Asymptotics for L_1-wavelet method for nonparametric regression

Xingcai Zhou[1,2] and Fangxia Zhu[3*]

*Correspondence:
zhufangxia75@163.com
[3]School of Mathematics and
Finance, Chuzhou University,
Chuzhou, Anhui 239001, China
Full list of author information is
available at the end of the article

Abstract

Wavelets are particularly useful because of their natural adaptive ability to characterize data with intrinsically local properties. When the data contain outliers or come from a population with a heavy-tailed distribution, L_1-estimation should obtain a better fit. In this paper, we propose a L_1-wavelet method for nonparametric regression, and derive the asymptotic properties of the L_1-wavelet estimator, including the Bahadur representation, the rate of convergence and asymptotic normality. The rate of convergence of it is comparable with the optimal convergence rate of the nonparametric estimation in nonparametric models, and it does not require the continuously differentiable conditions of a nonparametric function.

MSC: 62G05; 60F05

Keywords: L_1-wavelet; Nonparametric regression; Bahadur representation; Rate of convergence; Asymptotic normality

1 Introduction

Consider the problem of estimating the underlying regression function from a set of noisy data. The nonparametric regression is an underlying framework. It has the following standard form:

$$y_i = g(t_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where the y_i are the noisy samples of an unknown function $g(\cdot)$ defined on $[0, 1]$, $\{t_i\}$ are non-random design points with $0 \le t_1 \le \cdots \le t_n \le 1$, and the ϵ_i are i.i.d. random errors with mean zero.

Nonparametric regression is a classic smoothing technique for recovering a signal function from data, without having a strong prior restriction on its form [1, 2]. There is much literature on nonparametric regression. Most methods developed so far are based on the mean regression function by using a least-squares estimation (L_2). For example, Gasser and Müller [3] proposed a kernel estimation by a Gasser–Müller kernel weight; Fan [4, 5] added more insights to the local linear method for the mean regression; and Braun and Huang [6] proposed a kernel spline regression by replacing the polynomial approximation for local polynomial kernel regression with a spline basis. The least-square-based meth-

Springer

ods certainly have some nice properties as regards Gaussian errors, but this method will not perform well because of the high sensitiveness to extreme outliers, specially to the errors having a heavy-tailed distribution. More robust estimation methods are required. The local median and M-estimators have been studied; see, for example, [7–13]. Also see [14, 15] for more details on quantile regression and robust estimation, respectively. As pointed out in [9], among many robust estimation methods, the L_1 method based on least absolute deviations behaves quite well because of downweight outliers, unique solutions and no transition point in the influence function (such as the additional parameter c in Huber's ρ_c function). The above methods basically require that the unknown function g has high smoothness. But in reality, the condition may not be satisfied. In fact, objects of some practical areas, such as signal and image processing, are frequently inhomogeneous. In this paper, we consider the wavelet technique to recover the signal function g based on L_1 method for the robust case.

We aim to study the asymptotic properties on L_1-wavelet estimator for the nonparametric model (1.1). Wavelet techniques, due to their ability to adapt to local features of curves, have received a lot of attention, and have been used to estimate the nonparametric curve. See, for example, [16–19]. Wavelet methods are prominent because of their computational ease and having the minimax results over very wide classes of function spaces for the signal function g. For linear wavelet smoothing, Antoniadis et al. [16] is a key reference that introduces wavelet versions of some classical kernel and orthogonal series estimators, and studies their asymptotic properties such as mean square consistent, bias, variance and asymptotic normal. Huang [20] also gave asymptotic bias and variance of the wavelet density estimator by wavelet-based reproducing kernels. Zhou and You [21] constructed wavelet estimators for varying-coefficient partially linear regression models, and established their asymptotic normalities and some convergence rates. For varying-coefficient models, the convergence rate and asymptotic normality of wavelet estimators were considered by [22, 23] provided asymptotic bias and variance of wavelet estimator for regression function under a mixing stochastic process. Recently, Chesneau et al. [24] proposed the nonparametric wavelet estimators of the quantile density function and its consistency. Li and Xiao [25] considered a wavelet estimator for the mean regression function with strong mixing errors and investigated their asymptotic rates of convergence by using the thresholding of the empirical wavelet coefficients. Berry–Esseen type bounds for wavelet estimators for semiparametric regression models were studied by [26, 27]. For the nonparametric models (1.1), as we learned, no study on L_1-wavelet estimators is reported. For this model, the estimation should be combined with the special feature of the model.

