

Article

# Quantile-Wavelet Nonparametric Estimates for Time-Varying Coefficient Models

Xingcai Zhou \*, Guang Yang and Yu Xiang

School of Statistics and Data Science, Nanjing Audit University, Nanjing 211085, China; mg2008123@stu.nau.edu.cn (G.Y.); mg2008121@stu.nau.edu.cn (Y.X.)

\* Correspondence: xczhou@nau.edu.cn

**Abstract:** The paper considers quantile-wavelet estimation for time-varying coefficients by embedding a wavelet kernel into quantile regression. Our methodology is quite general in the sense that we do not require the unknown time-varying coefficients to be smooth curves of a common degree or the errors to be independently distributed. Quantile-wavelet estimation is robust to outliers or heavy-tailed data. The model is a dynamic time-varying model of nonlinear time series. A strong Bahadur order  $O\left\{\left(\frac{2^m}{n}\right)^{3/4}(\log n)^{1/2}\right\}$  for the estimation is obtained under mild conditions. As applications, the rate of uniform strong convergence and the asymptotic normality are derived.

**Keywords:** quantile-wavelet; nonparametric estimation; time-varying coefficient; Bahadur representation; strong mixing

**MSC:** 62G05; 62G08; 62G20



**Citation:** Zhou, X.; Yang, G.; Xiang, Y. Quantile-Wavelet Nonparametric Estimates for Time-Varying Coefficient Models. *Mathematics* **2022**, *10*, 2321. <https://doi.org/10.3390/math10132321>

Academic Editor: Stefano Bonnini

Received: 22 May 2022

Accepted: 29 June 2022

Published: 2 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The trending time-varying time series models have gained a lot of attention during the last three decades because they are increasingly used to observe changes in their dynamic structure, which have been applied to economics and finance [1,2]. When the time span of interest covers different economic periods, the parameters of the corresponding statistical models should be allowed to change with time; see references [3–7], among others. Varying coefficient models are flexible models for describing the dynamic structure of data, which have been extensively studied based on mean regression, for example, references [8–12], among others and the references therein. However, the important features of the joint distribution of the response and the covariates can not be well captured by mean regression.

In this paper, we consider the time-varying coefficient quantile regression models: the conditional  $\tau$ th quantile of  $Y_i$  given  $\mathbf{X}_i$  and a prespecified quantile  $\tau$ ,

$$Q_\tau(Y_i|\mathbf{X}_i) = \inf\{s : F_{Y_i|\mathbf{X}_i}(s|\mathbf{X}_i) > \tau\} = \mathbf{X}_i^T \boldsymbol{\beta}_\tau(t_i), i = 1, \dots, n, \quad (1)$$

where  $t_i = i/n$ ,  $Y_i$  are responses,  $\boldsymbol{\beta}_\tau(\cdot) = (\beta_{1,\tau}(\cdot), \dots, \beta_{p,\tau}(\cdot))^T$  is a  $p$ -dimensional vector of unspecified functions defined on  $[0, 1]$ ,  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$  are  $p$ -dimensional explanatory variables. We can write Equation (1) in the form

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta}_\tau(t_i) + \epsilon_{\tau,i}, i = 1, \dots, n, \quad (2)$$

where the errors  $\epsilon_{\tau,i}$  satisfy  $Q_\tau(\epsilon_{\tau,i}|\mathbf{X}_i) = 0$  almost surely. Quantile regression has also been emerged as an essential statistical tool in many scientific fields; see [13]. To simplify notations, we omit the subscript  $\tau$  from the function  $\boldsymbol{\beta}_\tau(\cdot)$  and  $\epsilon_{\tau,i}$  subsequently.

To estimate the varying coefficients in quantile regression models in Equation (2), references [7,14,15] used the kernel method, and references [16,17] considered regression

splines. To our knowledge, wavelet has not been considered in quantile regression with varying coefficients. Wavelet techniques can detect and represent localized features and can also create localized features on synthesis while being efficient in terms of computational speed and storage. It has received much attention from mathematicians, engineers and statisticians. Wavelet has been introduced for nonparametric regression, for example, references [18–22], and so on. For wavelet smoothing applied to nonparametric models, reference [18] is a key reference that introduces wavelet versions of some classical kernel and orthogonal series estimators and studies their asymptotic properties. Reference [23] provided asymptotic bias and variance of a wavelet estimator for a regression function under a stochastic mixing process. Reference [24] 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. References [25,26] showed Berry–Esseen-type bounds on wavelet estimators for semiparametric models under dependent errors. For varying coefficient models, references [27,28] provided wavelet estimation and studied convergence rate and asymptotic normality under i.i.d. errors, and reference [29] discussed these asymptotics of the wavelet estimators based on censored dependent data. However, all of the above wavelet estimators are based on mean regression by the least-squares method.

In the paper, we propose a quantile-wavelet method for time-varying coefficient models in Equation (2) with an  $\alpha$ -mixing errors stochastic process. The proposed methodology is quite general in the sense that we do not require coefficient  $\beta(\cdot)$  to be smooth curves of a common degree, it does not suffer from the curse of dimensionality, it is robust to outlier or heavy-tailed data, and it is a dynamic model for nonlinear time series. Bahadur representation, rate of convergence and asymptotic normality of quantile-wavelet estimators will be established. Bahadur representation theory seeks to approximate a statistical estimate by a sum of variables with a higher-order remained. It has been a major statistical theory issue since Bahadur’s pioneering work on quantiles [30]; see [31], among others. Recently, reference [32] investigated the Bahadur representation for sample quantiles under  $\varphi$ -mixing sequence; reference [33] gave an  $M$ -estimation for time-varying coefficient models with  $\alpha$ -mixing errors and established Bahadur representation in probability. In the paper, we will establish Bahadur representation with probability 1 (almost surely) for quantile-wavelet estimates in the models Equations (1) and (2).

Throughout, we assume that  $\{\mathbf{X}_i, \epsilon_i\}$  is a stationary  $\alpha$ -mixing sequence. Recall that a sequence  $\{\zeta_k, k \geq 1\}$  is said to be  $\alpha$ -mixing (or strong mixing) if the mixing coefficients

$$\alpha(m) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^k, B \in \mathcal{F}_{k+m}^\infty\}$$

converge to zero as  $m \rightarrow \infty$ , where  $\mathcal{F}_l^k$  denotes the  $\sigma$ -field generated by  $\{\zeta_i, l \leq i \leq k\}$ . We refer to the monograph of [34,35] for some properties or more mixing conditions.

The rest of this article is organized as follows. In Section 2, we present wavelet kernel and quantile-wavelet estimation for the model (2). Bahadur representation of quantile-wavelet estimators and their applications are given in Section 3. Technical proofs are provided in Section 4. Some simulation studies are conducted in Section 5.

## 2. Quantile-Wavelet Estimation

In the paper, the time-varying coefficient function is retrieved by a wavelet-based reproducing kernel Hilbert space (RKHS). An RKHS is a Hilbert space in which all the point evaluations are bounded linear functions. Let  $\mathcal{H}$  be a Hilbert space of functions on some domain  $\mathcal{I}$ . For  $t \in \mathcal{I}$ , then there exists an element  $k_t \in \mathcal{H}$ , such that

$$f(t) = \langle k_t, f \rangle, \forall f \in \mathcal{H},$$

where  $\langle \cdot, \cdot \rangle$  is the inner product in  $\mathcal{H}$ . Set  $\langle k_t, k_s \rangle = K(t, s)$ , which is called the reproducing kernel. Let  $k_t = K(t, \cdot)$ , and then  $\langle K(t, \cdot), K(s, \cdot) \rangle = K(t, s)$ .

