AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series

Jiti Gao, Shin Kanaya, Degui Li and Dag Tjøstheim

CREATES Research Paper 2013-29

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series[1]

Jiti Gao[2], Shin Kanaya[3], Degui Li[4] and Dag Tjøstheim[5]

Abstract

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

JEL subject classifications: C13, C14, C22.

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

[1]The authors would like to thank the Editor, Professor Peter C. B. Phillips, and referees for their constructive and valuable comments, which have greatly improved the original version of this paper. Two Australian Research Council Discovery Grants (DP0558602 and DP0879088) and an Australian Research Council Discovery Early Career Researcher Award (DE120101130) provided financial support for this work. Kanaya gratefully acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78).

[2]Jiti Gao (corresponding author) is from Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Email address: jiti.gao@monash.edu.

[3]Adress: Department of Economics and Business, Aarhus University and CREATES, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark. Email address: skanaya@creates.au.dk.

[4]Address: Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK. Email Address: degui.li@york.ac.uk.

[5]Address: Department of Mathematics, University of Bergen, Johannes Bruns gate 12, N-5008 Bergen, Norway. Email Address: Dag.Tjostheim@math.uib.no.

1 Introduction

In this paper, we consider kernel-based nonparametric density and regression estimators, and present their uniform consistency results. The results provide theoretical justification for the use of the estimators, and are also useful for deriving asymptotic results in various estimation and testing problems. Previous studies focused mainly on the cases in which observed time series satisfy some stationarity/ergodicity conditions, as found in Liero (1989), Roussas (1990), Andrews (1995), Liebscher (1996), Masry (1996), Bosq (1998), Fan and Yao (2003), Ould-Saïd and Cai (2005) and others. Most of these studies considered uniform convergence over fixed compact sets. Recently, Hansen (2008) made significant progress towards establishing sharp uniform convergence with sharp rates over unbounded sets for a general class of nonparametric functionals when the time series are stationary and α-mixing. Kristensen (2009) extended Hansen's result to a heterogeneous dependent case with an α-mixing condition. By contrast, little work has been done 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. They developed a local-time-based approach and established asymptotic theory. Simultaneously but independently, Karlsen and Tjøstheim (1998, 2001) considered nonparametric kernel estimation in the nonstationary case in which time series regressors are nonstationary null-recurrent Markov chains. These authors established 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), Schienle (2008, 2011), Cai, Li and Park (2009), Wang and Phillips (2009a, 2009b), Chen, Li and Zhang (2010), Chen, Gao and Li (2012) and the references therein. In the field of model specification testing, Gao *et al.* (2009a, 2009b) presented asymptotically consistent tests in both autoregression and co-integration cases. We also note that supplement materials to Gao *et al.* (2009a, 2009b) provide brief discussions on uniform weak consistency of a nonparametric kernel density estimator for the case where the time series follows a random walk process.

This paper is the first to systematically study strong and weak uniform consistency results for a class of nonparametric kernel density and regression estimators when the time series data involved are nonstationary null-recurrent Markov chains. Our weak uniform consistency result indicates a sharp rate of convergence of the order $O_P\big(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\big)$ when the regressors are β-null recurrent Markov processes (L_s is a slowly varying function defined in Section 2). This sharp rate of convergence is in contrast to $O_P\big(\sqrt{\frac{\log n}{nh}}\big)$, which is a well-known uniform convergence rate for nonparametric kernel estimators in the stationary time series case. Our strong uniform consistency result indicates a rate of convergence of the order $o\big(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\big)$ for some small $0 < \varepsilon_0 < \beta$, where ε_0 depends on the existence of moments of the process. This strong rate shall be close to the sharp weak convergence rate when ε_0 is close to zero. The uniform consistency results established in this paper not only strengthen existing pointwise consistency results given in Karlsen and Tjøstheim (2001), but also extend some corresponding results in Hansen (2008) for the stationary time series case. In a recent paper by Schienle (2011), under the null recurrent setting, the uniform consistency for the nonparametric kernel estimator was established over a compact set, which can be seen as a special case of our results.

The rest of the paper is organized as follows. Some basic definitions and results for Markov chains are introduced in Section 2. The main convergence results are presented in Section 3. Applications of the main results to the density, Nadaraya-Watson, and local-linear kernel estimators are given in Section 4. The conclusions are provided in Section 5. Some basic results on Markov theory are summarized in Appendix A. All 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 its state space $(\mathbb{E}, \mathcal{E})$, transition probability P $(\mathbb{E} \times \mathcal{E} \to [0,1])$, where $\mathcal{E}$ is the sigma algebra on $\mathbb{E}$, ϕ is a measure on $(\mathbb{E}, \mathcal{E})$, and $\mathsf{P}\,(x, \mathbb{A})$ stands for the probability that the chain falls into a set $\mathbb{A}$ in the next period when the current state is x. The ϕ-irreducibility means that there exists a nontrivial measure ϕ on $(\mathbb{E}, \mathcal{E})$ such that each ϕ-positive set $\mathbb{A}$ is attainable from any point x in $\mathbb{E}$ with positive probability, that is,

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

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, we let $\mathbb{E} \subset \mathbb{R}$. Denote by $\mathcal{E}^+$ the class of nonnegative measurable functions with ϕ-positive support. 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

$$\mathsf{P}^m \geq b\eta \otimes \lambda, \tag{2.2}$$

where $\eta \otimes \lambda\,(x, \mathbb{A}) = \eta\,(x)\,\lambda\,(\mathbb{A})$ for any $x \in \mathbb{E}$, $\mathbb{A} \in \mathcal{E}$. If λ satisfies this 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 an indicator function $1_{\mathbb{B}}$ is a small function. By (2.2), Theorem 2.1, and Proposition 2.6 in Nummelin (1984), we know that for a ϕ-irreducible Markov chain, there exists a minorization inequality, i.e., there are a small function s, a probability measure ν and an integer $m_0 \geq 1$ such that $\mathsf{P}^{m_0} \geq s \otimes \nu$. As argued in Karlsen and Tjøstheim (2001), it is not a severe restriction to assume $m_0 = 1$. Therefore, throughout this paper, we assume that the minorization inequality

$$\mathsf{P} \geq s \otimes \nu \tag{2.3}$$

holds with $\nu(\mathbb{E}) = 1$, $0 \le s(x) \le 1$ for any $x \in \mathbb{E}$. We note that the extensions to general m_0 are straightforward (with some involved notations/proofs), but are not treated in this paper.

To develop asymptotic results for the nonparametric estimators, 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 known that a Markov chain defined on a countable $\mathbb{E}$ which has a point of recurrence (a point which the chain can reach within some finite time with probability one; also called an *atom*) can be split into independent and identically distributed (i.i.d.) blocks (c.f., Chung, 1967). To see the idea of this i.i.d. blocking, note that by the Markov property, the behavior of the chain after it reached the point of recurrence is independent of its previous history. The chain is therefore said to be *renewed* or *regenerated* on every visit to the point of recurrence. By splitting the chain by the *regeneration times* (when the chain reaches at the point of recurrence), we can construct i.i.d. blocks. For a general Markov chain whose value space $\mathbb{E}$ is an uncountable set (say, $\mathbb{R}$), we cannot in general find an obvious point of recurrence. However, if the chain possesses the property of Harris recurrence, we can still consider the decomposition of a partial sum of the original Markov chain $\{X_t\}$ into of i.i.d. block parts and remaining negligible parts by using Nummelin's (1984) method.

We here outline Nummelin's decomposition method, while some more details are provided in the Appendix A (see also Section 4.4 of Nummelin, 1984, or Sections 3.1-3.2 of Karlsen and Tjøstheim, 2001). Now, let us introduce an auxiliary chain $\{T_t, t \ge 0\}$ where T_t is a random variable whose value is only 0 or 1: given $X_t = x$, it takes the value 1 with probability $s(x)$ and 0 with $1 - s(x)$, where $s(x)$ is the small function given in (2.3). Under the condition (2.3), we can let the time when $T_t = 1$ be a regeneration time of the augmented/compound chain $\{(X_t, T_t), t \ge 0\}$, i.e., $\alpha = \mathbb{E} \times \{1\}$ be an atom of the chain (we can interpret that $\{(X_t, T_t)\}$ is initialized every time when $T_t = 1$, as explained into details in Appendix A). This chain $\{(X_t, T_t)\}$ is called a split chain (in the terminology of Nummelin 1984's book), since it allows us to split the original chain $\{X_t\}$ into i.i.d. parts and the other negligible parts by the times when $T_t = 1$. The distribution of $\{(X_t, T_t), t \ge 0\}$ is fully characterized by the initial distribution λ of X_0, the transition probability P of $\{X_t\}$ and (s, ν) (see (A.1) in Appendix A).

To explicitly consider the i.i.d. decomposition of $\{X_t\}$, let us define stopping times as follows:

$$\tau_k = \begin{cases} \inf\{t \ge 0 : T_t = 1\}, & k = 0, \\ \inf\{t > \tau_{k-1} : T_t = 1\}, & k \ge 1, \end{cases} \tag{2.4}$$

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

$$N(n) = \begin{cases} \max\{k : \tau_k \leq n\}, & \text{if } \tau_0 \leq n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Let f be a real-valued function on $\mathbb{R}$. By using these $\tau_0, \tau_1, \ldots$ and $N(n)$, we can decompose a sum $\sum_{t=0}^{n} f(X_t)$ into three parts with a main one being a sum of i.i.d. random variables and the other two being asymptotically negligible:

$$\sum_{t=0}^{n} f(X_t) = Z_0 + \sum_{k=1}^{N(n)} Z_k + Z_{(n)}, \tag{2.6}$$

where

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

We can check that $\{Z_k, k \geq 1\}$ is a sequence of i.i.d. random variables (see Appendix A). In this decomposition (2.6), $N(n)$ plays a role as if it were the number of observations in a standard stationary setting. It follows from Lemma 3.2 in Karlsen and Tjøstheim (2001) that Z_0 and $Z_{(n)}$, divided by $N(n)$, converge to zero almost surely. We note that $N(n)$ is stochastic in our general Harris recurrent setting, which fully depends on the structure of the underlying $\{X_t\}$ and its realization. The stochastic growing rate of $N(n)$ is generally unknown while it can be controlled by a condition of the β-null recurrency presented in Definition 2.2 below, which we impose for our subsequent convergence theorems.

As a notable property of a Harris recurrent Markov chain satisfying (2.3), we can find an invariant measure of the chain in terms of s and ν. If we let

$$\pi_s = \nu G_{s,\nu}, \quad \text{where} \quad G_{s,\nu} = \sum_{n=0}^{\infty} (\mathsf{P} - s \otimes \nu)^n, \tag{2.7}$$

then

$$\pi_s = \pi_s \mathsf{P}, \tag{2.8}$$

where $\nu G_{s,\nu}(\mathbb{A}) = \int \nu(dx) G_{s,\nu}(x, \mathbb{A})$ and $\pi_s \mathsf{P}(\mathbb{A}) = \int \pi_s(dx) \mathsf{P}(x, \mathbb{A})$ for any $\mathbb{A} \in \mathcal{E}$. This implies that π_s is an invariant measure of $\{X_t\}$. The formula (2.7) can be interpreted as a natural extension of the standard stationary case, as explained in Appendix A (see also Section 5.2 of Nummelin, 1984, and Section 3.2 of Karlsen and Tjøstheim, 2001). This π_s may not be necessarily a finite measure. It is known that the Harris recurrence includes two sub-cases: positive recurrence and null recurrence (c.f., Meyn and Tweedie, 2009). The former corresponds to the ergodicity with π_s a finite measure, which can be normalized to a probability measure (an ergodic process is strictly stationary if it is initialized by the invariant probability measure); and the latter is the focus of this paper in which π_s is not finite but only σ-finite. The concept of null recurrence provides a general framework for nonstationary time series analyses. It is generally uncertain if we can develop sensible asymptotic theory for nonstationary time series. However, given the null recurrence condition, which allows for quite flexible time series dynamics, we can develop asymptotic results. We note that our subsequent results are also applicable to positive recurrent chains in which π_s can be regarded as the probability measure and the growing rate of $N(n)$ is of the order n (see Remark 3.2).

In the standard stationary time series setting, various nonparametric estimators include a component which estimates the invariant probability density, e.g., the denominator of a Nadaraya-Watson regression estimator. In our general Harris recurrent (nonstationary) setting, $\{X_t\}$ does not necessarily have its invariant probability measure or probability density. However, the density of π_s, which we denote by p_s, plays the same role as the invariant probability density in the standard setting: e.g., the denominator of our Nadaraya-Watson type estimator can be seen as an estimator of p_s upon suitable normalization (the invariant density of the Harris Markov chain is unique up to a multiplicative constant; see Section 3.2 of Karlsen and Tjøstheim, 2001). We note that the ϕ-irreducibility of the Markov chain has the following implication for the invariant measure π_s: ϕ is absolutely continuous with respect to π_s (i.e., if $\phi(\mathbb{A}) > 0$ for any $\mathbb{A} \in \mathcal{E}$, then $\pi_s(\mathbb{A}) = \int_{\mathbb{A}} p_s(x)\,dx > 0$; see Propositions 2.4 and 5.6 of Nummelin, 1984).

The following definition imposes further restrictions on the behavior of the Markov chain.

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

$$\mathsf{E}_\lambda\left[\sum\nolimits_{t=0}^{n} f(X_t)\right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n), \tag{2.9}$$

where E_λ stands for the expectation with initial distribution λ of X_0, $\Gamma(\cdot)$ is the usual Gamma function, $a_n \sim b_n$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$, and a function $l(\cdot)$ is called to be slowly varying (at infinity) if

$$\lim_{x\to\infty} \frac{l(cx)}{l(x)} = 1, \quad \textit{for all } c > 0.$$

For L_f in (2.9), some slowly varying function L_s can be chosen so that for any small function f,

$$L_f = L_s \int f(u)\,\pi_s(du), \tag{2.10}$$

where this result is given as Lemma 3.1 of Karlsen and Tjøstheim (2001). Let $N(n)$ be the number of regenerations of the β-null recurrent Markov chain $\{X_t\}$ as defined in (2.5). The

β-null recurrence condition has the following implication for this $N(n)$, which is important for our subsequent analyses: given the minorization inequality (2.3), $N(n)$ has the asymptotic distribution:

$$\frac{N(n)}{n^{\beta}L_s(n)} \xrightarrow{d} M_{\beta}(1), \tag{2.11}$$

as $n \to \infty$, where $M_{\beta}(1)$ is the Mittag-Leffler distribution with parameter β. The above result is given as Theorem 3.2 of Karlsen and Tjøstheim (2001) (see also Lemma B.4 in Appendix A). For a (strictly) 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

$$X_t = X_{t-1} + u_t, \quad t = 1, 2, \cdots, \quad X_0 = 0, \tag{2.12}$$

where $\{u_t\}$ is a sequence of i.i.d. random variables. Kallianpur and Robbins (1954) showed that this random walk process 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

$$X_t = \alpha_1 X_{t-1} I\{X_{t-1} \in \mathbb{C}\} + \alpha_2 X_{t-1} I\{X_{t-1} \in \mathbb{C}^c\} + v_t, \tag{2.13}$$

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 $\mathsf{E}[v_1] = 0$, $0 < \mathsf{E}[v_1^2] < \infty$, $\mathsf{E}[v_1^4] < \infty$, and the distribution of $\{v_t\}$ is absolutely continuous with respect to the Lebesgue measure with $f(\cdot)$ being its density function satisfying $\inf_{x \in \mathbb{S}} f(x) > 0$ for all compact sets $\mathbb{S}$. Recently, Gao, Tjøstheim and Yin (2013) have shown that $\{X_t\}$ generated by (2.13) is a $\frac{1}{2}$-null recurrent Markov chain.

3 Main results

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

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

where $L(\cdot)$ is a kernel function satisfying Assumption A2(i) below, h is a bandwidth and $N(n)$ is defined in (2.5). To derive uniform consistency results for this nonparametric quantity $\Phi_n(x)$ defined by (3.1), we impose the following assumptions.

ASSUMPTION 1 (i) The invariant measure of the β-null recurrent Markov chain $\{X_t\}$, $\pi_s(\cdot)$, has a uniformly continuous density function $p_s(\cdot)$ on $\mathbb{E} = \mathbb{R}$ with $\sup_{x \in \mathbb{R}} p_s(x) < \infty$.

(ii) For each $z \in \mathbb{R}$, there exists a transition density $p^{(1)}$ satisfying $\mathsf{P}(z, dy) = p^{(1)}(z, y)\, dy$. There exist some (sufficiently small) constants $\kappa, \delta \in (0, 1)$ such that for any z, $\int I\{y \in \mathbb{R} : p^{(1)}(z, y) \geq \kappa\} dy \geq \delta$ (κ and δ are independent of z).

(iii) $\{e_t\}$ is a sequence of independent random variables with $\mathsf{E}[e_t] = 0$ and $\sup_{t\geq 1} \mathsf{E}[e_t^2] < \infty$, and is independent of $\{X_t\}$.

Assumption 2 (i) The kernel function $L(\cdot)$ has compact support $C(L)$, and satisfies a Lipschitz-continuity condition: $|L(x) - L(y)| \leq C_L |x - y|$ for all $x, y \in C(L)$ and some constant $C_L > 0$.

(ii) For each $x \in [-\mathcal{T}_n, \mathcal{T}_n]$, $\mathcal{N}_x = \{z \in \mathbb{R} : |x - z| \leq 1\}$ is a small set of the Markov chain $\{X_t\}$, where $\mathcal{T}_n = n^{1-\beta}\mathcal{L}_*(n)$ and $\mathcal{L}_*(n)$ is a sequence of positive and slowly varying functions.

Remark 3.1. (i) Assumption 1(i) corresponds to the analogous conditions in the density function estimation for the stationary time series case. It can be satisfied by a random walk process $\{X_t\}$ as in Example 2.1, whose invariant density function is $p_s(x) = 1$, as argued in Nummelin (1984, page 75).

