AISM (2007) 59: 299–324
DOI 10.1007/s10463-006-0048-6

**Linyuan Li · Yimin Xiao**

# Mean integrated squared error of nonlinear wavelet-based estimators with long memory data

Received: 2 July 2004 / Revised: 14 November 2005 / Published online: 11 May 2006
© The Institute of Statistical Mathematics, Tokyo 2006

**Abstract** We consider the nonparametric regression model with long memory data that are not necessarily Gaussian and provide an asymptotic expansion for the mean integrated squared error (MISE) of nonlinear wavelet-based mean regression function estimators. We show this MISE expansion, when the underlying mean regression function is only piecewise smooth, is the same as analogous expansion for the kernel estimators. However, for the kernel estimators, this MISE expansion generally fails if an additional smoothness assumption is absent.

**Keywords** Mean integrated square error · Nonlinear wavelet-based estimator · Non-parametric regression · Long-range dependence · Hermite rank · Rates of convergence

## 1 Introduction

Consider nonparametric regression

$$Y_k = g(x_k) + \varepsilon_k, \quad k = 1, 2, \ldots, n, \tag{1}$$

where $x_k = k/n \in [0, 1]$, $\varepsilon_1, \ldots, \varepsilon_n$ are observational errors with mean 0 and $g$ is an unknown function to be estimated. Many authors have investigated various aspects of this model, under the assumptions that $\varepsilon_1, \ldots, \varepsilon_n, \ldots$ are i.i.d.

Research supported in part by the NSF grant DMS-0103939.

L. Li (✉)
Department of Mathematics and Statistics, University of New Hampshire,
Durham, NH 03824, USA
E-mail: linyuan@cisunix.unh.edu

Y. Xiao
Department of Statistics and Probability, Michigan State University,
Esat Lansing, MI 48824, USA
E-mail: xiao@stt.msu.edu

errors or a stationary process with short-range dependence such as classic ARMA processes (see, e.g., Hart, 1991; Tran et al. 1996; Truong and Patil, 2001); or a stationary Gaussian sequence with long-range dependence (see, e.g., Csörgö and Mielniczuk, 1995; Wang, 1996; Johnstone and Silverman, 1997; Johnstone, 1999); or a correlated and heteroscedastic noise sequence (Kovac and Silverman, 2000); or a correlated and nonstationary noise sequence (von Sachs and Macgibbon, 2000), just to mention a few. Regression models with long memory data are more appropriate for various phenomena in many fields which include agronomy, astronomy, economics, environmental sciences, geosciences, hydrology and signal and image processing. Moreover, many times series encountered in practice do not exhibit characteristics of Gaussian processes. For example, some economic time series, especially price series, are non-Gaussian to the extreme. Hence it is of interest to consider the regression model (1) with long-range dependent errors $\{\varepsilon_n\}$ that are not necessarily Gaussian.

Let $\{\varepsilon_k, k \geq 1\}$ be a stationary process with mean 0 and constant variance. Recall that $\{\varepsilon_k, k \geq 1\}$ is said to have long-range dependence or long memory, if $\sum_{k=1}^{\infty} |\rho(k)| = \infty$, where $\rho(k) = E(\varepsilon_1 \varepsilon_{1+k})$ is the autocovariance function of $\{\varepsilon_k\}$. This is the case if there exists $H \in (0, 1)$ such that

$$\rho(k) = k^{-H} L(k), \tag{2}$$

where $L(x)$ is a slowly varying function at $\infty$, i.e., for all $a > 0$,

$$\lim_{x \to \infty} \frac{L(ax)}{L(x)} = 1.$$

See Bingham, Goldie and Teugels (1987) for more information on slowly varying functions.

The literature on long-range dependence is very extensive, see, e.g., the monograph of Beran (1994) and the references cited therein. In particular, several authors have studied the kernel estimators of the regression function $g$ with long-range dependent errors. For example, Hall and Hart (1990) considered the model (1) with stationary errors for which $\rho(n) \sim C n^{-H}$ as $n \to \infty$ and established the convergence rates of mean regression function estimators. Csörgö and Mielniczuk (1995) have studied the weak convergence of the finite dimensional distributions and the suitably renormalized maximal deviations of the kernel estimators of $g$ in (1) with long-range dependent Gaussian errors. Robinson (1997) has established the central limit theorems for the kernel estimators of $g$ in (1) when the errors form a stationary martingale difference sequence. They all require that the regression function $g$ is continuously differentiable.

This paper is concerned with the asymptotic properties of wavelet-based estimators of mean regression function with long memory errors. In recent years, wavelet methods in nonparametric curve estimation have become a well-known and powerful technique. We refer to the monograph by Härdle, Kerkyacharian and Tsybakov (1998) for a systematic discussion of wavelets and their applications in statistics. The major advantage of the wavelet method is its adaptability to the degree of smoothness of the underlying unknown curve. These wavelet estimators typically achieve the optimal convergence rates over exceptionally large function spaces. For more information and related references, see Donoho, Johnstone, Kerkyacharian and Picard (1995); Donoho, et al. (1996) and Donoho and Johnstone (1998). Hall and Patil (1995, 1996a,b) also have demonstrated explicitly the extraordinary local adaptability of wavelet estimators in handling discontinuities. They showed that discontinuities of the unknown curve have a negligible effect on performance of nonlinear wavelet curve estimators. All of the above works are under the assumption that the errors are independent.

For correlated noise, Wang (1996), Johnstone and Silverman (1997) and Johnstone (1999) have examined the asymptotic properties of wavelet-based estimators of mean regression function with long memory Gaussian noise. They have shown that these estimators achieve minimax rates over wide range of function spaces. In this paper we consider a different approach by following the framework of Hall and Patil (1995). We focus on a fixed target function $g$ in (1), rather than describe performance uniformly over large classes of $g$'s, to identify more clearly the ways in which the choice of threshold affects the performance of a wavelet-based curve estimator (see Hall and Patil, 1996a,b). These results may be generalized to large classes of functions.

Specifically, we provide an asymptotic expansion for the mean integrated square error (MISE) of nonlinear wavelet-based mean regression function estimators with long memory data that are not necessarily Gaussian. More explicitly, we assume in (1) that $\varepsilon_k = G(\xi_k)$, where $\{\xi_k, k \geq 1\}$ is a stationary Gaussian sequence with mean 0 and variance 1 and $G$ is a function such that $EG(\xi_k) = 0$ and $EG^2(\xi_k) < \infty$. Under the assumption that the underlying mean regression function $g$ is only piecewise smooth, we show that the MISE expansion is of the same form as the analogous expansion for the kernel estimators. However, for the kernel estimators, this MISE expansion generally fails if the additional smoothness assumption is absent.

The rest of this paper is organized as follows. In the next section, we give some basic elements of the wavelet theory, provide nonlinear wavelet-based mean regression function estimators and recall some basic properties of the Hermite polynomials. The main results are described in Sect. 3, while their proofs are given in Sect. 4. Throughout we will use $C$ to denote a positive and finite constant whose value may change from line to line. Specific constants are denoted by $C_0, C_1, C_2$ and so on.

## 2 Notations and estimators

This section contains some facts about wavelets that will be used in the sequel. Let $\phi(x)$ and $\psi(x)$ be the father and mother wavelets, having the following properties: $\phi$ and $\psi$ are bounded and compactly supported; $\int \phi^2 = \int \psi^2 = 1$, $\nu_k \equiv \int y^k \psi(y)\,\mathrm{d}y = 0$ for $0 \leq k \leq r-1$ and $\nu_r = \int y^r \psi(y)\,\mathrm{d}y \neq 0$. Define

$$\phi_j(x) = p^{1/2}\phi(px - j), \quad \psi_{ij}(x) = p_i^{1/2}\psi(p_i x - j), \quad x \in \mathbb{R}$$

for arbitrary integers $p > 0$, $j \in \mathbb{Z}$ and $p_i = p2^i$, $\quad i \geq 0$. Then the system $\left\{\phi_j(x),\ \psi_{ij}(x), j \in \mathbb{Z}, i \geq 0\right\}$ satisfies

$$\int \phi_{j_1}\phi_{j_2} = \delta_{j_1 j_2}, \quad \int \psi_{i_1 j_1}\psi_{i_2 j_2} = \delta_{i_1 i_2}\delta_{j_1 j_2}, \quad \int \phi_{j_1}\psi_{i j_2} = 0, \tag{3}$$

where $\delta_{ij}$ denotes the Kronecker delta (i.e., $\delta_{ij} = 1$ , if $i = j$; and $\delta_{ij} = 0$, otherwise) and is an orthonormal basis for the space $L^2(\mathbb{R})$. For more information on wavelets, see Daubechies (1992) or Härdle, Kerkyacharian and Tsybakov (1998).

In our regression model (1), the mean function $g$ is supported on the unit interval [0, 1], thus we can select an index set $\Lambda \subset \mathbb{Z}$ and modify some of $\psi_{ij}(x),\ i, j \in \mathbb{Z}$, such that $\left\{\psi_{ij}(x),\ i, j \in \Lambda\right\}$ forms a complete orthonormal basis for $L^2[0, 1]$. We refer to Cohen, Daubechies and Vial (1993) for more details on wavelets on the interval. Hence, without loss of generality, we may and will assume that $\phi$ and $\psi$ are compactly supported on [0, 1]. We also assume that both $\phi$ and $\psi$ satisfy a uniform Hölder condition of exponent 1/2, i.e.,

$$|\psi(x) - \psi(y)| \le C|x - y|^{1/2}, \quad \text{for all } x, y \in [0, 1]. \tag{4}$$

Daubechies (1992, Chap. 6) provides examples of wavelets satisfying these conditions.

For every function $g$ in $L^2([0, 1])$, we have the following wavelet expansion:

$$g(x) = \sum_{j=0}^{p-1} b_j \phi_j(x) + \sum_{i=0}^{\infty} \sum_{j=0}^{p_i-1} b_{ij} \psi_{ij}(x), \quad x \in [0, 1], \tag{5}$$

where

$$b_j = \int g\phi_j, \quad b_{ij} = \int g\psi_{ij}$$

are the wavelet coefficients of the function $g$ and the series in (5) converges in $L^2([0, 1])$.