In this paper, we develop L_1-wavelet method for nonparametric regression model (1.1) by adopting wavelet to detect and represent localized features of the signal function g, and applying L_1 to yield better recovery for outliers or heavy-tailed data. The advantage of L_1-wavelet method is in avoiding the restrictive smoothness requirement for nonparametric function of the traditional smoothing approaches, such as kernel and local polynomial methods, and to robustify the usual mean regression. Last, we investigate asymptotic properties of the L_1-wavelet estimators, including the Bahadur representation, the rate of convergence and asymptotic normality.

The paper is organized as follows. In Sect. 2, we provide some necessary background on wavelet and develop L_1-wavelet estimation for the model (1.1). Asymptotic properties of L_1-wavelet estimators are presented in Sect. 3. Technical proofs are deferred to Sect. 4.

2 L_1-Wavelet estimation

Wavelet analysis requires a description of two related and suitably chosen orthonormal basic functions: the scaling function ϕ and the wavelet ψ. A wavelet system is generated by dilation and translation of ϕ and ψ through

$$\phi_{m,k}(t) = 2^{m/2}\phi\left(2^m t - k\right), \qquad \psi_{m,k}(t) = 2^{m/2}\psi\left(2^m t - k\right), \quad m, k \in \mathbb{Z}.$$

A multiresolution analysis of $\mathcal{L}^2(\mathbb{R})$ consists of a nested sequence of closed subspace V_m, $m \in \mathbb{Z}$ of $\mathcal{L}^2(\mathbb{R})$,

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots ,$$

where $\mathcal{L}^2(\mathbb{R})$ is the set of square integral functions over real line. Since $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthogonal family of $\mathcal{L}^2(\mathbb{R})$ and V_0 is the subspace spanned, $\{\phi_{0k}, k \in \mathbb{Z}\}$ and $\{\phi_{mk}, k \in \mathbb{Z}\}$ are the orthogonal bases of V_0 and V_m, respectively. From the Moore–Aronszajn theorem [28], it follows that

$$E(t,s) = \sum_k \phi(t-k)\phi(s-k)$$

is a reproducing kernel of V_0. By self-similarity of multiresolution subspaces,

$$E_m(t,s) = 2^m E\left(2^m t, 2^m s\right)$$

is a reproducing kernel of V_m. Thus, the projection of g on the space V_m is given by

$$\mathbb{P}_{V_m} g(t) = \int 2^m E\left(2^m t, 2^m s\right) g(s)\, ds.$$

This motivates us to define a L_1-wavelet estimator of g by

$$\hat{g}(t) = \operatorname{argmin}_a \sum_{i=1}^n |y_i - a| \int_{A_i} E_m(t,s)\, ds, \tag{2.1}$$

where A_i are intervals that partition $[0, 1]$, so that $t_i \in A_i$. One way of defining the intervals $A_i = [s_{i-1}, s_i)$ is by taking $s_0 = 0$, $s_n = 1$, and $s_i = (t_i + t_{i+1})/2$, $i = 1, \ldots, n-1$.

For the ith sample point, define e_i^+ and e_i^- to be the positive and negative parts of e_i. Then, with the noisy samples, problem (2.1) can be reduced to the following linear program:

$$\text{Minimize} \quad \sum_{i=1}^n \left(e_i^+ + e_i^-\right) \int_{A_i} E_m(t,s)\, ds,$$

$$\text{Subject to} \quad \boldsymbol{a} + \boldsymbol{e}^+ - \boldsymbol{e}^- = \boldsymbol{y},$$

$$\boldsymbol{e}^+, \boldsymbol{e}^- \geq 0,$$

where $\boldsymbol{a} = a\mathbf{1}_p$, $\mathbf{1}_p$ is a p dimensional vector whose each component is 1, $\boldsymbol{e}^+ = (e_1^+, \ldots, e_n^+)^T$, $\boldsymbol{e}^- = (e_1^-, \ldots, e_n^-)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^T$. In addition, $\int_{A_i} E_m(t,s)\,ds$ can be calculated by the cascade algorithm given by [16]. Thus, the L_1-wavelet estimator can be easily obtained. This linear program is just for calculating the estimator. To establish the asymptotic properties, we work only with (2.1).

3 Asymptotic properties

We begin with the following assumptions required to derive the asymptotic properties of the proposed estimator in Sect. 2.

(A1) The noisy errors ϵ_i are i.i.d. with median 0 and a continuous, positive density f_ϵ in a neighborhood of 0.

(A2) g belongs to the Sobolev space $\mathcal{H}^\nu(\mathbb{R})$ with order $\nu > 1/2$.

