WAVELET-BASED ESTIMATORS OF MEAN REGRESSION FUNCTION WITH LONG MEMORY DATA *

LI Lin-yuan (李林元)[1],   XIAO Yi-min (肖益民)[2]

(1. Department of Mathematics and Statistics, University of New Hampshire, 03824, USA;

2. Department of Statistics and Probability, Michigan State University, 48824, USA)

(Communicated by ZHOU Zhe-wei)

Abstract: This paper provides an asymptotic expansion for the mean integrated squared error (MISE) of nonlinear wavelet-based mean regression function estimators with long memory data. 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 the additional smoothness assumption is absent.

Key words: nonlinear wavelet-based estimator; nonparametric regression; long-range dependence

Introduction

Consider nonparametric regression

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

where $x_k = k/n \in [0,1], \varepsilon_1, \cdots, \varepsilon_n$ are observational errors with mean 0 and $g(x)$ is an unknown function to be estimated. Common assumptions on $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. errors or stationary processes with short-range dependence such as classic ARMA processes[1]. However, in many fields which include agronomy, astronomy, economics, environmental sciences, geosciences, hydrology and signal and image processing, it is unrealistic to assume that the observational errors are independent or short-range dependent. Instead, these observational errors exhibit slow decay in correlation which is often referred to as long-range dependence or long memory[2]. Suppose $\varepsilon_1, \cdots, \varepsilon_n$ are a stationary Gaussian process with mean 0 and variance 1. Then $\{\varepsilon_k, k \geq 1\}$ is said to have long-range dependence or long memory, if there exists $\alpha \in (0,1)$ such that

$$r(j) = E(\varepsilon_1 \varepsilon_{1+j}) \sim C_0 j^{-\alpha}, \tag{2}$$

where $C_0 > 0$ is a constant and $a_j \sim b_j$ means that $a_j/b_j \to 1$ when $j \to \infty$. Estimation for data with long-range dependence is quite different from that for observations with independence or short-range dependence. For example, Hall and Hart[3] showed that the convergence rates of mean regression function estimators differ from those with independence or short-range dependence. However, most of research so far assume that the regression function $g(x)$ is continuously differentiable.

Wavelet methods in nonparametric curve estimation have become a well-known technique. For a systematic discussion of wavelets and their applications in statistics, see the recent monograph by Hardle, *et al.*[4]. The major advantage of the wavelet method is its adaptability to the degree of smoothness of the underlying unknown curve. Hall and Patil[5] 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. However they consider the errors are independent. In this paper, we provide an asymptotic expansion for the mean integrated square error (MISE) of nonlinear wavelet-based mean regression function estimators under long memory Gaussian noise. 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 the additional smoothness assumption is absent.

In the next section, we give the elements of wavelet transform and provide nonlinear wavelet-based mean regression function estimators. The main results are described in Section 2, while their proofs appear in Sections 3.

1 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 properties: $\phi(x)$ and $\psi(x)$ are bounded and compactly supported; $\int \phi^2(x)dx = \int \psi^2(x)dx = 1$, $\nu_k \equiv \int y^k\psi(y)\,dy = 0$ for $0 \le k \le r-1$ and $\nu_r = \int y^r\psi(y)\,dy \ne 0$. Define

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

for arbitrary $p > 0$, $-\infty < j < \infty$ and $p_i = p2^i$, $i \ge 0$. Then, $\{\psi_{ij}(x),\ i,j \in \mathbb{Z}\}$ is a complete orthonormal basis for the space $L^2(\mathbb{R})$.

In our regression model, the mean function $g(x)$ is supported on a fixed unit interval $[0,1]$. Hence, without loss of generality, we may and will assume that $\phi(x)$ and $\psi(x)$ are compactly supported on $[0,1]$. We also assume that both $\phi(x)$ and $\psi(x)$ satisfy a uniform Hölder condition of exponent $1/2$, *i.e.*,

$$|\psi(x) - \psi(y)| \le C|x - y|^{1/2}, \text{ for all } \quad x, y \in [0,1]. \tag{3}$$

Daubechies[6] provides examples of wavelets satisfying these conditions.