A multiresolution analysis is a sequence of closed subspaces  $\{V_m, m \in \mathbb{Z}\}$  in  $\mathcal{L}_2(\mathbb{R})$  such that they lie in a containment hierarchy,

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots, \tag{3}$$

where  $\mathcal{L}_2(\mathbb{R})$  is the collection of square-integrable functions over the real line. The hierarchy Equation (3) is constructed such that (i)  $V$ -spaces are self-similar,  $f(2^m x) \in V_m$  iff  $f(x) \in V_0$ , and (ii) there exists a scaling function  $\phi \in V_0$  whose integer-translate  $V_0 = \{f \in \mathcal{L}_2(\mathbb{R}) | f(x) = \sum_{k \in \mathbb{Z}} c_k \phi(x - k)\}$ , and for which the set  $\{\phi(\cdot - k), k \in \mathbb{Z}\}$  is an orthonormal basis. Wavelet analysis requires a description of two related and suitably chosen orthonormal basic functions: the scaling function  $\phi$  and the wavelet  $\psi$ . A wavelet system is generated by dilation and translation of  $\phi$  and  $\psi$  through

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

Therefore,  $\{\phi_{0k}, k \in \mathbb{Z}\}$  and  $\{\psi_{mk}, k \in \mathbb{Z}\}$  are the orthogonal bases of  $V_0$  and  $V_m$ , respectively. From Moore-Aronszajn’s theorem [36], it follows that

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

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

$$K_m(t, s) = 2^m K(2^m t, 2^m s)$$

is a reproducing kernel of  $V_m$ , and then the projection of  $g$  on the space  $V_m$  is given by

$$P_{V_m} g(t) = \int K_m(t, s) g(s) ds.$$

It motivates us to define a quantile-wavelet estimator of  $\beta(t)$  by

$$\hat{\beta}(t) = \operatorname{argmin}_{\mathbf{b}} \sum_{i=1}^n \rho_{\tau} \left\{ y_i - \mathbf{X}_i^T \mathbf{b} \right\} \int_{A_i} K_m(t, s) ds, \tag{4}$$

where  $\rho_{\tau}(u) = u(\tau - I\{u < 0\})$  with  $u \in \mathbb{R}$  called the loss (“check”) function,  $I_B$  is the indicator function of any set  $B$ , and  $A_i$  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  $s_i = (t_i + t_{i+1})/2, i = 1, \dots, n - 1$ .

Note that many other nonparametric methods can be used here, including spline and Kernel approaches. However, they might not be rich enough to characterize the local properties of the time-varying coefficient function. In the following section, we will present the asymptotic properties of the quantile-wavelet estimator Equation (4).

### 3. Bahadur Representation and Its Applications

Let  $\mathcal{H}^{\nu}$  be the collection of all functions on  $[0,1]$  with order  $\nu > 0$  in Sobolev space, which is a very general space. The degree of smoothness of the true coefficient functions determines how well the functions can be approximated. Functions belonging to  $\mathcal{H}^{\nu}$  for  $1/2 < \nu < 3/2$  are not continuously differentiable. It is worth stressing that the wavelet approach allows us to obtain rates under much weaker assumptions than second-order differentiability. Denote  $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)$ ;  $\|\cdot\|$  is the  $L_2$  norm, and  $C$  is used to denote positive constants whose values are unimportant and may change from line to line in the proof.

Our main results will be established under the following assumptions.

(A1.) (i)  $\{\epsilon_i, \mathbf{X}_i\}$  is a stationary  $\alpha$ -mixing sequence with  $\alpha(i) = O(i^{-\kappa_0})$  with  $\kappa_0 > \frac{9(1+d)\delta+6}{2(1-d)\delta-4}$ ,  $\delta > \frac{2}{1-d}$ , where  $d$  is defined in (A7)(i); (ii) the noisy errors  $\epsilon_i$  has  $Q_{\tau}(\epsilon_i | \mathbf{X}_i) = 0$  almost surely, and a continuous, positive conditional density  $f_{\epsilon | \mathbf{X}}$  in a neighborhood of

0 given the  $\mathbf{X}_i$ ,  $E\|\mathbf{X}_1\|^{2\delta} < \infty$ , and  $E|\epsilon_1 | \mathbf{X}_1|^{2\delta} < \infty$ , a.s.; (iii)  $\Phi_x = E(\mathbf{X}_1\mathbf{X}_1^T)$ ,  $\Omega_x = E(f_{\epsilon|\mathbf{X}}(0)\mathbf{X}_1\mathbf{X}_1^T)$  are non-singular matrices.

(A2.)  $\beta_j$  belongs to Sobolev space  $\mathcal{H}^\nu([0, 1])$  with order  $\nu > 1/2$ .

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

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

(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.) The tuning parameter  $m$  satisfies (i)  $2^m = O(n^d)$  with  $0 < d < 1$ ; (ii) let  $v^* = \min(3/2, \nu, \gamma + 1/2) - \epsilon_1$  and  $\epsilon_1 = 0$  for  $\nu \neq 3/2$ ,  $\epsilon_1 > 0$  for  $\nu = 3/2$ . Assume that  $n2^{-2mv^*} \rightarrow 0$ .

**Remark 1.** *These conditions are mild. Condition (A1) is the standard requirement for moments and the mixing coefficient for an  $\alpha$ -mixing time series. Conditions (A2)–(A6) are the mild regularity conditions for wavelet smoothing, which have been adopted by [18]. In condition (A7),  $m$  acts as a tuning parameter, such as the bandwidth does for standard kernel smoothers; (A7) (i) is for Bahadur representation and rate of convergence, and (A7) (ii) combining with (A7) (i) is for asymptotic normality of the quantile-wavelet estimator.*

Our results are as follows.

**Theorem 1.** *(Bahadur representation) Support that (A1)–(A5) and (A7) (i) hold, then*

$$\hat{\beta}(t) - \beta(t) = \Omega_x^{-1}\mathbf{Z}_n(t) + R_n(m; \gamma, \nu), \text{ a.s.}$$

with

$$\mathbf{Z}_n(t) = \sum_{i=1}^n \varphi_\tau(\epsilon_i + \mathbf{X}_i^T[\beta(t_i) - \beta(t)])\mathbf{X}_i \int_{A_i} K_m(t, s)ds$$

and

$$R_n(m; \gamma, \nu) = O\left\{ \left(\frac{2^m}{n}\right)^{3/4} (\log n)^{1/2} \right\}.$$

**Remark 2.** *Theorem 1 presents the strong Bahadur representation of a quantile-wavelet estimator for a time-varying coefficient model. Here,  $R_n(m; \gamma, \nu) = O\left\{ \left(\frac{2^m}{n}\right)^{3/4} (\log n)^{1/2} \right\}$ , a.s., which is comparable with the Bahadur order  $O\left\{ \left(\frac{\log \log n}{nh}\right)^{3/4} \right\}$ , a.s., of [37], where the bandwidth  $h \rightarrow 0$ . Reference [37] is based on kernel local polynomial M-estimation, and requires that the function  $\beta$  has the second-order differentiability. However, we do not need the strong, smooth conditions. The function  $\beta$  is not differentiable when  $\beta \in \mathcal{H}^\nu$ ,  $1/2 < \nu < 3/2$ .*

The Bahadur representation of the quantile-wavelet estimator of Theorem 1 can be applied to obtain the following two results.

**Corollary 1.** *(Rate of uniform strong convergence) Assume that (A1)–(A5) and (A7) (i) hold, then*

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

**Remark 3.** Corollary 1 provides the rate of uniform strong convergence of quantile-wavelet estimator  $\hat{\beta}$  for model Equation (2). We consider the rate in the case of  $1/2 < \nu < 3/2$ , under which  $\eta_m$  is a lower rate of convergence than one of  $\nu \geq 3/2$ . If we take  $2^m = O(n^\gamma)$  with  $1/3 \leq \gamma < 1$ , then  $\sup_{t \in [0,1]} \|\hat{\beta}(t) - \beta(t)\| = O\left(n^{-(1-\gamma)/2}(\log n)^{1/2}\right)$ , a.s.. If we further take  $\gamma = 1/3$ , then one obtains

$$\sup_{t \in [0,1]} |\hat{\beta}(t) - \beta(t)| = O\left(n^{-1/3} \log n\right), \text{ a.s.,}$$

which is comparable with the optimal convergence rate of the nonparametric estimation in nonparametric models. The result is better than the ones of [27,28] based on the local linear estimator for the varying-coefficient model. In addition, we also do not require the unknown coefficient  $\beta$  to be smooth curves of a common degree.

**Corollary 2.** (Asymptotic normality) Support that (A1)–(A7) holds, then

$$\sqrt{n2^{-m}}(\hat{\beta}^d(t) - \beta(t)) \xrightarrow{D} N\left(0, \tau(1 - \tau)\kappa(t)\omega_0^2\Omega_x^{-1}\Psi_x\Omega_x^{-1}\right),$$

where  $\hat{\beta}^d(t) = \hat{\beta}(t^{(m)})$  with  $t^{(m)} = \lfloor 2^m t \rfloor / 2^m$  and  $\omega_0^2 = \int_{\mathbb{R}} E_0^2(0, u) du = \sum_{k \in \mathbb{Z}} \phi^2(k)$ .