(A3) g satisfies the Lipschitz of order condition of order $\gamma > 0$.

(A4) ϕ has a compact support, and is in the Schwarz space with order $l > \nu$, it satisfies the Lipschitz condition with order l. Furthermore, $|\hat{\phi}(\xi) - 1| = O(\xi)$ as $\xi \to 0$, where $\hat{\phi}$ is the Fourier transform of ϕ.

(A5) $\max_i |t_i - t_{i-1}| = O(n^{-1})$.

(A6) We also assume that, for some Lipschitz function $\kappa(\cdot)$,

$$\rho(n) = \max_i \left| s_i - s_{i-1} - \frac{\kappa(s_i)}{n} \right| = o(n^{-1}).$$

(A7) (i) $n2^{-m} \to \infty$; (ii) $2^m = O(n^{1-2p})$, $1/(2+\delta) \le p \le 1/2$ for $\delta > 0$. (iii) Let $\nu^* = \min(3/2, \nu, \gamma + 1/2) - \epsilon_1$ and $\epsilon_1 = 0$ for $\nu \ne 3/2$, $\epsilon_1 > 0$ for $\nu = 3/2$. Assume that $n2^{-2m\nu^*} \to 0$.

Remark 3.1 The above conditions are mild and easily satisfied. (A1) is crucial to the asymptotic behavior of $\hat{g}(\cdot)$ based on L_1 estimation (2.1). (A2)–(A6) and (A7)(i)(iii) have been used in [16]. Note that, if $g \in \mathcal{H}^\nu(\mathbb{R})$ with $\nu > 3/2$, then g is continuously differentiable; thus (A3) is redundant when $\nu > 3/2$. So, (A2) is weaker than smoothness. For (A7), m acts as a tuning parameter, such as the bandwidth does for standard kernel smoothers; (A7) is the standard assumption for asymptotic behaviors; For example, take $\delta = 2$ and $p = 1/4$, then $2^m = O(n^{1/2})$. Thus, (i) holds; furthermore, taking $\nu > 1$ and $\gamma > 1/2$, then (iii) holds.

Our results are as follows.

Theorem 3.1

(i) (*Bahadur representation*) *Suppose that* (A1)–(A5) *and* (A7)(i) *hold*, *then*

$$\hat{g}(t) = g(t) - \frac{1}{2}f_\epsilon^{-1}(0) \sum_{i=1}^n \operatorname{sign}(\epsilon_i) \int_{A_i} E_m(t,s)\,ds + R_n(m; \gamma, \nu),$$

with

$$R_n(m; \gamma, \nu) = O_p\left\{ n^{-\gamma} + \eta_m + \sqrt{\frac{2^m}{n^{1+\gamma}}} \right\},$$

where $\operatorname{sign}(\cdot)$ *is a sign function.*

(ii) (*Rate of convergence*) *Assume that* (A1)–(A5) *and* (A7)(ii) *hold, then*

$$\sup_{t\in[0,1]} \left|\hat{g}(t) - g(t)\right| = O_p\left\{\sqrt{\frac{2^m}{n}} \log n + n^{-\gamma} + \eta_m\right\}.$$

Remark 3.2 Theorem 3.1(i) gives the Bahadur representation of the L_1-wavelet estimation for a nonparametric model. For $1/2 < \nu < 3/2$, η_m is a lower rate of convergence than the one of $\nu \geq 3/2$; Meanwhile, $g \in \mathcal{H}^\nu(\mathbb{R})$ is not differentiable if $1/2 < \nu < 3/2$. If we take $2^m = O(n^{1-\gamma})$ and $\nu = (\gamma + 1)/[2(1 - \gamma)]$ with $0 < \gamma < 1/2$, then $g \in \mathcal{H}^\nu(\mathbb{R})$ ($1/2 < \nu < 3/2$), and $R_n(m; \gamma, \nu) = O_p(n^{-\gamma})$ ($0 < \gamma < 1/2$), It implies that $R_n(m; \gamma, \nu)$ has an order very close to $O_p(n^{-1/2})$, which is comparable with the Bahadur order $o_p((nh)^{-1/2})$ for kernel weighted local polynomial estimation [29], where the bandwidth $h \to 0$ and the function requires second-order differentiability. For example, the triangular function having Fourier transform $\sin^2(\xi/2)/(\xi/2)^2$ belongs to $\mathcal{H}^1(\mathbb{R})$ and is Lipschitz of order 1, so it satisfies our conditions for g but is not differentiable. Such a function has not been studied before.

