Research Article

Wavelet-M-Estimation for Time-Varying Coefficient Time Series Models

Xingcai Zhou1,2 and Fangxia Zhu3

1Institute of Statistics and Data Science, Nanjing Audit University, Nanjing, Jiangsu, China
2School of Economics and Management, Southeast University, Nanjing, Jiangsu, China
3School of Mathematics and Finance, Chuzhou University, Chuzhou, Anhui, China

Correspondence should be addressed to Fangxia Zhu; zhufangxia75@163.com

Received 8 May 2020; Revised 17 July 2020; Accepted 29 July 2020; Published 3 September 2020

Abstract

This paper proposes wavelet-M-estimation for time-varying coefficient time series models by using a robust-type wavelet technique, which can adapt to local features of the time-varying coefficients and does not require the smoothness of the unknown time-varying coefficient. The wavelet-M-estimation has the desired asymptotic properties and can be used to estimate conditional quantile and to robustify the usual mean regression. Under mild assumptions, the Bahadur representation and the asymptotic normality of wavelet-M-estimation are established.

1. Introduction

The analysis of nonlinear and nonstationary time series, particularly with a time trend, has been very popular over the last two decades, because most time series data, coming from economics and finance data, are nonlinear or nonstationary or trending. Some nonlinear and nonstationary parametric, semiparametric, and nonparametric time series models have been proposed in the econometrics and statistics literature, for example, [1–6] and the references therein. One of the most attracted models is the time-varying coefficient time series models, which was formulated as follows:

\[ Y_i = X_i^\top \beta(t_i) + \varepsilon_i, \quad t_i = \frac{i}{n}, \quad i = 1, \ldots, n, \]

where \( Y_i \) is the response, \( \beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^\top \) is a \( p \)-dimensional vector of unspecified coefficient function defined on \([0,1]\), \( X_i = (X_{i1}, \ldots, X_{ip})^\top \) is a \( p \)-dimensional random vector, and \( \varepsilon_i \) is the random error.

There are many smoothers proposed to estimate the time-varying coefficient \( \beta(\cdot) \) in models (1), and the estimator is analyzed in the large-sample theory. Robinson [4] developed the Nadaraya–Watson method and showed the consistency and asymptotic normality of local constant estimator under the assumptions that the time series \( \{X_i\} \) is a stationary \( \alpha \)-mixing and the errors \( \{\varepsilon_i\} \) are i.i.d. and independent of \( X_i \). Cai [1] proposed the local linear approach and established asymptotic properties of the proposed estimators under the \( \alpha \)-mixing conditions and without specifying the error distribution. Hoover et al. [7] gave a smoothing spline and a locally weighted polynomial methods for longitudinal data and presented asymptotic properties. Li et al. [8] and Fan et al. [9] made statistical inference for the partially time-varying coefficient (errors-in-variables) model, respectively. These estimations are all based on local least squares, which are efficient for Gaussian errors, but least-squares estimation may perform poorly in the presence of extreme outliers. In addition, these methods are all based on an important assumption that \( \beta(\cdot) \) is of high smoothness. In reality, the assumption may be not satisfied. In some practical areas, such as signal and image processing, objects are frequently inhomogeneous. More robust estimation methods are required.

In this paper, we propose an M-type regression based on wavelet technique, which is called wavelet-M-estimation (WME), for the time-varying coefficient time series models.
There is the considerable literature devoted to M-estimation for nonparametric regression models. Fan et al. [10] obtained asymptotic normality of the M-estimator for local linear fit under independent observations. Hong [11] established a Bahadur representation for the local polynomial estimates in nonparametric M-regression under the i.i.d. random errors. Jiang and Mack [12] and Cai and Ould-Saïd [13] considered the local polynomial M-estimator and the local linear M-estimator for dependent observations and showed some asymptotic theories of the proposed estimators. For varying coefficient models, Tang and Cheng [14] showed asymptotic normality of local M-estimators for longitudinal data, and so on. However, the above works required the smoothness of the function being estimated, for example, assuming that the function has continuous first/second derivative. With wavelets, such assumptions are relaxed considerably. Because wavelet bases can adapt to local features of curves in both time and frequency domains, wavelet provides a new technique to analyze functions with discontinuities or sharp spikes. Therefore, it is natural to have better estimators than the local kernel method in many cases. Great achievements have been made for wavelet in nonparametric models, for example, Antoniadis et al. [15];Donoho and Johnstone [16]; Hall and Patil [17]; Härdele et al. [18]; Vidakovic [19]; Zhou and You [20]; Lu and Li [21]; and Donoho and Johnstone [16]. To the best of our knowledge, however, there exists a scaling function 

\[ \phi \in C^\infty \cap C^2 \cap \mathcal{L}_2 \] 

where \( \phi \) is a given convex function, 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 \( s_0 = 0 \), \( s_1 = 1 \), and \( s_i = (t_i + t_{i+1})/2 \), \( i = 1, \ldots, n - 1 \). As an alternative to (4), the following equation is also used to define the WME of \( \beta(t) \), that is, to find \( \beta \) to satisfy 

\[ \sum_{i=1}^{n} \rho(Y_i - X_i^T \beta) \int_{A_i} K_m(t, s)ds = 0, \] 

where \( \rho(\cdot) \) is a given convex function, 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 \( s_0 = 0 \), \( s_1 = 1 \), and \( s_i = (t_i + t_{i+1})/2 \), \( i = 1, \ldots, n - 1 \). As an alternative to (4), the following equation is also used to define the WME of \( \beta(t) \), that is, to find \( \beta \) to satisfy 

\[ \sum_{i=1}^{n} \rho(Y_i - X_i^T \beta) \int_{A_i} K_m(t, s)ds = 0, \] 

where \( \rho \) is a \( p \)-dimensional zero vector. It is a natural method to obtain (5) by taking the partial derivatives of (4) with respect to \( \beta \) when \( \rho(\cdot) \) is continuously differentiable and equating it to null, i.e., \( \psi(\cdot) = \rho'(\cdot) \). In this paper, we shall apply the suitable choice function \( \psi(\cdot) \) to (5), which includes many interesting cases such as least-squares estimation, the least absolute distance estimation, and quantile regression. See the monographs Huber and Ronchetti [24] and Koenker [25] for more details about the robustness of M-estimations and quantile regression, respectively.

