

## Asymptotic properties of density estimates for linear processes: application of projection method

ARTUR BRYK<sup>†¶</sup> and JAN MIELNICZUK<sup>‡¶\*</sup>

<sup>†</sup>Warsaw University of Technology, Pl. Politechniki 1, 00-601 Warsaw, Poland

<sup>‡</sup>Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland

(Received December 2003; revised in final form June 2004)

We specify conditions under which kernel density estimate for linear process is weakly and strongly consistent, and establish rates of its pointwise and uniform convergence. In particular, it is proved that for short-range dependent data of size  $n$  and bandwidth  $b_n$ , the rate of convergence is  $\mathcal{O}((\log n/nb_n)^{1/2} + b_n^2)$ . The results are established using projection method introduced in this setup by Ho and Hsing (Ho, H. C. and Hsing, T. (1996). On asymptotic expansion of the empirical process of long-memory moving averages. *Annals of Statistics*, **24**, 992–1024.) and Wu (Wu, W. B. (2001). Nonparametric estimation for stationary processes, *Ph.D. thesis*, University of Michigan, available at <http://www.stat.uchicago.edu/research/techreports.html>).

**Keywords:** Long- and short-range dependence; Kernel density estimators; Linear process; Projection method; Rates of convergence

### 1. Introduction

Let  $(X_t)_{t=1}^\infty$  be a stationary sequence with marginal density  $f$ , and  $\hat{f}_n(x)$  be a kernel density estimate of  $f(x)$  based on  $n$  observations  $X_1, X_2, \dots, X_n$  given by

$$\hat{f}_n(x) = \frac{1}{nb_n} \sum_{t=1}^n K\left(\frac{x - X_t}{b_n}\right),$$

where the kernel  $K$  is some function, not necessarily positive, such that  $\int K(s) ds = 1$ , and bandwidths (smoothing parameters) satisfy natural conditions  $b_n \rightarrow 0$  and  $nb_n \rightarrow \infty$ . Majority of asymptotic properties of  $\hat{f}_n(x)$  are established for independent  $(X_t)$ ; however, recently, there has been an increasing interest in the dependent case. In one line of research, weak dependence assumptions (usually various types of mixing conditions) are imposed on  $(X_t)$ , and it is shown that the results for independent data carry over to this more general case;

---

<sup>¶</sup>Affiliated also with Polish–Japanese Institute of Information Technology, Warsaw, Poland

\*Corresponding author. Email: [miel@ipipan.waw.pl](mailto:miel@ipipan.waw.pl)

see ref. [1] for the overview of such results. Alternatively, one can consider a flexible sub-model which restricts a class of considered stationary processes but not the strength of their dependence, and try to check when the analogy with independent case breaks down. This is especially interesting in the case of estimation of local parameters as a density function, because the answer to this question depends not only on the strength of dependence, but also on the effective number of observations used in the estimation, which is determined by magnitude of smoothing parameter. For representative examples of this line of research, we refer to refs [2–5]. Studied models include transformed stationary Gaussian sequences and infinite order moving averages, among others. In this article, we consider the latter, assuming throughout that

$$X_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}, \quad t = 1, 2, \dots, \quad (1)$$

where  $(\eta_i)_{i=-\infty}^{\infty}$  are i.i.d. random variables with mean 0 and finite variance  $\sigma^2$ , and  $c_i$  is such that  $\sum_{i=0}^{\infty} c_i^2 < \infty$ . We additionally assume that  $c_0 = 1$  and that density  $f_1$  of  $\eta_1$  exists. If  $c_i = L(i)i^{-\beta}$ , where  $1/2 < \beta < 1$ , and  $L(\cdot)$  is slowly varying at  $\infty$ , routine calculation based on the Karamata theorem implies that  $r_X(i) := \text{Cov}(X_0, X_i) \sim C(\beta)L^2(i)i^{-(2\beta-1)}\sigma^2$ , where  $C(\beta) := \int_0^{\infty} (x+x^2)^{-\beta} dx$  and  $a_n \sim b_n$  means  $\lim_{n \rightarrow \infty} a_n/b_n \rightarrow 1$ . Thus, in this case, sum of absolute values of covariances diverge. This property is called long-range dependence (LRD) or long-memory, in contrast to short-range dependence (SRD) case of absolutely summable covariances. Note that if  $\sum_{i=0}^{\infty} |c_i| < \infty$ , or  $\beta > 1$  in the hyperbolic decay condition given earlier,  $(X_t)$  is SRD. Let  $\sigma_n^2 = \text{Var}(X_1 + \dots + X_n)$  and put  $c_i = 0$  for  $i < 0$ . Then, it is easily seen that  $\sigma_n^2 = \sigma^2 \sum_{k=-\infty}^n \left( \sum_{t=1}^n c_{t-k} \right)^2$  is  $\mathcal{O}(n)$  when  $\sum_{i=0}^{\infty} |c_i| < \infty$  and  $\sigma_n^2 \sim D(\beta)n^{2-(2\beta-1)}L^2(n)$ , where  $D(\beta) := \sigma^2[(2-2\beta)(3/2-\beta)]^{-1}C(\beta)$ .

In this article, we study conditions under which kernel density estimate is consistent in different senses and investigate rates of its a.s. convergence. In particular, we prove that, under natural conditions on bandwidths,  $\hat{f}_n(x)$  is always weakly (*i.e.*, in probability) consistent when density of  $\eta_1$  is Lipschitz and  $K \in \mathcal{L}^2$ . Moreover, for absolutely summable  $(c_i)$ , rate of a.s. convergence of  $\hat{f}_n(x) - f(x)$  to 0 is  $\mathcal{O}((\log n/nb_n)^{1/2} + b_n^2)$ . This parallels the result for independent data.

A method which turns out to be particularly well suited to investigate such problems is projection method, two variants of which were introduced in the considered setup by Ho and Hsing [3] and Wu [6]. The method is briefly discussed in section 2. Both variants are presented here, as both of them turn out to be useful depending on a particular problem considered. Applications of projection method to many statistical problems, including asymptotic behaviour of empirical processes and quantiles for linear processes and iterated dynamical systems, have been studied in depth by Wu in a series of papers [7–9].

Detailed analysis of asymptotic distributions of  $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$  is given by Wu and Mielniczuk [5]. Their results imply (cf. proof of their Theorem 2) that when density  $f_1$  of  $\eta_1$  is three times differentiable with bounded continuous and integrable derivatives, then  $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = \mathcal{O}_P((nb_n)^{-1/2} + \sigma_n/n)$ . The method of proof also relied on martingale representation given in equation (5). The present article indicates that projection method is useful as well with investigating rates of a.s. and uniform convergence.