Remark 3.3 Theorem 3.1(ii) states the rate of convergence of L_1-wavelet estimation for a nonparametric model. As in Remark 3.2, we consider the lower rate case of η_m, i.e., $1/2 < \nu < 3/2$. If we take $2^m = O(n^\gamma)$ ($\gamma \geq 1/3$), then $\sup_{t\in[0,1]} |\hat{g}(t) - g(t)| = O_p(n^{-(1-\gamma)/2} \log n)$. Furthermore, taking $\gamma = 1/3$, one gets

$$\sup_{t\in[0,1]} \left|\hat{g}(t) - g(t)\right| = O_p\left(n^{-1/3} \log n\right),$$

which is comparable with the optimal convergence rate of the nonparametric estimation in nonparametric models. Meanwhile, it is the same as the results of [21] (in probability) and [22] (almost sure) for any $t \in [0, 1]$ based on a least-square wavelet estimator, but they require that g is continuously differentiable, that is, $g \in \mathcal{H}^\nu(\mathbb{R})$ ($\nu > 3/2$).

To obtain an asymptotic expansion of the variance and an asymptotic normality result, we need to consider an approximation to $\hat{g}(t)$ based on its values at dyadic points of order m. That is, we define $\hat{g}^d(t) = \hat{g}_n(t^{(m)})$ with $t^{(m)} = \lfloor 2^m t \rfloor / 2^m$, where $\lfloor z \rfloor$ denotes the maximum integer not greater than z.

Theorem 3.2 (Asymptotic normality) *Support that* (A1)–(A6) *and* (A7)(iii) *hold, then*

$$\sqrt{n2^{-m}}\left(\hat{g}^d(t) - g(t)\right) \xrightarrow{D} N\left(0, 4^{-1}f_\epsilon^{-2}(0)\omega_0^2\kappa(t)\right),$$

where $\omega_0^2 = \int_{\mathbb{R}} E_0^2(0, u)\, du = \sum_{k\in\mathbb{Z}} \phi^2(k)$.

Remark 3.4 $\hat{g}^d(t)$ is the piecewise-constant approximation of $\hat{g}(t)$ at resolution 2^{-m}. The reason to consider this is that the variance of $\hat{g}$ is unstable as a function of t, because $\text{var}(\hat{g}(t)) = 2^m n^{-1}\kappa(t) \int_0^1 E_0^2(t_m, s)\, ds$, where $t_m = 2^m t - [2^m t]$. We know that, if t is non-dyadic, then the sequence t_m wanders around the unit interval and fails to converge. Also see [16].

4 Technical proofs

In order to prove the main results, we first present several lemmas.

Lemma 4.1 *Suppose that* (A4) *holds. We have*:

(i) $E_0(t,s) \le c_k/(1 + |t - s|)^k$ *and* $E_k(t,s) \le 2^k c_k/(1 + 2^k|t - s|)^k$, *where k is a positive integer and c_k is a constant depending on k only*.

(ii) $\sup_{0 \le t,s \le 1} |E_m(t,s)| = O(2^m)$.

(iii) $\sup_{0 \le t \le 1} \int_0^1 |E_m(t,s)| \, ds \le c$, *where c is a positive constant*.

(iv) $\int_0^1 E_m(t,s) \, ds \to 1$ *uniformly in $t \in [0,1]$, as $m \to \infty$*.

The proofs of (i) and (ii) can be found in [16], and (iii) follows from (i); the proof of (iv) can be found in [30].

Lemma 4.2 *Suppose that* (A4)–(A5) *hold and $h(\cdot)$ satisfies* (A2)–(A3). *Then*

$$\sup_{0 \le t \le 1} \left| h(t) - \sum_{i=1}^n h(t_i) \int_{A_i} E_m(t,s) \, ds \right| = O(n^{-\gamma}) + O(\eta_m),$$

where

$$\eta_m = \begin{cases} (1/2^m)^{\nu - 1/2} & \text{if } 1/2 < \nu < 3/2, \\ \sqrt{m}/2^m & \text{if } \nu = 3/2, \\ 1/2^m & \text{if } \nu > 3/2. \end{cases}$$

It follows easily from Theorem 3.2 of [16].

Lemma 4.3 *Let $\{V_i, i = 1, \dots, n\}$ be a sequence of independent random variables with mean zero and finite $2 + \delta$th moments, and $\{a_{ij}, i, j = 1, \dots, n\}$ a set of positive numbers such that $\max_{ij} |a_{ij}| \le n^{-p_1}$ for some $0 \le p_1 \le 1$ and $\sum_{i=1}^n a_{ij} = O(n^{p_2})$ for some $p_2 \ge \max(0, 2/(2 + \delta) - p_1)$. Then*

$$\max_{1 \le j \le n} \left| \sum_{i=1}^n a_{ij} V_{ij} \right| = O(n^{-(p_1 - p_2)/2} \log n), \quad a.s.$$

It can be found in [31].