Before stating the main results, we give the definition of \( \alpha \)-mixing dependence, which is necessary to establish our asymptotic theory for the time-varying coefficient model. Throughout, we assume that \( \{X_i, \varepsilon_i, k \geq 1 \} \) is stationary \( \alpha \)-mixing sequence. Recall that a sequence \( \{\varepsilon_k, k \geq 1 \} \) is said to be \( \alpha \)-mixing (or strong mixing) if the mixing coefficients, 

\[ \alpha(m) = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_k, B \in \mathcal{F}_{k+m} \}, \] 

close zero as \( m \to \infty \), where \( \mathcal{F}_k \) denotes the \( \sigma \)-field generated by \( \{\varepsilon_i, i \leq t \leq k \} \). The notion of \( \alpha \)-mixing is widely adopted in the study of nonparametric regression models. It is reasonably weak and is known to be fulfilled for many stochastic processes, including many familiar linear and nonlinear time series models. We refer to the monograph of Doukhan [26] and Fan and Yao [27] for some properties or more mixing conditions.
3. Asymptotic Theory

We first list the regularity conditions needed in the proof of the theorems although some of them might not be the weakest possible.

(A1) (i) The process \([X_n, \varepsilon_n]\) is a strictly stationary \(\alpha\)-mixing with \(\sum_{|j| \geq 1} \rho_j < \infty\) for some \(\alpha > (\delta/(2 + \delta))\) and \(\delta > 0\). (ii) \(\mathbb{E}[|X_1|^{2 + \delta}] < \infty\).

(A2) \(\rho(\cdot)\) is a convex function, and \(\psi(\cdot)\) is assumed to be any choice of the derivative of \(\rho(\cdot)\). Denote by \(\mathbb{D}\) the set of discontinuity points of \(\psi(\cdot)\). The common distribution function \(F\) of \(\varepsilon_i\) satisfying \(F(\mathbb{D}) = 0\).

(A3) \(\psi(\cdot)\) satisfies the following conditions:

(i) There exists some function \(B_1(\cdot)\) such that as \(u \to 0\),

\[
E[\psi(\varepsilon_1 + u) | X_1 = z] = B_1(z)u + o(|u|),
\]

where \(B_1(z)\) is continuous in a neighborhood of \(X\) with \(B_1(X) \neq 0\).

(ii) With probability 1,

\[
\max \left\{ E\left[ (\psi(\varepsilon_1 + u) - \psi(\varepsilon_i))^2 | X_1 = z \right] \right\} \leq B_2(|u|),
\]

holds uniformly for \(z\) in a neighborhood of \(X\), where \(B_2(\cdot)\) is continuous at \(u = 0\) with \(B_2(|u|) = O(|u|)\).

(iii) Let \(\gamma(z) = E(\psi^2(\varepsilon_1) | X_1 = z)\) and \(\gamma(z)\) be continuous in a neighborhood of \(X\) with \(\gamma(z) > 0\). Furthermore, \(E(\gamma^2(\varepsilon_1) | X_1 = z) < \infty\).

(iv) \(\Gamma_\varepsilon = E(\gamma(X_1)X_1X_1^T)\) and \(\Omega_\varepsilon = E(B_1(X_1)X_1X_1^T)\) are nonsingular matrices.

(A4) The time-varying coefficients \(\beta_j(\cdot)\) for \(j = 1, \ldots, p\), and the scaling function \(\phi(\cdot)\) in wavelet kernel satisfies the following conditions:

(i) \(\beta_j(\cdot)\) belongs to Sobolev space with order \(\nu > (1/2)\).

(ii) \(\beta_j(\cdot)\) satisfies the Lipschitz of order condition of order \(\nu > 0\).

(iii) \(\phi\) has a compact support and is in the Schwartz space with order \(\nu > 1\), satisfies the Lipschitz condition with order \(l\). Furthermore, \(|\phi(\xi) - 1| = O(\xi^n)\) as \(\xi \to 0\), where \(\phi\) is the Fourier transform of \(\phi\).

(A5) (i) For the designed points, \(\max_{|t_j|} |t_j - t_{j-1}| = O(n^{-l})\).

(ii) For some Lipschitz function \(\kappa(\cdot)\),

\[
\rho(n) = \max_i |s_i - s_{i-1} - \kappa(s_i)| = o(n^{-1}).
\]

(A6) The tuning parameter \(m\) satisfies (i) \(n^{2-m} \to \infty\). (ii) \(n^{2-m}(n^{-\nu} + \eta_m) \log n = o(1)\).

(iii) Let \(\nu^* = \min((5/3), (2(1 + \nu)/3), (1 + 2\nu/3)) - \varepsilon_1\) and \(\varepsilon_1 = 0\) for \(\nu^* \neq (3/2), \varepsilon_1 > 0\) for \(\nu = (3/2)\). Assume that \(n^{2-m^*} = O(1)\).

Some remarks on the conditions are in order.

Remark 1. Condition (A1) is the standard requirements for moments and the mixing coefficient for an \(\alpha\)-mixing sequence. It is well known that among various mixing conditions, for example, \(\alpha\), \(\rho\), and \(\psi\)-mixing, and \(\alpha\)-mixing is reasonably weak and can be used to depict many stochastic processes, including many familiar linear and nonlinear time series models. (A1) (i) is a very common condition, see Cai et al. [18]; Cai and Ould-Sa"id [13]; and Fan and Yao [27]; among others.