Similar results to Theorem 4(a) on rates of uniform convergence based on strong mixing condition were established by Bosq [1] and Fan and Yao [10]. In particular, it follows from Theorem 5.3 of Chanda [11] that if strong mixing coefficients  $\alpha(n)$  of an underlying stationary process decay geometrically,  $b_n \sim cn^{-\gamma}$  for  $0 < \gamma < 1$ ,  $f$  is bounded on  $[a, b]$  and  $K$  satisfies Lipschitz condition, then  $\sup_{x \in [a, b]} |\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| = \mathcal{O}((\log n/nb_n)^{1/2})$  in probability. Thus, the rate of convergence of the random component of  $\hat{f}_n(x)$  coincides with that

of Theorem 4(a). More generally,  $\alpha(n) \leq Cn^{-\beta}$  with  $\beta > 5/2$  implies the same rate of convergence, provided the sequence  $(b_n)$  tends to 0 sufficiently slowly. One gets further insight into this problem by considering conditions under which linear processes are strongly mixing. There are many such results in the literature, *e.g.*, refs [11–14]. In particular, for  $c_n \sim cn^{-\delta}$  under certain regularity conditions on the density of  $\eta_i$ , it follows from Withers [13] that the strong mixing coefficients  $\alpha(k) = \mathcal{O}\left[\sum_{n=k}^{\infty} \max(C_n^{1/3}, \sqrt{C_n |\log C_n|})\right] = \mathcal{O}(k^{(4-2\delta)/3})$ , where  $C_k = \sum_{i=k}^{\infty} c_i^2$ . Thus, the conditions of Fan and Yao when  $c_n \sim cn^{-\delta}$  are satisfied for  $\delta > 23/4$ . At the same time, Theorem 4(a) does not require any conditions on the decay rate of  $(c_n)$ ; the sole condition being  $(c_n) \in l^1$ . Let us also note that as mixing coefficients measure departure from independence, asymptotic results based on mixing approach concern weak dependence case, and handling LRD case under such conditions seems infeasible. In the article, we restrict ourselves to linear processes and try to quantify the effect of dependence on the rates of convergence of  $\hat{f}_n(\cdot)$ .

## 2. Projection method

Let  $(V_t)_{t=1}^{\infty}$  be a strictly stationary sequence of real random variables such that  $E|V_1| < \infty$ , and  $V_t$  is  $\mathcal{F}_t$ -measurable, where  $(\mathcal{F}_t)_{t=-\infty}^{\infty}$  is an increasing sequence of  $\sigma$ -fields such that  $\cap_{i=-\infty}^t \mathcal{F}_i$  is trivial for any  $t$ . In this section, we do not assume that  $V_t$  is an infinite-order moving average. The main tool used in the article is the projection method, which exploits in different ways the following basic equality

$$V_t - \mathbb{E}V_t = \sum_{k=-\infty}^t \mathbb{E}(V_t|\mathcal{F}_k) - \mathbb{E}(V_t|\mathcal{F}_{k-1}), \quad (2)$$

holding a.s. owing to the fact that  $\mathbb{E}(V_t|\mathcal{F}_{-j}) \rightarrow \mathbb{E}(V_t|\cap_{i=-\infty}^t \mathcal{F}_i) = \mathbb{E}(V_t)$  a.s. as  $j \rightarrow \infty$ . Let

$$\mathcal{P}_k U = \mathbb{E}(U|\mathcal{F}_k) - \mathbb{E}(U|\mathcal{F}_{k-1})$$

denote projection differences. Note that the equality  $\mathbb{E}(U|\mathcal{F}_k) = \mathcal{P}_k U + \mathbb{E}(U|\mathcal{F}_{k-1})$  yields orthogonal decomposition of  $\mathbb{E}(U|\mathcal{F}_k)$  with respect to  $\mathcal{F}_{k-1}$ . It follows from equation (2) that

$$\sum_{t=1}^n (V_t - \mathbb{E}V_t) = \sum_{t=1}^n \sum_{k=-\infty}^t \mathcal{P}_k V_t. \quad (3)$$

There are two possible ways of regrouping summands in equation (3). The first one consists in writing it formally as

$$\sum_{k=-\infty}^n \mathcal{P}_k \sum_{t=1}^n V_t := \sum_{k=-\infty}^n U_{n,k}, \quad (4)$$

using  $\mathcal{P}_k V_l = 0$  for  $l < k$ , whereas the second one represents equation (3) as

$$\sum_{k=0}^{\infty} \sum_{t=1}^n \mathcal{P}_{t-k} V_t := \sum_{k=0}^{\infty} W_{n,k}. \quad (5)$$

In order to observe the main difference between equations (4) and (5), note that  $U_{n,k}$  in equation (4) are sums of not necessarily uncorrelated summands such that  $U_{n,k}$  and  $U_{n,k'}$  are uncorrelated for  $k \neq k'$ . On the other hand, summands  $\mathcal{P}_{t-k} V_t$ ,  $t = 1, \dots, n$ , of  $W_{n,k}$  are

uncorrelated being martingale differences with respect to  $(\mathcal{F}_{t-k})_{t=1}^n$ , but  $W_{n,k}$  may be correlated with  $W_{n,k'}$  for  $k \neq k'$ .

In what follows,  $\|Y\| = (\mathbb{E}(Y^2))^{1/2}$  denotes  $\mathcal{L}^2$ -norm of a random variable  $Y$ . Observe that equality (4) leads to a more precise bound for  $\mathcal{L}^2$ -norm of centered sum of  $V_t$  than equation (5). Namely, using stationarity of  $(V_t)$  we get [6]

$$\left\| \sum_{t=1}^n (V_t - \mathbb{E}V_t) \right\|^2 = \sum_{k=-\infty}^n \|U_{n,k}\|^2 \leq \sum_{k=-\infty}^n \left( \sum_{t=\max(1,k)}^n \|\mathcal{P}_1 V_{t-k+1}\| \right)^2, \quad (6)$$

whereas analogous handling of equation (5) yields

$$\left\| \sum_{t=1}^n (V_t - \mathbb{E}V_t) \right\| \leq \sum_{k=0}^{\infty} \left\| \sum_{t=1}^n \mathcal{P}_{t-k} V_t \right\| = n^{1/2} \sum_{k=1}^{\infty} \|\mathcal{P}_1 V_k\|,$$

which is obviously weaker than equation (6). This is the reason why representation (4), but not equation (5), is used to bound the variance, *i.e.*,  $\mathcal{L}^2$ -norm of centred density estimate in Lemma 1. In particular, if  $\|\mathcal{P}_1 V_t\| \leq \theta_t$  for  $t \geq 1$ , we have that the last expression in equation (6) is not greater than

$$\sum_{k=-\infty}^0 \left( \sum_{t=1}^n \theta_{t-k+1} \right)^2 + \sum_{k=1}^n \left( \sum_{t=k}^n \theta_{t-k+1} \right)^2 \leq \sum_{k=1}^{\infty} (\Theta_{n+k} - \Theta_k)^2 + n\Theta_n^2 =: \Xi_n^2, \quad (7)$$

where  $\Theta_n = \sum_{i=1}^n \theta_i$ . In the case when  $\sum_{i=1}^{\infty} \theta_i < \infty$ , we have  $\Xi_n^2 = \mathcal{O}(n)$  as  $\sum_{k=1}^{\infty} (\Theta_{n+k} - \Theta_k)^2 \leq n(\sum_{i=1}^{\infty} \theta_i)^2$ .