Lemma 4.4 *Let $\{\lambda_n(\theta), \theta \in \Theta\}$ be a sequence of random convex functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \to \lambda(\theta)$ in probability, for each θ in Θ. Then for each compact subset K of Θ, in probability,*

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \to 0.$$

See [32].

Below, we give the proof of the main results. The proof of Theorem 3.1 uses the idea of [32] and the convex lemma (Lemma 4.4). To complete the proof of Theorem 3.2, it is enough to check the Lindeberg-type condition.

Proof of Theorem 3.1 (i) From (2.1), note that $\hat{g}(t) = \hat{a}$ and $\hat{a}$ minimizes

$$\sum_{i=1}^{n} |y_i - a| \int_{A_i} E_m(t,s)\, ds.$$

Let $\theta = a - g(t)$ and $\epsilon_i^* = \epsilon_i + [g(t_i) - g(t)]$. Then θ minimizes the function

$$G_n(\theta) = \sum_{i=1}^{n} \left\{ \left| \epsilon_i^* - \theta \right| - \left| \epsilon_i^* \right| \right\} \int_{A_i} E_m(t,s)\, ds.$$

The idea behind the proof, as in [32], is to approximate $G_n(\theta)$ by a quadratic function whose minima have an explicit expression, and then to show that $\hat{\theta}$ is close enough to those minima that share their asymptotic behavior.

We now set out to approximate $G_n(\theta)$ by a quadratic function of θ. Write

$$G_n(\theta) = W_n \theta + R_n(\theta),$$

where $W_n = - \sum_{i=1}^{n} \operatorname{sign}(\epsilon_i) \int_{A_i} E_m(t,s)\, ds$, which does not depend on θ, and

$$R_n(\theta) = \sum_{i=1}^{n} \left\{ \left| \epsilon_i^* - \theta \right| - \left| \epsilon_i^* \right| + \operatorname{sign}(\epsilon_i)\theta \right\} \int_{A_i} E_m(t,s)\, ds. \tag{4.1}$$

We have

$$G_n(\theta) = E\big(G_n(\theta)\big) + W_n \theta + \big[R_n(\theta) - E\big(R_n(\theta)\big)\big]. \tag{4.2}$$

From the error assumption (A1), it is ensured that the function $\Delta(t) = E[|\epsilon_i - t| - |\epsilon_i|]$ has a unique minimum at zero, and $\Delta(t) = t^2 f_\epsilon(0) + o(t^2)$. Therefore, by Lemmas 4.1 and 4.2,

$$E\big(G_n(\theta)\big) = \sum_{i=1}^{n} \left\{ f_\epsilon(0)\theta^2 - 2f_\epsilon(0)\big[g(t_i) - g(t)\big]\theta \right\} \int_{A_i} E_m(t,s)\, ds + o\big(\delta_n^2\big)$$

$$= f_\epsilon(0)\theta^2 - 2f_\epsilon(0)\theta \left\{ \sum_{i=1}^{n} g(t_i) \int_{A_i} E_m(t,s)\, ds - g(t) \right\} + o\big(\delta_n^2\big)$$

$$= f_\epsilon(0)\theta^2 + O\big[\big(n^{-\gamma} + \eta_m\big)\theta\big] + o\big(\delta_n^2\big), \tag{4.3}$$

where $\delta_n = \max\{(n^{-\gamma} + \eta_m), |\theta|\}$. For (4.1), note that $||\epsilon_i^* - \theta| - |\epsilon_i^*| + \operatorname{sign}(\epsilon_i)\theta| \le 2|\theta| I\{|\varepsilon_i| \le |\theta| + |g(t_i) - g(t)|\}$, then we obtain

$$ER_n^2(\theta) \le 4\theta^2 \sum_{i=1}^{n} EI\big\{|\varepsilon_i| \le |\theta| + |g(t_i) - g(t)|\big\} \left\{ \int_{A_i} E_m(t,s)\, ds \right\}^2$$

$$= 8\theta^2 f_\epsilon(0) \sum_{i=1}^{n} |g(t_i) - g(t)| \left\{ \int_{A_i} E_m(t,s)\, ds \right\}^2 \big(1 + o(1)\big)$$

$$= O\left(\frac{2^m}{n^{1+\gamma}} \theta^2 \right).$$

We get

$$R_n(\theta) - E\big(R_n(\theta)\big) = O_p\left(\theta\sqrt{\frac{2^m}{n^{1+\gamma}}}\right). \tag{4.4}$$

