called Markov chain splitting method of Karlsen and Tjøstheim (2001) to prove our results. In this method, an important role is played by the split chain under the minorization inequality (A.1). This allows for the decomposition of the chain into i.i.d. main parts and remaining parts that are asymptotically negligible. Denote

$$\mathbb{Q}(x, \mathbb{A}) = (1 - s(x))^{-1}(\mathbb{P}(x, \mathbb{A}) - s(x)\nu(\mathbb{A}))1(s(x) < 1) + 1_{\mathbb{A}}(x)1(s(x) = 1).$$

Then the transition probability $P(x, \mathbb{A})$ can be decomposed as

$$P(x, \mathbb{A}) = (1 - s(x))Q(x, \mathbb{A}) + s(x)\nu(\mathbb{A}).$$

When (A.1) holds, it can be verified that Q is a transition probability. As $0 \leq s(x) \leq 1$ and $\nu(\mathbb{E}) = 1$, P can be seen as a mixture of the transition probability Q and the small measure ν . Since ν is independent of x , the chain regenerates each time when ν is chosen with probability $s(x)$. For more details, we refer to Nummelin (1984). Now we introduce the split chain $\{(X_t, T_t), t \geq 0\}$, where $\{X_t\}$ is Harris recurrent and the auxiliary chain T_t only takes the values 0 and 1. Given $X_t = x$, $T_{t-1} = t_{t-1}$, T_t takes the value 1 with probability $s(x)$ and then the chain regenerates. Thus, $\alpha = \mathbb{E} \times \{1\}$ is a proper atom of the split chain. The distribution of $\{(X_t, T_t), t \geq 0\}$ is determined by its initial distribution λ , the transition probability P and (s, ν) . We use P_λ and E_λ for the distribution and expectation of the Markov chain with initial distribution λ . When $\lambda = \delta_x$ we write P_x instead of P_{δ_x} , which is the conditional distribution of $(T_0, \{(X_t, T_t), t \geq 1\})$ given $X_0 = x$. When $\lambda = \delta_\alpha(x, 1)$, i.e., $X_0 = x$ for arbitrary $x \in \mathbb{E}$ and $T_0 = 1$, then we write P_α and E_α . As shown in Karlsen and Tjøstheim (2001), if we let

$$\pi_s = \nu G_{s, \nu}, \quad \text{where } G_{s, \nu} = \sum_{n=0}^{\infty} (P - s \otimes \nu)^n, \quad (\text{A.2})$$

then $\pi_s = \pi_s P$, which implies that π_s is an invariant measure.

We then give some definitions of the stopping times of the Markov chain. Let

$$\tau = \min\{t \geq 0 : T_t = 1\} \quad (\text{A.3})$$

and

$$S_\alpha = \min\{t \geq 1 : T_t = 1\}. \quad (\text{A.4})$$

As $\{(X_t, T_t), t \geq 0\}$ is Harris recurrent, $P_\alpha(S_\alpha < \infty) = 1$. Let τ_k and $N(n)$ be defined as in (2.3) and (2.4), respectively. Following a standard result in Karlsen and Tjøstheim (2001), the number of regenerations $N(n)$ of the β -null recurrent Markov chain $\{X_t\}$ has the following asymptotic distribution

$$\frac{N(n)}{n^\beta L_s(n)} \xrightarrow{d} M_\beta(1) \quad (\text{A.5})$$

as $n \rightarrow \infty$, where L_s is defined in Theorem 3.1 and $M_\beta(1)$ is the Mittag-Leffler distribution with parameter β (cf., Kasahara 1984).

In addition, for a π_s -integrable function $g(\cdot)$ on \mathbb{R} , we have $\pi_s g = \int g(x) \pi_s(dx)$. Note that

$$G_{s, \nu} g(z) = E_z \left[\sum_{t=\tau_{k-1}+1}^{\tau_k} g(X_t) \right] = E_z \left[\sum_{t=0}^{\tau} g(X_t) \right], \quad (\text{A.6})$$

which, together with the definition of π_s , implies

$$\begin{aligned}
\pi_s g &= \int g(z) \pi_s(dz) = \int \left(\int g(z) G_{s,\nu}(x, dz) \right) \nu(dx) \\
&= \int \left(\int G_{s,\nu}(x, dz) g(z) \right) \nu(dx) = \int G_{s,\nu} g(x) \nu(dx) \\
&= \int \mathbb{E}_x \left[\sum_{t=0}^{\tau} g(X_t) \right] \nu(dx).
\end{aligned} \tag{A.7}$$

Equations (A.6) and (A.7) will be used in the proofs of our main results.

B: Proofs of the main results

To prove the main results in Sections 3 and 4, we need the following two lemmas.

LEMMA B.1. *Let*

$$g_h(x) = \sum_{t=\tau_{k-1}+1}^{\tau_k} L_{h,x}(X_t), \quad \text{with } L_{h,x}(X_t) = \frac{1}{h} L\left(\frac{X_t - x}{h}\right),$$

and $\{X_t\}$ be a β -null recurrent Markov process. If, in addition, A1(i) and A2 (i) are satisfied, then we have

$$\mathbb{E} \left[|g_h(x)|^{2m} \right] \leq M h^{-2m+1}, \tag{B.1}$$

where M is a positive constant which depends on m , but is independent of x and h .

PROOF. The main idea for the proof of (B.1) is similar to that for the proof of Lemma 5.2 in Karlsen and Tjøstheim (2001).

By the definition of $g_h(\cdot)$ and noting that the kernel function is nonnegative, we have

$$\begin{aligned}
\mathbb{E} \left[|g_h(x)|^{2m} \right] &= \mathbb{E}_\nu \left[\left(\sum_{t=0}^{\tau} L_{h,x}(X_t) \right)^{2m} \right] \\
&= \mathbb{E}_\nu \left\{ \sum_{t=0}^{\infty} \left[\prod_{s=0}^{t-1} I(T_s = 0) \right] L_{h,x}(X_t) \right\}^{2m} \\
&= \mathbb{E}_\nu \left[\sum_{t=0}^{\infty} \mathcal{B}_t L_{h,x}(X_t) \right]^{2m},
\end{aligned} \tag{B.2}$$

where $\mathcal{B}_t = \prod_{s=0}^{t-1} I(T_s = 0)$.

Letting

$$\Lambda_{2m,j} = \left\{ l = (l_1, \dots, l_j) \in (\mathbb{N}^+)^j : \sum_{k=1}^j l_k = 2m \right\}, \quad 1 \leq j \leq 2m,$$

where \mathbb{N}^+ is the set of positive integers, it is easy to check that

$$\mathbb{E}_\nu \left[\sum_{t=0}^{\infty} \mathcal{B}_t L_{h,x}(X_t) \right]^{2m} = \sum_{j=1}^{2m} \sum_{l \in \Lambda_{2m,j}} \frac{(2m)!}{l_1! \cdots l_j!} \mathbb{E}_\nu \left[\tilde{L}_{j,l}(x) \right], \quad (\text{B.3})$$

where

$$\tilde{L}_{j,l}(x) = \sum_{t_1=0}^{\infty} \sum_{t_2=t_1+1}^{\infty} \cdots \sum_{t_j=t_{j-1}+1}^{\infty} \mathcal{B}_{t_1} \mathcal{B}_{t_2} \cdots \mathcal{B}_{t_j} L_{h,x}^{l_1}(X_{t_1}) L_{h,x}^{l_2}(X_{t_2}) \cdots L_{h,x}^{l_j}(X_{t_j}).$$