We propose a nonlinear wavelet estimator for $g(x)$:

$$\widehat{g}(x) = \sum_{j=0}^{p-1} \widehat{b}_j \phi_j(x) + \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} \widehat{b}_{ij} I\left(|\widehat{b}_{ij}| > \delta_i\right) \psi_{ij}(x), \tag{6}$$

where $\delta_i > 0$ is a level-dependent "threshold", $q \ge 1$ is another smoothing parameter and the wavelet coefficients $\widehat{b}_j$ and $\widehat{b}_{ij}$ are defined as follows:

$$\widehat{b}_j = \frac{1}{n} \sum_{k=1}^{n} Y_k \phi_j(x_k), \quad \widehat{b}_{ij} = \frac{1}{n} \sum_{k=1}^{n} Y_k \psi_{ij}(x_k). \tag{7}$$

Let $\{\xi_k, k \ge 1\}$ be a stationary sequence of Gaussian random variables with mean zero and unit variance. We assume that

$$r(k) \mathrel{\hat{=}} E\left(\xi_1 \xi_{1+k}\right) \sim C_0 k^{-\alpha}, \quad \text{as } k \to \infty, \tag{8}$$

where $\alpha \in (0, 1)$ and $C_0 > 0$ is a constant and $a_k \sim b_k$ means that $a_k/b_k \to 1$ as $k \to \infty$. Here we have taken the slowing varying function $L$ in (2) to be the constant $C_0$ to simplify the presentations of our results. With a little modification, one can show that Theorems 3.1 and 3.2 still hold under the more general assumption (2). A typical example of stationary Gaussian sequences satisfying (8) is the fractional Gaussian noise:

$$\xi_k = B^H(k+1) - B^H(k), \quad (k = 0, 1, 2, \dots),$$

where $B^H = \left\{B^H(t), t \in \mathbb{R}\right\}$ is a real-valued fractional Brownian motion, that is $B^H$ is a centered Gaussian process with $B^H(0) = 0$ and

$$E\left(B^H(s)B^H(t)\right) = \frac{1}{2}\left(|s|^{2\alpha} + |t|^{2\alpha} - |t-s|^{2\alpha}\right), \quad \forall s,\ t \in \mathbb{R}.$$

In this case, it is easy to verify that if $H \neq 1/2$, then $r(k) \sim H(2H-1)k^{2H-2}$ as $k \to \infty$. Hence the fractional Gaussian noise satisfies (8) with $\alpha = 2-2H$ and has long-range dependence if and only if $H \in (1/2, 1)$. See Samorodnitsky and Taqqu (1994, Sect. 7.2) for more information on fractional Brownian motion and fractional Gaussian noise. Long-range dependent stationary Gaussian sequences can also be constructed as moving averages of i.i.d. standard normal random variables; see, e.g., Beran (1994).

For an integer $m \geq 1$, if the index $\alpha$ in (8) satisfies $0 < \alpha m < 1$, then, using the notation of Taqqu (1975, 1977), we write $\{\xi_k, k \geq 1\} \in (m)(\alpha, C_0)$.

Now we recall the notion of Hermite rank from Taqqu (1975, 1977). Let $\xi$ be an $N(0, 1)$ random variable and define

$$\mathcal{G} = \left\{G :\ EG(\xi) = 0,\ EG^2(\xi) < \infty\right\}.$$

Then $\mathcal{G}$ is a subspace of the Hilbert space

$$L^2\left(\mathbb{R}, \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\right) \hat{=} \left\{G :\ \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} G^2(x)\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x < \infty\right\}.$$

It is well known (see, e.g., Rozanov, 1967) that the Hermite polynomials

$$H_k(x) = (-1)^k \mathrm{e}^{\frac{x^2}{2}} \frac{\mathrm{d}^k}{\mathrm{d}x^k}\,\mathrm{e}^{-\frac{x^2}{2}} \quad (k \geq 0),$$

form a complete orthogonal system of functions of $L^2\left(\mathbb{R}, \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\right)$ and they satisfy the identity

$$E\left[H_k(\xi)H_\ell(\xi)\right] = \delta_{k\ell}\,k!, \tag{9}$$

where $\xi$ is a standard normal random variable. The first few Hermite polynomials are $H_0(x) \equiv 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$ and $H_3(x) = x^3 - 3x$.

Every function $G \in \mathcal{G}$ can be expanded in terms of the Hermite polynomials

$$G(x) = \sum_{k=0}^{\infty} \frac{J(k)}{k!}\,H_k(x),$$

where $J(k) = E\left(G(\xi)H_k(\xi)\right), \forall k \geq 0$, and the above series converges in $L^2\left(\mathbb{R}, \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\right)$. It is clear from (9) that $EG^2(\xi) = \sum_{k=0}^{\infty} \frac{J^2(k)}{k!} < \infty$.

For every $G \in \mathcal{G}$, the Hermite rank of $G$ is defined by

$$m_{_G} = \min\{k \geq 0 : \ J(k) \neq 0\}.$$

Since $J(0) = EG(\xi) = 0$, we have $m_{_G} \geq 1$ for all $G \in \mathcal{G}$. When no confusion can arise, we will omit the subscript $G$ and simply write the Hermite rank of $G$ as $m$.

As examples, we mention that for every integer $k \geq 0$, $G(x) = x^{2k+1}$ has Hermite rank 1; and the function $G(x) = x^{2k} - E(\xi^{2k})$ has Hermite rank 2. The Hermite polynomial $H_k(x)$ has Hermite rank $k$.

We will need the class of functions

$$\mathcal{G}_m = \{G \in \mathcal{G} : \ G \text{has Hermit rank } m\}.$$

For every $G \in \mathcal{G}_m$ and $\{\xi_n, n \geq 1\} \in (m)(\alpha, C_0)$, let $\varepsilon_n = G(\xi_n)$ for $n \geq 1$. Then $\{\varepsilon_n, n \geq 1\}$ is a stationary sequence. In order to verify its long-range dependence, we make use of the following formula

$$E\left[H_k(\xi_1)H_\ell(\xi_{1+n})\right] = \delta_{k\ell}\, k!\, r(n)^k \tag{10}$$

(see Rozanov, 1967, p. 183) to get

$$\begin{aligned} E\left(\varepsilon_1\varepsilon_{1+n}\right) &= \sum_{q=m}^{\infty} \frac{J^2(q)}{q!}\, r^q(n) \\ &\sim \frac{J^2(m)}{m!}\, r^m(n), \quad \text{as } n \to \infty, \end{aligned} \tag{11}$$

where the last statement follows from (8); see also Taqqu (1975, p. 293). Hence $\{\varepsilon_n, n \geq 1\}$ is long-range dependent with index $m\alpha$. Note that in this paper, we assume $0 < m\alpha < 1$. The case of $m\alpha \geq 1$ requires different methods and will be dealt elsewhere.

## 3 Main results

Throughout this paper, we will assume the following conditions [denoted by (SP)] are satisfied

(1) The errors $\{\varepsilon_k, k \geq 1\}$ in (1) are given by $\varepsilon_k = G(\xi_k)$, where $G \in \mathcal{G}_m$ and $\{\xi_k, k \geq 1\} \in (m)(\alpha, C_0)$.

(2) The smoothing parameters $p$, $q$, and $\delta_i$ are functions of $n$. We assume that $p \to \infty$, $q \to \infty$ as $n \to \infty$, and for every $i = 1, 2, \ldots, q-1$,

$$p_i\delta_i^2 \to 0, \quad p_i^{2r+1}\delta_i^2 \to \infty, \quad \delta_i^2 \geq \frac{C_1(4e)^m(\ln n)^{m+1}}{m^m n^{m\alpha} p_i^{1-m\alpha}},$$

where

$$C_1 = \frac{C_0 J^2(m)}{m!} \iint |x-y|^{m\alpha}\psi(x)\psi(y)\,\mathrm{d}x\,\mathrm{d}y. \tag{12}$$

Observe that, since $0 < m\alpha < 1$, the Fourier transform of the function $s \mapsto |s|^{-m\alpha}$ and the Plancherel's theorem yield

$$\iint |x-y|^{-m\alpha}\psi(x)\psi(y)\mathrm{d}x\mathrm{d}y = C_{m\alpha}\int_{\mathbb{R}} \frac{1}{|u|^{1-m\alpha}}\left|\widehat{\psi}(u)\right|^2 \mathrm{d}u > 0,$$

where $C_{m\alpha} > 0$ is a constant depending on $m\alpha$ only and $\widehat{\psi}$ is the Fourier transform of $\psi$. Hence $C_1$ is a positive and finite constant.

**Theorem 3.1** *If, in addition to the conditions on $\phi$ and $\psi$ stated in Sect. 2 and the condition (SP), we assume that the $r$-th derivative $g^{(r)}$ is continuous and bounded. Then, as $n \to \infty$,*

$$E\left|\int (\hat{g}-g)^2 - \left\{C_2\left(n^{-1}p\right)^{m\alpha} + p^{-2r}\kappa^2\left(1-2^{-2r}\right)^{-1}\int g^{(r)^2}\right\}\right|$$
$$= o\left(\left(n^{-1}p\right)^{m\alpha} + p^{-2r}\right), \qquad (13)$$

*where $\kappa = (r!)^{-1}\nu_r$ and $C_2$ is a positive and finite constant defined by*

$$C_2 = \frac{C_0 J^2(m)}{m!}\iint |x-y|^{m\alpha}\phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y. \qquad (14)$$

*Remark 3.1* Because of the long-range dependence, our choice of the thresholds must be level-dependent and depend on the unknown long memory parameter $\alpha$ and Hermite rank $m$. Hence, our result is mainly of theoretical significance and these parameters are assumed to be known constants. In practice, we need first to estimate these parameters. Wang (1996, p480) and Johnstone and Silverman (1997, p340) provide simple methods to estimate the long memory parameter $\alpha$. Delbeke and Van Assche (1998) and Abry and Veitch (1998) also provide wavelet-based, statistically and computationally efficient estimators of $\alpha$ based on the wavelet coefficients and show they are unbiased, consistent and has asymptotically a normal distribution. Thus, we assume that the parameter $\alpha$ is estimated and treated as known. An alternative way of choosing the threshold is the following: from Lemma 4.2 in the next section, the threshold in (SP) satisfy $\delta_i^2 \geq \sigma_i^2(2e)^m(\ln n)^{m+1}m^{-m}$, where $\sigma_i^2 = \mathrm{var}\left(\hat{b}_{ij}\right) = C_1 n^{-m\alpha}p_i^{-1+m\alpha}$. This noise variance $\sigma_i^2$ at each level $i$ can be estimated from the data, for example, the robust median absolute deviation estimator $\hat{\sigma}_i = \mathrm{MAD}\{\hat{b}_{ij}, j = 0, 1, \ldots, p_i - 1\}/0.6745$. Hence, we can also treat $\sigma_i^2$ as known (see Johnstone and Silverman, 1997 for more details).