Assumption 1(ii) might be seen as an unfamiliar condition on the transition density, but it is satisfied by various types of processes, including ones in Examples 2.1 and 2.2 (given the existence of the probability density of the error disturbances). An important implication of this condition is that it regulates tail decay properties of the transition density. To see this point, consider the following case in which the condition is violated:

$$\sup_{y\in\mathbb{R}} p^{(1)}(z, y) \to 0 \quad \text{as} \quad |z| \to \infty, \tag{3.2}$$

for example. If this holds true, we cannot find any super κ-level set of $p^{(1)}(z, \cdot)$ which is bounded away from zero by δ (note that even when (3.2) is true, we can still find κ and δ, if they are allowed to depend on z; but Assumption 1(ii) requires that κ and δ to be uniform over z). By the fact that $p^{(1)}(z, \cdot)$ is the probability density, (3.2) means that the tail decay rate of $p^{(1)}(z, \cdot)$ becomes slower as $|z| \to \infty$ (in other words, the tail of the conditional distribution is relatively thicker for larger $|z|$). If the degree of tail-thickness of the transition density is the same for all z, we can check that Assumption 1(ii) is indeed satisfied.

Assumption 1(iii) allows for the case where the compound process $\{(X_t, e_t)\}$ is not β-null recurrent. While the β-null recurrency of the compound process seems often necessary for obtaining sensible distribution theory of estimators as in Section 5 of Karlsen and Tjøstheim (2001), we can work without it for the purpose here to derive uniform convergence rates. While the independence between $\{e_t\}$ and $\{X_t\}$ greatly simplify our proofs, we can think of some ways to relax it. One way is to suppose a certain dependence structure such as

$$e_t = \sigma(X_t)\epsilon_t, \tag{3.3}$$

where $\{\epsilon_t\}$ is a sequence of independent random variables and is independent of $\{X_t\}$, and $\sup_{x\in\mathbb{R}} |\sigma(x)| < \infty$. Although some form of heterogeneity is already allowed (i.e., $e_1, e_2, \ldots$ need not to be identically distributed), this (3.3) allows heterogeneity through the dependence on X_t. We note that all the results stated in this and subsequent sections carry over to the case with (3.3). Another way to relax the independence is to simply impose the β-null recurrency condition on the compound process $\{(X_t, e_t)\}$ or that of $\{(X_t, \epsilon_t)\}$ for the case of (3.3), under which we can apply the Markov splitting technique to $\{(X_t, e_t)\}$ or $\{(X_t, \epsilon_t)\}$ (this condition is satisfied, for example, under the case where $\{e_t\}$ or $\{\epsilon_t\}$ is an i.i.d. sequence and is independent of the β-null recurrent chain $\{X_t\}$, while this case for $\{e_t\}$ is covered by Assumption 1(iii)). As another relaxation of Assumption 1(iii), we might be able to consider some time-series dependence structure of $\{e_t\}$ (say, stationarity and weak dependence, such as α-mixing), which is, however, probably not trivial to work with (requiring some different conditions and/or resulting in different convergence rates).

(ii) Assumption 2(i) is a commonly used condition on the kernel function. As discussed in condition B_2 in Section 5 of Karlsen and Tjøstheim (2001), Assumption 2(ii) is needed in this kind of kernel estimation of null-recurrent time series. The small set requirement is a weak condition when combined. For example, if $\{X_t\}$ is an autoregressive process given by $X_t = g(X_{t-1}) + v_t$, a sufficient condition for the smallness of $\mathcal{N}_x$ is that $g(\cdot)$ is bounded on compact sets and that $\{v_t\}$ has its density with respect to the Lebesgue measure which is strictly positive on any compact set (c.f., Doukhan and Ghindés, 1980; pages 589-590 of Tjøstheim, 1990). We can also find various other sufficient conditions in Section 2.3 of Nummelin (1984). We note that the compact support condition of the kernel function is important in our setup since it is associated with the small-set requirement (see the proof of Lemma B.2). While Hansen (2008) allows for kernels whose support is unbounded and whose tail decay rate is fast enough (say, the normal kernel), it is uncertain if we can work with such kernels.

Hansen (2008) considered uniform consistency for a nonparametric estimator of the form

$$\Psi_*(x) = \frac{1}{nh}\sum_{t=1}^{n} L\left(\frac{X_t - x}{h}\right) Y_t,$$

where $\{(X_t, Y_t) : t \geq 1\}$ is stationary and α-mixing, and n is the sample size. He showed both weak and strong uniform consistency results (Theorems 2-3). In Theorem 3.1 below, we establish a weak uniform consistency result for the nonparametric quantity defined in (3.1), which corresponds to Hansen's $\Psi_*(x)$.

Theorem 3.1. *Suppose that Assumptions 1-2 hold, and that*

$$\begin{aligned} &1/h < n^{\beta-\varepsilon_0} \quad \textit{for some } \varepsilon_0 \in (0, \beta)\,, \\ &\sup_{t\geq 1} \mathsf{E}[|e_t|^{2p_0}] < \infty \quad \textit{with some positive integer } p_0 > \frac{2}{\varepsilon_0} - \frac{1}{2}. \end{aligned} \tag{3.4}$$

Then, as $n \to \infty$ and $h \to 0$,

$$\sup_{|x|\leq \mathcal{T}_n} |\Phi_n(x)| = O_P\Big(\sqrt{\frac{\log n}{n^\beta L_s(n) h}}\Big), \tag{3.5}$$

where $\mathcal{T}_n = n^{1-\beta}L_*(n)$ (defined in Assumption 2(ii)), and L_s is a slowly varying function introduced in (2.10).

REMARK 3.2. (i) Theorem 3.1 can be seen as a nonstationary (and null recurrent) extension of the corresponding results in the stationary time series case. When $\beta = 1$ and $L_s(\cdot)$ is a non-zero constant, the result (3.5) is reduced to a standard result in the stationary case (c.f., Theorem 2 in Hansen, 2008). Letting $h \sim \left(\frac{\log n}{n^{\beta}L_s(n)}\right)^{1/5}$, the right hand side (RHS) of (3.5) is $\left(\frac{\log n}{n^{\beta}L_s(n)}\right)^{2/5}$, which is reduced to the optimal rate in the stationary case when $\beta = 1$ and $L_s(\cdot)$ is a non-zero constant (c.f., Stone, 1980).

To compare our result (3.5) with Hansen's (2008), consider an autoregressive Markov process $\{X_t\}$ described by $X_t = aX_{t-1} + u_t$, where $|a| < 1$, $\{u_t\}$ is a sequence of i.i.d. error disturbances whose probability density exists, and X_{t-1} and u_t are independent. We also assume that $\{e_t\}$ is i.i.d. and independent of $\{X_t\}$. Then, we can check that $\{(X_t, e_t)\}$ is geometrically strong (α) mixing satisfying Hansen (2008)'s condition (c.f., Section 2.4 of Doukhan, 1994). At the same time, this Markov process $\{X_t\}$ satisfies the minorizaton inequality (2.3) (this can be verified under several weak conditions on the density of u_t; see Example 3.1 of Karlsen and Tjøstheim, 2001), and $1/N(n) = O_P(1/n)$ with $\beta = 1$ and $L_s(\cdot)$ is a non-zero constant. In this example, given the geometric decay rate of α-mixing coefficients, Theorem 2 in Hansen (2008) tells us that $\Phi_n(x) = O_P\left(\frac{1}{nh}\sum_{t=0}^{n} L\left(\frac{X_t - x}{h}\right)e_t\right)$ has the sharp convergence rate of $O_P\left(\sqrt{\frac{\log n}{nh}}\right)$ under fairly weak moment and bandwidth conditions, say, it suffices to have the existence of the first-order moment of e_t and $(\log n)/nh \to 0$. On the other hand, for such a sharp rate of $\Phi_n(x)$, our Theorem 3.1 requires the existence of higher-order moments (in particular when we want to have a weaker bandwidth condition, $1/h < n^{\beta-\varepsilon_0}$ with a smaller ε_0). This contrast is due to the fact that our proof relies on the Markov properties of $\{X_t\}$, but does not utilize its mixing properties, meaning that in our theorem, the fast decay rate of the mixing coefficients does not help to improve moment/bandwidth conditions or convergence rates (we note that the same remark on the comparison between Hansen (2008)'s results and ours also applies to the subsequent theorems). While our results may be less sharp than Hansen (2008)'s for stationary and mixing processes, they can cover null recurrent processes, which are nonstationary and are not in the scope of Hansen (2008)'s theorems.

(ii) The condition in (3.4) indicates that there exists a trade-off between the bandwidth condition and the moment condition on $\{e_t\}$. As ε_0 decreases (the bandwidth condition becomes weaker), we need a stronger moment condition on $\{e_t\}$.

(iii) If $\beta = \frac{1}{2}$ with a non-zero constant $L_s(\cdot)$, for example (see the random walk process defined in Example 2.1), the rate of convergence in (3.5) is $O_P\left(\sqrt{\frac{\log n}{\sqrt{n}h}}\right)$, which is in contrast to $O_P\left(\sqrt{\frac{\log n}{nh}}\right)$ for the stationary time series case. This slower convergence rate occurs since the amount of time spent by a $\frac{1}{2}$-null recurrent process around any particular point is of order $\sqrt{n}$ rather than n for the stationary case.

The next theorem provides a strong uniform rate of convergence under slightly different conditions on the bandwidth and the moment of $\{e_t\}$.

Theorem 3.2. *Suppose that Assumptions 1-2 hold, and that*

$$\begin{aligned} &1/h < n^{\beta\delta-\varepsilon_0} \quad \textit{for some } \varepsilon_0 \in (0, \beta\delta) \textit{ and some } \delta \in (0,1)\,, \\ &\sup_{t\ge 1} \mathsf{E}[|e_t|^{2m_0}] < \infty \quad \textit{with some positive integer } m_0 > \frac{3-\varepsilon_0}{\beta\,(1-\delta)} - \frac{1}{2}. \end{aligned} \tag{3.6}$$

Then, as $n \to \infty$ and $h \to 0$,

$$\sup_{|x|\le \mathcal{T}_n} |\Phi_n(x)| = o\Big(\frac{1}{\sqrt{n^{\beta-\varepsilon_0} h}}\Big) \quad a.s., \tag{3.7}$$

where $\mathcal{T}_n$ is defined in Assumption 2(ii).

Remark 3.3. (i) The result (3.7) can be seen as an extension of some existing results for the stationary time series case (c.f., Theorem 3 of Hansen, 2008). The rate of convergence in (3.7) is very close to the sharp rate obtained in Theorem 3.1 when ε_0 is close to zero. We note that as in the previous theorem, we can see a trade-off between the bandwidth and moment conditions (m_0 has to be larger for smaller ε_0 in (3.6)).

(ii) The requirement for m_0 is not easy to compare with that for p_0 of Theorem 3.1, since the convergence rates of $\Phi_n(x)$, as well as the imposed bandwidth conditions, are different between two theorems. However, given (almost) the same bandwidth rate, Theorem 3.2 generally requires a stronger moment condition than Theorem 3.1, which seems to be a price for obtaining an almost sure result. To see this point, suppose that one selects some ϵ and a bandwidth h satisfying

$$1/h < n^{\beta-\epsilon} \tag{3.8}$$

in Theorem 3.1. Then, by having δ close to 1 in the bandwidth condition of Theorem 3.2, we can say that (3.8) is approximately satisfied, but this leads to a larger value of m_0 in (3.6) (typically than p_0), i.e., as $\delta \to 1$, $m_0 \to \infty$.

(iii) The parameter $\delta \in (0,1)$ is involved in our bandwidth condition $1/h < n^{\beta\delta-\varepsilon_0}$. Unlike the case in Theorem 3.1, this sort of parameter seems necessary for controlling the shrinking rate of h and obtaining the rate in (3.7). Our condition is similar to ones imposed in Karlsen and Tjøstheim (2001). For example, their strong consistency result (Theorem 5.2, page 406) requires that $1/h < n^{\beta\delta_m-\varepsilon}$, where the existence of the $2m$-order moment is assumed and $\delta_m \in (0,1/2]$ is a certain parameter which depends up on m (c.f., a bandwidth condition C6 on page 260 of Karlsen, Myklebust and Tjøstheim, 2007). Unlike Karlsen and Tjøstheim's δ_m, our δ can be an arbitrary value in $(0,1)$ (given m_0 sufficiently large), allowing for a more flexible choice of the bandwidth h.

4 Applications in density and regression estimation

In this section, we consider estimating the invariant density and regression functions. We first investigate the following object:

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

where $K(\cdot)$ is a kernel function. This can be regarded as an estimator of the invariant density function. The quantity $N(n)$ is unobservable and $\hat{p}_n(x)$ is an infeasible estimator. However, convergence properties of $\hat{p}_n(x)$ have important implications for characterizing those of some observable quantity (see Remark 4.1(ii) below). We derive weak and strong uniform convergence rates of $\hat{p}_n(x)$ with the following additional conditions imposed:

Assumption 3 (i) $p_s(x)$ is thrice continuously differentiable with $\sup_{x\in\mathbb{R}}[|p_s'(x)| + |p_s''(x)| + |p_s'''(x)|] \le C_p$ for some $C_p \in (0,\infty)$.

(ii) The kernel function $K(\cdot)$ is a symmetric probability density function, and satisfies Assumption 2 with $K = L$.

Theorem 4.1. *Suppose that Assumptions 1-3 hold.*

(i) If the bandwidth satisfies $1/h < n^{\beta-\varepsilon_0}$ for some $\varepsilon_0 \in (0,\beta)$, then, as $n \to \infty$ and $h \to 0$,

$$\sup_{|x|\le\mathcal{T}_n} |\hat{p}_n(x) - p_s(x)| = O_P\left(h^2\right) + O_P\Big(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\Big). \tag{4.2}$$

(ii) If the bandwidth satisfies $1/h < n^{\beta\delta-\varepsilon_0}$ for some $\varepsilon_0 \in (0,\beta\delta)$ and some $\delta \in (0,1)$, then, as $n \to \infty$ and $h \to 0$,

$$\sup_{|x|\le\mathcal{T}_n} |\hat{p}_n(x) - p_s(x)| = O(h^2) + o\Big(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\Big) \quad a.s. \tag{4.3}$$

Remark 4.1. (i) This theorem can be seen as a nonstationary extension of Theorem 5.3 in Fan and Yao (2003) and Theorems 6 and 7 in Hansen (2008). The bandwidth condition for the probability-convergence result (i) of Theorem 4.1 is slightly weaker than that for the almost sure result (ii), which is also the case in Hansen (2008)'s theorems. Karlsen and Tjøstheim (2001, Theorem 5.1) obtained the strong pointwise consistency of $\hat{p}_n(x)$ for the null recurrent time series case, with imposing

$$n^{(\beta/2)-\varepsilon_0}h \to \infty \quad \text{for } \varepsilon_0 \in (0,\beta/2).$$

Our strong uniform consistency result, Theorem 4.1(ii), not only weakens Karlsen and Tjøstheim (2001)'s bandwidth condition (our δ is allowed to be arbitrary as long as it is in $(0,1)$), but also strengthens the pointwise consistency to the uniform consistency with possible rates.

(ii) As stated above, $\hat{p}_n(x)$ may be thought of as a sort of theoretical object due to the unobservability of $N(n)$. However, we note that $N(n)$ is linked to an observable hitting time.

To see this, let $\mathbb{C}_* \in \mathcal{E}^+$ and $N_{\mathbb{C}_*}(n) = \sum_{t=0}^{n} I\{X_t \in \mathbb{C}_*\}$, the number of times that the process is visiting $\mathbb{C}_*$ up to the time $t = n$. By Lemma 3.2 in Karlsen and Tjøstheim (2001), $N_{\mathbb{C}_*}(n)$ satisfies

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

Let

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

and then observe that

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

In this case, by (4.2)-(4.5) with the assumption that $\pi_s I_{\mathbb{C}_*} > 0$, we have

$$\sup_{|x| \le \mathcal{T}_n} \left|\hat{p}_n^{\mathbb{C}_*}(x) - p_s(x)/(\pi_s I_{\mathbb{C}_*})\right| = O_P\left(h^2\right) + O_P\Big(\sqrt{\frac{\log n}{n^\beta L_s(n) h}}\Big) \tag{4.6}$$

and

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

We next consider a nonlinear nonstationary regression model:

$$Y_t = m(X_t) + e_t, \tag{4.8}$$

where $0 \le t \le n$, $\{X_t\}$ is a β-null recurrent Markov chain, $\{e_t\}$ is a sequence of independent errors with $\mathsf{E}[e_1] = 0$ and $0 < \mathsf{E}[e_1^2] < \infty$, $m(\cdot)$ is an unknown function, and $\{e_t\}$ is independent of $\{X_t\}$. This sort of nonlinear cointegration model has been studied by several authors. For example, Karlsen, Myklebust and Tjøstheim (2007), and Wang and Phillips (2009a) considered the Nadaraya-Watson (NW) estimator of the form

$$\hat{m}(x) = \sum_{t=0}^{n} w_{n,t}(x) Y_t, \tag{4.9}$$

where

$$w_{n,t}(x) = K\left(\frac{X_t - x}{h}\right) \Big/ \sum_{s=0}^{n} K\left(\frac{X_s - x}{h}\right),$$

and they obtained asymptotic distributions for $\hat{m}(x)$ using different methods. As an application of Theorems 3.1 and 3.2, we provide both the weak and strong uniform consistency results for the NW estimator $\hat{m}_n(x)$.

Theorem 4.2. *Suppose that Assumptions 1-3 hold. Let*

$$\delta_n = \inf_{|x| \le \mathcal{T}_n} p_s(x) > 0, \quad \textit{and} \quad \rho_{i,n} = \sup_{|x| \le \mathcal{T}_n} \left|\left(d^i/dx^i\right) m(x)\right| / \delta_n \quad \textit{for } i = 1, 2, \tag{4.10}$$

and suppose also that $\delta_n^{-1} h^2 = o(1)$.