We first consider the case of $j \geq 2$. Following the calculations in Karlsen and Tjøstheim (2001) (see also the more detailed version of Karlsen and Tjøstheim 1998), we have (omitting the identity function in the sequel)

$$\begin{aligned} \mathbb{E}_\nu \left[\tilde{L}_{j,l}(x) \right] &= \nu G_{s,\nu} \check{I}_{L_{h,x}^{l_1}} G_{s,\nu} \check{I}_{L_{h,x}^{l_2}} \cdots G_{s,\nu} \check{I}_{L_{h,x}^{l_j}} \\ &\leq \nu G_{s,\nu} \tilde{I}_{L_{h,x}^{l_1}} G_{s,\nu} \tilde{I}_{L_{h,x}^{l_2}} \cdots G_{s,\nu} \tilde{I}_{L_{h,x}^{l_j}}, \end{aligned} \quad (\text{B.4})$$

where $2 \leq j \leq 2m$, $l \in \Lambda_{2m,j}$, $G_{s,\nu}$ is defined in (A.2),

$$\check{I}_f(y, dz) = (\mathbb{P} - s \otimes \nu)(y, dz) f(y) \quad \text{and} \quad \tilde{I}_f(y, dz) = \mathbb{P}(y, dz) f(y).$$

By the compactness of $L(\cdot)$, the definition of $L_{h,x}(\cdot)$ and Remark 5.1 in Karlsen and Tjøstheim (2001), we have for $2 \leq k \leq j$,

$$G_{s,\nu} \tilde{I}_{L_{h,x}^{l_k}} \leq M_{j,l}(k) h^{-l_k}, \quad (\text{B.5})$$

where $M_{j,l}(k)$ is a positive constant independent of x .

Meanwhile, by A1(i) and A2(i), there exists a positive constant $M_{j,l}(1)$, independent of x , such that

$$\nu G_{s,\nu} \tilde{I}_{L_{h,x}^{l_1}} = \pi_s L_{h,x}^{l_1} \leq M_{j,l}(1) h^{-l_1+1}, \quad (\text{B.6})$$

where $\pi_s g = \int g(x) \pi_s(dx)$.

In view of (B.5) and (B.6), we have for $l \in \Lambda_{2m,j}$ and $j \geq 2$,

$$\begin{aligned} \mathbb{E}_\nu \left[\tilde{L}_{j,l}(x) \right] &\leq \left(\prod_{k=2}^j M_{j,l}(k) h^{-l_k} \right) \left(\nu G_{s,\nu} \tilde{I}_{L_{h,x}^{l_1}} \right) \\ &\leq \left(\prod_{k=1}^j M_{j,l}(k) \right) h^{(-\sum_{k=1}^j l_k)+1} = \left(\prod_{k=1}^j M_{j,l}(k) \right) h^{-2m+1}. \end{aligned}$$

For the case of $j = 1$, by (B.6), we have

$$\mathbb{E}_\nu \left[\tilde{L}_{j,l}(x) \right] \leq \nu G_{s,\nu} \tilde{I}_{L_{h,x}^{2m}} = \pi_s L_{h,x}^{2m} \leq M_{1,l}(1) h^{-2m+1}. \quad (\text{B.7})$$

where $M_{1,l}(1)$ is a positive constant independent of x .

Letting

$$M = \sum_{j=1}^{2m} \sum_{l \in \Lambda_{2m,j}} \frac{(2m)!}{l_1! \cdots l_j!} M_{j,l} \quad \text{with} \quad M_{j,l} = \prod_{k=1}^j M_{j,l}(k),$$

by (B.3), (B.7) and (B.7), we have shown that (B.1) holds. \blacksquare

LEMMA B.2. *Let $\{X_t\}$ be a null recurrent Markov process. Then, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ C_1 < \frac{N(n)}{n^\beta L_s(n)} < C_2 \right\} = 1, \quad (\text{B.8})$$

where $C_1 < C_2$ are two positive constants. Furthermore,

$$\mathbb{P} \left\{ n^{\beta-\epsilon} < N(n) < n^{\beta+\epsilon}, \quad i.o. \right\} = 1 \quad (\text{B.9})$$

for any $\epsilon > 0$.

PROOF. We only provide the detailed proof of (B.8) as the proof of (B.9) follows from Lemma 3.4 of Karlsen and Tjøstheim (2001).

By the definition of Mittag–Leffler distribution (see, for example, Lin 1998), there exist two positive constants $0 < C_1 < C_2 < \infty$ such that

$$\mathbb{P} (C_1 < M_\beta(1) \leq C_2) \geq 1 - \frac{\delta}{2} \quad (\text{B.10})$$

for any small $\delta > 0$.

Let $F_n(x) = \mathbb{P} \left\{ \frac{N(n)}{n^\beta L_s(n)} \leq x \right\}$ and $F(x) = \mathbb{P} \{M_\beta(1) \leq x\}$. Then, equation (A.5) implies that for n large enough, we have

$$F_n(C_2) - F(C_2) \geq -\frac{\delta}{4}, \quad (\text{B.11})$$

$$F_n(C_1) - F(C_1) \leq \frac{\delta}{4}. \quad (\text{B.12})$$

Thus, equations (B.10)–(B.12) imply for large enough n

$$\begin{aligned} \mathbb{P} \left(C_1 < \frac{N(n)}{n^\beta L_s(n)} \leq C_2 \right) &= \left(\mathbb{P} \left(\frac{N(n)}{n^\beta L_s(n)} \leq C_2 \right) - \mathbb{P} (M_\beta(1) \leq C_2) \right) \\ &- \left(\mathbb{P} \left(\frac{N(n)}{n^\beta L_s(n)} \leq C_1 \right) - \mathbb{P} (M_\beta(1) \leq C_1) \right) + \mathbb{P} (C_1 < M_\beta(1) \leq C_2) \geq 1 - \delta, \end{aligned}$$

which implies that (B.8) holds. \blacksquare

PROOF OF THEOREM 3.1. Here and in the sequel, let C denote a positive constant, which may change from line to line. Since $\{e_t\}$ is assumed to be i.i.d. and independent of $\{X_t\}$, $\{(X_t, e_t)\}$ is still β -null recurrent by Lemma 3.1 in Karlsen, Myklebust and Tjøstheim (2007).

Let

$$\eta_n = \sqrt{\frac{\log n}{n^\beta L_s(n)h}}, \quad \Gamma_t(x) = \frac{1}{h} L \left(\frac{X_t - x}{h} \right) e_t$$

and $J_n(\beta)$ be the event $\{C_1 n^\beta L_s(n) \leq N(n) \leq C_2 n^\beta L_s(n)\}$, where C_1 and C_2 are defined as in Lemma B.2.

To prove Theorem 3.1, we need only to show

$$\sup_{|x| \leq T_n} \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) = O_P(\eta_n^*), \quad (\text{B.13})$$

where $\eta_n^* = \eta \cdot \eta_n$, in which $\eta > 0$ is a constant and can be sufficiently large.