**Remark 4.** To obtain an asymptotic expansion of the variance and an asymptotic normality, we need to consider an approximation to  $\hat{\beta}(t)$  based on its values at dyadic points of order  $m$ , as reference [18] has done. The  $\hat{\beta}^d(t)$  is the piecewise-constant approximation of  $\hat{\beta}(t)$  at resolution  $2^{-m}$ , which can avoid the instability of the variance of  $\hat{\beta}(t)$ . From the proof of Corollary 2, it can see that the main term of the variance of  $\hat{\beta}(t)$  is  $\tau(1 - \tau)\kappa(t)\omega^2(t_m)2^m n^{-1}\Omega_x^{-1}\Psi_x\Omega_x^{-1}$  with  $t_m = 2^m t - \lfloor 2^m t \rfloor$  and  $\omega^2(t_m) = \int_0^1 E_0^2(t_m, s) ds$ . When the dyadic  $t$  and  $m$  sufficiently large,  $t_m = 0$ , the variance of  $\hat{\beta}(t)$  is asymptotically stable. See [18] for the details.

**4. Lemmas and Proofs**

**Lemma 1** ([18,38]). Suppose that (A4) holds. We have

- (i)  $K_0(t, s) \leq c_k / (1 + |t - s|)^k$  and  $K_k(t, s) \leq 2^k c_k / (1 + 2^k |t - s|)^k$ , where  $k$  is a positive integer and  $c_k$  is a constant depending on  $k$  only.
- (ii)  $\sup_{0 \leq t, s \leq 1} |K_m(t, s)| = O(2^m)$ .
- (iii)  $\sup_{0 \leq t \leq 1} \int_0^1 |K_m(t, s)| ds \leq c$ , where  $c$  is a positive constant.
- (iv)  $\int_0^1 K_m(t, s) ds \rightarrow 1$  uniformly in  $t \in [0, 1]$ , as  $m \rightarrow \infty$ .

**Lemma 2** ([18]). Suppose that (A4)–(A5) hold and  $h(\cdot)$  satisfies (A2)–(A3). Then

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

where

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

**Lemma 3** ([39]). Let  $\{\lambda_n(\theta), \theta \in \Theta\}$  be a sequence of random convex functions defined on a convex, open subset  $\Theta$  of  $\mathbb{R}^d$ . Suppose  $\lambda(\cdot)$  is a real-valued function on  $\Theta$  for which  $\lambda_n(\theta) \rightarrow \lambda(\theta)$  with probability 1, for each fixed  $\theta$  in  $\Theta$ . Then for each compact subset  $K$  of  $\Theta$ , with probability 1,

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

**Lemma 4.** Let  $\{(X_i, e_i), 1 \leq i \leq n\}$  be a stationary sequence satisfying the mixing condition  $\alpha(\ell) = O(\ell^{-\kappa_0})$  for some  $\kappa_0 > \frac{9(1+d)\delta+6}{2(1-d)\delta-4}$ ,  $\delta > \frac{2}{1-d}$ ; and  $2^m = O(n^d)$  with  $0 < d < 1$ . Further, assume that  $E\{|e_1 X_1|^\delta\} < \infty$ . If Conditions (A4) and (A5) hold. Then

$$\sup_{s \in [0,1]} \left| \sum_{i=1}^n \{e_i X_i - E(e_i X_i)\} \int_{A_i} K_m(t, s) ds \right| = O\left(\sqrt{\frac{2^m \log n}{n}}\right), a.s..$$

**Remark 5.** In Lemma 4, we assume that  $\{X_i, 1 \leq i \leq n\}$  is a sequence of a 1-dimensional random variable. In fact,  $\mathbf{X}_i \in \mathbb{R}^p$  for the fixed  $p$ , we also have the same result as Lemma 4.

**Proof.** The theorem is similar to Lemma A.4 in [29] but has some differences. We suppose  $E(e_i X_i) = 0$ . If  $E(e_i X_i) \neq 0$ , the method of the proof is the same. Let  $Q_m(t) = \sum_{i=1}^n e_i X_i \int_{A_i} K_m(t, s) ds$ . Partition the interval  $[0, 1]$  into  $N = \lfloor (n2^{3m})^{1/2} \rfloor$  subintervals  $I_j$  of equal length. Let  $t_j$  be the centers of  $I_j$ . Notice that

$$|Q_m(t) - Q_m(t')| \leq C2^{2m} |t - t'| \frac{1}{n} \sum_{i=1}^n |e_i X_i| \leq C2^{2m} |t - t'| E|eX|, a.s.. \tag{5}$$

One obtains

$$\begin{aligned} \sup_{t \in [0,1]} |Q_m(t)| &\leq \max_{1 \leq j \leq N} \sup_{t \in I_j} |Q_m(t) - Q_m(t_j)| + \max_{1 \leq j \leq N} |Q_m(t_j)| \\ &\leq \max_{1 \leq j \leq N} |Q_m(t_j)| + C\sqrt{2^m/n}. \end{aligned} \tag{6}$$

Let  $Q_m^B(s) = \sum_{i=1}^n e_i X_i I(|e_i X_i| \leq B_n) \int_{A_i} K_m(t, s) ds$ , and take  $B_n = n^{\delta-1+\epsilon}$  for some  $\epsilon > 0$ . Note that  $\sum_i P(|e_i X_i| > B_i) < \sum_i B_i^{-\delta} E|e_1 X_1|^\delta < \infty$ . By the Borel–Cantelli lemma,  $|e_i X_i| \leq B_i$ , a.s., for sufficiently large  $i$ . Hence,

$$|e_i X_i| \leq B_n, \text{ a.s., for all } i \leq n, \tag{7}$$

for all sufficiently large  $n$ . In addition,

$$\begin{aligned} \sup_{t \in [0,1]} |E[Q_m(t) - Q_m^B(t)]| &= \sup_{t \in [0,1]} \left| \sum_{i=1}^n E(e_i X_i I[|e_i X_i| > B_n]) \int_{A_i} K_m(t, s) ds \right| \\ &\leq CB_n^{1-\delta} E|e_1 X_1|^\delta \leq CB_n^{1-\delta}. \end{aligned} \tag{8}$$

From Equations (7) and (8), we respectively have

$$\sup_{t \in [0,1]} |Q_m(t) - Q_m^B(t) - E[Q_m(t) - Q_m^B(t)]| = O(B_n^{1-\delta}) = o(n^{-1/2}), a.s. \tag{9}$$

Further, we have

$$\sup_{t \in [0,1]} |Q_m(t)| \leq \max_{1 \leq j \leq N} |Q_m^B(t_j) - E(Q_m^B(t_j))| + C\sqrt{2^m/n}. \tag{10}$$

Let  $\tilde{X}_i = n(e_i X_i I(|e_i X_i| \leq B_n) \int_{A_i} -E[e_i X_i I(|e_i X_i| \leq B_n) \int_{A_i} K_m(t, s) ds])$ . Note that  $|\tilde{X}_i| \leq C2^m B_n$ . By Theorem 2.18 (ii) in Fan and Yao (2003) [35], take  $h_n = (M2^m \log n/n)^{1/2}$ , for any  $\eta > 0$  and sufficiently large  $M$ , for each  $q = [1, n/2]$ , we have

$$\begin{aligned} &\sum_{n=1}^\infty P(\max_{1 \leq j \leq N} |Q_m^B(t_j) - E(Q_m^B(t_j))| > \eta h_n) \\ &\leq C \sum_{n=1}^\infty N \left( \exp\left(-\frac{h_n^2 q}{v^2(q)}\right) + \left(1 + \frac{2^m B_n}{h_n}\right)^{1/2} q \alpha(k) \right), \end{aligned}$$

where  $k = \lfloor n / (2q) \rfloor$ ,  $v^2(q) = 2\sigma^2(q) / k^2 + C2^m B_n h_n$ ,

$$\sigma^2(q) = \max_{0 \leq j \leq 2q-1} \text{Var} \left\{ \tilde{X}_{jk+1} + \dots + \tilde{X}_{(j+1)k+1} \right\}.$$