(i) If the conditions in (3.4) of Theorem 3.1 are satisfied, and $\delta_n^{-1}\Big(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\Big) = o(1)$, *then, as* $n\to\infty$ *and* $h\to 0$,

$$\sup_{|x|\le\mathcal{T}_n}|\hat m_n(x)-m(x)| = O_P\Big(\left[\delta_n^{-1}+\rho_{1,n}h\right]\sqrt{\frac{\log n}{n^\beta L_s(n)h}}+\left[\rho_{1,n}+\rho_{2,n}\right]h^2\Big). \tag{4.11}$$

(ii) If the conditions in (3.6) of Theorem 3.2 are satisfied, and $\frac{\delta_n^{-1}}{\sqrt{n^{\beta-\varepsilon_0}h}} = O(1)$, *then, as* $n\to\infty$ *and* $h\to 0$,

$$\sup_{|x|\le\mathcal{T}_n}|\hat m_n(x)-m(x)| = o\Big(\frac{\delta_n^{-1}+\rho_{1,n}h}{\sqrt{n^{\beta-\varepsilon_0}h}}\Big)+O(\left[\rho_{1,n}+\rho_{2,n}\right]h^2)\quad a.s. \tag{4.12}$$

Remark 4.2. (i) The conditions imposed for the establishment of Theorem 4.2 are reasonable and justifiable. The condition of $\delta_n^{-1}h^2 = o(1)$ can be easily verified when the regressor $\{X_t\}$ is a process as in Example 2.1 or Example 2.2. We can check $p_s(x)\equiv 1$ in the first example, and $p_s(x)\to 1$ as $|x|\to\infty$ in the second one. For the consistency, the components involving $\rho_{i,n}$ on the RHS of (4.11) and (4.12) have to shrink to zero, which 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 $m(x) = O\left(|x|^{1+\zeta}\right)$ for some $0<\zeta<1$ when x is large enough. In particular, when $m(x) = a+bx$ and $\{X_t\}$ is generated by a data generating process in Example 2.1 or Example 2.2, the rates on the RHS of (4.11) and (4.12) are reduced to $O_P\Big(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}+h^2\Big)$ and $o\Big(\frac{1}{\sqrt{n^{\beta-\varepsilon_0}h}}\Big)+O(h^2)$, respectively, which do not involve the penalty terms δ_n and $\rho_{i,n}$. In contrast, the convergence rates in Theorems 8 and 9 of Hansen (2008) are also penalized by δ_n (the component due to the invariant density $p_s(\cdot)$) but not by $\rho_{i,n}$ (the ones due to $m(\cdot)$). This is because he supposes the uniform boundedness of $m(x)\,p_s(x)$, which is reasonable since $p_s(\cdot)$ is the probability density in the stationary case. If we work with this sort of boundedness, we can also write our convergence rates without $\rho_{i,n}$. However, as discussed above, in our null recurrent case, $p_s(x)$ is not generally a probability density and the boundedness of $m(x)\,p_s(x)$ is not likely to be satisfied. From this reason, our theorems are written with using $\rho_{i,n}$.

(ii) Theorem 4.2 can be viewed as a nonstationary extension of Theorem 3.3 in Bosq (1998) and Theorems 8 and 9 in Hansen (2008) for the stationary time series regression case. When $\{X_t\}$ is a random walk as in 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 consider local linear estimation of $m(\cdot)$ and present both weak and strong uniform consistency results of a proposed estimator. As in Fan and Gijbels (1996), we define our local linear estimator of $m(x)$ as

$$\tilde m_n(x) = \sum_{t=0}^{n}\tilde w_{n,t}(x)Y_t, \tag{4.13}$$

where

$$\tilde{w}_{n,t}(x) = \frac{\frac{1}{h}K\left(\frac{X_t - x}{h}\right)\left[S_{n,2}(x) - \left(\frac{X_t - x}{h}\right)S_{n,1}(x)\right]}{\sum_{s=0}^{n}\frac{1}{h}K\left(\frac{X_s - x}{h}\right)\left[S_{n,2}(x) - \left(\frac{X_s - x}{h}\right)S_{n,1}(x)\right]},$$

$$S_{n,j}(x) = \frac{1}{N(n)h}\sum_{s=0}^{n} K\left(\frac{X_s - x}{h}\right)\left(\frac{X_s - x}{h}\right)^j \quad \text{for } j = 1, 2.$$

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

THEOREM 4.3. *Suppose that the conditions of Theorem 4.2 hold. Let δ_n and $\rho_{2,n}$ be quantities defined in (4.10), and suppose also that $\delta_n^{-1}h = o(1)$.*

(i) If the conditions in (3.4) of Theorem 3.1 are satisfied, and $\delta_n^{-1}\left(\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\right) = o(1)$, then as $n \to \infty$ and $h \to 0$,

$$\sup_{|x| \le \mathcal{T}_n} |\tilde{m}_n(x) - m(x)| = O_P\Big(\delta_n^{-1}\sqrt{\frac{\log n}{n^\beta L_s(n)h}}\Big) + O_P(\rho_{2,n}\delta_n^{-1}h^2). \tag{4.14}$$

(ii) If the conditions in (3.6) of Theorem 3.2 are satisfied, and $\frac{\delta_n^{-1}}{\sqrt{n^{\beta-\varepsilon_0}h}} = O(1)$, then, as $n \to \infty$ and $h \to 0$,

$$\sup_{|x| \le \mathcal{T}_n} |\tilde{m}_n(x) - m(x)| = o\Big(\frac{\delta_n^{-1}}{\sqrt{n^{\beta-\varepsilon_0}h}}\Big) + O(\rho_{2,n}\delta_n^{-1}h^2) \quad a.s. \tag{4.15}$$

REMARK 4.3. (i) Comparing to (4.11) and (4.12) in the NW estimation case, we do not have a term involving $\rho_{1,n}$ in (4.14) and (4.15). This is due to the first-order bias correction property of the local linear method. The penalty associated with the h^2 components is strengthened to $\rho_{2,n}\delta_n^{-1}$ here (instead of $\rho_{2,n}$ in (4.11) and (4.12)). This is because, roughly speaking, the denominator of the weight $\tilde{w}_{n,t}(x)$ converges to the product of $p_s^2(x)$ and some constant. In conjunction with this, the required bandwidth condition here ($\delta_n^{-1}h = o(1)$) is slightly stronger than the condition of Theorem 4.2 ($\delta_n^{-1}h^2 = o(1)$). These sorts of stronger requirements in the local linear estimation can be also observed in Hansen (2008).

(ii) Note that Assumption 1(iii) does not allow for the autoregression case with $X_t = Y_{t-1}$ in (4.8) since it requires the independence between $\{X_t\}$ and $\{e_t\}$. However, if the functional form of $m(\cdot)$ and the sequence $\{e_t\}$ are chosen such that $\{Y_t\}$ is a β-null recurrent Markov chain, then we can still verify the conclusions of Theorems 4.2 and 4.3 under some modified conditions. For example, if $\{e_t\}$ is a sequence of independent and bounded random variables or we have $\{(Y_t, e_t)\}$ a β-null recurrent Markov chain, almost all the proofs remain the same, requiring only minor modifications.

5 Conclusions

We have established several results of both 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 case to the nonstationary 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 considered only the case where $\{X_t\}$ is univariate. If $\{X_t\}$ is multivariate and satisfies the Harris recurrent condition, we can still apply our Markov splitting technique and derive some corresponding uniform convergence rate. However, it is not necessarily easy to check such a condition for the multivariate case. For example, if $\{X_t\}$ is a multivariate random walk, it may be transient. We refer to Schienle (2008, 2011), and Cai, Li and Park (2009) for the case where multivariate nonstationary regressors are involved.

Acknowledgments

The authors would like to thank the Editor, Professor Peter C. B. Phillips, and referees for their constructive and valuable comments, which have greatly improved the original version of this paper. Two Australian Research Council Discovery Grants (DP0558602 and DP0879088) and an Australian Research Council Discovery Early Career Researcher Award (DE120101130) provided financial support for this work. Kanaya gratefully acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78).

REFERENCES

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

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 (2012) Estimation in semiparametric regression with nonstationary regressors. *Bernoulli* 18, 678–702

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. (1994) Mixing, Properties and Examples. Lecture Notes in Statistics 85, Springer–Verlag.

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.

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 (2013) Estimation in threshold autoregressive models with nonstationarity. *Journal of Econometrics 172,* 1–12.

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.

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.

Meyn, S.P. & R.L. Tweedie (2009) *Markov Chains and Stochastic Stability*, Second Edition. Cambridge

University Press.

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.* PhD Thesis, University of Mannheim, Germany.

Schienle, M. (2011) Nonparametric nonstationary regression with many covariates. SFB 649 Discussion Paper available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2011-076.pdf.

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 review some useful terms and facts of Markov theory in this Appendix. More details on the Markov splitting method and related results, can be found in Sections 4.4 and 5.2 of Nummelin (1984) and Section 3 of Karlsen and Tjøstheim (2001).

Throughout the Appendices, we use Pr and E to stand for unconditional probabilities and expectation of the Markov chain, as well as those of the (augmented) split chain. We often write E_λ for the expectation with the initial distribution λ of X_0. When $\lambda = \delta_x$, we write E_x instead of E_{δ_x}, which is the conditional expectation of the (split) chain given $X_0 = x$. When the split chain in the atom $\alpha = \mathbb{E} \times \{1\}$ (i.e., $X_0 = x$ for arbitrary $x \in \mathbb{E}$ and $T_0 = 1$), we write E_α.

Let $\{X_t, t \geq 0\}$ be a Markov chain with its state space $(\mathbb{E}, \mathcal{E})$, the transition probability $\mathsf{P}(\mathbb{E} \times \mathcal{E} \to [0,1])$, and ϕ be a measure on $(\mathbb{E}, \mathcal{E})$. As stated in Section 2, we assume that $\{X_t\}$ is a ϕ-irreducible Harris recurrent Markov chain. Throughout this Appendix and subsequent proofs, we often use the following notations and definitions. Let η be a nonnegative measurable function and λ be a measure on $(\mathbb{E}, \mathcal{E})$. We define a kernel $\eta \otimes \lambda$ $(\mathbb{E} \times \mathcal{E} \to \mathbb{R})$ as

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

For a kernel K, we define the function $K\eta$, the measure λK, and the number $\lambda\eta$ as

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

The convolution of two kernels K_1 and K_2 is defined as

$$K_1K_2(v, \mathbb{A}) = \int K_1(v, dy)K_2(y, \mathbb{A}),$$

which creates a new kernel: $\mathbb{E} \times \mathcal{E} \to \mathbb{R}$.

As outlined in Section 2, we will apply the so-called Markov chain splitting method to prove our asymptotic results. In this method, the minorization inequality (2.3), $\mathsf{P} \geq s \otimes \nu$ (with $\nu(E) = 1$, $0 \leq s(x) \leq 1$), plays an important role, under which we can consider the decomposition of the chain into i.i.d. main parts and the other asymptotically negligible parts.

We now provide the details on the construction of the split chain $\{(X_t, T_t)\}$. As stated in Section 2, we let $\{T_t\}$ be an auxiliary chain, and each T_t be a random variable which takes $T_t = 1$ with probability $s(x)$ and $T_t = 0$ with $1 - s(x)$ given $X_t = x$. Define

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

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

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

When the minorization inequality (2.3) holds, we can easily verify that Q is a transition probability. As $0 \leq s(x) \leq 1$ and $\nu(E) = 1$, P can be seen as a mixture of the transition probability Q and the measure ν. This means that X_{t+1} moves according to Q if $T_t = 0$, which happens with probability $1 - s(x)$, and it moves according to ν if $T_t = 1$, which happens with probability $s(x)$. We here note that ν is independent of x, and therefore, the behavior of $(X_{t+1}, X_{t+2}, \ldots)$ after the occurrence of $T_t = 1$ is independent of the previous history; in other words, $\{X_t\}$ regenerates (is initialized) every time when $T_t = 1$ (or ν) is chosen. From these arguments, we can see that the dynamics of $\{(X_t, T_t)\}$ is, equipped with the initial distribution λ of X_0, fully determined by P and (s, ν). That is, we can write

$$\begin{aligned}
\Pr(X_0 \in \mathbb{A}) &= \lambda(\mathbb{A}), \\
\Pr(T_t = y \mid X_t, X_{t-1}, \ldots, T_{t-1}, T_{t-2}, \ldots) &= s(X_t)\, y + (1 - s(X_t))(1 - y), \quad t \geq 0, \\
\Pr(X_t \in \mathbb{A} \mid X_{t-1}, X_{t-2}, \ldots, T_{t-1}, T_{t-2}, \ldots) &= \nu(\mathbb{A})\, T_{t-1} + \mathsf{Q}(X_{t-1}, \mathbb{A})(1 - T_{t-1}), \quad t \geq 1,
\end{aligned} \tag{A.1}$$

where $y \in \{0, 1\}$.

For computing some (conditional) expectations of $\{X_t\}$ and several other objects, it is useful to define the following stopping times:

$$\tau = \min\{t \geq 0 : \ T_t = 1\}, \tag{A.2}$$

$$S_\alpha = \min\{t \geq 1 : \ T_t = 1\}. \tag{A.3}$$

We note that $\tau = \tau_0$ (defined in (2.4)), and that by the recurrent property of the chain, we have $\tau < \infty$ and $S_\alpha < \infty$ almost surely. For any set $\mathbb{A} \in \mathcal{E}$, let

$$\pi_s(\mathbb{A}) = \mathsf{E}_\alpha \left[\sum\nolimits_{n=1}^{S_\alpha} I(X_n \in \mathbb{A}) \right]. \tag{A.4}$$

The measure π_s is identical to the invariant measure object defined in (2.7). The RHS of (A.4) is the expected number of visits of $\{X_t\}$ to $\mathbb{A}$ between two consecutive occurrences of the event $\{T_t = 1\}$. It can be written as in (2.7), $\pi_s(\mathbb{A}) = \nu G_{s,\nu}(\mathbb{A})$, since, given $T_0 = 1$ (or equivalently $(X_0, T_0) \in \alpha$), the behavior of X_1 is determined by ν and the probability that the split chain subsequently does not fall into the atom α is calculated based on $G_{s,\nu} = \sum_{n=0}^{\infty} (\mathsf{P} - s \otimes \nu)^n$, i.e.,

$$\begin{aligned}
&\mathsf{E}_\alpha \left[\sum\nolimits_{n=1}^{S_\alpha} I(X_n \in \mathbb{A}) \right] = \mathsf{E}_\alpha \left[\sum\nolimits_{n=1}^{\infty} I(X_n \in \mathbb{A}) I(S_\alpha \geq n) \right] \\
&= \sum\nolimits_{n=1}^{\infty} \nu(dx) (\mathsf{P} - s \otimes \nu)^{n-1}(x, \mathbb{A}) \\
&= \sum\nolimits_{n=0}^{\infty} \nu(dx) (\mathsf{P} - s \otimes \nu)^{n}(x, \mathbb{A}) = \int \nu(dx) G_{s,\nu}(x, \mathbb{A}).
\end{aligned} \tag{A.5}$$

We can also interpret (A.4) as a null recurrent analogue of the following stationary case:

$$\tilde{\pi}_s(\mathbb{A}) = \mathsf{E}[I(\tilde{X}_n \in \mathbb{A})],$$

where $\{\tilde{X}_n\}$ is a strictly stationary process with its stationary probability measure $\tilde{\pi}_s$. By noting the i.i.d. property of the blocks and the fact that $(X_{\tau_{k-1}}, T_{\tau_{k-1}}) \in \alpha$, we also have

$$\mathsf{E}_\alpha \left[\sum\nolimits_{t=1}^{S_\alpha} I(X_t \in \mathbb{A}) \right] = \mathsf{E} \left[\sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k} I(X_t \in \mathbb{A}) \right] = \mathsf{E}_\nu \left[\sum\nolimits_{t=0}^{\tau} I(X_t \in \mathbb{A}) \right], \tag{A.6}$$

for any $k \geq 1$.

Now, let g be a π_s-integrable function on $\mathbb{R}$, i.e., $\pi_s g = \int \pi_s(dz) g(z) < \infty$. Analogously to (A.5), we can see that

$$\mathsf{E}_z \left[\sum\nolimits_{t=0}^{\tau} g(X_t) \right] = \mathsf{E}_z \left[\sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k} g(X_t) \right] = G_{s,\nu} g(z). \tag{A.7}$$

For details of this, we refer to page 379 of Karlsen and Tjøstheim (2001). This, together with the definition of π_s in (2.7), implies

$$\begin{aligned}
\pi_s g &= \int g(z) \pi_s(dz) = \int \left(\int g(z) G_{s,\nu}(x, dz) \right) \nu(dx) \\
&= \int \left(\int G_{s,\nu}(x, dz) g(z) \right) \nu(dx) = \int G_{s,\nu} g(x) \nu(dx) \\
&= \int \mathsf{E}_x \left[\sum\nolimits_{t=0}^{\tau} g(X_t) \right] \nu(dx) = \mathsf{E}_\nu \left[\sum\nolimits_{t=0}^{\tau} g(X_t) \right]
\end{aligned} \tag{A.8}$$

The expressions of (A.7) and (A.8) are often used in subsequent proofs (we will also write $\int G_{s,\nu}g(x)\nu(dx) = \nu G_{s,\nu}g$ in the sequel, following the notations presented above).

Appendix B: Proofs of main results

To prove the main results in Sections 3 and 4, we use the following four lemmas.

Lemma B.1. *Suppose that $\{X_t\}$ is a β-null recurrent Markov chain, and Assumptions 1(i)-(ii) and 2 are satisfied. Let*

$$L_{h,x}(X_t) = \frac{1}{h}L\left(\frac{X_t - x}{h}\right). \tag{B.1}$$

Then, it holds that uniformly for $x \in \left[-n^{1-\beta}\mathcal{L}_*(n), n^{1-\beta}\mathcal{L}_*(n)\right]$,

$$\mathsf{E}\left[\left(\sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k} |L_{h,x}(X_t)|\right)^{2m}\right] \le Mh^{-2m+1}, \tag{B.2}$$

for any integer $m \ge 1$, where M is some positive constant which depends on m but is independent of k, h and x.

