Nonparametric density estimation for linear processes with infinite variance

Toshio HONDA

February 2006

August 2006 (Revised)

NONPARAMETRIC DENSITY ESTIMATION FOR LINEAR PROCESSES

WITH INFINITE VARIANCE

Toshio Honda

Graduate School of Economics, Hitotsubashi University

2-1 Naka, Kunitachi, Tokyo 186-8601, JAPAN

NONPARAMETRIC DENSITY ESTIMATION

Abstract. We consider nonparametric estimation of marginal density functions of linear processes by using kernel density estimators. We assume that the innovation processes are i.i.d. and have infinite-variance. We present the asymptotic distributions of the kernel density estimators with the order of bandwidths fixed as $h = cn^{-1/5}$, where n is the sample size. The asymptotic distributions depend on both the coefficients of linear processes and the tail behavior of the innovations. In some cases, the kernel estimators have the same asymptotic distributions as for i.i.d. observations. In other cases, the normalized kernel density estimators converge in distribution to stable distributions. A simulation study is also carried out to examine small sample properties.

Key words and phrases: linear processes, kernel density estimator, domain of attraction, stable distribution, noncentral limit theorem, martingale central limit theorem.

1. Introduction

Let $\{X_i\}_{i=1}^{\infty}$ be a linear process defined by

$$(1.1) \qquad X_i = \sum_{j=0}^{\infty} b_j \epsilon_{i-j}, \ i = 1, 2, \ldots,$$

where $\{\epsilon_i\}_{i=-\infty}^{\infty}$ is an i.i.d. process, $b_0 = 1$, and $b_j \sim c_0 j^{-\beta}$, $j = 1, 2, \ldots$, and c_0 is a positive constant. $a_j \sim a'_j$ means that $a_j/a'_j \to 1$ as $j \to \infty$. The marginal density function of $\{X_i\}_{i=1}^{\infty}$ is denoted by $f(x)$. We will specify the conditions on $\{\epsilon_i\}$ and β later in this section.

In this paper we estimate the marginal density function $f(x)$ by kernel density estimators and present the asymptotic properties when ϵ_1 has infinite variance. The asymptotic distributions depend on both the tail behavior of ϵ_1 and β.

A lot of authors have examined the asymptotic properties of kernel density estimators of marginal density functions of dependent observations. Most of them considered density estimation for mixing processes by imposing assumptions on joint density functions and the order of mixing coefficients until about the early 1990's. See Fan and Yao (2003) for a review of the results for strongly mixing processes. However, it is difficult to ensure that the order of mixing coefficients satisfies the assumptions unless the coefficients b_j decay sufficiently fast. See Doukhan (1994) for sufficient conditions for linear processes to be strongly mixing. Therefore attention has been focused on the asymptotic properties of kernel density estimators for subordinated Gaussian processes $\{G(X_i)\}$, where $\{X_i\}$ is a stationary Gaussian process, and linear processes, especially subordinated Gaussian processes and linear processes with long memory since the late 1980's or the early 1990's. Note that Hall and Hart (1990) pointed out that the asymptotic properties depend on the degree of long memory when we estimate the marginal density functions of linear

2

processes with long memory. This is true of subordinated Gaussian processes with long memory.

As for subordinated Gaussian processes with long memory, there are, for example, Cheng and Robinson (1991), Csörgő and Mielniczuk (1995), and Ho (1996). See also the references therein. They examined the asymptotic properties of kernel density estimators by exploiting Hermite expansions. Ho (1996) proved that kernel density estimators behave asymptotically in the same way as for i.i.d. observations when the degree of long memory does not exceed a level, or we can say when the degree of long memory is weak. He proved it by evaluating the moments

Hidalgo (1997) studied the asymptotic properties of kernel density estimators for linear processes by appealing to Appell expansions. However, very restrictive conditions are necessary to verify the validity of Appell expansions and the paper does not mention those conditions at all. See section 6 of Giraitis and Surgailis (1986) for the conditions.

Theoretical studies for linear processes with long memory have developed since the seminal papers of Ho and Hsing, Ho and Hsing (1996, 1997). They applied the martingale decomposition method to examine the properties of subordinated linear processes with long memory. See Koul and Surgailis (2002) for Hermite expansions and the martingale approach of Ho and Hsing (1996,1997).

Recently several authors considered the asymptotic properties of kernel density estimators for linear processes with short memory or long memory by using the martingale approach initiated by Ho and Hsing (1996, 1997), for example, Honda (2000), Wu and Mielniczuk (2002), Schick and Wefelmeyer (2004), and Bryk and Mielniczuk (2005). See also the references therein. Especially Wu and Mielniczuk (2002) fully examined the asymptotic properties of kernel density estimators for linear processes. However, all of

them assumed ϵ_1 has finite variance and that the distribution of ϵ_1 satisfies some restrictive assumptions, for example, the existence of the bounded and Lipschitz continuous density function. Under those conditions, Wu and Mielniczuk (2002) proved that kernel density estimators behave asymptotically in the same way as for i.i.d. observations in the cases of short memory and weak long memory and that kernel density estimators behave asymptotically in the same way as the sample means when the degree of long memory exceeds a level. It is well known that standardized sample means of linear processes with long memory converge in distribution to the standard normal distribution when ϵ_1 has finite variance. The standardization is different from that for linear processes with short memory. See Theorem 5.2.3 of Taniguchi and Kakizawa (2000). Note that Hallin and Tran (1996) considered kernel density estimation for linear processes with short memory by appealing to truncation arguments. They assumed that ϵ_1 has finite variance and $\beta > 4$.

From a theoretical point of view, the marginal density function of $\{X_i\}$, $f(x)$, exists without finite variance or the bounded density function of ϵ_1. It is strange that theoretical studies of kernel density estimators are limited to the cases where ϵ_1 has finite variance and the bounded density function. Besides recently a lot of attention is paid to heavy tailed time series data. Therefore it is important to investigate the asymptotic properties of kernel density estimators in the cases where ϵ_1 does not have finite variance.

We examine kernel density estimators in the cases where ϵ_1 does not have finite variance by exploiting the results of Hsing (1999), Koul and Surgailis (2001), Surgailis (2002), and Pipiras and Taqqu (2003). We treat the asymptotic properties in a comprehensive way. We briefly mention the results of the above papers later in this section. In addition **C2** below also allows unbounded or discontinuous density functions of ϵ_1. When ϵ_1 does

not have finite variance, the asymptotic distributions depend on both the tail behavior of ϵ_1 and β. When the effect of the heavy tail of ϵ_1 and dependence among observations does not appear, the asymptotic distributions are the same as for i.i.d. observations. When the effect of the heavy tail of ϵ_1 and dependence among observations appears, the asymptotic distributions are stable distributions. Hereafter we shall call the effect that of α and β. In order to see the differences between asymptotic properties and small sample properties, we carried out a simulation study and the result is given in section 3.

We describe the conditions on $\{\epsilon_i\}$. We suppose that $0 < \alpha < 2$ in this paper and $G(x)$ denotes the distribution function of ϵ_1.

C1: $\lim_{x \to -\infty} |x|^\alpha G(x) = \lim_{x \to \infty} x^\alpha (1 - G(x)) = c_1 > 0$. $\mathrm{E}\{\epsilon_1\} = 0$ when $1 < \alpha < 2$.

C2: Letting $\phi(\theta)$ denote the characteristic function of ϵ_1, we have $|\phi(\theta)| < C(1 + |\theta|)^{-\delta}$ for some positive δ.

C stands for generic positive constants whose values change from place to place and are independent of the sample size n.

C1 implies

$$(1.2) \qquad \mathrm{E}\{|\epsilon_1|^r\} < \infty, \quad 0 < r < \alpha, \quad \text{and} \quad \mathrm{E}\{|\epsilon_1|^r\} = \infty, \quad r \geq \alpha,$$

and that the distribution of ϵ_1 belongs to the domain of attraction of a symmetric α-stable distribution. The characteristic function of the α-stable distribution $S_\alpha(\sigma, \eta, \mu)$ has the form of

$$(1.3) \qquad \begin{cases} \exp\{-\sigma^\alpha |\theta|^\alpha (1 - \mathrm{i}\eta \operatorname{sign}(\theta) \tan(\pi\alpha/2)) + \mathrm{i}\mu\theta\} \,, \ \alpha \neq 1 \\[2mm] \exp\{-\sigma|\theta|(1 + 2\mathrm{i}\eta \operatorname{sign}(\theta) \log|\theta|/\pi) + \mathrm{i}\mu\theta\} \quad, \ \alpha = 1 \end{cases},$$

where i stands for the imaginary unit. It is called symmetric if $\eta = \mu = 0$. See Samorodnitsky and Taqqu (1994) for more details about stable distributions. **C2** is necessary for

5

the existence and regularity conditions of both $f(x)$ and joint density functions of some random variables. See **P1-3** in section 2.

We fix x_0 and estimate $f(x_0)$ by the kernel density estimator $\hat{f}(x_0)$ defined below.

(1.4)
$$\hat{f}(x_0) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big),$$

where h is a bandwidth and $K(u)$ is a kernel function. We take $h = c_2 n^{-1/5}$ for some positive c_2 because of simplicity of presentation and partly because this is the optimal order when $f(x)$ is twice differentiable at x_0 and the effect of α and β does not appear. An exposition of the effect of bandwidths is given in section 2. We assume that $K(u)$ is a symmetric bounded density function with compact support.

We examine the asymptotic properties of $\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}$ in the following three cases.

Case 1: $1 < \alpha < 2$ and $1/\alpha < \beta < 1$

Case 2: $1 < \alpha < 2$ and $1 < \beta < 2/\alpha$

Case 3: $0 < \alpha < 2$ and $2/\alpha < \beta$

When $1 < \alpha < 2$, we have by the von Bahr and Esseen inequality that

$$\mathrm{E}\{|X_1|^r\} \leq C \sum_{j=1}^{\infty} j^{-\beta r} < \infty$$

for any r such that $\beta r > 1$ and $1 < r < \alpha$. Note that X_1 has infinite variance. X_1 is well defined in Case 3, too. See the proof of Theorem 2.2 of Pipiras and Taqqu (2003). Some authors say that $\{X_i\}$ has long memory in Case 1.

