

# Wavelet-Based Density Estimation in Presence of Additive Noise under Various Dependence Structures

N. Hosseinioun

Statistics Department, Payame Noor University, 19395-4697 Tehran, Iran  
Email: Mails.students@gmail.com

Received 4 December 2015; accepted 16 January 2016; published 19 January 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

We study the following model:  $Y = X + \epsilon$ . The aim is to estimate the distribution of  $X$  when only  $Y_1, \dots, Y_n$  are observed. In the classical model, the distribution of  $\epsilon$  is assumed to be known, and this is often considered as an important drawback of this simple model. Indeed, in most practical applications, the distribution of the errors cannot be perfectly known. In this paper, the author will construct wavelet estimators and analyze their asymptotic mean integrated squared error for additive noise models under certain dependent conditions, the strong mixing case, the  $\beta$ -mixing case and the  $\rho$ -mixing case. Under mild conditions on the family of wavelets, the estimator is shown to be  $L_p$  ( $1 \leq p \leq \infty$ )-consistent and fast rates of convergence have been established.

## Keywords

Additive Noise, Density Estimation, Dependent Sequence, Rate of Convergence, Wavelet

---

## 1. Introduction

In practical situations, direct data are not always available. One of the classical models is described as follows:

$$Y_i = X_i + \epsilon_i$$

where  $X_i$  stands for the random samples with unknown density  $f_X$  and  $\epsilon_i$  denotes the *i.i.d.* random noise with density  $g$ . To estimate the density  $f_X$  is a deconvolution problem. Among the nonparametric methods of deconvolution, one can find estimation by model selection (e.g. Comte, Rozenhole and Taupin [1]), wavelet thresholding (e.g. [2]), kernel smoothing (e.g. Carroll and Hall, [3]), spline deconvolution or spectral cut-off (e.g. Johannes [4]) and Meister [5] basically on the effect of noise misspecification. However, a problem frequently

encountered is that the proposed estimator is not everywhere positive, and therefore is not a valid probability density.

Sometimes, this problem can be circumvented by repeated observations of the same variable of interest, each time with an independent error. This is the model of panel data (see for example Li and Vuong [6], Delaigle, Hall and Meister [7], or Neumann [8] and references therein). On the other hand, there are many application fields where it is not possible to do repeated measurements of the same variable. So, information about the error distribution can be drawn from an additional experiment: a training set is used by experimenters to estimate the noise distribution. Think of  $\epsilon$  as a measurement error due to the measuring device, then preliminary calibration measures can be obtained in the absence of any signal  $X$  (this is often called the instrument line shape of the measuring device).

Odiachi and Prieve [9] study the effect of additive noise in Total Internal Reflection Microscopy (TIRM) experiments. This is an optical technique for monitoring Brownian fluctuations in separation between a single microscopic sphere and a flat plate in aqueous medium. See Carroll and Hall [3], Devroye [10], Fan [11], Liu and Taylor [12], Masry [13], Stefanski and Carroll [14], Zhang [15], Hesse [16], Cator [17], Delaigle and Gijbels [18] for mainly kernel methods and Koo [19] for a spline method, Efromovich [20] for particular strategy in supersmooth case and Meister (2004), on the effect of noise misspecification.

In this paper, we extend Geng and Wang [21] (Theorems 4.1 and 4.2) for certain dependent. More precisely, we prove that the linear wavelet estimator attains the standard rate of convergence *i.e.* the optimal one with additive noise for more realistic and standard dependent conditions as polynomial strong mixing dependence, the  $\beta$ -mixing dependence and  $\rho$ -mixing dependence. The properties of wavelet basis allow us to apply sharp probabilistic inequalities which improve the performance of the considered linear wavelet estimator.

The organization of the paper is as follows. Assumptions on the model are presented in Section 2. Section 3 is devoted to our linear wavelet estimator and a general result. Applications are set in Section 5, while technical proofs are collected in Section 6.

