Local Linear Quantile Regression for Time Series under near epoch dependence

Xiaohang Ren

University of Southampton

x.ren@soton.ac.uk

Zudi Lu

University of Southampton

z.lu@soton.ac.uk

May, 2020

Abstract

This paper aims to establish asymptotic normality of the local linear kernel estimator for quantile regression under near epoch dependence, a useful concept in characterising time series dependence of extensive interests in Econometrics. In particular, near epoch dependence can cover a wide range of linear or nonlinear time series models that are even not of strong or α-mixing property (a property usually assumed in the nonlinear time series literature). Under the mild conditions, the Bahadur representation of the quantile regression estimators is established in weak convergence sense. The method provides much richer information than mean regression and covers much more processes, which do not satisfy general mixing conditions. Simulation and application to a real data set are studied, which demonstrate the usefulness of the introduced method for analysis of time series. The theoretical results of this paper will be of widely potential interest for time series econometric semiparametric quantile regression modelling.

Keywords: Local linear fitting; Quantile regression; Near epoch dependence; Bahadur representation

JEL: C10 C14 C21

1. Introduction

Nonlinear modelling of time series data has drawn a lot of attention from researchers in the past few decades (c.f., Fan & Yao, 2003; Härdle et al., 1997; Tjøstheim & Auestad, 1994; Tong, 1990, etc). In these literatures, most modellings are based on conditional mean regression perspective, while fewer work has focused on nonlinear modelling by quantile regression for time series. In this study, we are mainly concerned with the nonlinear modelling of time series data from view point of quantile regression. Specifically, we consider a (strictly) stationary time series that are near-epoch dependent (NED) introduced by Ibragimov (1962).

Assume that $\{(Y_t, X_t)\}$ is a stationary multivariate time series on a probability space $(\Omega, \mathcal{F}, P)$ in general context, where X_t and Y_t are random variables taking their values in $\mathbb{R}^p$ and $\mathbb{R}^1$ respectively. In time series econometrics, X_t may involve both the lags of endogenous and/or exogenous variables. We are here interested in the τ-th $(0 < \tau < 1)$ conditional quantile function of Y_t given $X_t = x$.

$$q_\tau(x) = \arg\min_{a \in R^1} E\left\{\rho_\tau(Y_i - a)|X_i = x\right\}, \tag{1.1}$$

where $\rho_\tau(y) = y(\tau - I_{\{y<0\}})$ with $y \in \mathbb{R}^1$ and I_A is the indicator function of set A. This *conditional quantile regression* was initially developed under *i.i.d.* samples for linear regression models in econometric literature (Koenker & Bassett Jr, 1978, 1982).

In comparison with the mean regression, the modelling based on quantile regression has some essential advantages. Firstly, a well known special case with $\tau = 1/2$, i.e., median regression, is much explored and more robust than the mean regression when the data distribution is typically skewed or possesses a few outliers. Secondly, a heteroscedastic model can be easily detected if the regression quantiles of the model are not parallel (c.f., Efron, 1991). Thirdly, a collection of conditional quantiles can describe the whole distribution of the independent variable (c.f., Yu & Jones, 1998). Fourthly, pairs of extreme conditional quantiles can be used to depict the conditional prediction interval, which is important in econometric forecasting (see, Granger et al., 1989; Koenker & Zhao, 1996;

Kuester et al., 2006, for example). Finally, the value-at-risk (VaR) has become a popular tool to measure market risk, which is just the quantile of the potential loss to be expected over a given future period (for a given probability) (c.f., Jorion, 1997). Therefore, the regression quantile would be helpful to factor analysis of the risk modelling based on VaR.

The asymptotic properties of kernel estimators for quantile regression have been investigated under *i.i.d.* or α-mixing conditions. To be specific, Yu & Jones (1998) used local linear fitting to estimate the quantile regression, Cai (2002) introduced an estimate of conditional quantile based on the inverse of conditional distribution function (Hall et al., 1999), and Honda (2000) estimate the quantile regression by local polynomial fitting. Recently, for time-varying coefficient models, Honda (2004) and Kim (2007) considered the nonparametric quantile estimation based on the i.i.d. errors term. Cai & Xu (2009) assumed that error is α-mixing process. In Wu & Zhou (2017), the errors are locally stationary processes with cross-dependent. Moreover, Cai & Xiao (2012) investigated the partially varying coefficients quantile regression models based on β-mixing assumption, and Wang et al. (2009) without any specification of the error distribution. Oberhofer & Haupt (2016) established the asymptotic properties of the nonlinear quantile regression, with allowing the error process to be heterogeneous and mixing. Differently, we will develop a theory for nonlinear modelling of the quantile regression function in nonparametric way for time series under NED condition. The main motivation for the NED condition is that α-mixing is a stricter assumption, which is hard to cover compound processes (Davidson, 1994; Lu & Linton, 2007). For this reason, the NED processes, which can cover more kinds of processes, have been considered extensively in econometrics (Andrews, 1995; Gallant, 2008; Jenish, 2012; Li et al., 2012; Lu, 2001).

We will introduce a local linear kernel method to estimate both the quantile regression and its derivatives under NED condition, which is an extension of the work of Welsh (1996) and Yu & Jones (1998) under *i.i.d.* samples. In addition, we establish the asymptotic distribution for quantile regression estimators, which is important both for assessing the estimates and constructing an asymptotic confidence interval as well as testing hypothesis on the quantile regression, where a powerful tool of Bahadur representation will be established in convergence in probability for our aim. Our asymptotic normality result generalizes Honda (2000) who obtained a Bahadur representation with rather strong conditions.

The rest of the paper is organized as follow. Section 2 shows the definition of setting and estimator we shall examine. Section 3 proves asymptotic normality of local linear estimator for quantile regression under NED. Section 4 contains some numerical results based on some common econometric models.

2. Local linear quantile regression estimators

The stationary processes Y_t and X_t are $\mathbb{R}^1$- and $\mathbb{R}^p$- valued random fields, respectively. And

$$Y_t = \Psi_Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots), \tag{2.1}$$

$$X_t = (X_{t1}, \cdots, X_{tp})' = \Psi_X(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots), \tag{2.2}$$

where the stationary process$\{\varepsilon_t\}$ may be vector-valued, X' denotes the transpose of the vector X, $\Psi_Y : \mathbb{R}^\infty \longrightarrow \mathbb{R}^1$ and $\Psi_X : \mathbb{R}^\infty \longrightarrow \mathbb{R}^p$ are two Borel measurable functions, respectively.

In this research, we are interested in estimating the τ-th ($0 < \tau < 1$) conditional quantile function of Y_t given $X_t = x$:

$$q_\tau(x) = \arg\min_{a \in R^1} E\left\{\rho_\tau(Y_i - a)|X_i = x\right\}, \tag{2.3}$$

where $\rho_\tau(y) = y(\tau - I_{\{y<0\}})$ with $y \in \mathbb{R}^1$ and I_A is the indicator function of set A. This *conditional quantile* is usually termed *quantile regression*, which under *i.i.d* samples was initially proposed by Koenker & Bassett Jr (1978, 1982). We will use the local linear estimation method.

The idea of the local linear fit (Fan & Gijbels, 1996; Loader, 1999) is to approximate the $q_\tau(z)$ in a neighborhood of x by a linear function

$$q_\tau(z) \approx q_\tau(x) + (\dot{q}_\tau(x))'(z - x) \equiv a_0 + a_1^{'}(z - x), \tag{2.4}$$

where $\dot{q}_\tau(x) = (\partial q_\tau(x)/\partial x_1, \cdots, \partial q_\tau(x)/\partial x_p)'$ is the vector of the first order partial derivatives of $q_\tau(x)$ at $x = (x_1, \cdots, x_p)' \in \mathbb{R}^p$. Locally, estimating $(q_\tau(x), \dot{q}_\tau(x))$ is equivalent to estimating (a_0, a_1). This motivates us to define an estimator by setting $\widehat{q}_\tau(x) \equiv \widehat{a}_0$ and $\widehat{\dot{q}}_\tau(x) \equiv \widehat{a}_1$. Then

$$\begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \end{pmatrix} = \arg\min_{(a_0, a_1)} \sum_{i=1}^{n} \rho_\tau(Y_i - a_0 - a_1'(X_i - x))K_h(X_i - x), \tag{2.5}$$

where $K_h(x) = h_n^{-p} K(x/h_n)$, K is a kernel function on $\mathbb{R}^p$, and $h_n > 0$ is the bandwidth. Note that if $q_\tau(x)$ has no derivative, $q_\tau(x)$ can still be estimated by (2.5), but the estimate of its derivative is vacuous.

To establish the asymptotic properties of quantile regression, we first give some definitions and assumptions about the dependence structure of the data-generating processes (DCP) $\{(X_t, Y_t)\}$. We consider the processes $\{(X_t, Y_t)\}$ that are near-epoch dependent on stationary α-mixing process $\{\varepsilon_t\}$. In the following, $|\cdot|$ and $\|\cdot\|$ denote the absolute value and the Euclidean norm of $\mathbb{R}^p$, respectively. $Y_t^{(m)} = \Psi_{Y,m}(\varepsilon_t, \ldots, \varepsilon_{t-m+1})$, $X_t^{(m)} = (X_{t1}^{(m)}, \cdots, X_{tp}^{(m)})^\tau = \Psi_{X,m}(\varepsilon_t, \ldots, \varepsilon_{t-m+1})$, and $\Psi_{Y,m}$ and $\Psi_{X,m}$ are $\mathbb{R}^1$- and $\mathbb{R}^p$-valued Borel measurable functions with m arguments, respectively. Let $\nu > 0$ be a positive real number.

Definition 1. *The stationary process $\{(Y_t, X_t)\}$ is said to be L_ν-NED on $\{\varepsilon_t\}$ if*

$$v_\nu(m) = E|Y_t - Y_t^{(m)}|^\nu + E\|X_t - X_t^{(m)}\|^\nu \to 0 \tag{2.6}$$

as $m \to \infty$. The $v_\nu(m)$ are called the stability coefficients of order ν of the process $\{(Y_t, X_t)\}$.

Definition 2. *Let $\mathcal{F}_{-\infty}^n$ and $\mathcal{F}_{n+k}^\infty$ be the σ–fields generated by $\{\varepsilon_t, t \leq n\}$ and $\{\varepsilon_t, t \geq n+k\}$, respectively. And let $\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |P(AB) - P(A)P(B)| \to 0$ as $k \to \infty$. The stationary sequence $\{\varepsilon_t, t = 0, \pm 1, \ldots\}$ is said to be α–mixing and the $\alpha(k)$ is termed mixing coefficient.*

Clearly, $\{(Y_t^{(m)}, X_t^{(m)})\}$ is an α-mixing process with mixing coefficient

$$\alpha_m(k) \leq \begin{cases} \alpha(k-m) & k \geq m+1, \\ 1 & k \leq m. \end{cases} \tag{2.7}$$

Throughout, we assume that the observations of the NED process $\{(Y_t, X_t)\}$ are (Y_t, X_t), $t = 1, 2, \ldots, n$. For the sake of convenience, we are summarizing here the main assumptions on the data generating process (DGP) (2.6), the kernel K and the bandwidth to be used in the estimation method. In what follows, we denote $F_{Y|X}(x) = P(Y_t < y|X_t = x)$ and $f_X(y|x)$ as the conditional distribution and the conditional density function of Y_t given $X_t = x$, respectively, and $f_X(x)$ as the marginal density function of X_t.

A1: (i) The DGP is a strictly stationary NED process L_1-NED on $\{\varepsilon_t\}$. For the mixing coefficient of ε_t, the function α is such that $\lim_{k\to\infty} k^a \sum_{j=k}^{\infty} \alpha(j) = 0$, for some positive real number a.

(ii) The marginal density function $f_X(\cdot)$ of X is continuous and $f_X(x) > 0$ at x.

(iii) The conditional density function $f_{Y|X}(y|\tilde{x})$ is continuous as a function of y in a neighborhood of $q_\tau(x)$ uniformly for $\tilde{x}$ in a neighborhood of x, and continuous as a function of $\tilde{x}$ in a neighborhood of x for all y in a neighborhood of $q_\tau(x)$. Also, $f_{Y|X}(q_\tau(x)|x) > 0$.

(iv) For all i and j in $\mathbb{Z}$, the vectors X_i and X_j admit a joint density (X_i, X_j) at $(x, \tilde{x})$; moreover,

$$\sup_{i,j} \sup_{(x,\tilde{x})} f_{ij}(x, \tilde{x}) \le C, \tag{2.8}$$

where C is a generic positive constant.

A2. The kernel function K: $\mathbb{R}^p \to \mathbb{R}$ satisfies:

(i) $K(\cdot)$ is a bounded and symmetric density kernel function such that $\int_{\mathbb{R}^p} K(u)du = 1$, $\int_{\mathbb{R}^p} uK(u)du = 0$ and $\mu_2 = \int_{\mathbb{R}^p} uu'K(u)du > 0$ (positive definite).

(ii) For any $c := (c_0, c_1^\tau)^\tau \in \mathbb{R}^{p+1}$, the function $K_c(u) := (c_0 + c_1^\tau u)K(u)$ satisfies: $\sup_{u\in\mathbb{R}^p} \|K_c(u)\| \le K_c^+$, and $\int_{\mathbb{R}^p} |K_c(x)|dx < \infty$.

(iii) For any $c \in \mathbb{R}^{p+1}$, $|K_c(u) - K_c(v)| \le C\|u - v\|$ for any $u, v \in \mathbb{R}^p$ *and* $C < \infty$.

A3. The quantile function $q_\tau(\cdot)$ is twice continuously differentiable. The ϑ-th order derivative $q_\tau^{(\vartheta)}(\cdot)$ of the quantile function $q_\tau(\cdot)$ satisfies the Lipschitz condition of degree δ $(0 < \delta \le 1)$, such that

$$\|q_\tau^{(\vartheta)}(x) - q_\tau^{(\vartheta)}(\tilde{x})\| \le C\|x - \tilde{x}\|^\delta, \text{ for any } x, \tilde{x} \in R^p, \tag{2.9}$$

where $q_\tau^{(0)}(x) = q_\tau(x)$, $q_\tau^{(1)}(x) = \dot{q}_\tau(x)$, and $\|\cdot\|$ is the Euclidean norm.