Koul and Surgailis (2001) deals with Case 1 and the weak convergence of empirical distribution functions is proved. The asymptotics of M-estimators of linear regression models are also examined there. Surgailis (2002) deals with Case 2. The asymptotic properties of empirical distribution functions of X_i and partial sums of $H(X_i)$, where

$H(x)$ is any bounded function, are given there. Hsing (1999) and Pipiras and Taqqu (2003) examined the asymptotic properties of partial sums of $H(X_i)$, where $H(x)$ is any bounded function, in Case 3. Those papers are crucial to our results and those papers are also based on the martingale decomposition method of Ho and Hsing (1996, 1997). We have to obtain necessary theoretical results to deal with the cases of infinite variance other than Cases 1-3 above. It is a subject of future research. Peng and Yao (2004) applied Hsing (1999), Koul and Surgailis (2001), and Surgailis (2002) to nonparametric estimation of trend functions, i.e. nonparametric regression with fixed design. There is a close similarity between nonparametric regression with random design and kernel density estimation. However, nonparametric estimation of trend functions and kernel density estimation are different problems.

In Case 1, the effect of the heavy tail of ϵ_1 and the dependence among observations, which we call that of α and β, appears when $\beta < 1/\alpha + 2/5$. In Case 2 when $\alpha\beta < 5/3$. However, we see no effect of α and β in Case 3. We repeat that the asymptotic distributions are the same as for i.i.d. observations when the effect of α and β does not appear. In Peng and Yao (2004), the effect of α and β always appears.

The paper is organized as follows. In section 2, we decompose $\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}$ into two components and give a heuristic argument of the asymptotic asymptotic properties of $\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}$. Then the main theorems of this paper are presented. We state the result of a simulation study in section 3. The main theorems are proved in section 4. The proofs of technical lemmas are confined to section 5.

2. The asymptotic distributions

We state the main results of this paper in Theorems 2.1-3. First we give definitions and notations. Then necessary properties of density functions are described. We present the asymptotic distributions of kernel density estimatiors after a heuristic argument. The proofs of the theorems are deferred to section 4.

Let $\xrightarrow{d}$ and $\xrightarrow{p}$ stand for convergence in distribution and convergence in probability, respectively. We omit $n \to \infty$ and a.s. for brevity.

We rewrite X_i as

$$(2.1) \qquad\qquad X_i = X_{i,j} + \tilde{X}_{i,j},$$

where

$$X_{i,j} = \sum_{l=0}^{j-1} b_l \epsilon_{i-l} \quad \text{and} \quad \tilde{X}_{i,j} = \sum_{l=j}^{\infty} b_l \epsilon_{i-l}.$$

We denote the distribution functions of $X_{i,j}$ and $\tilde{X}_{i,j}$ by $F_j(x)$ and $\tilde{F}_j(x)$, respectively. **C2** and Lemma 1 of Giraitis et al. (1996) imply the existence of the density functions and we denote them by $f_j(x)$ and $\tilde{f}_j(x)$, respectively.

We state necessary properties of density functions, which can be verified by using **C1** and **C2**. **P1** and **P2** are derived by following the proof of Lemmas 1-2 of Giraitis et al. (1996). **P3** is part of Lemma 4.2 of Koul and Surgailis (2001).

There exists a positive integer s_1 for which **P1**, **P2**, and **P3** hold.

P1: $f_s(x)$ is twice continuously differentiable and $f_s(x)$ and all the derivatives up to the second order are uniformly bounded for $s \geq s_1$. Note that we can take $s = \infty$.

P2: $(X_{1,s_1}, X_{i,s_1+i-1})$ has the bounded joint density function for any $i \geq 2$.

P3: When $1 < \alpha < 2$, $1 < r < \alpha$, and $r\beta > 1$, there exists a constant C such that $|f'(x) - f'_s(x)| \leq C|s|^{1/r-\beta}$ for any $s \geq s_1$. Note that C depends on α, β, and r.

When we use **P3** in the proofs of Lemmas 4.2 and 4.4, r is specified and satisfies the conditions in **P3**.

Before we state Theorems 2.1-3, we give a heuristic argument of the asymptotics of $\hat{f}(x_0)$. We need to decompose $\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}$ into two components for the argument. Let be s_2 a large positive integer and put $s_0 = s_1 + s_2$. We will be more specific about s_2 in the proofs of Theorems 2.1-3. We write $\mathcal{S}_i$ for the σ-field generated by $\{\epsilon_j \,|\, j \leq i\}$.

$$(2.2) \qquad\qquad \hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\} = S_a + S_b,$$

where

$$(2.3) \qquad S_a = \frac{1}{nh} \sum_{i=1}^{n} \Big[K\Big(\frac{X_i - x_0}{h}\Big) - \mathrm{E}\Big\{K\Big(\frac{X_i - x_0}{h}\Big)\Big|\mathcal{S}_{i-s_0}\Big\} \Big]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Big[\frac{1}{h} K\Big(\frac{X_i - x_0}{h}\Big) - \int K(\xi) f_{s_0}(x_0 + \xi h - \tilde{X}_{i,s_0}) d\xi \Big]$$

$$(2.4) \qquad S_b = \frac{1}{nh} \sum_{i=1}^{n} \Big[\mathrm{E}\Big\{K\Big(\frac{X_i - x_0}{h}\Big)\Big|\mathcal{S}_{i-s_0}\Big\} - \mathrm{E}\Big\{K\Big(\frac{X_i - x_0}{h}\Big)\Big\} \Big]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Big[\int K(\xi) f_{s_0}(x_0 + \xi h - \tilde{X}_{i,s_0}) d\xi - \frac{1}{h} \mathrm{E}\Big\{K\Big(\frac{X_i - x_0}{h}\Big)\Big\} \Big].$$

The domain of integration is $(-\infty, \infty)$ when it is omitted. Remember that $X_{i,j} = \sum_{l=0}^{j-1} b_l \epsilon_{i-l}$ and that $f_j(x)$ is the density function of $X_{i,j}$. Similar expressions can be found in (3)-(5) of Wu and Mielniczuk (2002). In Wu and Mielniczuk (2002) and Bryk and Mielniczuk (2005), $s_0 = 1$ and the Lipschitz continuous density function of ϵ_1 is assumed. They applied the martingale central limit theorem to S_a. A technique is devised to avoid such assumptions on ϵ_1 in this paper.

The asymptotic properties of S_a are examined in Lemma 4.1 below. We investigate the asymptotic properties of S_b in section 4 by using the results of Hsing (1999), Koul and Surgailis (2001), Surgailis (2002), and Pipiras and Taqqu (2003). The asymptotic distribution of (2.2) depends on which of S_a and S_b is stochastically larger.

9

We put $h = c_2 n^{-\gamma} (\gamma > 0)$ only in this heuristic argument. In either case, we have

$$
(2.5) \qquad S_a = O_p((nh)^{-1/2}) \quad \text{and} \quad \mathrm{E}\{\hat{f}(x_0)\} - f(x_0) \sim \frac{h^2}{2} f''(x_0) \nu,
$$

where $\nu = \int u^2 K(u) du$. The asymptotic properties of S_b are independent of h and depend only on α and β. In addition we have

$$
S_b = O_p(n^{1/\alpha - \beta}) \text{ in Case 1}, \quad S_b = O_p(n^{-1+1/(\alpha\beta)}) \text{ in Case 2}, \quad S_b = O_p(n^{-1/2}) \text{ in Case 3}.
$$

The stochastic order is exact in all the above expressions. Then (2.5) and a simple calculation imply that we cannot improve the rate of convergence of $\hat{f}(x_0)$ by choosing γ other than $1/5$. If there are three parameters, α, β, and γ, things may be complicated and we present the theorems with $\gamma = 1/5$ to avoid possible complications.

When S_b is stochastically larger in S_a, the effect of α and β appears in the asymptotic properties of $\hat{f}(x_0)$. Since the asymptotic properties of S_b are independent of h and depend only on α and β, we have no optimal bandwidth and we can choose larger bandwidths without affecting the asymptotic properties of $\hat{f}(x_0)$.

When S_a and S_b have the same stochastic order, we can say that the effect of α and β still appears. However, we have no result on the joint distribution of S_a and S_b and we do not refer to this case in this paper.

When $\mathrm{E}\{|\epsilon_1|^{2+\delta}\} < \infty$ for some positive δ and $h = c_2 n^{-1/5}$, the effect of dependence among observations does not appear in the case of $\beta > 9/10$ in contrast to Case 2 below.

Here we state the main results of this paper.

Case 1: When β is smaller than $1 \wedge (1/\alpha + 2/5)$, the effect of α and β appears in the asymptotic properties. When α is smaller than $5/3$, the effect of long memory always appears.

THEOREM 2.1. *Suppose that* **C1** *and* **C2** *hold and that* $1 < \alpha < 2$ *and* $1/\alpha < \beta < 1$.

Then we have

$1/\alpha - \beta < -2/5$: $\sqrt{nh}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} \mathrm{N}(0, \kappa f(x_0))$,

$1/\alpha - \beta > -2/5$: $n^{\beta - 1/\alpha}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} -f'(x_0)c_A Z$,

where $\kappa = \int K^2(u)du$,

$$c_A = c_0\Big(2c_1\frac{\Gamma(2-\alpha)\cos(\alpha\pi/2)}{1-\alpha}\int_{-\infty}^1\int_0^1(t-s)_+^{-\beta}dtds\Big)^{1/\alpha},$$

$x_+ = x \vee 0$, and Z is a random variable whose distribution is $S_\alpha(1,0,0)$. The asymptotic joint distributions of the kernel density estimators at different points are independent in case of $1/\alpha - \beta < -2/5$ and degenerate in case of $1/\alpha - \beta > -2/5$, respectively.

Case 2: When $1 < \beta < 5/(3\alpha)$, the effect of α and β appears in the asymptotic properties. When α is larger than $5/3$, the effect does not appear.