## 2. Estimation Procedure

The Fourier transform of  $f \in L_1(\mathbb{R})$  is defined as follows:

$$\tilde{f}(x) = \int f(t) \exp(-itx) dt$$

It is well known that  $\tilde{f}(y) = \tilde{f}_x(t) \tilde{g}(t)$  for  $\tilde{g}(t) \neq 0$ . Let  $N$  be a positive integer. We assume that there exist constants  $c > 0$  and  $\delta > 1$  such that, for any  $x$ ,

$$|\tilde{g}(x)| \frac{c}{(1+x^2)^{\delta/2}} \tag{1}$$

One can easily find an example

$$g(x) = 1/2 \exp\{-|x|\}, x \in \mathbb{R}$$

which is the Laplace density and  $\tilde{g}(x) = \frac{1}{1+x^2}$ , which satisfies (2.1) with  $\delta = 2$ .

We consider an orthonormal wavelet basis generated by dilations and translations of a father Daubechies-type wavelet and a mother Daubechies-type wavelet of the family db2N (see [22]) Further details on wavelet theory can be found in Daubechies [22] and Meyer [23]. For any  $j \geq 0$ , we set  $\Lambda_j = \{0, \dots, 2^j - 1\}$  and for  $k \in \Lambda_j$ , we define  $\phi$  and  $\psi$  as father and mother wavelet:

$$\phi_{jk}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$$

With appropriated treatments at the boundaries, there exists an integer  $\tau$  such that, for any integer  $l \geq \tau$ ,

$$\Gamma = \{\phi_{l,k}, k \in \Lambda_l; \psi_{j,k}, j \in N - \{0, \dots, L-1\}, K \in \Lambda_j\}$$

forms an orthonormal basis of  $L_2[0,1]$ . For any integer  $l \geq \tau$  and  $f \in L_2[0,1]$ , we have the following wavelet expansion:

$$f(x) = \sum_{k \in \Lambda_l} \alpha_{l,k} \phi_{l,k}(x) + \sum_{j=l}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x), \quad x \in [0,1]$$

where  $\alpha_{l,k} = \int f \phi_{l,k}$  and  $\beta_{j,k} = \int f \psi_{j,k}$ . Furthermore we consider the following wavelet sequential definition of the Besov balls. We say  $f \in B_{p,r}^s(M)$ , with  $s > 0, p, r \geq 1$ , and  $M > 0$  if there exists a constant  $C > 0$ , such that

$$2^{\tau(1/2-1/p)} \left( \sum_{k \in \Lambda_\tau} |\alpha_{\tau,k}|^p \right)^{1/p} + \left( \sum_{j=\tau}^{\infty} \left( 2^{j(s+1/2-1/p)} \left( \sum_{k \in \Lambda_j} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq C$$

with the usual modifications if  $p=1$  or  $r=1$ . Note that, for particular choices of  $s, p$  and  $r, B_{p,r}^s(M)$  contains the classical Holder and Sobolev balls. See, e.g., Meyer [23] and Hardle, Kerkycharian, Picard and Tsybakov [24]. We define the linear wavelet estimator  $\hat{f}_{x,n}(x)$  by

$$\hat{f}_{x,n}(x) = \sum_{k \in \Lambda_l} \hat{\alpha}_{l,k} \phi_{l,k}(x) \tag{2}$$

where

$$\hat{\alpha}_{l,k} = \frac{1}{n} \sum_{i=1}^n (H_l \phi)_{l,k}(x) Y_i \tag{3}$$

$$(H_l \phi)(y) = \frac{1}{2\pi} \int \exp(it y) \frac{\tilde{\phi}(t)}{\tilde{g}(-2^l t)} dt \tag{4}$$

Such an estimator is standard in nonparametric estimation via wavelets. For a survey on wavelet linear estimators in various density models, we refer to [25]. Note that by Plancherel formula, we have

$$\alpha_{l,k} = \frac{1}{2\pi} \int \tilde{f}(t) \overline{\tilde{\phi}_{l,k}(t)} dt = \frac{1}{2\pi} \int \frac{\tilde{f}(t)}{\tilde{g}(t)} \tilde{\phi}_{l,k}(t) dt$$