Remark 2. Conditions (A2) and (A3) are often imposed to establish the large-sample theory of M-estimation in parametric or nonparametric models, see, for example, Bai et al. [29]; Cai and Ould-Sa"id [13]; and Lin et al. [30]. They are mild and cover some well-known special cases, such as least-square estimation, Huber loss, and quantile. Some special examples are given as follows.

Remark 3. Conditions (A4) and (A5) are the mild regularity conditions for wavelet smoothing, which have been adopted by Antoniadis et al. [15]; Zhou and You [20]; and Zhou et al. [22]. In condition (A6), \(m\) acts as a tuning parameter, such as the bandwidth does for standard kernel smoothers; (A6) (i) and (ii) are for Bahadur representation and (A6) (i) and (iii) are for asymptotic normality of the WME. If (A6) (iii) holds, it implies (A6) (ii). There is a wide range of options to make A6 (i) and (iii) work. For example, if \(\nu = \gamma = 1\), then \(\nu^* = (4/3)\). Furthermore, take \(m^* = O(n^{1/5})\), then (A6) (i) and (iii) hold.

Recall that \(\beta_n(t) = \delta\) based on (1). Let \(\theta_n = \sqrt{n^{2-m} (\beta - \beta(t))}\), then \(\theta = \sqrt{n^{2-m} (\beta - \beta(t))}\). We can write the object function in equation (4) as

\[
\begin{aligned}
\Theta_n(\theta; t) = & \frac{1}{n^{1-2m}} \sum_{i=1}^n \rho \left( \psi_1(t) - \psi(t) \right) \\
& \cdot \int_{A_\theta} K_m(t, s) ds,
\end{aligned}
\]

and denote

\[
W_n(t) = \frac{1}{n^{1-2m}} \sum_{i=1}^n \psi_1(t)X_i \int_{A_\theta} K_m(t, s) ds.
\]

The first theorem is crucial for establishing the asymptotic properties of the WME.

**Theorem 1.** Under the conditions A1–A5 (i) and (A6) (i) and (ii), for any compact subset \(K \subset \mathbb{R}^p\), we have

\[
\sup_{\theta \in K} |\Theta_n(\theta; t) - \Theta(\theta; t) + W_n(t)^T \theta - \frac{1}{2} \theta^T \Omega_\varepsilon \theta| = O_p(b_n),
\]

where
Remark 4. Op shows that the WME has Bahadur order is WME for time-varying coefficient time series models (1), which normality of the WME.

Theorem 2. Under the conditions (A1)–(A5) (i) and (A6) (i) and (ii), we have

\[ \hat{\beta}(t) - \beta(t) = \Omega^{-1}_x \sum_{i=1}^n \psi(\varepsilon_i) X_i \int_{\alpha_i} K_m(t, s) ds + O_p(d_n), \]

uniformly in \( t \in [0, 1] \), where

\[ d_n = \left( n^{-\gamma} + \eta_m \right)^{1/4} + \left( \frac{2^m}{n} \right)^{1/4} \sqrt{\log n}. \]

If (A1)–(A5) (i) and (A6) (i) and (iii) hold, then

\[ d_n = \left( \frac{2^m}{n} \right)^{1/4} \sqrt{\log n}. \]

Remark 4. Theorem 2 gives the Bahadur representation of WME for time-varying coefficient time series models (1), which shows that the WME has Bahadur order is \( O_p \left( \left( \log \log(n/\rho(n)) \right)^{1/4} \right) \). It is slightly weaker than the Bahadur order \( O_p \left( \left( \log \log(n/\rho(n)) \right)^{1/4} \right) \) of Hong [11], where the bandwidth \( h \to 0 \). However, Hong [11] required strong smoothness: time-varying coefficient function \( \beta(\cdot) \) with the second-order differentiability. We have greatly relaxed this assumption. We do not need the strong assumption. Our degrees of smoothness of the function are less restrictive. See condition (A4) (i) and (ii).

With the help of Theorem 2, we can establish asymptotic normality of the WME.

\[ \sqrt{n^{2^{-m}} \left( \hat{\beta}^d(t) - \beta(t) \right)} \overset{D}{\to} N\left( 0, \kappa(t) \omega_0^2 \Omega^{-1}_x \Gamma_x \Omega^{-1}_x \right), \]

where \( f_\varepsilon(\cdot) \) is the density of \( \varepsilon_i \).

Theorem 3. Under the conditions (A1)–(A5) and (A6) (i) and (iii), we have

\[ \sqrt{n^{2^{-m}} \left( \hat{\beta}^d(t) - \beta(t) \right)} \overset{D}{\to} N\left( 0, \kappa(t) \omega_0^2 \Omega^{-1}_x \Gamma_x \Omega^{-1}_x \right), \]

where \( \hat{\beta}^d(t) = \hat{\beta}(t(\eta)) \) with \( \eta = \left( \frac{2^m t}{2^m} \right) \), \( \omega_0 = \int_0^1 E_0(0, u) du = \sum_k z_k^2 \), and \( \left[ u \right] \) denotes the maximum integer not greater than \( u \).

Remark 5. 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 Antoniadis et al. [15] have done. The main reason is that the variance of \( \hat{\beta}(t) \) as a function of \( t \) is unstable. It can be avoided by using dyadic points \( t(m) \). Also see Antoniadis et al. [15] for the details. From Theorems 2 and 3, we have the uniform weak consistency of WME:

\[ \sup_{t \in [0, 1]} \| \hat{\beta}(t) - \beta(t) \| = O_p \left( \left( \frac{2^m}{n} \right)^{1/2} \right). \]

Next, we shall give some special cases as corollaries of Theorem 3.