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

$$g(x) = \sum_{j=0}^{p-1} b_j\phi_j(x) + \sum_{i=0}^{\infty}\sum_{j=0}^{p_i-1} b_{ij}\psi_{ij}(x), \quad x \in [0,1], \tag{4}$$

where

$$b_j = \int g(x)\phi_j(x)dx, \quad b_{ij} = \int g(x)\psi_{ij}(x)dx$$

are the wavelet coefficients of the function $g(x)$ and the series in Eq.(4) converges in $L^2([0,1])$.

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

$$\hat{g}(x) = \sum_{j=0}^{p-1} \hat{b}_j\phi_j(x) + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} \hat{b}_{ij}I(|\hat{b}_{ij}| > \delta_i)\psi_{ij}(x), \tag{5}$$

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

$$\hat{b}_j = \frac{1}{n}\sum_{k=1}^{n} Y_k \phi_j(x_k), \qquad \hat{b}_{ij} = \frac{1}{n}\sum_{k=1}^{n} Y_k \psi_{ij}(x_k). \tag{6}$$

In this paper, we use C to denote positive and finite constants whose value may change from line to line. Specific constants are denoted by C_0, C_1, C_2 and so on.

2   Main Results

Throughout this paper, the smoothing parameters p, q, and δ_i are functions of n. We assume they satisfy the following condition: as $n \to \infty$,

(SP): $p \to \infty, \quad q \to \infty, \quad p_i \delta_i^2 \to 0, \quad p_i^{2r+1}\delta_i^2 \to \infty, \quad \delta_i \geq \sqrt{\frac{4C_1 \ln n}{n^\alpha p_i^{1-\alpha}}}, \quad i = 1, 2, \cdots, q-1,$

where $C_1 = C_0 \iint |x-y|^{-\alpha}\psi(x)\psi(y)dxdy.$

Theorem 1   *In addition to the conditions on $\phi(x)$ and $\psi(x)$ stated in Section 1, assume that the r-th derivative $g^{(r)}(x)$ is continuous and bounded. Also assume condition (SP) holds. Then, as $n \to \infty$,*

$$E\left| \int (\hat{g}(x) - g(x))^2 dx - \left\{ C_2(n^{-1}p)^\alpha + p^{-2r}\kappa^2(1 - 2^{-2r})^{-1} \int g^{(r)^2}(x)dx \right\} \right|$$
$$= o\left((n^{-1}p)^\alpha + p^{-2r} \right), \tag{7}$$

where $\kappa = (r!)^{-1}\nu_r$ and $C_2 = C_0 \iint |x-y|^{-\alpha}\phi(x)\phi(y)dxdy.$

In Theorem 1, we have assumed that the mean regression function $g(x)$ is r-times continuously differentiable for simplicity and convenience of the exposition. However, if $g^{(r)}(x)$ is only piecewise continuous, Theorem 1 still holds, as given in the following.

Theorem 2   *In addition to the conditions on $\phi(x)$ and $\psi(x)$ stated in Section 1, assume that the r-th derivative $g^{(r)}(x)$ 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 $g(x)$ exist and are bounded and continuous on (x_i, x_{i+1}) for $0 \leq i \leq N$, with left and right limits. In particular, $g(x)$ itself may be only piecewise continuous. Also assume that condition (SP) holds and $p_q^{2r+\alpha}n^{-2\alpha r} \to \infty$. Then Eq.(7) still holds.*

Remark 1   Hall and Hart[3] considered kernel estimator in fixed-design nonparametric regression when the error is a Gaussian long memory process, giving a similar asymptotic expansion for MISE. However they assume that the regression function g is continuously differentiable. Our result is stronger than the traditional asymptotic expansion for MISE. In particular, Eq.(7) implies a wavelet version of the MISE expansion:

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

However, for kernel estimators, the above expansion usually fails without the assumption that $g(x)$ is r-times continuously differentiable.