In 1999, Pensky and Vidakovic [26] investigate Meyer wavelet estimation over Sobolev spaces and  $L_2(R)$  risk under moderately and severely ill-posed noises. Three years later, Fan and Koo [2] extend those works to Besov spaces, but the given estimator is not computable since it depends on an integral in the frequency domain that cannot be calculated in practice. It should be pointed out that, by using different method, Lounici and Nickl [27] study wavelet optimal estimation over Besov spaces  $B_{\infty,\infty}^s$  and  $L_\infty$  risk under both noises. In [3], wavelet optimal estimation is provided over and risk under moderately ill-posed noise. Furthermore in 2014, Li and Liu [28] considered the wavelet estimation for random samples with moderately ill-posed noise.

Our work is related to the paper of Geng and Wang [21], since our estimator is similar and we borrow a useful Lemma from that study. Geng and Wang [21] prove that, under mild conditions on the family of wavelets, the estimators are shown to be  $L_p(1 \leq p \leq \infty)$ -consistent for additive noise model. We extend thier result to certain class of dependent observation and prove that the mean integrated squared error of linear wavelet estimator developed by [29] attains the standard rate of convergence *i.e.* the optimal one in the *i.i.d.* case.

### 3. Optimality Results

The main result of the paper is the upper bound for the mean integrated square error of the wavelet estimator  $f(x)$ , which is defined as usual by

$$MISE(f(x)) = E \int_R (\tilde{f}(x) - f(x))^2 dx$$

We refer to [24] and [30] for a detailed coverage of wavelet theory in statistics. The asymptotic performance of our estimator is evaluated by determining an upper bound of the MISE over Besov balls. It is obtained as sharp as possible and coincides with the one related to the standard *i.i.d.* framework.

**Theorem 3.1.** Consider  $\phi$  as Meyer scaling function,  $\tilde{g}(t) \succeq (1+|t|^2)^{-\delta/2}$ , ( $\beta \succeq 0$ ) and  $\hat{f}$  in (2). We suppose

a) there exists constants  $\gamma \in (0,1)$  and  $C > 0$  such that

$$\sum_{k=0}^{2^l-1} \sum_{m=0}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_i\right) \preceq C 2^{2l\delta}, \quad (5)$$

b) for any  $m \in Z$ , let  $h_{y_0, y_m}(x, y)$  be the joint distribution of  $(Y_0, Y_m)$ , then there exists a constant  $C > 0$  such that

$$\sup_{m \in Z} \sup_{(x,y) \in [0,1]^2} \left| h_{y_0, y_m}(x, y) - h(x)h(y) \right| \preceq C, \quad \sup_x f_X(x) \preceq C$$

Let  $f \in B_{p,r}^s(M)$ , with  $s > 0$ ,  $p, r \succeq 1$ , with  $2^l = n^{1/(2s+2\delta+1)}$ . Then there exists a constant  $C > 0$  such that

$$\text{MISE}(f(x)) \preceq C n^{-2s/(2s+2\delta+1)}$$

Naturally, the rate of convergence in Theorem 4.1 is obtained to be as sharp as possible.

## 4. Applications

The three following subsections investigate separately the strong mixing case, the  $\rho$ -mixing case and the  $\beta$ -mixing case, which occur in a large variety of applications.

### 4.1. Application to the Strong Mixing Dependence

We define the  $m$ -th strong mixing coefficient of  $(V_i)_{i \in Z}$  by

$$\alpha_m = \sup_{(A,B) \in \sigma_{-\infty,0} \times \sigma_{m,\infty}} |p(AB) - p(A)p(B)|$$

where  $\sigma_{-\infty,0}$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $\dots, V_{-1}, V_0$  and  $\sigma_{m,\infty}$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $V_m, V_{m+1}, \dots$ . We say that  $(V_i)_{i \in Z}$  is strong mixing if and only if  $\lim_{m \rightarrow \infty} \alpha_m = 0$ .