Corollary 1. Let \( \rho(z) = z^2 \) and \( \psi(z) = 2z \), which corresponds to the case of mean regression, which implies \( B_1 = 2 \), \( \gamma(X_i) = E(\varepsilon_i^2 | X_i) \), \( \Gamma_x = E(\gamma(X_i)X_i^T) \), and \( \Omega_x = E(X_iX_i^T) \). We have

\[ \sqrt{n^{2^{-m}} \left( \hat{\beta}^d(t) - \beta(t) \right)} \overset{D}{\to} N\left( 0, \frac{1}{4} \kappa(t) \omega_0^2 \Omega^{-1}_x \Gamma_x \Omega^{-1}_x \right). \]

Corollary 2. Let \( \rho(z) = z(\tau - I(z < 0)) \) and \( \psi(z) = z - I(z < 0) \), which corresponds to the case of quantile regression. Assume \( P(\varepsilon_i < 0 | X_i) = \tau \) a.s. and \( \varepsilon_i \) has a continuous positive conditional density \( f_{\varepsilon|X}(\cdot) \) and cumulative distribution function \( F_{\varepsilon|X}(\cdot) \) in a neighborhood of 0 given the \( X_i \). Thus, \( B_1 = f_{\varepsilon|X}(0) \) and \( \Gamma_x = [\tau^2 - (2\tau - 1) F_{\varepsilon|X}(0)]E(X_iX_i^T) \). Let \( \Omega_x = E(X_iX_i^T) \). We have

\[ \sqrt{n^{2^{-m}} \left( \hat{\beta}^d(t) - \beta(t) \right)} \overset{D}{\to} N\left( 0, \frac{1}{4} \kappa(t) \omega_0^2 \Omega^{-1}_x \Gamma_x \Omega^{-1}_x \right). \]

4. Technical Lemmas and Proofs

In the following sections, \( C \) is positive constant, which may be changed from line to line in the proof.

Lemma 1 (see Antoniadis et al. [15] and Walter [31]). Suppose that (A4) holds. We have
(i) \( E_0(t, s) \leq \left( c_k/(1+|t-s|)^k \right) \) and \( E_k(t, s) \leq \left( 2^k c_k/(1+2^k|t-s|)^k \right) \), where \( k \) is a positive integer and \( c_k \) is a constant depending on \( k \) only

(ii) \( \sup_{0 \leq s \leq t} |K_m(t, s)| = O(2^m) \)

(iii) \( \sup_{0 \leq s \leq t} \int_0^1 |K_m(t, s)|ds \leq c \), 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 (see Antoniadis et al. [15]). Suppose that (A4) and (A5) (i) hold and \( h(\cdot) \) satisfies (A4) (i) and (ii). 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

\[
\eta_m = \begin{cases} 
\left( \frac{1}{2m} \right)^{\gamma/2}, & \text{if } \frac{1}{2} < \gamma < \frac{3}{2}, \\
\frac{\sqrt{m}}{2m}, & \text{if } \gamma = \frac{3}{2}, \\
\frac{1}{2m}, & \text{if } \gamma > \frac{3}{2}.
\end{cases}
\]

Lemma 3 (see Lin and Lu [32]). Let \( \{X_i, i \geq 1\} \) be an \( \alpha \)-mixing sequence, \( X \in \mathcal{F}^{\leq \infty} \) and \( Y \in \mathcal{F}^{\leq \infty} \) with \( E|X|^p < \infty \) and \( E|Y|^q < \infty \), \( (1/p) + (1/q) < 1 \). Then,

\[
|EXY - EXEY| \leq 10(E|X|^p)^{1/p}(E|X|^p)^{1/q}(a(n))^{1-(1/p)-(1/q)}. \tag{25}
\]

Lemma 4 (see Pollard [33]). 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 \( \lambda_n(\theta) \rightarrow \lambda(\theta) \) in probability, for each fixed \( \theta \) in \( \Theta \). Then, for each compact subset \( K \) of \( \Theta \), in probability,

\[
\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \rightarrow 0. \tag{26}
\]

For simplicity, we introduce some notations before proving. We are interested in the asymptotic behaviors of \( \theta \), which can follow from the new optimization objective function:

\[
G_n(\theta; t) = \Theta_n(\theta; t) - \Theta_n(\theta; t). \tag{27}
\]

Furthermore, denote

\[
R_n(\theta; t) = G_n(\theta; t) - E[G_n(\theta; t)] + W_n(t)^T \theta.
\]

We rewrite \( G_n(\theta; t) \) as

\[
G_n(\theta; t) = E[G_n(\theta; t)] - W_n(t)^T \theta + R_n(\theta; t). \tag{29}
\]

Following Lemmas 5–7 give the asymptotic behaviors of

\[
R_n(\theta; t), E[G_n(\theta; t)] \text{ and } W_n(t) \text{ uniformly in } t \in [0, 1], \text{ respectively.}
\]

Lemma 5. Under the assumptions of Theorem 1, for fixed \( \theta \), it holds that

\[
\sup_{t \in [0, 1]} |R_n(\theta; t)| = O_p \left( \left( n^{2-\gamma} \right)^{1/2} \left( n^{-\gamma} + \eta_m \right)^{1/2} \right) + \left( n^{-1/2} \right) \sqrt{\log n}
\]

where \( R_n(\theta; t) \) is defined by (28).

Proof of Lemma 5. From the definition of \( R_n(\theta; t) \), we have

\[
R_n(\theta; t) = \left[ G_n(\theta; t) + W_n(t)^T \theta \right] - E \left[ G_n(\theta; t) + W_n(t)^T \theta \right]
\]

\[
= \frac{1}{n^{1/2}} \sum_{i=1}^n \left( U_i - E(U_i) \right) \int_{A_i} K_m(t, s)ds
\]

\[
= \frac{1}{n^{1/2}} \sum_{i=1}^n \left( U_i - E(U_i) \right),
\]

where

\[
U_i = U_i \int_{A_i} K_m(t, s)ds,
\]

\[
U_i = \rho \left( \epsilon_i + \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right) - \rho \left( \epsilon_i + \epsilon_i(t) \right) + \sqrt{n^{-1/2} \psi(\epsilon_i) \mathbf{X}_i^T \theta}
\]