Remark. Assumption **A1** suppose that the DGP $\{(Y_t, X_t)\}$ is L_1-NED, which is easily satisfied for econometric time series models and has larger fields than mixing random fields considered by Honda (2000). Different to Lu & Linton (2007) and Jenish (2012), who require uniform $L_{2+\delta}$ boundedness of DGP, we just need L_1-NED processes. **A2** is standard and mild in the nonparametric literature and similar to assumptions of Lu & Linton (2007) and Jenish (2012). **A3** describes the properties of quantile function $q_\tau(\cdot)$, which is used to establish the convergence rates.

B1.

$$h_n \to 0 \quad and \quad nh_n^{(1+2/a)p} \to \infty, \quad as \quad n \to \infty, \tag{2.10}$$

where a is the positive constant defined in **A1**(i).

B2. There is two positive integers $m = m_n \to \infty$ and $L = L_n = \sqrt{v_1(m)} \to 0$ such that the stability coefficients, defined in (2.3) with $\nu = 1$, satisfy

$$h_n^{-1-p} v_1(m) \to 0, \quad n^6 v_1(m) \to 0, \quad as \quad n \to \infty. \tag{2.11}$$

B3. There exist two sequences of positive integer vectors, $p^* = {p^*}_n \in \mathbb{Z}$ and $q^* = q_n^* = 2m_n \in \mathbb{Z}$, with $m = m_n \to \infty$ such that $p^* = p_n^* = o((\, nh_n^{p^*})^{1/2})$, $q^*/p^* \to 0$ and $n/p^* \to \infty$, and $\dfrac{n}{p^*}\alpha(m) \to 0$.

B4. h_n tends to zero in such a manner that $q^* h_n^p = O(1)$ such that

$$h_n^p \sum_{t=q}^{\infty} \alpha_m(t) \to 0 \quad \text{as } T \to \infty.$$

Remark. Assumption **B1** is standard on the bandwidth and same as in the α-mixing case; while **B2** is concerned with the conditions on the stability coefficients related to the bandwidth and sample size; and **B3** and **B4** describe the mixing coefficients associated with the bandwidth, which is similar to the assumptions of Lu & Linton (2007). Assumptions **B2**-**B4** specify the restrictions on the decay rates of the stability and mixing coefficients for a given bandwidth.

3. Asymptotic behaviours

In this section, we establish the asymptotic normality of the local linear quantile regression estimators under near-epoch dependence. To this purpose, we need to develop a powerful tool of Bahadur representation in weak convergence sense for the quantile regression estimators which are not of analytical expression as the mean regression.

3.1. Bahadur representation

In this subsection, we consider the weak conditions to ensure the Bahadur representation of the local linear estimators of $q_\tau(\cdot)$ and its derivatives. If $q_\tau(x)$ is first order differentiable, then its derivatives can be estimated reasonably well by the local linear fitting. We first introduce a notation. Let $Z_n = o_P(1)$ represent the random sequence $Z_n \xrightarrow{P} 0$, where $\xrightarrow{P}$ denotes convergence in probability.

Theorem 3.1 *Assume that Assumptions **A1**, **A2** and **A3** are satisfied for some $a \geq 1$, and that the quantile function $q_\tau(x)$ is twice continuously differentiable at x. Then*

$$\sqrt{nh_n^p}\begin{pmatrix} (\widehat{q}_\tau(x) - q_\tau(x)) \\ h_n\left(\widehat{\dot{q}}_\tau(x) - \dot{q}_\tau(x)\right)\end{pmatrix} = \phi_\tau(x)\frac{1}{\sqrt{nh_n^p}}\sum_{i=1}^{n}\psi_\tau(Y_i^*)\begin{pmatrix} 1 \\ \frac{X_i - x}{h_n}\end{pmatrix} K(\frac{X_i - x}{h_n}) + o_P(1), \tag{3.1}$$

as $n \to \infty$, where $\psi_\tau(y) = \tau - \boldsymbol{I}_{\{y<0\}}$, $Y_i^* = Y_i^*(\tau) = Y_i - q_\tau(x) - (\dot{q}_\tau(x))'(X_i - x)$ and $\phi_\tau(x) = (f_{Y|X}(q_\tau(x)|x)f_X(x))^{-1}$.

Remark. The proofs of theorem 3.1 are given in Appendix B.

If $q_\tau(x)$ has the first order derivatives which are Lipschitz continuous, then $q_\tau(x)$ and its derivatives can be estimated with optimal convergence rates of Stone (1980) as in the *i.i.d*. setting.

Theorem 3.2 *Under the conditions of Theorem 3.1, if **A3** with $v = 1$ is satisfied, then*

$$\widehat{q}_\tau(x) - q_\tau(x) = O_P(h_n^{1+\delta}) + O_P((nh_n^p)^{-1/2}), \tag{3.2}$$

and

$$\widehat{\dot{q}}_\tau(x) - \dot{q}_\tau(x) = O_P(h_n^{\delta}) + O_P((nh_n^{p+2})^{-1/2}). \tag{3.3}$$

*Furthermore, if in **A1**(i) $a > p/(v+\delta)$ and $h_n = n^{-1/[p+2(1+\delta)]}$, then*

$$\widehat{q}_\tau(x) - q_\tau(x) = O_P(n^{-(1+\delta)/[p+2(1+\delta)]}), \tag{3.4}$$

and

$$\widehat{\dot{q}}_\tau(x) - \dot{q}_\tau(x) = O_P(n^{-\delta/[p+2(1+\delta)]}) \tag{3.5}$$

as $n \to \infty$.

Remark. The condition $a > p/(z+\delta)$ in the theorem is to ensure that the requirement in Assumption **B1** can be satisfied with the optimal bandwidth, $h_n = n^{-1/[p+2(1+\delta)]}$.

3.2. Asymptotic normality

Based on the powerful tool of the weak Bahadur representation, we can establish the asymptotic distribution of the local linear quantile regression estimates under near-epoch dependence. Toward the asymptotic normality result, we prove the following lemmata. Suppose

$$W_n := \begin{pmatrix} w_{n0} \\ w_{n1} \end{pmatrix}, \quad (W_n)_j := (nh_n^p)^{-1} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{X_i - x}{h_n}\right)_j K\left(\frac{X_i - x}{h_n}\right), \quad j = 0, \ldots, p,$$

with $\left(\frac{X_i - x}{h_n}\right)_0 = 1$.

The usual Cramér-Wold device will be adopted. For all $c := (c_0, c_1')' \in \mathbb{R}^{1+p}$, let

$$A_n := (\ nh_n^p)^{1/2} c' W_n = \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^n \psi_\tau(Y_i^*) K_c(\frac{X_i - x}{h_n}),$$

with $K_c(u)$ defined in **A2**. The following lemma provides the expectation and asymptotic variance of A_n for all c.

Lemma 3.1 *Assume that Assumptions **A1** and **A2** hold, and that h_n satisfies Assumptions **B1** – **B2**. The expectation is as*

$$E\left[\phi_\tau(x) A_n\right] = \sqrt{nh_n^p} \left[(1 + o(1)) \begin{pmatrix} B_0(x) \\ B_1(x) \end{pmatrix} \right],$$

where $B_0(x) = \frac{1}{2} f_X^{-1}(x) tr[\ddot{q}_\tau(x) \int uu'K(u)du] h_n^2$, and $B_1(x) = (B_{11}(x), ..., B_{1p}(x))'$, $B_{1j}(x) = \frac{1}{2} f_X^{-1}(x) tr[\ddot{q}_\tau(x) \int uu'u_j K(u)du] h_n$, $\quad j = 1, ..., p$. The asymptotic variance is as

$$\lim_{n\to\infty} \text{Var}[\phi_\tau(x) A_n] = c'\Sigma c,$$

where

$$\Sigma := \phi_\tau^2(x)\tau(1-\tau)f_X(x)\begin{pmatrix}\int K^2(u)du & \int u'K^2(u)du\\ \int uK^2(u)du & \int uu'K^2(u)du\end{pmatrix}$$

Lemma 3.2 *Suppose that Assumptions in Lemma 3.1 hold. Denote by σ^2 the asymptotic variance of A_n. Then $(nh_n^p)^{1/2}(c'[W_n(x)-\mathrm{E}W_n(x)]/\sigma)$ is asymptotically standard normal as $n\to\infty$.*

The proof of Lemma 3.1-3.2 will be given in Appendix C. Based on these lemmata, we can show the main consistency and asymptotic normality result of the local linear quantile estimator.

Theorem 3.3 *If **A3** with $v=2$ is satisfied, $a>p/(1+\delta)$ in **A1**(i) and $nh_n^{p+2(1+\delta)}\to 0$, then for any $0<\tau_1,\tau_2<1$,*

$$\sqrt{nh_n^p}\begin{pmatrix}\hat{q}_{\tau_1}(x)-q_{\tau_1}(x)-B_1(x)\\ \hat{q}_{\tau_2}(x)-q_{\tau_2}(x)-B_2(x)\end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\begin{pmatrix}\sigma_{000}^2(x) & \sigma_{001}^2(x)\\ \sigma_{010}^2(x) & \sigma_{011}^2(x)\end{pmatrix}\right), \tag{3.6}$$

$$\sqrt{nh_n^{p+2}}\begin{pmatrix}\hat{\dot{q}}_{\tau_1}(x)-\dot{q}_{\tau_1}(x)\\ \hat{\dot{q}}_{\tau_2}(x)-\dot{q}_{\tau_2}(x)\end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\begin{pmatrix}\sigma_{100}^2(x) & \sigma_{101}^2(x)\\ \sigma_{110}^2(x) & \sigma_{111}^2(x)\end{pmatrix}\right), \tag{3.7}$$

as $n\to\infty$. Here $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution,

$B_i(x)=\frac{h_n^2}{2}f_X^{-1}(x)tr[\ddot{q}_{\tau_i}(x)\int uu'K(u)du]$, $i=1,2$, $\sigma_{000}^2(x)=\phi_1^*(x)\int K^2(u)du$, $\sigma_{011}^2(x)=\phi_2^*(x)\int K^2(u)du$, $\sigma_{001}^2(x)=\sigma_{010}^2(x)=\phi_3^*(x)\int K^2(u)du$, $\sigma_{100}^2(x)=\phi_1^*(x)\int uu'K^2(u)du$, $\sigma_{111}^2(x)=\phi_2^*(x)\int uu'K^2(u)du$, $\sigma_{101}^2(x)=\sigma_{110}^2(x)=\phi_3^*(x)\int uu'K^2(u)du$,

and where

$$\phi_i^*(x)=\phi_{\tau_i}^2(x)\tau_i(1-\tau_i)f_X(x)=\frac{\tau_i(1-\tau_i)}{f_X(x)f_{Y|X}^2(q_{\tau_i}(x)|x)},\quad i=1,2$$

$$\phi_3^*(x)=\phi_{\tau_1}(x)\phi_{\tau_2}(x)(\min(\tau_1,\tau_2)-\tau_1\tau_2)f_X(x)=\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2}{f_X(x)f_{Y|X}(q_{\tau_1}(x)|x)f_{Y|X}(q_{\tau_2}(x)|x)}$$

Remark. The proof of Theorem 3.3 is relegated to the Appendix C.

4. Numerical results

4.1. Simulation

In this section, we will show the results of a small Monte Carlo study of the finite sample performance of median regression and mean regression under near epoch dependence. The purpose is to illustrate that local linear median regression with a bandwidth choice pointed by Masry & Fan (1997), which is under α-mixing, can work more efficient and robust for the processes under near-epoch dependence.

We considered the following model(Reference to Lee, 2003):

$$Y_t = g(X_t) + a_3\xi_t, \tag{4.1}$$

where ξ_t was drawn from the Student t-distribution with the 2 degrees of freedom (denoted by t_2) and $a_3 = 0.2$. The Student t-distribution is of great interest in financial modeling of market volatility(c.f., Bollerslev et al., 1992).In particular the variance of the time series may not exist if the ξ_t is heavily tailed enough (e.g., t_2 distribution).

$$g(x) = a_0x + a_1sin(a_2x), \tag{4.2}$$

where $a_0 = a_1 = 1, a_2 = 4$. And X_t was drawn from the ARMA(1,1)-GARCH(1,1) model, which is as follows:

$$X_t = \mu + \phi X_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \tag{4.3}$$

$$\varepsilon_t = e_t h_t^{1/2}, \qquad h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{4.4}$$

where X_t is modeled by an ARMA(1,1) model (4.3), and h_t is the conditional variance of X_t, given the past information up to day $t-1$, modeled by a GARCH(1,1) model (4.4), with $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\{e_t\}$ being $i.i.d.$ random sequence with $Ee_t = 0$ and $Ee_t^2 = 1$ (taken to be standard normally distributed in this example). Referring to the estimations of the model (4.3) with (4.4) in section 4.2, we take the parameters $\mu = -0.0177$, $\phi = 0.6057$, $\theta = -0.6263$, $\alpha_0 = 0.3002$, $\alpha_1 = 0.1078$, and $\beta_1 = 0.8518$, which are the obtained from the real oil price return data by the maximum likelihood method procedure in the ARMA-GARCH module of R.

If $|a| < 1$, the model (4.3) can be expressed as (Lu & Linton, 2007)

$$X_t = \mu/(1-\phi) + \theta\varepsilon_{t-1} + \varepsilon_t + \sum_{j=1}^{\infty} \phi^j (\theta\varepsilon_{t-j-1} + \varepsilon_{t-j}),$$

and under $\alpha_1 + \beta_1 < 1$ with some suitably regular conditions (c.f. Carrasco & Chen, 2002), the ε_t in the GARCH(1,1) model (4.4) is α- mixing with a geometrically decaying mixing coefficient. Here it is difficult to show under such natural and mild conditions $|a| < 1$ and $\alpha_1 + \beta_1 < 1$ (to the best of our knowledge) that X_t is α- mixing, but it can be shown that X_t is NED of order 1 with regard to a α- mixing process, if $E|\varepsilon_t| < \infty$, with stable coefficients (owing to the convex property of $|\cdot|$)

$$\begin{aligned} v_1(k) = E\left|X_t - X_t^{(k)}\right| &= w_k E\left|\sum_{j=k+1}^{\infty} \frac{\phi^j}{w_k}\left(\theta\varepsilon_{t-j-1} + \varepsilon_{t-j}\right)\right| \\ &\le w_k E\left[\sum_{j=k+1}^{\infty} \frac{\phi^j}{w_k}|\left(\theta\varepsilon_{t-j-1} + \varepsilon_{t-j}\right)|\right] = O\left(|\phi|^k\right), \end{aligned}$$

decaying at a geometric rate, where $X_t^{(k)} = \mu/(1-\phi) + \theta\varepsilon_{t-1} + \varepsilon_t + \sum_{j=1}^{k} \phi^j (\theta\varepsilon_{t-j-1} + \varepsilon_{t-j})$, and $w_k = \sum_{j=k+1}^{\infty} \phi^j = O(\phi^k)$. $E|X_t| < \infty$ can be ensured by $E|\varepsilon_t| < \infty$, which Carrasco & Chen (2002) shows the conditions. Therefore, (X_t, Y_t) is a stationary NED of order 1 w.r.t. a strongly (α-) mixing process.