In Theorem 3.1, we have assumed that the mean regression function $g$ is $r$-times continuously differentiable for simplicity and convenience of the exposition. However, if $g^{(r)}$ is only piecewise continuous, Theorem 3.1 still holds, as given in the following:

**Theorem 3.2** *In addition to the conditions on $\phi$ and $\psi$ stated in Sect. 2, we assume that the $r$-th derivative $g^{(r)}$ is only piecewise smooth, i.e., there exist points $x_0 = 0 < x_1 < x_2 < \cdots < x_N < 1 = x_{N+1}$ such that the first $r$ derivatives of $g$ exist and are bounded and continuous on $(x_i, x_{i+1})$ for $0 \leq i \leq N$, with left- and right-hand limits. In particular, $g$ itself may be only piecewise continuous. Also assume that condition (SP) holds and $p_q^{2r+m\alpha}n^{-2rm\alpha} \to \infty$. Then (13) still holds.*

*Remark 3.2* Truong and Patil (2001) have considered wavelet estimation of mean regression function $\mu(x) = E(Y_0|X_0 = x)$ based on stationary sequences of random variables $(X_i, Y_i), i = 0, \pm1, \pm2, \ldots$. They constructed wavelet estimator of $\mu$ through $\hat{\mu} = \hat{g}/\hat{f}$, where $\hat{f}$ and $\hat{g}$ are wavelet estimators of density function of $X_0$ and $g(x) = \int yh(y, x)\mathrm{d}y$, where $h(y, x)$ is the joint density function of $(Y_0, X_0)$. One of results in Truong and Patil (2001) is that they derived estimator's MISE formulae $\int \mathbb{E}\left(\hat{\mu} - \mu\right)^2 \sim k_1 n^{-1} p + k_2 p^{-2r}$, where $k_1, k_2$ are constants. Since these stationary sequences are short range dependent, its MISE is analogous to those of Hall and Patil (1995, 1996a,b) for the independent case. However, in our case for long range dependence data, the MISE formulae is $\int \mathbb{E}\left(\hat{g} - g\right)^2 \sim k_3 \left(n^{-1} p\right)^{m\alpha} + k_2 p^{-2r}$, which depends on long memory parameter $\alpha$ and *Hermite rank* of $G$, and thus it is different from those of Hall and Patil (1995, 1996a,b). Wang (1996) and Johnstone and Silverman (1997) considered wavelet estimator of mean regression function with long memory Gaussian error and derived the minimax convergence rate $n^{-2r\alpha/(2r+\alpha)}$. In our paper, we considered that random error is a nonlinear function of Gaussian error. For the Gaussian error special case, the *Hermite rank* of $G$ is 1 (i.e., $m = 1$). In this case, if our smoothing parameter $p$ is chosen of size $n^{\alpha/(2r+\alpha)}$, then the convergence rate of our MISE is $n^{-2r\alpha/(2r+\alpha)}$, which is the same as those in Wang (1996) and Johnstone and Silverman (1997).

*Remark 3.3* Hall and Hart (1990) considered kernel estimator in fixed-design nonparametric regression when error is Gaussian long memory process, giving a similar asymptotic expansion for MISE. Robinson (1997) considered kernel nonparametric regression estimator when error is long memory moving average, providing a central limit theorem and a similar asymptotic expansion for MISE as well. However, they all assume that the regression function $g$ is continuously differentiable. Our result is stronger than the traditional asymptotic expansion for MISE. In particular, (13) implies a wavelet version of the MISE expansion:

$$E \int (\widehat{g} - g)^2 \sim C_2 \left(n^{-1} p\right)^{m\alpha} + p^{-2r} \kappa^2 \left(1 - 2^{-2r}\right)^{-1} \int g^{(r)^2}, \quad \text{as } n \to \infty.$$

For kernel estimators, the above expansion usually fails without the assumption that $g$ is $r$-times continuously differentiable.

## 4 Proofs

Observing that the orthogonality (3) implies

$$\int (\widehat{g} - g)^2 = \sum_{j=0}^{p-1} \left(\widehat{b}_j - b_j\right)^2 + \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} \left(\widehat{b}_{ij} - b_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right)$$
$$+ \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|\widehat{b}_{ij}| \le \delta_i\right) + \sum_{i=q}^{\infty} \sum_{j=0}^{p_i-1} b_{ij}^2, \tag{15}$$

we will break the proofs of Theorem 3.1 into several parts. The basic ideas of our proofs are similar to those of Theorems 2.1 and 2.2 in Hall and Patil (1995). The difference is that we consider the errors $\{\epsilon_k, k \geq 1\}$ to be long memory stationary noise here, instead of i.i.d. random variables in their paper. As is always the case of going from i.i.d. to long-range dependence, several technical difficulties have to be overcome. We will use different methods than those in Hall and Patil (1995). The importance of the results and techniques of Taqqu (1975, 1977), Fox and Taqqu (1985) to our proofs will be clear.

We start by collecting and proving some lemmas. Denote

$$d_j \hat{=} E\left(\widehat{b_j}\right) = \frac{1}{n}\sum_{k=1}^{n} g(x_k)\phi_j(x_k),$$

and

$$d_{ij} \hat{=} E\left(\widehat{b_{ij}}\right) = \frac{1}{n}\sum_{k=1}^{n} g(x_k)\psi_{ij}(x_k).$$

The following lemma will be used for proving Lemmas 4.4 and 4.5.

**Lemma 4.1** *Suppose the function g in (1) is continuously differentiable on* [0, 1] *and the wavelets $\phi$ and $\psi$ satisfy the uniform Hölder conditions (4). Then*

$$\sup_j |d_j - b_j| = O\left(n^{-1/2}\right) \tag{16}$$

*and*

$$\sup_j |d_{ij} - b_{ij}| = O\left(n^{-1/2}\right). \tag{17}$$

*Proof* We only prove (16). The proof of (17) is similar.

First we write

$$d_j = \frac{p^{1/2}}{n}\sum_{k=1}^{n} g\left(\frac{k}{n}\right)\phi\left(\frac{pk}{n} - j\right).$$

For fixed $n$, $p$ and $j$, we note that

$$0 \leq \frac{pk}{n} - j \leq 1 \quad \text{if and only if} \quad \frac{nj}{p} \leq k \leq \frac{n(j+1)}{p}.$$

Let $m_j = \lfloor \frac{nj}{p} \rfloor$, where $\lfloor x \rfloor$ denotes the smallest integer that is at least $x$. Since $\phi$ has its support in [0, 1], we see that

$$\begin{aligned} d_j &= \frac{p^{1/2}}{n}\sum_{k=m_j+1}^{m_{j+1}} g\left(\frac{k}{n}\right)\phi\left(\frac{pk}{n} - j\right) \quad \left(\text{let } k = m_j + \ell\right) \\ &= \frac{p^{1/2}}{n}\sum_{\ell=1}^{\lfloor n/p \rfloor} g\left(\frac{\ell}{n} + \frac{j}{p}\right)\phi\left(\frac{p\ell}{n}\right) \quad \left(\text{let } t_\ell = \frac{p\ell}{n}\right) \\ &= \frac{1}{p^{1/2}}\sum_{\ell=1}^{\lfloor n/p \rfloor} g\left(\frac{t_\ell + j}{p}\right)\phi\left(t_\ell\right)\,\frac{p}{n}. \end{aligned} \tag{18}$$

Similarly, by a simple change of variables, we have

$$\begin{aligned} b_j &= \int_0^1 g(x)\phi_j(x)\mathrm{d}x \\ &= p^{1/2} \int_{j/p}^{(j+1)/p} g(x)\,\phi(px-j)\,\mathrm{d}x \quad (\text{let } t = px - j) \\ &= \frac{1}{p^{1/2}} \int_0^1 g\left(\frac{t+j}{p}\right)\phi(t)\,\mathrm{d}t. \end{aligned} \tag{19}$$

Combining (18) and (19), we have

$$\begin{aligned} |d_j - b_j| &= \frac{1}{p^{1/2}} \sum_{\ell=1}^{\lfloor n/p \rfloor} \int_{\frac{p\ell}{n}}^{\frac{p(\ell+1)}{n}} \left[ g\left(\frac{t_\ell + j}{p}\right)\phi(t_\ell) - g\left(\frac{t+j}{p}\right)\phi(t)\right] \mathrm{d}t \\ &= J_1 + J_2, \end{aligned} \tag{20}$$

where

$$J_1 = \frac{1}{p^{1/2}} \sum_{\ell=1}^{\lfloor n/p \rfloor} \int_{\frac{p\ell}{n}}^{\frac{p(\ell+1)}{n}} \left[ g\left(\frac{t_\ell + j}{p}\right) - g\left(\frac{t_+ j}{p}\right)\right] \phi(t_\ell)\ \mathrm{d}t$$

and

$$J_2 = \frac{1}{p^{1/2}} \sum_{\ell=1}^{\lfloor n/p \rfloor} \int_{\frac{p\ell}{n}}^{\frac{p(\ell+1)}{n}} g\left(\frac{t_+ j}{p}\right) [\phi(t_\ell) - \phi(t)]\ \mathrm{d}t.$$

For $J_1$, we use the differentiability of $g$ (in fact, it is enough if $g$ is a Lipschitz function on [0, 1]) and the boundedness of $\phi$ to get

$$J_1 \le \frac{1}{p^{1/2}} \cdot \frac{C}{n} \le C\, n^{-1/2}, \tag{21}$$

where $C$ is a constant. For $J_2$, we use the boundedness of $g$ and the uniform 1/2-Hólder condition (4) for $\phi$ to derive

$$J_2 \le \frac{1}{p^{1/2}} \cdot C\, \left(\frac{p}{n}\right)^{1/2} = C\, n^{-1/2}. \tag{22}$$

It is clear that (16) follows from (21) and (22).