3   Proofs

The ideas of proving Theorems 1 and 2 come from the proofs of Theorems 2.1 and 2.2 in Ref.[5]. The difference is that we consider the errors $\{\epsilon_k, k \geq 1\}$ to be long memory Gaussian

noise here, instead of i.i.d. random variables in their paper. Before we present the proof of Theorem 1, we begin with some lemmas.

Denote

$$d_j = E(\hat{b}_j) = \frac{1}{n}\sum_{k=1}^{n}g(x_k)\phi_j(x_k), \quad d_{ij} = E(\hat{b}_{ij}) = \frac{1}{n}\sum_{k=1}^{n}g(x_k)\psi_{ij}(x_k).$$

The following lemma will be used for proving Lemmas 2 and 3.

Lemma 1 *Suppose the function $g(x)$ is continuously differentiable on $[0,1]$ and the wavelets $\phi(x)$ and $\psi(x)$ satisfy the uniform Hölder conditions (3). Then*

$$\sup_j|d_j - b_j| = O(n^{-1/2}), \tag{8}$$

$$\sup_j|d_{ij} - b_{ij}| = O(n^{-1/2}). \tag{9}$$

Proof We only prove Eq.(8). The proof of Eq.(9) is similar. First we write

$$d_j = \frac{p^{1/2}}{n}\sum_{k=1}^{n}g\Big(\frac{k}{n}\Big)\phi\Big(\frac{pk}{n} - j\Big). \tag{10}$$

For fixed n, p and j, we note that

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

Let $m_j = \lfloor\frac{nj}{p}\rfloor$, where $\lfloor x\rfloor$ denotes the smallest integer that is at least x. Since $\phi(x)$ has its support in $[0,1]$, the summation in Eq.(10) runs from m_j to $m_{j+1} - 1$. However, for simplicity of the notation, we will not distinguish between $\lfloor x\rfloor$ and x. Thus, let $k = m_j + \ell$,

$$
\begin{aligned}
d_j &= \frac{p^{1/2}}{n}\sum_{\ell=0}^{n/p-1}g\Big(\frac{\ell}{n} + \frac{j}{p}\Big)\phi\Big(\frac{p\ell}{n}\Big) \qquad \Big(\text{let} \quad t_\ell = \frac{p\ell}{n}\Big) \\
&= \frac{1}{p^{1/2}}\sum_{\ell=0}^{n/p-1}g\Big(\frac{t_\ell + j}{p}\Big)\phi(t_\ell)\frac{p}{n}. \tag{11}
\end{aligned}
$$

Similarly, by a simple change of variables, we have, let $t = px - j$,

$$
\begin{aligned}
b_j &= p^{1/2}\int_{j/p}^{(j+1)/p}g(x)\,\phi(px - j)\,dx \\
&= \frac{1}{p^{1/2}}\int_{0}^{1}g\Big(\frac{t + j}{p}\Big)\phi(t)\,dt. \tag{12}
\end{aligned}
$$

Combining Eq.(11) and Eq.(12), we have $|d_j - b_j| = J_1 + J_2$, where

$$J_1 = \frac{1}{p^{1/2}}\sum_{\ell=0}^{n/p-1}\int_{\frac{p\ell}{n}}^{\frac{p(\ell+1)}{n}}\Big[g\Big(\frac{t_\ell + j}{p}\Big) - g\Big(\frac{t + j}{p}\Big)\Big]\phi(t_\ell)\,dt,$$

$$J_2 = \frac{1}{p^{1/2}}\sum_{\ell=0}^{n/p-1}\int_{\frac{p\ell}{n}}^{\frac{p(\ell+1)}{n}}g\Big(\frac{t + j}{p}\Big)\Big[\phi(t_\ell) - \phi(t)\Big]\,dt.$$