Applications on strong mixing can be found in [15] [31] and [32]. Among various mixing conditions used in the literature,  $\alpha$ -mixing has many practical applications. Many stochastic processes and time series are known to be  $\alpha$ -mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with exponential mixing coefficients. The  $\alpha$ -mixing dependence is reasonably weak; it is satisfied by a wide variety of models including Markov chains, GARCH-type models and discretely observed discussions.

**Proposition 4.1.** Consider the strong mixing case as defined above. Suppose that there exist two constants  $\gamma \in (0,1)$ ,  $C > 0$  such that, for any integer  $m$ ,

$$\sum_{m=1}^{\infty} m^\gamma \alpha^m \preceq C,$$

then

$$\sum_{k=0}^{2^l-1} \sum_{m=0}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_i\right) \preceq C 2^{2l\delta}.$$

### 4.2. Application to the $\rho$ -Mixing Dependence

Let  $(Y_i)_{i \in Z}$  be a strictly stationary random sequence. For any  $m \in Z$ , we define the  $m$ -the maximal correlation coefficient of  $(V_i)_{i \in Z}$  by  $\rho$ -mixing:

$$\rho_{(m)} = \sup_{(U,V) \in L^2(\sigma_{-\infty,0}) \times L^2(\sigma_{m,\infty})} |\text{corr}(U,V)|$$

where  $\sigma_{-\infty,0}$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $\dots, Y_{-1}, Y_0$  and  $\sigma_{m,\infty}$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $Y_m, Y_{m+1}, \dots$ . We say  $(Y_i)_{i \in \mathbb{Z}}$  is  $\rho$ -mixing if and only if  $\lim_{m \rightarrow \infty} \rho_m = 0$ .

**Proposition 4.2.** Consider the  $\rho$ -mixing case as defined above. Furthermore, there exist two constants  $C > 0$  such that, for any integer  $m$ ,

$$\sum_{m=1}^{\infty} \rho_{(m)} \preceq C,$$

then

$$\sum_{k=0}^{2^l-1} \sum_{m=0}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_i\right) \preceq C 2^{2l\delta}.$$

### 4.3. Application to the $\beta$ -Mixing Dependence

Let  $(Y_i)_{i \in \mathbb{Z}}$  be a strictly stationary random sequence. For any  $m \in \mathbb{Z}$ , we define the  $m$ -th  $\beta$ -mixing coefficient of  $(Y_i)_{i \in \mathbb{Z}}$  by,

$$\beta_m = \frac{1}{2} \sup_{(A_i)_{i \in I}, (B_i)_{i \in J} \in \sigma_{-\infty,0}^Y \times \sigma_{m,\infty}^Y} \sum_{i \in I} \sum_{j \in J} \left| P(A_i \cap B_j) - P(A_i)P(B_j) \right|$$

where the supremum is taken over all finite partitions  $(A_i)_{i \in I}$  and  $(B_i)_{i \in J}$  of  $\Omega$ , which are respectively,  $\sigma_{-\infty,0}^Y$  and  $\sigma_{m,\infty}^Y$  are measurable,  $\sigma_{-\infty,0}^Y$  is the  $\sigma$ -algebra generated by  $\dots, Y_{-1}, Y_0$  and  $\sigma_{m,\infty}^Y$  is the one generated by  $Y_m, Y_{m+1}, \dots$ . We say that  $(Y_i)_{i \in \mathbb{Z}}$  is  $\beta$ -mixing if and only if  $\lim_{m \rightarrow \infty} \beta_m = 0$ .

Full details can be found in e.g. [29] [31] [33] and [34].