Observe that

$$\begin{aligned} & \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \eta_n^* \right) \\ = & \left\{ \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \eta_n^* \right) \cap J_n(\beta) \right\} \\ & \cup \left\{ \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \eta_n^* \right) \cap J_n^c(\beta) \right\} \\ \subset & \left\{ \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \eta_n^* \right) \cap J_n(\beta) \right\} \cup J_n^c(\beta). \end{aligned} \quad (\text{B.14})$$

By (B.8) in Lemma B.2, in order to prove (B.13), it suffices to show that as $n \rightarrow \infty$

$$\mathbb{P} \left\{ \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \eta_n^* \right) \cap J_n(\beta) \right\} \rightarrow 0. \quad (\text{B.15})$$

The set $\{x : |x| \leq T_n\}$ can be covered by a finite number of subsets $\{S_i\}$ centered at s_i with radius $(n^{\beta - \frac{\varepsilon_0}{2}} h)^{-(2p_0 - 1)}$, where $p_0 > \frac{1 + \beta}{\varepsilon_0} - 1$. Letting $Q(n)$ be the number of these sets, then $Q(n) = O\left(T_n (n^{\beta - \frac{\varepsilon_0}{2}} h)^{2p_0 - 1}\right)$. Hence, it is easy to check that

$$\begin{aligned} \sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| & \leq \max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(s_j) \right| \\ & + \max_{1 \leq j \leq Q(n)} \sup_{x \in S_j} \left| \frac{1}{N(n)} \sum_{t=0}^n (\Gamma_t(x) - \Gamma_t(s_j)) \right|. \end{aligned} \quad (\text{B.16})$$

Assumption A2(i) implies that there exists a constant $C_l > 0$ such that

$$\left| L\left(\frac{X_t - x}{h}\right) - L\left(\frac{X_t - s_j}{h}\right) \right| \leq C_l \left| \frac{s_j - x}{h} \right| \leq C_l \frac{(n^{\beta - \frac{\varepsilon_0}{2}} h)^{-(2p_0 - 1)}}{h}. \quad (\text{B.17})$$

By (B.17), it is easy to check that in $J_n(\beta) = \{C_1 n^\beta L_s(n) \leq N(n) \leq C_2 n^\beta L_s(n)\}$,

$$\begin{aligned}
& \max_{1 \leq j \leq Q(n)} \sup_{x \in S_j} \left\{ \left| \frac{1}{N(n)} \sum_{t=0}^n [\Gamma_t(x) - \Gamma_t(s_j)] \right| \right\} \\
&= O_P \left(n h^{-2} (n^{\beta - \frac{\varepsilon_0}{2}} h)^{-(2p_0 - 1)} \cdot N^{-1}(n) \right) \\
&= O_P \left(n^{1 - 2p_0 \beta + p_0 \varepsilon_0 - \frac{\varepsilon_0}{2}} h^{-2p_0 - 1} \cdot L_s^{-1}(n) \right) \\
&= O_P \left(n^{1 - 2p_0 \beta + p_0 \varepsilon_0 - \frac{\varepsilon_0}{2}} h^{-2p_0 - 1} \cdot L_s^{-1}(n) \eta_n^{-1} \right) \cdot O(\eta_n^*) \\
&= O_P \left(\frac{1}{(n^{\beta - \varepsilon_0} h)^{2p_0 + \frac{1}{2}} n^{p_0 \varepsilon_0 - 1 - \beta + \varepsilon_0} \sqrt{L_s(n) \log(n)}} \right) \cdot O(\eta_n^*) \\
&= o_P \left(\frac{1}{n^{p_0 \varepsilon_0 - 1 - \beta + \varepsilon_0} \sqrt{L_s(n) \log(n)}} \right) \cdot O(\eta_n^*) = o_P(\eta_n^*), \tag{B.18}
\end{aligned}$$

using $n^{\beta - \varepsilon_0} h \rightarrow \infty$ by Assumption A2(ii) and $p_0 > \frac{1 + \beta}{\varepsilon_0} - 1$.

In view of (B.16) and (B.18), in order to prove (B.15), it suffices to show that as $n \rightarrow \infty$

$$\mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(s_j) \right| > \eta_n^* \right) \cap J_n(\beta) \right\} \rightarrow 0. \tag{B.19}$$

We then apply the same independence decomposition technique as used in (2.5) to show (B.19). Define

$$Z_k(s_j) = \begin{cases} \sum_{t=0}^{\tau_0} \Gamma_t(s_j), & k = 0, \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} \Gamma_t(s_j), & k \geq 1, \\ \sum_{t=\tau_{N(n)}+1}^n \Gamma_t(s_j), & k = (n), \end{cases} \tag{B.20}$$

where $\tau_k, k \geq 0$, are defined as in Section 2. Then

$$\sum_{t=0}^n \Gamma_t(s_j) = Z_0(s_j) + \sum_{k=1}^{N(n)} Z_k(s_j) + Z_{(n)}(s_j). \tag{B.21}$$

From Nummelin (1984)'s result, we know that $\{Z_k(s_j), k \geq 1\}$ is a sequence of i.i.d. random variables for each fixed j . We first show that as $n \rightarrow \infty$

$$\mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} Z_k(s_j) \right| > \eta_n^* \right) \cap J_n(\beta) \right\} \rightarrow 0. \tag{B.22}$$

We prove (B.22) through using the Bernstein inequality and the truncation method. Similarly to the proof of Lemma B.1, we have for $p_0 = \left\lceil \frac{1 + \beta}{\varepsilon_0} \right\rceil$,

$$\max_{1 \leq j \leq Q(n)} \mathbf{E} \left[|Z_k(s_j)|^{4p_0} \right] \leq C h^{-4p_0 + 1} \tag{B.23}$$

where the constant C depends neither on s_j nor on n .

By the definition of $\Gamma_t(x) = \frac{1}{h}L\left(\frac{X_t-x}{h}\right) e_t$ and the mutual independence between $\{e_s : s \geq 1\}$ and $\{X_t : t \geq 1\}$, we have $\mathbf{E}[Z_k(s_j)] = 0$.

Note that $n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}} = \sqrt{n^{\beta-\varepsilon_0}h} n^{\frac{3}{8}\varepsilon_0}h^{-1} \rightarrow \infty$ by Assumption A2(ii) and define

$$\bar{Z}_k(s_j) = Z_k(s_j)I\left(|Z_k(s_j)| < n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}}\right) \quad \text{and} \quad \tilde{Z}_k(s_j) = Z_k(s_j) - \bar{Z}_k(s_j). \quad (\text{B.24})$$

By standard arguments, we have

$$\begin{aligned} & \mathbf{P}\left\{\left(\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=0}^{N(n)} Z_k(s_j) - \nu_*(s_j)\right| > \eta_n^*\right) \cap J_n(\beta)\right\} \\ & \leq \mathbf{P}\left\{\left(\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=0}^{N(n)} (\bar{Z}_k(s_j) - \mathbf{E}[\bar{Z}_k(s_j)])\right| > \eta_n^*/2\right) \cap J_n(\beta)\right\} \\ & + \mathbf{P}\left\{\left(\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=0}^{N(n)} (\tilde{Z}_k(s_j) - \mathbf{E}[\tilde{Z}_k(s_j)])\right| > \eta_n^*/2\right) \cap J_n(\beta)\right\}. \end{aligned} \quad (\text{B.25})$$

By (B.23) and (B.24), we have uniformly for $1 \leq j \leq Q(n)$,

$$\begin{aligned} \mathbf{E}\left|\tilde{Z}_k(s_j)\right| &= \mathbf{E}\left[|Z_k(s_j)|I\left(|Z_k(s_j)| \geq n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}}\right)\right] \\ &\leq \mathbf{E}\left[|Z_k(s_j)|^{4p_0} \left(n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}}\right)^{-4p_0+1}\right] = O\left(h^{-4p_0+1} \left(n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}}\right)^{-4p_0+1}\right) \\ &= O\left(\frac{\sqrt{L_s(n)}}{(n^{\beta-\varepsilon_0}h)^{2p_0-1} n^{\frac{\varepsilon_0(12p_0-7)}{8}}}\right) \cdot O(\eta_n^*) = o(\eta_n^*) \end{aligned} \quad (\text{B.26})$$

using similar arguments to the derivation of (B.18), $p_0 > 1$ and Assumption A2(ii).