We generate data X_t from the model (4.3) and (4.4), and Y_t from model (4.1) with (4.2), denoted as $\{(X_t, Y_t)\}$ for $t = 1, ..., n$. We consider two time series of sample size: $n = 100$ and $n = 500$. We are assessing the estimate of the median regression $q_{0.5}(x) = \text{argmin}_{a\in R^1} E\{\rho_{0.5}(Y_i - a)|X_i = x\}$ and mean regression $m(x) = E(Y_t|X_t = x)$, which are all equal to $g(x)$. We partition 50 points of x between the -2.5 and 2.5, which are approximately 10th and 90th percentiles of X_t.In the simulations, the bandwidth b_T was chosen by the conventional cross-validation rule of Stone (1984).

Figure 1 shows that the biases of 100 replications of median and mean estimates of $g(\cdot)$ for sample size 100 and 500 are acceptable. Figure 2 displays the the results of local linear estimators of median regression $q_{0.5}(x) = g(x)$ and mean regression function $m(x) = g(x)$ in 50 points of x, based on 100 replications with each sample size $n = 100$ and $n = 500$. Figure 3 assesses the accuracy of estimation by defining a squared estimation error (SEE)

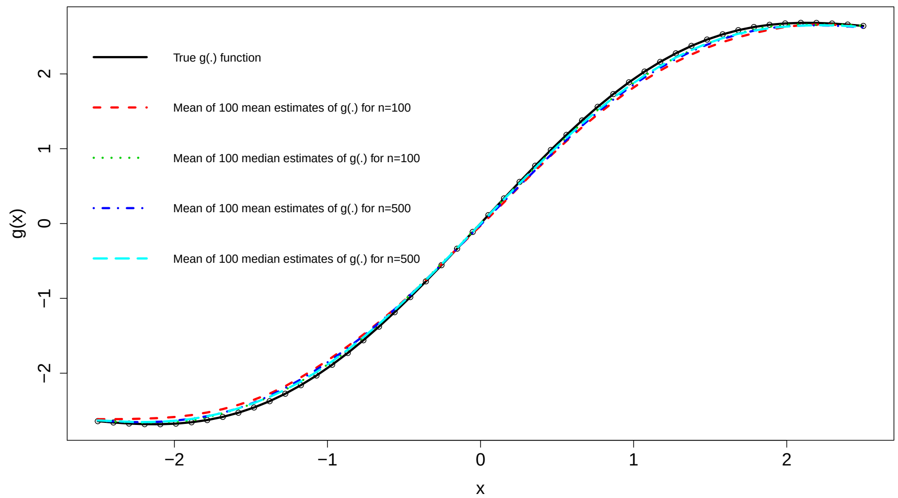


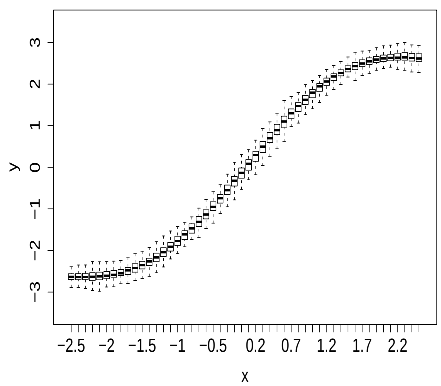
Figure 1: *Comparison of mean of 100 median and mean estimates of $g(\cdot)$ for sample size 100 and 500*

of mean regression and median regression estimation about $g(\cdot)$.

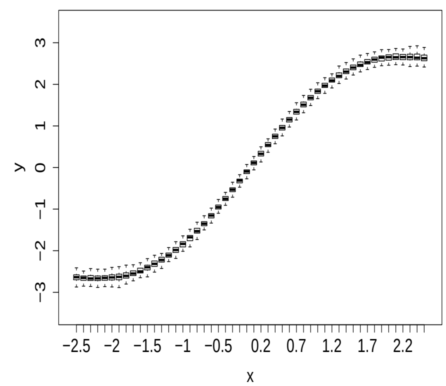
$$SEE(\widehat{q}_{0.5}(\cdot)) = \frac{1}{50}\sum_{k=1}^{50}\left\{\widehat{q}_{0.5}^{(r)}(x_k) - g(x_k)\right\}^2$$

$$SEE(\widehat{m}(\cdot)) = \frac{1}{50}\sum_{k=1}^{50}\left\{\widehat{m}^{(r)}(x_k) - g(x_k)\right\}^2$$

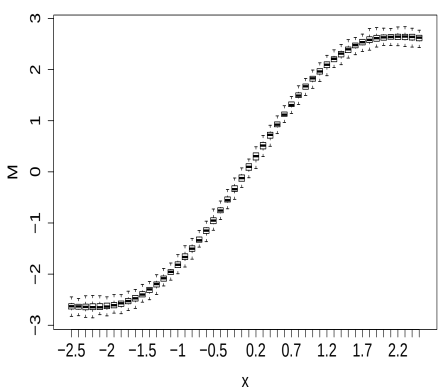
where $r = 1, ..., 100$ is the simulation times, for the sample size $n = 100$ and $n = 500$. Overall, the simulation results of Model (4.1) adapt very well to asymptotic theory: with the sample size increasing, the both mean and median regressions with a cross validation method for bandwidth selection become more stable and fit better to the true $g(\cdot)$ function, and even for the sample size of 100, the estimate procedure and bandwidth selection looks acceptable with the median regression. Clearly, the median regression has better performance in estimated results than the mean regression, and the median regression in sample size 500 works very well in all cases.



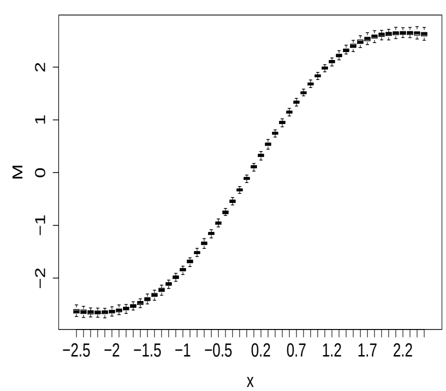
(a) mean regression, $n = 100$



(b) median regression, $n = 100$

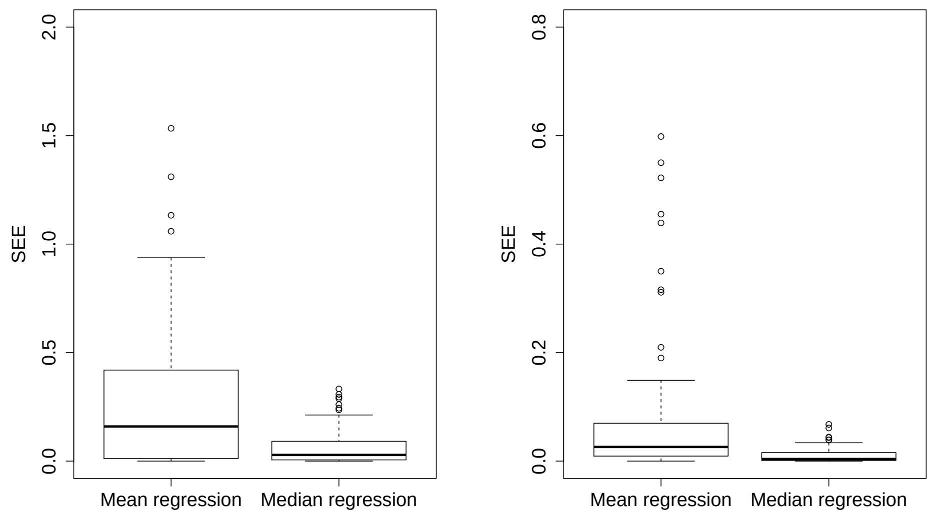


(c) mean regression, $n = 500$



(d) median regression, $n = 500$

Figure 2: *Simulation results-Boxplots of the local linear fitting for the median regression and mean regression, for* $n = 100$*(top) and* $n = 500$ *(bottom)*



(a) sample size=100 (b) sample size=500

Figure 3: *Simulation results-Boxplots of squared estimation error(SEE) of mean regression and median regression estimation of* $g(\cdot)$*, for* $n = 100$ *(left) and* $n = 500$ *(right)*

4.2. An Empirical Application

Climate change has received a great deal of attention in recent years and become one of the world's supreme policy challenges. In order to promote greener growth and internalise the costs of future environmental damage, the economists advise on the appropriate design of a price on the thing that causes it – namely carbon emissions. The European Union Emissions Trading System (EU ETS), launched in 2005, was the first large, and remains the biggest greenhouse gas emissions trading scheme in the world.

It is well known that oil is one of the biggest sources of world greenhouse gases. Low oil prices could discourage further innovation in and adoption of cleaner energy technologies, which result in higher emissions of carbon dioxide and other greenhouse gases (Balaguer & Cantavella, 2016). Through corrective carbon pricing, governments could restore appropriate price incentives, and lower the risk of irreversible and potentially devastating effects of climate change. Correspondingly, the interaction betweeen carbon price and oil price has been increasingly closer, due to the rapid development and steady expansion of carbon market. The relationship oil price and carbon market has drawn attention in many recent studies (Benz & Trück, 2009; Chevallier, 2011; Mansanet-Bataller et al., 2007). To capture well the underlying impact of oil price on the carbon market, we investigate the relationship between daily return of WTI crude oil price and carbon future price [1], with sample size 454 from 27th July 2015 to 15th May 2017, for an illustration.

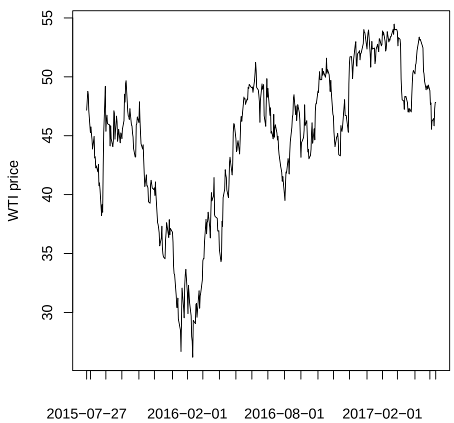
In Figure 4,(a) and (b) show WTI crude oil price(W_t) and carbon future price(C_t), and the daily return of WTI crude oil price X_t and daily return of carbon future price Y_t, defined by

$$X_t = \log(W_t/W_{t-1}) \times 100, \quad Y_t = \log(C_t/C_{t-1}) \times 100,$$

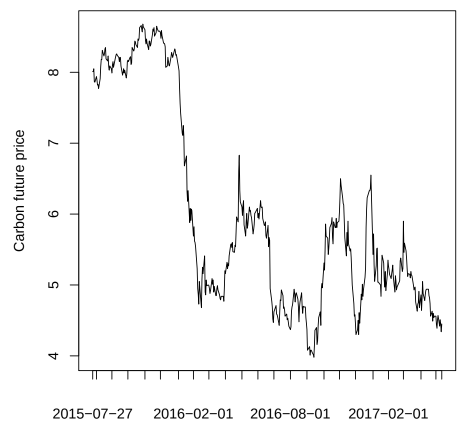
are plotted in (c) and (d), respectively.

ARMA-GARCH models have been used to describe the oil prices in many studies (Chang et al., 2010; Lee & Chiou, 2011; Zhang & Chen, 2011, 2014). Therefore, we used the daily return of WTI crude oil price data to estimate the model (4.3) with (4.4) and check

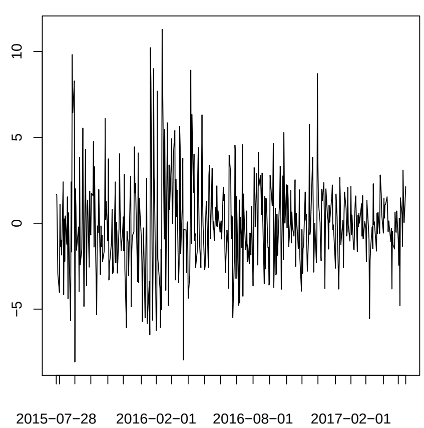
[1] Historical Futures Prices: ECX EUA Futures, Continuous Contract # 1. Non-adjusted price based on spot-month continuous contract calculations. Raw data from ICE.