The key lemma for the proof of Theorem 3.1 is the following Lemmas 4.3 and 4.4. To prove Lemma 4.3, we will make use of the following maximal inequality from Kôno (1983); see also Móricz (1976). □

**Lemma 4.2** *Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables. Assume that for integers $c \geq 0$ and $n \geq 1$, the partial sums $S_{c,n} = \sum_{j=c+1}^{c+n} X_j$ satisfy the following moment conditions:*

*(i) For $\gamma \geq 1$, there exist nonnegative numbers $\left\{h_{c,n},\ c \geq 0,\ n \geq 1\right\}$ such that*

$$E\left[\left|S_{c,n}\right|^{\gamma}\right] \leq h_{c,n} \quad \text{for all } c \text{ and } n.$$

*(ii) For all integers $c \geq 0, k, j \geq 1$,*

$$h_{c,j} + h_{c+j,k} \leq h_{c,j+k}.$$

*Then*

$$E\left[\max_{1 \leq k \leq n} \left|S_{c,k}\right|^{\gamma}\right] \leq (\log 2n)^{\gamma}\, h_{c,n}. \tag{23}$$

For every $j = 0, 1, 2, \ldots$, denote

$$s^*_{j,n} \mathrel{\hat{=}} \widehat{b}_j - E\left(\widehat{b}_j\right) = \frac{1}{n}\sum_{k=1}^{n} G(\xi_k)\phi_j(x_k),$$

$$s_{j,n} \mathrel{\hat{=}} \frac{J(m)}{m!}\,\frac{1}{n}\sum_{k=1}^{n} H_m(\xi_k)\phi_j(x_k),$$

$$s^*_{ij,n} \mathrel{\hat{=}} \widehat{b}_{ij} - E\left(\widehat{b}_{ij}\right) = \frac{1}{n}\sum_{k=1}^{n} G(\xi_k)\psi_{ij}(x_k),$$

$$s_{ij,n} \mathrel{\hat{=}} \frac{J(m)}{m!}\,\frac{1}{n}\sum_{k=1}^{n} H_m(\xi_k)\psi_{ij}(x_k).$$

Consider their second moments:

$$\sigma^{*2}_{j,n} = E\left(|s^*_{j,n}|^2\right), \quad \sigma^2_{j,n} = E\left(|s_{j,n}|^2\right),$$
$$\sigma^{*2}_{ij,n} = E\left(|s^*_{ij,n}|^2\right), \quad \sigma^2_{ij,n} = E\left(|s_{ij,n}|^2\right).$$

Using again the formula (10) for $E\left[H_k(\xi_1)H_\ell(\xi_{1+n})\right]$, we can derive

$$\begin{aligned}
\sigma^2_{j,n} &= \frac{J^2(m)}{m!}\,\frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} r(k-\ell)^m \phi_j(x_k)\,\phi_j(x_\ell) \\
&= \frac{J^2(m)}{m!}\,\frac{p}{n^2}\sum_{k=1}^{n/p-1}\sum_{\ell=1}^{n/p-1} r(k-\ell)^m \phi\left(\frac{pk}{n}\right)\phi\left(\frac{p\ell}{n}\right) \\
&\sim C_2\, p^{-1}\left(\frac{p}{n}\right)^{m\alpha} \quad \text{as } n \to \infty,
\end{aligned} \tag{24}$$

where $C_2$ is the constant in (14).

On the other hand, it follows from (30) below that we also have

$$\sigma_{j,n}^{*2} = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \sum_{q=m}^{\infty} \frac{J^2(q)}{q!}\, r(k-\ell)^q \phi_j(x_k)\, \phi_j(x_\ell)$$
$$\sim C_2\, p^{-1} \left(\frac{p}{n}\right)^{m\alpha} \quad \text{as } n \to \infty. \tag{25}$$

Similar relations hold for $\sigma_{ij,n}^2$ and $\sigma_{ij,n}^{*2}$ as well. For example, we mention that for fixed $i, j$,

$$\sigma_{ij,n}^2 = \frac{J^2(m)}{m!} \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} r(k-\ell)^m \psi_{ij}(x_k)\, \psi_{ij}(x_\ell)$$
$$\sim C_1\, p_i^{-1} \left(\frac{p_i}{n}\right)^{m\alpha} \quad \text{as } n \to \infty, \tag{26}$$

where $C_1$ is the constant in (12).

The following lemma is more than we actually need in this paper. We believe it will be useful elsewhere.

**Lemma 4.3** [Reduction Principal] *Let $G \in \mathcal{G}_m$ and $\{\xi_n, n \geq 1\} \in (m)(\alpha, C_0)$. Then for all integers $i, j \geq 0$,*

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{|s_{j,k}^* - s_{j,k}|}{\sigma_{j,n}} = 0 \quad a.s. \tag{27}$$

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{|s_{ij,k}^* - s_{ij,k}|}{\sigma_{ij,n}} = 0 \qquad a.s. \tag{28}$$

*Proof* We only prove (27), the proof of (28) follows from the same argument. Note that the orthogonality of $\{H_k(x)\}$ implies that

$$E\left(|s_{j,n}^* - s_{j,n}|^2\right) = E\left(|s_{j,n}^*|^2\right) - E\left(|s_{j,n}|^2\right)$$
$$= \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \sum_{q=m+1}^{\infty} \frac{J^2(q)}{q!}\, r(k-\ell)^q \phi_j(x_k)\, \phi_j(x_\ell). \tag{29}$$

Define

$$h_{c,j} \,\hat{=}\, \frac{1}{n^2} \sum_{k=c+1}^{c+j} \sum_{\ell=c+1}^{c+j} \sum_{q=m+1}^{\infty} \frac{J^2(q)}{q!}\, |r(k-\ell)|^q \left|\phi_j(x_k)\right| \left|\phi_j(x_\ell)\right|.$$

Then for all integers $k, n \geq 1$ we have $h_{0,k} + h_{k,n} \leq h_{0,k+n}$. Moreover, (29) implies that

$$E\left(|s_{j,n}^* - s_{j,n}|^2\right) \leq h_{0,n}.$$

Thus the two assumptions of Lemma 4.2 are satisfied with $c = 0$.

Let $0 < \varepsilon < \varepsilon' < 1$ be fixed and small. Then (8) implies that for $n$ large and all $|k| \ge n^{\varepsilon'}$, we have $|r(k)| \le n^{-\alpha\varepsilon}$. It follows that

$$\begin{aligned} h_{0,n} &= \frac{1}{n^2}\left[\sum_{k=1}^{n}\sum_{|k-\ell|\le n^{\varepsilon'}}\sum_{q=m+1}^{\infty}\frac{J^2(q)}{q!}\,|r(k-\ell)|^q\left|\phi_j(x_k)\right|\left|\phi_j(x_\ell)\right|\right.\\ &\qquad\left.+\sum_{k=1}^{n}\sum_{|k-\ell|>n^{\varepsilon'}}\sum_{q=m+1}^{\infty}\frac{J^2(q)}{q!}\,|r(k-\ell)|^q\left|\phi_j(x_k)\right|\left|\phi_j(x_\ell)\right|\right]\\ &\le \frac{C}{n^2}\left[n^{1+\varepsilon'}+n^{-\alpha\varepsilon}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\frac{J^2(m)}{m!}\,|r(k-\ell)|^m\left|\phi_j(x_k)\right|\left|\phi_j(x_\ell)\right|\right]. \end{aligned} \tag{30}$$

Similar to (24), we derive that

$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\frac{J^2(m)}{m!}\,|r(k-\ell)|^m\left|\phi_j(x_k)\right|\left|\phi_j(x_\ell)\right| \sim C_3\,p^{-1}\left(\frac{p}{n}\right)^{m\alpha} \tag{31}$$

as $n \to \infty$, where $C_3$ is the positive and finite constant defined by

$$C_3 = \frac{C_0\,J^2(m)}{m!}\int\int |x-y|^{m\alpha}|\phi(x)||\phi(y)\Big|\,\mathrm{d}x\,\mathrm{d}y.$$

This and (24) imply that the left hand side of (31) is comparable with $\sigma^2_{j,n}$, i.e.,

$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\frac{J^2(m)}{m!}\,|r(k-\ell)|^m\left|\phi_j(x_k)\right|\left|\phi_j(x_\ell)\right| \asymp \sigma^2_{j,n}. \tag{32}$$

here $a_n \asymp b_n$ means that $a_n/b_n$ is bounded above and below by positive and finite constants as $n \to \infty$. It follows from (30) and (32) that for $n$ large enough

$$h_{0,n} \le C\,n^{-\eta}\sigma^2_{j,n}, \tag{33}$$

where $\eta = \min\{\alpha\varepsilon,\ 1-m\alpha-\varepsilon\} > 0$. Hence by Lemma 4.2 with $c = 0$, we see that for any $q \ge 1$

$$E\left(\max_{1\le n\le 2^q}\left|s^*_{j,n}-s_{j,n}\right|^2\right) \le (q+1)^2 h_{0,2^q} \le Cq^2 2^{-\eta q/2}\,\sigma^2_{j,2^q}.$$

Hence,

$$\sum_{q=1}^{\infty} E\left[\frac{\max_{1\le n\le 2^q}\left|s^*_{j,n}-s_{j,n}\right|^2}{\sigma^2_{j,2^q}}\right] < \infty.$$

This implies that

$$\lim_{q\to\infty}\frac{\max_{1\le n\le 2^q}\left|s^*_{j,n}-s_{j,n}\right|}{\sigma_{j,2^q}} = 0 \quad \text{a.s.} \tag{34}$$

Since the sequence $\{G(\xi_k)\}$ is stationary, we also have

$$\lim_{q\to\infty} \frac{\max_{2^q \le n \le 2^{q+1}} \left| s^*_{j,n} - s^*_{j,2^q} - s_{j,n} + s_{j,2^q} \right|}{\sigma_{j,2^q}} = 0 \quad \text{a.s.} \tag{35}$$

Therefore, (27) follows from (32), (34), (35) and a standard monotonicity argument. □

**Lemma 4.4** *Under the assumptions of Theorem 3.1,*

$$S_1 \hat{=} E \left| \sum_{j=0}^{p-1} \left(\widehat{b}_j - b_j\right)^2 - C_2 \left(n^{-1} p\right)^{m\alpha} \right| = o\big((n^{-1} p)^{m\alpha}\big).$$

*Proof* We first consider the behavior of $\sum_{j=0}^{p-1} \left(\widehat{b}_j - E\widehat{b}_j\right)^2$.

