Generalized Kernel Smoothing for Non-negative Stationary Ergodic Processes

Yogendra P. CHAUBEY
Department of Mathematics and Statistics
Concodria University, Montreal, Canada H4B 1R6
E-mail:chaubey@alcor.concordia.ca

Naâmane LAIB*
L.S.T.A., University Paris 6, France
E-mail: naamane.laib@upmc.fr

Arusharka SEN
Department of Mathematics and Statistics
Concodria University, Montreal, Canada H4B 1R6
E-mail:asen@mathstat.concordia.ca

September 4, 2009

*Corresponding author

Abstract

In this paper, we consider a generalized kernel smoothing estimator of the regression function with non-negative support, using gamma probability densities as kernels, which are nonnegative and have naturally varying shapes. It is based on a generalization of Hille's lemma and a perturbation idea that allows us to deal with the problem at the boundary. Its uniform consistency and asymptotic normality are obtained at interior and boundary points, under a stationary ergodic process assumption, without using traditional mixing conditions. The asymptotic mean squared error of the estimator is derived and the optimal value of smoothing parameter is also discussed. Graphical illustrations of the proposed estimator are provided for simulated as well as for real data. A simulation study is also carried out to compare our method with the competing local linear method.

Key words and phrases: ergodic processes; Hille's lemma; gamma density; martingale difference; mixing; normality; prediction; regression function.

AMS 1991 subject classifications: Primary 62G05, 62M20, secondary 60J15.

1   Introduction

Various nonparametric estimators of regression function m have been proposed in the literature for linear and non-linear processes (see *e.g.* Laïb[23] and Tran[29] and references therein). Note, however, that when the support of the regression is restricted to a subset of the whole real line, most of these methods may not provide admissible values of the regression or its functionals at the boundaries. Even though, the usual kernel method may be used to estimate m for the restricted support of regression, the boundary bias (see Silverman [28]) inherent in the kernel estimator of the density gets imbedded in the regression estimator also (see Gasser and Müller [15] and Müller[24]). It can perform very well for densities that are not far from Gaussian in shape (see, e.g., Wand, Marron and Ruppert[31]), but it may not be able to consistently estimate discontinuity at the boundary, for example in the case of regressions on $[0, \infty)$ with $m(0) > 0$.

For identically and independently distributed (*iid*) observations, several methods have been developed in the past to cope with the boundary error. See for instance Zhang *et al.*[34], the reflection method of Hall and Wehrly[20] and, in fixed-design regression setting, the generalized jackknifing technique of Rice[25] [see also Härdle [21], pages 130-132].

2

In addition, there are a number of approaches to density estimation f exclusively for non-negative data, for instance, the transformation method (e.g., Wand, Marron and Ruppert[31]); the Bagai and Prakasa Rao[2] method which, unfortunately, uses only the first r order-statistics to estimate $f(x)$ if x lies between the r-th and $(r+1)$-st order-statistics; the Chaubey and Sen [4] method based on Hille's smoothing lemma; the Gamma-kernel estimator of Chen [6] and the inverse-Gaussian kernel estimator of Scaillet[26]; the Chaubey *et al.*[5] method based on a generalization of Hille's smoothing lemma, coupled with a perturbation idea to take care of the boundary bias.

Note also that most of the above papers deal with density or regression estimators in the setting of independent random variables. However, a great deal of data in econometrics, engineering and natural sciences, among other areas, occur in the form of time series in which observations are highly dependent. Also, there may be a high mass at zero, such as in the case of income data of a country with high unemployment or financial transaction data for frequently traded stocks.

In this paper we adapt the method proposed in Chaubey *et al.*[5] for density estimation for non-negative data. We further assume that the data are sampled from a stationary, ergodic process to incorporate the dependence structure in the data. This avoids the widely used strong mixing condition and its variants as a dependence measure, that is is generally not easy to calculate as it involves a complicated manipulation of taking the supremum over two sigma algebras. Moreover, the mixing properties (strong or not strong) of a number of well known processes (e.g. AR(1) and GARCH(1, 1)) are not fully known.

Additionally, many well-known processes are not strong mixing. Chernick [7] and Andrews[1] have given examples in which the first order linear autoregressive process with discrete valued random innovation is not strong mixing. In particular, if $(\epsilon_i)_{i\in\mathbb{Z}}$ is a sequence of independent Bernoulli random variables with parameter q, then the process $X_i = \rho X_{i-1} + \epsilon_i$, where $\rho \in (0, 1/2]$, is not strong mixing for all n (see Andrews [1]). In the same spirit, Guégan and Ladoucette [16] show that some long memory processes with Gaussian innovation are ergodic without being strong mixing. Another example is given in Bosq ([3], pp. 57-58), where the chaotic process of type $X_i = T(X_{i-1})$, with T being a measurable real function, is shown to be ergodic but not strongly mixing.

In Section 2, we first derive a raw estimator $m_n(x)$ without perturbation. It is then shown that this estimator can be *inconsistent* at $x = 0$ except in special cases. Following the idea in Chaubey *et al.*[5], we therefore consider the perturbed version $\tilde{m}_n(x)$. It appears that perturbation is a very useful new idea to deal with boundary bias in the case of nonnegative data, which also avoids the complication of some of the rigorous boundary correction methods mentioned above.

Section 3 is devoted to the study of asymptotic properties of the proposed estimator. We

establish the uniform almost sure convergence of the estimator $\tilde{m}_n$ when the observations are assumed to be stationary and ergodic, so that the results hold for both mixing and non mixing processes. However, the asymptotic normality is established under a weaker dependence condition. In comparison to strong mixing this dependence condition appears sufficiently mild. Also, in this section the asymptotic mean squared error is derived and the optimal choice of smoothing parameter is discussed.

Section 4 deals with the generalization of our results to higher dimensional case. In Section 5, we present some graphical illustrations of the proposed estimator on simulated as well as on real data, the latter pertaining to hardwood sapling height-growth in a boreal forest. A simulation study is also carried out to compare our estimator to the local linear estimator that demonstrates that our estimator is a good competitor to the local linear method. The proofs are deferred to the Appendix. In this context, the martingale techniques play a vital role that allow us to obtain optimal results as in the *iid* case.

2   A smooth estimator of the regression function

Let $Z_i = (X_i, Y_i)_{i \in \mathbb{N}^*}$ be a $\mathbb{R}^+ \times \mathbb{R}^+$-valued strictly stationary ergodic process defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let f be the common density function of the sample $X_1, \ldots, X_n$, which is assumed to be bounded and continuous on $[0, \infty)$. Let ϕ be a `Borel` function of $\mathbb{R}^+$ into $\mathbb{R}$ such that $\mathbb{E}(|\phi(Y_1)|) < \infty$. For $x \in \mathbb{R}^+$, let $m(x) := \mathbb{E}(\phi(Y_1)|X_1 = x)$ be the conditional mean function of $\phi(Y_1)$ given $X_1 = x$, which is assumed to be bounded.

The problem of interest is to construct a smooth estimator of the regression function m. This estimator is based on the following generalization of the Hille's lemma (see Feller (1965), Eq. (5.1), p.227).

Lemma 2.1 [Feller (1965), Lemma 1, §VII.1)] *Let h be any bounded and continuous function and Q_{x,v_n}, $n = 1, 2, \ldots$, be a family of density functions with mean $\mu_n(x)$ and variance $\sigma_{v_n}^2(x)$. Then if $\mu_n(x) \to x$ and $\sigma_{v_n}(x) \to 0$ we have*

$$\int_{-\infty}^{\infty} h(t) Q_{x,v_n}(t) dt \to h(x) \quad as \quad n \to \infty. \tag{2.1}$$

The convergence is uniform in every subinterval in which $\sigma_{v_n}(x) \to 0$ and h is uniformly continuous.

4

Taking now in (2.1), $h(t) = m(t)f(t)$, observe that the left hand side of (2.1) can be written as $\mathbb{E}_f\left(\phi(Y_1)Q_{x,v_n}(X_1)\right)$, where the expectation is taken with respect to f.

This motivates the introduction of the following smooth generalized kernel type estimator

$$m_n(x) \;=\; \frac{\sum_{i=1}^{n}\phi(Y_i)Q_{x,v_n}(X_i)}{\sum_{i=1}^{n}Q_{x,v_n}(X_i)} := \frac{h_n(x)}{f_n(x)}, \quad x \geq 0, \tag{2.2}$$

when the denominator is not 0; here $h_n(x)$ and $f_n(x)$ denote the estimators of $h(x)$ and $f(x)$, respectively and v_n $(0 < v_n < 1)$ is the bandwidth parameter sequence satisfying $v_n \to 0$ and $nv_n \to \infty$ as $n \to \infty$. The function Q_{x,v_n} may be generated by considering a density function $q_{v_n}(x)$ on $[0,\infty)$ with mean 1 and variance v_n^2, giving $Q_{x,v_n}(t) = \frac{1}{x}q_{v_n}\left(\frac{t}{x}\right)$ $(x \neq 0)$. This makes Q_{x,v_n} a density function with mean x and variance $(xv_n)^2$. Note that the usual kernel estimator is a special case of the representation given by (2.2), by taking $Q_{x,v_n}(t) = \frac{1}{v_n}K\left((t-x)/v_n\right)$, where K is a density function with mean zero and variance 1, in which case the density $Q_{x,v_n}(t)$ has mean x and variance v_n^2 that is appropriate for the case where the support is the whole real line.

The above estimator, however, may not be defined at $x = 0$, except in cases where $m_n(0) = \lim_{x\to 0+} m_n(x)$ exists. To see this, suppose, for instance q_{v_n} is a Gamma $(\alpha_n = \frac{1}{v_n^2},\ \beta_n = \frac{1}{\alpha})$ density function given by

$$q_{v_n}(t) = \frac{1}{\beta_n^{\alpha_n}\Gamma(\alpha_n)}t^{\alpha_n-1}e^{-t/\beta_n}, \quad t > 0.$$

Then $Q_{x,v_n}(t)$ is a Gamma $(\alpha_n = \frac{1}{v_n^2},\ \beta_{n,x} := \beta_x = v_n^2 x)$ density function on $[0,\infty)$ given by

$$Q_{x,v_n}(t) \;=\; \frac{1}{\beta_x^{\alpha_n}\,\Gamma(\alpha_n)}t^{\alpha_n-1}\,e^{-\alpha_n t/x}, \quad x > 0, \tag{2.3}$$

which has mean x and variance $(xv_n)^2 \to 0$ as $n \to \infty$.

The limit $m_n(0)$ of the gamma kernel estimator (2.2) may now be computed as follows,

$$
\begin{aligned}
m_n(0) &= \lim_{x\to 0+} \frac{\sum_{i=1}^{n}\phi(Y_{[i]})X_{(i)}^{\alpha_n-1}\,e^{-\alpha_n X_{(i)}/x}}{\sum_{i=1}^{n}X_{(i)}^{\alpha_n-1}\,e^{-\alpha_n X_{(i)}/x}} \\
&= \lim_{x\to 0+} \frac{\sum_{i=1}^{n}\phi(Y_{[i]})X_{(i)}^{\alpha_n-1}\,e^{-\alpha_n[X_{(i)}-X_{(1)}]/x}}{\sum_{i=1}^{n}X_{(i)}^{\alpha_n-1}e^{-\alpha_n[X_{(i)}-X_{(1)}]/x}} \\
&= \lim_{x\to 0+} \frac{\phi(Y_{[1]})X_{(1)}^{\alpha_n-1} + \sum_{i=2}^{n}\phi(Y_{[i]})X_{(i)}^{\alpha_n-1}\,e^{-\alpha_n[X_{(i)}-X_{(1)}]/x}}{X_{(1)}^{\alpha_n-1} + \sum_{i=2}^{n}X_{(i)}^{\alpha_n-1}\,e^{-\alpha_n[X_{(i)}-X_{(1)}]/x}} \\
&= \phi(Y_{[1]}),
\end{aligned}
$$

5

where $X_{(i)}$ stands for the order statistic of X_i and $Y_{[i]}$ the corresponding concommitant, i.e., $Y_{[i]} = Y_j$ if $X_{(i)} = X_j$. In this case $m_n(0)$ may not consistently estimate $m(0)$ as demonstrated below in the following example.

Let $(X_i, Y_i), i = 1, ..., n$ be a sequence of *iid* random variables with joint density $f(x, y) = e^{-y}$ for $0 \leq x \leq y$. Thus $f(x) = e^{-x}$, $f(y|x) = e^{-(y-x)}$, $m(x) = \int_x^\infty y f(y|x) dy = x + 1$ and $G_x(y) = \mathbb{P}(Y \leq y | X = x) = 1 - e^{-y+x}$. Since for all $t > 0$

$$\mathbb{P}(Y_{[1]} \leq t) = \int_{-\infty}^\infty G_x(t) f_{(1)}(x) I(t \geq x) dx,$$

where $f_{(1)}$ stands for the density of $X_{(1)}$ and $I(A)$ denotes the indicator function of the set A, we have

$$\begin{aligned}
\mathbb{P}\left(\sqrt{n}(Y_{[1]} - m(0)) \leq t\right) &= n \int_0^\infty G_x\left(\frac{t}{\sqrt{n}} + m(0)\right)(1 - F(x))^{n-1} f(x) dx \\
&= n \int_0^{1 + t/sqrtn}\left(1 - e^{-1 - tn^{-1/2} + x}\right) e^{-(n+1)x} dx \\
&\to 1 - e^{-1} \quad as \quad n \to \infty.
\end{aligned}$$

That is, when $\phi(y) = y$, $m_n(0)$ does not consistently estimate $m(0) = 1$. This would be the case in general, unless the conditional distribution of Y, given $X = 0$, is degenerate. To alleviate this situation we consider the following perturbed version of the above regression estimator,

$$\tilde{m}_n(x) := m_n(x + \epsilon_n) = \frac{\sum_{=1}^n \phi(Y_i) Q_{x+\epsilon_n, v_n}(X_i)}{\sum_{=1}^n Q_{x+\epsilon_n, v_n}(X_i)} := \frac{h_n(x + \epsilon_n)}{f_n(x + \epsilon_n)}, \quad x \geq 0, \qquad (2.4)$$

where ϵ_n is a positive real number that goes to 0 at an appropriate (sufficiently slow) rate as $n \to \infty$.