**Proposition 4.3.** Consider the  $\beta$  mixing case as defined above. Furthermore, there exist two constants  $C > 0$  such that, for any integer  $m$ ,

$$\sum_{m=1}^{\infty} \beta_{(m)} \preceq C.$$

then

$$\sum_{k=0}^{2^l-1} \sum_{m=0}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_i\right) \preceq C 2^{2l\beta}.$$

## 5. Proofs

In this section, we investigate the results of Section 3 under the assumptions of Section 4.

Moreover,  $C$  denotes any constant that does not depend on  $l, k$  and  $n$ .

**Proof of Theorem 3.1.** Since we set  $f \in L^2([0;1])$ , we have

$$E\left(\int (\hat{f}(x) - f(x))^2 dx\right) = \sum_{k \in \Lambda_l} (\hat{\alpha}_{l,k} - \alpha_{l,k})^2 + \sum_{j=l}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}^2$$

Following the lines of Geng and Wang [21], with Plancherel formula, it is easy to say  $\hat{\alpha}_{l,k}$  is the unbiased estimation of  $\alpha_{l,k}$ , furthermore

$$\text{Var}(\hat{\alpha}_{l,k}) = \frac{1}{n} \text{Var}\left((H_l \phi)_{l,k}(x) Y_i\right) + \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_i, (H_l \phi)_{l,k}(x) Y_j\right) = S_1 + S_2 \quad (7)$$

where

$$S_1 = \frac{1}{n} \text{Var}\left((H_l \phi)_{l,k}(x) Y\right)$$

and

$$\begin{aligned}
 S_2 &= \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_i, (H_l \phi)_{l,k}(x) Y_j\right) \\
 \text{Var}\left((H_l \phi)_{l,k}(x) Y\right) &\leq E\left|\left((H_l \phi)_{l,k}(x) Y\right)^2\right| = \int_{-\infty}^{\infty} \left(\left((H_l \phi)_{l,k}(x) y\right)^2 f_Y(y)\right) dy \\
 &\leq 2\pi \int_{\mathbb{R}} \left| \frac{\tilde{\phi}(t)}{\tilde{g}(-2^l t)} \right|^2 dt \leq \int \left| (1+|2^l t|)^{\delta/2} \tilde{\phi}(t) \right|^2 dt \leq 2^{2l\delta}
 \end{aligned} \tag{8}$$

on the other hand, it follows from the stationarity of  $(Y_i)_{i \in \mathbb{Z}}$  that

$$\begin{aligned}
 &\sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left((H_l \phi)_{l,k}(x) Y_i, (H_l \phi)_{l,k}(x) Y_j\right) \\
 &= \sum_{m=1}^n (n-m) \left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right| \\
 &\leq n \sum_{m=1}^n \left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right| = T_1 + T_2
 \end{aligned} \tag{9}$$

where

$$\begin{aligned}
 T_1 &= n \sum_{m=1}^{m=2^l-1} \left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right| \\
 T_2 &= n \sum_{m=2^l}^n \left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right|
 \end{aligned}$$

For upper bound of  $T_1$ , one can only consider the change of variables  $y = 2^j x - k$ , and we obtain

$$\begin{aligned}
 &\left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right| \\
 &\leq \iint |f(y_0, y_m) - f(y_0) f(y_m)| \left| (H_l \phi)_{l,k}(y_0) \right| \left| (H_l \phi)_{l,k}(y_m) \right| dy_0 dy_m \\
 &\leq \left( \int \left| (H_l \phi)_{l,k}(y) \right| dy \right)^2
 \end{aligned}$$

By (6) and inequality obtained in Lemma 6 in [2], we have,

$$\left| \text{Cov}\left((H_l \phi)_{l,k}(x) Y_0, (H_l \phi)_{l,k}(x) Y_m\right) \right| \leq 2^{l(2\delta-1)}$$

Therefore

$$T_1 \leq Cn2^j 2^{l(2\delta-1)} \leq Cn2^{2l\delta} \tag{10}$$

It follows from (5) that

$$T_2 \leq Cn2^{2l\delta} \tag{11}$$

Therefore, combining (7) to (11), we obtain

$$\sum_{k \in \Lambda_l} (\hat{\alpha}_{l,k} - \alpha_{l,k})^2 \leq C2^l 2^{2l\delta} \frac{1}{n} \tag{12}$$