Taking  $k = (B_n h_n)^{-1}$ , we obtain  $\sigma^2(q) = O(2^m k)$ . Assume  $2^m = O(n^d)$ ,  $0 < d < 1$ . We have

$$\begin{aligned} & \sum_{n=1}^{\infty} P \left( \max_{1 \leq j \leq N} |Q_m^B(t_j) - E(Q_m^B(t_j))| > \eta h_n \right) \\ & \leq C \sum_{n=1}^{\infty} N \left( \exp \left( -\frac{C n h_n^2}{2^m} \right) + C \sqrt{2^m n} B_n^{\kappa_0+1.5} h_n^{\kappa_0+0.5} \right) \\ & = C \left( \sum_{n=1}^{\infty} n^{-CM+2} \right) + C \left( \sum_{n=1}^{\infty} n^{\frac{5}{4} + (\delta-1+\epsilon)(\kappa_0+1.5) - \frac{\kappa_0}{2}} (2^m)^{\frac{(9+2\kappa_0)}{4}} (\log n)^{\frac{\kappa_0+0.5}{2}} \right) \\ & \leq C \left( \sum_{n=1}^{\infty} n^{-CM+2} \right) + C \left( \sum_{n=1}^{\infty} n^{\frac{5}{4} + \delta^{-1}(\kappa_0+1.5) - \frac{\kappa_0}{2}} (2^m)^{\frac{(9+2\kappa_0)}{4}} n^{2\epsilon(\kappa_0+1.5)} \right) \\ & \leq C \left( \sum_{n=1}^{\infty} n^{-CM+2} \right) + C \left( \sum_{n=1}^{\infty} n^{(\frac{1}{\delta} - \frac{1}{2} + \frac{d}{2})\kappa_0 + \frac{5}{4} + \frac{9d}{4} + \frac{3\delta}{2}} n^{2\epsilon(\kappa_0+1.5)} \right) < \infty, \end{aligned}$$

since we have  $(\frac{1}{\delta} - \frac{1}{2} + \frac{d}{2})\kappa_0 + \frac{5}{4} + \frac{9d}{4} + \frac{3\delta}{2} + \epsilon(\kappa_0 + 1.5) < -1$  when  $\kappa_0 > \frac{9(1+d)\delta+6}{2(1-d)\delta-4}$  and  $\epsilon$  are small enough with  $\delta > 2/(1-d)$ . By the Borel–Cantelli Lemma and Equation (10), we prove Lemma 4.  $\square$

**Lemma 5.** Under conditions in Theorem 1, we have

$$\sup_{t \in [0,1]} \left\| \sum_{i=1}^n \{ \varphi_{\tau}(\epsilon_i^*) - \varphi_{\tau}(\epsilon_i) \} \mathbf{X}_i \int_{A_i} K_m(t,s) ds \right\| = O \left( n^{-\gamma} + \eta_m + \sqrt{\frac{2^m \log n}{n}} \right), \text{ a.s.,}$$

where  $\epsilon_i^* = \epsilon_i + \mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]$ .

**Proof.** Let  $e_i = I[\epsilon_i \leq -\mathbf{X}_i^T (\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t))] - I[\epsilon_i \leq 0]$ . By Lemmas 2 and 4, we have

$$\begin{aligned} & \sup_{t \in [0,1]} \left\| \sum_{i=1}^n \{ \varphi_{\tau}(\epsilon_i^*) - \varphi_{\tau}(\epsilon_i) \} \mathbf{X}_i \int_{A_i} K_m(t,s) ds \right\| = \sup_{t \in [0,1]} \left\| \sum_{i=1}^n e_i \mathbf{X}_i \int_{A_i} K_m(t,s) ds \right\| \\ & \leq \sup_{t \in [0,1]} \left\| \sum_{i=1}^n E(e_i \mathbf{X}_i) \int_{A_i} K_m(t,s) ds \right\| + \sup_{t \in [0,1]} \left\| \sum_{i=1}^n [e_i \mathbf{X}_i - E(e_i \mathbf{X}_i)] \int_{A_i} K_m(t,s) ds \right\| \\ & \leq \sup_{t \in [0,1]} \left\| \sum_{i=1}^n E \left[ \left\{ F_{\epsilon|\mathbf{X}}(-\mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]) - F_{\epsilon|\mathbf{X}}(0) \right\} \mathbf{X}_i \right] \int_{A_i} K_m(t,s) ds \right\| + O \left( \sqrt{\frac{2^m \log n}{n}} \right) \tag{11} \\ & = \sup_{t \in [0,1]} \left\| \sum_{i=1}^n E \left\{ f_{\epsilon|\mathbf{X}}(0) \mathbf{X}_i \mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)] \right\} \int_{A_i} K_m(t,s) ds \right\| + O \left( \sqrt{\frac{2^m \log n}{n}} \right) \\ & = \sup_{t \in [0,1]} \left\| \boldsymbol{\Omega}_x \left\{ \boldsymbol{\beta}(t) - \sum_{i=1}^n \boldsymbol{\beta}(t_i) \int_{A_i} K_m(t,s) ds \right\} \right\| + O \left( \sqrt{\frac{2^m \log n}{n}} \right) \\ & = O \left( n^{-\gamma} + \eta_m + \sqrt{\frac{2^m \log n}{n}} \right), \text{ a.s..} \end{aligned}$$

This completes the proof of Lemma 5.  $\square$

**Lemma 6.** Under conditions in Theorem 1, for fixed  $\theta$ , we have

$$\sup_{t \in [0,1]} \left\| E(R_n(\theta, t)) - \frac{1}{2} \theta^T \Omega_x \theta \right\| = O\left(\sqrt{\frac{2^m \log n}{n}}\right) \tag{12}$$

and

$$\sup_{t \in [0,1]} \|R_n(\theta, t) - E(R_n(\theta, t))\| = O\left(\sqrt{\frac{2^m \log n}{n}}\right), \text{ a.s.,} \tag{13}$$

where  $R_n(\theta, t)$  is defined by (18) in the proof of Theorem 1.

**Proof.** For Equation (12), by Condition (A1) (iii) and Lemma 4, one obtains

$$\begin{aligned} & E(R_n(\theta, t) | \mathbf{X}) \\ &= n2^{-m} \sum_{i=1}^n \int_0^{v_i} F_{\epsilon | \mathbf{X}}(s - \mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]) - F_{\epsilon | \mathbf{X}}(-\mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]) ds \int_{A_i} K_m(t, s) ds \\ &= n2^{-m} \sum_{i=1}^n \int_0^{v_i} f_{\epsilon | \mathbf{X}}(0) ds \int_{A_i} K_m(t, s) ds (1 + o(1)) \\ &= n2^{-m} \frac{1}{2} \theta^T \left\{ \frac{2^m}{n} \sum_{i=1}^n f_{\epsilon | \mathbf{X}}(0) \mathbf{X}_i \mathbf{X}_i^T \int_{A_i} K_m(t, s) ds \right\} \theta (1 + o(1)) \\ &= \frac{1}{2} \theta^T \Omega_x \theta + o(\|\theta\|^2) + O\left(\sqrt{\frac{2^m \log n}{n}}\right), \text{ a.s..} \end{aligned} \tag{14}$$

Note that  $E(R_n(\theta, t)) = E[E(R_n(\theta, t) | \mathbf{X})]$ . Thus, Equation (12) holds. Now, let us prove Equation (13). Notice that, for any  $k > 0$ ,

$$|I(\epsilon_i^* \leq s) - I(\epsilon_i^* \leq 0)|^k = I(d_1 \leq \epsilon_i \leq d_2), \tag{15}$$

where  $d_1 = \min(c_1, s + c_1)$ , and  $d_2 = \max(c_1, s + c_1)$  with  $c_1 = -\mathbf{X}_i^T [\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]$ . Let  $\tilde{v}_i = n2^{-m} \int_0^{v_i} \{I(\epsilon_i^* \leq s) - I(\epsilon_i^* \leq 0)\} ds$ . To prove Equation (13), we only need to show that  $E(|\tilde{v}_i|^\delta) < \infty$  by the proof of Lemma 4. By Equation (15) and Jensen’s inequality, one obtains

$$\begin{aligned} E(|\tilde{v}_i|^\delta) &= E\{E(|\tilde{v}_i|^\delta | \mathbf{X})\} \\ &\leq E\left\{ \left( \left[ n2^{-m} \int_0^{v_i} E|I(\epsilon_i^* \leq s) - I(\epsilon_i^* \leq 0)| ds \right]^\delta | \mathbf{X} \right) \right\} \\ &= E\left\{ \left( \left[ n2^{-m} \int_0^{v_i} |F_{\epsilon | \mathbf{X}}(d_2) - F_{\epsilon | \mathbf{X}}(d_1)| ds \right]^\delta | \mathbf{X} \right) \right\} \\ &= E\left\{ \left( \left[ n2^{-m} \int_0^{v_i} f_{\epsilon | \mathbf{X}}(0) ds \right]^\delta | \mathbf{X} \right) \right\} (1 + o(1)) \\ &= E(\mathbf{X}_1^T \theta)^{2\delta} (1 + o(1)) < \infty. \end{aligned}$$

By Lemma 4, we obtain Equation (13).