To estimate J_1, we use the continuous differentiability of $g(x)$ (in fact, it is enough if $g(x)$ is a Lipschitz function on $[0, 1]$) and the boundedness of $\phi(x)$ to get $J_1 \leq p^{-1/2} \cdot Cn^{-1}$, where $C > 0$ is a finite constant. For J_2, we use the boundedness of $g(x)$ and the uniform 1/2-Hölder condition (3) for $\phi(x)$ to derive $J_2 \leq p^{-1/2} \cdot C \left(pn^{-1}\right)^{1/2} = C\,n^{-1/2}$. It is clear that Eq.(8) follows from J_1 and J_2.

Lemma 2   *Under the assumptions of Theorem* 1,

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

Proof   Put $Q_n = \sum_{j=0}^{p-1} (\hat{b}_j - d_j)^2$ and $\mu_n = E(Q_n)$. Note that

$$\mu_n = \frac{p}{n^2} \sum_{j=0}^{p-1} \sum_{k=1}^{n} \sum_{l=1}^{n} r(k - l)\phi(px_k - j)\phi(px_l - j).$$

For each fixed $j = 0, 1, \ldots, p - 1$, similar to Eq.(11), we have

$$\frac{p}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} r(k - l)\phi(px_k - j)\phi(px_l - j)$$
$$= p^{-1}C_0\left(n^{-1}p\right)^\alpha \left[\iint |x - y|^{-\alpha}\phi(x)\phi(y)dxdy + o(1) \right], \tag{13}$$

where the last equality follows from Eq.(2) and a standard limiting argument. Hence $\mu_n = C_2\left(n^{-1}p\right)^\alpha + o\left((n^{-1}p)^\alpha\right)$. Using the triangle inequality and the Cauchy-Schwartz inequality, we can obtain

$$S_1 \leq E^{1/2}(Q_n - \mu_n)^2 + \sum_{j=0}^{p-1}(d_j - b_j)^2$$
$$+ 2\left[\sum_{j=0}^{p-1} E(\hat{b}_j - d_j)^2 \sum_{j=0}^{p-1}(d_j - b_j)^2 \right]^{1/2} + o\left((n^{-1}p)^\alpha\right)$$
$$= S_{11} + S_{12} + S_{13} + o\left((n^{-1}p)^\alpha\right). \tag{14}$$

It follows from Eq.(8) in Lemma 1 that $\sup_j |d_j - b_j| = O(n^{-1/2})$. Thus,

$$S_{12} = O(n^{-1}p) = o\left((n^{-1}p)^\alpha\right),$$
$$S_{13} = O\left((n^{-1}p)^{1/2}\mu_n^{1/2}\right) = o\left((n^{-1}p)^\alpha\right). \tag{15}$$

As to the term S_{11}, we put $a_{kl} = pn^{-2} \sum_{j=0}^{p-1} \phi(px_k - j)\phi(px_l - j)$. Then we can write $Q_n - \mu_n = \sum_{k=1}^{n} \sum_{l=1}^{n} a_{kl}\varepsilon_k\varepsilon_l - \mu_n$. Note that Q_n is a quadratic form of a stationary sequence of Gaussian random variables 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 Refs.[7] and [8] for more information.

Since $\{\varepsilon_k, k \geq 1\}$ is a stationary Gaussian process, it follows from the S. Bochner theorem that its covariance function $r(k)$ has the following spectral representation $r(k) = \int_{[-\pi,\pi]} e^{ikx}G(dx)$, $k \in \mathbb{Z}$, where G is a Borel measure on $[-\pi, \pi]$ which is called the spectral measure of $\{\varepsilon_k, k \geq 1\}$.