Let $a_n = O_p\{n^{-\gamma} + \eta_m + \sqrt{\frac{2^m}{n^{1+\gamma}}}\}$. Combining (4.2)–(4.4), for each fixed θ, we have

$$\begin{aligned}
G_n(\theta) &= f_\epsilon(0)\theta^2 + W_n\theta + O\big[\big(n^{-\gamma} + \eta_m\big)\theta\big] + O_p\left(\theta\sqrt{\frac{2^m}{n^{1+\gamma}}}\right) \\
&= f_\epsilon(0)\theta^2 + (W_n + a_n)\theta,
\end{aligned} \tag{4.5}$$

with $a_n = o_p(1)$ uniformly. Note that

$$W_n = -\sum_{i=1}^{n} \text{sign}(\epsilon_i) \int_{A_i} E_m(t,s)\,ds.$$

It is easy to see that W_n has a bounded second moment and hence is stochastically bounded. Since the convex function $G_n(\theta) - (W_n + a_n)\theta$ converges in probability to the convex function $f_\epsilon(0)\theta^2$, it follows from the convexity lemma, Lemma 4.4, that, for every compact set K,

$$\sup_{\theta \in K}\big|G_n(\theta) - (W_n + a_n)\theta - f_\epsilon(0)\theta^2\big| = o_p(1). \tag{4.6}$$

Thus, the quadratic approximation to the convex function $G_n(\theta)$ holds uniformly for θ in any compact set. So, using the convexity assumption again, the minimizer $\hat{\theta}$ of $G_n(\theta)$ converges in probability to the minimizer

$$\bar{\theta} = -\frac{1}{2}f_\epsilon^{-1}(0)(W_n + a_n), \tag{4.7}$$

that is,

$$P\big(|\hat{\theta} - \bar{\theta}| > \delta\big) \to 0.$$

The assertion can be proved by some elementary arguments, which is similar to the proof of Theorem 1 in [32]. Based on (4.6), let $G_n(\theta) = (W_n + a_n)\theta + f_\epsilon(0)\theta^2 + r_n(\theta)$ which can be written as

$$G_n(\theta) = f_\epsilon(0)\big\{|\theta - \bar{\theta}|^2 - |\bar{\theta}|^2\big\} + r_n(\theta), \tag{4.8}$$

with $\sup_{\theta \in K}|r_n(\theta)| = o_p(1)$. Because $\bar{\theta}$ is stochastically bounded. The compact set K can be chosen to contain a closed ball $B(n)$ with center $\bar{\theta}$ and radius δ in probability, thereby implying that

$$\Delta_n = \sup_{\theta \in B(n)} \big|r_n(\theta)\big| = o_p(1).$$

Now consider the behavior of $G_n(\theta)$ outside $B(n)$. Suppose $\theta = \bar{\theta} + \beta\mu$, with $\beta > \delta$ and μ is a unit vector. Define θ^* as the boundary point B_n that lies on the line segment from $\bar{\theta}$ to θ, i.e. $\theta^* = \bar{\theta} + \delta\mu$. Convexity of $G_n(\theta)$, (4.8) and the definition of Δ_n imply

$$\frac{\delta}{\beta} G_n(\theta) + \left(1 - \frac{\delta}{\beta}\right) G_n(\bar{\theta}) \geq G_n(\theta^*)$$
$$\geq f_\epsilon(0)\delta^2 - f_\epsilon(0)|\bar{\theta}|^2 - \Delta_n$$
$$\geq f_\epsilon(0)\delta^2 + G_n(\bar{\theta}) - 2\Delta_n.$$

It follows that

$$\inf_{|\theta - \bar{\theta}| > \delta} G_n(\theta) \geq G_n(\bar{\theta}) + \frac{\beta}{\delta}\left[f_\epsilon(0)\delta^2 - 2\Delta_n\right],$$

when $2\Delta_n < f_\epsilon(0)\delta^2$, which happens with probability tending to one, the minimum of $G_n(\theta)$ cannot occur at any θ with $|\theta - \bar{\theta}| > \delta$. This implies that, for any $\delta > 0$ and for large enough n, the minimum of $G_n(\theta)$ must be achieved with $B(n)$, i.e., $|\hat{\theta} - \bar{\theta}| \leq \delta$ with probability tending to one. Thus, it completes the proof of (i).