Proof. The following arguments are similar to those in the proof of Lemma 5.2 of Karlsen and Tjøstheim (2001). By the i.i.d. property of the blocks of the split chain, (A.1) and (A.6), we have

$$\begin{aligned}
&\mathsf{E}\left[\left(\sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k} |L_{h,x}(X_t)|\right)^{2m}\right] = \mathsf{E}_\nu\left[\left(\sum\nolimits_{t=0}^{\tau} |L_{h,x}(X_t)|\right)^{2m}\right] \\
&= \mathsf{E}_\nu\left[\left\{\sum\nolimits_{t=0}^{\infty}[\prod\nolimits_{s=0}^{t-1} I(T_s = 0)]\,|L_{h,x}(X_t)|\right\}^{2m}\right] \le \mathsf{E}_\nu\left[\left\{\sum\nolimits_{t=0}^{\infty} \mathcal{B}_t\,|L_{h,x}(X_t)|\right\}^{2m}\right],
\end{aligned} \tag{B.3}$$

where $\mathcal{B}_0 = 1$ and $\mathcal{B}_t = \prod_{s=0}^{t-1} I(T_s = 0)$. The first equality on the left-hand side (LHS) of (B.3) holds by the same arguments as those for (A.6), and the second equality holds by the definition of τ.

Let $\mathbb{N}^+$ be the set of positive integers, and

$$\Lambda_{2m,j} = \{l = (l_1, \cdots, l_j) \in (\mathbb{N}^+)^j : \sum\nolimits_{k=1}^{j} l_k = 2m\}, \quad \text{for } j = 1, \ldots, 2m.$$

Then, we can write

$$\mathsf{E}_\nu\left[\left\{\sum\nolimits_{t=0}^{\infty} \mathcal{B}_t L_{h,x}(X_t)\right\}^{2m}\right] = \sum\nolimits_{j=1}^{2m}\sum\nolimits_{l\in\Lambda_{2m,j}} \frac{(2m)!}{l_1!\cdots l_j!}\mathsf{E}_\nu[\tilde{L}_{j,l}(x)], \tag{B.4}$$

where

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

To find the bound of $\mathsf{E}_\nu[\tilde{L}_{j,l}(x)]$, we first consider the case of $j \ge 2$. Following calculations in Karlsen and Tjøstheim (2001, pp. 403-404) (see also its working paper version with more details, pp. 52-53, Karlsen and Tjøstheim, 1998), we have

$$\mathsf{E}_\nu[\tilde{L}_{j,l}(x)] = \nu G_{s,\nu}\,|L_{h,x}|^{l_1}\,G_{s,\nu}\,|L_{h,x}|^{l_2}\cdots G_{s,\nu}|L_{h,x}|^{l_j}, \tag{B.5}$$

where $2 \le j \le 2m$, $l \in \Lambda_{2m,j}$ and $G_{s,\nu}$ is defined in (2.7).

Observe that for $2 \le k \le j$,

$$G_{s,\nu}|L_{h,x}|^{l_k}(y) = \mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} |L_{h,x}(X_t)|^{l_k}\right]$$
$$\le h^{-l_k} \sup\nolimits_{z\in\mathbb{R}} |1+L(z)|^{l_k} \mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\right] \le M_{j,l}(k)h^{-l_k}, \qquad \text{(B.6)}$$

where $\mathcal{N}_x = \{z \in \mathbb{R} : |x-z| \le 1\}$, $1_{\mathcal{N}_x}(z)$ is an indicator function ($=1$ if $z \in \mathcal{N}_x$; $=0$ otherwise), and the first inequality holds for h small enough (since the support of L is compact), and the last inequality holds by Lemma B.2 with some positive constant $M_{j,l}(k)$ independent of x.

Meanwhile, by Assumptions 1(i) and 2(i), there exists a positive constant $M_{j,l}(1)$, independent of x, such that

$$\nu G_{s,\nu}|L_{h,x}|^{l_1} = \pi_s|L_{h,x}|^{l_1} = h^{-l_1}\int_{\mathbb{R}}\left|L\left(\frac{u-x}{h}\right)\right|^{l_1} p_s(u)\,du$$
$$= h^{-l_1+1}\int_{\mathbb{R}} |L(r)|^{l_1} p_s(rh+x)\,dr \le M_{j,l}(1)h^{-l_1+1}, \qquad \text{(B.7)}$$

where the first equality holds by (A.8) and the last equality holds with $M_{j,l}(1) = \int_{\mathbb{R}} |L(r)|^{l_1}dr \times \sup_{x\in\mathbb{R}} p_s(x) (< \infty)$.

In view of (B.6) and (B.7), we have for $l \in \Lambda_{2m,j}$ and $j \ge 2$,

$$\mathsf{E}_\nu[\tilde{L}_{j,l}(x)] \le \left(\prod\nolimits_{k=2}^{j} M_{j,l}(k)h^{-l_k}\right)\left(\nu G_{s,\nu}|L_{h,x}|^{l_1}\right)$$
$$\le \left(\prod\nolimits_{k=1}^{j} M_{j,l}(k)\right) h^{\left(-\sum_{k=1}^{j} l_k\right)+1} = \left(\prod\nolimits_{k=1}^{j} M_{j,l}(k)\right) h^{-2m+1}. \qquad \text{(B.8)}$$

For the case of $j = 1$, by (B.7), we have

$$\mathsf{E}_\nu[\tilde{L}_{j,l}(x)] \le \nu G_{s,\nu}\tilde{I}_{L_{h,x}^{2m}} = \pi_s L_{h,x}^{2m} \le M_{1,l}(1)h^{-2m+1}. \qquad \text{(B.9)}$$

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

Letting

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

by (B.4), (B.8) and (B.9), we have shown that (B.2) holds. ■

Lemma B.2. *Suppose that $\{X_t\}$ is a β-null recurrent Markov chain, and Assumptions 1(i)-(ii) and 2 are satisfied. Then, it holds that uniformly for $x \in \left[-n^{1-\beta}\mathcal{L}_*(n), n^{1-\beta}\mathcal{L}_*(n)\right]$,*

$$\mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\right] \le \tilde{M},$$

where $\mathcal{N}_x = \{z \in \mathbb{R} : |x-z| \le 1\}$ for each $x \in \mathbb{R}$ as defined in Assumption 2(ii), and $\tilde{M}$ is some positive constant which is independent of x and y.

Proof. To derive the uniform moment bound, we consider a regeneration scheme with the set $\mathcal{N}_x$ as a hitting set for each x. That is, when the process hits $\mathcal{N}_x$ regeneration may happen (but hitting $\mathcal{N}_x$ does not necessarily imply regeneration). For each x, we let $\{(X_t, \tilde{T}_{x,t})\}_{t\geq 0}$ be a split chain and $\tilde{\alpha}_x = \mathcal{N}_x \times \{1\}$, and define the following (stopping) times when regeneration occurs:

$$\tilde{\tau}_{x,0} = \min\{t \geq 0 : \tilde{T}_{x,t} = 1\}, \quad \tilde{\tau}_{x,k} = \min\{t \geq \tilde{\tau}_{x,k-1} : \tilde{T}_{x,t} = 1\} \text{ for } k \geq 1.$$

This regeneration scheme with $\mathcal{N}_x$ can be constructed in a manner similar to that in Example of 3.1 of Karlsen and Tjøstheim (2001) by finding a small function $\tilde{s}_x \in [0, 1]$ and a probability measure $\tilde{\nu}_x$ satisfying the following minorization inequality:

$$\mathsf{P} \geq \tilde{s}_x \otimes \tilde{\nu}_x, \tag{B.10}$$

for each x. To check the existence of $\tilde{s}_x$ and $\tilde{\nu}_x$, let $\rho_{x,0}(y) = (1/2) \inf_{z\in\mathcal{N}_x} p^{(1)}(z, y)$ and $a_x = \int \rho_{x,0}(y)\, dy \,(\leq 1/2)$. Then, we have

$$\mathsf{P}(z, dy) \geq a_x 1_{\mathcal{N}_x}(z) a_x^{-1} \rho_{x,0}(y)\, dy.$$

By setting $\tilde{s}_x(z) = a_x 1_{\mathcal{N}_x}(z)$ and $\tilde{\nu}_x(dy) = a_x^{-1}\rho_{x,0}(y)\, dy$, we have $0 \leq \tilde{s}_x(z) \leq 1$ and $\int \tilde{\nu}_x(dy) = 1$, which confirms the minorization inequality (B.10).

We note that for any x, the auxiliary process $\{\tilde{T}_{x,t}\}$ of this split chain is independent of another auxiliary process $\{T_t\}$ of the split chain $\{(X_t, T_t)\}$, conditionally on (any realization of) the process $\{X_t\}$, where transition dynamics of $\{(X_t, T_t, \tilde{T}_{x,t})\}$ can be described in the same way as those of $\{(X_t, T_t)\}$ in (A.1).

Note also that by the condition on the transition density in Assumption 1(i), we have

$$\inf_{x\in\mathbb{R}} a_x = \inf_{x\in\mathbb{R}} \int \rho_{x,0}(y)\, dy \geq \kappa\delta/2 = q \in (0, 1)$$

and therefore,

$$\inf_{x\in\mathbb{R}} \inf_{z\in\mathcal{N}_x} \tilde{s}_x(z) \geq \kappa\delta/2, \tag{B.11}$$

for any x and z.

Now, we derive the desired moment bound by using the regeneration scheme with the hitting set $\mathcal{N}_x$. To this end, let

$$s_{x,0} = \min\{t \geq 0 : X_t \in \mathcal{N}_x\}, \quad s_{x,k} = \min\{t > s_{x,k-1} : X_t \in \mathcal{N}_x\} \text{ for } k \geq 1.$$

At time $s_{x,0}$ or $s_{x,k}$, the process may regenerate or may not (these are simply stopping times when the process hits $\mathcal{N}_x$). By using these stopping times, we can obtain the following decomposition:

$$\mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\right] = \sum\nolimits_{i=0}^{\infty} \mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\right| \tilde{\tau}_{x,0} = s_{x,i}\right] \mathsf{E}_y\left[I\{\tilde{\tau}_{x,0} = s_{x,i}\}\right], \tag{B.12}$$

where we below derive bounds of components of the summands on the RHS.

For any $i \geq 0$, we have the following bound:

$$
\begin{aligned}
&\mathsf{E}_y\left[\sum_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\middle|\, \tilde{\tau}_{x,0} = s_{x,i}\right] \\
&\leq \mathsf{E}_y\left[\sum_{t=0}^{\infty} 1_{\mathcal{N}_x}(X_t) I\left\{t \leq \tau\right\}\middle|\, \tilde{\tau}_{x,0} = s_{x,i}\right] \\
&= \mathsf{E}_y\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} 1_{\mathcal{N}_x}(X_t)\middle|\, \tilde{\tau}_{x,0} = s_{x,i}\right] + \mathsf{E}_y\left[\sum_{t=\tilde{\tau}_{x,0}+1}^{\infty} 1_{\mathcal{N}_x}(X_t) I\left\{t \leq \tau\right\}\middle|\, \tilde{\tau}_{x,0} = s_{x,i}\right] \\
&\leq (1+i) + \mathsf{E}_{\tilde{\alpha}_x}\left[\sum_{k=0}^{\infty}\sum_{t=\tilde{\tau}_{x,k}+1}^{\tilde{\tau}_{x,k+1}} 1_{\mathcal{N}_x}(X_t) I\left(\tilde{\tau}_{x,k} < \tau\right)\right] \\
&= (1+i) + \mathsf{E}_{\tilde{\alpha}_x}\left[\sum_{k=0}^{\infty} I\left(\tilde{\tau}_{x,k} < \tau\right) \mathsf{E}_{\tilde{\alpha}_x}\left[\sum_{t=\tilde{\tau}_{x,k}+1}^{\tilde{\tau}_{x,k+1}} 1_{\mathcal{N}_x}(X_t)\middle|\, \{(X_t, \tilde{T}_{x,t}) : t \leq \tilde{\tau}_{x,k}\}\right]\right] \\
&= (1+i) + \sum_{k=0}^{\infty} \mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tilde{\tau}_{x,k} < \tau\right)\right] \times \mathsf{E}_{\tilde{\alpha}_x}\left[\sum_{t=\tilde{\tau}_{x,k}+1}^{\tilde{\tau}_{x,k+1}} 1_{\mathcal{N}_x}(X_t)\right] \\
&= (1+i) + \sum_{k=0}^{\infty} \mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tilde{\tau}_{x,k} < \tau\right)\right] \times \mathsf{E}_{\tilde{\nu}_x}\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} 1_{\mathcal{N}_x}(X_t)\right], \qquad \text{(B.13)}
\end{aligned}
$$

where the equality in the fourth line follows from the law of iterated expectations, the equality in the fifth line holds by the fact that the event $\{\tilde{\tau}_{x,k} < t\}$ is $\mathfrak{F}_{x,\tilde{\tau}_{x,k}}^{(X,\tilde{T}_x)}$-measurable (since $\tilde{\tau}_{x,k}$ is a stopping time) and the independence of $\tilde{\tau}_{x,k}$ and τ conditionally on $\mathfrak{F}_{x,t}^{(X,\tilde{T}_x)}$ ($\mathfrak{F}_{x,t}^{(X,\tilde{T}_x)} = \sigma[\{(X_s, \tilde{T}_{x,s}) : s \leq t\}]$ and $\mathfrak{F}_t^X = \sigma[\{X_s : s \leq t\}]$, filtrations generated by $\{(X_s, \tilde{T}_{x,s})\}$ and $\{X_s\}$, respectively). We can verify the following results:

$$
\mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tilde{\tau}_{x,k} < \tau\right)\right] \leq (1/2)^k (1-q)^{-1} \quad \text{for any } k \geq 0, \qquad \text{(B.14)}
$$

and

$$
\begin{aligned}
\mathsf{E}_{\tilde{\nu}_x}\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} 1_{\mathcal{N}_x}(X_t)\right] &= a_x^{-1}\mathsf{E}_{\tilde{\nu}_x}\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} a_x 1_{\mathcal{N}_x}(X_t)\right] \\
&= a_x^{-1}\mathsf{E}_{\tilde{\nu}_x}\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} \tilde{s}_x(X_t)\right] = a_x^{-1} \leq q^{-1}, \qquad \text{(B.15)}
\end{aligned}
$$

where the proof of the former result (B.14) will be provided later, and (B.15) holds since

$$
\mathsf{E}_{\tilde{\nu}_x}\left[\sum_{t=0}^{\tilde{\tau}_{x,0}} \tilde{s}_x(X_t)\right] = 1,
$$

which follows from the arguments on page 379 of Karlsen and Tjøstheim (2001). From these, we have

$$
\mathsf{E}_y\left[\sum_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\middle|\, \tilde{\tau}_{x,0} = s_{x,i}\right] \leq (1+i) + \left[(1-q)\, q\right]^{-1} \sum_{k=0}^{\infty} (1/2)^k \leq i + \tilde{c}, \qquad \text{(B.16)}
$$

for any $i \geq 0$, with some $0 < \tilde{c} < \infty$ (independent of i, x and y). We also have the following bound:

$$
\begin{aligned}
\mathsf{E}_y\left[I\left\{\tilde{\tau}_{x,0} = s_{x,i}\right\}\right] &= \mathsf{E}_y\left[\mathsf{E}_y\left[I\left\{\tilde{\tau}_{x,0} = s_{x,i}\right\}\middle|\, X_0 = y, \{X_t\}_{t \geq 0}\right]\right] = \mathsf{E}_y\left[\prod_{k=0}^{i-1}(1 - \tilde{s}_x\left(X_{s_k}\right)) \times \tilde{s}_x\left(X_{\tilde{\tau}_{x,0}}\right)\right] \\
&\leq \left[\sup_{z \in \mathcal{N}_x}\left(1 - \tilde{s}_x\left(z\right)\right)\right]^i \times \sup_{z \in \mathcal{N}_x} \tilde{s}_x\left(z\right) \leq (1-q)^i, \qquad \text{(B.17)}
\end{aligned}
$$

for any $i \geq 0$.

By (B.12), (B.16) and (B.17), we can now obtain the desired result:

$$\mathsf{E}_y\left[\sum\nolimits_{t=0}^{\tau} 1_{\mathcal{N}_x}(X_t)\right] \le \sum\nolimits_{i=0}^{\infty} (i+\tilde{c})(1-q)^i = \tilde{M} < \infty,$$

where the majorant side is independent of x and y. To complete the proof, we next give the detailed verification of (B.14).