Let Z_G be the corresponding random spectral measure, *i.e.*, the complex-valued Gaussian scattered measure with $E(Z_G(A))^2 = G(A)$ for all Borel set $A \subset [-\pi, \pi]$, then for all $k \geq 1$, $\varepsilon_k = \int_{[-\pi,\pi]} e^{ikx} dZ_G(x)$. By Lemma 3.5 of Ref.[8] with $m = 1$, we can write $Q_n - \mu_n = \sum_{k=1}^{n} \sum_{l=1}^{n} a_{kl} K_1(k, l)$, where $K_1(k, l)$ is the double Wiener-Itô-Dobrushin integral,

$$K_1(k, l) = \int_{[-\pi,\pi]^2}'' e^{ikx_1 + ilx_2} dZ_G(x_1) dZ_G(x_2).$$

Furthermore, it follows from Lemma 3.6 of Ref.[8] with $m = 1$ that for any integers k, s, l, t,

$$E\Big[K_1(k, k+s) K_1(l, l+t)\Big]$$
$$= r(k - l + s) r(k - l - t) + r(k - l) r(k - l + s - t).$$

Thus, we have

$$E(Q_n - \mu_n)^2 = \sum_{k_1} \sum_{k_2} \sum_{l_1} \sum_{l_2} a_{k_1 l_1} a_{k_2 l_2} \Big[r(l_1 - k_2) r(k_1 - l_2) + r(k_1 - k_2) r(l_1 - l_2) \Big]$$
$$=: A_{n1} + A_{n2}.$$

We first estimate A_{n1}. Since $r(k) \sim C_0 k^{-\alpha}$ as $k \to \infty$, similar arguments to the proof of Eq.(13) yield

$$A_{n1} = C_0^2 p^{2\alpha - 2} n^{-2\alpha} \sum_{j_1} \sum_{j_2} \Big[\iint |x - y + j_1 - j_2|^{-\alpha} \phi(x)\phi(y) dx dy \Big]^2 + o\big((n^{-1}p)^{2\alpha}\big)$$
$$=: A_{n1}(1) + o\big((n^{-1}p)^{2\alpha}\big).$$

Let $k = j_1 - j_2$ and change the order of the summation, we have

$$A_{n1}(1) = C_0^2 p^{2\alpha - 2} n^{-2\alpha} \bigg\{ \sum_{k=0}^{p-1} (p - k) \Big[\iint |x - y + k|^{-\alpha} \phi(x)\phi(y) dx dy \Big]^2$$
$$+ \sum_{k=-(p-1)}^{-1} (p + k) \Big[\iint |x - y + k|^{-\alpha} \phi(x)\phi(y) dx dy \Big]^2 \bigg\}.$$

Since $\phi(x)$ is bounded and compactly supported, we have

$$\Big| \iint |x - y + k|^{-\alpha} \phi(x)\phi(y) dx dy \Big| = O\Big(|k|^{-\alpha}\Big) \quad \text{as } k \to \infty.$$

Thus,

$$A_{n1}(1) = O\big(p^{2\alpha - 2} n^{-2\alpha} p^{2 - 2\alpha}\big) = O\Big(\big(n^{-1}p\big)^{2\alpha} p^{-2\alpha}\Big) = o\Big(\big(n^{-1}p\big)^{2\alpha}\Big).$$

Hence, we have $A_{n1} = o\big((n^{-1}p)^{2\alpha}\big)$. Similarly, we can show that $A_{n2} = o\big((n^{-1}p)^{2\alpha}\big)$ also holds. Thus, we obtain $S_{11} = o\big((n^{-1}p)^{\alpha}\big)$. Combining with Eqs.(14) and (15), we complete the proof.