By the convexity of \( \rho(\cdot) \), we have

\[
U_i \leq \psi \left( \epsilon_i + \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right) - \psi \left( \epsilon_i \right) \psi \left( \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right)
\]

\[
= \psi \left( \epsilon_i + \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right) \psi \left( \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right)
\]

\[
\leq \psi \left( \epsilon_i(t) - \sqrt{n^{-1/2} \mathbf{X}_i^T \theta} \right).
\]

(33)

Furthermore, we have

\[
\text{Var}(R_n(\theta; t)) = \left( \frac{n}{2m} \right)^2 \text{Var} \left( \sum_{i=1}^n U_i \right)
\]

\[
\leq \left( \frac{n}{2m} \right)^2 \left\{ \sum_{i=1}^n E \left[ U_i^2 \right] + 2 \sum_{i \in K, j \in S} \text{Cov}(U_i, U_j) \right\}.
\]

(34)

Let \( a_n = (2^m/n) \left( (n^{-\gamma} + \eta_m) + (2^m/n)^{3/2} \right) \). From conditions A1 (ii), A3 (ii), and A4 (ii), and Lemma 1, we have
\[ \sum_{i=1}^{n} E \left[ \tilde{U}_i^2 \right] = \sum_{i=1}^{n} E \left[ \psi \left( e_i + e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right) - \psi (e_i) \right] X_i^2 \times \left( e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right)^2 \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^2 \]
\[ \leq \sum_{i=1}^{n} E \left[ B_{2\epsilon} \left( e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right) \left( e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right)^2 \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^2 \right] \]
\[ \leq C \sum_{i=1}^{n} E \left[ \left( X_i^T (\beta(t) - \beta(t)) - \sqrt{n^{-1}2m} X_i^T \theta \right)^3 \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^2 \right] \]
\[ \leq C \sum_{i=1}^{n} E \left[ \left( X_i^T (\beta(t) - \beta(t)) \right)^3 + \left( \sqrt{n^{-1}2m} X_i^T \theta \right)^3 \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^2 \right] \]
\[ \leq C \sum_{i=1}^{n} E \left[ \|X_i\|^3 \left( \beta(t) - \beta(t) \right)^3 \left( \sqrt{n^{-1}2m} X_i^T \theta \right)^3 \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^2 \right] \]
\[ \leq O \left( \frac{2^{m} \left( n^{-\gamma} + \eta_{m} \right) + \left( \frac{m}{n} \right)^{5/2}}{n} \right) = O[a_n]. \]

To obtain an upper bound for the second term on the right-hand side of (34), we split it into two terms as follows. 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 $d_n \to \infty$ specified later. We have
\[ \sum_{1 \leq i < j \leq n} |\text{Cov}(\tilde{U}_i, \tilde{U}_j)| = \sum_{S_1} |\text{Cov}(\tilde{U}_i, \tilde{U}_j)| + \sum_{S_2} |\text{Cov}(\tilde{U}_i, \tilde{U}_j)| = J_{n1} + J_{n2}, \]
(36)

where $d_n$ is a sequence of positive integers such that $d_n = O(\log n)$ as $n \to \infty$.

For $J_{n1}$, by (35) and choice of $d_n$, we have
\[ J_{n1} \leq C \sum_{j=1}^{d_n} \sum_{i=1}^{n-j} |\text{Cov}(\tilde{U}_i, \tilde{U}_{i+j})| \leq d_n \sum_{i=1}^{n} E(U_i^2) = d_n O[a_n] = O[a_n \log n]. \]
(37)

By conditions A1 (ii), A3 (ii), and A4, and Lemmas 1 and 2, we have
\[ \begin{align*}
E \left( \tilde{U}_i^{2+\delta} \right) &= E \left[ \psi \left( e_i + e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right) - \psi (e_i) \right] X_i^{2+\delta} \times \left( e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right)^{2+\delta} \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^{2+\delta} \\
&\leq CE \left[ \left( e_i(t) - \sqrt{n^{-1}2m} X_i^T \theta \right)^{3+\delta} \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^{2+\delta} \right] \]
\[ \leq CE \left[ \left\|X_i\right\|^{3+\delta} \left\|\beta(t) - \beta(t)\right\|^{3+\delta} \left( \sqrt{n^{-1}2m} \right)^{3+\delta/2} \left\|X_i\right\|^{3+\delta} \left\|\theta\right\|^{3+\delta} \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^{2+\delta} \right] \\
&\leq C \left[ \left\|\beta(t) - \beta(t)\right\|^{3+\delta} + \left( n^{-1}2m \right)^{(3+\delta)/2} \left( \int_{A_{\epsilon}} K_m(t,s)ds \right)^{2+\delta} \right].
\]
(38)

For $J_{n2}$, by (38), conditions A1 (ii) and A3 (ii), and Lemmas 1–3, we have
\[
J_{\alpha} \leq C \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \sum_{i=1}^{n-d_n} \left[ E\left(U^{2+\delta}_i\right) \right]^{1/(2+\delta)} \left[ E\left(U^{2+\delta}_{ir_d+1}\right) \right]^{1/(2+\delta)} \\
\leq C \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \\
\times \sum_{i=1}^{n-d_n-1} \left\{ \| \beta(t_i) - \beta(t) \|^{3+\delta} + \left( n^{-1} 2^m \right)^{(3+\delta)/2} \right\}^{1/(2+\delta)} \left( \int_{A_i} K_m(t,s)ds \right) \\
\times \left\{ \| \beta(t_{ir_d+1}) - \beta(t) \|^{3+\delta} + \left( n^{-1} 2^m \right)^{(3+\delta)/2} \right\}^{1/(2+\delta)} \left( \int_{A_{ir_d+1}} K_m(t,s)ds \right) \\
\leq C n \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \\
\times \left\{ \sum_{i=1}^{n-d_n-1} \left\{ \| \beta(t_i) - \beta(t) \|^{3+\delta} + \left( n^{-1} 2^m \right)^{(3+\delta)/2} \right\} \left( \int_{A_i} K_m(t,s)ds \right)^2 \right\}^{2/(2+\delta)} \\
\leq C n \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \sum_{i=1}^{n-d_n-1} \left\{ \| \beta(t_i) - \beta(t) \|^{3+\delta} + \left( n^{-1} 2^m \right)^{(3+\delta)/2} \right\} \left( \int_{A_i} K_m(t,s)ds \right)^2 \\
\leq C n \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \left\{ n^{-\gamma} + \eta_m \right\}^{(5+\delta)/(2+\delta)} \\
\leq C n \sum_{j=1}^{d_n} [\alpha(j)]^{\delta/(2+\delta)} \left\{ n^{-\gamma} + \eta_m \right\}^{5/12} \\
\leq C n [\alpha(j)]^{\delta/(2+\delta)} \left\{ n^{-\gamma} + \eta_m \right\}^{5/12} \\
\leq C n \left\{ d_n + 1 \right\}^{-a} \sum_{j=1}^{n-d_n+1} \left\{ j^a [\alpha(j)]^{\delta/(2+\delta)} \right\} = o(n). 
\]