Proof of (B.14). For any k, we have

$$\begin{aligned}
&\mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tilde{\tau}_{x,k} < \tau\right)\right] = \mathsf{E}_{\tilde{\alpha}_x}\left[\sum\nolimits_{l=1}^{\infty} I\left(\tilde{\tau}_{x,k} = l\right) I\left(\tau > l\right)\right] \\
&= \sum\nolimits_{l=1}^{\infty} \mathsf{E}_{\tilde{\alpha}_x}\left[\mathsf{E}\left[I\left(\tilde{\tau}_{x,k} = l\right) I\left(\tau > l\right) \mid \{X_t : t \le l\}\right]\right] \\
&= \sum\nolimits_{l=1}^{\infty} \mathsf{E}_{\tilde{\alpha}_x}\left[\mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tilde{\tau}_{x,k} = l\right) \mid \ \{X_t : t \le l\}\right] \times \mathsf{E}_{\tilde{\alpha}_x}\left[I\left(\tau > l\right) \mid \{X_t : t \le l\}\right]\right] \\
&\le \sum\nolimits_{k \le l < \infty} \mathsf{E}_{\tilde{\alpha}_x}\left[\mathsf{E}\left[I\left(\tilde{\tau}_{x,k} = l\right) \mid (X_0, \tilde{T}_{x,0}) \in \tilde{\alpha}_x, \ \{X_t : t \le l\}\right]\right] \\
&= \sum\nolimits_{k \le l < \infty} \mathsf{E}_{\tilde{\alpha}_x}\left[\prod\nolimits_{m=1}^{k-1} \tilde{s}_x\left(X_{\tilde{\tau}_{x,m}}\right) \times \prod\nolimits_{t \ne \tilde{\tau}_{x,1},\ldots,\tilde{\tau}_{x,k-1},l} \left[1 - \tilde{s}_x\left(X_t\right)\right] \times \tilde{s}_x\left(X_l\right)\right] \\
&= \sum\nolimits_{k \le l < \infty} \left[\sup\nolimits_{z \in \mathcal{N}_x}\left(1 - \tilde{s}_x(z)\right)\right]^{l-k-1} \times \left[\sup\nolimits_{z \in \mathcal{N}_x} \tilde{s}_x(z)\right]^k \\
&\le \sum\nolimits_{k \le l < \infty} (1-q)^{l-k+1} \times (1/2)^k \le (1/2)^k \sum\nolimits_{l=0}^{\infty} (1-q)^l = (1/2)^k / (1-q), \qquad \text{(B.18)}
\end{aligned}$$

where the third equality holds since $I\left(\tilde{\tau}_{x,k} = l\right)$ and $I\left(\tau > l\right)$ are independent conditionally on $\{X_t : t \le l\} \in \mathfrak{F}_{x,l}^X$ (note that $\{\tilde{\tau}_{x,k} = l\}$ and $\{\tau > l\}$ are $\mathfrak{F}_{x,l}$-measurable since $\tilde{\tau}_{x,k}$ and τ are stopping times, where $\mathfrak{F}_{x,t} = \sigma\{(X_s, T_s, \tilde{T}_{x,s}) : s \le t\}$). To see the validity of the equality in the fifth line, note that $\{\tilde{\tau}_{x,k} = l\}$ (conditionally on $(X_0, \tilde{T}_{x,0}) \in \tilde{\alpha}_x$ and the history $\{X_t : t \le l\}$) is the event that "$\tilde{T}_{x,t} = 1$" happens $(k-1)$ times for $1 \le t < l$ and it also happens at $t = l$, and its (conditional) probability is bounded as

$$\mathsf{E}\left[I\left(\tilde{\tau}_{x,k} = l\right) | (X_0, \tilde{T}_{x,0}) \in \tilde{\alpha}_x, \ \{X_t : t \le l\}\right] \le \left[\sup\nolimits_{z \in \mathbb{E}}\left(1 - \tilde{s}(z)\right)\right]^{l-k-1} \left[\sup\nolimits_{z \in \mathbb{E}} \tilde{s}(z)\right]^{k+1}.$$

The last two inequalities on the RHS of (B.18) hold by the construction of $\tilde{s}_x(z)$ and by the lower bound (B.11). Now, the proof is completed. ■

Lemma B.3. *Suppose that $\{X_t\}$ is a β-null recurrent Markov chain, and Assumptions 1-2 are satisfied. Then, it holds that uniformly for $x \in \left[-n^{1-\beta}\mathcal{L}_*(n), n^{1-\beta}\mathcal{L}_*(n)\right]$,*

$$\mathsf{E}\left[\left(\sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k} L_{h,x}(X_t) e_t\right)^{2m}\right] \le \bar{M} h^{-2m+1},$$

for any integer $m \ge 1$, where $L_{h,x}$ is defined in (B.1), $\bar{M}$ is some positive constant which depends on m but is independent of k, h and k.

Proof. Let $C_e = \max\{1, \sup_{t\geq 1} E\left[|e_t|\right], \sup_{t\geq 1} E[|e_t|^2], \ldots, \sup_{t\geq 1} E[|e_t|^{2m}]\}$. Then, look at

$$\mathsf{E}\Big[\Big(\sum_{t=\tau_{k-1}+1}^{\tau_k} L_{h,x}(X_t)e_t\Big)^{2m}\Big] = \mathsf{E}\Big[\sum_{\substack{0\leq l_1,l_2,\ldots,l_{\tau_k-\tau_{k-1}}\leq 2m,\\ l_1+l_2+\cdots+l_{\tau_k-\tau_{k-1}}=2m}} \frac{(2m)!}{l_1!l_2!\cdots l_{\tau_k-\tau_{k-1}}!}$$

$$\times \left|L_{h,x}(X_{\tau_{k-1}+1})e_{\tau_{k-1}+1}\right|^{l_1}\left|L_{h,x}(X_{\tau_{k-1}+2})e_{\tau_{k-1}+2}\right|^{l_2}\cdots\left|L_{h,x}(X_{\tau_k})e_{\tau_k}\right|^{l_{\tau_k-\tau_{k-1}}}\Big]$$

$$= \mathsf{E}\Big[\sum_{\substack{0\leq l_1,l_2,\ldots,l_{\tau_k-\tau_{k-1}}\leq 2m,\\ l_1+l_2+\cdots+l_{\tau_k-\tau_{k-1}}=2m}} \frac{(2m)!}{l_1!l_2!\cdots l_{\tau_k-\tau_{k-1}}!}\mathsf{E}\Big[\,|e_{\tau_{k-1}+1}|^{l_1}|e_{\tau_{k-1}+2}|^{l_2}\cdots|e_{\tau_k}|^{l_{\tau_k-\tau_{k-1}}}\Big|\,\{(X_t,T_t)\}\,\Big]$$

$$\times \left|L_{h,x}(X_{\tau_{k-1}+1})\right|^{l_1}\left|L_{h,x}(X_{\tau_{k-1}+2})\right|^{l_2}\cdots\left|L_{h,x}(X_{\tau_k})\right|^{l_{\tau_k-\tau_{k-1}}}\Big]$$

$$\leq C_e^{2m}\mathsf{E}\Big[\sum_{\substack{0\leq l_1,l_2,\ldots,l_{\tau_k-\tau_{k-1}}\leq 2m,\\ l_1+l_2+\cdots+l_{\tau_k-\tau_{k-1}}=2m}} \frac{(2m)!}{l_1!l_2!\cdots l_{\tau_k-\tau_{k-1}}!}$$

$$\times \left|L_{h,x}(X_{\tau_{k-1}+1})\right|^{l_1}\left|L_{h,x}(X_{\tau_{k-1}+2})\right|^{l_2}\cdots\left|L_{h,x}(X_{\tau_k})\right|^{l_{\tau_k-\tau_{k-1}}}\Big]$$

$$= C_e^{2m}\mathsf{E}\Big[\Big(\sum_{t=\tau_{k-1}+1}^{\tau_k}|L_{h,x}(X_t)|\Big)^{2m}\Big] \leq C_e^{2m}Mh^{-2m+1},$$

where the equalities use the multinomial theorem and the law of iterated expectations. The last inequality uses the result (B.2) of Lemma B.1, while the inequality in the fifth line uses

$$\mathsf{E}[|e_{\tau_{k-1}+1}|^{l_1}|e_{\tau_{k-1}+2}|^{l_2}\cdots|e_{\tau_k}|^{l_{\tau_k-\tau_{k-1}}}\Big|\{(X_t,T_t)\}] \leq C_e^{2m},$$

which holds since $\{e_t\}$ is independent of $\{(X_t, T_t)\}$, $\{e_t\}$ is an independent sequence, and $l_1 + l_2 + \cdots + l_{\tau_k-\tau_{k-1}} = 2m$. Now, by setting $\bar{M} = C_e^{2m}M$, we obtain the desired result. ■

Lemma B.4. *Let $\{X_t\}$ be a β-null recurrent Markov process. Then, it holds that*

$$\lim_{n\to\infty}\Pr\left\{C_1 < \frac{N(n)}{n^\beta L_s(n)} \leq C_2\right\} = 1, \tag{B.19}$$

for two positive constants $C_1 < C_2$. Furthermore,

$$\Pr\left\{n^{\beta-\epsilon} < N(n) < n^{\beta+\epsilon} \quad \textit{infinitely often}\right\} = 1, \quad \textit{for any } \epsilon > 0. \tag{B.20}$$

Proof. We here provide only the proof of (B.19). The second assertion (B.20) is given as a part of Lemma 3.4 of Karlsen and Tjøstheim (2001).

By the definition of the Mittag-Leffler distribution (c.f., Lin, 1998), for any small $\delta > 0$, there exist two positive constants $0 < C_1 < C_2 < \infty$ such that

$$\Pr\left\{C_1 < M_\beta(1) \leq C_2\right\} \geq 1 - \delta/2. \tag{B.21}$$

The convergence result (A.8) implies that for n large enough,

$$\Pr\{\tfrac{N(n)}{n^\beta L_s(n)} \leq C_2\} - \Pr\{M_\beta(1) \leq C_2\} \geq -\delta/4, \tag{B.22}$$

$$\Pr\{\tfrac{N(n)}{n^\beta L_s(n)} \leq C_1\} - \Pr\{M_\beta(1) \leq C_1\} \leq \delta/4. \tag{B.23}$$

Now, equations (B.21)-(B.23) imply for large enough n,

$$\Pr\{C_1 < \tfrac{N(n)}{n^\beta L_s(n)} \le C_2\} = \left[\Pr\{\tfrac{N(n)}{n^\beta L_s(n)} \le C_2\} - \Pr\{M_\beta(1) \le C_2\}\right]$$
$$- \left[\Pr\{\tfrac{N(n)}{n^\beta L_s(n)} \le C_1\} - \Pr\{M_\beta(1) \le C_1\}\right] + \Pr\left\{C_1 < M_\beta(1) \le C_2\right\} \ge 1-\delta,$$

which leads to the desired result (B.19). ■

Given Lemmas B.1-B.4, we are now ready to prove Theorem 3.1.

Proof of Theorem 3.1.

Let

$$\eta_n = \sqrt{(\log n)/[n^\beta L_s(n) h]} \quad \text{and} \quad \bar{\eta}_n = \eta \times \eta_n, \tag{B.24}$$

where $\eta > 0$ is a positive constant, and also let

$$\Gamma_t(x) = \frac{1}{h} L\left(\frac{X_t - x}{h}\right) e_t.$$

For the desired result, it is sufficient to show that for n and η large enough,

$$\Pr\left\{\sup_{|x| \le \mathcal{T}_n} \left|\tfrac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x)\right| > 3\bar{\eta}_n\right\} \text{ can be made arbitrarily small.} \tag{B.25}$$

To show this, we consider the following event:

$$J_n(\beta) = \left\{C_1 n^\beta L_s(n) \le N(n) \le C_2 n^\beta L_s(n)\right\},$$

where C_1 and C_2 are constants defined in Lemma B.4. By an inclusion relation, we have

$$\left\{\sup_{|x| \le \mathcal{T}_n} \left|\tfrac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x)\right| > 3\bar{\eta}_n\right\}$$
$$\subset \left\{\left(\sup_{|x| \le \mathcal{T}_n} \left|\tfrac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x)\right| > 3\bar{\eta}_n\right) \cap J_n(\beta)\right\} \cup J_n^c(\beta). \tag{B.26}$$

Therefore, by the result (B.19) in Lemma B.4, it is sufficient to show that for n and η large enough,

$$\Pr\left\{\left(\sup_{|x| \le \mathcal{T}_n} \left|\tfrac{1}{N(n)} \sum_{t=0}^n \Gamma_t(x)\right| > 3\bar{\eta}_n\right) \cap J_n(\beta)\right\} \text{ can be made arbitrarily small.} \tag{B.27}$$

To verify this result, we consider a finite covering of the set $\{x : |x| \le \mathcal{T}_n\}$, i.e., a finite number of sets $\{S_j\}_{i=1}^{Q(n)}$, such that each S_j is an open ball in $\mathbb{R}$ with center s_i and radius r_n, Q_n is the number of these sets, and $\{x : |x| \le \mathcal{T}_n\} \subset \bigcup_{j=1}^{Q(n)} S_j$, where we let

$$r_n = n^{(\beta/2)-1}\left[h^3 (\log n) L_s(n)\right]^{1/2} \quad \text{and} \quad Q_n = [\mathcal{T}_n / r_n] + 1. \tag{B.28}$$

By this definition, we can bound Q_n as

$$C_3 \mathcal{L}_*(n) \left[(\log n) L_s(n)\right]^{-1/2} h^{-3/2} n^{2-(3/2)\beta}$$
$$\le Q_n \le C_4 \mathcal{L}_*(n) \left[(\log n) L_s(n)\right]^{-1/2} h^{-3/2} n^{2-(3/2)\beta}, \tag{B.29}$$

with constant $C_3, C_4 > 0$, since $\mathcal{T}_n = n^{1-\beta}\mathcal{L}_*(n)$.

By using the covering $\{S_j\}_{i=1}^{Q(n)}$, we can obtain the following decomposition:

$$\sup_{|x|\leq \mathcal{T}_n}\left|\frac{1}{N(n)}\sum\nolimits_{t=0}^{n}\Gamma_t(x)\right| \leq \max_{1\leq j\leq Q_n}\sup_{x\in S_j}\frac{1}{N(n)}\sum\nolimits_{t=0}^{n}|\Gamma_t(x)-\Gamma_t(s_j)| + \max_{1\leq j\leq Q_n}\left|\frac{1}{N(n)}\sum\nolimits_{t=0}^{n}\Gamma_t(s_j)\right|. \tag{B.30}$$

We next derive the bounds of the first and second terms on the RHS of (B.30). As for the first term, by Assumption 2(i), we have

$$\left|L\left(\frac{X_t - x}{h}\right) - L\left(\frac{X_t - s_j}{h}\right)\right| \leq C_L\left|\frac{s_j - x}{h}\right| \leq C_L\frac{r_n}{h}, \tag{B.31}$$

for any $x \in S_i$. This, together with the definition of r_n in (B.28), implies that in the event $J_n(\beta)$,

$$\begin{aligned}\max_{1\leq j\leq Q(n)}\sup_{x\in S_j}\frac{1}{N(n)}\sum\nolimits_{t=0}^{n}|\Gamma_t(x)-\Gamma_t(s_j)| &\leq \frac{1}{N(n)}\sum\nolimits_{t=0}^{n}|e_t|\times C_L\frac{r_n}{h^2}\\ &= O_P(nr_n/N(n)\,h^2) = O_P(\sqrt{(\log n)/n^{\beta}L_s(n)h}).\end{aligned} \tag{B.32}$$

In view of (B.30) and (B.32), for the result (B.27), it suffices to show that

$$\Pr\left\{\left(\max\nolimits_{1\leq j\leq Q_n}\left|\tfrac{1}{N(n)}\sum\nolimits_{t=0}^{n}\Gamma_t(s_j)\right| > 3\bar{\eta}_n\right)\cap J_n(\beta)\right\} \text{ can be made arbitrarily small,} \tag{B.33}$$

for n and η large enough.

To verify (B.33), we consider the decomposition based on the split chain $\{(X_t, T_t)\}$ as in (2.6). Define

$$Z_k(s_j) = \begin{cases}\sum\nolimits_{t=0}^{\tau_0}|\Gamma_t(s_j)|, & k=0,\\ \sum\nolimits_{t=\tau_{k-1}+1}^{\tau_k}|\Gamma_t(s_j)|, & k\geq 1,\\ \sum\nolimits_{t=\tau_{N(n)}+1}^{n}|\Gamma_t(s_j)|, & k=(n),\end{cases} \tag{B.34}$$

where $\Gamma_t(x) = \frac{1}{h}L\left(\frac{X_t - s_j}{h}\right)e_t = L_{h,s_j}(X_t)e_t$, and $\{\tau_k\}_{k\geq 0}$ is the stopping times as defined in (2.4). Then, we have

$$\left|\frac{1}{N(n)}\sum\nolimits_{t=0}^{n}\Gamma_t(s_j)\right| \leq \frac{1}{N(n)}Z_0(s_j) + \frac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}Z_k(s_j) + \frac{1}{N(n)}Z_{(n)}(s_j). \tag{B.35}$$

We next investigate properties of the three terms on RHS of (B.35) separately.

We first consider the second term on the RHS of (B.35) and verify that for n and η large enough,

$$\Pr\left\{\left(\max\nolimits_{1\leq j\leq Q_n}\left|\tfrac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}Z_k(s_j)\right| > \bar{\eta}_n\right)\cap J_n(\beta)\right\} \text{ can be made arbitrarily small.} \tag{B.36}$$

To this end, we claim that $\{Z_k(s_j)\}_{k\geq 1}$ is a sequence of independent random variables, which follows from Assumption 1(iii) and the fact that $\{X_t\}_{t=\tau_{k-1}+1}^{\tau_k}$ and $\{X_t\}_{t=\tau_{\bar{k}-1}+1}^{\tau_{\bar{k}}}$ are independent for different k and $\bar{k}$.