(ii) In the following, we will prove that

$$W_n = O\left\{\left(\frac{2^m}{n}\right)^{1/2} \log n\right\}, \quad \text{a.s.} \tag{4.9}$$

By Lemma 4.1, we have

$$\max_{i,m}\left|\int_{A_i} E_m(t,s)\,ds\right| = O(2^m/n) = O(n^{-2p})$$

and

$$\sum_{i=1}^n \int_{A_i} E_m(t,s)\,ds = \int_0^1 E_m(t,s)\,ds = O(1) = O(n^{p_2}),$$

where $p_1 = 2p$ with $0 \leq p_1 \leq 1$ and $p_1 \geq 2/(2 + \delta)$, and $p_2 = 0$, which can be satisfied by Conditions (A1) and (A7)(ii). By Lemma 4.3, $W_n = O(n^{-p} \log n)$. Furthermore, we get (4.9). So, (ii) holds. □

Proof of Theorem 3.2 From Theorem 3.1(i), we have

$$2f_\epsilon(0)\{\hat{g}(t) - g(t)\} = Z_n(t) + R_n(m; \gamma, \nu), \tag{4.10}$$

where $Z_n(t) = \sum_{i=1}^n \text{sign}(\epsilon_i)\int_{A_i} E_m(t,s)\,ds$ and $R_n(m; \gamma, \nu) = O_p\{n^{-\gamma} + \eta_m + \sqrt{\frac{2^m}{n^{1+\gamma}}}\}$. From (A7)(iii) $n2^{-2m\nu^*} \to 0$, one gets

$$\sqrt{n2^{-m}}R_n(m; \gamma, \nu) = o_p(1). \tag{4.11}$$

Now, let us verify the asymptotic normality of $\sqrt{n2^{-m}}Z_n(t^{(m)})$. First, we calculate the variance of it. By the proofs of Theorem 3.3 and Lemma 6.1 of [16], we have

$$
\begin{aligned}
&\left|\operatorname{var}\left(\sqrt{n2^{-m}}Z_n(t^{(m)})\right) - \kappa(t)\omega_0^2\right| \\
&= \left|n2^{-m}\sum_{i=1}^{n}\left(\int_{A_i}E_m(t^{(m)},s)\,ds\right)^2 - \kappa(t)\omega_0^2\right| \\
&\leq \left|n2^{-m}\sum_{i=1}^{n}\left(\int_{A_i}E_m(t^{(m)},s)\,ds\right)^2 - 2^{-m}\int_0^1 E_m^2(t^{(m)},s)\kappa(s)\,ds\right| \\
&\quad + \left|2^{-m}\int_0^1 E_m^2(t^{(m)},s)\kappa(s)\,ds - \kappa(t)\omega_0^2\right| \\
&\leq n2^{-m}\left|\sum_{i=1}^{n}(s_i - s_{i-1})^2 E_m^2(t^{(m)},u_i) - \frac{1}{n}(s_i - s_{i-1})E_m^2(t^{(m)},v_i)k(v_i)\right| + o(1)
\end{aligned}
$$

(where u_i and v_i belong to A_i)

$$
\begin{aligned}
&= n2^{-m}O(n^{-1})O(n2^{-m})\left(\rho(n)2^{2m} + \frac{2^{2m}}{n^2} + \frac{2^{2m}}{n}\frac{2^m}{n}\right) + o(1) \\
&\leq O(n\rho(n) + 2^m/n) = o(1).
\end{aligned}
$$

So,

$$
\operatorname{var}\left(\sqrt{n2^{-m}}Z_n(t^{(m)})\right) = \kappa(t)\omega_0^2 + o(1). \tag{4.12}
$$

To complete the proof, we only need to check the Lindeberg-type condition

$$
\max_{1\leq i\leq n}\frac{n2^{-m}(\int_{A_i}E_m(t,s)\,ds)^2}{\operatorname{var}(\sqrt{n2^{-m}}Z_n(t^{(m)}))} \to 0.
$$

From (4.12) and Lemma 4.1, one sees that the order is $O(2^m/n) \to 0$. Thus, we complete Theorem 3.2. □

Availability of data and materials
No applicable.

Competing interests
The authors declare to have no competing interests.

Authors' contributions
The authors contributed equally in writing the final version of this article. All authors read and approved the final manuscript.

Author details
[1]Institute of Statistics and Data Science, Nanjing Audit University, Nanjing, Jiangsu 211815, China. [2]School of Economics and Management, Southeast University, Nanjing, Jiangsu 211189, China. [3]School of Mathematics and Finance, Chuzhou University, Chuzhou, Anhui 239001, China.