In this paper, we focus on the special case where $Q_{x+\epsilon_n, v_n}$ is a gamma density function with mean $x + \epsilon_n$ and variance $v_n^2(x + \epsilon_n)^2$. Namely, for $x \geq 0$,

$$Q_{x+\epsilon_n, v_n}(t) = \frac{1}{\beta_{x+\epsilon_n}^{\alpha_n} \Gamma(\alpha_n)} t^{\alpha_n - 1} e^{-\alpha_n t/(x+\epsilon_n)}, \qquad (2.5)$$

where

$$\alpha_n = 1/v_n^2, \quad \text{and} \quad \beta_{x+\epsilon_n} = v_n^2(x + \epsilon_n).$$

The modified gamma kernel estimator (2.4) is a generalization of the standard kernel estimator by replacing the fixed symmetric kernel with the general gamma kernels $Q_{x+\epsilon_n, v_n}(\cdot)$, which is nonnegative and naturally asymmetric to cope with discontinuity at $t = 0$.

6

2.1 Notations and Assumptions

In order to state our results we introduce the following notations. Let $\mathcal{F}_i$ be the σ-field generated by $((X_1, Y_1), \ldots, (X_i, Y_i))$ and $\mathcal{G}_i$ be that generated by $((X_1, Y_1), \ldots, (X_i, Y_i), X_{i+1})$. For $i \in \mathbb{N}$, let $f(\cdot | \mathcal{F}_{i-1})$ be the conditional density of X_i given $\mathcal{F}_{i-1}$.

Let $\mathcal{C}_0(\mathbb{R})$ be the space of continuous functions going to zero at infinity and $\| \cdot \|$ be the sup norm and set $J := [a, b] \subset \mathbb{R}^+$ ($0 \leq a < b$). From now on, the notation $\xrightarrow{\mathcal{D}}$, and $\xrightarrow{\mathbb{P}}$ stand for the convergence of random variables in distribution and in probability, respectively. For a random variable ξ, write $\xi \in \mathcal{L}^p$ ($p > 0$) if $\|\xi\|_p := (\mathbb{E}|\xi|^p)^{1/p} < \infty$ and define the projection $\mathcal{P}_k$ by $\mathcal{P}_k \xi := \mathbb{E}(\xi | \mathcal{F}_k) - \mathbb{E}(\xi | \mathcal{F}_{k-1})$, $k \in \mathbb{N}$.

Denote by $o_{\text{a.s.}}(u)$ a random function l such that $l(u)/u$ converges to zero almost surely as $u \to 0$. Similarly, define $O_{\text{a.s.}}(u)$ as a random function l such that $l(u)/u$ is almost surely bounded.

Our results are stated under some assumptions which are gathered here for easy reference:

(A0) $v_n \to 0$, $\epsilon_n \to 0$ and $nv_n \to \infty$ as $n \to \infty$.

(A1) For all $i \in \mathbb{N}$, $f \in \mathcal{C}_0(\mathbb{R})$ and $f(\cdot \,|\, \mathcal{F}_{i-1}) \in \mathcal{C}_0(\mathbb{R})$.

(A2) The sequence $\{n^{-1} \sum_{i=1}^{n} f^k(x \,|\, \mathcal{F}_{i-1})\}$ converges uniformly and almost surely (a.s.) to $f(x)$ as $n \to \infty$ ($k = 1, 2$)

(A3) $\inf_{x \in J} |f(x)| > 0$.

(A4) The conditional means of the quantities $\phi^k(Y_i)$ given $\mathcal{G}_{i-1}$ ($k = 1, 2$) exist and only depends on X_i, i.e., for any $i \geq 1$,

(i) $\mathbb{E}\left(\phi(Y_i)\,\middle|\, \mathcal{G}_{i-1}\right) = \mathbb{E}\left(\phi(Y_i)\,\middle|\, X_{i-1}\right) = m(X_i)$ a.s.

(ii) $\mathbb{E}\left(\phi^2(Y_i)\,\middle|\, \mathcal{G}_{i-1}\right) = \mathbb{E}\left(\phi^2(Y_i)\,\middle|\, X_{i-1}\right) := W_2(X_i)$ a.s.

(A5) There exists some $\gamma > 0$ such that

i) $\mathbb{E}\left(|\phi(Y_1)|^{\gamma+1}\right) < \infty$ and

ii) $\mathbb{E}\left(|\phi(Y_i)|^{\gamma+1}|\mathcal{G}_{i-1}\right) = \mathbb{E}\left(|\phi(Y_i)|^{\gamma+1}|X_i\right) := L(X_i)$ a.s., where the function $L(\cdot)$ is continuous uniformly bounded on $\mathbb{R}^+$.

(A6) (i) For some $\delta > 0$, $\mathbb{E}(|\phi(Y_1)|^{2+\delta}) < \infty$, and the function

$$\overline{W}_{2+\delta}(x) := \mathbb{E}\left(|\phi(Y_1)|^{2+\delta}\Big|X_1 = x\right)$$

is differentiable at $x \in J$ and bounded in a neighborhood of x.

(ii) The functions $m(\cdot)$ and $f(\cdot)$ have bounded derivatives up to order two.

(A7) The following type of mixing condition holds:

$$\sup_y \sum_{i=1}^{\infty} \|\mathcal{P}_1 f(y|\mathcal{F}_i)\|_2 < \infty. \tag{2.6}$$

2.2 Comments on the assumptions

The conditions stated in the above assumptions are fairly mild. Condition (A0) is about the constants involved in the estimator (*i.e.* for the sequence of bandwidth and perturbation constant. Conditions (A1) and (A2) are justified by the work of Györfi and Lugosi[17] where the authors have pointed out that the ergodic condition alone is not sufficient to ensure the L^1 consistency of kernel or histogram density estimates. A complementary assumption is therefore needed like the existence and the absolutely continuity lmost surely of the conditional density.

Conditions (A3) and (A6) are common in nonparametric regression estimation. Condition (A4) is satisfied, for instance, by letting $Y_i = X_{i+1}$ with $\{X_i\}$ being a Markov process. It is also satisfied when we consider the usual regression model $Y_i = m(X_i) + \epsilon_i$, where ϵ_i's are iid, and for any $i \geq 1$, ϵ_i is independent of X_i, since in this case, $\mathcal{G}_i$ is the sigma field generated by $(X_1, \epsilon_1), \ldots, (X_i, \epsilon_i), X_i$, and therefore, $\mathbb{E}(\phi(Y_i)|G_{i-1}) = m(X_i)$, $\mathbb{E}(\phi^2(Y_i)|G_{i-1}) = W_2(X_i)$ and $\mathbb{E}\left(|\phi(Y_i)|^{\gamma+1}|\mathcal{G}_{i-1}\right) = L(X_i)$ a.s. (A5) is a weaker condition than those proposed elsewhere in the literature.

The condition (2.6) in (A7) replaces the strong mixing condition and allows us to give an estimate of the convergence rate of the conditional bias term $\tilde{B}_n$ defined in (3.3). It holds for linear as well as many nonlinear processes, such as threshold autoregressive(AR) models and AR models with conditionally heteroscedastic errors (see, e.g., Wu [32] and Wu and Shao [33]). To make this statement clearer, take $\phi(Y) = Y$, $Y_i = X_{i+1}$ and consider the following examples :

Example 2.1 Nonlinear models.

a) Let $p \geq 1$ be a fixed integer and suppose that the X_i's are generated by the nonlinear AR(p) model

$$X_i = R_{\eta_i}(X_{i-1}, \ldots, X_{i-p}), \quad i \geq p, \tag{2.7}$$

where R is a measurable function, $\{X_i\}$ is a stationary process and $\{\eta_i\}$ is a sequence of *iid* random variables. For different forms of R in (2.7) one can obtain threshold autoregressive models TAR (Tong[30]), AR models with conditionally heteroscedastic errors ARCH (Engle[10]) and exponential autoregressive models EAR (Haggan and Ozaki [18]) among others. By iterating R in (2.7) one can see that X_i may be written as $X_i = F(\ldots, \eta_{i-1}, \eta_i)$ where F is a measurable function. The process $\{X_i\}$ is a causal stationary process and represents a large class of time series models. For $p = 1$, X_i admits a unique stationary distribution whenever

$$\mathbb{E}(\log L_\eta) < 0, \quad \mathbb{E}(L_\eta^\alpha) + \mathbb{E}(|x_0 - R\eta(x_0)|^\alpha) < \infty, \tag{2.8}$$

where $L_\eta = \sup_{x \neq y} \frac{|R_\eta(x) - R_\eta(y)|}{|x-y|}$, for some $\alpha > 0$ and some x_0 in $\mathbb{R}$ (see, e.g., Diaconis and Freedman[9]).

Let $f(u|X_i)$ be the conditional density of X_{i+1} at u given X_i such that

$$\sup_{u \in \mathbb{R}} |f(u|X_1)| < \infty \quad \text{and} \quad \text{for all} \quad (z, z') \in \mathbb{R}^2, \; \sup_{u \in \mathbb{R}} |f(u|z) - f(u|z')| \leq C|z - z'|^\beta,$$

where C and β are positives constants. Then we have by the analogous proof as that of Theorem 3 in Wu [32] that

$$\sup_{u \in \mathbb{R}} \|\mathcal{P}_0 f(u|X_n)\|_2 = O(r^n) \quad \text{for some} \quad r \in]0, 1[.$$

Therefore, the condition (2.6) in (A7) is satisfied.

b) Suppose that X_i's are generated following an AR(1) model:

$$X_i = \theta X_{i-1} + \sqrt{a_0 + a_1 X_{i-1}^2}\, \eta_i \tag{2.9}$$

where $a_0 \geq 0$ and $0 \leq a_1 < 1$, the sequence $\{\eta_i\}$ is iid and for any $i \geq 1$, η_i is independent of X_{i-1}. Using the statement (2.8), one can see that sufficient conditions for the existence of stationary distribution of model (2.9) are $\mathbb{E}\left(\log(|\theta| + |a_1\eta|)\right) < 0$ and $\mathbb{E}(|\eta|^\alpha) < \infty$.

Let f_η and f'_η be the density function of η_1 and its derivative, respectively. It is easily seen that the conditional density of X_i given $X_{i-1} = x$ is $f(z|x) = \frac{1}{\sqrt{a_0 + ax}} f_\eta\left(\frac{z - \theta x}{\sqrt{a_0 + ax}}\right)$.

9

Therefore, using Theorem 3 of Wu [32] one can see that the condition (2.6) in (A7) is satisfied whenever $\sup_{z \in \mathbb{R}}[|zf'_\eta(z) + f_\eta(z)] < \infty$ and $\sup_{z \in \mathbb{R}} |f(z|x)| < \infty$.

Example 2.2 Linear models.

Let $X_i = \sum_{i=0}^{\infty} a_j \eta_{i-j}$, where $\sum_{j=0}^{\infty} |a_j| < \infty$, $\mathbb{E}(\eta_0) = 0$ and $\mathbb{E}(\eta_0^2) < \infty$. The process X_i includes many useful special cases such that the causal ARMA models. By the analogous proof as that of Theorem 4 in Wu [32], we can show that the condition (2.6) holds true whenever

$$\sup_x |f_\eta(x)| < \infty$$

and

$$\sup_x |f'(x)_\eta| < \infty.$$

3 Main Results

In order to state our results, we introduce some further notations. For any $x \geq 0$, define

$$\overline{h}_n(x + \epsilon_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\phi(Y_i) Q_{x+\epsilon_n, v_n}(X_i) | \mathcal{F}_{i-1}\right] \tag{3.1}$$

$$\overline{f}_n(x + \epsilon_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[Q_{x+\epsilon_n, v_n}(X_i) | \mathcal{F}_{i-1}\right], \tag{3.2}$$

so that

$$\tilde{m}_n(x) := m_n(x + \epsilon_n) = \frac{h_n(x + \epsilon_n)}{f_n(x + \epsilon_n)}.$$

We define also the centralizing conditional parameter

$$\tilde{B}_n(x) := \frac{\left[\overline{h}_n(x + \epsilon_n) - h(x)\right] - m(x)\left[\overline{f}_n(x + \epsilon_n) - f(x)\right]}{\overline{f}_n(x + \epsilon_n)}, \tag{3.3}$$

which can be viewed as the "conditional bias" of $\tilde{m}_n$.

3.1 Uniform strong consistency

First we establish uniform strong consistency of the regression estimator $\tilde{m}_n$ as given in the following theorem. As can be seen from this theorem, the uniform strong consistency holds true even for $x = 0$.

Theorem 3.1 *In addition to conditions (A0)-(A5), assume that there exist sequences of real numbers $M_n \to \infty$ and $\nu_n \to \infty$ as $n \to \infty$ that satisfy*

$$a_n^{-2}\, b_n^{\alpha_n - 1}\, \alpha_n^{7/2} M_n\, \nu_n^{-1} \to 0 \quad as \quad n \to \infty \tag{3.4}$$

and

$$\sum_{n \geq 1} \nu_n \exp\left(-\pi \lambda^2 a_n^2 n M_n^{-2} v_n^2\right) < \infty, \tag{3.5}$$

where, $\lambda > 0$ and α_n is specified in (2.3). Then we have

$$\sup_{x \in [a,b]} |\tilde{m}_n(x) - m(x)| = o_{a.s.}(1) \quad as \quad n \to \infty.$$

Remark 3.1 *The sequence M_n may be chosen as $M_n = \left\{ n \ln n \left[\ln \ln n\right]^{1+\zeta} \right\}^{1/\gamma}$, where ζ is a positive constant and γ is as in (A5). The sequence ν_n is used to obtain the finite number of subintervals to cover the compact interval J. Condition (3.4) is satisfied if we choose, for instance, $\nu_n = \left[a_n^{-2} b_n^{\alpha_n - 1} \alpha_n^{7/2} M_n \log n \right] + 1$ (where $[t]$ stands for the integer part of t) whereas (3.5) holds true by taking $\lambda = \lambda_n = \sqrt{Cn} \frac{M_n}{a_n \sqrt{\pi n v_n}}$, where C is a large positive constant.*

Remark 3.2 *Let $\{U_i;\ i \in \mathbb{N}\}$ be a real-valued strictly stationary Markov process. Theorem 1 may be used to construct a nonparametric predictor, say $\hat{U}_{N+1}$, to evaluate the value of U_{N+1} given the past $U_1, \ldots, U_N$. For this purpose let $\mathbf{X}_i = (U_1, \ldots, U_{i+p-1})$, $Y_i = U_{i+p}$, for $i = 1, 2, \ldots, n$, and define $\hat{U}_{N+1} := \tilde{m}_n(\mathbf{X}_n)$, where $n = N - p + 1$ and p is appropriately chosen.*

3.2   Asymptotic normality

The following theorem gives the asymptotic normality of the regression estimator at interior and boundaries of the support, and gives the form of its asymptotic variance in both cases.