Define

$$\bar{Z}_k(s_j) = Z_k(s_j)I(|Z_k(s_j)| < \kappa_n) \quad \text{and} \quad \tilde{Z}_k(s_j) = Z_k(s_j) - \bar{Z}_k(s_j), \tag{B.37}$$

where $\{\kappa_n\}$ is a deterministic sequence of real numbers defined by

$$\kappa_n = \sqrt{\eta} h^{-1+1/2p_0} \left[Q_n n^\beta L_s(n)\right]^{1/2p_0}. \tag{B.38}$$

By these definitions, we can write

$$Z_k(s_j) = \bar{Z}_k(s_j) + \tilde{Z}_k(s_j) = (\bar{Z}_k(s_j) - \mathsf{E}[\bar{Z}_k(s_j)]) + (\tilde{Z}_k(s_j) - \mathsf{E}[\tilde{Z}_k(s_j)]),$$

since $\mathsf{E}[Z_k(s_j)] = 0$, which follows from the independence between $\{e_t\}$ and $\{X_t\}$ and the mean-zero property of $\{e_t\}$. This decomposition leads to

$$\begin{aligned}
&\Pr\left\{\left(\max_{1\leq j\leq Q_n}\left|\frac{1}{N(n)}\sum_{k=1}^{N(n)} Z_k(s_j)\right| > \bar{\eta}_n\right)\cap J_n(\beta)\right\} \\
&\leq \Pr\left\{\left(\max_{1\leq j\leq Q_n}\left|\frac{1}{N(n)}\sum_{k=1}^{N(n)} (\bar{Z}_k(s_j) - \mathsf{E}[\bar{Z}_k(s_j)])\right| > \bar{\eta}_n/2\right)\cap J_n(\beta)\right\} \\
&+ \Pr\left\{\left(\max_{1\leq j\leq Q_n}\left|\frac{1}{N(n)}\sum_{k=1}^{N(n)} (\tilde{Z}_k(s_j) - \mathsf{E}[\tilde{Z}_k(s_j)])\right| > \bar{\eta}_n/2\right)\cap J_n(\beta)\right\}.
\end{aligned} \tag{B.39}$$

We use the Bernstein inequality for deriving the convergence of the first term on the RHS of (B.39). To do so, we note that $\{\bar{Z}_k(s_j) - \mathsf{E}[\bar{Z}_k(s_j)]\}$ is an independent sequence of zero-mean random variables, satisfying

$$|\bar{Z}_k(s_j) - \mathsf{E}\left[\bar{Z}_k(s_j)\right]| \leq 2\kappa_n, \quad \text{uniformly over } k \text{ and } j, \tag{B.40}$$

and

$$\mathsf{Var}\left[\bar{Z}_k(s_j)\right] = \mathsf{E}\left[\bar{Z}_k^2(s_j)\right] - \{\mathsf{E}\left[\bar{Z}_k(s_j)\right]\}^2 \leq \mathsf{E}\left[\bar{Z}_k^2(s_j)\right] \leq 2\mathsf{E}\left[Z_k^2(s_j)\right] \leq M_1 h^{-1}, \tag{B.41}$$

where M_1 is independent of k and j (the last inequality follows form Lemma B.3). From (B.41), for any $q \geq 1$, we have uniformly for $1 \leq j \leq Q_n$,

$$\sum_{k=1}^{q} \mathsf{Var}[\bar{Z}_k(s_j)] \leq M_1 q h^{-1}. \tag{B.42}$$

Recalling the definitions of $\bar{\eta}_n$ and κ_n, the upper bound of Q_n, and the condition $(1/h) < n^{\beta-\varepsilon_0}$ in Assumption 2(iii), we have

$$\begin{aligned}
h\kappa_n\bar{\eta}_n &= h \times \sqrt{\eta} h^{-1+1/2p_0}\left[Q_n n^\beta L_s(n)\right]^{1/2p_0} \times \eta\sqrt{(\log n)/[n^\beta L_s(n) h]} \\
&\leq \eta^{3/2} \times \left[C_4 \mathcal{L}_*(n) (\log n)^{-1/2} (L_s(n))^{1/2}\right]^{1/2p_0} \sqrt{(\log n)/L_s(n)} \times n^{(-2p_0\varepsilon_0 - \varepsilon_0 + 4)/4p_0}.
\end{aligned}$$

Therefore, by the property of slowly varying functions, we can easily see that

$$h\kappa_n\bar{\eta}_n = \eta^{3/2} \times o(1), \tag{B.43}$$

since $-2p_0\varepsilon_0 - \varepsilon_0 + 4 < 0 \Longleftrightarrow p_0 > (2/\varepsilon_0) - (1/2)$.

Given (B.40), (B.42) and (B.43), we use the Bernstein inequality for independent random variables (page 102, van der Vaart and Wellner, 1996). That is, for some $C_1, C_2 \in (0, \infty)$ given in Lemma B.4, we have

$$
\begin{aligned}
&\Pr\left\{\left(\max_{1\le j\le Q_n}\left|\tfrac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}\left(\bar{Z}_k(s_j) - \mathsf{E}\left[\bar{Z}_k(s_j)\right]\right)\right| > \bar{\eta}_n/2\right)\cap J_n(\beta)\right\} \\
&\le \sum\nolimits_{j=1}^{Q_n}\sum\nolimits_{q=C_1n^\beta L_s(n)}^{C_2n^\beta L_s(n)}\Pr\left\{\left|q^{-1}\sum\nolimits_{k=1}^{q}\left(\bar{Z}_k(s_j) - \mathsf{E}\left[\bar{Z}_k(s_j)\right]\right)\right| > \bar{\eta}_n/2\right\} \\
&\le \sum\nolimits_{j=1}^{Q_n}\sum\nolimits_{q=C_1n^\beta L_s(n)}^{C_2n^\beta L_s(n)} 2\exp\left\{-\frac{q^2(\bar{\eta}_n/2)^2}{2\sum_{k=1}^{q}\mathsf{Var}\left[\bar{Z}_k(s_j)\right] + (2/3)\left(2\kappa_n\right)q(\bar{\eta}_n/2)}\right\} \\
&\le \sum\nolimits_{j=1}^{Q_n}\sum\nolimits_{q=C_1n^\beta L_s(n)}^{C_2n^\beta L_s(n)} 2\exp\left\{-\frac{q(\bar{\eta}_n/2)^2h}{2M_1 + (2/3)h\kappa_n\bar{\eta}_n}\right\} \\
&\le Q_n\sum\nolimits_{q=C_1n^\beta L_s(n)}^{C_2n^\beta L_s(n)} 2\exp\left\{-\frac{q\eta^2\left(\log n\right)/[4n^\beta L_s(n)]}{2M_1 + \eta^{3/2}\times o\left(1\right)}\right\} \\
&\le Q_n \times C_2n^\beta L_s(n) \times\, 2\exp\left\{-\left(1/4\right)\eta^{1/2}\times q\left(\log n\right)/n^\beta L_s(n)\right\}\Big|_{q=C_1n^\beta L_s(n)} \\
&\le Q_n \times C_2n^\beta L_s(n) \times 2 \times n^{-(C_1/4)\eta^{1/2}} \to 0 \quad \text{as } n\to\infty,
\end{aligned}
\tag{B.44}
$$

where the convergence in the last line holds for η large enough since the growing rate of Q_n is at most of polynomial order of n.

We now investigate the second term on the RHS of (B.39), using the Markov inequality. To do so, notice that by Lemma B.3 and (B.37),

$$
\mathsf{E}[|\tilde{Z}_k(s_j)|] = \mathsf{E}\left[|Z_k(s_j)I(|Z_k(s_j)| \ge \kappa_n)|\right] \le \mathsf{E}[|Z_k(s_j)|^{2p_0}\,\kappa_n^{-2p_0+1}] \le C_{\tilde{Z}}\left(h\kappa_n\right)^{-2p_0+1},
$$

for some positive constant $C_{\tilde{Z}}$ which is independent of k and j. Therefore, for n large enough, we have

$$
\begin{aligned}
\mathsf{E}[|\tilde{Z}_k(s_j)|] \le C_{\tilde{Z}}\left(h\kappa_n\right)^{-2p_0+1} \le C_{\tilde{Z}}\eta^{-p_0+1/2}\times C_3^{-1+1/2p_0}[(\log n)/L_s\left(n\right)\left(\mathcal{L}_*(n)\right)^{2/3}]^{1/2-1/4p_0} \\
\times h^{(1/2)-1/4p_0}n^{-(4-\beta)(2p_0-1)/4p_0} \le \bar{\eta}_n/6,
\end{aligned}
\tag{B.45}
$$

where the second inequality uses the lower bound of Q_n (given in (B.29)), and the last inequality holds (for n large enough) since $\bar{\eta}_n = \eta\left(\log n\right)^{1/2}\left[n^\beta L_s(n)h\right]^{-1/2}$,

$$
(1/2) - 1/4p_0 > 0 \quad\text{and}\quad \left(4-\beta\right)\left(2p_0-1\right)/4p_0 > \beta/2,
$$

for any positive integer $p_0 \ge 1$ and $\beta \in [0,1]$. The inequality (B.45) implies that

$$
\begin{aligned}
&\Pr\{|\tilde{Z}_k(s_j) - \mathsf{E}[\tilde{Z}_k(s_j)]| > \bar{\eta}_n/2\} \le \Pr\{|\tilde{Z}_k(s_j)| > \bar{\eta}_n/3\} \\
&\le \Pr\left\{|Z_k(s_j)| \ge \kappa_n\right\} \le \mathsf{E}[|Z_k(s_j)|^{2p_0}]\kappa_n^{-2p_0} \le M_2h^{-2p_0+1}\kappa_n^{-2p_0},
\end{aligned}
$$

for some M_2 (independent of j, k and n), where the second inequality follows from the definition of $\tilde{Z}_k(s_j)$; the third follows from the Markov inequality, and the last follows from the result of

Lemma B.3. Using this inequality, we can obtain the following bound:

$$\begin{aligned}
&\Pr\left\{\left(\max_{1\le j\le Q_n}\left|\tfrac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}(\tilde{Z}_k(s_j)-\mathsf{E}[\tilde{Z}_k(s_j)])\right|>\bar{\eta}_n/2\right)\cap J_n(\beta)\right\}\\
&\le \Pr\left\{\max_{1\le j\le Q_n}\max_{1\le k\le C_2 n^\beta L_s(n)}|\tilde{Z}_k(s_j)-\mathsf{E}[\tilde{Z}_k(s_j)]|>\bar{\eta}_n/2\right\}\\
&\le \sum\nolimits_{j=1}^{Q_n}\sum\nolimits_{k=1}^{C_2 n^\beta L_s(n)}\Pr\left\{|\tilde{Z}_k(s_j)-\mathsf{E}[\tilde{Z}_k(s_j)]|>\bar{\eta}_n/2\right\}\\
&\le Q_n\times C_2 n^\beta L_s(n)\times M_2 h^{-2p_0+1}\kappa_n^{-2p_0}=C_2M_2\eta^{-p_0}, \qquad \text{(B.46)}
\end{aligned}$$

where the last equality follows from the definition of κ_n in (B.38). By letting η large enough, the majorant side can be made arbitrarily close to zero. Now, (B.39), (B.44) and (B.46) implies the desired result (B.36).

Finally, we consider the two edge terms on the RHS of (B.35): $Z_0(s_j)/N(n)$ and $Z_{(n)}(s_j)/N(n)$. For the former, observe that

$$\begin{aligned}
\mathsf{E}[|Z_0(s_j)|^{2p_0}] &\le \frac{(C_e\bar{L})^{2p_0}}{h^{2p_0}}\int \mathsf{E}_y\left[\left\{\sum\nolimits_{t=0}^{\tau}1_{\mathcal{N}_{s_j}}(X_t)\right\}^{2p_0}\right]\lambda(dy)\\
&= O\left(h^{-2p_0}\right)\times\int \mathsf{E}_y\left[\left\{\sum\nolimits_{t=0}^{\infty}\mathcal{B}_t 1_{\mathcal{N}_{s_j}}(X_t)\right\}^{2p_0}\right]\lambda(dy)\\
&= O\left(h^{-2p_0}\right)\times\sum\nolimits_{r=1}^{2p_0}\sum\nolimits_{l\in\Lambda_{2p_0,r}}\frac{(2p_0)!}{l_1!\cdots l_r!}\int \mathsf{E}_y[\tilde{I}_r(x)]\lambda(dy), \qquad \text{(B.47)}
\end{aligned}$$

where λ is the initial distribution of X_0, the inequality holds by the same arguments as in the proof of B.3 (the independence between $\{(X_t,T_t)\}$ and $\{e_t\}$ are used and C_e defined in that proof), the two equalities hold with $\mathcal{B}_t$ and $\Lambda_{2p_0,r}$ as defined in the proof of Lemma B.1, and

$$\tilde{I}_r(x)=\sum\nolimits_{t_1=0}^{\infty}\sum\nolimits_{t_2=t_1+1}^{\infty}\cdots\sum\nolimits_{t_r=t_{r-1}+1}^{\infty}\mathcal{B}_{t_1}\mathcal{B}_{t_2}\cdots\mathcal{B}_{t_r}1_{\mathcal{N}_{s_j}}(X_{t_1})\,1_{\mathcal{N}_{s_j}}(X_{t_2})\cdots 1_{\mathcal{N}_{s_j}}(X_{t_r}).$$

Given (B.47), by arguments analogous to those in the proof of Lemma B.1 (with the aid of the result of Lemma B.2), as well as by the fact that λ is a (probability) measure, we have

$$\mathsf{E}[|Z_0(s_j)|^{2p_0}]\le \bar{M}_0 h^{-2p_0},$$

for some constant $\bar{M}_0$ which depends on p_0 but is independent of h and s_j, and thus

$$\max_{1\le j\le Q(n)}\mathsf{E}[|Z_0(s_j)|^{2p_0}]\le \bar{M}_0 h^{-2p_0}, \qquad \text{(B.48)}$$

Then, by Lemma B.4 and the Markov inequality with this result (B.48), we have

$$\begin{aligned}
&\Pr\left\{\left(\tfrac{1}{N(n)}\max_{1\le j\le Q_n}|Z_0(s_j)|>\bar{\eta}_n\right)\cap J_n(\beta)\right\}\le\sum\nolimits_{j=1}^{Q_n}\Pr\left\{\tfrac{1}{C_1n^\beta L_s(n)}|Z_0(s_j)|>\bar{\eta}_n\right\}\\
&\le\sum\nolimits_{j=1}^{Q_n}\frac{\mathsf{E}[|Z_0(s_j)|^{2p_0}]}{[C_1n^\beta L_s(n)\bar{\eta}_n]^{2p_0}}=O\Big(\frac{Q_nh^{-2p_0}}{[n^\beta L_s(n)\bar{\eta}_n]^{2p_0}}\Big)=O\Big(\frac{\mathcal{L}_*(n)n^{2-(3/2)\varepsilon_0-p_0\varepsilon_0}}{\eta^{2p_0}\,[L_s(n)\,(\log n)]^{p_0+1/2}}\Big)=o(1), \qquad \text{(B.49)}
\end{aligned}$$

where the equalities use the bound of Q_n in (B.29) and the condition that $(1/h)<n^{\beta-\varepsilon_0}$, and the last convergence result holds for any $\eta>0$ since $2-(3/2)\,\varepsilon_0-p_0\varepsilon_0<0\Leftrightarrow p_0>(2/\varepsilon_0)-(3/2)$, which is implied by the stated condition $p_0>(2/\varepsilon_0)-(1/2)$.

To investigate the component $Z_{(n)}(s_j)/N(n)$, we note that

$$\sum_{t=\tau_{N(n)}+1}^{n} |\Gamma_t(s_j)| \le \sum_{t=\tau_{N(n)}+1}^{\tau_{N(n)+1}} |\Gamma_t(s_j)|, \tag{B.50}$$

since $n \le \tau_{N(n)+1}$. This, together with Lemma B.3, implies that

$$\max_{1\le j\le Q_n} \mathsf{E}[\left|Z_{(n)}(s_j)\right|^{2p_0}] \le \max_{1\le j\le Q_n} \mathsf{E}\left[\left(\sum_{t=\tau_{N(n)}+1}^{\tau_{N(n)+1}} L_{h,x}(X_t)e_t\right)^{2p_0}\right] = O(h^{-2p_0+1}).$$

In the same way as for deriving (B.49), we can obtain

$$\Pr\left\{\left(\frac{1}{N(n)}\max_{1\le j\le Q_n}|Z_{(n)}(s_j)| > \bar{\eta}_n\right)\cap J_n(\beta)\right\} = o(1). \tag{B.51}$$

In view of (B.35), (B.36), (B.49) and (B.51), we have proved the result (B.33), completing the proof of Theorem 3.1. ■

Proof of Theorem 3.2.

Let

$$\eta_n^* = 1/\sqrt{n^{\beta-\varepsilon_0}h} \quad \text{and} \quad \bar{\eta}_n^* = \eta_n^* \times \eta, \tag{B.52}$$

where η is an arbitrary positive constant, and $\Gamma_t(x)$ be defined as in the proof of Theorem 3.1. Define $J_n^*(\beta)$ as the following event:

$$J_n^*(\beta) = \{n^{\beta-\xi_1\varepsilon_0} \ll N(n) \ll n^{\beta+\xi_1\varepsilon_0}\},$$

where $\xi_1 \in (0,1)$ is a positive constant which is chosen later, and the symbol "$a_n \ll b_n$" means that $\lim_{n\to\infty}\frac{a_n}{b_n} = 0$. By (B.20) of Lemma B.4, in order to prove (3.7), it suffices to show that for any $\eta > 0$,

$$\Pr\left\{\left(\sup_{|x|\le\mathcal{T}_n}\left|\frac{1}{N(n)}\sum_{t=0}^{n}\Gamma_t(x)\right| > 3\bar{\eta}_n^*\right)\cap J_n^*(\beta), \ \ \text{infinitely often}\right\} = 0. \tag{B.53}$$

As in the proof of Theorem 3.1, we construct a finite number of subsets $\{S_j^*\}_{j=1}^{U_n}$ such that $\{x : |x| \le \mathcal{T}_n\} \subset \bigcup_{j=1}^{U_n} S_j^*$, each S_j^* is centered at s_i^* with radius r_n^*, where U_n is the number of these subsets, and

$$r_n^* = n^{\beta-\xi_1\varepsilon_0}\frac{h^2}{n} \times \eta_n^* \quad \text{and} \quad U_n = [\mathcal{T}_n/r_n^*] + 1. \tag{B.54}$$

From this definition, we have

$$C_5\mathcal{L}_*(n)\,h^{-3/2}n^{2-(3/2)\beta+\xi_1\varepsilon_0-(1/2)\varepsilon_0} \le U_n \le C_6\mathcal{L}_*(n)\,h^{-3/2}n^{2-(3/2)\beta+\xi_1\varepsilon_0-(1/2)\varepsilon_0}, \tag{B.55}$$

for some constants $0 < C_5 < C_6 < \infty$.