However, observe that it is often beneficial to exploit martingale structure of  $W_{n,k}$  in equation (5), *e.g.*, by taking advantage of known exponential inequalities for sums of martingale differences [15]. This is the approach used in Theorem 3 to study the rates of convergence of  $\hat{f}_n(x)$ . When using representation (5), summand  $W_{n,0}$  usually has to be handled differently from the terms  $W_{n,k}$  for  $k > 0$  as it involves random variable  $V_t$  which is not conditionally averaged in contrast to remaining terms. This leads to natural decomposition  $W_{n,0} + \sum_{k>0} W_{n,k}$  in which the second term is the sum of random variables  $V'_t = E(V_t | \mathcal{F}_{t-1}) - \mathbb{E}V_t$ ,  $t = 1, \dots, n$  for which decomposition (4) may be employed.

The idea to use martingale approximations to investigate properties of  $S_n = \sum_{t=1}^n V_t$  dates back to Gordin and Lifszic [16] who considered this problem for  $V_t = g(Y_t)$ , where  $Y_t$  is an ergodic Markov chain. Note that for  $V_t$  defined in equation (1),  $\hat{f}_n(x)$  is a special case of this setup by letting  $Y_t = (\dots, \eta_{t-1}, \eta_t)$  and defining  $g = g_n$ , by  $g_n(Y_t) = n^{-1}K_{b_n}(x - X_t)$ . They proved that if  $\mathbb{E}(S_n | X_1) \rightarrow h(X_1)$  in  $\mathcal{L}^2$ , then  $h$  is a solution of the Poisson equation  $g(x) = h(x) - Qh(x)$ , where  $Qh(x) = \int h(y)Q(x; dy)$  and  $Q(x, B)$  is a transition function of  $(Y_t)$ . Then  $S_n = \sum_{k=1}^n M_k + R_n$ , where  $M_k = h(Y_k) - Qh(Y_{k-1})$  is the martingale difference with respect to  $\mathcal{F}_{k-1}$  and  $R_n = Qh(Y_0) - Qh(Y_n)$  is a remainder term. However, for problems considered in this article, weaker conditions are obtained when representation (5) is studied directly than those pertaining to existence of solution of the Poisson equation for  $g_n$  defined earlier.

### 3. Consistency and rates of almost sure convergence of kernel density estimates

In Theorem 1, we investigate conditions under which weak and strong consistencies of  $\hat{f}_n$  hold. In both cases, no decay rate of coefficients  $c_i$  is needed. Theorem 2 provides asymptotic

representation of centred  $\hat{f}_n(x)$ . In Theorem 3, rate of a.s. convergence of  $\hat{f}_n(x)$  data is established for SRD and LRD case. Rate of uniform weak consistency over any finite interval is established in Theorem 4.

Projection method is applied to rows of row-wise stationary array  $Y_{tn} = n^{-1}(K_{b_n}(x - X_t) - \mathbb{E}\hat{f}_n(x))$ ,  $t = 1, 2, \dots, n$ ,  $n \in \mathbb{N}$ , and  $\mathcal{F}_t = \sigma(\dots, \eta_{t-1}, \eta_t)$  is the  $\sigma$ -field generated by all innovations up to the moment  $t$ .

#### THEOREM 1

- (a) Assume that the density  $f_1$  of  $\eta_1$  exists and is Lipschitz continuous with Lipschitz constant  $A$ ,  $nb_n \rightarrow \infty$  and  $\int K^2(s) ds < \infty$ . Then  $\hat{f}_n(x) \rightarrow f(x)$  in probability.  
 (b) Assume, moreover, that  $f_1$  is bounded,  $nb_n/\log n \rightarrow \infty$  and  $\int |vK(v)| dv < \infty$ . Then  $\hat{f}_n(x) \rightarrow f(x)$  a.s.

The following theorem provides decomposition of  $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$  with explicit bound for a remainder term for  $c_i$ s decaying hyperbolically. A weaker result under more stringent conditions on density  $f_1$  and with remainder term  $\mathcal{O}_P(n^{-1/2}) + o_P(\sigma_n/n)$  has been proved by Wu and Mielniczuk (2002). Let  $\Xi_n$  be defined as in equation (7) with  $\theta_t = c_{t-1}(\sum_{i=t-1}^{\infty} c_i^2)^{1/2}$  and let  $f \star g$  denote convolution of  $f(\cdot)$  and  $g(\cdot)$ . A sequence  $(Z_n)_{n=1}^{\infty}$  of random variables converges to 0 completely if  $\sum_{n=1}^{\infty} P(|Z_n| > \varepsilon) < \infty$  for any  $\varepsilon > 0$ .

**THEOREM 2** Assume that  $f_1$  is bounded two times continuously differentiable function with bounded derivatives, and  $K$  satisfies assumptions of Theorem 1(b). Then

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = M_n(x) - K_b \star f'(x)n^{-1} \sum_{t=1}^n X_{t,t-1} + \mathcal{O}\left(\frac{\Xi_n}{n}\right), \quad (8)$$

where  $M_n(x) := \sum_{t=1}^n Y_{tn} - \mathbb{E}(Y_{tn}|\mathcal{F}_{t-1})$ ,  $X_{t,s} := \mathbb{E}(X_t|\mathcal{F}_s)$  and  $\Xi_n$  is defined in equation (7).  $M_n(x)$  is a sum of martingale differences converging completely to 0 provided  $nb_n/\log n \rightarrow \infty$ . If  $c_i = L(i)i^{-\beta}$  for some  $\beta > 1/2$ , then  $\Xi_n = \mathcal{O}(n^{1/2} \sum_{i=1}^{2n} i^{1/2-2\beta} L^2(i) + n^{2-2\beta} L^2(n))$  which is  $\mathcal{O}(n^{2-2\beta} L^2(n))$  for  $\beta < 3/4$  and  $\mathcal{O}(n^{1/2})$  for  $\beta > 3/4$ . If  $\beta > 3/4$  and  $\mathbb{E}\eta_1^4 < \infty$ , the second term in equation (8) converges completely to 0.