Theorem 3.2 . *Assume that conditions (A0)-(A4) and (A6) hold. (i) Let $f(x) > 0$ at given $x \geq 0$ and let*

$$\sigma^2(x) := \frac{1}{2\sqrt{\pi}} \frac{W_2(x) - m^2(x)}{x f(x)}. \tag{3.6}$$

Then, we have

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(x)\right).$$

(ii) Suppose in addition the condition (A7) is satisfied and that

$$nv_n^5 \to 0 \qquad \text{and} \qquad (nv_n)^{\frac{1}{2}}\epsilon_n \longrightarrow 0 \quad \text{as} \quad n \to \infty. \tag{3.7}$$

Then, we have

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(x)\right).$$

(iii) Let $x = 0$ and suppose moreover that

$$\epsilon_n v_n \to 0, \quad nv_n^5\epsilon_n \to 0, \quad nv_n\epsilon_n^3 \to 0 \quad \text{and} \quad nv_n\epsilon_n \to \infty \quad \text{as} \quad n \to \infty. \tag{3.8}$$

Then,

$$\sqrt{nv_n\epsilon_n}(\tilde{m}_n(0) - m(0)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma_0^2(0)\right),$$

where $\sigma_0^2(0) = \frac{1}{2\sqrt{\pi}}\frac{W_2(0) - m^2(0)}{f(0)}$ whenever $f(0) > 0$.

Remark 3.3 *Some discussion on how the convergence rate obtained in (ii) and (iii) depends on the choice of the bandwidth v_n and the sequence ϵ_n is in order. Indeed, we have from (i) that*

$$\tilde{m}_n(x) - m(x) + \tilde{B}_n(x) = O_{\mathbb{P}}((nv_n)^{-1/2}) \qquad \text{and} \tag{3.9}$$
$$\tilde{m}_n(0) - m(0) + \tilde{B}_n(0) = O_{\mathbb{P}}((\epsilon_n nv_n)^{-1/2}). \tag{3.10}$$

Using Lemma 6.9 with the choice, $\epsilon_n = O(v_n^2)$, yields

$$\tilde{m}_n(x) - m(x) = O_{\mathbb{P}}((nv_n)^{-1/2}) + O_{\mathbb{P}}(v_n^2).$$

To make the rates of convergence explicit, take, for instance, $v_n = O\left(n^{-\frac{s}{1+2s}}\right)$ where s is a nonnegative constant. The condition (A0) and the statements (3.7) are then satisfied provided that $s > 1/3$, whereas the statements in (3.8) hold true if $1/5 < s < 1$. Therefore the pointwise convergence rates given in (ii) and (iii) are $O_{\mathbb{P}}(n^{-\frac{1+s}{2(1+2s)}})$ and $O_{\mathbb{P}}(n^{-\frac{1-s}{2(1+2s)}})$, respectively, which fail to be optimal. However, if we choose $v_n = O\left((\frac{1}{n\log n})^{1/5}\right)$, the optimal bandwidth obtained in Section 3, and $\epsilon_n = O(v_n^2)$, then the conditions of the statements (3.7) and (3.8) hold true. And the optimal convergence rates given in (ii) and (iii) are $O_{\mathbb{P}}\left(n^{-2/5}(\log n)^{1+(3/5)}\right)$ and $O_{\mathbb{P}}\left(n^{-1/5}(\log n)^{3/10}\right)$, respectively.

Remark 3.4 *Using Theorem 2, one may define the following asymptotic* $100(1 - \alpha)\%$ *confidence band for the function* m

$$\tilde{m}_n(x) \pm c_\alpha \left(\frac{\sigma_n(x)}{nv_n} \right)^{1/2} \qquad \text{when } x > 0$$

and

$$\tilde{m}_n(0) \pm c_\alpha \left(\frac{\sigma_n(0)}{\epsilon_n nv_n} \right)^{1/2} \qquad \text{when} \quad x = 0,$$

where c_α *is the upper* $\alpha/2$ *quantile of the standard normal distribution,* $\sigma_n(\cdot)$ *and* $\sigma_n(0)$ *are appropriate estimates of* $\sigma(\cdot)$ *and* $\sigma_n(0)$, *respectively.*

3.3   Asymptotic mean squares error (AMSE) of the regression estimator $\tilde{m}_n(x)$

Here we consider only asymptotic mean squared error (AMSE) of $\tilde{m}_n(x)$ computed at one single positive point x. In a future paper we shall consider asymptotic mean integrated squared error (AMISE) as well as data-driven choice of both the smoothing parameters (ϵ_n, v_n) via an empirical, *cross-validation* function derived from AMISE.

The following proposition gives the MSE of the estimator $\tilde{m}_n(x)$ in the interior as well as in the boundary of the interval, which permits to determine the optimal (in the sense of minimizing MSE of $\tilde{m}_n(x)$) rates of convergence of $v_n \to 0, \epsilon_n \to 0$ (which is necessary for $\tilde{m}_n(x)$ to be a consistent estimator of $m(x)$) as $n \to \infty$.

Proposition 3.1 . *In addition to conditions (A1), (A4), (A7), assume that the functions* f *and* m *have bounded derivatives up to order 3. Then we have*

$$
\begin{aligned}
MSE(x) \quad \approx \quad & \frac{W_2(x) + m^2(x)}{nv_n x f(x)} + \frac{m^2(x)}{f^2(x)} \left[\epsilon_n \, m'(x) + \frac{1}{2} v_n^2 x^2 m''(x) \right]^2 \\
+ \quad & \frac{1}{f^2(x)} \left[\epsilon_n \, (m(x)f(x))' + \frac{1}{2} v_n^2 x^2 (m(x)f(x))'' \right]^2 \qquad\qquad \text{if } x \neq 0 \\
\approx \quad & \frac{W_2(0) + m^2(0)}{nv_n \epsilon_n f(0)} + \frac{\epsilon_n}{f^2(0)} \left[(m(0)f(0))' + m'(0)m^2(0) \right]^2 \quad \text{if } \ x = 0 \text{(3.11)}
\end{aligned}
$$

Remark 3.5 *When* x *is an* interior *point, i.e.,* $x > 0$, *(3.11) shows that the optimal choice of* ϵ_n *is* $\epsilon_n = 0$, *which gives the optimal choice of* v_n *to be* $v_n = O(n^{-1/5})$ *and the optimal*

13

order of MSE (x) is then the usual $O(n^{-4/5})$. In fact, the same optimal order is obtained by setting $\epsilon_n = O(v_n^2)$ in (3.11).

On the other hand when x is a boundary *point, i.e., $x = 0$ or $x = O(\epsilon)$, there is no optimal choice for $v_n > 0$, while that for ϵ_n is $\epsilon_n = O((nv_n)^{-1/3})$. In this case ϵ is the setting $\epsilon_n = O(v_n^2)$ as above leads to $v_n = O(n^{-1/7})$ and MSE $(x) = O(n^{-4/7})$ which is suboptimal. Note, however, that we can recover the optimal order of the MSE by setting $v_n = O(\epsilon_n^{-1/2})$. This leads to $\epsilon_n = O(n^{-2/5})$, so that the order of MSE is $n^{-4/5}$, but v_n becomes $O(n^{1/5})$. Hence, unlike the usual kernel-based methods, there is no universal optimal choice for the smoothing parameters, in terms of point-wise MSE.*

Remark 3.6 *As pointed out by Seifert and Gasser [27], the local linear method (Fan and Gijbles [12]) has a drawback as the variance is unbounded in finite sample, because the denominator of the local linear regression estimate has positive probability of being zero or arbitrarily small. Moreover, this method uses only symmetric compact as kernel. Chen[**?**] adapted the local linear smoothing method by replacing the classical symmetric kernels by the Gamma kernel smoother. He demonstrated that the asymmetric gamma kernel can eliminate the problem of the finite sample variance. Moreover, the local gamma kernel estimate may also remove the problem of an increased mean square error near the boundary, since its support matches that of the curve in the local linear estimator in order to increase the effective sample size and then to reduce the variance of local polynomial smoothers. This method has the usual good properties of standard local linear smoothing with a fixed symmetric kernel and offers some additional extra advantages in aspects of having finite variance and resistance to sparse design. In our context the numerical study carried out in Section 5 demonstrates that our estimator is a good competitor to the local linear estimator. Theoretical study of the local linear method combining with our estimator is the subject of future research.*

4   Generalization to the p-dimensional case $(p > 1)$

We briefly discuss a generalization of our result to the case where the independent random variable is p-dimensional. For $p \geq 1$, let $\mathbf{Z}_i = (\mathbf{X}_i, Y_i)_{i\in\mathbb{N}^*}$ be a $\mathbb{R}^{+p} \times \mathbb{R}^+$-valued strictly stationary ergodic sequence, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$. Let $\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^{+p}$ and $\epsilon_n = (\epsilon_{1n}, \ldots, \epsilon_{pn})$ such that for any $1 \leq i \leq p$, $\epsilon_{in} \to 0$. Then for any $\mathbf{t} \in \mathbb{R}^{+p}$, the density function defined in (2.3) takes the forme

$$Q_{\mathbf{x}+\epsilon_n,v}(\mathbf{t}) \quad = \quad \frac{1}{(\prod_{i=1}^{d}\beta_{x_i+\epsilon_{in}})^{\alpha}\,(\Gamma(\alpha))^{p}}\left(\prod_{i=1}^{p}t_i\right)^{\alpha-1}\exp\left(-\alpha\sum_{i=1}^{p}\frac{t_i}{x_i+\epsilon_{in}}\right),(4.1)$$

where $\alpha := \alpha_n = 1/v^2$, $\beta_{x_i+\epsilon_{in}} = v^2(x_i + \epsilon_{in})$ and $v := v_n$. Consider the following estimator for the regression function m

$$\tilde{m}_n(\mathbf{x}) \quad = \quad \frac{\sum_{i=1}^{n}\phi(Y_i)Q_{\mathbf{x}+\epsilon_n,v}(\mathbf{X}_i)}{\sum_{i=1}^{n}Q_{\mathbf{x}+\epsilon_n,v_n}(\mathbf{X}_i)}. \tag{4.2}$$

To state our results for the estimator (4.2), we introduce further notations. Let $\mathfrak{F}_i = \sigma(\mathbf{Z}_1, \ldots, \mathbf{Z}_i)$ and $\mathfrak{G}_i = \sigma(\mathbf{Z}_1, \ldots, \mathbf{Z}_i, \mathbf{X}_{i+1})$, denote the σ-algebras generated by $(\mathbf{Z}_1, \ldots, \mathbf{Z}_i)$ and $(\mathbf{Z}_1, \ldots, \mathbf{Z}_i, \mathbf{X}_{i+1})$, respectively. For $i \in \mathbb{N}$, let $f_{\mathbf{X}_i}(\cdot|\mathfrak{F}_i)$ be the conditional density of $\mathbf{X}_i$ given $\mathfrak{F}_{i-1}$ and $\mathbf{f}$ be the unconditional density of $\mathbf{X}_i$. We denote by **(A1)-(A6)**, the same conditions as (A1)-(A6) given in Section 2.1 where the quantities $\mathcal{F}_{i-1}, \mathcal{G}_{i-1}, f_{X_i}(\cdot|\mathcal{F}_i)$ and f are replaced by $\mathfrak{F}_i, \mathfrak{G}_i, f_{\mathbf{X}_i}(\cdot|\mathfrak{F}_i)$ and $\mathbf{f}$, respectively. We denote by $\tilde{B}_n(\mathbf{x})$ and $W_2(\mathbf{x})$ the functions defined in (A4)-(ii) and (3.3) where x is replaced by the vector $\mathbf{x}$.

Theorem 4.1 *In addition to conditions* **(A1)-(A4)** *and* **(A6)**(*i*)*, suppose that the functions* m *and* $\mathbf{f}$ *have bounded partial derivatives up to order* p *and*

$$v_n^p \to 0 \quad and \quad nv_n^p \to \infty \quad as \quad n \to \infty. \tag{4.3}$$

Then we have for $\mathbf{f}(\mathbf{x}) > 0$ *at given* $\mathbf{x} = (x_1, \ldots, x_p)$ *where* $x_j > 0$ $(j = 1, \ldots, p)$

i) $\sqrt{nv_n^p}(\tilde{m}_n(\mathbf{x}) - m(\mathbf{x}) - \tilde{B}_n(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(\mathbf{x})\right)$

where

$$\sigma^2(\mathbf{x}) := \frac{1}{(2\sqrt{\pi})^p}\frac{W_2(\mathbf{x}) - m^2(\mathbf{x})}{(\prod_{i=1}^{p}x_i)\mathbf{f}(\mathbf{x})}.$$

ii) Suppose in addition that (2.6) holds true and $nv_n^{5p} \to 0$ *and* $(nv_n^p)^{\frac{1}{2}}\epsilon_n^p \to 0$ *as* $n \to \infty$, *then*

$$\sqrt{nv_n^p}(\tilde{m}_n(\mathbf{x}) - m(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(\mathbf{x})\right).$$

iii) If $x = 0$ *and if*

$$\epsilon_n^p v_n^p \to 0, \ nv_n^{5p}\epsilon_n^p \to 0, \ nv_n^p\epsilon_n^{3p} \to 0, \ and \ n(v_n\epsilon_n)^p \to \infty, \ as \ n \to \infty,$$

then we have

$$\sqrt{n v_n^p \epsilon_n^p}(\tilde{m}_n(\mathbf{0}) - m(\mathbf{0}))) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma_0^2(\mathbf{0})\right)$$

where

$$\sigma_0^2(\mathbf{0}) := \frac{1}{(2\sqrt{\pi})^p} \frac{W_2(\mathbf{0}) - m^2(\mathbf{0})}{\mathbf{f}(\mathbf{0})},$$

whenever $\mathbf{f}(\mathbf{0}) > 0$.