In the sequence, we will give the proofs of the main results.  $\square$

**Proof of Theorem 1.** Recall that  $\hat{\boldsymbol{\beta}}_n(t) = \hat{\mathbf{b}}$  and  $\hat{\mathbf{b}}$  minimize

$$\sum_{i=1}^n \rho_\tau \{y_i - \mathbf{X}_i^T \mathbf{b}\} \int_{A_i} K_m(t, s) ds.$$

Let  $\hat{\theta} = \sqrt{n2^{-m}}(\hat{\mathbf{b}} - \boldsymbol{\beta}(t))$  and  $\epsilon_i^* = \epsilon_i + \mathbf{X}_i^T[\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t)]$ . The behavior of  $\hat{\theta}$  follows from consideration of the objective function

$$G_n(\boldsymbol{\theta}; t) = n2^{-m} \sum_{i=1}^n \left\{ \rho_\tau(\epsilon_i^* - \sqrt{2^m/n} \mathbf{X}_i^T \boldsymbol{\theta}) - \rho_\tau(\epsilon_i^*) \right\} \int_{A_i} K_m(t, s) ds.$$

The function  $G_n(\boldsymbol{\theta}; t)$  is obviously convex and is minimized at  $\hat{\theta}$ . It is sufficient to show that  $G_n(\boldsymbol{\theta}; t)$  converges to its expectation since it follows from Lemma 3 that the convergence is also uniform on any compact set  $K$  of  $\boldsymbol{\theta}$ . Using the identity of [40],

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\varphi_\tau(u) + \int_0^v \{I(u \leq s) - I(u \leq 0)\} ds,$$

where  $\varphi_\tau(u) = \tau - I\{u < 0\}$ ; we may write

$$G_n(\boldsymbol{\theta}; t) = \boldsymbol{\theta}^T \mathbf{W}_n(t) + E(R_n(\boldsymbol{\theta}, t)) + (R_n(\boldsymbol{\theta}, t) - E[R_n(\boldsymbol{\theta}, t)]), \tag{16}$$

where

$$\mathbf{W}_n(t) = -\sqrt{n/2^m} \sum_{i=1}^n \varphi_\tau(\epsilon_i^*) \mathbf{X}_i \int_{A_i} K_m(t, s) ds, \tag{17}$$

$$R_n(\boldsymbol{\theta}, t) = n2^{-m} \sum_{i=1}^n \int_0^{v_i} \{I(\epsilon_i^* \leq s) - I(\epsilon_i^* \leq 0)\} ds \int_{A_i} K_m(t, s) ds, \tag{18}$$

with  $v_i = \sqrt{2^m/n} \mathbf{X}_i^T \boldsymbol{\theta}$ .

To obtain strong Bahadur representation of the quantile-wavelet estimator, we first need to show the uniform approximation of  $G_n(\boldsymbol{\theta}; t)$  on  $t \in [0, 1]$  with probability 1 for the fixed  $\boldsymbol{\theta}$  by the three terms in Equation (16). By Lemma 6, we have

$$G_n(\boldsymbol{\theta}; t) = \boldsymbol{\theta}^T \mathbf{W}_n(t) + \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Omega}_x \boldsymbol{\theta} + O\left(\sqrt{\frac{2^m \log n}{n}}\right) \text{ a.s.} \tag{19}$$

uniformly on  $t \in [0, 1]$ . Let  $a_n^2 = \sqrt{\frac{2^m}{n}} \log n$ . We have

$$\sup_{t \in [0,1]} a_n^{-2} \left| G_n(\boldsymbol{\theta}; t) - \boldsymbol{\theta}^T \mathbf{W}_n(t) - \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Omega}_x \boldsymbol{\theta} \right| = O\left(\frac{1}{\sqrt{\log n}}\right) = o(1), \text{ a.s.} \tag{20}$$

Second, the strong Bahadur representation requires the existence of one compact subset  $K$  with probability measure 1 that will suffice for all  $\boldsymbol{\theta}$ . We prove it by applying a stronger convexity lemma (See Lemma 3) than one of [41]. However, the arguments to prove it are essentially the same as in [41].

Let  $\bar{\theta} = -\boldsymbol{\Omega}_x^{-1} \mathbf{W}_n(t)$ . It is easy to see that  $\mathbf{W}_n(t)$  has a bounded second moment and hence is stochastically bounded. Since the convex function  $\lambda_n(\boldsymbol{\theta}) = G_n(\boldsymbol{\theta}; t) - \boldsymbol{\theta}^T \mathbf{W}_n(t)$  converges with probability 1 to the convex function  $\lambda(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Omega}_x \boldsymbol{\theta}$ , it follows from the convexity Lemma 3 that for any compact subset  $K \subset \mathbb{R}^p$ ,

$$\sup_{\boldsymbol{\theta} \in K} a_n^{-2} \left| G_n(\boldsymbol{\theta}; t) - \boldsymbol{\theta}^T \mathbf{W}_n(t) - \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Omega}_x \boldsymbol{\theta} \right| = o(1), \text{ a.s.}$$

The argument will be complete if we can show for each  $\epsilon > 0$  that, with probability 1,

$$\|\hat{\theta} - \bar{\theta}\| < a_n \epsilon.$$

Because  $\bar{\theta}$  is almost surely converged by Lemmas 4 and 5, it is bounded with probability 1. The compact subset  $K$  can be chosen to contain  $B(n)$ , a closed ball with center  $\bar{\theta}$  and radius  $a_n \varepsilon$ , thereby implying that

$$\Delta_n \equiv \sup_{\theta \in B(n)} a_n^2 \left| G_n(\theta; t) - \theta^T \mathbf{W}_n(t) - \frac{1}{2} \theta^T \Omega_x \theta \right| = o(1), \text{ a.s.} \tag{21}$$

Now, we consider the behavior of  $G_n(\theta; t)$  outside  $K$ . For any  $\theta = \bar{\theta} + a_n \varrho \mathbf{v}$ , with  $\varrho > \varepsilon$  and  $\mathbf{v}$  a unit vector. Define  $\theta^*$  as the boundary point of  $B_n$  that lies on the line segment from  $\bar{\theta}$  to  $\theta$ , that is,  $\theta^* = \bar{\theta} + a_n \varepsilon \mathbf{v}$ . Convexity of  $G_n(\theta; t)$  and the definition of  $\Delta_n$  imply

$$\begin{aligned} \frac{\varepsilon}{\varrho} G_n(\theta; t) + \left(1 - \frac{\varepsilon}{\varrho}\right) G_n(\bar{\theta}; t) &\geq G_n(\theta^*; t) \\ &= -(\theta^*)^T \Omega_x \bar{\theta} + \frac{1}{2} (\theta^*)^T \Omega_x \theta^* - a_n^2 \Delta_n \\ &= \frac{1}{2} a_n^2 \mathbf{v}^T \Omega_x \mathbf{v} \varepsilon^2 - \frac{1}{2} \bar{\theta}^T \Omega_x \bar{\theta} - a_n^2 \Delta_n \\ &\geq G_n(\bar{\theta}; t) + a_n^2 \left\{ \frac{1}{2} \mathbf{v}^T \Omega_x \mathbf{v} \varepsilon^2 - 2\Delta_n \right\}. \end{aligned}$$

The last expression does not depend on  $\theta$ . It follows that

$$\inf_{\|\theta - \bar{\theta}\| > a_n \varepsilon} G_n(\theta; t) \geq G_n(\bar{\theta}; t) + a_n^2 \frac{\varrho}{\varepsilon} \left[ \frac{1}{2} \mathbf{v}^T \Omega_x \mathbf{v} \varepsilon^2 - 2\Delta_n \right].$$

As  $\Omega_x$  is positively defined, then according to (21), with probability 1,  $2\Delta_n < \frac{1}{2} \mathbf{v}^T \Omega_x \mathbf{v} \varepsilon^2$  for enough  $n$ . This implies that for any  $\varepsilon > 0$  and for large enough  $n$ , the minimum of  $G_n(\theta; t)$  must be achieved with  $B(n)$ , i.e.,  $\|\hat{\theta} - \bar{\theta}\| < a_n \varepsilon$ , that is,

$$\sqrt{n} 2^{-m} (\hat{\beta}(t) - \beta(t)) = -\Omega_x^{-1} \mathbf{W}_n(t) + O(a_n), \text{ a.s.}$$

One obtains

$$R_n(m; \gamma, \nu) = O \left\{ \left( \frac{2^m}{n} \right)^{3/4} (\log n)^{1/2} \right\}.$$

We complete the proof of Theorem 1.  $\square$

**Proof of Corollary 1.** From Theorem 1, and Lemmas 4 and 5, it is easy to obtain the result of Corollary 1.  $\square$

**Proof of Corollary 2.** From Theorem 1 and Lemma 5, we only verify the asymptotic normality of

$$\mathbf{U}_n(t) = \sqrt{n/2^m} \sum_{i=1}^n \varphi_\tau(\varepsilon_i) \mathbf{X}_i \int_{A_i} K_m(t, s) ds.$$