Note that

$$\widehat{b}_j - E\widehat{b}_j = \frac{1}{n} \sum_{k=1}^{n} G(\xi_k)\phi_j(x_k) = s^*_{j,n}.$$

It follows from the triangle inequality and the Cauchy–Schwarz inequality that

$$\begin{aligned} E\left|\sum_{j=0}^{p-1} \left(\widehat{b}_j - E\widehat{b}_j\right)^2 - C_2\left(n^{-1}p\right)^{m\alpha}\right| &\le E\left|\sum_{j=0}^{p-1} s^2_{j,n} - C_2\left(n^{-1}p\right)^{m\alpha}\right| \\ +2\left[\sum_{j=0}^{p-1} E(s^2_{j,n})\right]^{1/2} &\left[\sum_{j=0}^{p-1} E\left(s^*_{j,n} - s_{j,n}\right)^2\right]^{1/2} + \sum_{j=0}^{p-1} E\left(s^*_{j,n} - s_{j,n}\right)^2. \end{aligned} \tag{36}$$

By (the proof of) Lemma 4.3, the last two terms in (36) are of the order $o\left((n^{-1}p)^{m\alpha}\right)$. We claim that

$$E\left|\sum_{j=0}^{p-1} s^2_{j,n} - C_2\left(n^{-1}p\right)^{m\alpha}\right| = o\left((n^{-1}p)^{m\alpha}\right). \tag{37}$$

To prove this, put $Q_n = \sum_{j=0}^{p-1} s^2_{j,n}$ and $\mu_n = E(Q_n)$. Then by (24) we have

$$\mu_n = C_2\left(n^{-1}p\right)^{m\alpha} + o\left((n^{-1}p)^{m\alpha}\right).$$

Thus, the left hand side of (37) is at most

$$S_{11} = E^{1/2}(Q_n - \mu_n)^2 + o\left((n^{-1}p)^{m\alpha}\right). \tag{38}$$

Note that $Q_n$ is a quadratic form of the stationary sequence $\{G(\xi_k), k \ge 1\}$ with long-range dependence. In order to evaluate its variance, we apply the "diagram formula" for multiple Wiener-Itô-Dobrushin integrals, which gives a convenient way to calculate the expectation of the products of Gaussian random variables or multiple Wiener-Itô-Dobrushin integrals. See Major (1981, Sect. 5), or (Fox and Taqqu, 1985, Sect. 3) for more information.

Since $\{\xi_k, k \geq 1\}$ is a stationary Gaussian process with mean 0 and variance 1, it follows from the S. Bochner theorem that its covariance function $r(k)$ has the following spectral representation

$$r(k) = \int_{[-\pi,\pi]} e^{iky} F(dy), \quad \forall\, k \in \mathbb{Z},$$

where $F$ is a Borel probability measure on $[-\pi, \pi]$ which is called the spectral measure of $\{\xi_k, k \geq 1\}$. Let $Z_F$ be the corresponding random spectral measure, i.e. the complex-valued Gaussian scattered measure such that $E(Z_F(A))^2 = F(A)$ for all Borel set $A \subset [-\pi, \pi]$. Then for all $k \geq 1$

$$\xi_k = \int_{[-\pi,\pi]} e^{iky} dZ_F(y).$$

It follows from Theorem 4.2 of Major (1981) that

$$H_m(\xi_k) = \int''_{[-\pi,\ \pi]^m} e^{ik(y_1+\cdots+y_m)}\, dZ_F(y_1)\cdots dZ_F(y_m), \tag{39}$$

where $\int''$ is the multiple Wiener-Itô-Dobrushin integral. Such a representation and multiple Wiener-Itô-Dobrushin integrals played important rôles in Fox and Taqqu (1985), who investigated the limiting processes of the quadratic forms $\sum_{k=1}^{\lfloor nt \rfloor} \sum_{\ell=1}^{\lfloor nt \rfloor} H_m(\xi_k) H_m(\xi_\ell)$. In particular, by applying the "Diagram Formula", (Fox and Taqqu, 1985, Lemma 3.4) showed

$$\begin{aligned} H_m(\xi_k)H_m(\xi_\ell) &= m!\, r(k-\ell)^m \\ &\quad + \sum_{h=1}^{m} \left[ (m-h)! \binom{m}{h}^2 r(k-\ell)^{m-h}\, K_h(k,\ell) \right], \end{aligned} \tag{40}$$

where

$$K_h(k,\ell) = \int''_{[-\pi,\ \pi]^{2h}} e^{ik(y_1+\cdots+y_h)+i\ell(y_{h+1}+\cdots+y_{2h})}\, dZ_F(y_1)\cdots dZ_F(y_{2h}). \tag{41}$$

In order to express $Q_n - \mu_n$ in terms of the multiple Wiener-Itô-Dobrushin integrals $K_h(k, \ell)$, we denote

$$a_{k\ell} = \frac{J^2(m)}{(m!)^2}\, pn^{-2} \sum_{j=0}^{p-1} \phi(px_k - j)\phi(px_\ell - j).$$

Similar to the proof of Lemma 3.5 in Fox and Taqqu (1985), we see that (40) implies

$$\begin{aligned} Q_n - \mu_n &= \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k,\ell} \left[ H_m(\xi_k) H_m(\xi_\ell) - E\left( H_m(\xi_k) H_m(\xi_\ell) \right) \right] \\ &= \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k,\ell} \sum_{i=1}^{m} (m-i)! \binom{m}{i}^2 r(k-\ell)^{m-i} K_i(k,\ell) \\ &= \sum_{i=1}^{m} (m-i)! \binom{m}{i}^2 \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k,\ell}\, r(k-\ell)^{m-i} K_i(k,\ell). \end{aligned}$$

By the Cauchy–Schwarz inequality, we have

$$E\,(Q_n - \mu_n)^2 \le m \sum_{i=1}^{m} ((m-i)!)^2 \binom{m}{i}^4 E\left[ \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k,\ell}\, r(k-\ell)^{m-i} K_i(k,\ell) \right]^2 . \tag{42}$$

Let $1 \le i \le m$ be fixed. We note that

$$\begin{aligned} &E\left[ \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k,\ell}\, r(k-\ell)^{m-i} K_i(k,\ell) \right]^2 \\ &= \sum_{k_1=1}^{n} \sum_{\ell_1=1}^{n} \sum_{k_2=1}^{n} \sum_{\ell_2=1}^{n} a_{k_1,\ell_1} a_{k_2,\ell_2}\, r(k_1-\ell_1)^{m-i} r(k_2-\ell_2)^{m-i} E\left[ K_i(k_1,\ell_1) K_i(k_2,\ell_2) \right]. \end{aligned} \tag{43}$$

Lemma 3.6 of Fox and Taqqu (1985) states that

$$\begin{aligned} &E\left[ K_i\,(k_1,\ell_1)\, K_i(k_2,\ell_2) \right] \\ &= \sum_{q=0}^{i} (i!)^2 \binom{i}{q}^2 r\,(k_1-k_2)^q\; r\,(\ell_1-\ell_2)^q\; r\,(\ell_1-k_2)^{i-q}\; r\,(k_1-\ell_2)^{i-q}\,. \end{aligned}$$

It follows that (43) is a sum of the terms $(i!)^2 \binom{i}{q}^2 T_q (q = 0, \ldots, i)$, where

$$\begin{aligned} T_q = \sum_{k_1=1}^{n} \sum_{\ell_1=1}^{n} \sum_{k_2=1}^{n} \sum_{\ell_2=1}^{n} & a_{k_1,\ell_1} a_{k_2,\ell_2}\, r\,(k_1-\ell_1)^{m-i}\, r\,(k_2-\ell_2)^{m-i} \\ & \times r\,(k_1-k_2)^q\; r\,(\ell_1-\ell_2)^q\; r\,(\ell_1-k_2)^{i-q}\; r\,(k_1-\ell_2)^{i-q}\,. \end{aligned} \tag{44}$$

Some elementary calculations similar to those in Lemmas 4.4 show that $T_q$ can be written as

$$\begin{aligned}
&\frac{J^4(m)}{(m!)^4}\,\frac{p^2}{n^4}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}\sum_{k_1}\sum_{\ell_1}\sum_{k_2}\sum_{\ell_2}\phi\left(\frac{pk_1}{n}-j_1\right)\phi\left(\frac{p\ell_1}{n}-j_1\right)\\
&\quad\times\phi\left(\frac{pk_2}{n}-j_2\right)\phi\left(\frac{p\ell_2}{n}-j_2\right)\times r\,(k_1-\ell_1)^{m-i}\,r\,(k_2-\ell_2)^{m-i}\\
&\quad\times r\,(k_1-k_2)^q\;r\,(\ell_1-\ell_2)^q\;r\,(\ell_1-k_2)^{i-q}\;r\,(k_1-\ell_2)^{i-q}\\
&\le C\left(\frac{p}{n}\right)^{2m\alpha}\,p^{-2}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}\iiiint|x_1-y_1|^{-(m-i)\alpha}|x_2-y_2|^{-(m-i)\alpha}\\
&\quad\times|x_1-x_2-(j_1-j_2)|^{-q\alpha}\,|y_1-y_2-(j_1-j_2)|^{-q\alpha}\\
&\quad\times|y_1-x_2-(j_1-j_2)|^{-(i-q)\alpha}\\
&\quad\times|x_1-y_2-(j_1-j_2)|^{-(i-q)\alpha}\,\phi(x_1)\phi(y_1)\phi(x_2)\phi(y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2
\end{aligned}\tag{45}$$

for all $n$ large enough.