Similarly to the derivation in (B.30), we have

$$\sup_{|x|\le\mathcal{T}_n}\left|\frac{1}{N(n)}\sum_{t=0}^{n}\Gamma_t(x)\right| \le \max_{1\le j\le U_n}\sup_{x\in S_j^*}\frac{1}{N(n)}\sum_{t=0}^{n}\left|\Gamma_t(x)-\Gamma_t(s_j^*)\right|$$

$$+\max_{1\le j\le U_n}\left|\frac{1}{N(n)}\sum_{t=0}^{n}\Gamma_t(s_j^*)\right| = \Pi_{n,1}+\Pi_{n,2}. \tag{B.56}$$

The first component $\Pi_{n,1}$ can be bounded as

$$\Pi_{n,1} \leq \frac{1}{N(n)} \sum\nolimits_{t=0}^{n} |e_t| \times C_L \frac{r_n^*}{h^2} = \frac{n^{\beta-\xi_1\varepsilon_0}}{N(n)} \times O_{a.s.}(\eta_n^*) = o_{a.s.}(\eta_n^*) \quad \text{in } J_n^*(\beta), \tag{B.57}$$

which follows from the fact that $\sum\nolimits_{t=0}^{n} |e_t| = O_{a.s.}(n)$ (by the strong law of large numbers), the definition of r_n^* in (B.54), and (B.20) in Lemma B.4. We next consider the component $\Pi_{n,2}$. Here, we again apply the independence-decomposition and truncation techniques as in the proof of Theorem 3.1. Let Z_k be defined as in (B.34). Then, we have

$$\Pi_{n,2} \leq \max_{1\leq j\leq U_n} \frac{1}{N(n)} \left|Z_0(s_j^*)\right| + \max_{1\leq j\leq U_n} \frac{1}{N(n)} \left|\sum\nolimits_{k=1}^{N(n)} Z_k(s_j^*)\right| + \max_{1\leq j\leq U_n} \frac{1}{N(n)} \left|Z_{(n)}(s_j^*)\right|. \tag{B.58}$$

We subsequently show that for any $\eta > 0$ (recall that $\bar{\eta}_n^* = \eta_n^* \times \eta$),

$$\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \left|\tfrac{1}{N(n)} \sum\nolimits_{k=1}^{N(n)} Z_k(s_j^*)\right| > \bar{\eta}_n^*\right) \cap J_n^*(\beta)\right\} < \infty, \tag{B.59}$$

$$\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \tfrac{1}{N(n)} |Z_0(s_j^*)| > \bar{\eta}_n^*\right) \cap J_n^*(\beta)\right\} < \infty, \tag{B.60}$$

$$\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \tfrac{1}{N(n)} |Z_{(n)}(s_j^*)| > \bar{\eta}_n^*\right) \cap J_n^*(\beta)\right\} < \infty, \tag{B.61}$$

These three results, together with the Borel-Cantelli lemma, (B.56) and (B.57), imply the desired result (B.53). We complete the proof by verifying (B.59)-(B.61), respectively.

Proof of (B.59). Let

$$\hat{Z}_k(s_j^*) = Z_k(s_j^*) I\left(|Z_k(s_j^*)| < \kappa_n^*\right), \quad \underline{Z}_k(s_j^*) = Z_k(s_j^*) - \hat{Z}_k(s_j^*), \tag{B.62}$$

and $\{\kappa_n^*\}$ be a sequence of positive real numbers defined by

$$\kappa_n^* = \left[n^{(1+c)} U_n n^{\beta+\xi_1\varepsilon_0} h^{-2m_0+1}\right]^{1/2m_0}, \tag{B.63}$$

where c is a positive constant (which can be very close to zero). Given (B.62), we have

$$\begin{aligned}
&\Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \left|\tfrac{1}{N(n)} \sum\nolimits_{k=1}^{N(n)} Z_k(s_j^*)\right| > \bar{\eta}_n^*\right) \cap J_n^*(\beta)\right\} \\
&\leq \Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \left|\tfrac{1}{N(n)} \sum\nolimits_{k=1}^{N(n)} \left(\hat{Z}_k(s_j^*) - \mathsf{E}[\hat{Z}_k(s_j^*)]\right)\right| > \bar{\eta}_n^*/2\right) \cap J_n^*(\beta)\right\} \\
&+ \Pr\left\{\left(\max\nolimits_{1\leq j\leq U_n} \left|\tfrac{1}{N(n)} \sum\nolimits_{k=1}^{N(n)} \left(\underline{Z}_k(s_j^*) - \mathsf{E}[\underline{Z}_k(s_j^*)]\right)\right| > \bar{\eta}_n^*/2\right) \cap J_n^*(\beta)\right\}.
\end{aligned} \tag{B.64}$$

We first show the convergence of the first term on the RHS of (B.64) by using the Bernstein inequality. For this, note that by the previous arguments in the proof of Theorem 3.1, $\{\hat{Z}_k(s_j^*)\}$ is a sequence of i.i.d. random variables, $|\hat{Z}_k(s_j^*)|$ is bounded by $2\kappa_n^*$, and the variance bound is given by

$$\sum\nolimits_{k=1}^{q} \mathsf{Var}[\hat{Z}_k(s_j^*)] \leq M_3 q h^{-1},$$

for any $q \geq 1$ with some constant $M_3 > 0$ (independent of j and n). By the definitions of $\bar{\eta}_n^*$ and κ_n^*, the upper bound of U_n and the condition that $1/h < n^{\beta\delta-\varepsilon_0}$, we have

$$\begin{aligned}
h\kappa_n^*\bar{\eta}_n^* &= h \times \left[n^{(1+c)} U_n n^{\beta+\xi_1\varepsilon_0} h^{-2m_0+1}\right]^{1/2m_0} \times \eta\sqrt{1/(n^{\beta-\varepsilon_0} h)} \\
&\leq \eta \times \left[C_6 \mathcal{L}_*(n)\right]^{1/2m_0} n^{[3-\varepsilon_0-(1/2)\beta(1-\delta)-\beta(1-\delta)m_0+c+2\xi_1\varepsilon_0]/2m_0}.
\end{aligned}$$

Since it is supposed that $m_0 > \frac{3-\varepsilon_0}{\beta(1-\delta)} - \frac{1}{2}$, we can let $\xi_1 > 0$ and $c > 0$ small enough such that

$$m_0 > \tfrac{3-\varepsilon_0}{\beta(1-\delta)} - \frac{1}{2} + \tfrac{c+2\xi_1\varepsilon_0}{\beta(1-\delta)},$$

which is equivalent to

$$3 - \varepsilon_0 - (1/2)\,\beta\,(1-\delta) - \beta\,(1-\delta)\,m_0 + c + 2\xi_1\varepsilon_0 < 0$$

Therefore, given such a choice of (ξ_1, c), it holds that

$$h\kappa_n^*\bar{\eta}_n^* = \eta \times o\,(1)\,.$$

Then, by the Bernstein inequality,

$$\begin{aligned}
&\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\max\nolimits_{1\le j\le U_n}\left|\tfrac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}(\hat{Z}_k(s_j^*) - \mathsf{E}[\hat{Z}_k(s_j^*)])\right| > \bar{\eta}_n^*/2\right) \cap J_n^*(\beta)\right\}\\
&\le \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n}\sum\nolimits_{q=n^{\beta-\xi_1\varepsilon_0}}^{n^{\beta+\xi_1\varepsilon_0}} \Pr\left\{\left|q^{-1}\sum\nolimits_{k=1}^{q}(\hat{Z}_k(s_j^*) - \mathsf{E}[\hat{Z}_k(s_j^*)])\right| > \bar{\eta}_n^*/2\right\}\\
&\le \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n}\sum\nolimits_{q=n^{\beta-\xi_1\varepsilon_0}}^{n^{\beta+\xi_1\varepsilon_0}} 2\exp\left\{-\frac{q^2(\bar{\eta}_n^*/2)^2}{2\sum_{k=1}^{q}\mathsf{Var}[\hat{Z}_k(s_j^*)] + (2/3)\,(2\kappa_n^*)\,q(\bar{\eta}_n^*/2)}\right\}\\
&\le \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n}\sum\nolimits_{q=n^{\beta-\xi_1\varepsilon_0}}^{n^{\beta+\xi_1\varepsilon_0}} 2\exp\left\{-\frac{q(\bar{\eta}_n^*/2)^2 h}{2M_3 + (2/3)\kappa_n^*\bar{\eta}_n^* h}\right\}\\
&= \sum\nolimits_{n=1}^{\infty} U_n \sum\nolimits_{q=n^{\beta-\xi_1\varepsilon_0}}^{n^{\beta+\xi_1\varepsilon_0}} 2\exp\left\{-\frac{q\eta^2/[4n^{\beta-\varepsilon_0}]}{2M_3 + \eta\times o\,(1)}\right\}\\
&\le \sum\nolimits_{n=1}^{\infty} U_n \times n^{\beta+\xi_1\varepsilon_0} \times 2\,\exp\left\{-[\eta^2/(9M_3)]q/n^{\beta-\varepsilon_0}\right\}\Big|_{q=n^{\beta-\xi_1\varepsilon_0}}\\
&\le \sum\nolimits_{n=1}^{\infty} U_n \times n^{\beta+\xi_1\varepsilon_0} \times 2\exp\left\{-[\eta^2/(9M_3)]n^{(1-\xi_1)\varepsilon_0}\right\} \to 0,
\end{aligned} \tag{B.65}$$

where the equality in the fifth line holds for sufficiently large n, and the last convergence holds for any $\eta > 0$ since the growing rate of U_n is at most of polynomial order of n and $\xi_1 \in (0,1)$.

To check the convergence of the second term on the RHS of (B.64), we note the moment bound:

$$\mathsf{E}\left[\underline{Z}_k(s_j^*)\right] = \mathsf{E}\left[Z_k(s_j^*)I\left(|Z_k(s_j^*)| \ge \kappa_n^*\right)\right] \le C_{\underline{Z}}\,(h\kappa_n^*)^{-2m_0+1}\,,$$

where $C_{\underline{Z}}$ is some positive constant, whose existence is due to Lemma B.3. This implies

$$\begin{aligned}
\mathsf{E}\left[\underline{Z}_k(s_j^*)\right] &\le C_{\underline{Z}}\,(h\kappa_n^*)^{-2m_0+1}\\
&\le [C_5\mathcal{L}_*\,(n)]^{(-1+1/2m_0)}\,h^{-(1/2)+1/4m_0}n^{-[3-(1/2)\varepsilon_0-(1/2)\beta+c+2\xi_1\varepsilon_0](1-1/2m_0)}\\
&\le \bar{\eta}_n^*/6(=\eta h^{-1/2}n^{-(\beta-\varepsilon_0)/2}/6)
\end{aligned} \tag{B.66}$$

where the second inequality uses the lower bound of U_n given in (B.55), and the third inequality holds since

$$[3-(1/2)\,\varepsilon_0 - (1/2)\,\beta + c + 2\xi_1\varepsilon_0]\,(1-1/2m_0) > (\beta-\varepsilon_0)\,/2 \Leftrightarrow m_0 > \tfrac{1}{2} + \tfrac{\beta-\varepsilon_0}{4(3-\beta+c+2\xi_1\varepsilon_0)},$$

which is always satisfied for $m_0 \geq 1$ (since $\frac{1}{8} > \frac{\beta-\varepsilon_0}{4(3-\beta+c+2\xi_1\varepsilon_0)}$). Using (B.66) and the Markov inequality, we have

$$\begin{aligned}&\Pr\left\{|\underline{Z}_k(s_j^*) - \mathsf{E}[\underline{Z}_k(s_j^*)]| > \bar{\eta}_n^*/2\right\} \leq \Pr\left\{|\underline{Z}_k(s_j^*)| > \bar{\eta}_n^*/3\right\}\\&\leq \Pr\left\{|\underline{Z}_k(s_j^*)| \geq \kappa_n^*\right\} \leq \mathsf{E}[|\underline{Z}_k(s_j^*)|^{2m_0}](\kappa_n^*)^{-2m_0} \leq M_4 h^{-2m_0+1}(\kappa_n^*)^{-2m_0},\end{aligned}$$

uniformly over j, k, where the last inequality holds uses

$$\mathsf{E}[|Z_k(s_j^*)|^{2m_0}] \leq M_4 h^{-2m_0+1}, \tag{B.67}$$

for some positive constant M_4 (independent of j, k and n), which follows from Lemma B.3. Therefore,

$$\begin{aligned}&\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\max\nolimits_{1\leq j\leq U(n)}\left|\tfrac{1}{N(n)}\sum\nolimits_{k=1}^{N(n)}\left(\underline{Z}_k(s_j^*) - \mathsf{E}[\underline{Z}_k(s_j^*)]\right)\right| > \bar{\eta}_n^*/2\right) \cap J_n^*(\beta)\right\}\\&\leq \sum\nolimits_{n=1}^{\infty} \Pr\left\{\max\nolimits_{1\leq j\leq U(n)} \max\nolimits_{1\leq k\leq n^{\beta+\xi_1\varepsilon_0}} \left|\underline{Z}_k(s_j^*) - \mathsf{E}[\underline{Z}_k(s_j^*)]\right| > \bar{\eta}_n^*/2\right\}\\&\leq \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n}\sum\nolimits_{k=1}^{n^{\beta+\xi_1\varepsilon_0}} \Pr\left\{\left|\underline{Z}_k(s_j^*) - \mathsf{E}[\underline{Z}_k(s_j^*)]\right| > \bar{\eta}_n^*/2\right\}\\&\leq \sum\nolimits_{n=1}^{\infty} U_n \times n^{\beta+\xi_1\varepsilon_0} \times M_4 h^{-2m_0+1}(\kappa_n^*)^{-2m_0} = M_4 \sum\nolimits_{n=1}^{\infty} n^{-(1+c)} < \infty,\end{aligned} \tag{B.68}$$

for any arbitrary $\eta > 0$. By (B.64), (B.65) and (B.68), we now have obtained the desired result (B.59).

Proofs of (B.60) and (B.61). We derive the convergence properties of the two edge terms on the RHS of (B.58). By the same arguments as those for deriving (B.48), we have

$$\max\nolimits_{1\leq j\leq U(n)} \mathsf{E}[|Z_0(s_j^*)|^{2m_0}] \leq \tilde{M}_0 h^{-2m_0},$$

for some positive constant $\tilde{M}_0$. Given this, similarly to the derivation of (B.49), we have

$$\begin{aligned}&\sum\nolimits_{n=1}^{\infty} \Pr\left\{\left(\tfrac{1}{N(n)}\max\nolimits_{1\leq j\leq U_n}|Z_0(s_j^*)| > \bar{\eta}_n^*\right) \cap J_n^*(\beta)\right\}\\&\leq \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n} \Pr\left\{\tfrac{1}{n^{\beta-\xi_1\varepsilon_0}}|Z_0(s_j^*)| > \bar{\eta}_n^*\right\} \leq \sum\nolimits_{n=1}^{\infty}\sum\nolimits_{j=1}^{U_n} \frac{\mathsf{E}[|Z_0(s_j^*)|^{2m_0}]}{(\bar{\eta}_n^* n^{\beta-\xi_1\varepsilon_0})^{2m_0}}\\&\leq \sum\nolimits_{n=1}^{\infty} \frac{U_n \tilde{M}_0 h^{-2m_0}}{(\bar{\eta}_n^* n^{\beta-\xi_1\varepsilon_0})^{2m_0}} \leq \frac{C_6}{\eta^{2m_0}} \sum\nolimits_{n=1}^{\infty} \mathcal{L}_*(n)\, n^{(1/2)(1+3\beta\delta)-2\varepsilon_0+\xi_1\varepsilon_0-m_0[\beta(1-\delta)+2\varepsilon_0(1-\xi_1)]} < \infty,\end{aligned} \tag{B.69}$$

where the inequalities use the bound of U_n in (B.55) and the condition that $(1/h) < n^{\beta\delta-\varepsilon_0}$, and the last inequality can be verified by the stated condition $m_0 > \frac{3-\varepsilon_0}{\beta(1-\delta)} - \dfrac{1}{2} = \frac{3-\varepsilon_0-\beta(1-\delta)/2}{\beta(1-\delta)}$ and $2\varepsilon_0(1-\xi_1) > 0$ since

$$\begin{aligned}&(1/2)(1+3\beta\delta) - 2\varepsilon_0 + \xi_1\varepsilon_0 - m_0[\beta(1-\delta) + 2\varepsilon_0(1-\xi_1)] < -1\\&\Leftrightarrow m_0 > \frac{3-\varepsilon_0-\beta(1-\delta)/2}{\beta(1-\delta)+2\varepsilon_0(1-\xi_1)} + \underbrace{\frac{-(3/2)+\beta[(1/2)+\delta]-\varepsilon_0(1-\xi_1)}{\beta(1-\delta)+2\varepsilon_0(1-\xi_1)}}_{<0}.\end{aligned}$$

Now, we have obtained the desired result (B.60).

We can verify (B.61) by arguments similar to those for deriving (B.51) and (B.69), and thus omit details for brevity. The proof of Theorem 3.2 is now completed. ■

Proof of Theorem 4.1.

Similarly to the decomposition of (B.35) in the proof of Theorem 3.1, we have

$$\hat{p}_n(x) = \frac{1}{N(n)}\sum_{k=1}^{N(n)} V_k(x) + \frac{1}{N(n)}V_0(x) + \frac{1}{N(n)}V_{(n)}(x), \tag{B.70}$$

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 $\{V_k(x), k \geq 1\}$ is a sequence of i.i.d. random functions of x. Then, we have

$$\begin{aligned}\hat{p}_n(x) - p_s(x) = {} & \frac{1}{N(n)}\sum_{k=1}^{N(n)} (V_k(x) - \mathsf{E}[V_1(x)]) + (\mathsf{E}[V_1(x)] - p_s(x)) \\ & + \frac{1}{N(n)}V_0(x) + \frac{1}{N(n)}V_{(n)}(x).\end{aligned} \tag{B.71}$$

We next consider the four terms on the RHS of (B.71) separately.

Applying Lemma B.1 to $V_k(x)$, and following the same argument as in the proof of Theorem 3.1, given $1/h < n^{\beta-\varepsilon_0}$, we can prove

$$\sup_{|x|\leq \mathcal{T}_n}\left|\frac{1}{N(n)}\sum_{k=1}^{N(n)} (V_k(x) - \mathsf{E}[V_1(x)])\right| = O_P(\sqrt{(\log n)/n^{\beta}L_s(n)h}). \tag{B.72}$$

Similarly, following the same argument as in the proof of Theorem 3.2, we can also obtain

$$\sup_{|x|\leq \mathcal{T}_n}\left|\frac{1}{N(n)}\sum_{k=1}^{N(n)} (V_k(x) - \mathsf{E}[V_1(x)])\right| = o(1/\sqrt{n^{\beta-\varepsilon_0}h})\ a.s., \tag{B.73}$$

under the stated condition $1/h < n^{\beta\delta-\varepsilon_0}$.