Remark 4.1 *To make explicit the rates of convergence, chosen as in the discussion following Theorem 2, $v_n = O\left(\left(\frac{1}{n \log n}\right)^{1/(2+3p)}\right)$ and $\epsilon_n^p = O(v_n^{2p})$. Therefore, all conditions in i), ii) and iii) are satisfied, and the optimal point-wise convergence rates in ii) and iii) are*

$$O_{\mathbb{P}}\left(\frac{\log^{\frac{p}{2(1+p)}} n}{n}\right)^{\frac{1+p}{2+3p}} \quad and \quad O_{\mathbb{P}}\left(\frac{\log^{\frac{3p}{2}} n}{n}\right)^{\frac{1}{2+3p}}, \quad respectively.$$

Note that the rates of convergence are obviously affected by the dimension p, and this result will be practically useful only when $p \leq 3$ since for $p > 3$ the curse of dimensionality will kick in.

5 Illustrations and simulation

We illustrate our method with a simulated data-set from an autoregressive model as well as a real data-set on hardwood sapling height-growth. For comparison, we also include in our plots the *local linear estimator* (Fan[11], Fan and Gijbels[13]) which is well-known for its boundary correction and robustness properties.

Recall that the local linear estimator is defined as the solution $\hat{m}$ for β_0 that minimizes the sum $\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1(X_i - x))^2 W_i(x)$, where $W_i(x) = K\left(\frac{X_i - x}{w}\right)$ and $w = w_n$ is the bandwidth parameter. The solution of this problem is given by

$$\hat{m}(x) = \frac{s_2(x)t_0(x) - s_1(x)t_1(x)}{s_2(x)s_0(x) - s_1^2(x)},$$

where $s_k = n^{-1}\sum_{i=1}^{n}(X_i - x)^k W_i(x)$ and $t_k = n^{-1}\sum_{i=1}^{n} Y_i(X_i - x)^k W_i(x)$, $k = 0, 1, 2$.

The smoothing parameters for both the estimators were chosen using the following, leave-one-out *cross-validation* (CV) procedure:

$$\arg\min n^{-1}\sum_{i=1}^{n}(Y_i - M_i(X_i))^2$$

where M_i stands for either $\tilde{m}_n$ or $\hat{m}$, evaluated at X_i but based on the X_i-deleted ('leave-one-out') sample $\{X_1, \ldots, X_{i-1}, X_i, \ldots, X_n\}$, $1 \leq i \leq n$. The minimization is over (ϵ, v) in case of $\tilde{m}_n(x)$ and over h in case of $\hat{m}(x)$.

Autoregressive data. Here $X_1, \ldots, X_n$ are generated as:

$$X_i = 0.5X_{i-1} + \left(\sqrt{0.2 + 0.1X_{i-1}^2} \right) \varepsilon_i, \ X_1 \text{ Exponential (1)},$$

where ε_i, $1 \leq i \leq n$, are i.i.d Weibull $(1, 3)$, i.e., with density $g(\varepsilon) = 3\varepsilon^2 \exp(-\varepsilon^3)$, $\varepsilon \geq 0$. We consider two models for $Y_1, \ldots, Y_n$:
a) $Y_i = 0.5(6 - 4X_i + X_i^2)\varepsilon_i$;
b) $Y_i = \sin(1/X_i)\varepsilon_i$, $1 \leq i \leq n$.
Here ε_i, $1 \leq i \leq n$, are taken to be iid Weibull $(1, 2)$, i.e., with density $g(\varepsilon) = 2\varepsilon \exp(-\varepsilon^2)$, $\varepsilon \geq 0$. The illustration/comparison is provided in Figure 1 and Figure 2 for Model-a and Model-b, respectively. The sample-size was chosen to be $n = 200$. The optimal smoothing parameter values are given in the respective captions.

As is clear from the figures, both $\tilde{m}_n$ and $\hat{m}$ are close to each other. However, $\tilde{m}_n$ appears to be slightly less affected by noisy observations and more adaptive to the shape of the true regression function.

Hardwood sapling data. We apply our method to data on initial height (X) versus 5-year height-growth (Y) of naturally-occurring hardwood saplings in gap areas of the boreal forest around Lake Duparquet in north-western Quebec. Both the initial height (as of 1998) and the height-growth (over 1998–2003) were obtained from multi-temporal LIDAR (LIght Detection And Ranging) surveys. (Data courtesy: Prof. Benoit St-Onge and Ms. Uday-alakshmi Vepakomma, University of Quebec at Montreal.) All measurements are in meters, and the sample consists of $n = 94$ saplings.

The scatter-plot and the two estimators are presented in Figure 3. The smoothing parameters were again chosen based on the CV criterion above. The local linear estimator in this case appears to be smoother and more monotonic than ours.
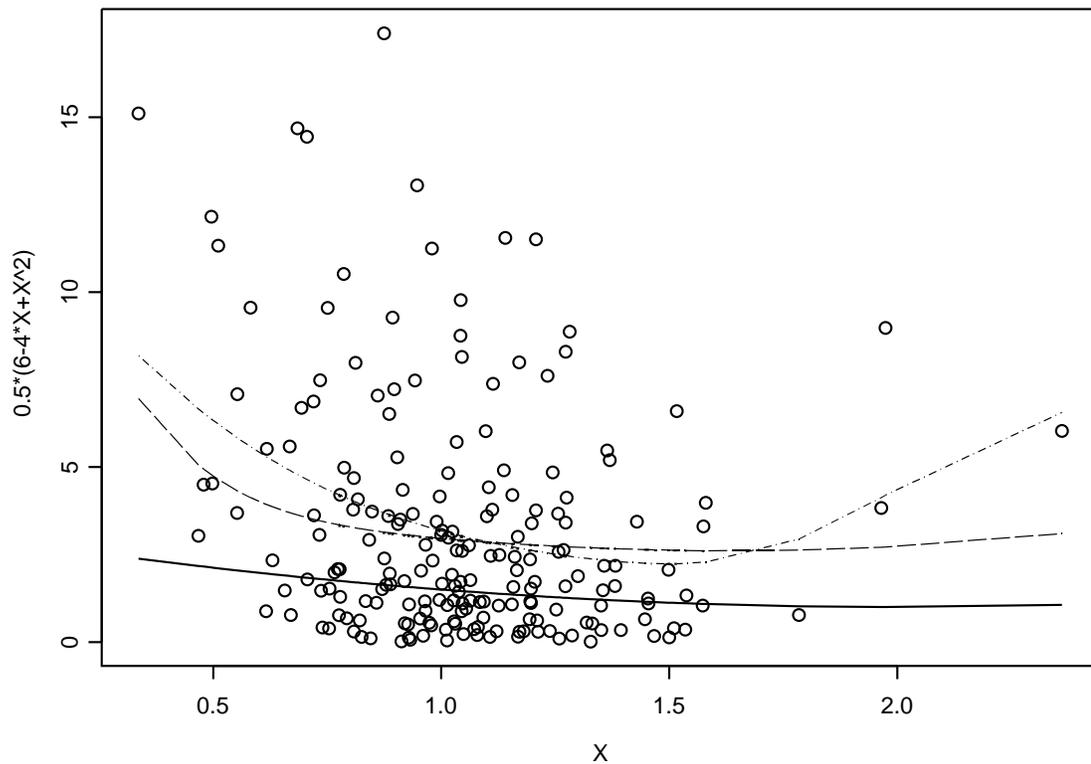
Figure 1: Scatterplot and regression estimators for autoregressive data with $Y = 0.5(6 - 4X + X^2)\varepsilon$, true regression (—), $\tilde{m}_n$ with $\epsilon_n = 0, v_n = 0.24$ (– –), local linear estimator with $w_n = 0.35$ (- · -)
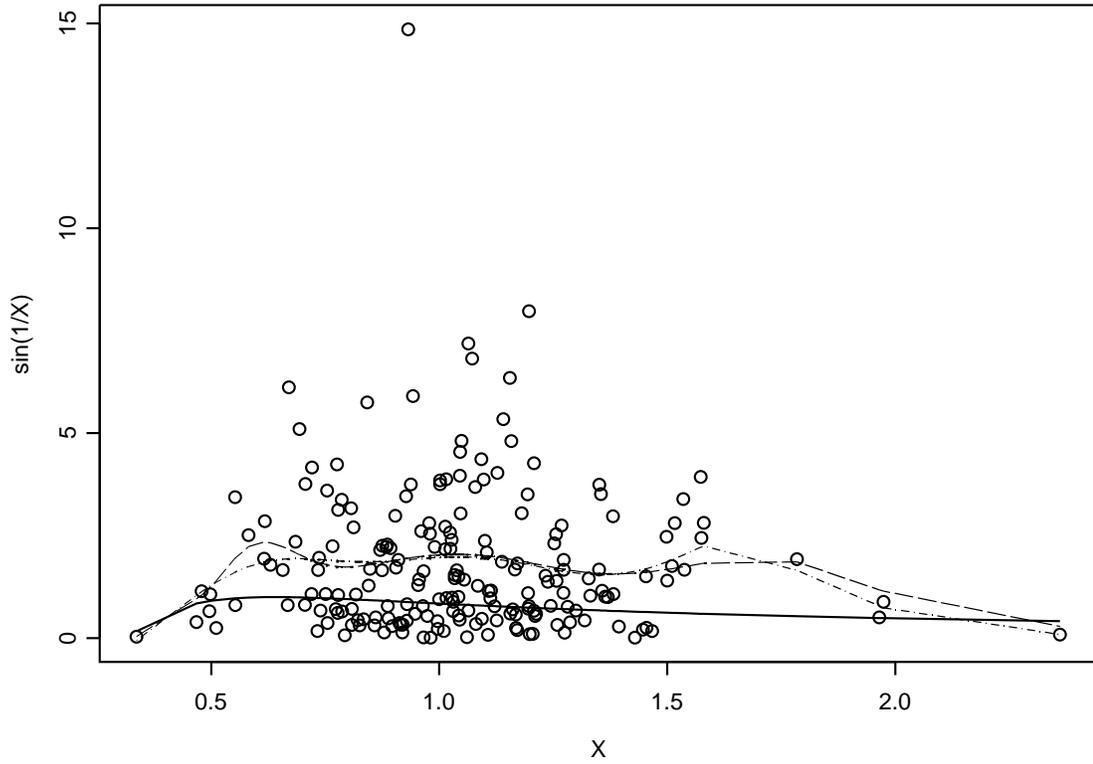
Figure 2: scatterplot and regression estimators for autoregressive data with $Y = \sin(1/X)\varepsilon$: true regression (—), $\tilde{m}_n$ with $\epsilon_n = 0.02, v_n = 0.95$ (– –), local linear estimator with $w_n = 0.14$ (-··-)
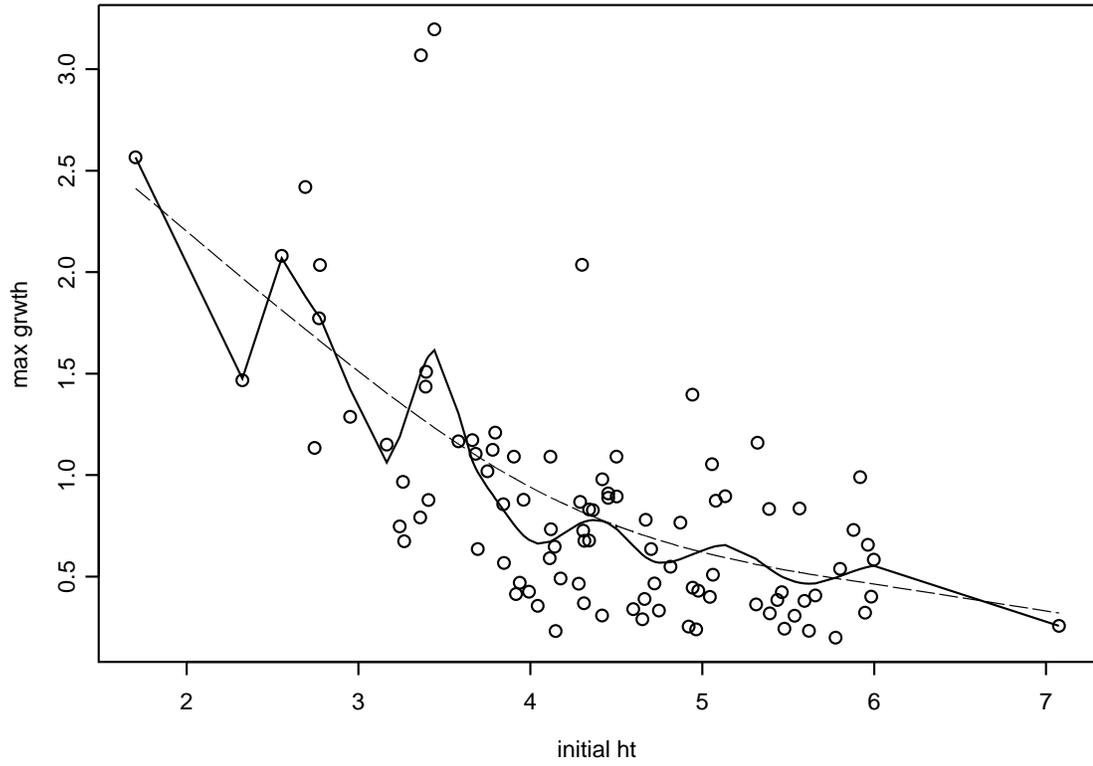
Figure 3: scatterplot and regression estimators for height-growth data: (a) $\tilde{m}_n$ with $\epsilon_n = 0$, $v_n = 0.331$ (—), local linear estimator with $w_n = 0.95$ (– –).

Comparison to local linear method: a simulation study

We now present a simulation study to compare our method to the local linear one. The comparison is done for four choices of distributions of (X, ε), and two different regression functions $m(\cdot)$ taken from Fan and Gijbels[13]. We use an additive model, $Y = m(X) + \varepsilon$, and the same smoothing parameter selection method as above. For each sample-size 1000 samples were generated, and the tables below list the average values of the minimum $n^{-1} \sum_{i=1}^{n} (Y_i - M_i(X_i))^2$ (see above) for each estimator.