Let $k=j_1-j_2$. We will make use of the following inequality:

$$\begin{aligned}
&\iiiint|x_1-y_1|^{-(m-i)\alpha}|x_2-y_2|^{-(m-i)\alpha}|x_1-x_2-k|^{-q\alpha}\,|y_1-y_2-k|^{-q\alpha}\\
&\quad\times|y_1-x_2-k|^{-(i-q)\alpha}\,|x_1-y_2-k|^{-(i-q)\alpha}\\
&\quad\times\phi(x_1)\phi(y_1)\phi(x_2)\phi(y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2\\
&\le C\,k^{-2\alpha i},\quad\forall\,k\ge 1.
\end{aligned}\tag{46}$$

We only need to verify this for $k\ge 2$. It follows from the easy fact that $|y_1-y_2-k|\ge k-1\ge k/2$ for all $y_1,y_2\in[0,1]$ and all $k\ge 2$.

By changing the order of summation in (45) and applying (46), we derive

$$\begin{aligned}
|T_q|&\le C\left(\frac{p}{n}\right)^{2m\alpha}p^{-2}\sum_{k=1}^{p-1}(p-k)\,k^{-2i\alpha}\\
&\le C\left(\frac{p}{n}\right)^{2m\alpha}p^{-2i\alpha}.
\end{aligned}\tag{47}$$

Combining (38), (42)–(47), we have established (37).

The rest of the proof of Lemma 4.4 is quite standard. Using the triangle inequality and the Cauchy–Schwarz inequality, we can obtain

$$\begin{aligned}
S_1&\le E^{1/2}\,(Q_n-\mu_n)^2+\sum_{j=0}^{p-1}\left(d_j-b_j\right)^2\\
&\quad+2\left[\sum_{j=0}^{p-1}E\left(\widehat{b}_j-d_j\right)^2\sum_{j=0}^{p-1}\left(d_j-b_j\right)^2\right]^{1/2}+o\left((n^{-1}p)^{m\alpha}\right)\\
&=S_{11}+S_{12}+S_{13}+o\left((n^{-1}p)^{m\alpha}\right).
\end{aligned}\tag{48}$$

It follows from (16) in Lemma 4.1 that

$$\sup_j |d_j - b_j| = O(n^{-1/2}).$$

Thus,

$$S_{12} = O(n^{-1}p) = o\left((n^{-1}p)^{m\alpha}\right),$$
$$S_{13} = O\left((n^{-1}p)^{1/2}\mu_n^{1/2}\right) = o\left((n^{-1}p)^{m\alpha}\right). \tag{49}$$

The proof of Lemma 4.4 is finished. □

Now we turn to the second sum in (15).

**Lemma 4.5** *Under the assumption of Theorem 3.1,*

$$S_2 \hat{=} \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{\left(\widehat{b}_{ij} - b_{ij}\right)^2 I(|\widehat{b}_{ij}| > \delta_i)\right\} = o\left((n^{-1}p)^{m\alpha} + p^{-2r}\right).$$

*Proof* The way of breaking $S_2$ into several parts is analogous to that in the proof of Theorem 2.1 of Hall and Patil (1995, p. 916–918). However, we have to overcome some complications caused by the fact that the errors $\{G(\xi_k), k \geq 1\}$ are long-range dependent and are, in general, non-Gaussian. Let $\lambda$ and $\beta$ denote positive numbers satisfying $2\lambda + \beta = 1$, and set

$$S'_{21} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{\left(\widehat{b}_{ij} - b_{ij}\right)^2\right\} I(|b_{ij}| > \lambda\delta_i),$$

$$S'_{22} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{(\widehat{b}_{ij} - b_{ij})^2 I\left(|\widehat{b}_{ij} - b_{ij}| > (\lambda+\beta)\delta_i\right)\right\}.$$

The triangle inequality implies $S_2 \leq S'_{21} + S'_{22}$. Replacing $b_{ij}$ in the above expressions with $d_{ij} = E(\widehat{b}_{ij})$, we define

$$S_{21} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{\left(\widehat{b}_{ij} - d_{ij}\right)^2\right\} I\left(|b_{ij}| > \lambda\delta_i\right), \tag{50}$$

$$S_{22} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{\left(\widehat{b}_{ij} - d_{ij}\right)^2 I\left(|\widehat{b}_{ij} - d_{ij}| > (\lambda+\beta)\delta_i\right)\right\}. \tag{51}$$

Using (17) in Lemma 4.1 and the fact $n\delta_i^2 \to \infty$ we can show that $S_{21}$ and $S_{22}$ are the leading terms of $S'_{21}$ and $S'_{22}$ and $S'_{21} = O(S_{21})$ and $S'_{22} = O(S_{22})$ [the arguments are the same as (36) and (48) above]. Hence, in order to prove the Lemma, it suffices to prove

$$S_{21} = o\left(\left(n^{-1}p\right)^{m\alpha} + p^{-2r}\right) \quad \text{and} \quad S_{22} = o\left(\left(n^{-1}p\right)^{m\alpha}\right), \tag{52}$$

respectively. Since the mean regression function $g$ is $r$-times continuously differentiable, by using the Taylor expansion of $g$ and the moment condition on the mother wavelet, we have

$$\begin{aligned} b_{ij} &= p_i^{-1/2} \int \psi(y) \left[ \sum_{v=0}^{r-1} \frac{1}{v!} \left( \frac{y}{p_i} \right)^v g^{(v)} \left( \frac{j}{p_i} \right) \right. \\ &\quad \left. + \frac{1}{(r-1)!} \left( \frac{y}{p_i} \right)^r \int_0^1 (1-t)^{r-1} g^{(r)} \left( \frac{j+ty}{p_i} \right) \mathrm{d}t \right] \mathrm{d}y \\ &= \kappa p_i^{-(r+1/2)} \left( g_{ij} + \eta_{ij} \right), \end{aligned} \tag{53}$$

where

$$g_{ij} = g^{(r)} \left( \frac{j}{p_i} \right) \quad \text{and} \quad \sup_{0 \le j \le p_i - 1;\, 0 \le i \le q-1} |\eta_{ij}| \to 0.$$

Hence, we have $|b_{ij}| \le C p_i^{-(r+1/2)}$. Similar to (26), we have that for all $i$, $j$,

$$E \left( \widehat{b}_{ij} - d_{ij} \right)^2 \le C p_i^{-1} \left( n^{-1} p_i \right)^{m\alpha}. \tag{54}$$

It follows from (50) and (54) that

$$\begin{aligned} S_{21} &\le \sum_{i=0}^{q-1} C (n^{-1} p_i)^{m\alpha} I \left( p_i \le C \delta_i^{-2/(2r+1)} \right) \\ &= C \left( n^{-1} p \right)^{m\alpha} \sum_{i=0}^{q-1} 2^{m\alpha i}\, I \left( p_i \le C \delta_i^{-2/(2r+1)} \right). \end{aligned} \tag{55}$$

Note that $p_i \le C \delta_i^{-2/(2r+1)}$ implies $(p2^i)^{2r+1} \le C n^{m\alpha} p^{1-m\alpha} (\ln\, n)^{-m}\, 2^{(1-m\alpha)i}$, we have

$$2^{(2r+m\alpha)i} \le C \frac{n^{m\alpha} p^{-(2r+m\alpha)}}{(\ln\, n)^m}.$$

There are only finitely many $i$'s satisfying this inequality. We denote the largest such $i$ by $t$. Since by (SP), $n/(p2^q) \to \infty$, it follows from (55) that

$$\begin{aligned} S_{21} &\le C \left( n^{-1} p \right)^{m\alpha} \sum_{i=0}^{t} 2^{m\alpha i} \\ &\le C \left( n^{-1} p \right)^{m\alpha} 2^{tm\alpha} \\ &\le C \left( n^{-1} p \right)^{m\alpha} \left[ \frac{n^{m\alpha} p^{-(2r+m\alpha)}}{(\ln\, n)^m} \right]^{\frac{m\alpha}{2r+m\alpha}} \\ &= C n^{-\frac{2rm\alpha}{2r+m\alpha}} (\ln\, n)^{-\frac{m^2\alpha}{2r+m\alpha}} \\ &= o \left( n^{-\frac{2rm\alpha}{2r+m\alpha}} \right) \\ &= o \left( \left( n^{-1} p \right)^{m\alpha} + p^{-2r} \right). \end{aligned} \tag{56}$$

In order to estimate $S_{22}$, we have to use a different truncation than that in Lemma 4.4. Recall that the condition of smoothing parameters $p$ and $q$ implies $q = O(\ln n)$. Thus, there exists a constant $C_4 > 0$ such that $q \le C_4 \log_2 n$. We choose an integer $\tau \ge 2$ such that $\tau\eta > C_4$ [recall $\eta = \min\{\alpha\varepsilon, 1 - m\alpha - \varepsilon\}$] and split $\widehat{b}_{ij} - d_{ij}$ as:

$$\widehat{b}_{ij} - d_{ij} = \frac{1}{n}\sum_{k=1}^{n}\sum_{q=m}^{m+\tau-1}\frac{J(q)}{q!}H_q(\xi_k)\psi_{ij}(x_k) + \frac{1}{n}\sum_{k=1}^{n}\sum_{q=m+\tau}^{\infty}\frac{J(q)}{q!}H_q(\xi_k)\psi_{ij}(x_k)$$
$$\hat{=}\, t_{ij,n} + t^*_{ij,n}. \tag{57}$$