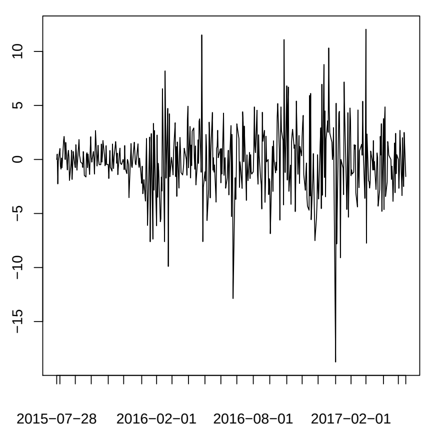
(a) WTI price



(b) Carbon future price



(c) Daily return of WTI crude oil price



(d) Daily return of carbon future price

Figure 4: *Real example – WTI crude oil daily price and carbon future daily price*

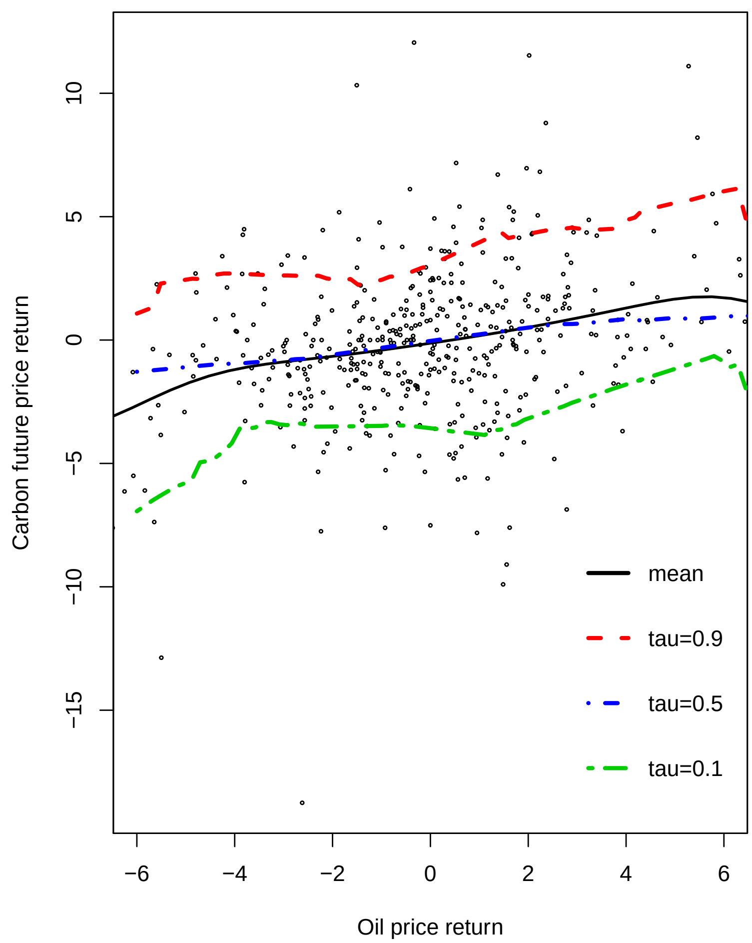


Figure 5: *The local linear fitting for the conditional mean and conditional quantile of the relationship between daily return of WTI crude oil price and carbon future price, from 27th July 2015 to 15th May 2017 with 454 observations.*

Table 1: *Estimation results of ARMA-GARCH*

Coefficients	Estimate	Std. Error	t value	$Pr(>\|t\|)$
μ	-0.0177	0.1208	-0.1462	0.8838
ϕ	0.6057	0.2222	2.7259	0.0064
θ	-0.6263	0.2230	-2.8087	0.0050
α_0	0.3002	0.1449	2.0713	0.0383
α_1	0.1078	0.0442	2.4403	0.0147
β_1	0.8518	0.0489	17.4335	0.0000

whether the conditional variance follows the GARCH process. The estimation results show in Table 1, which all the coefficients are statistically significant except the intercept.

Table 2: *Estimation results of mean squared prediction error for two methods*

Methods	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
parametirc quantile regression	0.3937184	0.97913	0.4405764
local linear quantile regression	0.3806771	0.98394	0.4240236
nonlinear threshold quantile regression	0.3511665	0.96354	0.4210472

The local linear estimates of the conditional mean and the conditional quantiles based on the asymptotic normality in Section 3 are plotted in Figure 5, where the bandwidths used for the conditional mean and conditional quantiles are 1.85 and 1.370132 ($\tau = 0.9$), 2.907967 ($\tau = 0.5$), and 1.185506 ($\tau = 0.1$), respectively, chosen by cross-validation rule. From Figure 5, we can observe that both the conditional upper and lower conditional quantiles (e.g. $\tau = 0.1$ and $\tau = 0.9$) functions appear to be nonlinear.

To further evaluate the local linear quantile method, we consider comparison of the prediction based on different parametric forms of quantile regression. The first is a linear quantile function $Q^0_{Y_t}(\tau|X_t) = a^0_\tau + \alpha^0_\tau X_t$, where a^0_τ and α^0_τ are linear quantile coefficients (Koenker, 2005). Then, we consider the local linear quantile regression. In general, although nonparametric specification can help to explore the possibly nonlinear relationship, it may not give optimal prediction. Therefore, we consider a nonlinear threshold quantile function based on Figure 5.

We set aside the last 50 quarters for prediction and use the first T=404 quarters for model estimation. A quantile prediction error (QPE) of the one-step ahead prediction is computed for the linear quantile regression and nonlinear threshold quantile regression at 10th, 50th and 90th quantile levels. Here, we define QPE as:

$$QPE(\hat{q}_\tau(\cdot)) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\hat{q}_\tau(x_i) - q_\tau(x_i)), \tag{4.5}$$

where $\rho_\tau(y) = y(\tau - I_{\{y<0\}})$ and I_A is the indicator function of set A.

The results show in Table 2. The QPE values are 0.3937184 and 0.3511665 for linear and nonlinear threshold quantile regression at 10th quantile level, respectively. Compared with the linear quantile function, the threshold quantile regression outperforms in prediction, with a relative improvement of 11%. This result further illustrates that local linear quantile regression can help to uncover the relationship between daily returns of WTI crude oil price and carbon future price which is more complex than linear.

References

Andrews, D. W. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, *11*, 560–586.

Balaguer, J., & Cantavella, M. (2016). Estimating the environmental kuznets curve for spain by considering fuel oil prices (1874–2011). *Ecological Indicators*, *60*, 853–859.

Benz, E., & Trück, S. (2009). Modeling the price dynamics of co 2 emission allowances. *Energy Economics*, *31*, 4–15.

Bickel, P. J. (1975). One-step huber estimates in the linear model. *Journal of the American Statistical Association*, *70*, 428–434.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, *52*, 5–59.

Cai, Z. (2002). Regression quantiles for time series. *Econometric theory*, *18*, 169–192.

Cai, Z., & Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, *167*, 413–425.

Cai, Z., & Xu, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, *104*, 371–383.

Carrasco, M., & Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, *18*, 17–39.

Chang, C.-L., McAleer, M., & Tansuchat, R. (2010). Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets. *Energy Economics*, *32*, 1445–1455.

Chevallier, J. (2011). A model of carbon price interactions with macroeconomic and energy dynamics. *Energy Economics*, *33*, 1295–1312.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, *1*, 93–125.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability* volume 66. CRC Press.

Fan, J., & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Gallant, A. R. (Ed.) (2008). *Nonlinear Statistical Models*. John Wiley & Sons, Inc. doi:10.1002/9780470316719.

Granger, C. W. J., White, H., & Kamstra, M. (1989). Interval forecasting: an analysis based upon arch-quantile estimators. *Journal of Econometrics*, *40*, 87–96.

Hall, P., & Heyde, C. C. (2014). *Martingale limit theory and its application*. Academic press.

Hall, P., Wolff, R. C., & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, *94*, 154–163.

Hallin, M., Lu, Z., & Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli*, *15*, 659–686.

Härdle, W., Lütkepohl, H., & Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, *65*, 49–72.

He, X., & Shao, Q.-M. (1996). A general bahadur representation of m -estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, *24*, 2608–2630. doi:10.1214/aos/1032181172.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Annals of the Institute of Statistical Mathematics*, *52*, 459–470.

Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of statistical planning and inference*, *121*, 113–125.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, *7*, 349–382.

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics*, *167*, 224–239.

Jorion, P. (1997). *Value at risk*. McGraw-Hill, New York.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, *35*, 92–108. doi:10.1214/009053606000000966.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. doi:10.1017/cbo9780511754098.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, *50*, 33–50.

Koenker, R., & Bassett Jr, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, *50*, 43–61.

Koenker, R., & Zhao, Q. (1996). Conditional quantile estimation and inference for arch models. *Econometric Theory*, *12*, 793–813.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*, 53–89.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric theory*, *19*, 1–31.

Lee, Y.-H., & Chiou, J.-S. (2011). Oil sensitivity and its asymmetric impact on the stock market. *Energy*, *36*, 168–174.

Li, D., Lu, Z., & Linton, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory*, *28*, 935–958.

Loader, C. R. (1999). Bandwidth selection: classical or plug-in? *The Annals of Statistics*, *27*, 415–438. doi:10.1214/aos/1018031201.

Lu, Z. (2001). Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics*, *53*, 447–468.

Lu, Z., & Linton, O. (2007). Local linear fitting under near epoch dependence. *Econometric Theory*, *23*, 37–70.

Mansanet-Bataller, M., Pardo, A., & Valor, E. (2007). CO2 prices, energy and weather. *The Energy Journal*, *28*. doi:10.5547/issn0195-6574-ej-vol28-no3-5.

Masry, E., & Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, *24*, 165–179.

Oberhofer, W., & Haupt, H. (2016). Asymptotic theory for nonlinear quantile regression under weak dependence. *Econometric Theory*, *32*, 686–713.

Ruppert, D., & Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, *75*, 828–838.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, *8*, 1348–1360. doi:10.1214/aos/1176345206.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, *12*, 1285–1297. doi:10.1214/aos/1176346792.

Tjøstheim, D., & Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association*, *89*, 1410–1419.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

Wang, H. J., Zhu, Z., & Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, *37*, 3841–3866. doi:10.1214/09-aos695.

Welsh, A. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica*, *6*, 347–366.

Wu, W., & Zhou, Z. (2017). Nonparametric inference for time-varying coefficient quantile regression. *Journal of Business & Economic Statistics*, *35*, 98–109.

Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, *93*, 228–237.

Zhang, C., & Chen, X. (2011). The impact of global oil price shocks on china's stock returns: Evidence from the arji (-ht)-egarch model. *Energy*, *36*, 6627–6633.

Zhang, C., & Chen, X. (2014). The impact of global oil price shocks on china's bulk commodity markets and fundamental industries. *Energy policy*, *66*, 32–41.

. Appendix: Proofs

A: Proofs of lemmata

Some basic lemmata are given in this section for later reference.

Lemma A.1 *Let X and Y are two random variables on two σ–algebras generated fields $\mathcal{A}$ and $\mathcal{B}$, respectively. For two constants C_1 and C_2, there exist $|X| \leq C_1$ and $|Y| \leq C_2$, then*

$$|EXY - EXEY| \leq 4C_1C_2\alpha(\mathcal{A}, \mathcal{B}), \tag{A.1}$$

and

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |P(AB) - P(A)P(B)|.$$

Proof. *See the appendix of Hall & Heyde (2014).*

Lemma A.2 *Let $m = m_n$ be a positive integer and tending to $m_n \to \infty$ as $n \to \infty$. $b(\cdot)$: $\mathbb{R}^p \to \mathbb{R}^1$ is continuous function. Then under Assumptions **A1**(ii,iv) and **A2**, if $E|b(X_i^{(m)})| < \infty$ and $h_n \to 0$ as $n \to \infty$,*

$$h_n^{-p}Eb(X_i^{(m)})K((x - X_i^{(m)})/h_n) \to b(x)f_X(x)\int K(u)du + O\left(h_n^{-1-p}v_1(m)\right), \tag{A.2a}$$

$$h_n^{-p}Eb(X_i^{(m)})K^2((x - X_i^{(m)})/h_n) \to b(x)f_X(x)\int K^2(u)du + O\left(h_n^{-1-p}v_1(m)\right), \tag{A.2b}$$

as $n \to \infty$. for $i \neq j$,

$$\begin{aligned} &Eb(X_i^{(m)})b(X_j^{(m)})K((x - X_i^{(m)})/h_n)K((x - X_j^{(m)})/h_n) \\ &\quad = Eb(X_i)b(X_j)K((x - X_i)/h_n)K((x - X_j)/h_n) + O\left(h_n^{-1-p}v_1(m)\right) \end{aligned} \tag{A.2c}$$

Furthermore, for $j > 0$,

$$h_n^{-p}EK((x - X_1^{(m)})/h_n)K((x - X_{j+1}^{(m)})/h_n) = O(h_n^{\min(p,j)}) + O\left(h_n^{-1-p}v_1(m)\right), \tag{A.2d}$$

where (A.2c) holds uniformly for $j \geq p$.

Proof. *The main idea of the proof is similar to that of Lemma A.2 of Lu & Linton (2007), though details are different. We only briefly sketch it here.*

To prove (A.2a), note first that

$$\begin{aligned}
&Eb(X_i^{(m)})K((x-X_i^{(m)})/h_n)\\
=&Eb(X_i)K((x-X_i^{(m)})/h_n)+E(b(X_i^{(m)})-b(X_i))K((x-X_i^{(m)})/h_n)\\
=:&\ Eb(X_i)K((x-X_i^{(m)})/h_n)+\delta_{1T}. \qquad \text{(A.3)}
\end{aligned}$$

Here, using the bounded property of the kernel function $K(\cdot)$,

$$\begin{aligned}
|\delta_{1T}| \leq& E|b(X_i^{(m)})-b(X_i))|\, K((x-X_i^{(m)})/h_n)\\
\leq& CE|b(X_i^{(m)})-b(X_i))| \leq CE||X_i^{(m)}-X_i|| = O\left(v_1(m)\right).
\end{aligned}$$

Next,

$$\begin{aligned}
&Eb(X_i)K((x-X_i^{(m)})/h_n)\\
&= Eb(X_i)K((x-X_i)/h_n)+Eb(X_i)(K((x-X_i^{(m)})/h_n)-K((x-X_i)/h_n))\\
&=: \delta_{2T}+\delta_{3T},
\end{aligned}$$

$$\delta_{2T} = Eb(X_i)K((x-X_i)/h_n) = h_n^p b(x) f_X(x)\int K(u)du; \qquad \text{(A.4)}$$

*under **A2**(iii),*

$$\begin{aligned}
|\delta_{3T}| \leq& E|b(X_i)|\, |K((x-X_i^{(m)})/h_n)-K((x-X_i)/h_n)|\\
\leq& CE\left\|\frac{X_i^{(m)}-X_i}{h_n}\right\| = O\left(h_n^{-1}v_1(m)\right); \qquad \text{(A.5)}
\end{aligned}$$

Then we can get

$$Eb(X_i)K((x-X_i^{(m)})/h_n) = h_n^p b(x) f_X(x)\int K(u)du + O\left(h_n^{-1}v_1(m)\right). \qquad \text{(A.6)}$$

For (A.2b), (A.2c) and (A.2d), it can be proved in an argument similar to that in the above.