Table 1: Values[a] of minimum $n^{-1} \sum_{i=1}^{n} (Y_i - M_i(X_i))^2$ for the gamma kernel and local linear estimators for the model, $m(x) = x + 2 \exp(-16x^2)$.

Sample Size(n)	$X \sim \exp(1)$ $\varepsilon \sim \mathcal{N}(0,\ 0.7^2)$	$X \sim \exp(1)$ $\varepsilon \sim$ doubleExp	$X \sim$ Weibull $\varepsilon \sim \mathcal{N}(0,\ 0.7^2)$	$X \sim$ Weibull $\varepsilon \sim$ doubleExp
100	0.03584061 0.03913404	0.1958467 0.2006739	0.02308521 0.02178451	0.1190155 0.1110448
200	0.02258581 0.02616064	0.1168053 0.1304714	0.01371159 0.01323952	0.07219315 0.07009928
500	0.01231901 0.01350157	0.05675914 0.05971275	0.007551803 0.00613061	0.03658063 0.03428299

[a]First number in the column is corresponding to the gamma kernel and the second one is for the local linear method.

Table 2: Values[a] of minimum $n^{-1} \sum_{i=1}^{n} (Y_i - M_i(X_i))^{2*}$ for the gamma kernel and local linear estimators for the model, $m(x) = \sin(2x) + 2\exp(-16x^2)$.

Sample Size(n)	$X \sim \exp(1)$ $\varepsilon \sim \mathcal{N}(0,\ 0.7^2)$	$X \sim \exp(1)$ $\varepsilon \sim$ doubleExp	$X \sim$ Weibull $\varepsilon \sim \mathcal{N}(0,\ 0.7^2)$	$X \sim$ Weibull $\varepsilon \sim$ doubleExp
100	0.01544821 0.01415603	0.19257 0.1838041	0.008892057 0.007312179	0.1377363 0.1037686
200	0.009834903 0.009768867	0.1283950 0.1291265	0.005304467 0.00489389	0.08062876 0.06583907
500	006080469 0.005155077	0.06577962 0.05971275	0.002839715 0.0020642	0.04228712 0.03178110

[a]First number in the column is corresponding to the gamma kernel and the second one is for the local linear method.

The results show that the two estimators have very similar performances, each being better than the other in some cases. Our estimator is thus a good competitor to even an established procedure such as the local linear method.

6   Appendix: Proofs

The detailed proofs are provided here by means of a series of lemmas and propositions. However, first we introduce some more notations that will be used in the proofs to follow. For $x \in J := [a, b]$, let $x^+ = x + \epsilon_n$, $a_n = a + \epsilon_n$ and $b_n = b + \epsilon_n$. We denote by $\mathbb{R}_*^+ := \mathbb{R}^+ - \{0\}$.

Clearly we have form (2.4), (3.1) and (3.3) that

$$\tilde{m}_n(x) - m(x) - \tilde{B}_n(x) \quad = \quad \frac{1}{f_n(x^+)}[(h_n(x^+) - \overline{h}_n(x^+))$$
$$-(m(x) + \tilde{B}_n(x))(f_n(x^+) - \overline{f}_n(x^+))]. \qquad (6.1)$$

The major thrust of the decomposition (6.1) is due to the fact that both the terms $h_n(x^+) - \overline{h}_n(x^+)$ and $f_n(x^+) - \overline{f}_n(x^+)$ in the the summands form a martingale difference. The following result gives the almost sure uniform convergence of the bias term $\tilde{B}_n(x)$.

Proposition 6.1 *Assume that (A0), (A1), (A2), (A3) and (A4)(i) hold true. Then we have*

$$\sup_{x\in[a,b]} \left|\tilde{B}_n(x)\right| = o_{a.s.}(1) \quad as \quad n \to \infty.$$

Proof of Proposition 6.1. It suffices to show that $\overline{h}_n(x^+) - h(x)$ converges uniformly in x to 0 a.s. and $\overline{f}_n(x^+)$ is uniformly bounded over. Making use of (A4)(i) and the law of iterated conditional expectation we may write

$$\mathbb{E}\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \quad = \quad \mathbb{E}\left(\mathbb{E}\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{G}_{i-1}\right]|\mathcal{F}_{i-1}\right)$$
$$= \quad \mathbb{E}\left[Q_{x+\epsilon_n,v_n}(X_i)m(X_i)|\mathcal{F}_{i-1}\right].$$

Thus,

$$|\overline{h}_n(x^+) - h(x)| \quad \leq \quad \left|\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)m(t)f(t|\mathcal{F}_{i-1})dt - h(x)\right|$$
$$\leq \quad \left\|\frac{1}{n}\sum_{i=1}^n f(\cdot|\mathcal{F}_{i-1}) - f\right\|\left|\int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)m(t)dt\right|$$
$$+ \left|\int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)h(t)dt - h(x)\right|. \qquad (6.2)$$

Using Lemma 2.1 and (A0), the second integral goes to 0 uniformly in x, provided that h is bounded. The first term is bounded above by

$$\left\|\frac{1}{n}\sum_{i=1}^n f(\cdot|\mathcal{F}_{i-1}) - f\right\| \sup_{x\in\mathbb{R}^+} |m(x)|,$$

that also tends to to 0 a.s. as $n \to \infty$ in view of (A2) and the fact that $m(\cdot)$ is bounded. By the same arguments we can conclude by Lemma 2.1, (A0) and (A2) that $\overline{f}_n(x^+)$ converges uniformly in x to $f(x)$ a.s., which is bounded over J uniformly in x in view of (A3). □

The following proposition gives an a.s. asymptotic lower bound for $\inf_{x \in J} |f_n(x^+)|$.

Proposition 6.2 . *Assuming (A0), (A2) and (A3) hold true, then we have*

$$
(i) \quad \sup_{x \in J} |f_n(x^+) - f(x)| = o_{a.s.}(1) \quad as \quad n \to \infty
$$

$$
(ii) \quad \inf_{x \in J} |f_n(x^+)| > 0 \quad a.s. \quad as \quad n \to \infty.
$$

Proof of Proposition 6.2. To prove (i) above, first note that we have

$$
|f_n(x^+) - f(x)| \le |f_n(x^+) - \overline{f}_n(x^+)| + |\overline{f}_n(x^+) - f(x)|.
$$

Next we argue as in the proof of Proposition 6.1, and observe that the second term in the right hand side of the above inequality tends to 0 a.s. as $n \to \infty$. Additionally, we use the same arguments as used to prove Proposition 6.3 below, and conclude that the first term also converges to 0 a.s., uniformly in x, as $n \to \infty$.

Finally, we have for any $x \in J$,

$$
\inf_{x \in J} |f_n(x^+)| \ge \inf_{x \in J} f(x) - \sup_{x \in J} |f_n(x^+) - f(x)|.
$$

And (ii) follows from (i) and condition (A3). □

The main task now is to establish the uniform almost sure convergence for $\overline{h}_n(x^+) - h(x)$. Making use of the Stirling's formula we can easily see that, for any fixed x, the function $t \mapsto Q_{x+\epsilon_n, v_n}(t)$ is bounded above by $\frac{1}{\sqrt{2\pi}(x+\epsilon_n)v_n}$ for every $t \ge 0$ whenever $v_n \to 0$. By contrast, the function $\phi(y)$ is not necessarily bounded, it can thus be handled by a suitable truncation. For this purpose, let $M_n = \left\{ n \ln n \left[\ln \ln n\right]^{1+\zeta} \right\}^{1/\gamma}$, where ζ is a positive constant and γ is as in (A5).

Let us now define the processes

$$
h_n^b(x^+) \quad := \quad \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) I\{|\phi(Y_i)| \le M_n\} Q_{x+\epsilon_n, v_n}(X_i) \tag{6.3}
$$

and

$$\overline{h}_n^b(x^+) \;\; := \;\; \frac{1}{n}\sum_{i=1}^n E\left[\phi(Y_i)I\{|\phi(Y_i)| \le M_n\}Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right], \tag{6.4}$$

where $I(A)$ stands for the indicator function of the set A. Then, we have

$$
\begin{aligned}
h_n(x^+) - \overline{h}_n(x^+) \;\; = \;\; & (h_n(x^+) - h_n^b(x^+)) + (h_n^b(x^+) - \overline{h}_n^b(x^+)) \\
& + (\overline{h}_n^b(x^+) - \overline{h}_n(x^+)).
\end{aligned} \tag{6.5}
$$

The asymptotic behavior of the first and the last terms on the right hand side of (6.5) is given in the following two Lemmas and that for the middle term is given in Proposition 4.

Lemma 6.1 *Assuming (A5)(i) holds, then, for each ω outside a null set D, there exists a positive integer $n_0(\omega)$ such that*

$$h_n(x^+) = h_n^b(x^+) \quad for\ n \ge n_0(\omega) \quad and\ all\ x \in \mathbb{R}^+.$$

Proof of Lemma 6.1. We have

$$h_n(x^+) - h_n^b(x^+) = n^{-1}\sum_{i=1}^n I(|\phi(Y_i)| > M_n)\phi(Y_i)Q_{x+\epsilon_n}(X_i).$$

Let $p > 1$ and $q > 1$ be real numbers such that $p^{-1} + q^{-1} = 1$, then from stationarity of the data, Cauchy-Schwarz and Markov inequalities, condition (A5)(i) and the fact that $Q_{x+\epsilon_n}(\cdot)$ is bounded, it follows that $\forall \lambda > 0$,

$$
\begin{aligned}
\mathbb{P}\left(|h_n(x^+) - h_n^b(x^+)| > \lambda\right) \;\; \le \;\; & (\mathbb{P}(|\phi(Y_1)| > M_n))^{1/p}\left(\mathbb{E}(|\phi(Y_1)Q_{x+\epsilon_n}(X_1)|^b)\right)^{1/q} \\
\le \;\; & CM_n^{-q/a}\mathbb{E}(|\phi(Y_1)|^a) = O(M^{-\gamma}),
\end{aligned}
$$

by taking $p = (\gamma+1)/\gamma$ and $q = \gamma + 1$ $(\gamma > 0)$. Now the use of Borel Contelli Lemma and the summability of $\{M_n^{-\gamma}\}$ imply that for each ω outside of a set D such that $\mathbb{P}(D^c) = 1$, there exits $n_0(\omega)$ such that for any $n \ge n_0(\omega)$, $h_n(x^+) = h_n^b(x^+)$ a.s. $\qquad\square$

We deal now with the asymptotic behavior of the third term in (6.5).

Lemma 6.2 *Assume (A2) and (A5) hold, then we have*

$$\sup_{x\in\mathbb{R}^+} |\overline{h}_n^b(x^+) - \overline{h}_n(x^+)| = O_{a.s.}\left(M_n^{-\gamma}\right) \quad as \quad n \to \infty. \tag{6.6}$$

Proof of Lemma 6.2. Observe that

$$\overline{h_n^b}(x^+) - \overline{h}_n(x^+) = n^{-1} \sum_{i=1}^n \mathbb{E}\left[Q_{x+\epsilon_n,v_n}(X_i)\phi(Y_i)I(|\phi(Y_i)| > M_n) \,|\mathcal{F}_{i-1}\right]. \qquad (6.7)$$

We have by the properties of conditional expectation that

$$\mathbb{E}\left[Q_{x+\epsilon_n,v_n}(X_i)\phi(Y_i)I(|\phi(Y_i)| > M_n) \,\Big|\mathcal{F}_{i-1}\right]$$
$$= \mathbb{E}\left[Q_{x+\epsilon_n,v_n}(X_i)\mathbb{E}\left(\phi(Y_i)I(|\phi(Y_i)| > M_n) \,\Big|\mathcal{G}_{i-1}\right) \,\Big|\mathcal{F}_{i-1}\right]. \qquad (6.8)$$

Next, using the Cauchy-Schwarz inequality twice along with Markov inequality, we obtain

$$\mathbb{E}\left(\phi(Y_i)I(|\phi(Y_i)| > M_n) \,\Big|\mathcal{G}_{i-1}\right)$$
$$\leq \left(\mathbb{E}\left(|\phi(Y_i)|^{\gamma+1} \,\Big|\mathcal{G}_{i-1}\right)\right)^{\frac{1}{\gamma+1}} \left(\mathbb{P}(|\phi(Y_i)| > M_n) \,\Big|\mathcal{G}_{i-1}\right)^{\frac{\gamma}{\gamma+1}}$$
$$\leq M_n^{-\gamma} \times \mathbb{E}\left(|\phi(Y_i)|^{\gamma+1} \,\Big|\mathcal{G}_{i-1}\right) = M_n^{-\gamma}L(X_i) \quad a.s.,$$

in view of condition (A5)(ii). Therefore, using condition (A2) in (6.7, we have

$$
\begin{aligned}
|\overline{h_n^b}(x^+) - \overline{h}_n(x^+)| &\leq M_n^{-\gamma} \int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)L(t)\left(n^{-1}\sum_{i=1}^n f(t|\mathcal{F}_{i-1})\right) dt \\
&\simeq M_n^{-\gamma} \int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)L(t)f(t)dt \quad \text{as } n \to \infty \\
&= O(M_n^{-\gamma})
\end{aligned}
$$

uniformly in x by Lemma 2.1 provided that $f(x)$ and $L(x)$ are uniformly continuous and bounded. This proves Lemma 6.2. □

following proposition gives the asymptotic behavior of the middle term on the right side of (6.5).

Proposition 6.3 .