We next consider the second term on the RHS of (B.71). By using (A.7) and (A.8) in Appendix A, and the Taylor expansion, we have

$$\begin{aligned}\mathsf{E}\left[V_k(x)\right] &= \mathsf{E}\left[\sum_{t=\tau_{k-1}+1}^{\tau_k}\frac{1}{h}K\left(\frac{X_t - x}{h}\right)\right] = \mathsf{E}_\nu\left[\sum_{t=0}^{\tau}\frac{1}{h}K\left(\frac{X_t - x}{h}\right)\right] \\ &= \int \frac{1}{h}K\left(\frac{u-x}{h}\right)\nu G_{s,\nu}(du) = \int \frac{1}{h}K\left(\frac{u-x}{h}\right)\pi_s(du) \\ &= \int K(u)\, p_s(x+hu)du = p_s(x) + p_s'(x)h\int uK(u)du \\ &\quad + p_s''(x)h^2\int u^2K(u)du + \int (hu)^3\, p_s'''(\tilde{x})K(u)du,\end{aligned}$$

where $G_{s,\nu}$ and π_s are defined in (2.7), and $\tilde{x}$ is on the line segment connecting x to $x+hu$ ($\tilde{x}$ may depend on u and h). This then implies that

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\mathsf{E}\left[V_k(x)\right] - p_s(x)\right| \\ &\leq h^2\sup_{x\in\mathbb{R}}\left|p_s''(x)\right|\int u^2K(u)du + h^3\sup_{x\in\mathbb{R}}\left|p_s'''(x)\right|\int u^3K(u)du = O\left(h^2\right).\end{aligned} \tag{B.74}$$

The convergence of the last two terms on the RHS of (B.71) can be proved by the same arguments as those for the corresponding terms in the proofs of Theorems 3.1 and 3.2. For the case where $1/h < n^{\beta-\varepsilon_0}$, following the proofs of (B.49) and (B.51), we can in the same way verify that the two last terms are $O_P(\sqrt{(\log n)/n^{\beta}L_s(n)h})$. Given $1/h < n^{\beta\delta-\varepsilon_0}$, we can verify the two terms are $o(1/\sqrt{n^{\beta-\varepsilon_0}h})$ a.s. analogously to the proofs of (B.60) and (B.61). Given these, we now have verified the desired results (4.2) and (4.3). ■

Proof of Theorem 4.2.

Here, we only prove the a.s. convergence result (4.12) with the help of Theorem 3.2. The proof of the probability-convergence result (4.11) can be done quite analogously by using Theorem 3.1, and is thus omitted for brevity.

By the definition of $\hat{m}_n(x)$, we can write

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

By Theorem 3.2 and (4.3) of Theorem 4.1, we have

$$\sup_{|x|\leq \mathcal{T}_n} \left| \frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) e_t \right| = o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s.,$$
$$\sup_{|x|\leq \mathcal{T}_n} \left| \frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) - p_s(x) \right| = O(h^2) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s.,$$

respectively. These two, together with the definition of $\delta_n = \inf_{|x|\leq \mathcal{T}_n} p_s(x)$, imply that

$$\begin{aligned}
&\sup_{|x|\leq \mathcal{T}_n} \left|\sum_{t=0}^{n} w_{n,t}(x)e_t\right| \\
&= \sup_{|x|\leq \mathcal{T}_n} \left| \frac{1}{(1/N(n)h)\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right)} \right| \times \left| (1/N(n)h) \sum_{t=0}^{n} K\left(\tfrac{X_t-x}{h}\right) e_t \right| \\
&\leq \sup_{|x|\leq \mathcal{T}_n} \frac{1}{p_s(x) + o_{a.s.}(1)} \times \sup_{|x|\leq \mathcal{T}_n} \left| (1/N(n)h) \sum_{t=0}^{n} K\left(\tfrac{X_t-x}{h}\right) e_t \right| = o(\delta_n^{-1}/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s.
\end{aligned} \tag{B.75}$$

On the other hand, by using the Taylor expansion, we have

$$\begin{aligned}
&\sum_{t=0}^{n} w_{n,t}(x)m(X_t) - m(x) = \frac{1}{\hat{p}_s(x)} \left[\frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) m(X_t) - m(x)\hat{p}_s(x) \right] \\
&= \frac{1}{\hat{p}_s(x)} \frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) [m(X_t) - m(x)] \\
&= \frac{m'(x)h}{\hat{p}_s(x)} \frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) \left(\frac{X_t - x}{h}\right) \\
&+ \frac{h^2}{2\hat{p}_s(x)} \frac{1}{N(n)h} \sum_{t=0}^{n} m''(x + \vartheta_t(X_t - x)) K\left(\frac{X_t - x}{h}\right) \left(\frac{X_t - x}{h}\right)^2 \\
&= \Xi_{n,1}(x) + \Xi_{n,2}(x),
\end{aligned} \tag{B.76}$$

for some $0 \leq \vartheta_t \leq 1$.

By using the fact that $\int uK(u)du = 0$, we can verify

$$\sup_{|x|\leq \mathcal{T}_n} \left| \frac{1}{N(n)h} \sum_{t=0}^{n} K\left(\frac{X_t - x}{h}\right) \left(\frac{X_t - x}{h}\right) \right| = O(h) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s., \tag{B.77}$$

whose proof is quite analogous to the proof of Theorem 4.1, and is omitted for brevity. Therefore, we have

$$\sup_{|x|\leq \mathcal{T}_n} |\Xi_{n,1}(x)| = \rho_{1,n} h \Big[O(h) + o(1/\sqrt{n^{\beta-\varepsilon_0}h})\Big] \ \ a.s. \tag{B.78}$$

By using analogous arguments to those in the proof of Theorem 4.1 with the fact that $\int u^3 K(u)\,du = 0$, we also have

$$\sup_{|x|\le\mathcal{T}_n}\left|\frac{1}{N(n)h}\sum\nolimits_{t=0}^{n}K\left(\frac{X_t-x}{h}\right)\left(\frac{X_t-x}{h}\right)^2 - p_s(x)\int u^2K(u)\,du\right|$$
$$= O(h^2) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \quad a.s., \tag{B.79}$$

and therefore,

$$\sup_{|x|\le\mathcal{T}_n}|\Xi_{n,2}(x)|$$
$$\le h^2\frac{\sup_{|x|\le\mathcal{T}_n}|m''(x)|}{2\hat{p}_s(x)}\times\underbrace{\sup_{|x|\le\mathcal{T}_n}\frac{1}{N(n)h}\sum\nolimits_{t=0}^{n}K\left(\frac{X_t-x}{h}\right)\left(\frac{X_t-x}{h}\right)^2}_{=O_{a.s.}(1)} = O\left(\rho_{2,n}h^2\right) \quad a.s. \tag{B.80}$$

By the definition of $\hat{m}_n(x)$ and (B.75)-(B.80), we can obtain

$$\begin{aligned}|\hat{m}_n(x) - m(x)| &\le \sup_{|x|\le\mathcal{T}_n}\left|\sum\nolimits_{t=0}^{n} w_{n,t}(x)e_t\right| + \sup_{|x|\le\mathcal{T}_n}|\Xi_{n,1}(x)| + \sup_{|x|\le\mathcal{T}_n}|\Xi_{n,2}(x)|\\ &= o(1/\delta_n\sqrt{n^{\beta-\varepsilon_0}h}) + \rho_{1,n}h\Big[O(h) + o(1/\sqrt{n^{\beta-\varepsilon_0}h})\Big] + O\left(\rho_{2,n}h^2\right)\\ &= o([\delta_n^{-1} + \rho_{1,n}h]/\sqrt{n^{\beta-\varepsilon_0}h}) + O([\rho_{1,n}+\rho_{2,n}]\,h^2) \quad a.s.,\end{aligned}$$

which is the desired result (4.12). Now, the proof is completed. ■

Proof of Theorem 4.3.

We only prove (4.15) since the proof of (4.14) is analogous. By the definition of $\tilde{m}_n(x)$, we can write

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

We next consider the two terms on the RHS of the above equation separately. To do so, recall that

$$S_{n,j}(x) = \frac{1}{N(n)h}\sum\nolimits_{t=0}^{n}K\left(\frac{X_t-x}{h}\right)\left(\frac{X_t-x}{h}\right)^j$$

for $j = 0, 1, 2$. Then, by Theorem 4.1, and (B.77) and (B.79) in the proof of Theorem 4.2, we have uniformly over $|x| \le \mathcal{T}_n$,

$$S_{n,0}(x) = p_s(x) + O(h^2) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \quad a.s., \tag{B.81}$$

$$S_{n,1}(x) = O(h) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \quad a.s., \tag{B.82}$$

$$S_{n,2}(x) = p_s(x)\mu_2 + O(h^2) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \quad a.s., \tag{B.83}$$

where $\mu_2 = \int_{-\infty}^{\infty} x^2K(x)dx$.

By arguments analogous to those in the proof of Theorem 4.1, we also have uniformly over $|x| \le \mathcal{T}_n$,

$$\frac{1}{N(n)h}\sum\nolimits_{t=0}^{n}\left|\left(\frac{X_t-x}{h}\right)^jK\left(\frac{X_t-x}{h}\right)\right| = p_s(x)\int\left|u^jK(u)\right|du + O(h) + o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \quad a.s., \tag{B.84}$$

for $j = 2, 3$.

Given these preparations, we may show that

$$\sum\nolimits_{t=0}^{n} \tilde{w}_{n,t}(x)e_t = o(\delta_n^{-1}/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s., \tag{B.85}$$

$$\sum\nolimits_{t=0}^{n} \tilde{w}_{n,t}(x)m(X_t) = O(\rho_{2,n}\delta_n^{-1}h^2) \ \ a.s., \tag{B.86}$$

whose proofs are provided below. These two convergence results imply the conclusion of the theorem (4.15) as desired.

Proof of (B.85). Recalling the definitions of $\tilde{w}_{n,t}(x)$ and $S_{n,j}(x)$, we can write

$$\sum\nolimits_{t=0}^{n} \tilde{w}_{n,t}(x)e_t = \frac{[\frac{1}{N(n)h}\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right) e_t]S_{n,2}(x)}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)} + \frac{[\frac{1}{N(n)h}\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right)\left(\frac{X_t-x}{h}\right) e_t]S_{n,1}(x)}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)}.$$

By applying Theorem 3.2 with $L(u) = K(u)\, u^j$, we have as $n \to \infty$,

$$\sup_{|x|\le \mathcal{T}_n} \left| \frac{1}{N(n)h} \sum\nolimits_{t=0}^{n} K\left(\frac{X_t - x}{h}\right)\left(\frac{X_t - x}{h}\right)^j e_t \right| = o(1/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s., \tag{B.87}$$

for $j = 1, 2$. Let $b_{n,0}(x) = S_{n,0}(x) - p_s(x)$ and $b_{n,2}(x) = S_{n,2}(x) - p_s(x)\mu_2$. Note that

$$\inf\nolimits_{|x|\le \mathcal{T}_n} \left|p_s^{-1}(x) b_{n,j}(x)\right| = o(1) \quad \text{for } j = 0, 2, \quad \inf\nolimits_{|x|\le \mathcal{T}_n} \left|p_s^{-1}(x) S_{n,1}(x)\right| = o(1),$$

by (B.81)–(B.83). Then, using (B.87), we have

$$\sup_{|x|\le \mathcal{T}_n} \left| \frac{[\frac{1}{N(n)h}\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right) e_t]S_{n,2}(x)}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)} \right| = \sup_{|x|\le \mathcal{T}_n} \left| \frac{[\frac{1}{N(n)h}\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right) e_t][p_s(x)\mu_2 + b_{n,2}(x)]}{[p_s(x) + b_{n,0}(x)]\,[p_s(x)\mu_2 + b_{n,2}(x)] + S_{n,1}^2(x)} \right|$$

$$= \sup_{|x|\le \mathcal{T}_n} \left| \frac{[p_s^{-1}(x) \frac{1}{N(n)h}\sum_{t=0}^{n} K\left(\frac{X_t-x}{h}\right) e_t][\mu_2 + p_s^{-1}(x) b_{n,2}(x)]}{\left[1 + p_s^{-1}(x) b_{n,0}(x)\right]\left[\mu_2 + p_s^{-1}(x) b_{n,2}(x)\right] + p_s^{-2}(x) S_{n,1}^2(x)} \right| = o(\delta_n^{-1}/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s., \tag{B.88}$$

and

$$\sup_{|x|\le \mathcal{T}_n} \left| \frac{[\frac{1}{N(n)h}\sum_{t=0}^{n} \left(\frac{X_t-x}{h}\right) K\left(\frac{X_t-x}{h}\right) e_t]S_{n,1}(x)}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)} \right|$$

$$= \sup_{|x|\le \mathcal{T}_n} \left| \frac{[p_s^{-1}(x) \frac{1}{N(n)h}\sum_{t=0}^{n} \left(\frac{X_t-x}{h}\right) K\left(\frac{X_t-x}{h}\right) e_t] p_s^{-1}(x) S_{n,1}(x)}{\left[1 + p_s^{-1}(x) b_{n,0}(x)\right]\left[\mu_2 + p_s^{-1}(x) b_{n,2}(x)\right] + p_s^{-2}(x) S_{n,1}^2(x)} \right| = o(\delta_n^{-1}/\sqrt{n^{\beta-\varepsilon_0}h}) \ \ a.s., \tag{B.89}$$

which lead to the desired result (B.85).

Proof of (B.86). Let

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

Then, look at

$$\sum\nolimits_{t=0}^{n} \tilde{w}_{n,t}(x) m(X_t) - m(x) = \frac{\frac{1}{N(n)h}\sum_{t=0}^{n} \tilde{K}_{x,h}(X_t)\,[m(X_t) - m(x)]}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)}.$$

The denominator of the RHS can be written as

$$\begin{aligned}&\left[S_{n,0}(x)S_{n,2}(x)-S_{n,1}^2(x)\right]^{-1}\\&=\frac{p_s^{-2}(x)}{\left[1+p_s^{-1}(x)b_{n,0}(x)\right]\left[\mu_2+p_s^{-1}(x)b_{n,2}(x)\right]+p_s^{-2}(x)S_{n,1}^2(x)}=O(\delta_n^{-2})\ \ a.s.,\end{aligned}$$

uniformly over $|x|\leq\mathcal{T}_n$, where the last equality uses (B.81)-(B.83) and $\delta_n=\inf_{|x|\leq\mathcal{T}_n}p_s(x)$. We next look at the numerator:

$$\begin{aligned}&\sup_{|x|\leq\mathcal{T}_n}\left|\frac{1}{N(n)}\sum_{t=0}^n\tilde{K}_{x,h}(X_t)\left[m(X_t)-m(x)\right]\right|\\&=\sup_{|x|\leq\mathcal{T}_n}\frac{1}{N(n)}\left|m'(x)\sum_{t=0}^n(X_t-x)\tilde{K}_{x,h}(X_t)+\frac{1}{2}\sum_{t=0}^n m''(x+\vartheta_t'(X_t-x))(X_t-x)^2\tilde{K}_{x,h}(X_t)\right|\\&\leq\sup_{|x|\leq\mathcal{T}_n}\left|m''(x)\right|\times\sup_{|x|\leq\mathcal{T}_n}\frac{1}{2N(n)}\sum_{t=0}^n(X_t-x)^2\left|\tilde{K}_{x,h}(X_t)\right|\\&=\sup_{|x|\leq\mathcal{T}_n}\left|m''(x)\right|\times\frac{1}{2N(n)}\sum_{t=0}^n(X_t-x)^2\left|\frac{1}{h}K\left(\frac{X_t-x}{h}\right)\left[S_{n,2}(x)-\left(\frac{X_t-x}{h}\right)S_{n,1}(x)\right]\right|\\&\leq\sup_{|x|\leq\mathcal{T}_n}\left|m''(x)\right|\times\frac{h^2}{2}\left\{\frac{1}{N(n)h}\sum_{t=0}^n\left|\left(\frac{X_t-x}{h}\right)^2K\left(\frac{X_t-x}{h}\right)\right||S_{n,2}(x)|\right.\\&\left.+\frac{1}{N(n)h}\sum_{t=0}^n\left|\left(\frac{X_t-x}{h}\right)^3K\left(\frac{X_t-x}{h}\right)\right||S_{n,1}(x)|\right\}\\&=\sup_{|x|\leq\mathcal{T}_n}\left|m''(x)\right|\times O\left(h^2\right)\end{aligned}$$

where the first equality holds by the Taylor expansion and the continuity of $m''(\cdot)$ with $0\leq\vartheta_t'\leq 1$, the inequality in the third line uses the fact that $\sum_{t=0}^n(X_t-x)\tilde{K}_{x,h}(X_t)=0$, and the last equality uses the results (B.82)-(B.84). Therefore,

$$\begin{aligned}\sup_{|x|\leq\mathcal{T}_n}\left|\sum_{t=0}^n\tilde{w}_{n,t}(x)m(X_t)-m(x)\right|&=O(\delta_n^{-2})\times\sup_{|x|\leq\mathcal{T}_n}\left|m''(x)\right|\times O\left(h^2\right)\\&=O(\rho_{2,n}\delta_n^{-1}h^2)\ \ a.s.,\end{aligned}$$

obtaining the result (B.86). Now, the proof of (4.15) is completed. ■

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications

2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions

2013-14: Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments

2013-15: Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation

2013-16: Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression

2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model

2013-18: Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models

2013-19: Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange

2013-20: Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models

2013-21: Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement

2013-22: Johannes Tang Kristensen: Diffusion Indexes with Sparse Loadings

2013-23: Asger Lunde and Anne Floor Brix: Estimating Stochastic Volatility Models using Prediction-based Estimating Functions

2013-24: Nima Nonejad: A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory

2013-25: Nima Nonejad: Time-Consistency Problem and the Behavior of US Inflation from 1970 to 2008

2013-26: Nima Nonejad: Long Memory and Structural Breaks in Realized Volatility: An Irreversible Markov Switching Approach

2013-27: Nima Nonejad: Particle Markov Chain Monte Carlo Techniques of Unobserved Compdonent Time Series Models Using Ox

2013-28: Ulrich Hounyo, Sílvia Goncalves and Nour Meddahi: Bootstrapping pre-averaged realized volatility under market microstructure noise

2013-29: Jiti Gao, Shin Kanaya, Degui Li and Dag Tjøstheim: Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series