THEOREM 2.2. *Suppose that **C1** and **C2** hold and that $1 < \alpha < 2$ and $1 < \beta < 2/\alpha$. Then we have*

$1/(\alpha\beta) < 3/5$: $\sqrt{nh}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} \mathrm{N}(0, \kappa f(x_0))$,

$1/(\alpha\beta) > 3/5$: $n^{1-1/(\alpha\beta)}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} (c_1 c_0^\alpha/(\sigma_{\alpha\beta}\beta^{\alpha\beta}))^{1/(\alpha\beta)}(c_f^+ L_+ + c_f^- L_-)$,

where L_+ and L_- are mutually independent random variables whose distributions are $S_{\alpha\beta}(1,1,0)$,

$$c_f^+ = \int_0^\infty(f(x_0-t) - f(x_0))t^{-1-1/\beta}dt,$$

$$c_f^- = \int_0^\infty(f(x_0+t) - f(x_0))t^{-1-1/\beta}dt,$$

$$\sigma_{\alpha\beta} = \Gamma(2-\alpha\beta)|\cos(\pi\alpha\beta/2)|/(\alpha\beta-1).$$

The asymptotic joint distributions of the kernel density estimators at different points are independent in case of $1/(\alpha\beta) < 3/5$ and degenerate in case of $1/(\alpha\beta) > 3/5$, respectively.

Case 3: In this case, we see no effect of α and β in the asymptotic properties.

THEOREM 2.3. *Suppose that* **C1** *and* **C2** *hold and that* $0 < \alpha < 2$ *and* $2/\alpha < \beta$. *Then we have*

$$\sqrt{nh}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} \mathrm{N}(0, \kappa f(x_0)).$$

The asymptotic joint distributions of the kernel density estimators at different points are independent.

When the effect of α and β does not appear, we can define the asymptotically optimal bandwidth in the same way as for i.i.d. observations.

By combining Theorems 2.1-2, we know the effect of α and β appears in the following cases.

• $1 < \alpha < 5/3$

  $1/\alpha < \beta < 1$ in Case 1 and $1 < \beta < 5/(3\alpha)$ in Case 2

• $5/3 < \alpha < 2$

  $1/\alpha < \beta < 1/\alpha + 2/5$ in Case 1

Then we can only say that larger bandwidths will improve small sample properties and it will be difficult to conduct statistical inference. The same problem happens for linear processes with long memory, too. However, it is important to know the statistical properties of such often used estimators as kernel density estimators.

3. Simulation study

We carried out a simulation study to examine the small sample properties. The result is presented in Tables 1-3 below. In this simulation study ϵ_1 follows a standard symmetric α-stable distribution, $S_\alpha(1, 0, 0)$. We estimate $f(x_0)$ by using the Epanechnikov kernel.

We took

$$\alpha = 1.2, \ 1.5, \ 1.8, \quad \beta = 0.9, \ 1.3, \ 1.7, \ 2.1, \ \infty,$$

$$b_j = \begin{cases} c_0(j+1)^{-\beta}, & 0 \le j \le 999 \\ \\ 0, & 1000 \le j \end{cases}.$$

We mean i.i.d. observations by $\beta = \infty$ and c_0 is chosen so that X_1 also follows $S_\alpha(1,0,0)$.

We tried $h = 0.2, \ 0.3, \ 0.4$ to see the effect of bandwidths.

We conducted the simulation study by using R2.3.1 and the fBasics package. The sample size is 200 and each entry of Tables 1-3 are based on 2,000 repetitions. In Tables 1-3, mean, var, and mse stand for the sample means, the sample variances, and the sample mean squared errors of the repetitions, respectively. The values of β are on the left margins of the Tables 1-3. The true values of $f(x_0)$ are as follows:

$$\alpha = 1.2 : f(0.0) = 0.2994, \ f(0.75) = 0.2130, \ f(1.5) = 0.1097.$$

$$\alpha = 1.5 : f(0.0) = 0.2874, \ f(0.75) = 0.2264, \ f(1.5) = 0.1287.$$

$$\alpha = 1.8 : f(0.0) = 0.2831, \ f(0.75) = 0.2381, \ f(1.5) = 0.1478.$$

Tables 1-3 are around here.

Theorem 2.1 tells that the effect of α and β appears in the asymptotic properties in the cases of $(\alpha, \beta) = (1.2, 0.9), \ (1.5, 0.9), \ (1.8, 0.9)$. Theorem 2.2 tells that the effect appears in the asymptotic properties in the case of $(\alpha, \beta) = (1.2, 1.3)$.

We obtained the following implications from Tables 1-3.

(i) The variance is more serious than the bias in each pair of (α, β). Thus larger bandwidths will be better.

(ii) The effect of α and β is seen in the cases of $(\alpha, \beta) = (1.2, 0.9)$, $(1.2, 1.3)$, $(1.5, 0.9)$, $(1.8, 0.9)$. This is conformable with Theorems 2.1-2. Especially the effect is remarkable in the case of $(\alpha, \beta) = (1.2, 0.9)$. Even when the effect is seen, larger bandwidths seem to perform better contrary to Theorems 2.1-2.

(iii) The effect of α and β rapidly disappears as β becomes larger.

4. Proofs of theorems

We prove Theorems 2.1-3 in this section. The proofs of all the lemmas are postponed to section 5.

We begin with Lemma 4.1 which deals with S_a in (2.2) and (2.3). We reproduce S_a here for reference.

$$S_a = \frac{1}{nh} \sum_{i=1}^{n} \left[K\left(\frac{X_i - x_0}{h}\right) - \mathrm{E}\left\{ K\left(\frac{X_i - x_0}{h}\right) \Big| \mathcal{S}_{i-s_0} \right\} \right]$$
$$= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{h} K\left(\frac{X_i - x_0}{h}\right) - \int K(\xi) f_{s_0}(x_0 + \xi h - \tilde{X}_{i,s_0}) d\xi \right]$$

Remember $s_0 = s_1 + s_2$ and that s_1 is fixed in **P1-3**. In the proofs of the theorems, we take a large s_2 and temporarily fix it. Then we let n tend to ∞. Thus we can take $n = k s_0$ for simplicity of presentation without affecting the asymptotic properties. Since the summands in S_a do not form martingale differences, we cannot apply the martingale central limit theorem directly and we need Lemma 4.1.

We further decompose S_a into four components.

$$(4.1) \qquad S_a = \sum_{l=1}^{k} N_{1l} + \sum_{l=1}^{k} N_{2l} + \sum_{l=1}^{k} N_{3l} + \sum_{l=1}^{k} N_{4l},$$

where

$$(4.2) \quad N_{1l} = \frac{1}{nh} \sum_{i=1+(l-1)s_0}^{s_1+(l-1)s_0} \left[K\left(\frac{X_i - x_0}{h}\right) - \mathrm{E}\left\{ K\left(\frac{X_i - x_0}{h}\right) \Big| \mathcal{S}_{1+(l-1)s_0 - s_1} \right\} \right],$$

$$(4.3) \quad N_{2l} = \frac{1}{nh} \sum_{i=1+(l-1)s_0}^{s_1+(l-1)s_0} \left[E\left\{ K\left(\frac{X_i - x_0}{h}\right) \big| \mathcal{S}_{1+(l-1)s_0-s_1} \right\} - E\left\{ K\left(\frac{X_i - x_0}{h}\right) \big| \mathcal{S}_{i-s_0} \right\} \right],$$

$$(4.4) \quad N_{3l} = \frac{1}{nh} \sum_{i=s_1+1+(l-1)s_0}^{ls_0} \left[K\left(\frac{X_i - x_0}{h}\right) - E\left\{ K\left(\frac{X_i - x_0}{h}\right) \big| \mathcal{S}_{1+(l-1)s_0} \right\} \right],$$

$$(4.5) \quad N_{4l} = \frac{1}{nh} \sum_{i=s_1+1+(l-1)s_0}^{ls_0} \left[E\left\{ K\left(\frac{X_i - x_0}{h}\right) \big| \mathcal{S}_{1+(l-1)s_0} \right\} - E\left\{ K\left(\frac{X_i - x_0}{h}\right) \big| \mathcal{S}_{i-s_0} \right\} \right].$$

N_{1l}, N_{2l}, N_{3l}, and N_{4l} are $\mathcal{S}_{s_1+(l-1)s_0}$-, $\mathcal{S}_{1+(l-1)s_0}$-, $\mathcal{S}_{1+ls_0}$-, and $\mathcal{S}_{1-s_2+ls_0}$ -measurable, respectively. In addition,

$$(4.6) \qquad E\{N_{1l}|\mathcal{S}_{s_1+(l-2)s_0}\} = E\{N_{2l}|\mathcal{S}_{1+(l-2)s_0}\}$$

$$= E\{N_{3l}|\mathcal{S}_{1+(l-1)s_0}\} = E\{N_{4l}|\mathcal{S}_{1-s_2+(l-1)s_0}\} = 0.$$

LEMMA 4.1. *Suppose that* **C1** *and* **C2** *hold. Then we have*

$$(4.7) \qquad \mathrm{Var}\left(\sum_{l=1}^{k} N_{2l}\right) = O\left(\frac{1}{n}\right), \quad \mathrm{Var}\left(\sum_{l=1}^{k} N_{4l}\right) = O\left(\frac{1}{n}\right),$$

$$(4.8) \qquad \sqrt{nh} \sum_{l=1}^{k} N_{1l} \overset{d}{\to} \mathrm{N}\left(0, \frac{s_1}{s_0}\kappa f(x_0)\right), \quad \sqrt{nh} \sum_{l=1}^{k} N_{3l} \overset{d}{\to} \mathrm{N}\left(0, \frac{s_2}{s_0}\kappa f(x_0)\right).$$

Remark 1. Take an arbitrary positive integer m. Then the proof of Lemma 4.1 in section 5 and standard arguments imply that $\sqrt{nh} \sum_{l=1}^{k} N_{3l}$ for $x_{01}, \ldots, x_{0m}$ are asymptotically mutually independent if $x_{0k} \neq x_{0l}(k \neq l)$.

We go on to the proofs of Theorems 2.1-3. We investigate the asymptotic properties of S_b in (2.4) for Cases 1-3 in Propositions 4.1-3, respectively. Then by combining Lemma 4.1 and Propositions 4.1-3, we derive the asymptotic distributions of $\hat{f}(x_0) - E\{\hat{f}(x_0)\}$.

Case 1: The proof of Proposition 4.1 below is based on the arguments in Koul and Surgailis (2001). Especially the proof of Lemma 4.2 is a modified and simplified argument of those of Koul and Surgailis (2001).