$\square$

First, we compute its variance-covariance matrix. Let  $\mathbf{V}_i = \sqrt{n/2^m} \varphi_\tau(\varepsilon_i) \mathbf{X}_i \int_{A_i} K_m(t, s) ds$ . We know  $E(\mathbf{V}_i) = 0$ . Let  $S_1 = \{(j, i) : 1 \leq j - i \leq d_n; 1 \leq i < j \leq n\}$  and  $S_2 = \{(j, i) : 1 \leq i < j \leq n\} - S_1$  with  $d_n \rightarrow \infty$  specified later. We have

$$\begin{aligned} \text{Var}(\mathbf{U}_n(t)) &= \text{Var} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\} \leq \sum_{i=1}^n E(\mathbf{V}_i \mathbf{V}_i^T) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(\mathbf{V}_i, \mathbf{V}_j) \\ &= \sum_{i=1}^n E(\mathbf{V}_i \mathbf{V}_i^T) + 2 \sum_{S_1} \text{Cov}(\mathbf{V}_i, \mathbf{V}_j) + 2 \sum_{S_2} \text{Cov}(\mathbf{V}_i, \mathbf{V}_j) \\ &= I_1 + I_2 + I_3. \end{aligned} \tag{22}$$

For  $I_1$ , by Theorem 3.3 and Lemma 6.1 in [18], one obtains

$$\begin{aligned}
 I_1 &= \sum_{i=1}^n E(E(\mathbf{V}_i \mathbf{V}_i^T | \mathbf{X}_i)) = n2^{-m} \tau(1 - \tau) E(\mathbf{X}_1 \mathbf{X}_1^T) \sum_{i=1}^n \left( \int_{A_i} K_m(t, s) ds \right)^2 \\
 &= \tau(1 - \tau) E(\mathbf{X}_1 \mathbf{X}_1^T) \left\{ n2^{-m} \sum_{i=1}^n \left( \int_{A_i} K_m(t, s) ds \right)^2 - 2^{-m} \int_0^1 E_m^2(t, s) \kappa(s) ds \right\} \\
 &\quad + \tau(1 - \tau) E(\mathbf{X}_1 \mathbf{X}_1^T) 2^{-m} \int_0^1 E_m^2(t, s) \kappa(s) ds \\
 &= \tau(1 - \tau) \kappa(t) E(\mathbf{X}_1 \mathbf{X}_1^T) \int_0^1 E_0^2(0, s) ds + O(n\rho(n) + 2^m/n).
 \end{aligned} \tag{23}$$

For  $I_2$ , since  $n2^{-m} \rightarrow \infty$ , take  $d_n \rightarrow \infty$  such that  $d_n/(n2^{-m}) \rightarrow 0$ . Let  $(I_2)_{k,l}$  be the  $(k, l)$ th component of  $I_2$ . By Lemma 1, we have

$$\begin{aligned}
 |(I_2)_{k,l}| &\leq 2 \sum_{S_1} \left\{ E|V_{ik}V_{jl}| + E|V_{ik}|E|V_{jl}| \right\} \\
 &\leq Cn2^{-m} \sum_{S_1} \left\{ E|X_{ik}X_{jl}| \left| \int_{A_i} K_m(t, s) ds \int_{A_j} K_m(t, s) ds \right| + E^2\|X_1\| \left| \int_{A_i} K_m(t, s) ds \right|^2 \right\} \\
 &= n2^{-m} O\left( (2^m/n)^2 d_n \right) = O\left( \frac{2^m}{n} d_n \right) = o(1).
 \end{aligned} \tag{24}$$

For  $I_3$ , let  $(I_3)_{k,l}$  be the  $(k, l)$ th component of  $I_3$ . Noting that  $\kappa_0 > (2 + \delta)/\delta$ , by Proposition 2.5 (i) in [35], we have

$$\begin{aligned}
 \|(I_3)_{k,l}\| &\leq C \left\| \sum_{S_2} [E|V_{ik}|^{2+\delta}]^{2/(2+\delta)} \alpha^{\delta/(2+\delta)} (j - i) \right\| \\
 &\leq Cn2^{-m} \left( \frac{2^m}{n} \right)^2 \left\{ E|X_{ik}|^{2+\delta} \right\}^{2/(2+\delta)} \sum_{j=d_n}^{\infty} \alpha^{\delta/(2+\delta)} (j) \\
 &\leq O\left( \frac{2^m}{n} d_n^{-\kappa_0 \delta / (2+\delta) + 1} \right) = O\left\{ \left( \frac{2^m}{n} d_n \right) d_n^{-\kappa_0 \delta / (2+\delta)} \right\} = o(1).
 \end{aligned} \tag{25}$$

From Equations (22)–(25), we obtain

$$\text{Var}(\mathbf{U}_n(t)) = \tau(1 - \tau) \kappa(t) E(\mathbf{X}_1 \mathbf{X}_1^T) \int_0^1 E_0^2(0, s) ds + o(1). \tag{26}$$

Similar to the proof of Theorem 2 in [42], by using the small-block and large-block technique and the Cramér–Wold device, one can show that

$$\mathbf{Z}_n(t) \rightarrow N\left(0, \tau(1 - \tau) \kappa(t) \omega_0^2 \Psi_x\right). \tag{27}$$

This, in conjunction with Theorem 1 and the Slutsky Theorem, proves Corollary 2.

### 5. Simulation Study

To explore the numerical performances of quantile wavelet estimation, we compare our estimator with a local linear kernel [43] and Spline [44] by quantile regression. We call them QR-Wavelet, QR-Local-linear and QR-Spline methods, respectively. In the simulation study, our goal is to show that the QR-wavelet is robust to heavy-tailed data and more adaptive to a nonsmooth nonparametric function than local linear and spline methodologies. The data are of the form:

$$Y_i = X_i \beta(i/n) + \epsilon_{\tau, i}, \quad i = 1, \dots, n,$$

where  $X_i$  are random design points generated from the normal distribution  $N(2, 0.1^2)$  independently, and  $\epsilon_{\tau,i} = \epsilon_i - F^{-1}(\tau)$  with  $F$  being the distribution function of  $\epsilon_i$ . Here,  $F^{-1}(\tau)$  is subtracted from  $\epsilon_i$  to make the  $\tau$ th quantile  $\epsilon_{\tau,i}$  zero for identifiability.

We set  $n = 200, 300$  and  $500$ ;  $\tau = 0.10, 0.25, 0.50, 0.75$  and  $0.90$ ;  $\epsilon_i$  comes from the normal distribution  $N(0, \sigma^2)$  with  $\sigma = 0.1$  and the  $t$  distribution with  $d$  degrees of freedom (denoted as  $t(d)$ ) with  $d = 5$ , respectively. We consider two special curves of  $\beta(t)$  for  $t \in [0, 1]$  as follows:

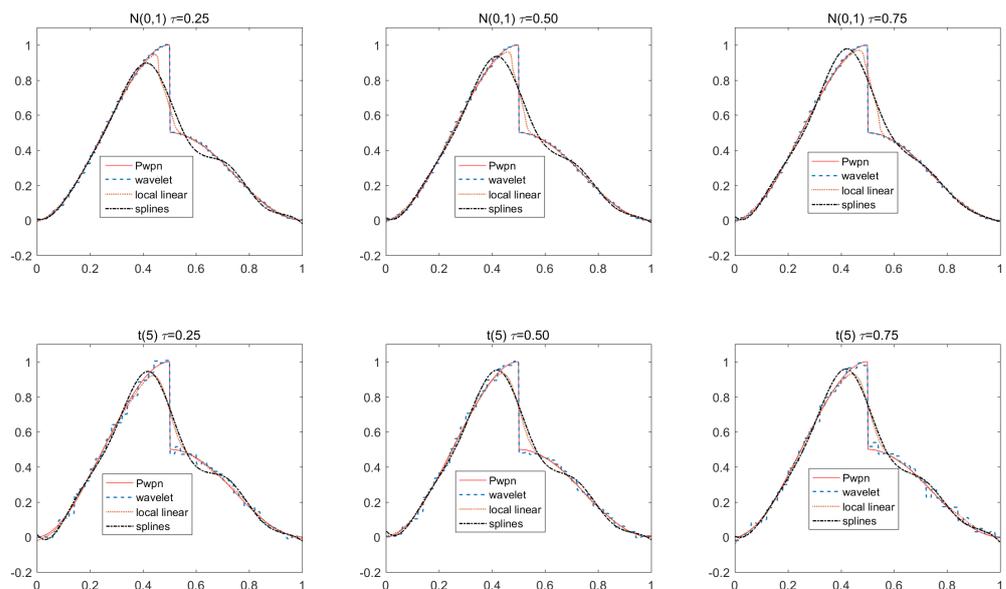
Case 1. *Pwpn* (piecewise polynomial function):  $\beta(t) = [4t^2(3 - 4t)]1_{[0,0.5]}(t) + [\frac{4}{3}t(4t^2 - 10t + 7) - 1.5]1_{(0.5,0.75]}(t) + [\frac{16}{3}t(t - 1)^2]1_{(0.75,1]}(t)$ . The function is generally smooth except for a jump at  $t = 0.5$ .