The triangle inequality implies

$$\begin{aligned} S_{22} &\le 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\left\{\left[t^2_{ij,n} + \left(t^*_{ij,n}\right)^2\right] I(|\widehat{b}_{ij} - d_{ij}| > (\lambda+\beta)\delta_i)\right\} \\ &\le 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}\left\{E\left[t^2_{ij,n} I\left(|t_{ij,n}| > \beta\delta_i\right)\right] + E\left[t^2_{ij,n} I\left(|t^*_{ij,n}| > \lambda\delta_i\right)\right]\right. \\ &\qquad\qquad \left. + E\left(t^*_{ij,n}\right)^2\right\} \\ &\hat{=}\, 2(S_{23} + S_{24} + S_{25}). \end{aligned}$$

Applying the Cauchy–Schwarz inequality to $S_{23}$, we derive

$$S_{23} \le \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E^{1/2}\left(t^4_{ij,n}\right)\, P^{1/2}\left(|t_{ij,n}| > \beta\delta_i\right). \tag{58}$$

By using (39), we write $t_{ij,n}$ as a sum of the $q$-tuple Wiener-Itô-Dobrushin integrals:

$$t_{ij,n} = \sum_{q=m}^{m+\tau-1} \int''_{[-\pi,\pi]^q} \frac{J(q)}{q!}\frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{ik(y_1+\cdots+y_q)}\psi_{ij}(x_k)\mathrm{d}Z_F(y_1)\ldots\mathrm{d}Z_F(y_q). \tag{59}$$

Denote the above integrands by $f_q\left(y_1, \ldots, y_q\right)$ $(q = m, \ldots, m+\tau-1)$. It is clear that every $f_q$ satisfies the following conditions:

(1) $f_q\left(-y_1, \ldots, -y_q\right) = \overline{f_q(y_1, \ldots, y_q)}$;
(2) $\|f_q\|^2_F = \int |f_q(y_1, \ldots, y_q)|^2\, \mathrm{d}F(y_1)\ldots\mathrm{d}F(y_q) < \infty$;
(3) For every permutation $\pi$ of $\{1, 2, \ldots, q\}$,

$$f_q\left(y_{\pi(1)}, \ldots, y_{\pi(q)}\right) = f_q(y_1, \ldots, y_q).$$

In the notation of Major (1981, p. 22–23), we have $f_q \in H^q_F$. Hence, by applying Corollary 5.6 in Major (1981, p. 53) together with an estimate of the constant $\overline{C}(q, N)$ in Major (1981, p. 69) to each $q$-tuple Wiener-Itô-Dobrushin integral $I_F(f_q)$ in (59), we deduce that for all integer $N \ge 1$,

$$E\left[I_F(f_q)^{2N}\right] \le (2N)^{qN}\left[I_F(f_q)^2\right]^N. \tag{60}$$

It follows Jensen's inequality and (60) that

$$
\begin{aligned}
E\left[(t_{ij,n})^{2N}\right] &= E\left[\left(\sum_{q=m}^{m+\tau-1} I_F(f_q)\right)^{2N}\right] \\
&\le \tau^{2N-1}\sum_{q=m}^{m+\tau-1} E\left[I_F(f_q)^{2N}\right] \\
&\le \tau^{2N-1}\sum_{q=m}^{m+\tau-1} (2N)^{qN}\left[I_F(f_q)^2\right]^N \\
&\sim C_1^N\tau^{2N-1}\sum_{q=m}^{m+\tau-1} (2N)^{qN}\left[p_i^{-1}\left(\frac{p_i}{n}\right)^{q\alpha}\right]^N, \quad \text{as } n\to\infty,
\end{aligned}
$$

where the last inequality follows from (26). Hence for all positive integers $N$ such that

$$N\left(\frac{p_i}{n}\right)^{\alpha} \le \frac{1}{4}, \tag{61}$$

we have

$$E\left[(t_{ij,n})^{2N}\right] \le (2N)^{mN}\tau^{2N}\left[\sigma_{ij,n}\right]^{2N}. \tag{62}$$

Let $\Lambda = |t_{ij,n}|/\sigma_{ij,n}$, then by (62), we derive that for all integers $N$ satisfying (61),

$$E\left(\Lambda^{2N}\right) \le (2N)^{mN}\,\tau^{2N}.$$

For any $u > 0$, we take

$$N = \frac{1}{2e}\left(\frac{u}{\tau}\right)^{2/m}. \tag{63}$$

A simple argument using Chebyshev's inequality shows that, as long as the $N$ defined in (63) satisfies (61), we have

$$
\begin{aligned}
P(\Lambda > u) &\le \frac{(2N)^{mN}\,\tau^{2N}}{u^{2N}} \\
&= \exp\left(-\frac{m}{2e}\left(\frac{u}{\tau}\right)^{2/m}\right).
\end{aligned} \tag{64}
$$

Now let

$$u = \frac{\beta\delta_i}{\sigma_{ij,n}}.$$

Then it is easy to verify that for all $n$ large enough, the $N$ defined in (63) satisfies (61). It follows from (64) that

$$P\left(|t_{ij,n}| > \beta\delta_i\right) \le \exp\left(-\frac{m}{2e}\,\sigma_{ij,n}^{-2/m}\left(\frac{\beta\delta_i}{\tau}\right)^{2/m}\right).$$

By the choice of $\delta_i$, we derive that for some $\varepsilon > 0$,

$$P\left(|t_{ij,n}| > \beta\delta_i\right) \le \exp\left(-2\beta^{(2/m)} \ln n\right) = n^{-2\beta^{(2/m)}}. \tag{65}$$

Combining (58), (61) and (65), we obtain

$$\begin{aligned} S_{23} &\le C \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} p_i^{-1} \left(n^{-1} p_i\right)^{m\alpha} n^{-\beta^{(2/m)}} \\ &= C\left(n^{-1} p_q\right)^{m\alpha} n^{-\beta^{(2/m)}} \\ &= o\left(n^{-\frac{2rm\alpha}{2r+m\alpha}}\right) \\ &= o\left(\left(n^{-1} p\right)^{m\alpha} + p^{-2r}\right), \end{aligned} \tag{66}$$

the third equality follows from $n^{-1} p_q \to 0$ and choose $\beta < 1$ so close to 1.

To estimate $S_{24}$, we note that

$$S_{24} \le \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} E^{1/2}\left(t_{ij,n}^4\right) P^{1/2}\left(|t_{ij,n}^*| > \lambda\delta_i\right). \tag{67}$$

The same argument as in proof of Lemma 4.3 [cf. (33)] gives that

$$E\left[\left(t_{ij,n}^*\right)^2\right] \le Cn^{-\tau\eta} \cdot \sigma_{ij,n}^2.$$

Thus, Chebyshev's inequality implies that

$$P\left(|t_{ij,n}^*| > \lambda\delta_i\right) \le Cn^{-\tau\eta}.$$

Hence, similar to (66), we have

$$\begin{aligned} S_{24} &\le C \left(n^{-1} p\right)^{m\alpha} \cdot n^{-\tau\eta} \sum_{i=0}^{q-1} 2^{m\alpha i} \\ &\le C \left(n^{-1} p\right)^{m\alpha} \cdot n^{-\tau\eta} \cdot 2^{qm\alpha} \\ &= o\left((n^{-1} p)^{m\alpha}\right), \end{aligned} \tag{68}$$

the last equality follows from our choice of $\tau$ which implies $n^{-\tau\eta} \cdot 2^{qm\alpha} \to 0$.

In the same way, we have

$$\begin{aligned} S_{25} &\le C\left(n^{-1} p\right)^{m\alpha} \cdot n^{-\tau\eta} \cdot 2^{qm\alpha} \\ &= o\left((n^{-1} p)^{m\alpha}\right). \end{aligned} \tag{69}$$

Combining the estimates on $S_{21}$ and $S_{22}$ together, we obtain (52). This finishes the proof of Lemma 4.5. □

**Lemma 4.6** *Under the assumption of Theorem 3.1,*

$$S_3 \hat{=} E\left|\sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|\widehat{b}_{ij}| \le \delta_i\right) - p^{-2r}\kappa^2\left(1 - 2^{-2r}\right)^{-1} \int g^{(r)^2}\right| = o\left(p^{-2r}\right).$$

*Proof* Let $\epsilon > 0$, and define

$$S_{30} = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|\widehat{b}_{ij}| \le \delta_i\right),$$

$$S_{31} = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|b_{ij}| \le (1+\epsilon)\delta_i\right),$$

$$S_{32} = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|b_{ij}| \le (1-\epsilon)\delta_i\right),$$

$$\Delta = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I\left(|\widehat{b}_{ij} - b_{ij}| > \epsilon\delta_i\right).$$

Then the triangle inequality implies

$$S_{32} - \Delta \le S_{30} \le S_{31} + \Delta. \tag{70}$$

Using the Eq. (53) for $b_{ij}$ and the assumption $p_i^{2r+1}\delta_i^2 \to \infty$ in (SP), we see that $I\{|b_{ij}| \le (1+\epsilon)\delta_i\} = I\{|b_{ij}| \le (1-\epsilon)\delta_i\} = 1$ for $n$ sufficiently large, and

$$S_{31} = S_{32} = p^{-2r}\kappa^2\left(1 - 2^{-2r}\right)^{-1} \int g^{(r)^2} + o(p^{-2r}). \tag{71}$$

For more details, see Hall and Patil (1995, p. 920). On the other hand, applying the argument analogous to that for $S_{22}$, we have

$$E\Delta = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 P\left(|\widehat{b}_{ij} - b_{ij}| > \epsilon\delta_i\right) \le Cn^{-\epsilon^{2/m}} \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 = o\left(p^{-2r}\right).$$

Combining (70) and (71), we have proved the lemma. □

**Lemma 4.7** *Under the assumption of Theorem 3.1,*

$$S_4 \equiv \sum_{i=q}^{\infty} \sum_{j=0}^{p_i-1} b_{ij}^2 = o(p^{-2r}).$$

*Proof* The proof follows from the $b_{ij}$'s Taylor expansion (53) and $q \to \infty$.

We are now in the position to give the proof of the Theorems 3.1 and 3.2. □

*Proof of the Theorem 3.1* By using (15), we have

$$E\left|\int (\widehat{g} - g)^2 - \left\{C_2(n^{-1}p)^{m\alpha} + p^{-2r}\kappa^2(1 - 2^{-2r})^{-1}\int g^{(r)^2}\right\}\right|$$
$$\le S_1 + S_2 + S_3 + S_4.$$

Hence the proof follows from Lemmas 4.2 to 4.5.