On the other hand, as we define  $f \in B_{p,r}^s(M)$  and since for  $p \geq 2$ ,  $B_{p,r}^s(M) \subseteq B_{p,\infty}^s(M)$ , then there exists a constant  $C > 0$ , such that

$$\sum_{j=k \in \Lambda_l} \sum_{j=k \in \Lambda_l} \beta_{j,k}^2 \leq C2^{-2ls} \tag{13}$$

It follows from (13) and (14) and the assumption on  $2^l$  that

$$MISE(f(x)) \leq Cn^{-2s/(2s+2\delta+1)}$$

Now the proof of Theorem 3.1 is complete.

**Proof of Proposition 5.1.** We apply the Davydov inequality for strongly mixing processes (see [29]); for any  $q \in (0,1)$ , we have

$$\begin{aligned} & \left| Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_m\right) \right| \leq C\alpha_m^q \left( E\left| (H_l\phi)_{l,k}(x)Y_0 \right|^{2/(1-\gamma)} \right)^{1-\gamma} \\ & \leq C\alpha_m^q \left( \sup_{x \in [0,1]} \left| (H_l\phi)_{l,k}(x)Y_0 \right| \right)^{2\gamma} E\left| (H_l\phi)_{l,k}(x)Y_0 \right|^{2(1-\gamma)} \leq C2^{2\delta l} 2^{\gamma l} \alpha_m^\gamma \end{aligned} \quad (14)$$

Since we have  $\sup_{x \in [0,1]} \left| (H_l\phi)_{l,k}(x)Y_0 \right| \leq C2^{l(\delta+1/2)} \frac{1}{(1+|2^l y - k|)^2}$  and

$$E\left| (H_l\phi)_{l,k}(x)Y_0 \right|^2 \leq C \int \left| (H_l\phi)_{l,k}(x)Y_0 \right|^2 dy \leq C2^{l2\delta} \int \frac{1}{(1+|u|)^2} du \leq C2^{2l\delta}. \quad (15)$$

therefore

$$\begin{aligned} & \sum_{m=2^l}^n \left| Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_m\right) \right| \\ & \leq C2^{2\delta l} 2^{\gamma l} \sum_{m=2^l}^n \alpha_m^\gamma \leq C2^{2l\delta} \sum_{m=2^l}^n m^\gamma \alpha^\gamma \leq C2^{2l\delta} \sum_{m=1}^{\infty} m^\gamma \alpha^\gamma \leq C2^{2l\delta}. \end{aligned} \quad (16)$$

Now the proof is finished by (14), (15) and (16).

**Proof of Proposition 5.2.** Applying the covariance inequality for  $\rho$ -mixing processes (see Doukahn [32]), we have

$$\begin{aligned} & \left| Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_i\right) \right| \\ & \leq 2\rho_i \left( E\left| (H_l\phi)_{l,k}(x)Y_0 \right|^2 \right)^{1/2} \left( E\left| (H_l\phi)_{l,k}(x)Y_i \right|^2 \right)^{1/2}, \\ & \left| Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_i\right) \right| \leq C\rho_i 2^{2lm} \end{aligned}$$

Hence by the same technique we use in (8), we obtain

$$\sum_{m=2^l}^n \left| Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_m\right) \right| \leq C2^{2l\beta} \sum_{i=1}^m \rho(i) \leq C2^{2l\beta}$$