PROPOSITION 4.1. *Suppose that* **C1** *and* **C2** *hold and that* $1 < \alpha < 2$ *and* $1/\alpha < \beta < 1$. *Then we have*

$$n^{\beta-1/\alpha} S_b \xrightarrow{d} -f'(x_0)c_A Z.$$

Remark 2. The proof of Proposition 4.1 after that of Theorem 2.1 implies that Z in Proposition 4.1 comes from the sample mean of $X_1, \ldots, X_n$. Therefore two $n^{\beta-1/\alpha} S_b$ for any pair of (x_{01}, x_{02}) are asymptotically degenerate.

We prove Theorem 2.1.

PROOF OF THEOREM 2.1. First let $1/\alpha - \beta + 2/5$ be smaller than 0. Then Proposition 4.1 implies that $\sqrt{nh} S_b = o_p(1)$. By taking a sufficiently large s_2 in Lemma 4.1, we can make s_2/s_0 and s_1/s_0 arbitrarily close to 1 and 0, respectively. These yield the convergence in distribution of

$$\sqrt{nh}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} \mathrm{N}(0, \kappa f(x_0)).$$

Next let $1/\alpha - \beta + 2/5$ be larger than 0. By Lemma 4.1, we have $n^{\beta-1/\alpha} S_a = o_p(1)$. Thus the convergence in distribution of the latter follows from Proposition 4.1. Hence the proof of the theorem is complete.

PROOF OF PROPOSITION 4.1. We adopt the notations of Koul and Surgailis (2001) and prove the proposition by following the arguments there.

Provided that

$$(4.9) \qquad n^{\beta-1/\alpha}\Big(S_b + \frac{1}{n}\int K(\xi)f'(x_0 + \xi h)d\xi \sum_{j=-\infty}^{n-s_0} \sum_{i=1\vee(j+s_0)}^{n} b_{i-j}\epsilon_j\Big) = o_p(1),$$

Proposition 4.1 follows from (1.9) of Koul and Surgailis (2001),

$$n^{\beta-1/\alpha-1} \sum_{i=1}^{n} \sum_{j=s_0}^{\infty} b_j \epsilon_{i-j} \xrightarrow{d} c_A Z.$$

We establish (4.9). As in Koul and Surgailis (2001), we represent S_b as

(4.10)
$$S_b = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j=s_0}^{\infty} (\mathrm{E}\{K_i | \mathcal{S}_{i-j}\} - \mathrm{E}\{K_i | \mathcal{S}_{i-j-1}\})$$
$$= \frac{1}{nh} \sum_{j=-\infty}^{n-s_0} \sum_{i=1 \vee (j+s_0)}^{n} (\mathrm{E}\{K_i | \mathcal{S}_j\} - \mathrm{E}\{K_i | \mathcal{S}_{j-1}\}).$$

Then left hand side of (4.9) is rewritten as

(4.11)
$$\int K(\xi) [n^{\beta - 1/\alpha - 1} \sum_{j=-\infty}^{n-s_0} \sum_{i=1 \vee (j+s_0)}^{n} U_{i,i-j}(\xi)] d\xi,$$

where

(4.12) $\quad U_{i,j}(\xi) = f_j(x_0 + \xi h - b_j \epsilon_{i-j} - \tilde{X}_{i,j+1})$

$$- \int f_j(x_0 + \xi h - b_j u - \tilde{X}_{i,j+1}) dG(u) + f'(x_0 + \xi h) b_j \epsilon_{i-j}.$$

We define $R_n(\xi)$ by

(4.13)
$$R_n(\xi) = n^{\beta - 1/\alpha - 1} \sum_{j=-\infty}^{n-s_0} \sum_{i=1 \vee (j+s_0)}^{n} U_{i,i-j}(\xi),$$

then by Jensen's inequality, we obtain for any $r \geq 1$,

(4.14)
$$\mathrm{E}\left\{ \left| \int K(\xi) R_n(\xi) d\xi \right|^r \right\} \leq \int K(\xi) \mathrm{E}\{|R_n(\xi)|^r\} d\xi.$$

We evaluate $\mathrm{E}\{|R_n(\xi)|^r\}$ in Lemma 4.2 below.

LEMMA 4.2. *For any positive M and r such that $1 < r = \alpha/(1 + \eta)$ and $0 < \eta < (\alpha\beta - 1)/2$, we have*

$$\lim_{n \to \infty} \sup_{|\xi| < M} \mathrm{E}\{|R_n(\xi)|^r\} = 0.$$

(4.9) follows from (4.11), (4.13), (4.14), and Lemma 4.2. Hence the proof of the proposition is complete.

Case 2: Theorem 2.2 follows from Lemma 4.1 and Proposition 4.2 below. The proof of Proposition 4.2 below is based on the arguments in Surgailis (2002). We adopt the

17

notations. Especially the proof of Lemma 4.4 is a modified and simplified argument of those in Surgailis (2002). We adopt the notations.

PROPOSITION 4.2. *Suppose that* **C1** *and* **C2** *hold and that* $1 < \alpha < 2$ *and* $1 < \beta < 2/\alpha$. *Then we have*

$$n^{1-1/(\alpha\beta)} S_b \xrightarrow{d} \Big(\frac{c_1 c_0^\alpha}{\sigma_{\alpha\beta} \beta^{\alpha\beta}} \Big)^{1/(\alpha\beta)} (c_f^+ L_+ + c_f^- L_-).$$

Remark 3. The proof of Lemma 3.1 of Surgailis (2002) implies that L^+ and L^- come from

$$n^{-1/(\alpha\beta)} \sum_{i=1}^n ((\epsilon_i \vee 0)^{1/\beta} - \mathrm{E}\{(\epsilon_i \vee 0)^{1/\beta}\}) \quad and$$

$$n^{-1/(\alpha\beta)} \sum_{i=1}^n (|\epsilon_i \wedge 0|^{1/\beta} - \mathrm{E}\{|\epsilon_i \wedge 0|^{1/\beta}\}),$$

respectively. As in Theorem 2.1 of Surgailis (2002), L^+ and L^- are common to every x_0. It also follows from the proof of Lemma 3.1 of Surgailis (2002) that the result of Proposition 4.2 does not depend on s_0.

The proof of Proposition 4.2 is given after Theorem 2.2 is proved.

PROOF OF THEOREM 2.2. First let $1/(\alpha\beta)$ be smaller than $3/5$. Then Proposition 4.2 implies that $\sqrt{nh} S_b = o_p(1)$. By taking a sufficiently large s_2 in Lemma 4.1, we can make s_2/s_0 and s_1/s_0 arbitrarily close to 1 and 0, respectively. These yield the convergence in distribution of

$$\sqrt{nh}(\hat{f}(x_0) - \mathrm{E}\{\hat{f}(x_0)\}) \xrightarrow{d} \mathrm{N}(0, \kappa f(x_0)).$$

Next let $1/(\alpha\beta)$ be larger than $3/5$. By Lemma 4.1, we have $n^{1-1/(\alpha\beta)} S_a = o_p(1)$. Thus the convergence in distribution of the latter follows from Proposition 4.2. Hence the proof of the theorem is complete.

18

PROOF OF PROPOSITION 4.2. We begin the proof with several definitions.

$$H_{n,\infty}(t) = \int K(\xi)(f(x_0 + \xi h - t) - f(x_0 + \xi h))d\xi,$$

$$\mathcal{T}_n = \sum_{i=1}^{n}\sum_{j=s_0}^{\infty}(H_{n,\infty}(b_j\epsilon_i) - \mathrm{E}\{H_{n,\infty}(b_j\epsilon_i)\}),$$

$$T_n = \sum_{i=1}^{n}\sum_{j=s_0}^{\infty}(H_{n,\infty}(b_j\epsilon_{i-j}) - \mathrm{E}\{H_{n,\infty}(b_j\epsilon_{i-j})\}).$$

We prove the following two lemmas in section 5. The argument in p.265 of Surgailis (2002), which corresponds to Lemma 4.3 here and deals with a kind of uniformity, is rather difficult to understand. If another approach based on Lemma 4.3 is possible, the argument in Surgailis (2002) may become easier to understand.

LEMMA 4.3.

$$T_n - \mathcal{T}_n = o_p(n^{1/(\alpha\beta)}).$$

LEMMA 4.4.

$$nS_b - T_n = o_p(n^{1/(\alpha\beta)}).$$

We consider $\mathcal{T}_n$ since Lemmas 4.3-4 imply that $nS_b = \mathcal{T}_n + o_p(n^{1/(\alpha\beta)})$. By the Taylor series expansion,

$$(4.15) \quad H_{n,\infty}(b_j\epsilon_i) = f(x_0 - b_j\epsilon_i) - f(x_0)$$

$$+h^2\int K(\xi)(\int_0^{\xi}(\xi - \eta)(f''(x_0 + \eta h - b_j\epsilon_i) - f''(x_0 + \eta h))d\eta)d\xi$$

$$= f(x_0 - b_j\epsilon_i) - f(x_0) + h^2\bar{H}_{n,\infty}(b_j\epsilon_i),$$

where $\bar{H}_{n,\infty}(u)$ is clearly defined.

$\bar{H}_{n,\infty}(u)$ is a continuously differentiable bounded function and the derivative is also bounded. In addition,

$$\lim_{n\to\infty}\bar{H}_{n,\infty}(u) = \frac{1}{2}\int \xi^2 K(\xi)d\xi(f''(x_0 - u) - f''(x_0)).$$

19

By applying the arguments in Lemma 3.1 of Surgailis (2002) to $\bar{H}_{n,\infty}(u)$, we can show that for any $r < \alpha\beta$,

$$\mathrm{E}\{|\sum_{j=s_0}^{\infty} \bar{H}_{n,\infty}(b_j\epsilon_i)|^r\} < C_r,$$

where C_r depends on r. This means that $h^2\bar{H}_{n,\infty}(b_j\epsilon_i)$ in (4.15) is negligible in $n^{-1/(\alpha\beta)}\mathcal{T}_n$ and the result of Proposition 4.2 follows from Lemma 3.1 of Surgailis (2002). Hence the proof of Proposition 4.2 is complete.