From Theorem 2, it follows that for twice continuously differentiable  $f$  we have  $\hat{f}_n(x) - f(x) = \mathcal{O}_P((nb_n)^{-1/2} + \sigma_n/n + b_n^2)$ . Next, we study a.s. rates and uniform rates in probability. Let  $\mathcal{C}^{(2)}(\mathbb{R})$  denote a family of twice continuously differentiable real functions.

**THEOREM 3** Assume that conditions of Theorem 1(b) are satisfied,  $f_1 \in \mathcal{C}^2(\mathbb{R})$ ,  $\mathbb{E}|\eta_1|^p < \infty$  for some  $p > 2$ ,  $K$  is symmetric and  $\int |K(v)|v^2 dv < \infty$ .

- (a) If  $\sum_{i=1}^{\infty} |c_i| < \infty$  then

$$|\hat{f}_n(x) - f(x)| = \mathcal{O}\left(\left(\frac{\log n}{nb_n}\right)^{1/2} + b_n^2\right) \text{ a.s.} \quad (9)$$

- (b) If  $c_i = L(i)i^{-\beta}$  for  $1/2 < \beta < 1$ , then

$$|\hat{f}_n(x) - f(x)| = \mathcal{O}\left(\left(\frac{\log n}{nb_n}\right)^{1/2} + \frac{\sigma_n(\log n)^{1/2}}{n} + b_n^2\right) \text{ a.s.} \quad (10)$$

**THEOREM 4** Assume that conditions of Theorem 1(b) are satisfied,  $K$  is bounded, symmetric and Lipschitz continuous with Lipschitz constant  $L$ ,  $f_1$  is twice continuously differentiable with bounded derivatives,  $f'_1 \in \mathcal{L}^1(\mathbb{R})$  and  $f^2 \in \mathcal{L}^2(\mathbb{R})$ .

(a) If  $\sum_{i=1}^{\infty} |c_i| < \infty$  then

$$\sup_{x \in [a, b]} |\hat{f}_n(x) - f(x)| = \mathcal{O}_P \left( \left( \frac{\log n}{nb_n} \right)^{1/2} + b_n^2 \right)$$

(b) If  $c_i = L(i)i^{-\beta}$  for  $1/2 < \beta < 1$  then

$$\sup_{x \in [a, b]} |\hat{f}_n(x) - f(x)| = \mathcal{O}_P \left( \left( \frac{\log n}{nb_n} \right)^{1/2} + \frac{\sigma_n}{n} + b_n^2 \right)$$

for any finite interval  $[a, b]$ .

#### 4. Proofs

The proofs of Theorems 1–4 proceed through the following lemmas. Let us stress that both variants of projection method turn out to be useful here. We apply bound (6) for  $\|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\|$  to prove Theorem 1(a). Decomposition (5)

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = W_{n,0} + \sum_{k>0} W_{n,k} =: M_n(x) + N_n(x)$$

is employed to prove Theorem 1(b). Martingale structure of  $M_n(x)$  is used to establish its a.s. convergence to 0, whereas term  $N_n(x)$  is analysed using the ergodic theorem. Furthermore, we take advantage of bound (6) again to find suitable approximation for  $N_n(x)$  needed to prove Theorem 2. Exponential inequality for sums of martingale differences  $M_n(x)$  is used to prove Theorem 3, whereas the Burkholder inequality is employed to study the term  $N_n(x)$ .

**LEMMA 1** Assume that assumptions of Theorem 1(a) are satisfied. Then

$$\text{Var } \hat{f}_n(x) \leq \frac{8C_1^2 \|\eta_1\|^2}{n^2} \sum_{k=-\infty}^n \left( \sum_{t=1}^n |c_{t-k}| \right)^2 + \frac{2}{n} \text{Var } K_{b_n}(x - X_1), \quad (11)$$

where  $C_1 = A \int |K(s)| ds$ .

Observe that the bound in equation (11) can be written as  $C \text{Var}(\bar{X}_1 + \dots + \bar{X}_n)/n + D \text{Var } \hat{f}_n^0(x)$ , where  $\bar{X}_t = \sum_{i=0}^{\infty} |c_i| \eta_{t-i}$  and  $\hat{f}_n^0(x)$  is a kernel density estimate based on i.i.d. sample of size  $n$  having density  $f$ . This follows upon noting that  $\bar{X}_1 + \dots + \bar{X}_n = \sum_{k=-\infty}^n \sum_{t=1}^n |c_{t-k}| \eta_k$ . Only the first term in the bound is influenced by dependence of underlying sequence  $(X_t)$ .

**LEMMA 2** Let  $f_1$  be a bounded density,  $\int K^2(s) ds < \infty$  and  $nb_n/\log n \rightarrow \infty$ . Then

$$M_n(x) = \sum_{t=1}^n Y_{tn} - \mathbb{E}(Y_{tn} | \mathcal{F}_{t-1}) \longrightarrow 0 \quad \text{completely.}$$

LEMMA 3 Assume that conditions of Theorem 2 are satisfied. Then

$$N_n(x) + K_b \star f'(x)n^{-1} \sum_{t=1}^n X_{t,t-1} = \mathcal{O}_P \left( \frac{\Xi_n}{n} \right),$$

where  $N_n(x) = \hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) - M_n(x)$  and  $X_{t,s} = E(X_t|\mathcal{F}_S)$ .

$$\text{Let } S_n = \sum_{t=1}^n X_t.$$

LEMMA 4 Assume that  $\mathbb{E}(|\eta_1|^{2p}) < \infty$  for some  $p \geq 1$ . Then  $\mathbb{E}|S_n|^{2p} = \mathcal{O}((\text{Var } S_n)^p)$ .

*Proof of Theorem 1(a).* Writing  $X_t = X_{t,t-1} + \eta_t$ , it is easy to see that continuity of  $f_1$  implies continuity of  $f$ , and thus in view of the Bochner lemma  $\text{Var } K_b(x - X_1) \sim f(x)b^{-1} \int K^2(s) ds$  for  $b \rightarrow 0$  whence the second term on the right-hand side of Eq. (11) tends to 0. Moreover,

$$\begin{aligned} \frac{1}{n^2} \sum_{k=-\infty}^n \left( \sum_{t=1}^n |c_{t-k}| \right)^2 &= \frac{1}{n^2} \sum_{k=-\infty}^n \sum_{1 \leq t, t' \leq n} |c_{t-k} c_{t'-k}| \\ &\leq \frac{2}{n} \sum_{i=0}^n \sum_{j=0}^{\infty} |c_j c_{j+i}| \leq \left( \sum_{j=0}^{\infty} c_j^2 \right)^{1/2} \frac{2}{n} \sum_{i=0}^n \left( \sum_{j=0}^{\infty} c_{j+i}^2 \right)^{1/2} \rightarrow 0 \end{aligned}$$

as  $\sum_{j=0}^{\infty} c_j^2 < \infty$ . Thus in view of equation (11),  $\text{Var } \hat{f}_n(x) \rightarrow 0$  and weak consistency of  $\hat{f}_n(x)$  follows as  $\mathbb{E}\hat{f}_n(x) \rightarrow f(x)$  in view of continuity of  $f$  at  $x$ . ■