Lemma 3 *Under the assumption of Theorem 1,*

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

Proof   The proof is analogous to part 1 of the proof of Theorem 2.1 of Ref.[5]. Let λ and β denote positive numbers satisfying $\lambda + \beta = 1$, and set

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

$$S_{22}^{'} = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} E\Big\{(\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij} - b_{ij}| > \beta\delta_i)\Big\}.$$

The triangle inequality implies $S_2 \leq S_{21}^{'} + S_{22}^{'}$. Replacing b_{ij} in the expectation by $d_{ij} = E(\hat{b}_{ij})$, we define

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

$$S_{22} = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} E\Big\{(\hat{b}_{ij} - d_{ij})^2 I(|\hat{b}_{ij} - d_{ij}| > \beta\delta_i)\Big\}.$$

Using Eq.(9) in Lemma 1, the fact $n\delta_i^2 \to \infty$ and applying the argument as that in Eq.(14), we see that S_{21} and S_{22} are the leading terms of $S_{21}^{'}$ and $S_{22}^{'}$. Thus, we can show that $S_{21}^{'} = O(S_{21})$ and $S_{22}^{'} = O(S_{22})$. Hence, in order to prove the lemma, it suffices to prove $S_{21} = o(n^{-2\alpha r/(2r+\alpha)})$ and $S_{22} = o(n^{-2\alpha r/(2r+\alpha)})$, respectively. Since the mean regression function $g(x)$ is r-times continuously differentiable, by using the Taylor expansion of $g(x)$ and the moment condition on the mother wavelet, we have

$$b_{ij} = \kappa p_i^{-(r+1/2)}(g_{ij} + \xi_{ij}), \tag{16}$$

where $g_{ij} = g^{(r)}\left(\frac{j}{p_i}\right)$ and $\displaystyle\sup_{0\leq j\leq p_i-1; \, 0\leq i\leq q-1} |\xi_{ij}| \to 0.$

Hence, we have $|b_{ij}| \leq Cp_i^{-(r+1/2)}$. Also, through direct calculation like μ_n in Lemma 2, we have $E(\hat{b}_{ij} - d_{ij})^2 \leq Cp_i^{-1}(n^{-1}p_i)^\alpha$. Hence,

$$S_{21} \leq \sum_{i=0}^{q-1} C(n^{-1}p_i)^\alpha I\big(p_i \leq C\delta_i^{-2/(2r+1)}\big)$$

$$= Cn^{-\alpha}p^\alpha \sum_{i=0}^{q-1} 2^{\alpha i} I\big(p_i \leq C\delta_i^{-2/(2r+1)}\big). \tag{17}$$

Note that $p_i \leq C\delta_i^{-2/(2r+1)}$ implies $(p2^i)^{2r+1} \leq Cn^\alpha p^{1-\alpha}(\ln n)^{-1} 2^{(1-\alpha)i}$, that is

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

There are only finitely many i satisfying this inequality. We denote the largest such i by m. Then, Eq.(17) becomes

$$S_{21} \leq Cn^{-\alpha}p^\alpha \sum_{i=0}^{m} 2^{\alpha i} \leq Cn^{-\alpha}p^\alpha \Big[\frac{n^\alpha p^{-(2r+\alpha)}}{\ln n}\Big]^{\frac{\alpha}{2r+\alpha}} = o\big(n^{-2\alpha r/(2r+\alpha)}\big). \tag{18}$$

In order to estimate S_{22}, we apply the Cauchy-Schwartz inequality to derive

$$S_{22} \leq \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} E^{1/2}\Big[\big(\hat{b}_{ij} - d_{ij}\big)^4\Big] P^{1/2}\big(|\hat{b}_{ij} - d_{ij}| > \beta\delta_i\big).$$