Publisher's Note

References

1. Müller, H.G.: Nonparametric Regression Analysis for Longitudinal Data. Springer, Berlin (1988)
2. Fan, J., Yao, Q.: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York (2003)
3. Gasser, T., Müller, H.G.: Kernel estimation of regression function. In: Smoothing Techniques for Curve Estiamtion. Springer, New York (1979)
4. Fan, J.: Design-dadaptive nonparametric regression. J. Am. Stat. Assoc. **87**, 998–1004 (1992)
5. Fan, J.: Local linear regression smoothers and their minimax efficiencies. Ann. Stat. **21**, 196–216 (1993)
6. Braun, W.J., Huang, L.S.: Kernel spline regression. Can. J. Stat. **33**, 259–278 (2005)
7. Härdle, W., Gasser, T.: Robust non-parametric function fitting. J. R. Stat. Soc., Ser. B, Stat. Methodol. **46**, 42–51 (1984)
8. Truong, Y.K.: Asymtotic properties of kernel estimators based on local medians. Ann. Stat. **17**, 606–617 (1989)
9. Wang, F.T., Scott, D.W.: The l_1 method for robust nonparametric regression. J. Am. Stat. Assoc. **89**, 65–76 (1994)
10. Fan, J., Hu, T.C., Truong, Y.K.: Robust non-parametric function estimation. Scand. J. Stat. **21**, 433–446 (1994)
11. Tang, Q., Wang, J.: L_1-Estimation for varying coefficient models. Statistics **39**, 389–404 (2005)
12. Honda, T.: Nonparametric estimation of conditional medians for linear and related processes. Ann. Inst. Stat. Math. **62**, 995–1021 (2010)
13. Zhao, Z.B., Wei, Y., Lin, D.K.J.: Asymptotics of nonparametric l_1 regression models with dependent data. Bernoulli **20**, 1532–1559 (2014)
14. Koenker, R.: Quantile Regression. Cambridge University Press, Cambridge (2005)
15. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. Wiley, New York (2009)
16. Antoniadis, A., Gregoire, G., McKeague, I.W.: Wavelet methods for curve estimation. J. Am. Stat. Assoc. **89**, 1340–1353 (1994)
17. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. J. Am. Stat. Assoc. **90**, 1200–1224 (1994)
18. Hall, P., Patil, P.: Formulae for mean integrated squared error of non-linear wavelet-based density estimators. Ann. Stat. **23**, 905–928 (1995)
19. Härdle, W., Kerkycharian, G., Picard, D., Tsybakov, A.: Wavelet, Approximation and Statistical Application. Springer, New York (1998)
20. Huang, S.Y.: Density estimation by wavelet-based reproducing kernels. Stat. Sin. **9**, 137–151 (1999)
21. Zhou, X., You, J.: Wavelet estimation in varying-coefficient partially linear regression models. Stat. Probab. Lett. **68**, 91–104 (2004)
22. Lu, Y., Li, Z.: Wavelet estimation in varying-coefficient models. Chinese J. Appl. Probab. **25**, 409–420 (2009)
23. Doosti, H., Afshari, M., Niroumand, H.A.: Wavelets for nonparametric stochastic regression with mixing stochastic process. Commun. Stat., Theory Methods **37**, 373–385 (2008)
24. Chesneau, C., Dewan, I., Doosti, H.: Nonparametric estimation of a quantile density function by wavelet methods. Comput. Stat. Data Anal. **94**, 161–174 (2016)
25. Li, L., Xiao, Y.: Wavelet-based estimation of regression function with strong mixing errors under fixed design. Commun. Stat., Theory Methods **46**, 4824–4842 (2017)
26. Zhou, X.C., Lin, J.G.: Asymptotic properties of wavelet estimators in semiparametric regression models under dependent errors. J. Multivar. Anal. **122**, 251–270 (2013)
27. Ding, L., Chen, P., Li, Y.: Berry–Esseen bound of wavelet estimator in heteroscedastic regression model with random errors. Int. J. Comput. Math. **96**, 821–852 (2019)
28. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
29. Welsh, A.H.: Robust estimation of smooth regression and spread functions and their derivatives. Stat. Sin. **6**, 347–366 (1996)
30. Walter, G.G.: Wavelets and Orthogonal Systems with Applications. CRC Press, Boca Raton (1994)
31. Härdle, W., Liang, H., Gao, J.: Partially Linear Models. Physica-Verlag, Heidelberg (2000)
32. Pollard, D.: Asymptotics for least absolute deviation regression estimators. Econom. Theory **7**, 186–199 (1991)