*Proof of the Theorem 3.2* We use the same notation as in Hall and Patil (1995). Notice that, by the orthogonality properties of $\phi$ and $\psi$,

$$\int (\widehat{g} - g)^2 = I_q(\mathbb{Z}, \mathbb{Z}, \ldots),$$

where $\mathbb{Z}$ denotes the corresponding set of integers (for instance, $\Psi_i = \{0, 1, \ldots, p_i - 1\}$) and

$$\begin{aligned}
I_q(\Psi, \Psi_0, \Psi_1, \ldots) &= \sum_{j\in\Psi} \left(\widehat{b}_j - b_j\right)^2 + \sum_{i=0}^{q-1} \sum_{j\in\Psi_i} \left(\widehat{b}_{ij} - b_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right) \\
&\quad + \sum_{i=0}^{q-1} \sum_{j\in\Psi_i} b_{ij}^2 I\left(|\widehat{b}_{ij}| \le \delta_i\right) + \sum_{i=q}^{\infty} \sum_{j\in\Psi_i} b_{ij}^2 \\
&= \sum_{j\in\Psi} \left(\widehat{b}_j - d_j\right)^2 + \sum_{j\in\Psi} \left(d_j - b_j\right)^2 + 2\sum_{j\in\Psi} \left(\widehat{b}_j - d_j\right)\left(d_j - b_j\right) \\
&\quad + \sum_{i=0}^{q-1} \sum_{j\in\Psi_i} \left(\widehat{b}_{ij} - d_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right) \\
&\quad + \sum_{i=0}^{q-1} \sum_{j\in\Psi_i} \left(d_{ij} - b_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right) \\
&\quad + 2\sum_{i=0}^{q-1} \sum_{j\in\Psi_i} \left(\widehat{b}_{ij} - d_{ij}\right)(d_{ij} - b_{ij}) I\left(|\widehat{b}_{ij}| > \delta_i\right) \\
&\quad + \sum_{i=0}^{q-1} \sum_{j\in\Psi_i} b_{ij}^2 I\left(|\widehat{b}_{ij}| \le \delta_i\right) + \sum_{i=q}^{\infty} \sum_{j\in\Psi_i} b_{ij}^2 \\
&=: I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8.
\end{aligned}$$

From Lemma 4.1, it is easy to see $I_2 = o\left((n^{-1}p)^{m\alpha}\right)$ and $E(I_5) = O\left(E(I_4)\right)$. We will show below that $E(I_1) = O\left((n^{-1}p)^{m\alpha}\right)$ and $E(I_4) = o\left(n^{-2rm\alpha/(2r+m\alpha)}\right)$, whether $g$ is smooth or only piecewise smooth. Hence, applying the Cauchy-Schwarz inequality, we can show $E(I_3)$ and $E(I_6)$ are of the order $o\left(\left(n^{-1}p\right)^{m\alpha}\right)$ and $o\left(n^{-2rm\alpha/(2r+m\alpha)}\right)$, which is negligible compared to the main terms of MISE.

So, we focus on terms $I_1,\ I_4,\ I_7$ and $I_8$. When $g$ is only piecewise smooth, let $\Pi$ denotes the finite set of points where $g^{(s)}$ has discontinuities for some $0 \le s \le r$. Suppose supp $\phi \subseteq (0, 1)$, supp $\psi \subseteq (0, 1)$ and let

$$\begin{aligned}
\mathbb{K} &= \{k : k \in (px, px + 1) \quad \text{for some} \quad x \in \Pi\}, \\
\mathbb{K}_i &= \{k : k \in (p_i x, p_i x + 1) \quad \text{for some} \quad x \in \Pi\}.
\end{aligned}$$

Also let $\mathbb{K}^c$, $\mathbb{K}_i^c$ denote their complements. Then, unless $j \in \mathbb{K}_i$, $b_{ij}$ and $\widehat{b}_{ij}$ are constructed entirely from an integral over or an average of data values from an interval where $g^{(r)}$ exists and is bounded. Also, unless $j \in \mathbb{K}$, $b_j$ and $\widehat{b}_j$ are constructed solely from such regions. Thus we may write

$$\begin{aligned} I_q(\Psi, \Psi_0, \Psi_1, \ldots) = & \; I_1(\mathbb{K}) + I_2 + I_3 + I_4(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) + I_5 + I_6 \\ & + I_7(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) + I_8(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) \\ & + I_1\left(\mathbb{K}^c\right) + I_4\left(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots\right) \\ & + I_7\left(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots\right) + I_8\left(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots\right), \end{aligned} \tag{72}$$

where

$$I_1(\mathbb{K}) = \sum_{j \in \mathbb{K}} \left(\widehat{b}_j - d_j\right)^2, \quad I_1\left(\mathbb{K}^c\right) = \sum_{j \in \mathbb{K}^c} \left(\widehat{b}_j - d_j\right)^2,$$

$$I_4(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) = \sum_{i=0}^{q-1} \sum_{j \in \mathbb{K}_i} \left(\widehat{b}_{ij} - d_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right),$$

$$I_4\left(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots\right) = \sum_{i=0}^{q-1} \sum_{j \in \mathbb{K}_i^c} \left(\widehat{b}_{ij} - d_{ij}\right)^2 I\left(|\widehat{b}_{ij}| > \delta_i\right),$$

the rest of the terms are defined similarly. However, for our compactly supported wavelets $\phi$ and $\psi$, both $\mathbb{K}$ and $\mathbb{K}_i$ have no more than $3(\#\Pi)$ elements for each $i$. Considering $q = O(\ln n)$, we can show $I_1(\mathbb{K})$, $I_4(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots)$, and $I_7(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots)$ are of the lower order $o\left((n^{-1}p)^{m\alpha}\right)$. Thus it is negligible compared to the main terms of MISE. Although $b_{ij}$ is only of the order $p_i^{-1/2}$ when $g$ is not $r$-times smooth, based on theorem's additional assumption $p_q^{2r+m\alpha} n^{-2rm\alpha} \to \infty$, we readily see that $I_8(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) = o\left(n^{-2rm\alpha/(2r+m\alpha)}\right)$. By tracing the whole proof of Theorem 3.1 carefully, noticing that when the error $\epsilon_k = H_m(\xi_k)$, there is no need to have terms $S_{24}$ and $S_{25}$ in Lemma 4.5 (Hence we don't need assumption $q = o(\ln n)$, which is contradicted with assumption $p_q^{2r+m\alpha} n^{-2rm\alpha} \to \infty$). Therefore, we will see the rest of the terms of the right hand side of (72) have precisely the asymptotic properties claimed for $\int(\widehat{g} - g)^2$ in Theorem 3.2.

**Acknowledgements** The authors thank the referees and the associate editor for their careful reading of the earlier version of the manuscript and for their valuable suggestions, which have greatly improved the presentation of this paper.

## References

Abry, P., Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory 44*, 2–15.

Beran, J. (1994). *Statistics for long memory processes*. New York: Chapman and Hall.

Bingham, N., Goldie, C., Teugels, J. (1987). *Regular variation*. Cambridge: Cambridge University Press.

Cohen, A., Daubechies, I., Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis 1*, 54–82.

Csörgö, S., Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics 23*, 1000–1014.

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia: SIAM.
Delbeke, L., Van Assche, W. (1998). A wavelet based estimator for the parameter of self-similarity of fractional Brownian motion. *Mathematical Contributions: Communications 24*, 65–76
Donoho, D. L., Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics 26*, 879–921.
Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *Journal of the Royal Statistical Society. Series B. Methodological 57*, 301–369.
Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics 24*, 508–539.
Fox, R., Taqqu, M. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *The Annals of Probability 13*, 428–446.
Hall, P., Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications 36*, 339–351.
Hall, P., Patil, P. (1995). Formulae for mean integated squared error of non-linear wavelet-based density estimators. *The Annals of Statistics 23*, 905–928.
Hall, P., Patil, P. (1996a). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *Journal of the Royal Statistical Society. Series B. Methodological 58*, 361–377.
Hall, P., Patil, P. (1996b). Effect of threshold rules on performance of wavelet-based curve estimators. *Statistica Sinica 6*, 331–345.
Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998). *Wavelets, approximation and statistical applications*. Lecture notes in statistics, vol. *129*. Berlin Heidelberg New York: Springer.
Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B. Methodological 53*, 173–187.
Johnstone, I. M. (1999). Wavelet threshold estimators for correlated data and inverse problems: Adaptivity results. *Statistica Sinica 9*, 51–83.
Johnstone, I. M., Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society. Series B. Methodological 59*, 319–351.
Kôno, N. (1983). Classical limit theorems for dependent random sequences having moment conditions. In: *Probability theory and mathematical statistics (Tbilisi, 1982)*, pp. 315–319, Lecture notes in mathematics, vol. *1021*. Berlin Heidelberg New York: Springer.
Kovac, A., Silverman, B. W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journal of the American Statistical Association*. *95*, 172–183.
Major, P. (1981). *Multiple Wiener-Itô Integrals*. Lecture notes in mathematics, vol. *849*. Berlin Heidelberg New York: Springer.
Móricz, F. (1976). Moment inequalities and the laws of large numbers. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete 35*, 299–314.
Robinson, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics 25*, 2054–2083.
Rozanov, Y. A. (1967). *Stationary random processes*. San Francisco: Holden Day.
Samorodnitsky, G., Taqqu, M. S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*. New York: Chapman & Hall.
Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete 31*, 287–302.
Taqqu, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete 40*, 203–238.
Tran, L. T., Roussas, G. G., Yakowitz, S., Truong Van, B. (1996). Fixed-design regression for linear time series. *The Annals of Statistics 24*, 975–991.
Truong, Y. K., Patil, P. N. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series. *The Annals of Institute of Statistical Mathematics 53*, 159–178.
von Sachs, R., Macgibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics 27*, 475–499.
Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *The Annals of Statistics 24*, 466–484.