Case 2. *Blocks*:  $\beta(t) = \frac{0.6}{9.2}\{4\text{ssgn}(t - 0.1) - 5\text{ssgn}(t - 0.13) + 3\text{ssgn}(t - 0.15) - 4\text{ssgn}(t - 0.23) + 5\text{ssgn}(t - 0.25) - 4.2\text{ssgn}(t - 0.4) + 2.1\text{ssgn}(t - 0.44) + 4.3\text{ssgn}(t - 0.65) - 3.1\text{ssgn}(t - 0.76) + 2.1\text{ssgn}(t - 0.78) - 4.2\text{ssgn}(t - 0.81)\} + 0.2$ , where  $\text{ssgn}(t) = [1 + \text{sgn}(t)]/2$  with  $\text{sgn}(t) = 1_{(0,\infty)}(t) - 1_{(-\infty,0)}(t)$ . It is a step function with many jumps. Many jumps bring difficulties for the local linear and spline smoothing methods.

Pwpn and Blocks are shown in Figures 1 and 2, respectively. In the study, we use the Haar wavelet, which is the simplest of the wavelets. For a given sample size, take  $2^m = 0.6n^{2/3}$  in the QR-wavelet,  $h$  is chosen by “leave-one-out” cross-validation procedures and use the Gaussian kernel  $K(t) = \exp(-t^2/2)/\sqrt{2\pi}$  in the QR-local-linear. In the QR-splines, we use three degrees of the piecewise polynomial (cubic splines) and the knots  $[0.5n^{1/2} \log n]$ . The performances of the estimators are evaluated via the mean square error (MSE) based on the 200 repetitions, which is defined by

$$\text{MSE} = \frac{1}{M} \sum_{k=1}^M (\hat{\beta}_\tau(t_k) - \beta(t_k))^2,$$

where  $\{t_k, k = 1, \dots, M\}$  is a sequence of regular grid points. These results of MSE for  $\hat{\beta}_\tau(\cdot)$  are listed in Table 1 for Case I (Pwpn) and Table 2 for Case II (Blocks). The actual functions and their estimated curves for the two cases are depicted in Figures 1 and 2 when  $n = 300$ , and  $\tau = 0.25, 0.5$  and  $0.75$ , respectively.



**Figure 1.** The true time-varying coefficient function Papw, and their QR-wavelet, QR-local-linear and QR-splines estimates when  $n = 300$ , and  $\tau = 0.25, 0.50$  and  $0.75$  for the errors  $N(0, 1)$  and  $t(5)$ .

**Table 1.** MSEs of wavelet, local linear and spline smoothing in Case I: Pwpm.

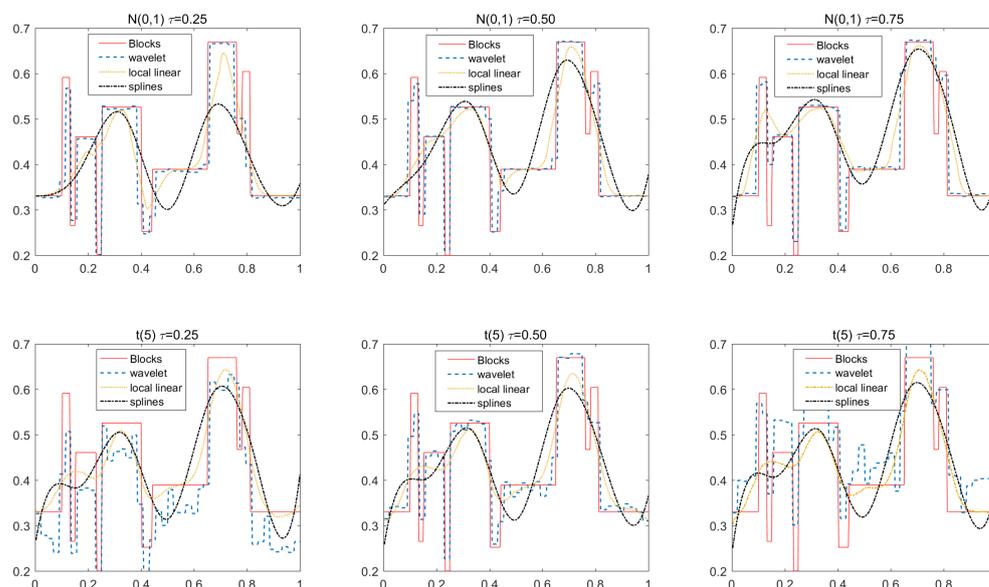
<i>n</i>	$\tau$	QR-Wavelet		QR-Local-Linear		QR-Splines	
		$N(0, 0.1^2)$	$t(5)$	$N(0, 0.1^2)$	$t(5)$	$N(0, 0.1^2)$	$t(5)$
200	0.10	0.00073	0.02273	0.00644	0.00479	0.01032	0.00824
	0.25	0.00065	0.00125	0.00383	0.00299	0.00859	0.00584
	0.50	0.00062	0.00085	0.00213	0.00293	0.00530	0.00529
	0.75	0.00065	0.00104	0.00293	0.00262	0.00860	0.00532
	0.90	0.00074	0.02208	0.00474	0.00345	0.01011	0.00660
300	0.10	0.00042	0.00966	0.00605	0.00287	0.01049	0.00557
	0.25	0.00040	0.00157	0.00355	0.00271	0.00868	0.00539
	0.50	0.00037	0.00055	0.00197	0.00250	0.00529	0.00524
	0.75	0.00039	0.00141	0.00279	0.00244	0.00867	0.00534
	0.90	0.00041	0.01121	0.00455	0.00254	0.01017	0.00561
500	0.10	0.00011	0.00319	0.00516	0.00262	0.00514	0.00319
	0.25	0.00010	0.00041	0.00311	0.00227	0.00358	0.00309
	0.50	0.00010	0.00026	0.00176	0.00213	0.00303	0.00300
	0.75	0.00010	0.00045	0.00255	0.00221	0.00354	0.00306
	0.90	0.00011	0.00354	0.00411	0.00252	0.00532	0.00339

**Table 2.** MSEs of wavelet, local linear and spline smoothing in Case II: Blocks.

<i>n</i>	$\tau$	QR-Wavelet		QR-Local-Linear		QR-Splines	
		$N(0, 0.1^2)$	$t(5)$	$N(0, 0.1^2)$	$t(5)$	$N(0, 0.1^2)$	$t(5)$
200	0.10	0.00573	0.01904	0.01114	0.00714	0.01730	0.00897
	0.25	0.00503	0.00368	0.00661	0.00593	0.00921	0.00778
	0.50	0.00378	0.00347	0.00599	0.00587	0.00803	0.00765
	0.75	0.00418	0.00384	0.00721	0.00613	0.00894	0.00773
	0.90	0.00574	0.02662	0.00865	0.00822	0.01150	0.01086
300	0.10	0.00624	0.01358	0.01059	0.00622	0.01751	0.00777
	0.25	0.00611	0.00553	0.00647	0.00556	0.00925	0.00764
	0.50	0.00484	0.00448	0.00584	0.00567	0.00800	0.00761
	0.75	0.00674	0.00557	0.00704	0.00570	0.00897	0.00775
	0.90	0.00682	0.01421	0.00852	0.00592	0.01139	0.00794
500	0.10	0.00562	0.00426	0.00958	0.00562	0.01541	0.00694
	0.25	0.00385	0.00368	0.00634	0.00528	0.00751	0.00671
	0.50	0.00310	0.00285	0.00557	0.00528	0.00706	0.00669
	0.75	0.00357	0.00325	0.00679	0.00550	0.00757	0.00681
	0.90	0.00408	0.00481	0.00813	0.00545	0.01000	0.00684

From Tables 1 and 2, we can make the following observations: (i) The MSEs of each time-varying coefficient  $\beta(\cdot)$  obtained by wavelet, local linear and spline techniques decrease with increasing the sample size  $n$ . The accuracy of QR-wavelet is obviously higher than that of QR-local linear and QR-splines. (ii) All QR methods work well when the random error comes from the  $t$  distribution with five freedoms; that is, QR is robust to heavy-tailed data. Based on MSE only, for Pwpm that is generally smooth except for a jump, all three methods perform almost equally well, but for Blocks with many jumps, the QR-wavelet performs better than QR-local linear and QR-splines. (iii) All three methods perform well for different quantile levels, especially for high and low quantile levels ( $\tau = 0.9$  (high),  $\tau = 0.1$  (low)). We also conducted some experiments based on the above setting by using least square estimation, and found that their estimators lead to systematic bias, especially at high and low quantile levels. However, QR-wavelet generally performs better than QR-local linear and QR-splines at these extreme quantile levels when there is a large sample size (e.g.,  $n = 500$ ) and multiple jump points (e.g., Blocks). From Figures 1 and 2, the estimators based on QR-local-linear and QR-splines are not better than those of QR-wavelet. For example, in Pwpm, QR-local linear and QR-splines cannot characterize the shape of the function in the interval (0.4, 0.6); in Blocks, they cannot find their jumping

points, while QR-wavelet detects these jumping points and represents the localized features of Pwpm and Blocks as a whole. Compared with the local linear and spline methods, the wavelet technique has great advantages in characterizing the local features of underlying curves. Therefore, the QR-wavelet overwhelmingly outperforms the QR-local-linear and QR-splines for the discontinuous/irregular functional coefficients in our time-varying coefficient models.