Assuming (A0) holds true and the conditions (3.4) and (3.5) are satisfied, we have

$$\sup_{x\in[a,b]} |h_n^b(x^+) - \overline{h_n^b}(x^+)| = o_{a.s.}(1) \quad as \quad n \to \infty. \qquad (6.9)$$

26

Proof of Proposition 6.3. Divide the interval $J = [a, b]$ into subintervals each of length $\delta_n = (b - a)/\nu_n$. Since the set $J_n = \{x; |x| \leq |b - a|\}$ is compact, it can be covered by a finite number of bounded intervals with centers x_{nj} whose sides are of length δ_n. That is $J = \bigcup_{j=1}^{\nu_n} J_{nj}$, where

$$J_{nj} = \left\{ x \; ; \; |x - x_{nj}| \leq (b - a)\nu_n^{-1} \right\}, \; j = 1, \ldots, \nu_n. \tag{6.10}$$

Let $V_n(x^+) = h_n^b(x^+) - \overline{h_n^b}(x^+)$, then we have, for $x_{nj} \in J_{nj}$, that

$$
\begin{aligned}
\sup_{x \in J} |V_n(x^+)| &= \max_{1 \leq j \leq \nu_n} \sup_{x \in J \cap J_{nj}} |V_n(x^+)| \\
&\leq \max_{1 \leq j \leq \nu_n} \sup_{x \in J \cap J_{nj}} |V_n(x^+) - V_n(x_{nj}^+)| + \max_{1 \leq j \leq \nu_n} |V_n(x_{nj}^+)| \\
&:= T_{1n} + T_{2n} + T_{3n},
\end{aligned}
$$

where

$$T_{1n} = \max_{1 \leq j \leq \nu_n} \sup_{x \in J_n \cap J_{nj}} |h_n^b(x^+) - h_n^b(x_{nj}^+)| \tag{6.11}$$

$$T_{2n} = \max_{1 \leq j \leq \nu_n} \sup_{x \in J_n \cap J_{nj}} |\overline{h_n^b}(x^+) - \overline{h_n^b}(x_{nj}^+)| \tag{6.12}$$

$$T_{3n} = \max_{1 \leq j \leq \nu_n} |h_n^b(x_{nj}^+) - \overline{h_n^b}(x_{nj}^+)|. \tag{6.13}$$

The proof is completed from Lemmas 6.3 and 6.4 below which provide an upper bound of each term in the above inequalities. □

Lemma 6.3 *Under (A0) we have*

$$
\begin{aligned}
(i) &\quad T_{1n} = O(\xi_n), \\
(ii) &\quad T_{2n} = O(\xi_n)
\end{aligned}
$$

where

$$\xi_n = C_1 a_n^{-4} \left(\frac{b_n^2}{a_n} \right)^{\alpha_n} \alpha_n^{3/2} M_n \, \nu_n^{-1}.$$

Proof of Lemma 6.3. We prove only (i), the proof of (ii) is similar. We have

$$|h_n^b(x^+) - h_n^b(x_{nj}^+)| \leq \frac{1}{n} \sum_{i=1}^{n} |\phi(Y_i)| I\{|\phi(Y_i)| \leq M_n\} \left| Q_{x+\epsilon_n, v_n}(X_i) - Q_{x_{nj}+\epsilon_n, v_n}(X_i) \right|.$$

27

Observe that

$$Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i) = \frac{X_i^{\alpha_n-1}}{\Gamma(\alpha_n)} \left[\frac{e^{-\alpha_n X_i/x^+}}{\beta_{x^+}^{\alpha_n}} - \frac{e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x_{nj}^+}^{\alpha_n}} \right], \qquad (6.14)$$

where $\beta_{x^+}^{\alpha} = (v_n^2 x^+)^{\alpha}$ and $\alpha_n = \frac{1}{v_n^2}$. The term in brackets in (6.14) can be written as

$$\frac{e^{-\alpha_n X_i/x^+} - e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x^+}^{\alpha_n}} + \frac{\left(\beta_{x_{nj}^+}^{\alpha_n} - \beta_{x^+}^{\alpha_n} \right) e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x^+}^{\alpha_n} \beta_{x_{nj}^+}^{\alpha_n}}. \qquad (6.15)$$

Since for $c \geq 0$ and for any t, $(a_n \leq t \leq b_n)$, the function $f_c(t) = e^{-c/t}$ is a K_c Lipshitz of order one with $K_c = \frac{c}{a_n^2} e^{-c/b_n}$, it follows by taking $c = c_n = \alpha_n X_i$, that for all $x, x_n^j \in [a, b]$,

$$\left| e^{-\alpha_n X_i/x^+} - e^{-\alpha_n X_i/x_{nj}^+} \right| \leq \frac{\alpha_n X_i \, e^{-\alpha_n X_i/b_n}}{a_n^2} \, |x - x_{nj}|. \qquad (6.16)$$

Moreover, making use of the mean value theorem, we can writefor x_* between x^+ and x_{nj}^+,

$$\begin{aligned} |\beta_{x^+}^{\alpha_n} - \beta_{x_{nj}^+}^{\alpha_n}| &\leq \alpha_n v_n^{2\alpha_n} |x - x_{nj}| x_*^{\alpha_n-1} \\ &\leq b_n^{\alpha_n-1} \alpha_n v_n^{2\alpha_n} |x - x_{nj}|. \end{aligned} \qquad (6.17)$$

Hence, combining (6.15), (6.16) and (6.17) we have,

$$\begin{aligned} &\left| Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i) \right| &(6.18) \\ &\leq \left[\frac{\beta_{b_n}^{\alpha_n} \, \alpha_n X_i}{\beta_{x^+}^{\alpha_n} \, a_n^2} Q_{b+\epsilon_n,v_n}(X_i) + \frac{b_n^{\alpha_n-1} \, \alpha_n \, v_n^{2\alpha_n}}{\beta_{x^+}^{\alpha_n}} Q_{x_{nj}+\epsilon_n,v_n}(X_i) \right] |x_i - x_{nj}| \\ &\leq \left[\frac{b_n \alpha_n}{a_n^3} \, X_i Q_{b+\epsilon_n,v_n}(X_i) + \frac{b_n^{\alpha_n-1} \, \alpha_n \, v_n^{2\alpha_n-2}}{a_n} Q_{x_{nj}+\epsilon_n,v_n}(X_i) \right] |x_i - x_{nj}|. (6.19) \end{aligned}$$

Using the Stirling's formula, we see that for x fixed and $\tau \geq 0$, the function $t \mapsto t^\tau Q_{x+\epsilon_n,v_n}(t)$

is bounded above by $\frac{(x+\epsilon_n)^{\tau-1}}{\sqrt{2\pi v_n}}$ whenever $v_n \to 0$. It then follows that

$$
\begin{aligned}
\left|Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i)\right| &\leq \frac{\alpha_n}{\sqrt{2\pi}\, v_n a_n^2}\left[\frac{b_n}{a_n} + b_n^{\alpha_n-1}\, v_n^{2\alpha_n-2}\right]|x - x_{nj}| \\
&\leq \frac{1}{\sqrt{2\pi}\, v_n^3 a_n^3}\frac{b_n}{a_n}\left[1 + \frac{a_n}{b_n^2}b_n^{\alpha_n}\,\alpha_n\right]|x - x_{nj}| \\
&\leq \frac{1}{\sqrt{2\pi}a_n^2}b_n^{\alpha_n-1}\,\alpha_n^{7/2}\,|x - x_{nj}|. \qquad (6.20)
\end{aligned}
$$

Therefore,

$$
|h_n^b(x^+) - h_n^b(x_{nj}^+)| \leq \frac{(b-a)}{\sqrt{2\pi}a_n^2}b_n^{\alpha_n-1}\alpha_n^{7/2}\nu_n^{-1}M_n = O(\xi_n)
$$

and hence $T_{1n} = O(\xi_n)$, which completes the proof of Lemma 6.3. $\qquad\square$

The following lemma deals with the asymptotic behavior of T_{3n}.

Lemma 6.4 . *Suppose that (A0) holds and the condition (3.5) is satisfied. Then we have*

$$
T_{3n} = o_{a.s.}(1) \quad as \quad n \to \infty. \qquad (6.21)
$$

Proof of Lemma 6.4. Write

$$
|h_n^b(x_{nj}^+) - \overline{h_n^b}(x_{nj}^+)| = \sum_{i=1}^{n} L_{n,i}(x_{nj}^+),
$$

where

$$
L_{n,i}(x_{nj}^+) = n^{-1}\{\phi(Y_i)I\{|\phi(Y_i)| \leq M_n\}Q_{x_{nj}+\epsilon_n,v_n}(X_i).
$$

It is then clear that for $x_{nj} \in [a, b]$,

$$
|L_{n,i}(x_{nj}^+)| \leq \frac{1}{\sqrt{2\pi}x_{nj}^+ v_n}n^{-1}M_n \leq \frac{1}{\sqrt{2\pi}a_n v_n}n^{-1}M_n,
$$

whenever $v_n \to 0$. Moreover, for any fixed j, $1 \leq j \leq \nu_n$, $L_{n,i}(x_{nj}^+)$ is a bounded triangular array of martingale differences with respect to $\mathcal{F}_i$. Behavior of this sequence may be studied using the following lemma due to Laïb [22].

Lemma 6.5 *Let $\{(X_i, \mathcal{S}_i) : i \geq 1\}$ be a sequence of martingale difference such that $|X_i| \leq B$ a.s. for $1 \leq i \leq n$. For all $\epsilon > 0$, one has*

$$
P\left\{\max_{1\leq i\leq n}\left|\sum_{j=1}^{i}X_j\right| > \epsilon\right\} \leq 2\exp\left(-\frac{\epsilon^2}{2nB^2}\right).
$$

29

Applying the above lemma we have for any $\lambda > 0$

$$P\{T_{3n} \geq \lambda\} \leq 2\nu_n \exp\left(-\pi\lambda^2 a_n n v_n^2 M_n^{-2}\right). \tag{6.22}$$

The result in (6.21) now follows from Borel-Cantelli Lemma and the condition (3.5). □

Proof of Theorem 3.1. The proof follows from decomposition (6.1), propositions 6.1 to 6.3 and lemmas 6.1 to 6.4. □

Proof of Theorem 3.2.

Part (i). We have from (6.1) that for any $x > 0$

$$\sqrt{nv_n}\left(\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)\right) = \frac{R_n(x^+)}{f_n(x^+)} - A_n(x^+), \tag{6.23}$$

where

$$R_n(x^+) = \sqrt{nv_n}\left((h_n(x^+) - \overline{h_n}(x^+)) - m(x)(f_n(x^+) - \overline{f_n}(x^+))\right)$$
$$\text{and}$$
$$A_n(x^+) = \sqrt{nv_n}\frac{\tilde{B}_n(x)(f_n(x^+) - \overline{f_n}(x^+))}{f_n(x^+)}.$$

Let

$$\eta_{ni} = \left(\frac{v_n}{n}\right)^{1/2}[(\phi(Y_i) - m(x))Q_{x+\epsilon_n,v_n}(X_i)] \quad \text{and} \quad \xi_{ni} = \eta_{ni} - E\left[\eta_{ni}|\mathcal{F}_{i-1}\right].$$

Then $R_n(x^+) = \sum_{i=1}^{n} \xi_{ni}$.

We will show that $A_n(x^+) = o_{\mathbb{P}}(1)$ and $f_n(x^+) \xrightarrow{\mathbb{P}} f(x)$ as $n \to \infty$. Therefore, a central limit theorem for the left hand-side of (6.23) may be obtained if we establish the asymptotic normality for the quantity $R_n(x^+)$. Lemma 6.6 below gives the weak consistency of the estimator $f_n(x^+)$.

Lemma 6.6 . *Assuming that (A0)-(A2) hold true, we have*

$$f_n(x^+) - f(x) = o_{\mathbb{P}}(1) \quad as \quad n \to \infty.$$

30

Proof of Lemma 6.6. The result given in the above lemma follows from a direct applications of Lemma 2.1 in conjunction with Lemma 6.5. □

The following Lemma gives the asymptotic behavior of $A_n(x^+)$.

Lemma 6.7 . *In addition to conditions (A0)-(A3), assume that $f(x) > 0$ at a given $x \geq 0$. Then we have*

$$A_n(x^+) = o_{\mathbb{P}}(1) \quad as \quad n \to \infty.$$

Proof of Lemma 6.7. Arguing as in the proof of Proposition 6.3 below, letting $m(x) = 0$ and $\phi(Y_i) \equiv 1$, we get for any $x > 0$, under condition (A2), the following central limit theorem for the density estimator,

$$\sqrt{nv_n}(f_n(x^+) - \overline{f}_n(x^+) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \frac{f(x)}{2\sqrt{\pi x}}\right).$$

Thus, $f_n(x^+) - \overline{f}_n(x^+) = O_{\mathbb{P}}(1/\sqrt{nv_n})$ and Lemma 6.6 yields $A_n(x^+) = O_{\mathbb{P}}(1)|\tilde{B}_n(x)|$ as $n \to \infty$. The result then follows from Proposition 6.1. □

Proposition 6.4 *Assuming conditions (A0), (A2) and (A6)(i) hold true, we have for a given $x \geq 0$ that*

$$R_n(x^+) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \tau^2(x)\right), \tag{6.24}$$

where

$$\tau^2(x) = \frac{f(x)}{2\sqrt{\pi x}}(W_2(x) - m^2(x)).$$

Proof of Proposition 6.4. Observe that for any fixed x, the summands in $R_n(x^+)$ form a triangular array stationary martingale differences with respect the sigma field $\mathcal{F}_{i-1}$, we can then apply a CLT for discrete-time arrays of real-valued martingales, as given for instance in [Hall and Heyde[19], page 23], to prove the asymptotic normality of $R_n(x^+)$. This will be done, if we check the following two conditions:

a) $\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \tau^2(x)$ (limiting variance)

31

b) $nE\left[\xi_{ni}^2 I_{[|\xi_{ni}|>\varepsilon]}\right] = o(1)$ holds true for any $\varepsilon > 0$ (Lindeberg condition).