*Proof of Theorem 1(b)* Observe that in view of Lemma 2 it is enough to prove that  $N_n(x) \rightarrow 0$  a.s. Note that

$$N_n(x) = \int K(v)n^{-1} \sum_{t=1}^n f_1(x - X_{t,t-1} - b_nv) dv - \mathbb{E}\hat{f}_n(x).$$

Observe that

$$\begin{aligned} |N_n(x)| &\leq \left| \int K(v)n^{-1} \sum_{t=1}^n (f_1(x - X_{t,t-1} - b_nv) - f_1(x - X_{t,t-1})) dv \right| \\ &\quad + \left| n^{-1} \sum_{t=1}^n f_1(x - X_{t,t-1}) - f(x) \right| + |\mathbb{E}\hat{f}_n(x) - f(x)|. \end{aligned} \quad (12)$$

The last term on the right-hand side tends to 0 in view of continuity of  $f$ . As  $(f_1(x - X_{t,t-1}))_{t=1}^{\infty}$  is ergodic as an instantaneous transformation of an ergodic sequence  $(X_{t,t-1})_{t=1}^{\infty}$ , the second term also tends to 0 a.s. in view of ergodic theorem as  $\mathbb{E}f_1(x - X_{t,t-1}) = f(x)$ . The first term is bounded by  $Ab_n \int K(v)v| dv \rightarrow 0$  using the fact that  $f_1$  is Lipschitz. ■

*Proof of Theorem 2* Observe that in view of Lemmas 2 and 3 we have that

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) - M_n(x) + K_b \star f'(x)n^{-1} \sum_{t=1}^n X_{t,t-1} = \mathcal{O}_P \left( \frac{\Xi_n}{n} \right)$$

and  $M_n(x) \rightarrow 0$  completely. In order to establish the bound on  $\Xi_n$ , observe that  $\Xi_n^2 = \mathcal{O}(2n\Theta_{2n}^2 + \sum_{i=n+1}^{\infty} (\Theta_{n+i} - \Theta_i)^2)$  and in view of the Karamata theorem  $\theta_i = \mathcal{O}(i^{1/2-2\beta} L^2(i))$ . Thus, we have

$$\sum_{i=n+1}^{\infty} (\Theta_{n+i} - \Theta_i)^2 = \sum_{i=n+1}^{\infty} \mathcal{O}((n\theta_i)^2) = n^2 \sum_{i=n+1}^{\infty} \mathcal{O}(i^{1-4\beta} L^4(i)) = \mathcal{O}(n^{4-4\beta} L^4(n)).$$

We now prove that for  $\beta > 3/4$  and  $\mathbb{E}\eta_1^4 < \infty$ ,  $n^{-1} \sum_{t=1}^n X_{t,t-1}$  tend to 0 completely. Applying Lemma 4 with  $X_{t,t-1}$  in place of  $X_t$ ,  $S'_n = \sum_{t=1}^n X_{t,t-1}$  and  $p = 2$ , we get

$$P\left(\left|\frac{S'_n}{n}\right| \geq \varepsilon\right) \leq \frac{\mathbb{E}S_n'^4}{(n\varepsilon)^4} = \mathcal{O}\left(\frac{(\text{Var } S'_n)^2}{n^4}\right).$$

Observe that reasoning analogously to proof of Theorem 1(a), we get  $\text{Var } S'_n = \mathcal{O}(n)$  when  $c_i$ s are absolutely summable, and as discussed in section 1  $\text{Var } S'_n \sim n^{2-(2\beta-1)} L^2(n)$  for  $1/2 < \beta < 1$  and whence for  $\beta > 3/4$ ,  $(\text{Var } S'_n)^2/n^4$  is summable. Thus for such cases  $P(|S'_n/n| \geq \varepsilon)$  is summable and conclusion follows from the Borel–Cantelli lemma. The case  $\beta \geq 1$  is treated analogously. ■

*Proof of Lemma 1* In order to prove inequality (11) we use bound (6). Let  $U_t = K_{b_n}(x - X_t) - \mathbb{E}K_{b_n}(x - X_t)$  and note that  $\mathcal{P}_k U_t = 0$  for  $t < k$ . Thus

$$\begin{aligned} n^2 \text{Var } \hat{f}_n(x) &\leq \sum_{k=-\infty}^n \left( \sum_{t=1}^n \|\mathcal{P}_1 U_{t-k+1}\| I\{t > k\} + \|\mathcal{P}_1 U_1\| I\{t = k\} \right)^2 \\ &\leq 2 \sum_{k=-\infty}^n \left( \sum_{t>k} \|\mathcal{P}_1 U_{t-k+1}\| \right)^2 + 2n \|\mathcal{P}_1 U_1\|^2. \end{aligned}$$

As  $\|\mathcal{P}_1 U_1\|^2 = \mathbb{E}U_1^2 - \mathbb{E}(\mathbb{E}(U_1|\mathcal{F}_0))^2 \leq \mathbb{E}U_1^2$ , it is enough to prove for  $t > 1$

$$\|\mathcal{P}_1 U_t\| \leq C_1(|c_{t-1}|(\mathbb{E}\eta_1^2))^{1/2} + |c_{t-1}|\mathbb{E}|\eta_1| \leq 2C_1|c_{t-1}|\|\eta_1\|.$$

To this end, observe that writing

$$X_t = \sum_{i=t-s}^{\infty} c_i \eta_{t-i} + \sum_{i=0}^{(t-s)-1} c_i \eta_{t-i} =: X_{t,s} + R_{t,s},$$

we have for  $t > 1$

$$\mathcal{P}_i U_t = \int K_{b_n}(x - z - X_{t,1}) f_{t-1}(z) dz - \int K_{b_n}(x - z - X_{t,0}) f_t(z) dz, \quad (13)$$

where  $f_{t-s}(\cdot)$  denotes a density of  $R_{t,s}$ . Moreover, an easy induction argument implies that  $f_t$  is Lipschitz continuous with Lipschitz constant  $A$  when  $f_1$  has this property. Note that



equation (13) can be written as

$$\int K(z)\{f_{t-1}(x - X_{t,1} - b_n z) - f_t(x - X_{t,1} - b_n z)\} dz + \int K(z)\{f_t(x - X_{t,1} - b_n z) - f_t(x - X_{t,0} - b_n z)\} dz =: \text{I} + \text{II}.$$