Case 3: The proof of Proposition 4.3 below is a modified and simplified argument of part of the arguments in Hsing (1999) and Pipiras and Taqqu (2003). We adopt the notations of the papers. Since $S_b = O_p(n^{-1/2})$ from Proposition 4.3 below and we can treat S_a in the same way as in Theorems 2.1-2, it is easy to prove the result of Theorem 2.3. We omit the proof of Theorem 2.3.

PROPOSITION 4.3. *Suppose that* **C1** *and* **C2** *hold and that* $0 < \alpha < 2$ *and* $2/\alpha < \beta$. *Then we have*

$$n^{1/2}S_b = O_p(1).$$

PROOF. What we have to prove is

(4.16) $$n\mathrm{E}\{S_b^2\} = O(1).$$

Let ϵ be a random variable which is distributed as ϵ_1 and independent of $\{\epsilon_j\}_{j=-\infty}^{\infty}$. Then we have

(4.17) $\mathrm{P}(|b_j\epsilon_i - b_j\epsilon| \geq 1) \leq C|b_j|^{\alpha}$ and $\mathrm{E}\{|b_j\epsilon_i - b_j\epsilon|^2 I(|b_j\epsilon_i - b_j\epsilon| < 1)\} \leq C|b_j|^{\alpha}$.

See (3.35) and (3.36) of Pipiras and Taqqu (2003) about (4.17).

S_b is written as

(4.18) $$S_b = \frac{1}{n}\sum_{i=1}^{n} H_n(\tilde{X}_{i,s_0}),$$

where

$$H_n(\zeta) = \int K(\xi)(f_{s_0}(x_0 + \xi h - \zeta) - E\{f(x_0 + \xi h - X_1)\})d\xi.$$

As in the proofs of Propositions 4.1-2, we define $U_{i,j}(\xi)$. In Case 3, it is defined by

$$U_{i,j}(\xi)$$

$$(4.19) \quad = f_j(x_0 + \xi h - b_j\epsilon_{i-j} - \tilde{X}_{i,j+1}) - E\{f_j(x_0 + \xi h - b_j\epsilon_{i-j} - \tilde{X}_{i,j+1})|\mathcal{S}_{i-j-1}\}$$

$$= \int (f_j(x_0 + \xi h - b_j\epsilon_{i-j} - \tilde{X}_{i,j+1}) - f_j(x_0 + \xi h - b_j\epsilon - \tilde{X}_{i,j+1}))dG(\epsilon).$$

As in the proofs of Propositions 4.1-2, $U_{i,j}(\xi)$ is $\mathcal{S}_{i-j}$-measurable and $E\{U_{i,j}(\xi)|\mathcal{S}_{i-j-1}\} = 0$. In addition, $H_n(\tilde{X}_{i,s_0})$ is written as

$$(4.20) \qquad\qquad H_n(\tilde{X}_{i,s_0}) = \int K(\xi) \sum_{j=s_0}^{\infty} U_{i,j}(\xi)d\xi.$$

By (4.18) and (4.20), we have

$$(4.21) \qquad\qquad nS_b = \sum_{i=1}^{n} H_n(\tilde{X}_{i,s_0}) = \int K(\xi) \sum_{i=1}^{n}\sum_{j=s_0}^{\infty} U_{i,j}(\xi)d\xi.$$

The properties of $U_{i,j}(\xi)$, (4.21), Jensen's inequality, and the Cauchy-Schwarz inequality imply that

$$(4.22) \qquad E\{(nS_b)^2\} \le \int K(\xi) \sum_{i=1}^{n}\sum_{j=s_0}^{\infty}\sum_{j'=s_0}^{\infty} (E\{U_{i,j}^2(\xi)\})^{1/2}(E\{U_{i',j'}^2(\xi)\})^{1/2}d\xi,$$

where $i' = i - j + j'$.

Provided that for any positive M,

$$(4.23) \qquad\qquad \sup_{|\xi| \le M} E\{U_{i,j}^2(\xi)\} \le C|b_j|^\alpha,$$

(4.16) follows from (4.22).

The integrand in (4.19) is bounded by $C(1 \wedge (|(\epsilon_{i-j} - \epsilon)b_j|))$. Therefore by (4.17), we get

$$E\{U_{i,j}^2(\xi)\} \le C(E\{|(\epsilon_{i-j} - \epsilon)b_j|^2 I(|(\epsilon_{i-j} - \epsilon)b_j| < 1)\} + P(|(\epsilon_{i-j} - \epsilon)b_j| \ge 1))$$

$$\le C|b_j|^\alpha.$$

Hence (4.23) is established and the proof of the proposition is complete.

5. Proofs of technical lemmas

In this section we prove technical lemmas. The proofs of Lemmas 4.2 and 4.4 are modified and simplified arguments of those in Koul and Surgailis (2001) and Surgailis (2002), respectively.

PROOF OF LEMMA 4.1. We write K_i for $K((X_i - x_0)/h)$ in the proof for notational simplicity.

Note that for any $s \geq s_1$,

$$(5.1) \qquad \frac{1}{h}E\{K_i|\mathcal{S}_{i-s}\} = \int K(\xi) f_s(x_0 + \xi h - \tilde{X}_{i,s}) d\xi$$

and that the above expression is a bounded continuous function of $\tilde{X}_{i,s}$. Hence (4.7) follows from (4.6) by appealing to the properties of martingale differences.

We prove the latter of (4.8) by using (4.6) and applying the martingale central limit theorem (e.g. Theorem 9.5.2 of Chow and Teicher (1988)). The former can be treated in the same way and the proof is omitted.

Since $|\sqrt{nh}N_{3l}| \leq C/\sqrt{nh}$, we have only to show that

$$(5.2) \qquad nh \sum_{l=1}^{k} E\{N_{3l}^2|\mathcal{S}_{1+(l-1)s_0}\} \xrightarrow{p} \frac{s_2}{s_0}\kappa f(x_0).$$

(4.6) and **P2** imply that

$$(5.3) \quad nh \sum_{l=1}^{k} E\{N_{3l}^2|\mathcal{S}_{1+(l-1)s_0}\}$$

$$= \sum_{l=1}^{k} \sum_{i=s_1+1+(l-1)s_0}^{ls_0} \frac{1}{nh} E\{K_i^2|\mathcal{S}_{1+(l-1)s_0}\}$$

$$+ 2 \sum_{l=1}^{k} \sum_{i_1=s_1+1+(l-1)s_0}^{ls_0} \sum_{i_2=i_1+1}^{ls_0} \frac{1}{nh} E\Big\{K_{i_1}K_{i_2}|\mathcal{S}_{1+(l-1)s_0}\Big\}$$

$$-\sum_{l=1}^{k}\sum_{i_1=s_1+1+(l-1)s_0}^{ls_0}\sum_{i_2=s_1+1+(l-1)s_0}^{ls_0}\frac{1}{nh}\mathrm{E}\big\{K_{i_1}|\mathcal{S}_{1+(l-1)s_0}\big\}\mathrm{E}\big\{K_{i_2}|\mathcal{S}_{1+(l-1)s_0}\big\}$$

$$=\sum_{l=1}^{k}\sum_{i=s_1+1+(l-1)s_0}^{ls_0}\frac{1}{nh}\mathrm{E}\{K_i^2|\mathcal{S}_{1+(l-1)s_0}\}+O_p(h).$$

By **P1** and the ergodic thereom, we obtain

$$(5.4)\qquad \sum_{l=1}^{k}\sum_{i=s_1+1+(l-1)s_0}^{ls_0}\frac{1}{nh}\mathrm{E}\{K_i^2|\mathcal{S}_{1+(l-1)s_0}\}$$

$$=\frac{1}{n}\sum_{l=1}^{k}\sum_{i=s_1+1+(l-1)s_0}^{ls_0}\int K^2(\xi)f_{i-1-(l-1)s_0}(x_0-\tilde{X}_{i,i-1-(l-1)s_0})d\xi+O_p(h)$$

$$\xrightarrow{p}\frac{s_2}{s_0}\kappa f(x_0)$$

since $\mathrm{E}\{f_s(x_0-\tilde{X}_{i,i-s})\}=f(x_0)$. (5.2) follows from (5.3) and (5.4).

PROOF OF LEMMA 4.2. We adopt the notations of Koul and Surgailis (2001), for example, $U_{i,j}^{(k)}$, $M_{i,j}^{(k)}$, $D_{i,j}^{(k)}$, $\chi_{i,j}^{(k)}$, and $W_{i,j}^{(k)}$. Note that we can do without (2.1) there. We fix a positive γ such that $(1/r-1/\alpha)/\beta<\gamma<((\alpha-r)/r)\wedge(1-1/(r\beta))$. Then $r(1+\gamma)<\alpha$.

$\{M_{j,n}(\xi)\}$ defined in (5.5) form a martingale difference sequence with respect to $\{\mathcal{S}_j\}$.

$$(5.5)\qquad\qquad\qquad M_{j,n}(\xi)=\sum_{i=1\vee(j+s_0)}U_{i,i-j}(\xi).$$

Then by the von Bahr and Esseen inequality, we have

$$(5.6)\qquad\qquad\qquad \mathrm{E}\{|R_n(\xi)|^r\}\leq 2\sum_{j=-\infty}^{n-s_0}\mathrm{E}\{|M_{j,n}(\xi)|^r\}.$$

To evaluate $\mathrm{E}\{|M_{j,n}(\xi)|^r\}$, we decompose $U_{i,j}(\xi)$ into three components, $U_{i,j}^{(1)}$, $U_{i,j}^{(2)}$, and $U_{i,j}^{(3)}$. Hereafter we suppress the dependence on ξ for notational convenience.

$$(5.7)\quad U_{i,j}^{(1)}=f_j(x_0+\xi h-b_j\epsilon_{i-j}-\tilde{X}_{i,j+1})-\int f_j(x_0+\xi h-b_j u-\tilde{X}_{i,j+1})dG(u)$$

$$+f_j'(x_0+\xi h-\tilde{X}_{i,j+1})b_j\epsilon_{i-j}$$

$$(5.8)\quad U_{i,j}^{(2)}=b_j\epsilon_{i-j}(f'(x_0+\xi h)-f'(x_0+\xi h-\tilde{X}_{i,j+1}))$$

$$(5.9)\quad U_{i,j}^{(3)}=b_j\epsilon_{i-j}(f'(x_0+\xi h-\tilde{X}_{i,j+1})-f_j'(x_0+\xi h-\tilde{X}_{i,j+1}))$$

Then

$$M_{j,n} = \sum_{k=1}^{3} M_{j,n}^{(k)}, \quad \text{where } M_{j,n}^{(k)} = \sum_{i=1\vee(j+s_0)}^{n} U_{i,i-j}^{(k)}.$$

Provided that

(5.10)
$$\mathrm{E}\{|M_{j,n}(\xi)|^r\} \le C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-\beta(1+\gamma)})^r,$$

(5.6), (5.10) and some calculation as on p.321 of Koul and Surgailis (2001) imply the result of Lemma 3.1. We establish (5.10).