Checking of the condition a). Observe that

$$\left|\sum_{i=1}^n \mathbb{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] - \sum_{i=1}^n \mathbb{E}\left[\xi_{ni}^2|\mathcal{F}_{i-1}\right]\right| \leq \sum_{i=1}^n \left(\mathbb{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right)^2.$$

Using Cauchy-Schwarz inequality we may write

$$|\mathbb{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]|^2 \leq \frac{v_n}{n}\left(\int_0^\infty (m(t)-m(x))^2 Q_{x+\epsilon_n,v_n}(t)dt\right)$$
$$\times \left(\int_0^\infty f^2(t|\mathcal{F}_{i-1})Q_{x+\epsilon_n,v_n}(t)dt.\right)$$

Moreover, conditions (A1) and (A2) lead to

$$\sum_{i=1}^n \left(\mathbb{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right)^2 \leq v_n \int_0^\infty (m(t)-m(x))^2 Q_{x+\epsilon_n,v_n}(t)dt$$
$$\times \int_0^\infty \left(n^{-1}\sum_{i=1}^n f^2(t|\mathcal{F}_{i-1})\right) Q_{x+\epsilon_n,v_n}(t)dt$$
$$\to 0 \quad \text{as } n \to \infty.$$

This is because, by Lemma 2.1 the first integral goes to 0 and the second one converges almost surely to $f^2(x)$, which is bounded, in view of condition (A2) together with Lemma 2.1. Condition a) follows then , if we prove

$$\sum_{i=1}^n \mathbb{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \tau^2(x). \tag{6.25}$$

We have

$$\sum_{i=1}^n \mathbb{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] = \frac{v_n}{n}\sum_{i=1}^n \mathbb{E}\left[W_2(X_i)Q_{x+\epsilon_n,v_n}^2(X_i)|\mathcal{F}_{i-1}\right]$$
$$-\frac{v_n}{n}\sum_{i=1}^n \mathbb{E}\left[m(x)(2m(X_i)-m(x))Q_{x+\epsilon_n,v_n}^2(X_i)|\mathcal{F}_{i-1}\right]$$
$$= J_{1n}(x) + J_{2n}(x). \tag{6.26}$$

32

The quantity $J_{1n}(x)$ can be be split as follows

$$J_{1n}(x) = v_n \int_{\mathbb{R}_*^+} W_2(t) Q^2_{x+\epsilon_n, v_n}(t) \left[\frac{1}{n} \sum_{i=1}^n f(t|\mathcal{F}_{i-1}) - f(t) \right] dt$$

$$+ v_n \int_{\mathbb{R}_*^+} W_2(t) Q^2_{x+\epsilon_n, v_n}(t) f(t) dt. \tag{6.27}$$

By (A2), the term in brackets in (6.27) goes to 0 a.s. uniformly in t. Moreover, $v_n \int_{\mathbb{R}_*^+} W_2(t) Q^2_{x+\epsilon_n, v_n}(t) dt$ is bounded above by

$$v_n \sup_t Q_{x+\epsilon_n, v_n}(t) \int_{\mathbb{R}_*^+} W_2(t) Q_{x+\epsilon_n, v_n}(t) dt \approx \frac{1}{\sqrt{2\pi x}} W_2(x)$$

since by Lemma 2.1 $\int_{\mathbb{R}_*^+} W_2(t) Q_{x+\epsilon_n, v_n}(t) dt \to W_2(x)$ as $n \to \infty$. This implies that the first member in $J_{1n}(x)$ goes to 0 a.s. as $n \to 0$. The second member in (6.27) can be split as

$$v_n \int_{\mathbb{R}_+^*} (W_2(t) f(t) - W_2(x) f(x) Q^2_{x+\epsilon_n, v_n}(t) dt + v_n \int_{\mathbb{R}_+^*} W_2(x) f(x) Q^2_{x+\epsilon_n, v_n}(t) dt. \tag{6.28}$$

To analyze the first term we prove the following lemma that can be easily established using Stirling's approximation.

Lemma 6.8 *We have for any $p \geq 0$ and $m \geq 1$*

$$\int_0^\infty t^p Q^m_{x+\epsilon_n, v_n}(t) dt = \frac{\left(\frac{1}{v_n^2(x+\epsilon 8n)} \right)^{m/v_n^2}}{\left(\frac{m}{v_n^2(x+\epsilon_n)} \right)^{((m/v_n^2)+p+1-m)}} \cdot \frac{\Gamma(m/v_n^2 + p + 1 - m)}{\Gamma^m(1/v_n^2)}$$

$$\approx \frac{1}{\sqrt{m(2\pi)^{m-1}}} \frac{1}{v_n^{m-1}(x+\epsilon_n)^{m-p-1}} \frac{1}{\sqrt{1 - v_n^2\left(\frac{m-p-1}{m} \right)}}, \quad as \ (v_n, \ \epsilon_n) \to (0, \ 0). \tag{6.29}$$

Use of a Taylor expansion of order one for the function $t \mapsto (W_2 f)$ and formula (6.29) one can show that the first member in (6.28) tends to 0 as $n \to \infty$. Therefore, as $\epsilon_n \to 0$, we have

$$J_{1n}(x) \approx \frac{f(x) W_2(x)}{2\sqrt{\pi} x} \quad \text{a.s.} \quad \text{as} \quad n \to \infty. \tag{6.30}$$

33

We now need to study the asymptotic behavior of the second member $J_{2n}(x)$. Observe that

$$
\begin{aligned}
J_{2n}(x) &= -v_n \int_{\mathbb{R}_*^+} (2m(t) - m(x)) \left(n^{-1} \sum_{i=1}^{n} f(t|\mathcal{F}_{i-1}) - f(t) \right) dt \\
&\quad - v_n m(x) \int_{\mathbb{R}_*^+} (m(t) - m(x)) Q_{x+\epsilon_n, v_n}^2(t) f(t) dt \\
&\quad - v_n m(x) \int_{\mathbb{R}_*^+} (m(t)f(x) - m(x)f(x)) Q_{x+\epsilon_n, v_n}^2(t) f(t) dt \\
&\quad - v_n m(x)^2 f(x) \int_{\mathbb{R}_*^+} Q_{x+\epsilon_n, v_n}^2(t) dt.
\end{aligned} \tag{6.31}
$$

Using the same argument as above we can easily see that $J_{2n}(x)$ is a.s. asymptotically equivalent, as $\epsilon_n \to 0$, to $-\frac{f(x)m^2(x)}{2\sqrt{\pi}\,x}$. That is

$$
J_{2n}(x) \approx -\frac{f(x)m^2(x)}{2\sqrt{\pi}\,x} \quad \text{a.s.} \quad n \to \infty. \tag{6.32}
$$

Then (6.25) follows from (6.30) and (6.32).

The Lindeberg condition results from Corollary 9.5.2 in Chow and Teicher[8] which implies that $nE[\xi_{ni}^2 I(|\xi_{ni}| > \epsilon)] \leq 4nE[\eta_{ni}^2 I(|\eta_{ni}| > \varepsilon/2)]$.

To check the above condition, let $p > 1$ and $q > 1$ be real numbers such that $p^{-1} + q^{-1} = 1$. Making use of Hölder and Markov inequalities one can write for all $\varepsilon > 0$

$$
E[\eta_{ni}^2 I(|\eta_{ni}| > \varepsilon/2)] \leq \frac{E|\eta_{ni}|^{2p}}{(\varepsilon/2)^{2p/q}}.
$$

Taking $2p = 2 + \delta$ and conditioning by X_1, we get by condition (A6)(i) that

$$
\begin{aligned}
4n\mathbb{E}[\eta_{ni}^2 I(|\eta_{ni}| > \varepsilon/2)] &= O(1)\, n^{-\delta} v_n^{(2+\delta)/2}\, \mathbb{E}\left[|(\phi(Y_i) - m(x))Q_{x+\epsilon_n, v_n}|^{2+\delta} \right] \\
&= O(1)\, n^{-\delta/2} v_n^{(2+\delta)/2} \left[\mathbb{E}\left(\phi(Y_1)\mathbb{E}\left(Q_{x+\epsilon_n, v_n}(X_1)\,|X_1\right)\right)^{2+\delta} + \mathbb{E}(m(x)Q_{x+\epsilon_n, v_n})^{2+\delta} \right] \\
&= O(1)\, n^{-\frac{\delta}{2}} v_n^{\frac{2+\delta}{2}} \left[\int_{\mathbb{R}_+^*} \overline{W}_{2+\delta}(t) Q_{x+\epsilon_n, v_n}^{2+\delta}(t) f(t) dt + m^{2+\delta}(x) \int_{\mathbb{R}_+^*} Q_{x+\epsilon_n, v_n}^{2+\delta}(t) f(t) dt \right]. \tag{6.33}
\end{aligned}
$$

The first term in square brackets in (6.33) can be written $U_{1n} + U_{2n}$ where

$$
U_{1n} = O(1)\, n^{-\frac{\delta}{2}} v_n^{\frac{2+\delta}{2}} \int_{\mathbb{R}_+^*} \overline{W}_{2+\delta}(x) f(x) Q_{x+\epsilon_n, v_n}^{2+\delta}(t) dt. \tag{6.34}
$$

34

and

$$U_{2n} = O(1)\, n^{-\frac{\delta}{2}} v_n^{\frac{2+\delta}{2}} \left[\int_{\mathbb{R}_+^*} \left(\overline{W}_{2+\delta}(t) f(t) - \overline{W}_{2+\delta}(x) f(x) \right) Q_{x+\epsilon_n,v_n}^{2+\delta}(t) dt \right] \quad (6.35)$$

Thus using the approximation formula given in (6.29), U_{1n} becomes

$$O(1) n^{-\delta/2} v_n^{(2+\delta)/2} \left[\int_{\mathbb{R}_+^*} \overline{W}_{2+\delta}(x) f(x) Q_{x+\epsilon_n,v_n}^{2+\delta}(t) \right] = O(1)(nv_n)^{-\delta/2}, \quad (6.36)$$

that converges to 0 as $n \to \infty$, since then $nv_n \to \infty$. Also, by continuity of the function $t \to \overline{W}_{2+\delta}(t) f(t)$ one can see that U_{2n} converges to 0 as $n \to \infty$. This implies that the first term of (6.33) converges to 0 as $n \to \infty$. Similarly, one can show that the second term in (6.33) is asymptotically negligible. This completes part (i) of the proof of Theorem 3.1.

Proof of Part (ii). To prove (ii) we need an estimate of the rate of convergence in probability of the bias term. This is the subject of the following Lemma.

Lemma 6.9 *Suppose that (A0), (A1), (A2), (A4)(i) and (A6)(ii) hold and the condition (2.6) is satisfied. Then we have*

$$|\tilde{B}_n(x)| = O_{\mathbb{P}} \left(\max \left(\max(v_n^2, \epsilon_n), n^{-1} \right) \right) = O_{\mathbb{P}} \left(\max(v_n^2, \epsilon_n) \right). \quad (6.37)$$

Proof of Lemma 6.9. It suffices to give an estimate of the convergence rate of the quantity $\overline{h}_n(x^+) - h(x)$ appears in (3.3). To this end write

$$\overline{h}_n(x^+) - h(x) = \left(\overline{h}_n(x^+) - \mathbb{E}(h_n(x^+)) \right) + \left(\mathbb{E}(h_n(x^+)) - h(x) \right). \quad (6.38)$$

Making use of (A4)(i) one may write

$$\overline{h}_n(x^+) - \mathbb{E}(h_n(x^+)) = \frac{1}{n} \int_{\mathbb{R}^+} m(t) Q_{x+\epsilon_n}(t) \left[\sum_{i=1}^n f(t \mid \mathcal{F}_{i-1}) - nf(t) \right] dt$$

$$= \frac{1}{n} \int_{\mathbb{R}^+} m(t) Q_{x+\epsilon_n,\, v_n}(t) H_n(t) dt, \quad (6.39)$$

where $H_n(t) := \sum_{i=1}^n f(t \mid \mathcal{F}_{i-1}) - nf(t)$. Here we quote a result from Wu [32] that is useful in the analysis of the difference in (6.39).

Lemma 6.10 [Wu [32]] *For any* $\mathbf{y} \in \mathbb{R}^d$ *let* $H_n(\mathbf{y}) = \sum_{i=1}^n f(\mathbf{y}|\mathcal{F}_i) - nf(\mathbf{y})$. *Then condition (2.6) implies that* $\sup_y \|H_n(\mathbf{y})\|_2^2 = O(n)$.

It follows from Cauchy-Schwarz inequality and Lemma (6.10) that

$$
\mathbb{E}\left[\left|\bar{h}_n(x^+) - \mathbb{E}(h_n(x^+))\right|^2\right] \leq \frac{1}{n^2} \cdot \left[\mathbb{E}\left(H_n^2(t)\right)\right]\left(\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n}(t)dt\right)^2
$$

$$
= O(n^{-1})\left(\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n}(t)dt\right)^2. \tag{6.40}
$$

In order to deal with the last term in small brackets in (6.40) recall that $Q_{x+\epsilon_n,v_n}(t) = (1/(x+\epsilon_n))q_{v_n}(t/(x+\epsilon_n))$, where $q_{v_n}(\cdot)$ is the gamma density function with mean 1 and variance v_n^2. Thus, we have, by the Taylor expansion of $m(\cdot)$

$$
\int_0^\infty Q_{x+\epsilon_n,v_n}(t)m(t)dt = \int_0^\infty q_{v_n}(s)m((x+\epsilon_n)s)ds
$$

$$
= \int_0^\infty q_{v_n}(s)[m(x) + (x(s-1)+s\epsilon_n))m'(x) + \frac{(x(s-1)+s\epsilon_n))^2}{2}m''(x)
$$

$$
+ o\left((x(s-1)+s\epsilon_n))^2\right)]ds
$$

$$
= O(1) + O(v_n^2) + O(\epsilon_n) + O\left(\max(v_n^2,\epsilon_n)\right)
$$

$$
= O(1) + O\left(\max(v_n^2,\epsilon_n)\right) = O(1). \tag{6.41}
$$

Thus $\mathbb{E}\left[\left|\bar{h}_n(x^+) - \mathbb{E}(h_n(x^+))\right|^2\right] = O(n^{-1})$. By the same argument as above one can easily see that $(\mathbb{E}(h_n(x^+)) - h(x)) = O\left(\max(v_n^2,\epsilon_n)\right)$. These lead to the desired result.

Proof of Part (iii). The proof is similar to that of part (ii). □

Proof of Proposition 3.1.