From (36), (37), and (39), one gets \( \sum_{1 \leq i < j \leq n} \text{Cov}(U_i, U_j) = O(\alpha_n \log n) \). Combining with (34) and (35), we have

\[
\text{Var}(R_n(\Theta; t)) = O\left( (n^{-2} - m)^2 \alpha_n \log n \right), \quad (40)
\]

uniform in \( t \in [0,1] \). Note that \( E(R_n(\Theta; t)) = 0 \). We have

\[
R_n(\Theta; t) = O_P\left( \sqrt{\left( n^{-2} - m \right)^2 \alpha_n \log n} \right) \\
= O_P\left( \left( n^{-2} - m \right)^{1/2} \left( n^{-\gamma} + \eta_m \right)^{1/2} + \left( n^{-1} 2^m \right)^{1/4} \right) \sqrt{\log n}, \quad (41)
\]

uniform in \( t \in [0,1] \).

\[\Box\]

**Lemma 6.** Under the assumptions of Theorem 1, for fixed \( \Theta \), it holds that

\[
\sup_{t \in [0,1]} \left| E\left( G_n(\Theta; t) \right) \right| = \frac{1}{2} \Theta^T \Sigma \Theta = O\left( \sqrt{n^{2-m}} \left( n^{-\gamma} + \eta_m \right) \right) = o(1), \quad (42)
\]

where \( G_n(\Theta; t) \) is defined by (1).

**Proof of Lemma 6.** By condition A2 (i), Lemma 2, and Lemma 1 in Bai et al. [29], we have
\[ E[G_n(\theta; t)] = \frac{1}{n^{-2m}} \sum_{i=1}^{n} E \left\{ E \left[ \rho \left( \varepsilon_i + \varepsilon_i(t) - \sqrt{n^{-1/2}} X_i^\top \theta \right) - \rho(\varepsilon_i + \varepsilon_i(t)) \right] X_i \right\} \int_A K_m(t, s) ds \]

\[ = \frac{1}{n^{-2m}} \sum_{i=1}^{n} E \left\{ \frac{1}{2} \left( \sqrt{n^{-1/2}} X_i^\top \theta \right)^2 - \varepsilon_i(t) \sqrt{n^{-1/2}} X_i^\top \theta \right\} B_1(X_i) (1 + o(1)) \int_A K_m(t, s) ds \]

\[ = \frac{1}{2} \theta^T \Omega_{\theta, \theta} - \sqrt{n} \frac{\Gamma \cdot \eta_m}{\sqrt{\pi}}. \]

(43)

By using the same argument as those used in Proof of Lemma 5, we obtain

\[ \sum_{1 \leq i < j \leq m} \text{Cov}(\xi_i, \xi_j) = o(1). \]  

(47)

Now, we prove the first term on the right-hand side of (46). Using the compact support and Lipschitz properties of \( \phi \), one can show that \( E_0(t, \cdot) \) is Lipschitz uniformly in \( t \), so that

\[ \sup_{r \in [0, 1]} \left| E_0(t, 2^m v_r) - E_0(t, 2^m u_t) \right| = O(2^m n^{-1}), \]  

(48)

and the Lipschitz property of \( \kappa(\cdot) \) implies

\[ \kappa(v_t) = \kappa(s_j) + O(n^{-1}), \]  

(49)