Lemma A.3 (Cross-term lemma) *Let $\{(Y_j^{(m)}, X_j^{(m)}); 1 \leq j \leq q\}$ be a stationary sequence with mixing coefficient*

$$\alpha_m(j) := \sup\left\{|\mathrm{P}(AB)-\mathrm{P}(A)\mathrm{P}(B)| :\ A \in \mathcal{B}(\{Y_i^{(m)}, X_i^{(m)}\}),\ B \in \mathcal{B}(\{Y_{i+j}^{(m)}, X_{i+j}^{(m)})\}\right\}.$$

Let $(y,x) \mapsto \tilde{b}(y,x)$ *be a bounded measurable function defined on* $\mathbb{R} \times \mathbb{R}^p$. *Set* $\eta_j^{(m)}(x) = \tilde{b}(Y_j^{(m)}, X_j^{(m)})\tilde{K}\left((x - X_j^{(m)})/h_n\right)$, *where* $\tilde{K}$ *is a kernel function satisfying Assumption* ***A2***, *and* $\Delta_j^{(m)}(x) = \eta_j^{(m)}(x) - E\eta_j^{(m)}(x)$, $\widetilde{R}(x) = \sum_{1\le i<j\le n} E\Delta_i^{(m)}(x)\Delta_j^{(m)}(x)$. *Then, under Assumptions* ***A1***, ***A2*** *and* ***A3***, *there exists a constant* $C > 0$ *such that*

$$|\widetilde{R}(x)| \le Cnh_n^p \left[\tilde{J}_1(x) + \tilde{J}_2(x)\right]. \tag{A.7}$$

where $\tilde{J}_1(x) := h_n^p N_n max\left\{1, h_n^{-2-2p} v_1(m)\right\}$ *and*

$$\tilde{J}_2(x) := h_n^p \left(\sum_{j=N_n}^{n} \alpha_m(j)\right) max\left\{1, h_n^{-2-2p} v_1(m)\right\}.$$

Proof. *The main idea of the proof is similar to that of Lemma A.3 of Lu & Linton (2007), though details are different. We only briefly sketch it here.*

$$\mathrm{E}\Delta_j^{(m)}(x)\Delta_i^{(m)}(x) = \left\{\mathrm{E}\left(\eta_j^{(m)}(x)\eta_i^{(m)}(x)\right) - \mathrm{E}\left(\eta_j^{(m)}(x)\right)\mathrm{E}\left(\eta_i^{(m)}(x)\right)\right\}$$

Then applying Lemma A.2,

$$\begin{aligned}
&\mathrm{E}\left[\Delta_j^{(m)}(x)\Delta_i^{(m)}(x)\right] \\
&= \left[\mathrm{E}Z_iK((x-X_i)/h_n)Z_jK((x-X_j)/h_n) - EZ_iK((x-X_i)/h_n)EZ_jK((x-X_j)/h_n)\right] \\
&\quad + \left[O\left(h_n^{-2}v_1^2(m)\right) + O\left(h_n^{-1}v_1(m)\right)\right] + h_n^p\left[O\left(h_n^{-1}v_1(m)\right)\right] \\
&\le Ch_n^{2p}\left[1 + O\left(h_n^{-2p-2}v_1(m)\right)\right],
\end{aligned} \tag{A.8}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j-i=1}^{N_n} \mathrm{E}\left[\Delta_j^{(m)}(x)\Delta_i^{(m)}(x)\right] &\le Ch_n^{2p}(nN_n)max\left\{1, h_n^{-2p-2}v_1(m)\right\} \\
&= nh_n^p\tilde{J}_1(x).
\end{aligned} \tag{A.9}$$

where N_n *is a positive integer depending on* n *to be specified later. On the other hand,*

$$\sum_{i=1}^{n}\sum_{j-i=N_n}^{n} \mathrm{E}\left[\Delta_j^{(m)}(x)\Delta_i^{(m)}(x)\right] \le Ch_n^{2p}\left(n\sum_{j=N_n}^{n}\alpha_m(j)\right) max\left\{1, h_n^{-2p-2}v_1(m)\right\}$$

$$= nh_n^p \tilde{J}_2(x). \tag{A.9}$$

B: Proofs for Subsection 3.1

In this section, we will mainly focus on the proof of Theorem 3.1. C will denote a generic constant in the proof.

We first introduce some notations. Denote $X_{hi} = (X_i - x)/h_n$, $\mathcal{X}_{hi} = (1, X'_{hi})'$, $K_i = K(X_{hi})$, $X_{hi}^{(m)} = (X_i^{(m)} - x)/h_n$, $\mathcal{X}_{hi}^{(m)} = (1, X_{hi}'^{(m)})'$, $K_i^{(m)} = K(X_{hi}^{(m)})$, $H_n = \sqrt{nh_n^p}$, $\overline{\theta}_n = H_n\left(\widehat{a}_0 - q_\tau(x), h_n(\widehat{a}_1 - \dot{q}(x))'\right)'$, $\theta = H_n\left(a_0 - q_\tau(x), h_n(a_1 - \dot{q}(x))'\right)'$, $\tilde{\theta} = H_n(\tilde{a}_0 - q_\tau(x), h_n(\tilde{a}_1 - \dot{q}(x))')'$ *where* $(a_0, a_1')'$, $(\tilde{a}_0, \tilde{a}_1')' \in R^{1+p}$. *Let* Y_i^* *be defined in Theorem 3.1, and set* $Y_{ni}^*(\theta) = Y_i^* - \theta'\mathcal{X}_{hi}/H_n$, $T_{ni} = (\dot{q}(x))'X_{hi}h_n$, $U_{ni} = U_{ni}(\theta) = T_{ni} + \theta'\mathcal{X}_{hi}/H_n$. $Y_{ni}^{*(m)}(\theta) = Y_i^{*(m)} - \theta'\mathcal{X}_{hi}^{(m)}/H_n$, $T_{ni}^{(m)} = (\dot{q}(x))'X_{hi}^{(m)}h_n$, $U_{ni}^{(m)} = U_{ni}^{(m)}(\theta) = T_{ni}^{(m)} + \theta'\mathcal{X}_{hi}^{(m)}/H_n$.

The following properties are useful in the discussion.

$$Y_i^* = Y_i - q_\tau(x) - T_{ni}, \tag{B.1a}$$

$$Y_i^{*(m)} = Y_i^{(m)} - q_\tau(x) - T_{ni}^{(m)},$$

$$Y_{ni}^*(\theta) = Y_i - q_\tau(x) - U_{ni}(\theta) = Y_i - a_0 - a_1'(X_i - x), \tag{B.1b}$$

$$Y_{ni}^{*(m)}(\theta) = Y_i^{(m)} - q_\tau(x) - U_{ni}^{(m)}(\theta) = Y_i^{(m)} - a_0 - a_1'(X_i^{(m)} - x).$$

Since $K(\cdot)$ *is a bounded density function with a bounded support,*

$$\|X_{hi}\| \le C, \quad \|\mathcal{X}_{hi}\| \le C \text{ when } K_i > 0, \tag{B.1c}$$

and when $\|\theta\| \le M$ *and* $K_i > 0$,

$$|T_{ni}| \le Ch_n, \quad |U_{ni}| \le Ch_n + CH_n^{-1} \to 0 \text{ as } n \to \infty. \tag{B.1d}$$

It follows from (2.5) that

$$\overline{\theta}_n = \text{argmin}_{\theta \in R^{1+p}} \sum_{i=1}^n \rho_\tau(Y_{ni}^*(\theta))K_i \stackrel{\Delta}{=} \text{argmin}_{\theta \in R^{1+p}} G_n(\theta). \tag{B.1e}$$

Set

$$V_n(\theta) = H_n^{-1} \sum_{i=1}^n \psi_\tau(Y_{ni}^*(\theta))\mathcal{X}_{hi}K_i. \tag{B.2}$$

Note that

$$V_n(\theta) - V_n(0) = H_n^{-1} \sum_{i=1}^{n} [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)] \mathcal{X}_{hi} K_i$$
$$\triangleq H_n^{-1} \sum_{i=1}^{n} V_{ni}(\theta). \tag{B.3}$$

Set $V_{ni}(\theta) = (V_{ni}^0(\theta), (V_{ni}^1(\theta))')'$ *and* $\Delta_i = V_{ni}(\theta) - EV_{ni}(\theta)$, *where* $V_{ni}^0(\theta) = [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)]K_i$ *and* $V_{ni}^1(\theta) = [\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)]X_{hi}K_i$. *And set* $V_{ni}^{(m)}(\theta) = (V_{ni}^{0(m)}(\theta), (V_{ni}^{1(m)}(\theta))')'$ *and* $\Delta_i^{(m)} = V_{ni}^{(m)}(\theta) - EV_{ni}^{(m)}(\theta)$, *where* $V_{ni}^{0(m)}(\theta) = [\psi_\tau(Y_{ni}^{*(m)}(\theta)) - \psi_\tau(Y_i^{*(m)})]K_i^{(m)}$ *and* $V_{ni}^{1(m)}(\theta) = [\psi_\tau(Y_{ni}^{*(m)}(\theta)) - \psi_\tau(Y_i^{*(m)})]X_{hi}^{(m)}K_i^{(m)}$.

Lemma B.1 *Let* $V_n(\pi)$ *be a vector function that satisfies*

(i) $-\pi' V_n(\lambda\pi) \geq -\pi' V_n(\pi)$, $\lambda \geq 1$,

(ii) $\sup_{\|\pi\| \leq M} \|V_n(\pi) + f_{Y|X}(q_\tau(x)|x) D\pi - A_n\| = o_P(1)$,

where $\|A_n\| = O_P(1)$, $0 < M < \infty$, $f_{Y|X}(q_\tau(x)|x) > 0$, *and* D *is a positive-definite matrix. Suppose that* π_n *such that* $\|V_n(\pi_n)\| = o_P(1)$. *Then,* $\|\pi_n\| = O_P(1)$ *and*

$$\pi_n = [f_{Y|X}(q_\tau(x)|x)]^{-1} D^{-1} A_n + o_P(1).$$

Proof. *This is Lemma A.4 of Koenker & Zhao (1996).*

The following lemmata is to check the conditions of Lemma B.1. We then can proof Theorem 3.1 based on Lemma B.1.

Lemma B.2 *Under Assumptions* ***A1****(ii, iii) and* ***A2***,

$$E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde\theta))|K_i \leq CEI_{(|Y_{ni}^*(\tilde\theta)| < C\|\theta - \tilde\theta\|/H_n)} K_i \leq C\|\theta - \tilde\theta\| h_n^p / H_n,$$
$$E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(\tilde\theta))|^2 K_i^2 \leq CEI_{(|Y_{ni}^*(\tilde\theta)| < C\|\theta - \tilde\theta\|/H_n)} K_i^2 \leq C\|\theta - \tilde\theta\| h_n^p / H_n,$$
$$E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_i^*)|^2 K^2((X_i^{(m)} - x)/h_n) \leq CEI_{(|Y_{ni}^*(\tilde\theta)| < C\|\theta\|/H_n)} K_i^{(m)2} \leq C\|\theta\| h_n^p / H_n.$$

for any θ, $\tilde\theta \in \{\theta : \|\theta\| \leq M\}$.

$$E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^{*(m)}(\theta))|^2 K^2((X_i^{(m)} - x)/h_n) \leq CL + \frac{C}{L} v_1(m)$$
$$E|\psi_\tau(Y_i^*) - \psi_\tau(Y_i^{*(m)})|^2 K^2((X_i^{(m)} - x)/h_n) \leq CL + \frac{C}{L} v_1(m),$$

where $L = L_n \to 0$

Proof. *From (B.1c)*

$$|\psi_\tau(Y^*_{ni}(\theta)) - \psi_\tau(Y^*_{ni}(\tilde\theta))|K_i = |I_{\{Y^*_{ni}(\theta)<0\}} - I_{\{Y^*_{ni}(\tilde\theta)<0\}}|K_i$$
$$=|I_{\{Y^*_{ni}(\tilde\theta)<(\theta-\tilde\theta)'\mathcal{X}_{hi}/H_n\}} - I_{\{Y^*_{ni}(\tilde\theta)<0\}}|K_i \le I_{\{|Y^*_{ni}(\tilde\theta)|<C\|\theta-\tilde\theta\|/H_n\}}K_i.$$

We have

$$E|\psi_\tau(Y^*_{ni}(\theta)) - \psi_\tau(Y^*_{ni}(\tilde\theta))|\ K_i \le CEI_{(|Y^*_{ni}(\tilde\theta)|<C\|\theta-\tilde\theta\|/H_n)}\ K_i.$$

From (B.1b), there exists $0 < \xi < 1$ *such that*

$$\begin{aligned}
&EI_{(|Y^*_{ni}(\tilde\theta)|<C\|\theta-\tilde\theta\|/H_n)}\ K_i\\
\le&CE[F_{Y|X}(q_\tau(x) + U_{ni}(\tilde\theta) + C\|\theta-\tilde\theta\|/H_n|X_i) - F_{Y|X}(q_\tau(x) + U_{ni}(\tilde\theta) - C\|\theta-\tilde\theta\|/H_n|X_i)]K_i\\
\le&CE[f_{Y|X}(q_\tau(x) + U_{ni}(\tilde\theta) - C\|\theta-\tilde\theta\|/H_n + 2\xi C\|\theta-\tilde\theta\|/H_n)|X_i]2C\|\theta-\tilde\theta\|/H_nK_i\\
\le&C\|\theta-\tilde\theta\|H_n^{-1}E[f_{Y|X}(q_\tau(x) + U_{ni}(\tilde\theta) - C\|\theta-\tilde\theta\|/H_n + 2\xi C\|\theta-\tilde\theta\|/H_n)|X_i]K_i.
\end{aligned}$$

for any $\theta,\ \tilde\theta \in \{\theta:\ \|\theta\| \le M\}$.