Write $m_n(x^+) = r_n(x^+)/f_n(x^+)$, $x \geq 0$ and note that

$$
m_n(x^+) - m(x) = (h_n(x^+)/f_n(x^+)) - m(x)
$$

$$
= (1/f_n(x^+))(h_n(x^+) - m(x)f(x)) - (m(x)/f_n(x^+))(f_n(x^+) - f(x)).
$$

Thus we may approximate

$$
\mathrm{MSE}(x) : = \mathbb{E}(m_n(x^+) - m(x))^2
$$

$$
\approx \frac{1}{(f(x))^2}\mathbb{E}(h_n(x^+) - m(x)f(x))^2 + \frac{m^2(x)}{(f(x))^2}\mathbb{E}(f_n(x^+) - f(x))^2
$$

$$
= \frac{\mathbb{E}(A_n^2(x))}{f^2(x)} + \frac{m^2(x)}{(f(x))^2}\mathbb{E}(D_n^2(x)). \tag{6.42}
$$

36

Here, the product term $-2(m(x)/f^2(x))\mathbb{E}[(h_n(x^+)-m(x)f(x))(f_n(x^+)-f(x))]$ has been ignored because by the Cauchy-Schwarz inequality

$$|\mathbb{E}[(r_n^+(x) - m(x)f(x))(f_n(x^+) - f(x))]|$$
$$\leq \sqrt{\mathbb{E}(h_n(x^+) - m(x)f(x))^2}\sqrt{\mathbb{E}(f_n(x^+) - f(x))^2}$$
$$\leq \max\{\mathbb{E}[(h_n(x^+) - m(x)f(x))]^2, \mathbb{E}[(f_n(x^+) - f(x))]^2\}.$$

Now write
$$A_n(x) = (h_n(x^+) - \overline{h}_n(x^+)) + (\overline{h}_n(x^+) - h(x)),$$

where $h(x) = m(x)f(x)$. We have then approximately (ignoring the product term as above)

$$\mathbb{E}(A_n^2(x)) \simeq \mathbb{E}(h_n(x^+) - \overline{h}_n(x^+))^2 + \mathbb{E}(\overline{h}_n(x^+) - h(x))^2 = I_{1n}(x) + I_{2n}(x).$$

Moreover, it follows from (6.38), (6.39), (6.40) and (6.41) that

$$I_{2n}(x) = O(n^{-1}) + (\mathbb{E}[h_n(x^+) - h(x^+)]^2) = O(n^{-1}) + (\text{bias}(h_n(x^+))^2.$$

Therefore,
$$\mathbb{E}(A_n^2(x)) \simeq \mathbb{E}(h_n(x^+) - \overline{h}_n(x^+))^2 + +(\text{bias}(h_n(x^+))^2.$$

Next let $W_2(t) = E(Y_1^2|X_1 = t)$ and $Z_{in} = Y_iQ_{x+\epsilon_n,v_n}(x) - \mathbb{E}\left[Y_iQ_{x+\epsilon_n,v_n}(x)\,|\mathcal{F}_{i-1}\right]$, then we have $\mathbb{E}(Z_{ni}Z_{nj}) = 0$ for $i \neq j$, and

$$\begin{aligned}
\mathbb{E}(h_n(x^+) - \overline{h}_n(x^+))^2 &\simeq n^{-2}\sum_{i=1}^{n}\mathbb{E}(Z_{in}^2) \\
&\approx n^{-1}E(Y_1^2Q_{x+\epsilon_n,v_n}^2(X_1)) \\
&= n^{-1}(x + \epsilon_n)^{-2}\int_0^\infty W_2(t)q_{v_n}^2(t/(x + \epsilon_n))f(t)dt, \\
&= n^{-1}(x + \epsilon_n)^{-1}\int_0^\infty W_2(t(x + \epsilon_n))f(t(x + \epsilon_n))q_{v_n}^2(t)dt
\end{aligned}$$
(6.43)

To further approximate the integral in the final expression above, we provide a result which can be easily proved using the Stirling approximation and routine calculations.

Lemma 6.11 *The $q_{v_n}(t)$ gamma density (with mean 1 and variance v_n^2) satisfies*

$$\begin{aligned}
q_{v_n}^2(t) &= \frac{(v_n^2/2)^{(2/v_n^2)-1}\Gamma((2/v_n^2) - 1)}{(v_n^2)^{(2/v_n^2)}\Gamma^2(1/v_n^2)}\frac{t^{(2/v_n^2)-2}\exp(-2t/v^2)}{(v_n^2/2)^{(2/v_n^2)-1}\Gamma((2/v_n^2) - 1)} \\
&= O(v_n^{-1})g_{\alpha,\beta}(t),
\end{aligned}$$
(6.44)

(6.45)

where $g_{\alpha,\beta}(t)$ is a Gamma density with $\alpha = (2/v_n^2) - 1$ and $\beta = v_n^2/2$.

Provided $W_2(\cdot)$ and $f(\cdot)$ are continuous, $\int_0^\infty W_2(t(x + \epsilon_n))f(t(x + \epsilon_n))g_{\alpha,\beta}(t) \rightarrow W_2(x)f(x)$ as $n \rightarrow \infty$, then it follows from Lemma (6.11) that we can write the above equation as

$$\mathbb{E}(h_n(x^+) - \overline{h}_n(x^+))^2 \approx \begin{cases} (nv_n x)^{-1}W_2(x)f(x) & \text{if } x > 0, \\ (nv_n\epsilon_n)^{-1}W_2(0)f(0) & \text{if } x = 0, \end{cases}$$

Further,

$$\begin{aligned} \text{bias}(h_n^+(x)) &= E(Y_1 Q_{x+\epsilon_n,v_n}(X_1)) - m(x)f(x) \\ &= (x + \epsilon_n)^{-1} \int_0^\infty m(t)q_{v_n}(t/(x + \epsilon_n))f(t)dt - m(x)f(x) \\ &= \int_0^\infty [h(t(x + \epsilon_n)) - h(x)]q_{v_n}(t)dt, \quad \text{where } h(x) = m(x)f(x), \\ &= \int_0^\infty [(x(t - 1) + \epsilon_n t)h'(x) + (1/2)(x(t - 1) + \epsilon_n t)^2 h''(x)]q_{v_n}(t)dt + \theta_n(x), \end{aligned}$$

where $\theta_n(x)$ denotes the 3rd and higher order terms of the Taylor's expansion used above. This can be further simplified to yield

$$\begin{aligned} \text{bias}(h_n^+(x)) &= xh'(x)\int_0^\infty (t - 1)q_{v_n}(t)dt + \epsilon_n h'(x)\int_0^\infty tq_{v_n}(t)dt \\ &\quad + (1/2)x^2 h''(x)\int_0^\infty (t - 1)^2 q_{v_n}(t)dt + xh''(x)\epsilon_n \int_0^\infty t(t - 1)q_{v_n}(t)dt \\ &\quad + (1/2)\epsilon_n^2 h''(x)\int_0^\infty t^2 q_{v_n}(t)dt + \theta_n(x) \\ &= \epsilon_n h'(x) + (1/2)v_n^2 x^2 h''(x) + \epsilon_n v_n^2 xh''(x) + (1/2)\epsilon_n^2(1 + v_n^2)h''(x) + \theta_n(x) \\ &\approx \epsilon_n h'(x) + (1/2)v_n^2 x^2 h''(x). \end{aligned} \tag{6.46}$$

The final expression is obtained using the facts that

$$\int_0^\infty tq_{v_n}(t)dt = 1, \quad \int_0^\infty (t - 1)^2 q_{v_n}(t)dt = v_n^2, \quad \int_0^\infty t^2 q_{v_n}(t)dt = v_n^2 + 1,$$

and ignoring higher order terms. The expectation $\mathbb{E}(D_n^2(x))$ in the right hand side of (6.42) may be handled similarly by taking $Y = 1$ and $h(x) = f(x)$. The expression of the MSE then follows from (6.42), (6.43) and (6.46). □

Proof of Theorem 4.1. We only give the proof when $p = 2$. The proofs of Lemma 6.6 and Lemma 6.7 remain still intact since Lemma 2.1 also holds on $\mathbb{R}^{+p}$. Let now g be a function

defined on $\mathbb{R}^{+p}$ that possesses continuous bounded partial derivatives of order one at each point of an open set $S \subset \mathbb{R}^{+p}$. Then for each point (s,t), $(s,t) \neq (x + \epsilon_{1n}, y + \epsilon_{2n}) :=$ (x^+, y^+), for the line segment $L((s,t), (x^+, y^+))$ joining (s,t) and (x^+, y^+) in S, we have

$$\int_{\mathbb{R}^{+d}} g(s,t) Q^2_{(x+\epsilon_{n1}, y+\epsilon_{n2}),v}(s,t) ds dt = I_{1n} + I_{2n}, \tag{6.47}$$

where

$$
\begin{aligned}
I_{1n} &= \int_{\mathbb{R}^{+p}} [g(s,t) - g(x^+, y^+)] Q^2_{(x+\epsilon_{1n}, y+\epsilon_{2n}),v}(s,t) ds dt \\
&\approx \int_{\mathbb{R}^{+p}} \left[(s - x^+) \frac{\partial g}{\partial x^+}(x^+, y^+) - (t - y^+) \frac{\partial g}{\partial y^+}(x^+, y^+) \right] Q^2_{x+\epsilon_{1n},v}(s) Q^2_{y+\epsilon_{2n},v}(t) ds dt
\end{aligned}
$$

converges to 0 as $n \to \infty$, in view of (6.29), whenever the partial derivatives of g are bounded. Using again (6.29) and the continuity of g, we get

$$I_{2n} = \int_{\mathbb{R}^{+d}} g(x^+, y^+) Q^2_{(x+\epsilon_{n1}, y+\epsilon_{n2}),v}(s,t) ds dt \approx g(x,y) \frac{1}{4\pi v^2 xy}$$

as $(\epsilon_{1n}, \epsilon_{2n}) \to (0,0)$. To complete the proof, it suffices to replace in the proof of Proposition 6.4, v_n by v_n^p and to apply the above result with $g(s,t) = W_2(s,t) f(s,t)$ in (6.28) and $g(s,t) = m(s,t) f(s,t)$ in (6.31) and finally $g(s,t) = W_{2+\delta}(s,t) f(s,t)$ in (6.33). $\qquad \square$

References

[1] Andrews, D.W.K., 1984, Non-strong mixing autoregressive processes. *J. Appl. Probab,* **21**, 930–934.

[2] Bagai, I. & Prakasa Rao, B.L.S., 1996, Kernel type density estimates for positive valued random variables. *Sankhy_a*, **A57**, 56–67.

[3] Bosq, D., 1998, *Nonparametric Statistics for Stochastic Processes*. Lecture Note in Statistics. Springer Verlag, New York.

[4] Chaubey, Y.P. & Sen, P.K., 1996, On smooth estimation of survival and density functions. *Statist. Decisions*, **14** , 1–22.

[5] Chaubey,Y.P., Sen, A. & Sen, P.K., 2007, A new smooth density estimator for nonnegative random variables, Technical Report, Department of Mathematics & Statistics, Concordia University, Montreal, Canada.

[6] Chen, S.X., 2000, Probability density function estimation using gamma kernels. *Ann.Inst. Statist. Math.*, **52**, 471-480.

[7] Chernick, M.R., 1981, A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Ann. Probab.*, **9**, 145–149.

[8] Chow, Y.S. & Teicher, H. *Probabilty Theory*, 2nd ed. Springer Verlag, New York.

[9] Diaconis, P. & Freedman, D., 1999, Iterated random functions. *SIAM Rev*, **41**, 41–76.

[10] Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

[11] Fan, J., 1993, Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.,* **21**, 196-216.

[12] Fan, J. Q. & Gijbles, I., 1992, variable bandwidth and local linear regression smoothers, *Ann. Statist*., **20**, 2008–2036.

[13] Fan, J. & Gijbles, I., 1996, *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall

[14] Feller, W., 1965, *An Introduction to Probability Theory and its Applications*, vol II. New York, John Wiley and Sons.

[15] Gasser, T. & Müller, H.G., 1979, *Kernel estimation of regression functions*. In: Smoothing Techniques for Curve Estimation, eds. Gasser and Rosenblatt. Heidelberg, Springer-Verlag.

[16] Guégan, D. & Ladoucette, S., 2001, Non-mixing properties of long memory processes.*C. R. Acad. Sci. Paris. Sér. I*, **333**, 373–376.

[17] Györfi, L. & Lugosi, G., 1992, Kernel density estimation from ergodic sample is not universally consistent. *Comp. Stat. and Dta Analysis.*, **14**, 437–442.

[18] Haggan , V. & Ozaki, T., 1981, Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, **68**, 189–196.

[19] Hall, P. & Heyde, C., 1980, *Martingale Limit Theory and its Applications*. Academic Press, New York.

[20] Hall, P. & Wehrly, T.E., 1991, A geometrical method for removing edge effects from kernel-type nonparametric estimators. *Journal of the American Statistical Association*, **86**, 665–672.

[21] Härdle, W., 1990, Applied nonparametric regression. *Cambridge University Press*.

[22] Laïb, N., 1999, Exponential-type inequalities for martingale difference sequences: Application to nonparametric regression estimation. *Communication in Statistics, Theory and Methods*, **Series A. 28**, 1565–1576.

[23] Laïb, N., 2005, Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *Journal of Statistical Planing and inference*, **134**, 116–139.

[24] Müller, H.G., 1984, Boundary effects in nonparametric curve estimation models. *In: COMPSTAT. Physica Verlag*, 84–89

[25] Rice, J.A., 1984, Boundary modification for kernel regression. *Communication in Statistics*, **Series A, 13**, 893–900.

[26] Scaillet, O., 2004, Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics,* **16**, 217–226.

[27] Seifert, B. & Gasser, T., 1996, Finite sample variance of local polynomials: Analysis and solutions, *J. Amer. Statist. Assoc.*, **96**, 267–275.

[28] Silverman, B.W., 1986, *Density estimation for statistics and data analysis.* London: Chapman and Hall.

[29] Tran, L.T., 1994, Density estimation for time series by histograms. *Journal of Statistical Planing and inference*, **40**, 61–79.

[30] Tong, H., 1990, *Non-linear time series: a dynamical system approach.* Oxford University Press.

[31] Wand, M.P., Marron J.S. & Ruppert, D., 1991, Transformations in density estimation. *J. Amer. Statist. Assoc.*, **86**, 343–361.

[32] Wu, W.B., 2003, Nonparametric estimation for stationary processes. *Technical Report. University of Chicago.*, **No 536**.

[33] Wu, W.B. & Shao, X., 2004, Limit theorems for iterated random functions. *J. Appl. Prob.*, **41**, 425–436.

[34] Zhang, S., Karunamuni, R.J. & Jones, M.C., 1999, An improved estimator of the density function at the boundary. *J. Amer. Statist. Assoc.*, **94**, 1231–1241.