Meanwhile, note that

$$\begin{aligned} & \max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{Z}_k(s_j)\right| \\ & \geq \max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=1}^{N(n)} (\tilde{Z}_k(s_j) - \mathbf{E}[\tilde{Z}_k(s_j)])\right| - \max_{1 \leq j \leq Q(n)} \mathbf{E}\left|\tilde{Z}_k(s_j)\right|, \end{aligned}$$

which, together with $\max_{1 \leq j \leq Q(n)} \mathbf{E}\left|\tilde{Z}_k(s_j)\right| \leq \frac{\eta_n^*}{6}$ by (B.26), leads to

$$\begin{aligned} & \left\{\max_{1 \leq j \leq Q(n)} \max_{1 \leq k \leq N(n)} |Z_k(s_j)| < n^{\frac{\beta}{2}-\frac{\varepsilon_0}{8}}h^{-\frac{1}{2}}\right\} \\ & \subset \left\{\tilde{Z}_k(s_j) = 0 \text{ for } 1 \leq j \leq Q(n), 1 \leq k \leq N(n)\right\} \\ & \subset \left\{\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{Z}_k(s_j)\right| = 0\right\} \subset \left\{\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{Z}_k(s_j)\right| \leq \frac{\eta_n^*}{3}\right\} \\ & \subset \left\{\max_{1 \leq j \leq Q(n)} \left|\frac{1}{N(n)} \sum_{k=1}^{N(n)} (\tilde{Z}_k(s_j) - \mathbf{E}[\tilde{Z}_k(s_j)])\right| \leq \frac{\eta_n^*}{2}\right\}. \end{aligned} \quad (\text{B.27})$$

By (B.23), (B.24), (B.27) and the Markov inequality, we have for $p_0 > \frac{2\beta}{\varepsilon_0} - 1$

$$\begin{aligned}
& \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\tilde{Z}_k(s_j) - \mathbb{E} \left[\tilde{Z}_k(s_j) \right] \right) \right| > \eta_n^*/2 \right) \cap J_n(\beta) \right\} \\
&= \mathbb{P}(J_n(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\tilde{Z}_k(s_j) - \mathbb{E} \left[\tilde{Z}_k(s_j) \right] \right) \right| \leq \eta_n^*/2 \right) \cap J_n(\beta) \right\} \\
&\leq \mathbb{P}(J_n(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{Z}_k(s_j) \right| \leq \eta_n^*/3 \right) \cap J_n(\beta) \right\} \\
&\leq \mathbb{P}(J_n(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{Z}_k(s_j) \right| = 0 \right) \cap J_n(\beta) \right\} \\
&\leq \mathbb{P}(J_n(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \max_{1 \leq k \leq N(n)} |Z_k(s_j)| < n^{\frac{\beta}{2} - \frac{\varepsilon_0}{8}} h^{-\frac{1}{2}} \right) \cap J_n(\beta) \right\} \\
&= \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \max_{1 \leq k \leq N(n)} |Z_k(s_j)| \geq n^{\frac{\beta}{2} - \frac{\varepsilon_0}{8}} h^{-\frac{1}{2}} \right) \cap J_n(\beta) \right\} \\
&\leq \sum_{k=1}^{C_2 n^\beta L_s(n)} \sum_{j=1}^{Q(n)} \mathbb{P} \left\{ |Z_k(s_j)| \geq n^{\frac{\beta}{2} - \frac{\varepsilon_0}{8}} h^{-\frac{1}{2}} \right\} \leq CQ(n)n^\beta L_s(n)h^{1-2p_0}n^{-2p_0(\beta - \frac{\varepsilon_0}{4})} \\
&\leq Cn^{2\beta - \varepsilon_0} L_s^2(n)(n^{\beta - \frac{\varepsilon_0}{2}} h)^{2p_0 - 1} h^{1-2p_0} n^{-2p_0(\beta - \frac{\varepsilon_0}{4})} \\
&= O \left(n^{-\left(\frac{1}{2}(p_0+1)\varepsilon_0 - \beta\right)} L_s^2(n) \right) = o(1). \tag{B.28}
\end{aligned}$$

By Lemma B.1, we have

$$\text{Var} \left[\bar{Z}_k(s_j) \right] = \mathbb{E} \left[\bar{Z}_k^2(s_j) \right] - \left\{ \mathbb{E} \left[\bar{Z}_k(s_j) \right] \right\}^2 \leq \mathbb{E} \left[\bar{Z}_k^2(s_j) \right] \leq 2\mathbb{E} \left[Z_k^2(s_j) \right] \leq M_1 h^{-1}, \tag{B.29}$$

where M_1 is independent of k and j .

By (B.29), for any $q \geq 1$, we have uniformly for $1 \leq j \leq Q(n)$

$$\sum_{k=1}^q \text{Var} \left[\bar{Z}_k(s_j) \right] \leq M_1 q h^{-1}. \tag{B.30}$$

Meanwhile, by (B.30) and Bernstein inequality for i.i.d. random variables (see, for example,

van der Vaart and Wellner 1996), we have for some $0 < C_1 < C_2 < \infty$

$$\begin{aligned}
& \mathbb{P} \left\{ \left(\max_{1 \leq j \leq Q(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} (\bar{Z}_k(s_j) - \mathbb{E} [\bar{Z}_k(s_j)]) \right| > \eta_n^*/2 \right) \cap J_n(\beta) \right\} \\
& \leq \sum_{j=1}^{Q(n)} \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \mathbb{P} \left\{ \left| \frac{1}{q} \sum_{k=1}^q (\bar{Z}_k(s_j) - \mathbb{E} [\bar{Z}_k(s_j)]) \right| > \eta_n^*/2 \right\} \\
& \leq \sum_{j=1}^{Q(n)} \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \exp \left\{ - \frac{q^2 (\eta_n^*/2)^2}{2 \sum_{k=1}^q \text{Var} [\bar{Z}_k(s_j)] + (2/3) n^{\frac{\beta}{2} - \frac{\varepsilon_0}{8}} h^{-\frac{1}{2}} q (\eta_n^*/2)} \right\} \\
& \leq \sum_{j=1}^{Q(n)} \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \exp \left\{ - \frac{q^2 (\eta_n^*/2)^2}{C q h^{-1} \left(1 + \eta \sqrt{\frac{\log n}{L_s(n)}} n^{-\frac{\varepsilon_0}{8}} \right)} \right\} \\
& \leq \sum_{j=1}^{Q(n)} \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \exp \left\{ - \frac{c_0 q (\eta_n^*)^2 h}{1 + o(1)} \right\} \leq Q(n) \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \exp \{ -c_0 q \eta_n^{*2} h \} \\
& = Q(n) \sum_{q=C_1 n^\beta L_s(n)}^{C_2 n^\beta L_s(n)} \exp \left\{ - \frac{c_0 \eta^2 q \log(n)}{n^\beta L_s(n)} \right\} \\
& \leq C_2 Q(n) n^\beta L_s(n) \exp \{ -c_0 C_1 \eta^2 \log n \} = o(1)
\end{aligned} \tag{B.31}$$

for some suitably chosen $\eta > 0$ such that $\eta > \sqrt{\frac{2(2p_0+1)\beta - (2p_0-1)\varepsilon_0}{2C_1 c_0}}$, where c_0 is a positive constant.