**Proof of Proposition 5.3.** Since  $(S_i)_{i \in \mathbb{Z}}$  is  $\beta$ -mixing, for any bounded function  $g$  ([25], equation line 12, p. 479 and Lemma 4.2 with  $p = 1$ ) implies that

$$\sum_{m=0}^n C(g(S_m), g(S_0)) \leq 2 \int_0^1 b(x) g^2(x) f_s(x) dx$$

where  $b$  is a function such that  $\int_0^1 b(x) f_s(x) dx \leq \sum_{m=0}^{\infty} \beta_m < C$ . Following the lines of Geng and Wang [21], we obtain

$$\begin{aligned} & \sum_{k=0}^{2^l-1} \sum_{m=0}^n Cov\left((H_l\phi)_{l,k}(x)Y_0, (H_l\phi)_{l,k}(x)Y_i\right) \\ & \leq \int b(x) f_s(x) \sum_{k=0}^{2^l-1} \left| (H_l\phi)_{l,k}(x) \right|^2 \leq 2^{2l\delta} \int b(x) f_s(x) \leq C2^{2l\delta} \end{aligned}$$

## References

- [1] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006) Penalized Contrast Estimator for Density Deconvolution. *The Canadian Journal of Statistics*, **34**, 431-452. <http://dx.doi.org/10.1002/cjs.5550340305>
- [2] Fan, J. and Koo, J.Y. (2002) Wavelet Deconvolution. *IEEE Transactions on Information Theory*, **48**, 734-747. <http://dx.doi.org/10.1109/18.986021>
- [3] Carroll, R.J. and Hall, P. (1988) Optimal Rates of Convergence for Deconvolving a Density. *Journal of the American Statistical Association*, **83**, 1184-1186. <http://dx.doi.org/10.1080/01621459.1988.10478718>
- [4] Johannes, J. (2009) Deconvolution with Unknown Error Distribution. *Annals of Statistics*, **37**, 2301-2323. <http://dx.doi.org/10.1214/08-AOS652>
- [5] Meister, A. (2004) On the Effect of Misspecifying the Error Density in a Deconvolution. *Canadian Journal of Statistics*, **32**, 439-449.
- [6] Li, T. and Vuong, Q. (1998) Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators. *Journal of Multivariate Analysis*, **65**, 139-165. <http://dx.doi.org/10.1006/jmva.1998.1741>
- [7] Delaigle, A., Hall, P. and Meister, A. (2008) On Deconvolution with Repeated Measurements. *Annals of Statistics*, **36**, 665-685. <http://dx.doi.org/10.1214/009053607000000884>
- [8] Neumann, M.H. (2007) Deconvolution from Panel Data with Unknown Error Distribution. *Journal of Multivariate Analysis*, **98**, 1955-1968. <http://dx.doi.org/10.1016/j.jmva.2006.09.012>
- [9] Odiachi, P. and Prieve, D. (2004) Removing the Effects of Additive Noise in Term. *Journal of Colloid and Interface Science*, **270**, 113-122. [http://dx.doi.org/10.1016/S0021-9797\(03\)00548-4](http://dx.doi.org/10.1016/S0021-9797(03)00548-4)
- [10] Devroye, L. (1989) Consistent Deconvolution in Density Estimation. *Canadian Journal of Statistics*, **17**, 235-239. <http://dx.doi.org/10.2307/3314852>
- [11] Fan, J. (1991) On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Annals of Statistics*, **19**, 1257-1272. <http://dx.doi.org/10.1214/aos/1176348248>
- [12] Liu, M.C. and Taylor, R.L. (1989) A Consistent Nonparametric Density Estimator for the Deconvolution Problem. *Canadian Journal of Statistics*, **17**, 427-438. <http://dx.doi.org/10.2307/3315482>
- [13] Masry, E. (1991) Multivariate Probability Density Deconvolution for Stationary Random Processes. *IEEE Transactions on Information Theory*, **37**, 1105-1115. <http://dx.doi.org/10.1109/18.87002>
- [14] Stefanski, L. and Carroll, R.J. (1990) Deconvoluting Kernel Density Estimators. *Statistics*, **21**, 169-184. <http://dx.doi.org/10.1080/02331889008802238>
- [15] Zhang, C.-H. (1990) Fourier Methods for Estimating Mixing Densities and Distributions. *Annals of Statistics*, **18**, 806-831. <http://dx.doi.org/10.1214/aos/1176347627>
- [16] Hesse, C.H. (1999) Data-Driven Deconvolution. *Journal of Nonparametric Statistics*, **10**, 343-373. <http://dx.doi.org/10.1080/10485259908832766>
- [17] Cator, E.A. (2001) Deconvolution with Arbitrarily Smooth Kernels. *Statistics & Probability Letters*, **54**, Article ID: 205214. [http://dx.doi.org/10.1016/s0167-7152\(01\)00083-9](http://dx.doi.org/10.1016/s0167-7152(01)00083-9)
- [18] Delaigle, A. and Gijbels, I. (2004) Bootstrap Bandwidth Selection in Kernel Density Estimation from a Contaminated Sample. *Annals of the Institute of Statistical Mathematics*, **56**, 1947. <http://dx.doi.org/10.1007/BF02530523>
- [19] Koo, J.-Y. (1999) Logspline Deconvolution in Besov Space. *Scandinavian Journal of Statistics*, **26**, 73-86. <http://dx.doi.org/10.1111/1467-9469.00138>
- [20] Efromovich, S. (1997) Density Estimation for the Case of Super-Smooth Measurement Error. *Journal of the American Statistical Association*, **92**, 526-535. <http://dx.doi.org/10.1080/01621459.1997.10474005>
- [21] Geng, Z. and Wang, J. (2015) The Mean Consistency of Wavelet Density Estimators. *Journal of Inequalities and Applications*, **2015**, 111. <http://dx.doi.org/10.1186/s13660-015-0636-1>
- [22] Daubechies, I. (1992) Ten Lectures on Wavelets. *CBMS-NSF Regional Conference Series in Applied Mathematics*. <http://dx.doi.org/10.1137/1.9781611970104>
- [23] Meyer, Y. (1992) Wavelets and Operators. Cambridge University Press, Cambridge.
- [24] Hardle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1997) Wavelets, Approximation and Statistical Applications. Springer, New York.
- [25] Tsybakov, A.B. (2009) Introduction to Nonparametric Estimation. Springer, Berlin.
- [26] Pensky, M. and Vidakovic, B. (1999) Adaptive Wavelet Estimator for Nonparametric Density Deconvolution. *The Annals of Statistics*, **27**, 2033-2053.