where \( u_t \) and \( v_t \) belong to \( A_t \). By condition A3 (iii), we have
Since \( 2^{-m} \int \varphi(t,s) \omega(t) \, ds \) is unstable as a function of \( t \),
we need to compute it at dyadic points of order \( m \). By Lemma 6.1 in Antoniadis et al. [15], we have
\[
\lim_{m \to \infty} 2^{-m} \int_0^1 E^2(t^{(m)}, s) \omega(s) \, ds = \omega(t)\omega'_0. \tag{51}
\]
Combining with (46), (47), and (50), we have
\[
\Var(W_n(t)) = \omega^2_0 \omega(t) \Gamma_x + o(1). \tag{52}
\]
Second, we shall show (44). Redefine
\[
\xi_i = \sqrt{n} \psi^{\epsilon}(\varepsilon_i) \mathbf{X} \int_{A} K_m(t,s) \, ds
- E \left( \sqrt{n} \psi^{\epsilon}(\varepsilon_i) \mathbf{X} \int_{A} K_m(t,s) \, ds \right). \tag{53}
\]
Note that \( E(\psi(\varepsilon_i) | X_1) = o(1) \), and it turns (44) to
\[
\sum_{i=1}^{n} \xi_i^d \overset{D}{\to} N(0, \omega^2_0 \omega(t) \Gamma_x), \tag{54}
\]
where the definition of \( \xi_i^d \) is similar to \( \xi_i \) by using \( t^{(m)} \) to replace \( t \). As \( \psi(\cdot) \) is not necessarily bounded, we adopt a truncation method. Denote \( \psi^\ell(\cdot) = \psi(\cdot) I(\psi(\cdot) \leq L) \) and as before that \( \xi_i^d = \sqrt{n} \psi^\ell(\varepsilon_i) \mathbf{X} \int_{A} K_m(t^{(m)}, s) \, ds \)
\[= E \left( \sqrt{n} \psi^\ell(\varepsilon_i) \mathbf{X} \int_{A} K_m(t^{(m)}, s) \, ds \right) \text{ and let } \Gamma_x \text{ be defined as before with } \gamma^\ell(\cdot) \text{ replaced by } E(\psi^\ell(\varepsilon_i) | X_1). \tag{55}
\]
Similar to Proof of Theorem 2 in Cai et al. [28], by using Doob’s large-block and small-block technique, we can show that
\[
\sup_{t} \left| G_n(\theta; t) - \frac{1}{2} \theta^T \Omega_x \theta + W_n(t) \theta \right|
= O_P \left\{ \sqrt{\frac{n^{-2} \varphi}{(\eta + \eta_m)}} + \sqrt{\frac{n^{-2} \varphi}{(\eta + \eta_m)^{1/2} + (n^{-1} 2m)^{1/4}}} \right\} \tag{56}
\]
\[
= O_P \left( b_n \right) = o(1).
\]
From Lemma 7, it is easy to see that \( W_n(t) \) is stochastically bounded. Since the convex function \( \lambda_n(\theta) = G_n(\theta; t) - W_n(t) \) converges in probability to the convex function \( \lambda(\theta) = (1/2) \theta^T \Omega_x \theta \), it follows from convexity Lemma 4 that for any compact set \( K \),
\[
\sup_{\theta \in K} \sup_{t \in [0,1]} \left| G_n(\theta; t) - \frac{1}{2} \theta^T \Omega_x \theta + W_n(t) \right| = O_P \left( b_n \right) = o(1). \tag{59}
\]
Notice that the convexity lemma strengthens the pointwise result to uniform convergence on compact subsets of \( \mathbb{R}^p \). This completes Proof of Theorem 1. \( \square \)

Proof of Theorem 2. To obtain Bahadur representation of WME, the idea behind the proof is to approximate \( G_n(\theta; t) \)
by a quadratic function whose minimizing value has an asymptotic the behavior, and then to show that \( \theta_n \) lies close enough to the minimizing value to share its asymptotic behaviour. We have done the first step, that is, the results of Theorem 1 and Lemma 7. Let \( \bar{\theta}_n = \Omega^{-1} W_n(t) \) and \( c^2_n = b_n \sqrt{\log n} \). Now, we prove the second step. The argument will be complete if we can show for each \( \epsilon > 0 \) that
\[
P \left( \| \bar{\theta}_n - \bar{\theta}_m \| > c_n \epsilon \right) = o(1). \tag{61}
\]
The argument is similar to Proof of Theorem 1 in Pollard [33], whose method is extended to obtain the Bahadur representation of WME. From Theorem 1, the compact set \( K \) can be chosen to contain a closed ball \( B(n) \) with center \( \bar{\theta}_n \) and radius \( c_n \epsilon \). With probability arbitrarily close to one. Therefore, it implies that
\[ \Delta_n = c_n^{-2} \sup_{\theta \in B(n) \cap (0,1)} \sup_{t \in [0,1]} \left| G_n(\theta; t) - \frac{1}{2} \theta^T \Omega_x \theta + W_n(t)^T \theta \right| \]
\[ = O_P \left( \frac{1}{\sqrt{\log n}} \right) = o_P(1). \] (62)

Now, we consider the behavior of \( G_n(\theta; t) \) outside \( B(n) \). Suppose \( \theta = \theta_n + c_n \varepsilon \) with \( \varepsilon > 0 \) and \( \varepsilon \) a unit vector. Define \( \theta^* \) as the boundary point of \( B(n) \) that lies on the line segment from \( \theta_n \) to \( \theta \), that is, \( \theta^* = \theta_n + c_n \varepsilon \). Convexity of \( G_n(\theta; t) \) and the definition of \( \Delta_n \) imply

\[ \frac{\varepsilon}{\theta^*} G_n(\theta; t) + \left( 1 - \frac{\varepsilon}{\theta^*} \right) G_n(\theta_n; t) \geq G_n(\theta^*; t) \]
\[ \geq \frac{1}{2} (\theta^*)^T \Omega_x \theta^* + W_n(t)^T \theta^* - c_n^2 \Delta_n \]
\[ \geq \frac{1}{2} \theta_n^T \Omega_x \theta_n + W_n(t)^T \theta_n + c_n^2 \left( \frac{1}{2} \varepsilon^T \Omega_x \varepsilon - \Delta_n \right) \]
\[ \geq G_n(\theta_n; t) + c_n^2 \left( \frac{1}{2} \varepsilon^T \Omega_x \varepsilon - \Delta_n \right). \] (63)

Furthermore, we have
\[ G_n(\theta; t) \geq G_n(\theta_n; t) + c_n^2 \left( \frac{1}{2} \varepsilon^T \Omega_x \varepsilon - 2 \Delta_n \right), \] (64)
uniformly in \( t \in [0,1] \). In addition, the right-hand side of the above inequality does not depend on \( \theta \). Hence, for each \( \varepsilon > 0 \) and \( n \) large enough, we have
\[ P \left\{ \inf_{\| \theta - \theta_n \| > c_n \varepsilon} G_n(\theta; t) \geq G_n(\theta_n; t) + c_n^2 \left( \frac{1}{2} \varepsilon^T \Omega_x \varepsilon - 2 \Delta_n \right) \right\} \rightarrow 1, \] (65)
which leads to \( \| \theta - \theta_n \| \leq c_n \varepsilon \) with probability tending to one, because \( 2 \Delta_n < (1/2) \varepsilon^T \Omega_x \varepsilon \), which happens with probability tending to one, the minimum of \( G_n(\theta; t) \) cannot occur at any \( \theta \) with \( \| \theta - \theta_n \| > c_n \varepsilon \). So, (61) holds, which leads to
\[ \sqrt{n} \left( \hat{\beta}(t) - \beta(t) \right) = \Omega_x^{-1} W_n(t) + O_p(c_n), \] (66)
uniformly in \( t \in [0,1] \). Thus, Proof of Theorem 2 is completed. \( \square \)

**Proof of Theorem 3.** From Theorem 2 and Lemma 7, we can directly obtain the result of Theorem 3. Thus, we complete Proof of Theorem 3. \( \square \)

**Data Availability**

No data were used to support the findings of the study.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**Acknowledgments**

This work was partially supported by the National Social Science Fund of China (no. 19BTJ034), NSFC (no. 11971235 and 11831008), and China Postdoctoral Science Foundation (no. 2018T110422 and 2016M590396).