*Then using Assumptions **A1**(iii) and **A2** together with (B.1d) and Lemma A2 with n large enough, we have*

$$\begin{aligned}
&E|\psi_\tau(Y^*_{ni}(\theta)) - \psi_\tau(Y^*_{hi}(\tilde\theta))|\ K_i\\
\le&C\|\theta-\tilde\theta\|H_n^{-1}E[f_{Y|X}(q_\tau(x)|X_i)\\
&\qquad\times f_{Y|X}(q_\tau(x) + U_{ni}(\tilde\theta) - C\|\theta-\tilde\theta\|/H_n + 2\xi C\|\theta-\tilde\theta\|/H_n|X_i)/f_{Y|X}(q_\tau(x)|X_i)]K_i\\
\le&C\|\theta-\tilde\theta\|H_n^{-1}\ [Ef_{Y|X}(q_\tau(x)|X_i)K_i] = C\|\theta-\tilde\theta\|\ O(h_n^pH_n^{-1}).
\end{aligned}$$

This is the first inequality of Lemma B2. The second one and third one can be proved similarly. For the fourth inequality, we have

$$\begin{aligned}
&E|\psi_\tau(Y^*_{ni}(\theta)) - \psi_\tau(Y^{*(m)}_{ni}(\theta))|^2K^2((X_i^{(m)} - x)/h_n)\\
=&E|I_{\{Y^*_{ni}(\theta)<0\}} - I_{\{Y^{*(m)}_{ni}(\theta)<0\}}|^2K^2((X_i^{(m)} - x)/h_n)\\
\le&E|I_{\{Y^*_{ni}(\theta)<0\}} - I_{\{Y^*_{ni}(\theta)<Y^*_{ni}(\theta)-Y^{*(m)}_{ni}(\theta)\}}|^2K_i^{(m)2}\\
\le&CEI^2_{\{|Y^*_{ni}(\theta)|<|Y^*_{ni}(\theta)-Y^{*(m)}_{ni}(\theta)|\}}K_i^{(m)2}
\end{aligned}$$

$$
\begin{aligned}
\leq & C\{P(|Y^*_{ni}(\theta)| < |Y^*_{ni}(\theta) - Y^{*(m)}_{ni}(\theta)|, |Y^*_{ni}(\theta)| \leq L) \\
& \quad + P(|Y^*_{ni}(\theta)| < |Y^*_{ni}(\theta) - Y^{*(m)}_{ni}(\theta)|, |Y^*_{ni}(\theta)| > L)\} \\
\leq & C\{P(|Y^*_{ni}(\theta)| \leq L) + P(|Y^*_{ni}(\theta) - Y^{*(m)}_{ni}(\theta)| > L)\} \\
\leq & CL + \frac{C}{L} E|Y^*_{ni}(\theta) - Y^{*(m)}_{ni}(\theta)| \\
\leq & CL + \frac{C}{L}\{E|Y_i - Y_i^m| + E|X_i - X_i^m|\} \\
\leq & CL + \frac{C}{L} v_1(m),
\end{aligned}
$$

where $L = L_n$.

Lemma B.3 *Under the conditions of Lemma 3.1,*

$$
\sup_{\|\theta\| \leq M} \|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| = o_P(1). \tag{B.4}
$$

Proof. *The proof is divided into two steps. First we prove that for any fixed $\theta : \|\theta\| \leq M$,*

$$
\|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| = o_P(1). \tag{B.5}
$$

Then from (B.3), the left-hand side of (B.5) is bounded by

$$
H_n^{-1}|\sum_{i=1}^n (V^0_{ni}(\theta) - EV^0_{ni}(\theta))| + H_n^{-1}\|\sum_{i=1}^n (V^1_{ni}(\theta) - EV^1_{ni}(\theta))\| \triangleq V_n^0 + V_n^1. \tag{B.6}
$$

It follows from the stationarity and Lemma A.1 that

$$
\begin{aligned}
E(V_n^0)^2 &= (nh_n^p)^{-1}\left\{\sum_{j=1}^n E(\Delta_j^0)^2 + 2\sum_{1\leq i<j\leq n} E\Delta_i^0\Delta_j^0\right\} \\
&= h_n^{-p}E(\Delta_j^0)^2 + 2(nh_n^p)^{-1}\sum_{1\leq i<j\leq n} E\Delta_i^0\Delta_j^0 := A_{n1} + 2A_{n2}.
\end{aligned} \tag{B.7a}
$$

In order to bound (B.7a), we apply Lemma B1 with $\tilde{\theta} = 0$; for $\|\theta\| \leq M$,

$$
h_n^{-p}E(\Delta_j^0)^2 \leq h_n^{-p}\mathrm{var}(V^0_{n1}(\theta)) \leq h_n^{-p}E(V^0_{n1})^2 = h_n^{-p}E|\psi_\tau(Y^*_{ni}(\theta)) - \psi_\tau(Y_i^*)|^2 K_i^2 \leq CH_n^{-1}.
$$

Therefore, to complete the proof of this lemma, it suffices to show that $A_{n2} \to 0$ as $n \to \infty$. By noticing $E\Delta_i^0\Delta_j^0 = E\Delta_i^{0(m)}\Delta_j^{0(m)} + E\Delta_i^{0(m)}(\Delta_j^0 - \Delta_j^{0(m)}) + E(\Delta_i^0 - \Delta_i^{0(m)})\Delta_j^0$, we can further *separate A_{T2} into three parts: $A_{n2} = A_{n21} + A_{n22} + A_{n23}$,*

$$A_{n21} := (nh_n^p)^{-1} \sum_{1 \le i < j \le n} E\Delta_i^{0(m)} \Delta_j^{0(m)} \le C(nh_n^{(1+2/a)p})^{-1/2} + CN_n^a \sum_{j=N_n}^{\infty} \alpha(j-m) = o(1) \tag{B.7b}$$

where $N_n = h_n^{-p/a}$, and the equality is due to Lemma A.3 and B2.

$$\begin{aligned} A_{n22} &= (nh_n^p)^{-1} \sum_{1 \le i < j \le n} E\Delta_i^{0(m)} (\Delta_j^0 - \Delta_j^{0(m)}) \\ &\le (nh_n^p)^{-1} \sum_{1 \le i < j \le n} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2} \\ &= (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2}, \end{aligned} \tag{B.8}$$

and as a result $E(\Delta_i^{0(m)})^2 \le Ch_n^p H_n^{-1} = o(1)$, and

$$\begin{aligned} & E\left(\Delta_j^0 - \Delta_j^{0(m)}\right)^2 \\ \le & E\left(V_{nj}^0(\theta) - V_{nj}^{0(m)}(\theta)\right)^2 \\ = & E\Big\{ [\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*) - \psi_\tau(Y_{nj}^{*(m)}(\theta)) + \psi_\tau(Y_j^{*(m)})] K((X_j^{(m)} - x)/h_n) \\ & \quad + [\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*)][K((X_j - x)/h_n) - K((X_j^{(m)} - x)/h_n)] \Big\}^2 \\ \le & 2\Big\{ E[\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*) - \psi_\tau(Y_{nj}^{*(m)}(\theta)) + \psi_\tau(Y_j^{*(m)})]^2 K^2((X_j^{(m)} - x)/h_n) \\ & \quad + E[\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_j^*)]^2 [K((X_j - x)/h_n) - K((X_j^{(m)} - x)/h_n)]^2 \Big\} \\ \le & C\Big[E|\psi_\tau(Y_{nj}^*(\theta)) - \psi_\tau(Y_{nj}^{*(m)}(\theta))|^2 K^2((X_j^{(m)} - x)/h_n) + E|\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})|^2 K^2((X_j^{(m)} - x)/h_n) \\ & \quad + H_n^{-1} h_n^{p-1} E f_{Y|X}(q_\tau(x)|X_i) \left\| X_j - X_j^{(m)} \right\| \Big] \\ \le & C\left[L + \frac{1}{L} v_1(m) + H_n^{-1} h_n^{p-1} v_1(m) \right], \end{aligned}$$

by using the Lipschitz continuity and boundedness of the kernel $K(\cdot)$. Therefore, we have

$$\begin{aligned} A_{n22} \le & (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{ E(\Delta_i^{0(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j^0 - \Delta_j^{0(m)})^2 \right\}^{1/2} \\ \le & C(nh_n^p)^{-3/4} n^{3/2} \left[L + v_1(m) L^{-1} + H_n h_n^{p-1} v_1(m) \right]^{1/2} = o(1), \end{aligned} \tag{B.9}$$

*where the equality is due to Assumption **B2** and $L = L_n = v_1(m)$. And similarly to A_{n22}, it can be proved that*

$$A_{n23} := (nh_n^p)^{-1} \sum_{1\le i<j\le n} E(\Delta_i^0 - \Delta_i^{0(m)})\Delta_j^0 \to 0.$$

Therefore,

$$E(V_n^0)^2 = A_{n1} + 2A_{n2} = o(1), \tag{B.10}$$

Similarly to $E(V_n^0)^2$, we have

$$E(V_n^1)^2 = o(1). \tag{B.11}$$

Thus (B.4) follows from (B.6) together with (B.10) and (B.11).

The next step is to use standard chaining argument in Bickel (1975) and He & Shao (1996). We only give a sketch here. Decompose $\{\|\theta\| \le M\}$ into cubes based on the grid $(j_1\gamma M, \cdots, j_{p+1}\gamma M)$, $j_i = 0, \pm1, \cdots, \pm[1/\gamma]+1$, where $[1/\gamma]$ denotes the integer part of $1/\gamma$, and γ is a small positive number independent of n. Let $R(\theta)$ be the lower vertex of the cube that contains θ. Clearly, $\|R(\theta) - \theta\| \le C\gamma$ and the number of the elements of $\{R(\theta) : \|\theta\| \le M\}$ is finite. Then

$$\sup_{\|\theta\|\le M} \|V_n(\theta) - V_n(0) - E(V_n(\theta) - V_n(0))\| \le V_{n1}^* + V_{n2}^* + V_{n3}^*, \tag{B.12}$$

where it follows from (B.4) that $V_{n1}^* = \sup_{\|\theta\|\le M} \|V_n(R(\theta)) - V_n(0) - E(V_n(R(\theta)) - V_n(0))\| = o_P(1)$, and $V_{n2}^* = \sup_{\|\theta\|\le M} \|V_n(\theta) - V_n(R(\theta))\|$ and $V_{n3}^* = \sup_{\|\theta\|\le M} \|E(V_n(\theta) - V_n(R(\theta)))\|$. Using (B.1c) and for $\|\theta\| \le M$, applying Lemma B2 with $\tilde\theta = R(\theta)$ with n large, we have

$$V_{n3}^* \le CH_n^{-1} n \sup_{\|\theta\|\le M} E|\psi_\tau(Y_{ni}^*(\theta)) - \psi_\tau(Y_{ni}^*(R(\theta)))|K_i \le C \sup_{\|\theta\|\le M} \|\theta - R(\theta)\| \le C\gamma. \tag{B.13}$$

Therefore letting $n\to\infty$ and then $\gamma \to 0$ leads to $V_{n3}^* = o(1)$.

Set $B_i(\theta) = I_{(|Y_{ni}^*(\theta)|<C\gamma/H_n)} \|\mathcal{X}_{hi}\| K_i$. Then

$$V_{n2}^* \le \sup_{\|\theta\|\le M} \|V_n(\theta) - V_n(R(\theta))\| \le C \sup_{\|\theta\|\le M} H_n^{-1} \sum_{i=1}^n B_i(R(\theta)) \le B_{n1} + B_{n2}, \tag{B.14}$$

where a similar argument to (B.13) leads to $B_{n1} = C\sup_{\|\theta\|\le M} H_n^{-1}\sum_{i=1}^n EB_i(R(\theta)) = o(1)$, *and similarly to (B.7),* $B_{n2} = C\sup_{\|\theta\|\le M} |H_n^{-1}\sum_{i=1}^n (B_i(R(\theta)) - EB_i(R(\theta)))| = o_P(1)$. *Thus,* $V_{n2}^* = o_P(1)$. *Finally (B.3) follows from (B.12).*

Lemma B.4 *Under Assumptions **A1**(iii) and **A2**,*

$$\sup_{\|\theta\|\le M} \|E(V_n(\theta) - V_n(0)) + f_{Y|X}(q_\tau(x)|x)D\theta\| = o(1), \tag{B.15}$$

where $D = f_X(x)\mathrm{diag}(1, \int uu'K(u)du)$.

Proof. *It follows from (B.5) and (B.1) that*

$$\begin{aligned} E(V_n(\theta) - V_n(0)) =& H_n^{-1} nE[I_{(Y_i^*<0)} - I_{(Y_{ni}^*(\theta)<0)}]\mathcal{X}_{hi}K_i \\ =& H_n h_n^{-p} E[F(q_\tau(x) + T_{ni}|X_i) - F(q_\tau(x) + U_{ni}(\theta)|X_i)]\mathcal{X}_{hi}K_i. \end{aligned}$$

Then similar to the proof of Lemma B1 with $U_{ni} - T_{ni} = \theta'\mathcal{X}_{hi}/H_n$ *and* $\|\theta\| \le M$, *there exists a* $0 < \xi < 1$ *such that*

$$\begin{aligned} &\sup_{\|\theta\|\le M} \|E(V_n(\theta) - V_n(0)) + f_{Y|X}(q_\tau(x)|x)D\theta\| \\ =& \sup_{\|\theta\|\le M} \| - h_n^{-p} E[f(q_\tau(x) + T_{ni} + \xi\theta'\mathcal{X}_{hi}/H_n|X_i)]\theta'\mathcal{X}_{hi}\mathcal{X}_{hi}K_i + f_{Y|X}(q_\tau(x)|x)D\theta\| \\ =& \sup_{\|\theta\|\le M} \| - h_n^{-p} E[f(q_\tau(x) + T_{ni} + \xi\theta'\mathcal{X}_{hi}/H_n|X_i) - f(q_\tau(x)|X_i)]\theta'\mathcal{X}_{hi}\mathcal{X}_{hi}K_i \\ &- E[h_n^{-p} f(q_\tau(x)|X_i)]\mathcal{X}_{hi}\mathcal{X}_{hi}'K_i - f_{Y|X}(q_\tau(x)|x)D]\theta\| \\ \le& C\sup_{\|\theta\|\le M} h_n^{-p} E|f(q_\tau(x) + T_{ni} + \xi\theta'\mathcal{X}_{hi}/H_n|X_i) - f(q_\tau(x)|X_i)|\|\mathcal{X}_{hi}\mathcal{X}_{hi}'\|K_i \\ &+ C\|E[h_n^{-p} f(q_\tau(x)|X_i)]\mathcal{X}_{hi}\mathcal{X}_{hi}'K_i - f_{Y|X}(q_\tau(x)|x)D\| = o(1), \end{aligned}$$

*where the last inequality follows from Assumptions **A1**(iii) and **A2**, (B.1d) and Lemma A.2.*

Lemma B.5 *Let* $\overline{\theta}_n$ *be the minimizer of the function defined in (B.1e). Then*

$$\|V_n(\overline{\theta}_n)\| \le dim(\mathcal{X}_{hi})H_n^{-1} max_{i\le n}\|\mathcal{X}_{hi}'K_i\|.$$

Proof. *The proof follows from Ruppert & Carroll (1980).*

Lemma B.6 *Under Assumptions **A1** and **A2**, if* $a \ge 1$ *and* $h_n \to 0$, *then*

$$E(c'V_n(0) - c'EV_n(0))^2 \to \tau(1-\tau)f_X(x)\int K_c^2(u)du$$

as $n \to \infty$*, where* $c = (c_0, c_1')' \in R^{1+p}$.