Then, by (B.25), (B.28) and (B.31), equation (B.22) is proved.

To consider the edge terms $Z_0(s_j)$ and $Z_{(n)}(s_j)$, we first prove

$$\max_{1 \leq j \leq Q(n)} \mathbb{E} \left[|Z_0(s_j)|^{4p_0} \right] \leq M_2 h^{-4p_0+1} \tag{B.32}$$

and

$$\max_{1 \leq j \leq Q(n)} \mathbb{E} \left[|Z_{(n)}(s_j)|^{4p_0} \right] \leq M_3 h^{-4p_0+1} \tag{B.33}$$

where M_2 and M_3 are both positive constants independent of s_j and n .

If $T_0 = 1$ ($\{T_t\}$ is defined in Appendix A), $\tau_0 = \tau = 0$, which implies that

$$Z_0(x) = \frac{1}{h} L \left(\frac{X_0 - x}{h} \right) e_0.$$

Then, by some standard calculation, we have

$$\begin{aligned}
\mathbb{E} \left[|Z_0(s_j)|^{4p_0} \right] &= h^{-4p_0} \mathbb{E}[|e_0|^{4p_0}] \mathbb{E}_\nu \left[L^{4p_0} \left(\frac{X_0 - s_j}{h} \right) \right] \\
&= h^{-4p_0} \mathbb{E}[|e_0|^{4p_0}] \int_{\mathbb{E}} L^{4p_0} \left(\frac{z - s_j}{h} \right) \nu(dz) \\
&= h^{-4p_0+1} \mathbb{E}[|e_0|^{4p_0}] \int_{C(L)} L^{4p_0}(y) \nu(dy),
\end{aligned}$$

which implies that (B.32) holds. If $T_0 = 0$, without loss of generality, let $\tau_{-1} = -1$. Then, following the proof of Lemma B.1, we can also show that (B.32) holds.

On the other hand, noting that $n \leq \tau_{N(n)+1}$, it follows that

$$\sum_{t=\tau_{N(n)+1}}^n |\Gamma_t(s_j)| \leq \sum_{t=\tau_{N(n)+1}}^{\tau_{N(n)+1}} |\Gamma_t(s_j)|. \quad (\text{B.34})$$

By (B.34) and following the proof of Lemma B.1, we have

$$\max_{1 \leq j \leq Q(n)} \mathbb{E} \left[|Z_{(n)}(s_j)|^{4p_0} \right] \leq \max_{1 \leq j \leq Q(n)} \mathbb{E} \left(\sum_{t=\tau_{N(n)+1}}^{\tau_{N(n)+1}} |\Gamma_t(s_j)| \right)^{4p_0} \leq M_2 h^{-4p_0+1},$$

which implies that (B.33) holds.

Then, by (B.32), Lemma B.2 and the Markov inequality, we have

$$\begin{aligned} & \mathbb{P} \left\{ \left(\frac{1}{N(n)} \max_{1 \leq j \leq Q(n)} |Z_0(s_j)| > \eta_n^* \right) \cap J_n(\beta) \right\} \\ & \leq \sum_{j=1}^{Q(n)} \mathbb{P} \left\{ \left(\frac{1}{N(n)} |Z_0(s_j)| > \eta_n^* \right) \cap J_n(\beta) \right\} \\ & \leq C \sum_{j=1}^{Q(n)} \frac{\mathbb{E} \left[|Z_0(s_j)|^{4p_0} \right]}{(\eta_n^* n^\beta L_s(n))^{4p_0}} = O \left(\frac{Q(n) h^{-4p_0+1}}{(\eta_n^* n^\beta L_s(n))^{4p_0}} \right) \\ & = O \left(\frac{1}{n^{(p_0-\frac{1}{2})\varepsilon_0} L_s^{2p_0-1}(n) \log^{2p_0}(n)} \right) = o(1) \end{aligned} \quad (\text{B.35})$$

as $\varepsilon_0 > 0$ and $p_0 > \frac{1}{2}$.

Hence, we have

$$\frac{1}{N(n)} \max_{1 \leq j \leq Q(n)} |Z_0(s_j)| = O_P(\eta_n^*). \quad (\text{B.36})$$

By (B.33), similarly to the proof of (B.36), we have

$$\frac{1}{N(n)} \max_{1 \leq j \leq Q(n)} |Z_{(n)}(s_j)| = O_P(\eta_n^*). \quad (\text{B.37})$$

In view of (B.21), (B.22), (B.36) and (B.37), equation (B.19) holds. Hence, the proof of Theorem 3.1 is completed. \blacksquare

PROOF OF THEOREM 3.2. Let $\Gamma_t(x)$ be defined as in the proof of Theorem 3.1 and $J_n^*(\beta) = \{n^{\beta-\xi_1\varepsilon_0} \ll N(n) \ll n^{\beta+\xi_1\varepsilon_0}\}$, where ξ_1 will be chosen later and the symbol “ $a_n \ll b_n$ ” means that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$.

By (B.9), in order to prove (3.5), it suffices to show that for any $\epsilon > 0$,

$$\mathbb{P} \left\{ \left(\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| > \frac{\epsilon}{\sqrt{n^{\beta-\varepsilon_0} h}} \right) \cap J_n^*(\beta), \text{ i.o.} \right\} = 0. \quad (\text{B.38})$$

As in the proof of Theorem 3.1, the set $\{x : |x| \leq T_n\}$ can be covered by a finite number of subsets $\{S_i^*\}$ centered at s_i^* with radius $r_n = (n^{(\beta+\varepsilon_0-2\xi_1\varepsilon_0-2)/2}h^{3/2})$.

Letting $U(n)$ be the number of these sets, we then have $U(n) = O(T_n r_n^{-1})$. Similarly to the derivation in (B.16), we have

$$\begin{aligned} & \sup_{|x| \leq T_n} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x) \right| \leq \max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{t=0}^n \Gamma_t(s_j^*) \right| \\ & + \max_{1 \leq j \leq U(n)} \sup_{x \in S_j^*} \frac{1}{N(n)} \sum_{t=0}^n |\Gamma_t(x) - \Gamma_t(s_j^*)| =: \Pi_{n,1} + \Pi_{n,2}. \end{aligned} \quad (\text{B.39})$$

In a derivation similar to (B.18), Assumptions A1(ii) and A2(i) imply that

$$\Pi_{n,2} = O\left(\frac{n r_n}{N(n)h^2}\right) = o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\right). \quad (\text{B.40})$$

in $J_n^*(\beta)$.

In view of (B.39) and (B.40), in order to prove (B.38), we need only to consider $\Pi_{n,1}$. We will still apply the independence decomposition technique and truncation method as in the proof of Theorem 3.1. Letting $Z_k(s_j^*)$ be defined as $Z_k(s_j)$ in (B.20),

$$\Pi_{n,1} = \max_{1 \leq j \leq U(n)} \frac{1}{N(n)} \left| Z_0(s_j^*) + \sum_{k=1}^{N(n)} Z_k(s_j^*) + Z_{(n)}(s_j^*) \right|. \quad (\text{B.41})$$

We first show that

$$\mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} Z_k(s_j^*) \right| > \epsilon_n \right) \cap J_n^*(\beta), \text{ i.o.} \right\} = 0, \quad (\text{B.42})$$

where $\epsilon_n = \frac{\epsilon}{\sqrt{n^{\beta-\varepsilon_0}h}}$ for some $\epsilon > 0$.