Using Lipschitz continuity of  $f_t$ , we have  $|\text{II}| \leq C_1 |c_{t-1} \eta_1|$ . Moreover, as  $f_t(s)$  can be written as convolution  $f_{t-1} \star \tilde{f}(s)$ , where  $\tilde{f}$  is a density of  $c_{t-1} \eta_1$ , we have that for  $s = x - X_{t,1} - b_n z$

$$\begin{aligned} |\text{II}| &\leq \int |K(z)| \int |f_{t-1}(s) - f_{t-1}(s - y)| \tilde{f}(y) dy dz \\ &\leq A \int |K(z)| \int |y| \tilde{f}(y) dy dz \\ &= C_1 \mathbb{E} |c_{t-1} \eta_1|. \end{aligned}$$

*Proof of Lemma 2* We use a special case of Freedman exponential inequality [15] stating that for sum  $S_n = \sum_{i=1}^n T_i$  of bounded martingale differences  $T_i$ ,  $|T_i| \leq B$ , we have for  $\lambda > 0$

$$\mathbb{E} \exp(\lambda S_n) \leq \exp(\beta B^{-2} e(B\lambda)),$$

where  $e(\lambda) = e^\lambda - \lambda - 1$  and  $\beta$  is a bound of conditional variances of  $T_i$  such that  $P(\sum_{i=1}^n \mathbb{E}(T_i^2 | \mathcal{F}_{i-1}) \leq \beta) = 1$ . In our case

$$\begin{aligned} \mathbb{E}(T_i^2 | \mathcal{F}_{i-1}) &= \mathbb{E}(K_{b_n}^2(x - X_i) | \mathcal{F}_{i-1}) - (\mathbb{E}(K_{b_n}(x - X_i) | \mathcal{F}_{i-1}))^2 \\ &\leq \mathbb{E}(K_{b_n}^2(x - X_i) | \mathcal{F}_{i-1}) \\ &= \frac{1}{b_n^2} \int K^2\left(\frac{x - X_{i,i-1} - s}{b_n}\right) f_1(s) ds \\ &\leq \frac{\int K^2(s) ds \sup f_1}{b_n} \end{aligned}$$

and  $B = 2 \sup K/b_n$ . Thus  $\sum_{i=1}^n \mathbb{E}(T_i^2 | \mathcal{F}_{i-1}) \leq nC/b_n$  for some positive  $C$ , and for  $\lambda > 0$

$$P(M_n \geq \varepsilon) \leq \mathbb{E} \frac{\exp(\lambda S_n)}{\exp(n\lambda \varepsilon)} \leq \exp\left\{\frac{(CB^{-2}ne(B\lambda))}{b_n} - n\lambda \varepsilon\right\}. \quad (14)$$

Let  $\lambda_n = \varepsilon_n b_n / C$ , where  $\varepsilon_n = (8C \log n / nb_n)^{1/2}$ . Note that as  $e(\lambda) \sim 2^{-1} \lambda^2$  for  $\lambda \rightarrow 0$  we have that  $e(B\lambda_n) < (3/4)(B\lambda_n)^2$  for sufficiently large  $n$ . Thus, for such  $n$ , the last bound is smaller than

$$\exp(6 \log n - 8 \log n) = \exp(-2 \log n). \quad (15)$$

Thus the Borel–Cantelli lemma implies that  $M_n(x) \geq (8C \log n / nb_n)^{1/2}$  finitely often almost surely. Analogously, we show that  $\sum_n P(M_n(x) \leq -\varepsilon_n) < \infty$  and whence  $M_n(x) \rightarrow 0$  completely. ■

*Proof of Lemma 3* The proof uses the main ideas of the proof of Lemma 4 by Mielniczuk and Wu (2004). Observe that  $nN_n(x) = \sum_{t=1}^n K_b \star T_t(x)$ , where  $T_t(x) = f_1(x - X_{t,t-1}) - f(x)$ ,

where  $\star$  denotes convolution. We will prove that

$$\|\mathcal{P}_1 K_b \star T_t(x) + K_b \star f'(x) c_{t-1} \eta_1\| = \mathcal{O}(|c_{t-1}| C_{t-1}^{1/2}), \quad (16)$$

where  $C_t = \sum_{i=t}^{\infty} c_i^2$ . This will prove the lemma in view of equation (7). Writing

$$\left\| \int K_b(x-s) \xi_s \, ds \right\|^2 = \mathbb{E} \left( \int K_b(x-s) K_b(x-s') \xi_s \xi_{s'} \, ds \, ds' \right)$$

for any family of random variables  $(\xi_s)_{s \in \mathbb{R}}$  and using Cauchy inequality, we have

$$\left\| \int K_b(x-s) \xi_s \, ds \right\|^2 \leq \sup_s \|\xi_s\|^2.$$

Thus, in order to prove equation (16) in view of  $\mathcal{P}_1 T_t(x) = f_{t-1}(x - X_{t,1}) - f_t(x - X_{t,0})$ , it is enough to show that

$$\sup_s \|f_{t-1}(s - X_{t,1}) - f_t(s - X_{t,0}) + f'(s) c_{t-1} \eta_1\| = \mathcal{O}(|c_{t-1}| C_{t-1}^{1/2}). \quad (17)$$

In view of Lemma 1 by Mielniczuk and Wu [17],  $f'(x) = \mathbb{E} f'_{t-1}(x - X_{t,1})$ . Thus

$$\begin{aligned} \sup_s |f'(s) - f'_{t-1}(s)| &\leq \sup_s \mathbb{E} |f'_{t-1}(s - X_{t,1}) - f'_{t-1}(s)| \\ &\leq \sup_s \|f'_{t-1}(s - X_{t,1}) - f'_{t-1}(s)\| = \mathcal{O}(C_{t-1}^{1/2}) \end{aligned}$$

as  $\sup_{s,t} |f_t^{(2)}(s)| < \infty$ . Moreover

$$\sup_s \|f'_{t-1}(s) - f'_{t-1}(s - X_{t,0})\| = \mathcal{O}(C_t^{1/2}).$$

Thus equation (17) follows from

$$\sup_s \|f_{t-1}(s - X_{t,1}) - f_t(s - X_{t,0}) + f'_{t-1}(s - X_{t,0}) c_{t-1} \eta_1\| = \mathcal{O}(|c_{t-1}|^2). \quad (18)$$