**References**

[1] Z. Cai, "Trending time-varying coefficient time series models with serially correlated errors," *Journal of Econometrics*, vol. 136, no. 1, pp. 163–188, 2007.

[2] J. H. Cochrane, *Asset Pricing*. Princeton University Press, Englewood Cliffs, NJ, USA, 2001.

[3] P. C. B. Phillips, "Trending time series and macroeconomic activity: some present and future challenges," *Journal of Econometrics*, vol. 100, no. 1, pp. 21–27, 2001.

[4] P. M. Robinson, "Nonparametric estimation of time-varying parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 164–253, Springer, Berlin, Germany, 1989.

[5] R. Tsay, *Analysis of Financial Time Series*, Wiley, New York, NY, USA, 2002.

[6] K. Q. Wang, "Asset pricing with conditioning information: a new test," *The Journal of Finance*, vol. 58, no. 1, pp. 161–196, 2003.

[7] D. Hoover, J. A. Rice, C. O. Wu, and L. P. Yang, "Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data," *Biometrika*, vol. 85, no. 4, pp. 809–822, 1998.

[8] D. Li, J. Chen, and Z. Lin, "Statistical inference in partially time-varying coefficient models," *Journal of Statistical Planning and Inference*, vol. 141, no. 2, pp. 995–1013, 2011.

[9] G.-L. Fan, H.-Y. Liang, and J.-F. Wang, "Statistical inference for partially time-varying coefficient errors-in-variables models," *Journal of Statistical Planning and Inference*, vol. 143, no. 3, pp. 505–519, 2013.

[10] J. Fan, T. C. Hu, and Y. K. Truong, "Robust non-parametric function estimation," *Scandinavian Journal of Statistics*, vol. 21, pp. 433–446, 1994.

[11] S.-Y. Hong, "Bahadur representation and its applications for local polynomial estimators in nonparametric m regression," *Journal of Nonparametric Statistics*, vol. 15, no. 2, pp. 237–251, 2003.
[12] J. Jiang and Y. P. Mack, "Robust local polynomial regression for dependent data," *Statistica Sinica*, vol. 11, pp. 705–722, 2001.

[13] Z. Cai and E. Ould-Saïd, "Local m-estimator for nonparametric time series," *Statistics & Probability Letters*, vol. 65, no. 4, pp. 433–449, 2003.

[14] Q. Tang and L. Cheng, "Asymptotic normality of m-estimators for varying coefficient models with longitudinal data," *Communications in Statistics—Theory and Methods*, vol. 38, pp. 1422–1440, 2009.

[15] A. Antoniadis, G. Gregoire, and I. W. McKeague, "Wavelet methods for curve estimation," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1340–1353, 1994.

[16] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1994.

[17] P. Hall and P. Patil, "Formulae for mean integrated squared error of nonlinear wavelet-based density estimators," *Annals of Statistics*, vol. 23, no. 3, pp. 905–928, 1995.

[18] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov, *Wavelet, Approximation and Statistical Application*, Springer, New York, NY, USA, 1998.

[19] B. Vidakovic, *Statistical Modeling by Wavelet*, Wiley, New York, NY, USA, 1999.

[20] X. Zhou and J. You, "Wavelet estimation in varying-coefficient partially linear regression models," *Statistics & Probability Letters*, vol. 68, no. 1, pp. 91–104, 2004.

[21] Y. Lu and Z. Li, "Wavelet estimation in varying-coefficient models," *Chinese Journal of Applied Probability*, vol. 25, pp. 409–420, 2009.

[22] X.-c. Zhou, Y.-z. Xu, and J.-g. Lin, "Wavelet estimation in varying-coefficient models for censored dependent data," *Statistics & Probability Letters*, vol. 122, pp. 179–189, 2017.

[23] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[24] P. Huber and E. Ronchetti, *Robust Statistics*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2009.

[25] R. Koenker, *Quantile Regression*, Cambridge University Press, Cambridge, UK, 2005.

[26] P. Doukhan, "Mixing: properties and examples," in *Lecture Notes in Statistics* Vol. 85, Springer, Berlin, Germany, 1994.

[27] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York, NY, USA, 2003.

[28] Z. Cai, J. Fan, and Q. Yao, "Functional-coefficient regression models for nonlinear time series," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 941–956, 2000.

[29] Z. D. Bai, C. R. Rao, and Y. Wu, "M-estimation of multivariate linear regression parameters under a convex discrepancy function," *Statistica Sinica*, vol. 2, pp. 237–254, 1992.

[30] Z. Lin, D. Li, and J. Gao, "Local Linear M-estimation in nonparametric spatial regression," *Journal of Time Series Analysis*, vol. 30, no. 3, pp. 286–314, 2009.

[31] G. G. Walter, *Wavelets and Orthogonal Systems with Applications*, CRC Press Inc., Boca Raton, FL, USA, 1994.

[32] Z. Lin and C. R. Lu, *Limit Theory for Mixing Dependent Random Variables*, Science Press, Beijing, China, 1996.

[33] D. Pollard, "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, vol. 7, no. 2, pp. 186–199, 1991.