Similarly to the proof of Lemma B.1, we have

$$\max_{1 \leq j \leq U(n)} \mathbb{E} \left[|Z_k(s_j^*)|^{2m_0} \right] \leq M_3 h^{-2m_0+1}, \quad (\text{B.43})$$

where the constant M_3 depends neither on s_j^* nor on n . Define

$$\widehat{Z}_k(s_j^*) = Z_k(s_j^*) I\left(|Z_k(s_j^*)| < n^{(\beta-\xi_2\varepsilon_0)/2} h^{-1/2}\right) \quad \text{and} \quad \underline{Z}_k(s_j^*) = Z_k(s_j^*) - \widehat{Z}_k(s_j^*), \quad (\text{B.44})$$

where ξ_2 is chosen such that $0 < \xi_1 < \xi_2 < \frac{(2-\xi_1)\varepsilon_0}{\beta+1+\varepsilon_0} < 1$.

As in (B.25), we have

$$\begin{aligned} & \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} Z_k(s_j^*) \right| > \epsilon_n \right) \cap J_n^*(\beta) \right\} \\ & \leq \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\widehat{Z}_k(s_j^*) - \mathbb{E} \left[\widehat{Z}_k(s_j^*) \right] \right) \right| > \epsilon_n/2 \right) \cap J_n^*(\beta) \right\} \\ & + \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\underline{Z}_k(s_j^*) - \mathbb{E} \left[\underline{Z}_k(s_j^*) \right] \right) \right| > \epsilon_n/2 \right) \cap J_n^*(\beta) \right\}. \end{aligned} \quad (\text{B.45})$$

By (B.43) and similarly to the proof of (B.28), we have

$$\begin{aligned}
& \sum_{n=1}^{\infty} \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\underline{Z}_k(s_j^*) - \mathbb{E} \left[\underline{Z}_k(s_j^*) \right] \right) \right| > \epsilon_n/2 \right) \cap J_n^*(\beta) \right\} \\
&= \sum_{n=1}^{\infty} \left(\mathbb{P}(J_n^*(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\underline{Z}_k(s_j^*) - \mathbb{E} \left[\underline{Z}_k(s_j^*) \right] \right) \right| \leq \epsilon_n/2 \right) \cap J_n^*(\beta) \right\} \right) \\
&\leq \sum_{n=1}^{\infty} \left(\mathbb{P}(J_n^*(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \underline{Z}_k(s_j^*) \right| = 0 \right) \cap J_n^*(\beta) \right\} \right) \\
&\leq \sum_{n=1}^{\infty} \left(\mathbb{P}(J_n^*(\beta)) - \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \max_{1 \leq k \leq N(n)} |\underline{Z}_k(s_j^*)| < n^{(\beta-\xi_2\epsilon_0)/2} h^{-1/2} \right) \cap J_n^*(\beta) \right\} \right) \\
&= \sum_{n=1}^{\infty} \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \max_{1 \leq k \leq N(n)} |\underline{Z}_k(s_j^*)| \geq n^{(\beta-\xi_2\epsilon_0)/2} h^{-1/2} \right) \cap J_n^*(\beta) \right\} \\
&\leq \sum_{n=1}^{\infty} \sum_{j=1}^{U(n)} \sum_{k=1}^{n^{\beta+\xi_1\epsilon_0}} \mathbb{P} \left\{ |\underline{Z}_k(s_j^*)| \geq n^{(\beta-\xi_2\epsilon_0)/2} h^{-1/2} \right\} \\
&\leq C \sum_{n=1}^{\infty} U(n) n^{\beta+\xi_1\epsilon_0} h^{1-2m_0} n^{-m_0(\beta-\xi_2\epsilon_0)} h^{m_0} \\
&\leq C \sum_{n=1}^{\infty} T(n) r_n^{-1} n^{\beta+\xi_1\epsilon_0} h^{1-2m_0} n^{-m_0(\beta-\xi_2\epsilon_0)} h^{m_0} \\
&= C \sum_{n=1}^{\infty} \frac{L_s(n)}{n^{m_0\epsilon_0(1-\xi_2)-1-2\xi_1\epsilon_0-2(\epsilon_0-\beta)} (n^{\beta-\epsilon_0} h)^{m_0+\frac{1}{2}}} < \infty,
\end{aligned}$$

since $m_0 > \frac{2(1+\xi_1\epsilon_0)}{\epsilon_0(1-\xi_2)} + \frac{2(\beta-\epsilon_0)}{\epsilon_0(1-\xi_2)}$ by the choice of ξ_2 and $m_0 = 2 \left(1 + \left\lceil \frac{\beta+1}{\epsilon_0} \right\rceil \right)$.

Meanwhile, as in the proof of (B.30), we have

$$\text{Var} \left[\widehat{Z}_k(s_j) \right] \leq M_4 q h^{-1}, \quad (\text{B.46})$$

where M_4 is a positive constant independent of j and n .

Then, by (B.46) and the Bernstein inequality, we have

$$\begin{aligned}
& \sum_{n=1}^{\infty} \mathbb{P} \left\{ \left(\max_{1 \leq j \leq U(n)} \left| \frac{1}{N(n)} \sum_{k=1}^{N(n)} \left(\widehat{Z}_k(s_j^*) - \mathbb{E} \left[\widehat{Z}_k(s_j^*) \right] \right) \right| > \epsilon_n/2 \right) \cap J_n^*(\beta) \right\} \\
&\leq \sum_{n=1}^{\infty} \sum_{j=1}^{U(n)} \sum_{q=c_1 n^{\beta-\xi_1\epsilon_0}}^{c_2 n^{\beta+\xi_1\epsilon_0}} \mathbb{P} \left\{ \left| \frac{1}{q} \sum_{k=1}^q \left(\widehat{Z}_k(s_j^*) - \mathbb{E} \left[\widehat{Z}_k(s_j^*) \right] \right) \right| > \epsilon_n/2 \right\} \\
&\leq \sum_{n=1}^{\infty} U(n) \sum_{q=c_1 n^{\beta-\xi_1\epsilon_0}}^{c_2 n^{\beta+\xi_1\epsilon_0}} \exp \left\{ - \frac{q^2 (\epsilon_n/2)^2}{2 \sum_{k=1}^q \text{Var} \left[\widehat{Z}_k(s_j) \right] + (2/3) n^{(\beta-\xi_2\epsilon_0)/2} h^{-\frac{1}{2}} q (\epsilon_n/2)} \right\} \\
&\leq \sum_{n=1}^{\infty} U(n) \sum_{q=c_1 n^{\beta-\xi_1\epsilon_0}}^{c_2 n^{\beta+\xi_1\epsilon_0}} \exp \left\{ - \frac{q^2 (\epsilon_n/2)^2}{C q h^{-1} n^{\epsilon_0(1-\xi_2)/2} (1+o(1))} \right\} \\
&= \sum_{n=1}^{\infty} U(n) \sum_{q=c_1 n^{\beta-\xi_1\epsilon_0}}^{c_2 n^{\beta+\xi_1\epsilon_0}} \exp \left\{ -C_1 \frac{q}{n^{\beta-\epsilon_0} n^{\epsilon_0(1-\xi_2)/2} (1+o(1))} \right\} \\
&\leq c_2 \sum_{n=1}^{\infty} U(n) n^{\beta+\xi_1\epsilon_0} \exp \left\{ -C_2 n^{(1-2\xi_1+\xi_2)\frac{\epsilon_0}{2}} \right\} < \infty, \quad (\text{B.47})
\end{aligned}$$

since $0 < \xi_1 < \xi_2 < 1$ leads to $1 - 2\xi_1 + \xi_2 > 0$, where c_1, c_2, C_1 and C_2 are some positive constants. By (B.45), (B.46), (B.47) and the Borel–Cantelli lemma, equation (B.42) is proved.