Let  $\eta_1^*$  be a copy of  $\eta_1$  independent of  $(\eta_i)_{i=-\infty}^{\infty}$  and  $X_{t,1}^* = X_{t,1} - c_{t-1} \eta_1 + c_{t-1} \eta_1^*$ . Random variable on the left-hand side of equation (18) can be written as

$$\mathbb{E}(f_{t-1}(s - X_{t,1}) - f_{t-1}(s - X_{t,1}^*) + f'_{t-1}(s - X_{t,0}) c_{t-1} \eta_1 | \mathcal{F}_1) \quad (19)$$

Adding to equation (19)

$$\mathbb{E}(f_{t-1}(s - X_{t,0}) - f_{t-1}(s - X_{t,0}) - f'_{t-1}(s - X_{t,0}) c_{t-1} \eta_1^* | \mathcal{F}_1) = 0,$$

we see that equation (18) follows from two term Taylor expansion of  $f_{t-1}(s - X_{t,0}) - f_{t-1}(s - X_{t,1})$  and  $f_{t-1}(s - X_{t,0}) - f_{t-1}(s - X_{t,1}^*)$ .  $\blacksquare$

*Proof of Lemma 4*  $S_n = \sum_{i=-\infty}^n \sum_{t=1}^n c_{t-i} \eta_i =: \sum_{i=-\infty}^n d_{n-i} \eta_i$  is a sum of martingale differences as  $\eta_i$  is independent. Thus, the Burkholder inequality implies

$$E|S_n|^{2p} \leq C_{2p} \mathbb{E} \left( \sum_{i=-\infty}^n (d_{n-i} \eta_i)^2 \right)^p$$

and in view of the Minkowski inequality, we have

$$\begin{aligned} \|S_n\|_{2p}^2 &\leq C_{2p}^{1/p} \left\| \sum_{i=-\infty}^n (d_{n-i} \eta_i)^2 \right\|_p \leq C_{2p}^{1/p} \sum_{i=-\infty}^n \|(d_{n-i} \eta_i)^2\|_p \\ &= C_{2p}^{1/p} \sum_{i=-\infty}^n \|d_{n-i} \eta_i\|_{2p}^2 = C_{2p}^{1/p} (\text{Var } S_n) \|\eta_1\|_{2p}^2, \end{aligned}$$

from which the statement of the lemma follows. ■

*Proof of Theorem 3 (a)* Proof of Lemma 2 indicates that  $M_n(x) = \mathcal{O}((\log n/nb_n)^{1/2})$  a.s. As  $\mathbb{E}\hat{f}_n(x) - f(x) = \mathcal{O}(b_n^2)$ , it is enough to show that  $N_n(x) = \mathcal{O}((\log n/n)^{1/2} + b_n^2)$  a.s., provided  $c_i$ s are absolutely summable. We bound  $N_n(x)$  as in proof of Theorem 1(b) and note that the first and the third terms of the bound are  $\mathcal{O}(b_n^2)$ . Let

$$H_n(x) = \sum_{t=1}^n f_1(x - X_{t,t-1}) - f(x) =: \sum_{t=1}^n T_t(x).$$

We first show that

$$\|\mathcal{P}_1 T_t(x)\|_p = \mathcal{O}(|c_{t-1}|). \quad (20)$$

To this end, consider a coupled version of  $X_{t,t-1}$ :  $X_{t,t-1}^* = X_{t,t-1} - c_{t-1} \eta_1 + c_{t-1} \eta_1^*$ , where  $\eta_1^*$  has the same distribution as  $\eta_1$  and is independent of  $(\eta_i)$ . Then  $\mathbb{E}(f_1(x - X_{t,t-1}) | \mathcal{F}_0) = E(f_1(x - X_{t,t-1}^*) | \mathcal{F}_1)$ , and as  $f_1$  satisfies Lipschitz condition we have

$$\begin{aligned} \|\mathcal{P}_1 T_t(x)\|_p &= \|\mathbb{E}((f_1(x - X_{t,t-1}) - f_1(x - X_{t,t-1}^*)) | \mathcal{F}_1)\|_p \\ &\leq \|\mathbb{E}(A|c_{t-1}(\eta_1 - \eta_1^*) | \mathcal{F}_1)\|_p \leq \|A|c_{t-1}(\eta_1 - \eta_1^*)\|_p \\ &= \mathcal{O}(|c_{t-1}|) \end{aligned}$$

in view of  $E|\eta_1|^p < \infty$ . Consider decomposition of  $H_n(x) = \sum_{j=-\infty}^n \mathcal{P}_j H_n(x)$  into martingale differences  $\mathcal{P}_j H_n(x)$ . It follows from the Burkholder inequality that

$$\mathbb{E}|H_n(x)|^p \leq C_p \mathbb{E} \left( \sum_{j=-\infty}^n (\mathcal{P}_j H_n(x))^2 \right)^{p/2}$$

and, moreover,

$$\|\mathcal{P}_j H_n(x)\|_p \leq \sum_{t=j}^n \|\mathcal{P}_1 T_{t-j+1}(x)\|_p \leq C \sum_{t=j}^n |c_{t-j}|.$$

Thus

$$\begin{aligned}
 (\mathbb{E}|H_n(x)|^p)^{2/p} &\leq C \left\| \sum_{j=-\infty}^n (\mathcal{P}_j H_n(x))^2 \right\|_{p/2} \\
 &\leq C \sum_{j=-\infty}^n \|\mathcal{P}_j H_n(x)\|_p^2 \\
 &\leq C \sum_{j=-\infty}^n \left( \sum_{t=j}^n |c_{t-j}| \right)^2. \tag{21}
 \end{aligned}$$

As  $c_i$ s are absolutely summable we get  $\mathbb{E}|H_n(x)|^p \leq n^{p/2}$ . We apply Moricz [18] inequality stating that if for some non-negative  $a_i$  and  $r > 0$ ,  $q > 1$ ,  $\mathbb{E}|X_1 + \dots + X_i|^r \leq (\sum_{j=1}^i a_j)^q$ , then  $\mathbb{E}(\max_{i \leq n} |X_1 + \dots + X_i|^r) \leq C_{rq} (\sum_{j=1}^n a_j)^q$  with  $r = p$  and  $q = p/2$ . This together with equation (22) implies that  $\mathbb{E}(\max_{i \leq n} |H_i(x)|^p) \leq n^{p/2}$ . Thus  $H_n(x) = \mathcal{O}(n^{1/2}(\log n)^{1/2})$  a.s. It follows, by the Borel–Cantelli lemma from the observation, that

$$P(|\max_{i \leq 2^k} |H_i(x)| \geq \epsilon(2^k k)^{1/2}) \leq \frac{\mathbb{E} \max_{i \leq 2^k} |H_i(x)|^p}{(2^{kp/2} k^{p/2})} = \mathcal{O}(k^{-p/2})$$

which is summable as  $p > 2$ .