We give some definitions before we consider $M_{j,n}^{(1)}$ and $U_{i,j}^{(1)}$.

$$\chi_{i,j}^{(1)} = I(|b_j v| \le 1, \ |b_j \epsilon_{i-j}| \le 1), \ \chi_{i,j}^{(2)} = I(|b_j v| > 1, \ |b_j \epsilon_{i-j}| \le 1), \ \chi_{i,j}^{(3)} = I(|b_j \epsilon_{i-j}| > 1).$$

By using them, we have

(5.11)
$$\mathrm{E}\{|M_{j,n}^{(1)}|^r\} \le C \sum_{k=1}^{3} D_{j,n}^{(k)},$$

where

$$D_{j,n}^{(k)} = \mathrm{E}\{|\sum_{i=1\vee(j+s_0)}^{n} \chi_{i,i-j}^{(k)} U_{i,i-j}^{(1)}|^r\}.$$

We show that all of $D_{j,n}^{(k)}$ are bounded by $C(\sum_{i=1\vee(j+s_0)}^{n})(i-j)^{-\beta(1+\gamma)})^r$. We deal with $D_{j,n}^{(1)}$ and $D_{j,n}^{(3)}$. $D_{j,n}^{(2)}$ can be treated in the same way as $D_{j,n}^{(3)}$.

We can represent $U_{i,j}^{(1)}$ as

(5.12)
$$U_{i,j}^{(1)} = W_{i,j}^{(1)} - W_{i,j}^{(2)},$$

where

$$W_{i,j}^{(1)} = \int(\int_{-b_j u}^{-b_j \epsilon_{i-j}} f_j'(x_0 + v + \xi h - \tilde{X}_{i,j+1})dv)dG(u),$$
$$W_{i,j}^{(2)} = \int(\int_{-b_j u}^{-b_j \epsilon_{i-j}} f_j'(x_0 + \xi h - \tilde{X}_{i,j+1})dv)dG(u).$$

Since

$$|f_j'(x_0 + v + \xi h - \tilde{X}_{i,j+1}) - f_j'(x_0 + \xi h - \tilde{X}_{i,j+1})| \le C|v|^\gamma \quad \text{for } |v| \le 1,$$

we have

$$(5.13) \qquad |U_{i,j}^{(1)}\chi_{i,j}^{(1)}| \leq C\int(\int_{-b_ju}^{-b_j\epsilon_{i-j}}|v|^\gamma dv)dG(u) \leq C|b_j|^{1+\gamma}(1+|\epsilon_{i-j}|^{1+\gamma}).$$

By (5.12), (5.13) and Minkowski's inequality, we get

$$(5.14) \qquad \begin{aligned} D_{j,n}^{(1)} &\leq C(\sum_{i=1\vee(j+s_0)}^{n}|b_{i-j}|^{1+\gamma}(\mathrm{E}\{1+|\epsilon_j|^{r(1+\gamma)}\})^{1/r})^r \\ &\leq C(\sum_{i=1\vee(j+s_0)}^{n}(i-j)^{-\beta(1+\gamma)})^r. \end{aligned}$$

Next we deal with $D_{j,n}^{(3)}$. By exploiting (5.12), we have

$$(5.15) \qquad D_{j,n}^{(3)} \leq C\sum_{l=1}^{2}\sum_{k=1}^{2}\mathrm{E}\{|\sum_{i=1\vee(j+s_0)}^{n}W_{i,i-j}^{(l,k)}I(|b_{i-j}\epsilon_j|>1)|^r\},$$

where

$$W_{i,j}^{(1,1)} = \int(\int_{-b_ju}^{-b_j\epsilon_{i-j}}f_j'(x_0+v+\xi h-\tilde{X}_{i,j+1})dv)I(|b_j\epsilon_{i-j}|>1)I(|\epsilon_{i-j}|>|u|)dG(u),$$

$$W_{i,j}^{(1,2)} = \int(\int_{-b_ju}^{-b_j\epsilon_{i-j}}f_j'(x_0+v+\xi h-\tilde{X}_{i,j+1})dv)I(|b_j\epsilon_{i-j}|>1)I(|\epsilon_{i-j}|\leq|u|)dG(u),$$

$$W_{i,j}^{(2,1)} = \int(\int_{-b_ju}^{-b_j\epsilon_{i-j}}f_j'(x_0+\xi h-\tilde{X}_{i,j+1})dv)I(|b_j\epsilon_{i-j}|>1)I(|\epsilon_{i-j}|>|u|)dG(u),$$

$$W_{i,j}^{(2,2)} = \int(\int_{-b_ju}^{-b_j\epsilon_{i-j}}f_j'(x_0+\xi h-\tilde{X}_{i,j+1})dv)I(|b_j\epsilon_{i-j}|>1)I(|\epsilon_{i-j}|\leq|u|)dG(u).$$

Noticing that

$$|W_{i,j}^{(1,1)}| = C|b_j\epsilon_{i-j}|^{1+\gamma}I(|b_j\epsilon_{i-j}|>1),$$

$$|W_{i,j}^{(1,2)}| = C\int|b_ju|^{1+\gamma}I(|b_j\epsilon_{i-j}|>1)dG(u) \leq C|b_j|^{1+\gamma},$$

$$|W_{i,j}^{(2,1)}| = C|b_j\epsilon_{i-j}|^{1+\gamma}I(|b_j\epsilon_{i-j}|>1),$$

$$|W_{i,j}^{(2,2)}| = C\int|b_ju|^{1+\gamma}I(|b_j\epsilon_{i-j}|>1)dG(u) \leq C|b_j|^{1+\gamma},$$

we get

$$(5.16) \qquad D_{j,n}^{(3)} \leq C(\sum_{i=1\vee(j+s_0)}^{n}|b_{i-j}|^{1+\gamma})^r \leq C(\sum_{i=1\vee(j+s_0)}^{n}(i-j)^{-\beta(1+\gamma)})^r$$

25

Combining (5.11), (5.14), and (5.16), we have

$$(5.17) \qquad \mathrm{E}\{|M_{j,n}^{(1)}|^r\} \leq C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-\beta(1+\gamma)})^r.$$

We deal with $M_{j,n}^{(2)}$.

(5.18) $\mathrm{E}\{|M_{j,n}^{(2)}|^r\}$

$$\leq \mathrm{E}\{|\epsilon_j|^r\}\mathrm{E}\{|\sum_{i=1\vee(j+s_0)}^{n} b_{i-j}(f'(x_0+\xi h) - f'(x_0+\xi h - \tilde{X}_{i,i-j+1}))|^r\}$$

$$\leq C\mathrm{E}\{(\sum_{i=1\vee(j+s_0)}^{n} |b_{i-j}\tilde{X}_{i,i-j+1}|)^r\} \leq C\{\sum_{i=1\vee(j+s_0)}^{n} |b_{i-j}|(\mathrm{E}\{|\tilde{X}_{i,i-j+1}|^r\})^{1/r}\}^r$$

$$\leq C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-2\beta+1/r})^r \leq C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-\beta(1+\gamma)})^r.$$

See the definition of r and γ about the last line of (5.18).

Finally, by using **P3**, we have

$$(5.19) \qquad \mathrm{E}\{|M_{j,n}^{(3)}|^r\}$$

$$\leq C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-\beta}(i-j)^{-\beta+1/r})^r \leq C(\sum_{i=1\vee(j+s_0)}^{n} (i-j)^{-\beta(1+\gamma)})^r.$$

Hence (5.10) is proved for every k and the proof of the lemma is complete.

PROOF OF LEMMA 4.3. We prove Lemma 4.3 by using the results given in Surgailis (2002). However, we do not deal with the uniformity and we can do without (2.3) there in the proofs of Lemmas 4.3-4.

We write V_n for the difference of T_n and $\mathcal{T}_n$ and represent V_n as

$$(5.20) \qquad V_n = T_n - \mathcal{T}_n = -\sum_{k=1}^{n} A_n(k) + \sum_{k=1}^{\infty} B_n(k),$$

where

$$A_n(k) = \sum_{j=k}^{\infty} (H_{n,\infty}(b_j\epsilon_{n+1-k}) - \mathrm{E}\{H_{n,\infty}(b_j\epsilon_{n+1-k})\}),$$

$$B_n(k) = \sum_{j=k}^{k+n-1} (H_{n,\infty}(b_j\epsilon_{1-k}) - \mathrm{E}\{H_{n,\infty}(b_j\epsilon_{1-k})\}).$$

We choose two positive numbers, r and γ, such that $1 + r(1 - \beta) < r/(\alpha\beta)$, $1 + r - \beta(\alpha - \gamma) < r/(\alpha\beta)$, and $1 < r < \alpha$. It is not difficult to check the existence of r and γ. Then by the von Bahr and Esseen inequality, we obtain

$$
(5.21) \qquad \mathrm{E}\{|V_n|^r\} \leq 2(\sum_{k=1}^{n} \mathrm{E}\{|A_n(k)|^r\} + \sum_{k=1}^{\infty} \mathrm{E}\{|B_n(k)|^r\}).
$$

Provided that

$$
(5.22) \qquad \sum_{k=1}^{n} \mathrm{E}\{|A_n(k)|^r\} = o(n^{r/(\alpha\beta)}) \quad \text{and} \quad \sum_{k=1}^{\infty} \mathrm{E}\{|B_n(k)|^r\} = o(n^{r/(\alpha\beta)}),
$$

the result of the lemma follows from (5.21). We prove (5.22).