**Figure 2.** The true time-varying coefficient function Blocks and their QR-wavelet, QR-local-linear and QR-splines estimates when  $n = 300$ , and  $\tau = 0.25, 0.50$  and  $0.75$  for the errors  $N(0, 1)$  and  $t(5)$ .

**Author Contributions:** Conceptualization, X.Z.; methodology, X.Z.; validation, G.Y. and Y.X.; investigation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, X.Z.; supervision, X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Chinese National Social Science Fund (Grant No. 19BTJ034), the National Natural Science Foundation of China (Grant No. 12171242, 11971235) and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX22\_2210, KYCX21\_1940).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cochrane, J.H. *Asset Pricing*; Princeton University Press: Englewood Cliffs, NJ, USA, 2001.
2. Tsay, R. *Analysis of Financial Time Series*; Wiley: New York, NY, USA, 2002.
3. Robinson, P.M. Nonparametric Estimation of Time-Varying Parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*; Hackl, P., Ed.; Springer: Berlin, Germany, 1989; pp. 164–253.
4. Orbe, S.; Ferreira, E.; Rodríguez-Póo, J. Nonparametric estimation of time varying parameters under shape restrictions. *J. Econom.* **2005**, *126*, 53–77. [[CrossRef](#)]
5. Cai, Z. Trending time-varying coefficient time series models with serially correlated errors. *J. Econom.* **2007**, *136*, 163–188. [[CrossRef](#)]
6. Li, D.; Chen, J.; Lin, Z. Statistical inference in partially time-varying coefficient models. *J. Stat. Plan. Inference* **2011**, *141*, 995–1013. [[CrossRef](#)]
7. Wu, W.; Zhou, Z. Nonparametric inference for time-varying coefficient quantile regression. *J. Bus. Econ. Stat.* **2017**, *35*, 98–109. [[CrossRef](#)]
8. Hastie, T.; Tibshirani, R. Varying-coefficient models. *J. R. Stat. Soc. Ser. B* **1993**, *55*, 757–796. [[CrossRef](#)]

9. Fan, J.; Zhang, W. Statistical estimation in varying coefficient models. *Ann. Stat.* **1999**, *27*, 1491–1518. [[CrossRef](#)]
10. Hoover, D.R.; Rice, J.A.; Wu, C.O.; Yang, L.P. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **1998**, *85*, 809–822. [[CrossRef](#)]
11. Huang, J.Z.; Wu, C.O.; Zhou, L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat. Sin.* **2004**, *14*, 763–788.
12. Zhou, Z.; Wu, W.B. Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B* **2010**, *72*, 513–531. [[CrossRef](#)]
13. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.
14. Honda, T. Quantile regression in varying coefficient models. *J. Stat. Plan. Inference* **2004**, *121*, 113–125. [[CrossRef](#)]
15. Cai, Z.; Xu, X. Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Am. Stat. Assoc.* **2009**, *104*, 371–383. [[CrossRef](#)]
16. Kim, M.O. Quantile regression with varying coefficients. *Ann. Stat.* **2007**, *35*, 92–108. [[CrossRef](#)]
17. Andriyana, Y.; Gijbels, I. Quantile regression in heteroscedastic varying coefficient models. *ASTA Adv. Stat. Anal.* **2017**, *101*, 151–176. [[CrossRef](#)]
18. Antoniadis, A.; Gregoire, G.; McKeague, I.W. Wavelet methods for curve estimation. *J. Am. Stat. Assoc.* **1994**, *89*, 1340–1353. [[CrossRef](#)]
19. Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **1994**, *90*, 1200–1224. [[CrossRef](#)]
20. Hall, P.; Patil, P. Formulae for mean integrated squared error of non-linear wavelet-based density estimators. *Ann. Stat.* **1995**, *23*, 905–928. [[CrossRef](#)]
21. Härdle, W.; Kerkycharian, G.; Picard, D.; Tsybakov, A. *Wavelet, Approximation and Statistical Application*; Springer: New York, NY, USA, 1998.
22. Vidakovic, B. *Statistical Modeling by Wavelet*; Wiley: New York, NY, USA, 1999.
23. Doosti, H.; Afshari, M.; Niroumand, H.A. Wavelets for nonparametric stochastic regression with mixing stochastic process. *Commun. Stat.- Methods* **2008**, *37*, 373–385. [[CrossRef](#)]
24. Li, L.; Xiao, Y. Wavelet-based estimation of regression function with strong mixing errors under fixed design. *Commun. Stat.- Methods* **2017**, *46*, 4824–4842. [[CrossRef](#)]
25. Zhou, X.C.; Lin, J.G. Asymptotic properties of wavelet estimators in semiparametric regression models under dependent errors. *J. Multivar. Anal.* **2013**, *122*, 251–270. [[CrossRef](#)]
26. Ding, L.; Chen, P.; Li, Y. Berry-Esseen bound of wavelet estimator in heteroscedastic regression model with random errors. *Int. J. Comput. Math.* **2019**, *96*, 821–852. [[CrossRef](#)]
27. Zhou, X.; You, J. Wavelet estimation in varying-coefficient partially linear regression models. *Stat. Probab. Lett.* **2004**, *68*, 91–104. [[CrossRef](#)]
28. Lu, Y.; Li, Z. Wavelet estimation in varying-coefficient models. *Chin. J. Appl. Probab.* **2009**, *25*, 409–420.
29. Zhou, X.C.; Xu, Y.Z.; Lin, J.G. Wavelet estimation in varying coefficient models for censored dependent data. *Stat. Probab. Lett.* **2017**, *122*, 179–189. [[CrossRef](#)]
30. Bahadur, R.R. A note quantiles in large samples. *Ann. Math. Statist.* **1966**, *37*, 577–581. [[CrossRef](#)]
31. He, X.M.; Shao, Q.M. A general bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Stat.* **1996**, *24*, 2608–2630. [[CrossRef](#)]
32. Yang, W.; Hu, S.; Wang, X. The Bahadur representation for sample quantiles under dependent sequence. *Acta Math. Appl. Sin. Engl. Ser.* **2019**, *35*, 521–531. [[CrossRef](#)]
33. Zhou, X.; Zhu, F. Wavelet-M-estimation for time-varying coefficient time series models. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 1025452. [[CrossRef](#)]
34. Doukhan, P. Mixing: Properties and Examples. In *Lecture Notes in Statistics*; Springer: Berlin, Germany, 1994; Volume 85.
35. Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparametric and Parametric Methods*; Springer: New York, NY, USA, 2003.
36. Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc.* **1950**, *68*, 337–404. [[CrossRef](#)]
37. Hong, S.Y. Bahadur representation and its applications for local polynomial estimates in nonparametric M regression. *J. Nonparametr. Stat.* **2003**, *15*, 237–251. [[CrossRef](#)]
38. Walter, G.G. *Wavelets and Orthogonal Systems with Applications*; CRC Press Inc.: Boca Raton, FL, USA, 1994.
39. Kong, E.; Xia, Y. A single-index quantile regression model and its estimation. *Econom. Theory* **2012**, *28*, 730–768. [[CrossRef](#)]
40. Knight, K. Limiting distributions of  $L_1$  regression estimators under general conditions. *Ann. Stat.* **1998**, *26*, 755–770. [[CrossRef](#)]
41. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199. [[CrossRef](#)]
42. Cai, Z.; Fan, J.; Yao, Q. Functional-coefficient regression models for nonlinear time series. *J. Am. Stat. Assoc.* **2000**, *95*, 941–956. [[CrossRef](#)]
43. Wand, M.P.; Jones, M.C. *Kernel Smoothing*; Chapman and Hall: London, UK, 1995.
44. Chambers, J.M.; Hastie, T.J. *Statistical Models in S*; Wadsworth & Brooks/Cole: Salt Lake City, UT, USA, 1992.