Furthermore, by (B.32) with $4p_0$ replaced by $2m_0$, and similarly to the proof of (B.36), we have

$$\begin{aligned}
& \sum_{n=1}^{\infty} \mathbf{P} \left\{ \left(\frac{1}{N(n)} \max_{1 \leq j \leq U(n)} |Z_0(s_j^*)| > \epsilon_n \right) \cap J_n^*(\beta) \right\} \\
& \leq \sum_{n=1}^{\infty} \sum_{j=1}^{U(n)} \mathbf{P} \left\{ \left(\frac{1}{N(n)} |Z_0(s_j^*)| > \epsilon_n \right) \cap J_n^*(\beta) \right\} \\
& \leq C \sum_{n=1}^{\infty} \sum_{j=1}^{U(n)} \frac{\mathbf{E} \left[|Z_0(s_j^*)|^{2m_0} \right]}{(\epsilon_n n^{\beta - \xi_1 \epsilon_0})^{2m_0}} \leq C \sum_{n=1}^{\infty} \frac{U(n) h^{-2m_0+1}}{(\epsilon_n n^{\beta - \xi_1 \epsilon_0})^{2m_0}} \\
& = C \sum_{n=1}^{\infty} \frac{1}{n^{m_0(\beta + \epsilon_0(1 - 2\xi_1)) - 1 - \beta + \epsilon_0(1 - \xi_1)} \sqrt{n^{\beta - \epsilon_0} h}} < \infty,
\end{aligned}$$

by the definition of m_0 .

Then, we have

$$\frac{1}{N(n)} \max_{1 \leq j \leq U(n)} |Z_0(s_j^*)| = o \left(\frac{1}{\sqrt{n^{\beta - \epsilon_0} h}} \right) \quad a.s. \quad (\text{B.48})$$

Analogously, by (B.33), we have

$$\frac{1}{N(n)} \max_{1 \leq j \leq U(n)} |Z_{(n)}(s_j^*)| = o \left(\frac{1}{\sqrt{n^{\beta - \epsilon_0} h}} \right) \quad a.s. \quad (\text{B.49})$$

Then, by (B.42), (B.48) and (B.49), we have

$$\Pi_{n,1} = o \left(\frac{1}{\sqrt{n^{\beta - \epsilon_0} h}} \right) \quad a.s. \quad (\text{B.50})$$

In view of (B.39), (B.40) and (B.50), equation (B.38) is proved. The proof of Theorem 3.2 is completed. \blacksquare

PROOF OF THEOREM 4.1. Similarly to the decomposition of (B.21), we have

$$\widehat{p}_n(x) = \frac{1}{N(n)} \sum_{k=1}^{N(n)} V_k(x) + \frac{1}{N(n)} V_0(x) + \frac{1}{N(n)} V_{(n)}(x), \quad (\text{B.51})$$

where $V_k(x) = \frac{1}{h} \sum_{t=\tau_{k-1}+1}^{\tau_k} K \left(\frac{X_t - x}{h} \right)$. Note that $\{V_k(x)\}$ is a sequence of i.i.d. random functions of x .

We then have

$$\widehat{p}_n(x) - p_s(x) = \frac{1}{N(n)} V_0(x) + \frac{1}{N(n)} V_{(n)}(x) + \frac{1}{N(n)} \sum_{k=1}^{N(n)} (V_k(x) - \mathbf{E}[V_1(x)]) + \mathbf{E}[V_1(x)] - p_s(x). \quad (\text{B.52})$$

By Lemma B.1 for $V_k(x)$, and following the same argument as in the proof of Theorem 3.1 with some modification, we can prove

$$\sup_{|x| \leq T_n} \frac{1}{N(n)} \left| \sum_{k=1}^{N(n)} (V_k(x) - \mathbb{E}[V_1(x)]) \right| = O_P \left(\sqrt{\frac{\log n}{n^\beta L_s(n) h}} \right). \quad (\text{B.53})$$

Similarly, following the same argument as in the proof of Theorem 3.2 with some modification, we can also prove

$$\sup_{|x| \leq T_n} \frac{1}{N(n)} \left| \sum_{k=1}^{N(n)} (V_k(x) - \mathbb{E}[V_1(x)]) \right| = o \left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}} \right) \text{ a.s.} \quad (\text{B.54})$$

Meanwhile, by (A.6) and (A.7) in Appendix A, we have

$$\begin{aligned} \mathbb{E}[V_k(x)] &= \mathbb{E} \left[\frac{1}{h} \sum_{t=\tau_{k-1}+1}^{\tau_k} K \left(\frac{X_t - x}{h} \right) \right] = \mathbb{E}_\nu \left[\frac{1}{h} \sum_{t=0}^{\tau} K \left(\frac{X_t - x}{h} \right) \right] \\ &= \int \frac{1}{h} K \left(\frac{u - x}{h} \right) \nu G_{s,\nu}(du) = \int \frac{1}{h} K \left(\frac{u - x}{h} \right) \pi_s(du) \\ &= \int K(u) p_s(x + hu) du = p_s(x) + p_s''(x) \left(\int u^2 K(u) du \right) h^2 + o(h^2), \end{aligned} \quad (\text{B.55})$$

where $G_{s,\nu}$ and π_s are defined as in Appendix A.

Equation (B.55) then implies

$$\sup_{x \in \mathcal{R}} |\mathbb{E}[V_k(x)] - p_s(x)| \leq \sup_{x \in \mathcal{R}} |p_s''(x)| \left(\int u^2 K(u) du \right) h^2 + o(h^2). \quad (\text{B.56})$$

In a similar way to the proofs of (B.36) and (B.37), we have

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} V_0(x) \right| + \sup_{|x| \leq T_n} \left| \frac{1}{N(n)} V_{(n)}(x) \right| = O_P \left(\sqrt{\frac{\log n}{n^\beta L_s(n) h}} \right).$$

Analogously to the proofs of (B.48) and (B.49), we have

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)} V_0(x) \right| + \sup_{|x| \leq T_n} \left| \frac{1}{N(n)} V_{(n)}(x) \right| = o \left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}} \right) \text{ a.s.}$$

The above two results and (B.52)–(B.56) imply that both (4.2) and (4.3) hold. \blacksquare

PROOF OF THEOREM 4.2. We only prove (4.12) with the help of Theorem 3.2. The proof of (4.11) is similar by using Theorem 3.1.