First we deal with $A_n(k)$. Since

$$
H_{n,\infty}(0) = 0, \quad |H_{n,\infty}(x)| \leq C, \quad \text{and} \quad |H'_{n,\infty}(x)| \leq C,
$$

we have

$$
|H_{n,\infty}(b_j \epsilon_{n+1-k})| \leq C(1 \wedge (|\epsilon_{n+1-k}| j^{-\beta})).
$$

If

$$
(5.23) \qquad \mathrm{E}\{(\sum_{j=k}^{n}(1 \wedge (|\epsilon_{n+1-k}| j^{-\beta})))^r\} \leq C(k^{r(1-\beta)} + k^{r-\beta(\alpha-\gamma)}),
$$

we get

$$
(5.24) \qquad \mathrm{E}\{\sum_{k=1}^{n} |A_n(k)|^r\} \leq C(\sum_{k=1}^{n} k^{r(1-\beta)} + \sum_{k=1}^{n} k^{r-\beta(\alpha-\gamma)}) = o(n^{r/(\alpha\beta)}).
$$

See the definitions of r and γ. Note that $\mathrm{E}\{H_{n,\infty}(b_j \epsilon_{n+1-k})\}$ in the definition of $A_n(k)$ is treated by Jensen's inequality. We establish (5.23).

Some calculation as in Surgailis (2002) gives

$$
(5.25) \qquad (\sum_{j=k}^{n}(1 \wedge (|\epsilon_{n+1-k}| j^{-\beta})))^r
$$
$$
\leq C\{I(|\epsilon_{n+1-k}| \leq k^{\beta})|\epsilon_{n+1-k}| k^{1-\beta} + I(|\epsilon_{n+1-k}| > k^{\beta})|\epsilon_{n+1-k}|^{1/\beta}\}^r.
$$

By evaluating the expectation of (5.25), we obtain (5.23).

Next we deal with $B_n(k)$. As in the case of $A_n(k)$, what we should establish is

$$(5.26) \quad \sum_{k=1}^{Mn} \mathrm{E}\{(\sum_{j=k}^{n+k-1} (1 \wedge (|\epsilon_{1-k}|j^{-\beta})))^r\}$$

$$+ \sum_{k=Mn+1}^{\infty} \mathrm{E}\{(\sum_{j=k}^{n+k-1} (1 \wedge (|\epsilon_{1-k}|j^{-\beta})))^r\} = o(n^{r/(\alpha\beta)}),$$

where M is an arbitrary large integer.

The first term of the left hand side of (5.26) can be treated in the same way as in the case of $A_n(k)$. We consider only the second term of the left hand side of (5.26).

We closely examine the random variables in the sign of expectation. Then

$$\sum_{j=k}^{n+k-1} (1 \wedge (|\epsilon_{1-k}|j^{-\beta}))$$

$$(5.27) \qquad \leq C \begin{cases} (k^{1-\beta} - (k+n)^{1-\beta})|\epsilon_{1-k}|, & |\epsilon_{1-k}| \leq k^{\beta} \\[2mm] n + (k^{1-\beta} - (k+n)^{1-\beta})|\epsilon_{1-k}|, & k^{\beta} < |\epsilon_{1-k}| \leq (k+n)^{\beta} \\[2mm] n, & |\epsilon_{1-k}| > (k+n)^{\beta} \end{cases} .$$

We take the expectation by using (5.27) and get

$$(5.28) \quad \mathrm{E}\{(\sum_{j=k}^{n+k-1} (1 \wedge (|\epsilon_{1-k}|j^{-\beta})))^r\} \leq C\{(k^{1-\beta} - (k+n)^{1-\beta})^r + n^r \mathrm{P}(|\epsilon_{1-k}| > k^{\beta})$$

$$\leq C\{(k^{1-\beta} - (k+n)^{1-\beta})^r + n^r k^{-\alpha\beta}\}.$$

Then the second term of (5.26) is bounded by

$$C \sum_{k=Mn+1}^{\infty} \{(k^{1-\beta} - (k+n)^{1-\beta})^r + n^r k^{-\alpha\beta}\}$$

$$\leq C(n^{1+r(1-\beta)} \int_M^{\infty} (u^{1-\beta} - (1+u)^{1-\beta})^r du + n^{r-\alpha\beta+1} \int_M^{\infty} u^{-\alpha\beta} du) = o(n^{r/(\alpha\beta)}).$$

The last inequality follows from the definition of r. Hence the proof is complete.

PROOF OF LEMMA 4.4. We adopt the notations of Surgailis (2002), for example, $U_{i,j}^{(k)}$, $M_{n,j}^{(k)}$, and V_{nj}. We choose two positive numbers, λ and r, such that

$$1 \vee \frac{2}{2\beta - 1 + 1/(\alpha\beta)} < r < \alpha \quad \text{and} \quad \frac{\alpha\beta - r}{\alpha\beta(2\beta r - 1 - r)} < \lambda < 1 \wedge \frac{2 - \alpha\beta}{\alpha\beta(3 - 2\beta r)}.$$

28

The existence of λ and r is proved in Surgailis (2002). $[a]$ stands for the largest integer which is less than or equal to a.

We represent $nS_b - T_n$ as

(5.29) $$nS_b - T_n = V_{1n} + V_{2n},$$

where

$$V_{1n} = \sum_{i=1}^{n} \sum_{j=s_0}^{[n^\lambda]} \int K(\xi) U_{i,j}(\xi) d\xi,$$

$$V_{2n} = \sum_{i=1}^{n} \sum_{j=[n^\lambda]+1}^{\infty} \int K(\xi) U_{i,j}(\xi) d\xi,$$

$$U_{i,j}(\xi) = f_j(x_0 + \xi h - b_j \epsilon_{i-j} - \tilde{X}_{i,j+1}) - f_{j+1}(x_0 + \xi h - \tilde{X}_{i,j+1})$$

$$- f(x_0 + \xi h - b_j \epsilon_{i-j}) + E\{f(x_0 + \xi h - b_j \epsilon_{i-j})\}.$$

If

(5.30) $$E\{|V_{n1}|^2\} = o(n^{2/(\alpha\beta)}) \quad \text{and} \quad E\{|V_{n2}|^r\} = o(n^{r/(\alpha\beta)}),$$

the result of the lemma follows from (5.29). We establish (5.30).

As in the proof of Lemma 4.2, we rewrite V_{n1} and V_{n2} as

(5.31) $$V_{n1} = \int K(\xi) (\sum_{j=1-[n^\lambda]}^{n-s_0} M_{n,j}^{(1)}(\xi)) d\xi \quad \text{and} \quad V_{n2} = \int K(\xi) (\sum_{j=-\infty}^{n-1-[n^\lambda]} M_{n,j}^{(2)}(\xi)) d\xi,$$

where

$$M_{n,j}^{(1)}(\xi) = \sum_{i=1\vee(j+s_0)}^{n} U_{i,i-j}(\xi) I(i-j \le [n^\lambda]) \quad \text{and} \quad M_{n,j}^{(2)}(\xi) = \sum_{i=1\vee(j+s_0)}^{n} U_{i,i-j}(\xi) I(i-j > [n^\lambda]).$$

$\{M_{n,j}^{(l)}(\xi)\}$ form martingale difference sequences with respect to $\{\mathcal{S}_j\}$.

(5.30) follows from (5.31), Jensen's inequality and the von Bahr and Esseen inequality if we have shown that for any positive M, uniformly on $\{|\xi| \le M\}$,

(5.32) $$\sum_{j=1-[n^\lambda]}^{n-s_0} (\sum_{i=1\vee(j+s_0)}^{n} (E\{U_{i,i-j}^2(\xi) I(i-j \le [n^\lambda])\})^{1/2})^2 = o(n^{2/(\alpha\beta)}),$$

(5.33) $$\sum_{j=-\infty}^{n-1-[n^\lambda]} (\sum_{i=1\vee(j+s_0)}^{n} (E\{|U_{i,i-j}(\xi)|^r I(i-j > [n^\lambda])\})^{1/r})^r = o(n^{r/(\alpha\beta)}).$$

Provided that on $\{|\xi| \le M\}$,

(5.34) $$\mathrm{E}\{|U_{i,j}(\xi)|^r\} \le Cj^{1-2r\beta}, \quad j \ge s_0,$$

some calculation as on pp. 270-271 of Surgailis (2002) implies (5.32) and (5.33). In the calculation we use the fact that

$$\mathrm{E}\{|U_{i,j}(\xi)|^2\} \le C\mathrm{E}\{|U_{i,j}(\xi)|^r\}$$

since $|U_{i,j}(\xi)|$ is uniformly bounded in i, j, and ξ.

We establish (5.34). Hereafter we suppress the dependence on ξ.

We decompose $U_{i,j}$ into three components.

(5.35) $$U_{i,j} = U_{i,j}^{(1)} + U_{i,j}^{(2)} + U_{i,j}^{(3)},$$

where

$$U_{i,j}^{(1)} = f_j(x_0 + \xi h - b_j \epsilon_{i-j} - \tilde{X}_{i,j+1}) - f_{j+1}(x_0 + \xi h - \tilde{X}_{i,j+1})$$
$$\qquad - f_j(x_0 + \xi h - b_j \epsilon_{i-j}) + \mathrm{E}\{f_j(x_0 + \xi h - b_j \epsilon_{i-j})\},$$

$$U_{i,j}^{(2)} = f_j(x_0 + \xi h - b_j \epsilon_{i-j}) - f_j(x_0 + \xi h) - f(x_0 + \xi h - b_j \epsilon_{i-j}) + f(x_0 + \xi h),$$

$$U_{i,j}^{(3)} = -\mathrm{E}\{U_{i,j}^{(2)}\}.$$

First we evaluate $U_{i,j}^{(1)}$, which is written as

$$U_{i,j}^{(1)} = \int \left(\int_{-b_j u}^{-b_j \epsilon_{i-j}} (f_j'(x_0 + \xi h + z - \tilde{X}_{i,j+1}) - f_{j+1}'(x_0 + \xi h + z))dz \right) dG(u).$$

Since

$$\left| \int_{-b_j u}^{-b_j \epsilon_{i-j}} (f_j'(x_0 + \xi h + z - \tilde{X}_{i,j+1}) - f_{j+1}'(x_0 + \xi h + z))dz \right| \le C(|b_j u| + |b_j \epsilon_{i-j}|)|\tilde{X}_{i-j+1}|,$$

we have

(5.36) $$\mathrm{E}\{|U_{i,j}^{(1)}|^r\} \le Cj^{1-2\beta r}.$$

Next we deal with $U_{i,j}^{(2)}$, which is also written as

$$U_{i,j}^{(2)} = \int_0^{-b_j \epsilon_{i-j}} (f_j'(x_0 + \xi h + z) - f'(x_0 + \xi h + z))dz.$$

Using **P3** and the above expression, we obtain

(5.37) $$|U_{i,j}^{(2)}| \leq C|b_j \epsilon_{i-j}|j^{1/r-\beta}.$$

(5.37) implies that

(5.38) $$E\{|U_{i,j}^{(2)}|^r\} \leq Cj^{1-2\beta r}.$$

Finally, by Jensen's inequality, we obtain

(5.39) $$|U_{i,j}^{(3)}|^r \leq Cj^{1-2\beta r}.$$

(5.34) follows from (5.36), (5.38), and (5.39). Hence the proof of the lemma is complete.