Proof. *Set* $v_i = \psi_\tau(Y_i^*)K_c((X_i - x)/h_n)$, $v_i^{(m)} = \psi_\tau(Y_i^{*(m)})K_c((X_i^{(m)} - x)/h_n)$, *and* $\eta_i = v_i - Ev_i, \eta_i^{(m)} = v_i^{(m)} - Ev_i^{(m)}$. *A similar argument to (B.7) leads to*

$$\begin{aligned} E(c'V_n(0) - c'EV_n(0))^2 &= (nh_n^p)^{-1}\left\{\sum_{i=1}^{n} E\eta_i^2 + 2\sum_{1\le i<j\le n} E\eta_i\eta_j\right\} \\ &= h_n^{-p}E\eta_1^2 + 2h_n^{-p}\sum_{1\le i<j\le n} E\eta_i\eta_j \\ &\triangleq v_{n1} + 2v_{n2} \end{aligned} \tag{B.16}$$

Note that (A.2a) and (A.2b) of Lemma A.2 gives

$$\begin{aligned} EI_{(Y_1^*<0)}K_c^2((X_1 - x)/h_n) &= EF_{Y|X}(q_\tau(x) + \dot{q}_\tau(X_1 - x)|X_1)K_c^2((X_1 - x)/h_n) \\ &\to \tau f_X(x)\int K_c^2(u)du, \\ EI_{(Y_1^*<0)}K_c((X_1 - x)/h_n) &\to \tau f_X(x)\int K_c(u)du, \end{aligned}$$

which lead to

$$\begin{aligned} h_n^{-p}Ev_1^2 &= E[\tau^2 - 2\tau I_{(Y_1^*<0)} + I_{(Y_1^*<0)}]K_c^2((X_1 - x)/h_n) \\ &\to \tau(1-\tau)f_X(x)\int K_c^2(u)du, \end{aligned}$$

and

$$h_n^{-p}Ev_1 = E[\tau - I_{(Y_1^*<0)}]K_c((X_1 - x)/h_n) \to (\tau - \tau)f_X(x)\int K_c(u)du = 0.$$

Then

$$v_{n1} = h_n^{-p}Ev_1^2 - h_n^{-p}(Ev_1)^2 \to \tau(1-\tau)f_X(x)\int K_c^2(u)du. \tag{B.17}$$

Therefore, to complete the proof of this lemma, it suffices to show that $v_{n2} \to 0$ *as* $n \to \infty$*. By noticing* $E\eta_i\eta_j = E\eta_i^{(m)}\eta_j^{(m)} + E\eta_i^{(m)}(\eta_j - \eta_j^{(m)}) + E(\eta_i - \eta_i^{(m)})\eta_j$*, we can further separate* v_{n2} *into three parts:* $v_{n2} = v_{n21} + v_{n22} + v_{n23}$,

$$v_{n21} := (nh_n^p)^{-1}\sum_{1\le i<j\le n} E\eta_i^{(m)}\eta_j^{(m)} \le O(h_n) + \epsilon O(h_n^{p(1-1/a)}) + CN_n^a\sum_{j=N_n}^{\infty}\alpha(j-m) \to 0 \tag{B.18}$$

Take $N_n = \epsilon h_n^{-p/a}$, where ϵ is a small positive number, and $a \geq 1$.

$$
\begin{aligned}
v_{n22} &= (nh_n^p)^{-1} \sum_{1\leq i<j\leq n} E\eta_i^{(m)}(\eta_j - \eta_j^{(m)}) \\
&\leq (nh_n^p)^{-1} \sum_{1\leq i<j\leq n} \left\{E(\eta_i^{(m)})^2\right\}^{1/2} \left\{E(\eta_j - \eta_j^{(m)})^2\right\}^{1/2} \\
&= (nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{E(\eta_i^{(m)})^2\right\}^{1/2} \left\{E(\eta_j - \eta_j^{(m)})^2\right\}^{1/2},
\end{aligned} \tag{B.19}
$$

and as a result $E(\eta_i^{(m)})^2 \leq E(v_i^{(m)})^2 \leq Ch_n^p = o(1)$ and by using the properties of the kernel $K_c(\cdot)$

$$
\begin{aligned}
E\left(\eta_j - \eta_j^{(m)}\right)^2 \leq E\left(v_j - v_j^{(m)}\right)^2 \\
&= E\Big\{[\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})]K_c((X_j^{(m)} - x)/h_n) \\
&\quad + [\psi_\tau(Y_j^*)][K_c((X_j - x)/h_n) - K_c((X_j^{(m)} - x)/h_n)]\Big\}^2 \\
&\leq 2\Big\{E[\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})]^2 K_c^2((X_j^{(m)} - x)/h_n) \\
&\quad + E[\psi_\tau(Y_j^*)]^2[K_c((X_j - x)/h_n) - K_c((X_j^{(m)} - x)/h_n)]^2\Big\} \\
&\leq C\Big[E[\psi_\tau(Y_j^*) - \psi_\tau(Y_j^{*(m)})]^2 K_c^2((X_j^{(m)} - x)/h_n) \\
&\quad + h_n^{p-1} EF_{Y|X}(q_\tau(x)|X_j)\left\|X_j - X_j^{(m)}\right\|\Big] \\
&\leq C\left[L + L^{-1}v_1(m) + h_n^{p-1}v_1(m)\right],
\end{aligned}
$$

we have

$$
\begin{aligned}
v_{n22} \leq &(nh_n^p)^{-1} \frac{n(n-1)}{2} \left\{E(\eta_i^{(m)})^2\right\}^{1/2} \left\{E(\eta_j - \eta_j^{(m)})^2\right\}^{1/2} \\
\leq &C(nh_n^p)^{-1/2} n^{3/2} \left[L + v_1(m)L^{-1} + H_n h_n^{p-1} v_1(m)\right]^{1/2} = o(1).
\end{aligned} \tag{B.20}
$$

And similarly to v_{n22}, it can be proved that

$$
v_{n23} := (nh_n^p)^{-1} \sum_{1\leq i<j\leq n} E(\eta_i - \eta_i^{(m)})\eta_j \to 0.
$$

Therefore,

$$
v_{n2} = v_{n21} + v_{n22} + v_{n23} = o(1). \tag{B.21}
$$

Finally the lemma follows from (B.17) and (B.21).

Proof of Theorem 3.1 *We now check the conditions of Lemma B.1, lemmata B.3 and B.4 lead to (ii) of Lemma B.1* $\|V_n(\bar{\theta}_n)\| = o_P(1)$ *follows from Lemma B.5 together with assumptions A2 and A3. Take* $A_n = V_n(\mathbf{0})$. *It can be seen from Lemma B.6 that* $A_n = O_P(1)$. *Since* $\psi_\tau(y)$ *is an increasing function of* y, *the function*

$$-\theta' V_n(\lambda\theta) = H_n^{-1} \sum_{i=1}^{n} \psi_\tau(y_i^* - \lambda\theta'\mathcal{X}_{hi}/H_n)(-\theta'\mathcal{X}_{hi})K_i$$

is increasing as a function of λ. *Therefore, condition (i) of Lemma B.1 holds and the result follows.*

Proof of Theorem 3.2 *The arguments are similar. We only give the proof of (3.2).*

Lemma 3.1 entails that

$$\begin{aligned}
&\widehat{q}_\tau(x) - q_\tau(x) = (nh_n^p)^{-1}\phi_\tau \sum_{i=1}^{n} \psi_\tau(Y_i^*)K_i + o_P(1/H_n) \\
=&(nh_n^p)^{-1}\phi_\tau \sum_{i=1}^{n}[\psi_\tau(Y_i^*)K_i - E\psi_\tau(Y_i^*)K_i] + (nh_n^p)^{-1}\phi_\tau \sum_{i=1}^{n} E\psi_\tau(Y_i^*)K_i + o_P(1/H_n) \\
&\stackrel{\Delta}{=} Q_{n1} + Q_{n2} + o_P(1/H_n).
\end{aligned} \tag{B.22}$$

Note that $\tau = F(q_\tau(X_i)|X_i)$ *and by (B.1c) and Assumption* ***A3*** *with* $v = 1$ *that when* $K_i > 0$, *there exists* $0 < \tilde{\xi} < 1$ *such that*

$$\begin{aligned}
|\Delta_i(x)| &\stackrel{\Delta}{=} |q_\tau(X_i) - q_\tau(x) - T_{ni}| \\
&= |[\dot{q}(x + \tilde{\xi}X_{hi}h_n) - \dot{q}(x)]'X_{hi}h_n| \le Ch_n^{1+\delta}.
\end{aligned} \tag{B.23a}$$

There exists a $0 < \xi < 1$ *from (B.23a) that,*

$$\begin{aligned}
|Q_{n2}| &= |h_n^{-p}\phi_\tau E[\tau - I_{(Y_i^*<0)}]K_i| = h_n^{-p}\phi_\tau E[F(q_\tau(X_i)|X_i) - F(q_\tau(x) + T_{ni}|X_i)]K_i \\
&= |h_n^{-p}\phi_\tau E[f(q_\tau(x) + T_{ni} + \xi\Delta_i(x)|X_i)\Delta_i(x)K_i]| \\
&\le O(h_n^{1+\delta})h_n^{-p}\phi_\tau E[f(q_\tau(x) + T_{ni} + \xi\Delta_i(x)|X_i)K_i] = O(h_n^{1+\delta}),
\end{aligned} \tag{B.23b}$$

where the last inequality is derived from (B.1d) and Assumption ***A1****(iii).*

On the other hand, it easily follows from Lemma B.6 with $c = (1, \mathbf{0}')' \in R^{1+p}$ *that,*

$$EQ_{n1}^2 = (nh_n^p)^{-1}\phi_\tau^2 E(c'V_n(0) - c'EV_n(0))^2 = O_P((nh_n^p)^{-1}) = O_P(H_n^{-2}),$$

which entails $Q_{n1} = O_P(H_n^{-1})$. *The result of this theorem follows from (B.22).*

C: Proofs for Subsection 3.2

The proof of Theorem 3.3 is based on the local Bahadur representation given in Subsection 3.1. We first proof the Lemma 3.1 and Lemma 3.2.

Proof of Lemma 3.1 *Based on the Bahadur representation of Theorem 3.1, the proof is similar to the arguments in the corresponding proof for mean regression in Lu & Linton (2007) . We first derive the asymptotic variance and expectation, with the Lemma A.3 replacing the corresponding Lemma A.3 in Lu & Linton (2007). Suppose*

$$W_n := \begin{pmatrix} w_{n0} \\ w_{n1} \end{pmatrix}, \quad (W_n)_j := (nh_n^p)^{-1} \sum_{i=1}^n \psi_\tau(Y_i^*) \left(\frac{X_i - x}{h_n}\right)_j K\left(\frac{X_i - x}{h_n}\right), \quad j = 0, \ldots, p, \tag{C.1}$$

with $\left(\frac{X_i - x}{h_n}\right)_0 = 1$.

Denote by $K_j(x) = (x)_j K(x)$ *Then it can be noted that*

$$\begin{aligned}
&E\left|\left(W_n^{(m)}\right)_j - (W_n)_j\right| = E\left|(nh_n^p)^{-1} \sum_{i=1}^n \left[\psi_\tau(Y_i^{*(m)})K_j\left(\frac{X_i^{(m)} - x}{h_n}\right) - \psi_\tau(Y_i^*)K_j\left(\frac{X_i - x}{h_n}\right)\right]\right| \\
&\le (nh_n^p)^{-1} \sum_{i=1}^n E\left|\psi_\tau(Y_i^{*(m)})K_j\left(\frac{X_i^{(m)} - x}{h_n}\right) - \psi_\tau(Y_i^*)K_j\left(\frac{X_i - x}{h_n}\right)\right| \\
&= h_n^{-p}E\left|\psi_\tau(Y_i^{*(m)})K_j\left(\frac{X_i^{(m)} - x}{h_n}\right) - \psi_\tau(Y_i^*)K_j\left(\frac{X_i - x}{h_n}\right)\right| \\
&\le h_n^{-p}E\left|\psi_\tau(Y_i^{*(m)}) - \psi_\tau(Y_i^*)\right| K_j\left(\frac{X_i^{(m)} - x}{h_n}\right) + h_n^{-p}E|\psi_\tau(Y_i^*)| \left|K_j\left(\frac{X_i^{(m)} - x}{h_n}\right) - K_j\left(\frac{X_i - x}{h_n}\right)\right| \\
&= O_p\left(h_n^{-p}v_1(m)\right) + O_p\left(h_n^{-p-1}v_1(m)\right) \\
&= O_p\left(h_n^{-p-1}v_1(m)\right).
\end{aligned} \tag{C.2}$$

The usual Cramér-Wold device will be adopted. For all $c := (c_0, c_1')' \in \mathbb{R}^{1+p}$, *let*

$$A_n := (\ nh_n^p)^{1/2}c'W_n = \phi_\tau(x)\frac{1}{\sqrt{nh_n^p}}\sum_{i=1}^n \psi_\tau(Y_i^*)K_c(\frac{X_i - x}{h_n}),$$

with $K_c(u)$ *defined in* ***A2****(ii).*

The expectation of the first term on right-hand side of (3.1) is as

$$E\left[\phi_\tau(x)\frac{1}{\sqrt{nh_n^p}}\sum_{i=1}^n \psi_\tau(Y_i^*) \begin{pmatrix} 1 \\ \frac{X_i - x}{h_n} \end{pmatrix} K(\frac{X_i - x}{h_n})\right]$$

$$=\phi_\tau(x)\frac{1}{\sqrt{nh_n^p}}nE\left[\psi_\tau(Y_i^*)\begin{pmatrix}1\\ \frac{X_i-x}{h_n}\end{pmatrix}K(\frac{X_i-x}{h_n})\right]$$

$$=\phi_\tau(x)\sqrt{nh_n^p}h_n^{-p}E\left[\left(F_{Y|X}(q_\tau(X_i)|X_i)-F_{Y|X}(q_\tau(x)+(\dot{q}_\tau(x))'(X_i-x)|X_i)\right)\begin{pmatrix}1\\ \frac{X_i-x}{h_n}\end{pmatrix}K(\frac{X_i-x}{h_n})\right]$$

$$=\sqrt{nh_n^p}\left[(1+o(1))\begin{pmatrix}B_0(x)\\ B_1(x)\end{pmatrix}\right] \tag{C.3}$$