Since $\hat{b}_{ij} - d_{ij}$ is a normal random variable with zero mean and

$$E\big[(\hat{b}_{ij} - d_{ij})^2\big] = C_1 n^{-\alpha} p_i^{-1+\alpha} + o(n^{-\alpha} p_i^{-1+\alpha}),$$

we can find a constant C such that $E\big[(\hat{b}_{ij} - d_{ij})^4\big] \le CE^2\big[(\hat{b}_{ij} - d_{ij})^2\big] \le C(n^{-1}p_i)^{2\alpha} p_i^{-2}$. By the well known tail probability inequality for the normal random variable, we have, for any $\epsilon > 0$,

$$\begin{aligned}
P\big(|\hat{b}_{ij} - d_{ij}| > \beta\delta_i\big) &\le P\big(|Z| > (1-\epsilon)\beta\sqrt{4\ln n}\big)\\
&\le \frac{C}{(1-\epsilon)\beta\sqrt{4\ln n}} e^{-2(1-\epsilon)^2\beta^2 \ln n}\\
&= O(n^{-2(1-\epsilon)^2\beta^2}),
\end{aligned}$$

where Z is a standard normal random variable. Thus, we obtain

$$\begin{aligned}
S_{22} &\le C \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} \big(n^{-1}p_i\big)^\alpha p_i^{-1} n^{-(1-\epsilon)^2\beta^2}\\
&= Cn^{-(1-\epsilon)^2\beta^2} \big(n^{-1}p_q\big)^\alpha.
\end{aligned}$$

From the assumption that $p_i \delta_i^2 \to 0$ in (SP), we obtain $n^{-1}p_q \to 0$. Choose $\beta < 1$ close to 1 and ϵ close to 0, we have $S_{22} = o\big(n^{-2\alpha r/(2r+\alpha)}\big)$, which, combined with (18), proves the lemma.

The following two lemmas are analogous to these in Ref.[5], for the sake of brevity, their proofs are omitted.

Lemma 4   *Under the assumption of Theorem 1,*

$$\begin{aligned}
S_3 &\equiv E\bigg| \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} b_{ij}^2 I(|\hat{b}_{ij}| \le \delta_i) - p^{-2r}\kappa^2(1-2^{-2r})^{-1} \int g^{(r)^2}(x)dx \bigg|\\
&= o(p^{-2r}).
\end{aligned}$$

Lemma 5   *Under the assumption of Theorem 1, $S_4 \equiv \sum_{i=q}^{\infty} \sum_{j=0}^{p_i-1} b_{ij}^2 = o(p^{-2r})$.*

We are now in the position to give the proof of the Theorems 1 and 2.

Proof of Theorem 1   The proof follows from Lemmas 2–5 and the fact that

$$E\bigg| \int (\hat{g}(x) - g(x))^2 dx - \bigg\{ C_1(n^{-1}p)^\alpha + p^{-2r}\kappa^2(1-2^{-2r})^{-1} \int g^{(r)^2}(x)dx \bigg\} \bigg|$$
$$\le S_1 + S_2 + S_3 + S_4.$$

Proof of Theorem 2   We use the same notation as in Ref.[5]. Notice that, by the orthogonality properties of $\phi(x)$ and $\psi(x)$, $\int (\hat{g}(x) - g(x))^2 dx = I_q(\mathbb{Z}, \mathbb{Z}, \ldots)$, where $\mathbb{Z}$ denotes the corresponding set of integers (for instance, $\Psi_i = \{0, 1, \cdots, p_i - 1\}$) and

$$\begin{aligned}
I_q(\Psi, \Psi_0, \Psi_1, \ldots) =& \sum_{j \in \Psi} (\hat{b}_j - b_j)^2 + \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} (\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta_i)\\
&+ \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} b_{ij}^2 I(|\hat{b}_{ij}| \le \delta_i) + \sum_{i=q}^{\infty} \sum_{j \in \Psi_i} b_{ij}^2
\end{aligned}$$

$$= \sum_{j \in \Psi} (\hat{b}_j - d_j)^2 + \sum_{j \in \Psi} (d_j - b_j)^2 + 2 \sum_{j \in \Psi} (\hat{b}_j - d_j)(d_j - b_j)$$

$$+ \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} (\hat{b}_{ij} - d_{ij})^2 I(|\hat{b}_{ij}| > \delta_i) + \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} (d_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta_i)$$