- 
- [27] Lounici, K. and Nickl, R. (2011) Global Uniform Risk Bounds for Wavelet Deconvolution Estimators. *The Annals of Statistics*, **39**, 201-231. <http://dx.doi.org/10.1214/10-AOS836>
- [28] Li, R. and Liu, Y. (2014) Wavelet Optimal Estimations for a Density with Some Additive Noises. *Applied and Computational Harmonic Analysis*, **36**, 2.
- [29] Doukhan, P. (1994) Mixing. Properties and Examples. *Lecture Notes in Statistics*, **85**, Springer Verlag, New York.
- [30] Antoniadis, A. (1997) Wavelets in Statistics: A Review (with Discussion). *Journal of the Italian Statistical Society, Series B*, **6**, 97-144. <http://dx.doi.org/10.1007/BF03178905>
- [31] Bradley, R.C. (2005) Basic Properties of Strong Mixing Conditions. *A Survey and Some Open Questions, Probability Surveys*, **2**, 107-144.
- [32] Fryzlewicz, P. and Rao, S.S. (2011) Mixing Properties of ARCH and Time-Varying ARCH Processes. *Bernoulli*, **17**, 320-346. <http://dx.doi.org/10.3150/10-BEJ270>
- [33] Carrasco, M. and Chen, X. (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory*, **18**, 17-39. <http://dx.doi.org/10.1017/S0266466602181023>
- [34] Volkonskii, V.A. and Rozanov, Y.A. (1959) Some Limit Theorems for Random Functions I. *Theory of Probability and Its Applications*, **4**, 178-197. <http://dx.doi.org/10.1137/1104015>