By the definition of $\widehat{m}_n(x)$, we have

$$\widehat{m}_n(x) = \sum_{t=0}^n w_{n,t}(x) e_t + \sum_{t=0}^n w_{n,t}(x) m(X_t).$$

By using Theorem 3.2, we then have as $n \rightarrow \infty$

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) e_t \right| = o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\right) \quad a.s. \quad (\text{B.57})$$

Meanwhile, by (4.3) in Theorem 4.1 we have

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) - p_s(x) \right| = O(h^2) + o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\right) \quad a.s. \quad (\text{B.58})$$

In view of $\delta_n = \inf_{|x| \leq T_n} p_s(x)$, by (B.57) and (B.58) we then have

$$\begin{aligned} & \sup_{|x| \leq T_n} \left(\left| \frac{1}{\frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right)} \right| \cdot \left| \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) e_t \right| \right) \\ & \leq C \delta_n^{-1} \sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) e_t \right| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0}h}}\right) \quad a.s., \end{aligned} \quad (\text{B.59})$$

which implies

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n w_{n,t}(x) e_t \right| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0}h}}\right) \quad a.s. \quad (\text{B.60})$$

In the meantime, standard arguments imply

$$\begin{aligned} \sum_{t=0}^n w_{n,t}(x) m(X_t) - m(x) &= \frac{\frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) m(X_t)}{\hat{p}_n(x)} - \frac{m(x) \hat{p}_n(x)}{\hat{p}_n(x)} \\ &= \frac{\frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) (m(X_t) - m(x))}{\hat{p}_n(x)} = \frac{\frac{m'(x)h}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t - x}{h}\right) \left(\frac{X_t - x}{h}\right)}{\hat{p}_n(x)} \\ &+ \frac{\frac{h^2}{N(n)h} \sum_{t=0}^n m''(x + \vartheta_t(X_t - x)) K\left(\frac{X_t - x}{h}\right) \left(\frac{X_t - x}{h}\right)^2}{2\hat{p}_n(x)} \\ &=: \Xi_{n,1}(x) + \Xi_{n,2}(x), \end{aligned}$$

where $0 \leq \vartheta_t \leq 1$.

Similarly to the proof of Theorem 4.1 above, in view of $\int uK(u)du = 0$, the conditions of Theorem 4.2 imply

$$\sup_{|x| \leq T_n} \Xi_{n,1}(x) = o(\delta_{1n}^* h) \quad a.s. \quad (\text{B.61})$$

Analogously, we have

$$\sup_{|x| \leq T_n} \Xi_{n,2}(x) = O(\delta_{2n}^* h^2) \quad a.s. \quad (\text{B.62})$$

Therefore, equations (B.60)–(B.62) show that equation (4.12) in Theorem 4.2 holds. \blacksquare

PROOF OF THEOREM 4.3. We only prove (4.15) with the help of Theorem 3.2 as the proof of (4.14) is similar. By the definition of $\tilde{m}_n(x)$, we have

$$\tilde{m}_n(x) = \sum_{t=0}^n \tilde{w}_{n,t}(x) e_t + \sum_{t=0}^n \tilde{w}_{n,t}(x) m(X_t).$$

Note that $\tilde{w}_{n,t}(x) = \frac{\tilde{K}_{x,h}(X_t)}{\sum_{s=0}^n \tilde{K}_{x,h}(X_s)}$ with $\tilde{K}_{x,h}(X_t) = \frac{1}{h} \tilde{K}_n\left(\frac{X_t-x}{h}\right)$, in which

$$\tilde{K}_n\left(\frac{X_t-x}{h}\right) = K\left(\frac{X_t-x}{h}\right) \left[S_{n,2}(x) - \left(\frac{X_t-x}{h}\right) S_{n,1}(x) \right] \quad (\text{B.63})$$

with $S_{n,j}(x) = \frac{1}{N(n)h} \sum_{t=0}^n K\left(\frac{X_t-x}{h}\right) \left(\frac{X_t-x}{h}\right)^j$ for $j = 0, 1, 2$.

Using the same arguments as in the proof of Theorem 4.1, we have as $n \rightarrow \infty$

$$\sup_{|x| \leq T_n} |S_{n,j}(x) - p_s(x) \mu_j| = O(h^2) + o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s., \quad (\text{B.64})$$

where $\mu_j = \int_{-\infty}^{\infty} x^j K(x) dx$ for $j = 0, 1, 2$.

Similarly to the proof of equation (B.55) in the proof of Theorem 4.2, we have as $n \rightarrow \infty$

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n K_j(x, X_t) e_t \right| = o\left(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s., \quad (\text{B.65})$$

where $K_j(x, X_t) = K\left(\frac{X_t-x}{h}\right) \left(\frac{X_t-x}{h}\right)^j$ for $j = 0$ and 1 .

Noting that $\delta_n = \inf_{|x| \leq T_n} p_s(x)$, equations (B.63)–(B.65) imply

$$\begin{aligned} & \sup_{|x| \leq T_n} \left| \frac{[\sum_{t=0}^n K\left(\frac{X_t-x}{h}\right) e_t] S_{n,2}(x)}{N(n)h \left(S_{n,2}(x) S_{n,0}(x) - S_{n,1}^2(x) \right)} \right| \\ & \leq C \delta_n^{-1} \sup_{|x| \leq T_n} \left| \frac{\sum_{t=0}^n K\left(\frac{X_t-x}{h}\right) e_t}{N(n)h} \right| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s. \end{aligned} \quad (\text{B.66})$$

and

$$\begin{aligned} & \sup_{|x| \leq T_n} \left| \frac{[\sum_{t=0}^n \left(\frac{X_t-x}{h}\right) K\left(\frac{X_t-x}{h}\right) e_t] S_{n,1}(x)}{N(n)h \left(S_{n,2}(x) S_{n,0}(x) - S_{n,1}^2(x) \right)} \right| \\ & \leq C \delta_n^{-1} \sup_{|x| \leq T_n} \left| \frac{\sum_{t=0}^n \left(\frac{X_t-x}{h}\right) K\left(\frac{X_t-x}{h}\right) e_t}{N(n)h} \right| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s.. \end{aligned} \quad (\text{B.67})$$

By (B.66), (B.67) as well as the definition of $\tilde{w}_{n,t}(x)$, we have

$$\sup_{|x| \leq T_n} \left| \frac{1}{N(n)h} \sum_{t=0}^n \tilde{w}_{n,t}(x) e_t \right| = o\left(\frac{1}{\delta_n \sqrt{n^{\beta-\varepsilon_0} h}}\right) \quad a.s. \quad (\text{B.68})$$

In the meantime, observe that

$$\sum_{t=0}^n \tilde{w}_{n,t}(x) m(X_t) - m(x) = \frac{\frac{1}{N(n)} \sum_{t=0}^n \tilde{K}_{x,h}(X_t) (m(X_t) - m(x))}{\tilde{p}_n(x)},$$

where $\tilde{p}_n(x) = \frac{1}{N(n)} \sum_{t=0}^n \tilde{K}_{x,h}(X_t)$, and

$$\begin{aligned} & \frac{1}{N(n)} \sum_{t=0}^n \tilde{K}_n \left(\frac{X_t - x}{h} \right) (m(X_t) - m(x)) \\ &= \frac{m'(x)}{N(n)} \sum_{t=0}^n (X_t - x) \tilde{K}_n \left(\frac{X_t - x}{h} \right) + \frac{1}{2N(n)} \sum_{t=0}^n m''(x + \vartheta'_t(X_t - x)) \\ & \times (X_t - x)^2 \tilde{K}_n \left(\frac{X_t - x}{h} \right) = \frac{m''(x)}{2N(n)} \sum_{t=0}^n (X_t - x)^2 \tilde{K}_n \left(\frac{X_t - x}{h} \right) (1 + o(1)) \text{ a.s.}, \end{aligned}$$

where we have used the fact that $\sum_{t=0}^n (X_t - x) \tilde{K}_n \left(\frac{X_t - x}{h} \right) = 0$ and that $m''(\cdot)$ is continuous, and $0 \leq \vartheta'_t \leq 1$ for $t = 0, \dots, n$.

Finally, using the proof of (B.61), we have

$$\sum_{t=0}^n \tilde{w}_{n,t}(x) m(X_t) - m(x) = O(\delta_{2n}^* h^2) \text{ a.s.} \quad (\text{B.69})$$

By equations (B.68) and (B.69), the proof of (4.15) is therefore completed. ■