$$+ 2 \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} (\hat{b}_{ij} - d_{ij})(d_{ij} - b_{ij}) I(|\hat{b}_{ij}| > \delta_i)$$

$$+ \sum_{i=0}^{q-1} \sum_{j \in \Psi_i} b_{ij}^2 I(|\hat{b}_{ij}| \le \delta_i) + \sum_{i=q}^{\infty} \sum_{j \in \Psi_i} b_{ij}^2$$

$$=: I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8.$$

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

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

$$\mathbb{K} = \{k : k \in (px - v, px + v) \quad \text{for some} \quad x \in \Pi\},$$
$$\mathbb{K}_i = \{k : k \in (p_i x - v, p_i x + v) \quad \text{for some} \quad x \in \Pi\}.$$

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

$$I_q(\Psi, \Psi_0, \Psi_1, \ldots) = I_1(\mathbb{K}) + I_2 + I_3 + I_4(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) + I_5 + I_6$$
$$+ I_7(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) + I_8(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots)$$
$$+ I_1(\mathbb{K}^c) + I_4(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots)$$
$$+ I_7(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots) + I_8(\mathbb{K}_0^c, \mathbb{K}_1^c, \mathbb{K}_2^c, \ldots), \tag{19}$$

where

$$I_1(\mathbb{K}) = \sum_{j \in \mathbb{K}} (\hat{b}_j - d_j)^2, \qquad I_1(\mathbb{K}^c) = \sum_{j \in \mathbb{K}^c} (\hat{b}_j - d_j)^2,$$

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

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

the rest of the terms are defined similarly. However, for our compactly supported wavelets $\phi(x)$ and $\psi(x)$, both $\mathbb{K}$ and $\mathbb{K}_i$ have no more than $(2v + 1)(\#\Pi)$ 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 $o((n^{-1}p)^\alpha)$. Thus it is negligible compared to the main terms of MISE. Although b_{ij} is only of the order $p_i^{-1/2}$ when $g(x)$ is not r-times smooth, based on theorem's additional

assumption $p_q^{2r+\alpha}n^{-2\alpha r} \to \infty$, we readily see that $I_8(\mathbb{K}_0, \mathbb{K}_1, \mathbb{K}_2, \ldots) = o\big(n^{-2\alpha r/(2r+\alpha)}\big)$. By tracing the whole proof of Theorem 1 carefully, we will see the rest of the terms of the right hand side of Eq.(19) have precisely the asymptotic properties claimed for $\int(\hat{g}(x) - g(x))^2 dx$ in Theorem 2. The proof is finished.

References

[1] Hart J D. Kernel regression estimation with time series errors[J]. *J Roy Statist Soc, Ser B*, 1991, **53**(1):173–187.

[2] Beran J. Statistics for Long Memory Processes[M]. Chapman and Hall, New York, 1994, 248–260.

[3] Hall P, Hart J D. Nonparametric regression with long-range dependence[J]. *Stochastic Process Appl*, 1990, **36**(2):339–351.

[4] Hardle W, Kerkyacharian G, Picard D, Tsybakov A. Wavelets, Approximation and Statistical Applications[M]. Lecture Notes in Statistics, 129, Springer-Verlag, New York, 1998, 125–212.

[5] Hall P, Patil P. Formulae for mean integated squared error of non-linear wavelet-based density estimators[J]. *Ann Statist,* 1995, **23**(3):905–928.

[6] Daubechies I. Ten Lectures on Wavelets[M]. SIAM, Philadelphia, 1992, 170–185.

[7] Mojor P. Multiple Wiener-Itô Integrals[M]. Lecture Notes in Math, 849, Springer-Verlag, New York, 1981, 88–120.

[8] Fox R, Taqqu M. Noncentral limit theorems for quadratic forms in random variables having long-range dependence[J]. *Ann Probab,* 1985, **13**(2):428–446.