(b) Proof of (b) follows the same argument after noting that the analogue of equation (21) is

$$(\mathbb{E}|H_n(x)|^p)^{2/p} = \mathcal{O}(\sigma_n^2).$$

Moreover, two-fold application of the Karamata theorem yields that the Moricz inequality may be applied with  $a_i = i^{2-2\beta} L^2(i)$ . It yields  $\mathbb{E} \max_{i \leq n} |H_i(x)|^p \leq C \sigma_n^p$ . ■

*Proof of Theorem 4* We first prove that

$$\sup_{x \in [a, b]} |M_n(x)| = \mathcal{O}\left(\frac{\log n}{nb_n}\right)^{1/2}$$

almost surely. Let  $\epsilon_n = 3(16C \log n / nb_n)^{1/2}$ , where  $C$  is defined in the proof of Lemma 2 and  $a = x_0 < x_1 < \dots < x_{l_n} = b$  is an equipartition of  $[a, b]$  with  $l_n = \lceil 6L(b-a)/(b_n^2 \epsilon_n) \rceil + 1$ . Define  $k(x) = k$  to be the point of partition closest to  $x$ . Observe that

$$\begin{aligned}
 P\left(\sup_{x \in [a, b]} |M_n(x)| \geq \epsilon_n\right) &\leq P\left(\sup_{x \in [a, b]} |M_n(x) - M_n(x_k)| \geq \left(\frac{2}{3}\right) \epsilon_n\right) \\
 &\quad + P\left(\max_{0 \leq k \leq l_n} |M_n(x_k)| \geq \left(\frac{1}{3}\right) \epsilon_n\right).
 \end{aligned}$$

It is easy to see that in view of Lipschitz continuity of  $K$ , we have  $|M_n(x) - M_n(x_k)| \leq 2L(b-a)/(b_n^2 l_n) \leq \epsilon_n/3$  and whence the first probability on the right-hand side is 0. Second term is trivially bounded by

$$(l_n + 1) \max_{0 \leq k \leq l_n} P(|M_n(x_k)| \geq 3^{-1} \epsilon_n). \tag{22}$$

Choosing  $\lambda_n = \epsilon_n b_n / 3C$  and reasoning as in proof of Lemma 2, we get that equation (22) is bounded by  $(l_n + 1) \exp(-4 \log n) \sim (n/b_n^3 \log n)^{1/2} \exp(-4 \log n)$  which is summable, and

the conclusion follows from the Borel–Cantelli lemma. Next, we show that in the case of (b)  $\sup_{x \in [a, b]} |N_n(x)| = \mathcal{O}_P(\sigma_n/n + b_n^2)$ . Reasoning in case (a) is similar. Bounding  $N_n(x)$  as in proof of Theorem 1(b), we see that the first and the third terms of the bound are  $b_n^2$  uniformly in  $x \in [a, b]$ . In order to bound the second term let, as in the proof of Theorem 3,  $H_n(x) = \sum_{t=1}^n f_1(x - X_{t,t-1}) - f(x)$  and  $h_n(x) = H'_n(x)$ . From the proof of Lemma 5 in Wu [9], it follows that  $f'_1 \in \mathcal{L}^1(\mathbb{R})$  provided  $f^{(2)} \in \mathcal{L}^2(\mathbb{R})$ , and if  $f_1$  is bounded we have

$$\int_{\mathbb{R}} \mathbb{E} h_n^2(x) dx = \mathcal{O} \left( \sum_{j=-\infty}^t \left( \sum_{t=1}^n |c_{t-j}| \right)^2 \right) = \mathcal{O}(\sigma_n^2).$$

Moreover, observe that

$$\mathbb{E} \sup_{x \in [a, b]} |H_n(x) - H_n(a)|^2 \leq \mathbb{E} \left( \int_a^b |h_n(x)| dx \right)^2 \leq (b-a) \int_a^b \mathbb{E} h_n^2(x) dx.$$

The last two equations imply via the Markov inequality that  $\sup_{x \in [a, b]} |H_n(x) - H_n(a)| = \mathcal{O}_P(\sigma_n)$ . Proof of Theorem 3(a) implies that  $|H_n(a)| = \mathcal{O}(\sigma_n)$  and the conclusion follows via triangle inequality. ■

## Acknowledgements

We thank Wei Biao Wu for his insightful comments on handling the term  $H_n(x)$  in the proof of Theorem 4 and a referee for helpful remarks.

## References

- [1] Bosq, D., 1998, *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Vol. **110**, Lecture Notes in Statistics (New York: Springer).
- [2] 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.
- [3] Ho, H.C. and Hsing, T., 1996, On asymptotic expansion of the empirical process of long-memory moving averages. *Annals of Statistics*, **24**, 992–1024.
- [4] Csörgő, S. and Mielniczuk, J., 1999, Random design regression under long-range dependent errors. *Bernoulli*, **5**, 209–224.
- [5] Wu, W.B. and Mielniczuk, J., 2002, Kernel density estimation for linear processes. *Annals of Statistics*, **30**, 1441–1459.
- [6] Wu, W.B., 2001, Nonparametric estimation for stationary processes. PhD thesis, University of Michigan, available at <http://www.stat.uchicago.edu/research/techreports.html>.
- [7] Wu, W.B., 2002, Central limit theorems for functionals of linear processes. *Statistica Sinica*, **12**, 635–649.
- [8] Wu, W.B., 2003a, Empirical processes of long-memory sequences. *Bernoulli*, **9**, 809–831.
- [9] Wu, W.B., 2003b, On the Bahadur representation of sample quantiles for dependent sequences, preprint.
- [10] Fan, J. and Yao, Q., 2003, *Nonlinear Time Series. Nonparametric and Parametric Methods* (New York: Springer).
- [11] Chanda, K., 1974, Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, **11**, 401–408.
- [12] Gorodetskii, V.V., 1977, On the strong mixing property for linear sequences. *Theory of Probability and its Applications*, **22**, 411–412.
- [13] Withers, C.S., 1981, Conditions for linear process to be strongly mixing. *Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete*, **57**, 477–480.
- [14] Pham, T.D. and Tran, L.T., 1985, Some mixing properties of time series models. *Stochastic Process and their Applications*, **19**, 297–303.
- [15] Freedman, D., 1975, On tail probabilities for martingales. *Annals of Probability*, **3**, 100–118.
- [16] Gordin, M.I. and Lifszic, B., 1978, The central limit theorem for stationary Markov processes. *Doklady Akademii Nauk SSSR*, **19**, 392–394.
- [17] Mielniczuk, J. and Wu, W.B., 2004, On random-design model with dependent errors. *Statistica Sinica*, to appear.
- [18] Moricz, F., 1976, Moment inequalities and the strong laws of large numbers, *Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete*, **35**, 299–314.