Acknowledgements

References

Bryk, A. and Mielniczuk, J. (2005). Asymptotic properties of density estimates for linear processes: application of projection method, *Journal of Nonparametric Statistics*, **17**, 121–133.

Cheng, B. and Robinson, P. M. (1991). Density estimation in Strongly dependent non-linear time series, *Statistica Sinica*, **1**, 335–359.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory*, 2nd ed., Springer, New York.

Csörgő, S. and Mielniczuk, J. (1995). Density estimation under long-range dependence. *The Annals of Statistics*, **28**, 990–999.

Doukhan, P. (1994). *Mixing: Properties and Examples*, Lecture Notes in Statistics **85**, Springer, New York.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series*, Springer, New York.

Giraitis, L., Koul, H. L., and Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors, *Statistics & Probability Letters*, **29**, 317–335.

Giraitis, L. and Surgailis, D. (1986). Multivariate Appell polynomials and the central limit theorem, *Dependence in Probability and Statistics* (eds. E. Eberlein and M. S. Taqqu), pp.21–71, Birkhäser, Boston.

Hall, P. and Hart, J. D. (1990). Convergence rates in density estimation for data from infinite-order moving average process, *Probability Theory and Related Fields*, **87**, 253–274.

Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes: asymptotic normality and optimal bandwidth derivation, *The Annals of the Institute of Statistical Mathematics*, **48**, 429–449.

Hidalgo, J. (1997). Non-parametric estimation with strongly dependent multivariate time

series, *Journal of Time Series Analysis*, **18**, 95–122.

Ho, H.-C. (1996). On central and non-central limit theorems in density estimation for sequences of long-range dependence, *Stochastic Processes and their Applications*, **63**, 153–174.

Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages, *The Annals of Statistics*, **24**, 992–1024.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages, *The Annals of Probability*, **25**, 1636–1669.

Honda, T. (2000). Nonparametric density estimation for a long-range dependent linear process, *The Annals of the Institute of Statistical Mathematics*, **52**, 599–611.

Hsing, T. (1999). On the asymptotic distribution of partial sum of functionals of infinite-variance moving averages, *The Annals of Probability*, **27**, 1579–1599.

Koul, H. L. and Surgailis, D. (2001). Asymptotics of empirical processes of long memory moving averages with infinite variance, *Stochastic Processes and their Applications*, **91**, 309–336.

Koul, H. L., Surgailis, D. (2002). Asymptotic expansion of the empirical process of long memory moving averages, *Empirical process techniques for dependent data* (eds. H. Dehling, T. Mikosch, and M. Sørensen), pp.213–239, Birkhäser, Boston.

Peng, L. and Yao, Q. (2004). Nonparametric regression under dependent errors with infinite variance, *The Annals of the Institute of Statistical Mathematics*, **56**, 73–86.

Pipiras, V. and Taqqu, M. S. (2003). Central limit theorems for partial sums of bounded functionals of infinite-variance moving averages, *Bernoulli*, **9**, 833–855.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, London.

Schick, A. and Wefelmeyer, W. (2005). *Pointwise Convergence Rates for Kernel Density Estimators in Linear Processes*, forthcoming in *Statistics & Probability Letters*.

Surgails, D. (2002). Stable limits of empirical processes of moving averages with infinite variance, *Stochastic Processes and their Applications*, **100**, 255–274.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer, New York.

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes, *The Annals of Statistics*, **30**, 1441–1459.

Wuertz, D. and many others and see the SOURCE file (2006). fBasics: Rmetrics - Marketes and Basic Statistics, R package version 221.10065, http://www.rmetrics.org.

Table1: $\alpha = 1.2$

x_0		0.0			0.75			1.5		
h		0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
0.9	mean	0.2872	0.2867	0.2848	0.2204	0.2203	0.2198	0.1160	0.1163	0.1169
	var	0.0325	0.0299	0.0276	0.0280	0.0260	0.0243	0.0143	0.0135	0.0129
	mse	0.0326	0.0300	0.0278	0.0280	0.0261	0.0243	0.0143	0.0136	0.0130
1.3	mean	0.2984	0.2972	0.2953	0.2224	0.2220	0.2216	0.1132	0.1139	0.1147
	var	0.0092	0.0076	0.0067	0.0083	0.0070	0.0062	0.0039	0.0034	0.0031
	mse	0.0092	0.0077	0.0067	0.0083	0.0070	0.0063	0.0039	0.0034	0.0031
1.7	mean	0.2987	0.2963	0.2942	0.2219	0.2222	0.2222	0.1157	0.1158	0.1163
	var	0.0057	0.0041	0.0033	0.0044	0.0033	0.0027	0.0022	0.0017	0.0014
	mse	0.0057	0.0041	0.0033	0.0045	0.0034	0.0028	0.0023	0.0017	0.0014
2.1	mean	0.2990	0.2973	0.2946	0.2218	0.2219	0.2219	0.1138	0.1146	0.1153
	var	0.0048	0.0032	0.0024	0.0034	0.0024	0.0018	0.0018	0.0012	0.0010
	mse	0.0048	0.0032	0.0024	0.0034	0.0024	0.0019	0.0018	0.0013	0.0010
∞	mean	0.2987	0.2969	0.2946	0.2221	0.2213	0.2210	0.1146	0.1155	0.1161
	var	0.0040	0.0025	0.0018	0.0030	0.0019	0.0013	0.0017	0.0011	0.0008
	mse	0.0040	0.0025	0.0018	0.0031	0.0020	0.0014	0.0017	0.0011	0.0009

Table2: $\alpha = 1.5$

x_0		0.0			0.75			1.5		
h		0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
0.9	mean	0.2842	0.2827	0.2808	0.2313	0.2307	0.2301	0.1362	0.1370	0.1375
	var	0.0119	0.0103	0.0094	0.0129	0.0115	0.0106	0.0093	0.0086	0.0082
	mse	0.0119	0.0103	0.0094	0.0129	0.0115	0.0106	0.0094	0.0087	0.0082
1.3	mean	0.2890	0.2876	0.2858	0.2344	0.2339	0.2335	0.1350	0.1362	0.1366
	var	0.0054	0.0038	0.0031	0.0052	0.0039	0.0033	0.0034	0.0028	0.0024
	mse	0.0054	0.0038	0.0031	0.0052	0.0039	0.0033	0.0035	0.0028	0.0025
1.7	mean	0.2867	0.2860	0.2845	0.2371	0.2364	0.2351	0.1369	0.1370	0.1375
	var	0.0043	0.0029	0.0021	0.0037	0.0025	0.0020	0.0025	0.0017	0.0014
	mse	0.0043	0.0029	0.0021	0.0038	0.0026	0.0021	0.0026	0.0018	0.0014
2.1	mean	0.2865	0.2857	0.2845	0.2347	0.2341	0.2336	0.1371	0.1370	0.1371
	var	0.0042	0.0027	0.0020	0.0035	0.0023	0.0017	0.0021	0.0014	0.0011
	mse	0.0042	0.0027	0.0020	0.0036	0.0024	0.0018	0.0022	0.0015	0.0011
∞	mean	0.2867	0.2856	0.2846	0.2335	0.2335	0.2330	0.1373	0.1374	0.1376
	var	0.0038	0.0024	0.0017	0.0031	0.0020	0.0014	0.0021	0.0013	0.0010
	mse	0.0039	0.0024	0.0017	0.0032	0.0020	0.0014	0.0021	0.0014	0.0010

Table3: $\alpha = 1.8$

x_0		0.0			0.75			1.5		
h		0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
0.9	mean	0.2799	0.2788	0.2778	0.2391	0.2388	0.2384	0.1514	0.1520	0.1525
	var	0.0064	0.0050	0.0043	0.0073	0.0061	0.0055	0.0067	0.0060	0.0056
	mse	0.0065	0.0051	0.0043	0.0073	0.0061	0.0055	0.0067	0.0060	0.0057
1.3	mean	0.2836	0.2822	0.2812	0.2408	0.2402	0.2397	0.1517	0.1517	0.1518
	var	0.0043	0.0028	0.0021	0.0043	0.0032	0.0026	0.0032	0.0025	0.0021
	mse	0.0043	0.0028	0.0021	0.0044	0.0032	0.0026	0.0032	0.0025	0.0021
1.7	mean	0.2830	0.2825	0.2811	0.2415	0.2410	0.2406	0.1527	0.1528	0.1530
	var	0.0041	0.0026	0.0019	0.0033	0.0023	0.0018	0.0024	0.0017	0.0013
	mse	0.0041	0.0026	0.0019	0.0033	0.0023	0.0018	0.0025	0.0017	0.0013
2.1	mean	0.2833	0.2823	0.2810	0.2415	0.2414	0.2407	0.1508	0.1514	0.1518
	var	0.0039	0.0025	0.0018	0.0035	0.0023	0.0017	0.0022	0.0015	0.0011
	mse	0.0039	0.0025	0.0018	0.0035	0.0023	0.0017	0.0022	0.0015	0.0012
∞	mean	0.2819	0.2815	0.2810	0.2411	0.2410	0.2404	0.1512	0.1517	0.1523
	var	0.0038	0.0024	0.0017	0.0033	0.0021	0.0014	0.0021	0.0014	0.0010
	mse	0.0038	0.0024	0.0017	0.0033	0.0021	0.0014	0.0021	0.0014	0.0011