Based on Lemma B.6, the variance is as

$$\begin{aligned}\Sigma := \mathrm{Var}&\left[\phi_\tau(x)\frac{1}{\sqrt{nh_n^p}}\sum_{i=1}^n\psi_\tau(Y_i^*)\begin{pmatrix}1\\ \frac{X_i-x}{h_n}\end{pmatrix}K(\frac{X_i-x}{h_n})\right]\\ =&Var\left(\phi_\tau(x)V_n(0)\right)\\ =&\phi_\tau^2(x)\tau(1-\tau)f_X(x)\begin{pmatrix}\int K^2(u)du & \int u'K^2(u)du\\ \int uK^2(u)du & \int uu'K^2(u)du\end{pmatrix}\end{aligned} \tag{C.4}$$

Then

$$\lim_{n\to\infty}\mathrm{Var}[A_n]=(\phi_\tau(x))^{-2}c'\Sigma c, \tag{C.5}$$

Proof of Lemma 3.2 *The fundamental idea to prove the asymptotic normality of $W_n(x)$ is to divide $W_n(x)$ into two parts: with $m=m_n\to\infty$ (to be specified later),*

$$W_n(x)=W_n^{(m)}(x)+\left[W_n(x)-W_n^{(m)}(x)\right], \tag{C.6}$$

Then applying the approximation lemma 3.1,

$$(nh_n^p)^{1/2}\left[W_n(x)-W_n^{(m)}(x)\right]=O_p\left(n^{1/2}h_n^{-1/2p-1}v_1(m)\right)\to_P 0,$$

*following from Assumption **B2**; and similarly*

$$(nh_n^p)^{1/2}E\left[W_n(x)-W_n^{(m)}(x)\right]\to 0.$$

Therefore, like Lu & Linton (2007), it suffices to prove that

$$(nh_n^p)^{1/2}(c'[W_n^{(m)}(x)-EW_n^{(m)}(x)]/\sigma)$$

is asymptotically standard normal as $n \to \infty$, which is the main effort we will made in this paper. Define

$$\eta_i^{(m)}(x) := \psi_\tau(Y_i^{*(m)}) K_c((X_i^{(m)} - x)/h_n),$$
$$\zeta_{ni}^{(m)} := h_n^{-p/2}\left(\eta_i^{(m)}(x) - \mathrm{E}\eta_i^{(m)}(x)\right),$$

and let $S_n^{(m)} := \sum_{i=1}^n \zeta_{ni}^{(m)}$. Then,

$$n^{-1/2} S_n^{(m)} = (nh_n^p)^{1/2} c'(W_n^{(m)}(x) - EW_n^{(m)}(x)) = A_n^{(m)} - EA_n^{(m)},$$

Then, we decompose $n^{-1/2}S_n^{(m)}$ into smaller pieces involving "large" and "small" blocks. More specifically, consider

$$U^{(m)}(1, n, x, k) := \sum_{j=k(p^*+q^*)+1}^{k(p^*+q^*)+p^*} \zeta_{nj}^{(m)}(x),$$
$$U^{(m)}(2, n, x, k) := \sum_{j=k(p^*+q^*)+p^*+1}^{(k+1)(p^*+q^*)} \zeta_{nj}^{(m)}(x)$$

*where $p^* = p_n^*$ and $q^* = q_n^*$ are specified in Assumption **B3**. Without loss of generality, assume that, for some integer $r = r_n$, n is such that $n = r(p^* + q^*)$, with $r \to \infty$. For each integer $1 \le j \le 2$, define*

$$\mathrm{Y}^{(m)}(n, x, j) := \sum_{k=0}^{r-1} U^{(m)}(j, n, x, k).$$

Clearly $S_n^{(m)} = \mathrm{Y}^{(m)}(n, x, 1) + \mathrm{Y}^{(m)}(n, x, 2)$. Note that $\mathrm{Y}^{(m)}(n, x, 1)$ is the sum of the random variables $\zeta_{nj}^{(m)}$ over "large" blocks, whereas $\mathrm{Y}^{(m)}(n, x, 2)$ are sums over "small" blocks. If it is not the case that $n = r(p^* + q^*)$ for some integer r, then an additional term $\mathrm{Y}^{(m)}(n, x, 3)$, say, containing all the $\zeta_{ni}^{(m)}$'s that are not included in the big or small blocks, can be considered. This term will not change the proof much. The general approach consists in showing that, as $n \to \infty$,

$$Q_1^{(m)} := \left| \mathrm{E}exp[iu\,\mathrm{Y}^{(m)}(n, x, 1)] - \prod_{j=0}^{r-1} \mathrm{E}exp[iuU^{(m)}(1, n, x, k)] \right| \longrightarrow 0, \tag{C.7}$$

$$Q_2^{(m)} := n^{-1}\mathrm{E}\left(\mathrm{Y}^{(m)}(n, x, 2)\right)^2 \longrightarrow 0, \tag{C.8}$$

$$Q_3^{(m)} := n^{-1}\sum_{k=0}^{r-1} \mathrm{E}[U^{(m)}(1, n, x, k)]^2 \longrightarrow \sigma^2, \tag{C.9}$$

$$Q_4^{(m)} := n^{-1}\sum_{k=0}^{r-1} \mathrm{E}[(U^{(m)}(1, n, x, k))^2 I\{|U^{(m)}(1, n, x, k)| > \varepsilon\sigma\, n^{1/2}\}] \longrightarrow 0 \tag{C.10}$$

for every $\varepsilon > 0$. *Note that*

$$[A_n^{(m)} - \mathrm{E}A_n^{(m)}]/\sigma = (\ nh_n^p)^{1/2} c' [W_n^{(m)}(x) - \mathrm{E}W_n^{(m)}(x)]/\sigma = S_n^{(m)}/(\sigma\ n^{1/2})$$
$$= \mathrm{Y}^{(m)}(n,x,1)/(\sigma\ n^{1/2}) + \mathrm{Y}^{(m)}(n,x,2)/(\sigma\ n^{1/2}).$$

The term $\mathrm{Y}^{(m)}(n,x,2)/(\sigma\ n^{1/2})$ *is asymptotically negligible by (C.8). The random variables* $U^{(m)}(1,n,x,k)$ *are asymptotically mutually independent by (C.7). The asymptotic normality of* $\mathrm{Y}^{(m)}(n,x,1)/(\sigma\ n^{1/2})$ *follows from (C.9) and the Lindeberg-Feller condition (C.10). The lemma thus follows if we can prove (C.7)-(C.10). These proofs are similar to the arguments in the corresponding proofs for mean regression in Lu & Linton (2007), with the different Lemma A.3 and (C.2) established in the above. Here, we only briefly sketch it.*

Proof of *(C.7) See the Proof of A.41 in appendix of Lu & Linton (2007).*

Proof of *(C.8). The proof follows exactly as in the corresponding proof for mean regression in Lu & Linton (2007), with the Lemma A.3 and (C.2) replacing. Here, we just briefly show the proof.*

For notational simplicity, refer to the random variables $U^{(m)}(2,n,x,k)$, $k = 0,1,\cdots,r-1$, *as* $\widehat{U}_1, \ldots \widehat{U}_r$. *We have*

$$\mathrm{E}[\mathrm{Y}^{(m)}(n,x,2)]^2 = \sum_{i=1}^{r} Var(\widehat{U}_i) + 2 \sum_{1\le i<j\le r} Cov(\widehat{U}_i, \widehat{U}_j) := \widehat{V}_1 + \widehat{V}_2, \quad say. \tag{C.11}$$

Since X_n *is stationary,*

$$\mathrm{Var}(\widehat{U}_i) = \sum_{i=1}^{q} \mathrm{E}\left[\left(\zeta_{ni}^{(m)}(x)\right)^2\right] + \sum_{1\le i<j\le q} \mathrm{E}\left[\zeta_{nj}^{(m)}(x)\zeta_{ni}^{(m)}(x)\right] := \widehat{V}_{11} + \widehat{V}_{12}.$$

From Lemma A.2 and the Lebesgue density theorem,

$$\begin{aligned}
\widehat{V}_{11} &= q\ \mathrm{Var}\{\zeta_{ni}^{(m)}(x)\} = q\ \{h_n^{-p}\mathrm{E}\left(\Delta_i^{(m)}(x)\right)^2\} \\
&\le= q\{h_n^{-p}\mathrm{E}\ \left(\psi_\tau(Y_i^{*(m)})K((x - X_i^{(m)})/h_n)\right)^2\} \\
&\le q\left\{h_n^{-p}\mathrm{E}\ (\psi_\tau(Y_i^*)K((x - X_i)/h_n))^2 + O(h_n^{-1-p}v_1(m))\right\} \\
&= O(q),
\end{aligned}$$

where the final equality follows from $h_n^{-1-p}v_1(m) = o(1)$ *by Assumption* ***B2***.

We then need the cross lemma, Lemma A.3, for $\widehat{V}_{12}$ and then taking $N_n = q$ yields

$$\begin{aligned}\widehat{V}_{12} &= h_n^{-p} \sum_{1\le i<j\le q} \mathrm{E}\left[\Delta_j^{(m)}(x)\Delta_i^{(m)}(x)\right] \\ &\le Cqh_n^p \left[q + \sum_{t=q}^{\infty} \alpha_m(t)\right] \\ &:= Cq\pi_n.\end{aligned}$$

*It follows from Assumption **B3** that $\pi_n = O(1)$ and*

$$n^{-1}\widehat{V}_1 = n^{-1}r\left(\hat{V}_{11} + \hat{V}_{12}\right) \le n^{-1}rCq[1+\pi_n] \le C(\frac{q}{p^*+q})[1+\pi_n]. \tag{C.12}$$

Similarly, we can obtain

$$|\widehat{V}_2| \le Cnh_n^p \sum_{t=q}^{\infty} \alpha_m(t). \tag{C.13}$$

*Assumption **B4** implies that $q^* h_n^p = O(1)$ and $\pi_n = O(1)$. Thus, from (C.11), (C.12), and (C.13),*

$$n^{-1}\mathrm{E}[\mathrm{Y}^{(m)}(n,x,2)]^2 \le C(\frac{q}{p^*+q})[1+\pi_n] + Ch_n^p\left(\sum_{t=q}^{\infty}\alpha_m(t)\right),$$

*which tends to zero by $q/p^* \to 0$ and Assumption **B4**; (C.8) follows.*

Proof of *(C.9) and (C.10) The main idea of the proofs is similar to the proofs in Appendix of Lu & Linton (2007). These easily follow by checking the Assumptions and changing the Lemma A.3 and (C.2). The detail is omitted.*

Proof of Theorems 3.3 *Based on the Bahadur representation of Theorem 3.1, the main idea of the proof of Theorem 3.3 is similar to the corresponding proofs for mean regression in Lu & Linton (2007) and α-mixing condition in Hallin et al. (2009). Here, we only briefly sketch it.*

Then consider (3.6). Set $v_i = [d_1\phi_{\tau_1}\psi_{\tau_1}(y_i^*(\tau_1)) + d_2\phi_{\tau_2}\psi_{\tau_2}(y_i^*(\tau_2))]K_i$, $\tilde{\tau} = d_1\phi_{\tau_1}\tau_1 + d_2\phi_{\tau_2}\tau_2$ and $I_i(\tau) = I_{(y_i^*(\tau)<0)}$. Here $d_i \in R^1$, $i = 1, 2$. A simple calculation leads to

$$\begin{aligned}v_i^2 = &\Big\{\tilde{\tau}^2 - 2\tilde{\tau}(d_1\phi_{\tau_1}I_i(\tau_1) + d_2\phi_{\tau_2}I_i(\tau_2)) + d_1^2\phi_{\tau_1}^2 I_i(\tau_1) \\ &+ 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}I_i(\tau_1)I_i(\tau_2) + d_2^2\phi_{\tau_2}^2 I_i(\tau_2)\Big\}K_i^2.\end{aligned} \tag{C.14}$$

It follows from (3.1) that

$$\sqrt{nh_n^p}(d_1\widehat{q}_{\tau_1}(x) + d_2\widehat{q}_{\tau_2}(x) - d_1 q_{\tau_1}(x) - d_2 q_{\tau_2}(x)) = H_n^{-1}\sum_{i=1}^{n} v_i + o_P(1) \triangleq D_n + o_P(1). \tag{C.15}$$

Similar to (B.23b),

$$ED_n = H_n^{-1}nEv_1 = H_n(h_n^{-p}Ev_1) = H_n O(h_n^{1+\delta}) = O((nh_n^{p+2(1+\delta)})^{1/2}) \to 0.$$

as $n \to \infty$. An analogous argument to (B.16) gives

$$E(D_n - ED_n)^2 \quad \triangleq v_{n1} + 2v_{n2}. \tag{C.16}$$

Similar to the proof of (B.17), (C.14) together with Lemma A.2 ensures

$$\begin{aligned} v_{n1} =& h_n^{-p}Ev_1^2 - h_n^{-p}(Ev_1)^2 \\ \to& [\tilde{\tau}^2 - 2\tilde{\tau}^2 + d_1^2\phi_{\tau_1}^2\tau_1 + 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}\min(\tau_1,\, \tau_2) + d_2^2\phi_{\tau_2}^2\tau_2]f_X(x)\int K^2(u)du \\ =& [d_1^2\phi_{\tau_1}^2\tau_1 + 2d_1\phi_{\tau_1}d_2\phi_{\tau_2}\min(\tau_1,\, \tau_2) + d_2^2\phi_{\tau_2}^2\tau_2 - \tilde{\tau}^2]f_X(x)\int K^2(u)du. \end{aligned} \tag{C.17}$$

In addition, similar argument to (B.20) leads to

$$|v_{n2}| \to 0. \tag{C.18}$$

Then the asymptotic variance for (3.6) follows from (C.16)-(C.18). The asymptotic variance in (3.7) can be obtained similarly.