Nonparametric Inference for Time-varying Coefficient Quantile Regression

Weichi Wu[1] and Zhou Zhou

University of Toronto

September 27, 2014

Abstract

The paper considers nonparametric inference for quantile regression models with time-varying coefficients. The errors and covariates of the regression are assumed to belong to a general class of locally stationary processes and are allowed to be cross-correlated. Simultaneous confidence tubes (SCT) and integrated squared difference tests (ISDT) are proposed for simultaneous nonparametric inference of the latter models with asymptotically correct coverage probabilities and type I error rates. Our methodologies are shown to possess certain asymptotically optimal properties. Furthermore, we propose an information criterion which performs consistent model selection for nonparametric quantile regression models of non-stationary time series. For implementation, a wild bootstrap procedure is proposed which is shown to be robust to the dependent and non-stationary data structure. Our method is applied to studying the asymmetric and time-varying dynamic structures of the US unemployment rate since the 1940s.

1 Introduction

When the time span of a stochastic process increases, it is more and more common to observe changes in its dynamic structure. For instance, we have observed different economic periods caused by such exogenous shocks as the Korean war and the Vietnam war in the 1950s-1960s, oil crisis in the 1970s, the monetary policy changes, stable oil prices and rise in private investment in the 1980s, the Dot Com bubble and the related stock market boom in the 1990s, the home price bubble in the 2000s. Many economic structures are subject to

[1]Corresponding author. Department of Statistics, 100 St. George Street, Toronto, Ontario, M5S 3G3 Canada.
E-mail: weichi.wu@mail.utoronto.ca
Key words and phrases. Information criterion, heterogeneity, local linear quantile estimators, time-varying models, simultaneous confidence tubes, integrated squared difference test, local stationarity, variable selection.

changes during different periods. When the time span of interest covers different economic periods, the parameters of the corresponding statistical procedures should be allowed to change with time, leading to non-stationary time series models. Besides econometrics, time varying models have wide applications in areas such as environmental sciences, finance and transportation. See Robinson (1989, 1991), Orbe et al. (2005, 2006) and Cai (2007), among others. In the statistics literature, varying coefficient models in the mean have been extensively studied. See for instance Hastie and Tibshirani (1993), Fan and Zhang (1999), Hoover et al. (1998), Zhang et al. (2002), Zhou and Wu (2010) and Zhang and Wu (2012) among others and the references therein. These models are useful for characterizing varying relationships between the covariates and the conditional mean of the response.

In many applications, however, various important features of the joint distribution of the response and the covariates cannot be captured by analyzing the conditional mean alone. In particular, the lower and upper conditional quantiles of the response are oftentimes related to the covariates in drastically different ways than that of the mean. Empirical evidences of such asymmetric behaviors are demonstrated for time series of the GDP (Oka and Qu 2011), unemployment rates, gasoline prices (Koenker and Xiao 2006) and demands for electricity (Hendriks and Koenker 1992), et al.. Quantile regression (Koenker 2005) provides a simple, efficient and interpretable alternative in such situations.

The main purpose of the paper is to perform uniform nonparametric inference of the following time-varying coefficient quantile regression model: suppose for a $p+1$ dimensional time series $(\mathbf{x}_i = (x_{i,1}, \cdots, x_{i,p})', y_i)_{i=1}^n$ and a pre-specified quantile τ, the conditional τth quantile of y_i given $\mathbf{x}_i$,

$$Q_\tau(y_i|\mathbf{x}_i) = \theta_{1,\tau}(i/n)x_{i,1} + ...\theta_{p,\tau}(i/n)x_{i,p}, \tag{1}$$

where $\theta_\tau(t) = (\theta_{1,\tau}(t), ..., \theta_{p,\tau}(t))'$ is a smooth function on $[0, 1]$. It is common to write (1) in the form

$$y_i = \mathbf{x}_i'\theta_\tau(i/n) + e_{i,\tau}, \quad i = 1, 2, \cdots, n, \tag{2}$$

where the errors $e_{i,\tau}$ satisfy $Q_\tau(e_{i,\tau}|\mathbf{x}_i) = 0$ almost surely. We shall study model (2) in a general framework in which the processes $(\mathbf{x}_i)$ and $(e_{i,\tau})$ are locally stationary[2], short range dependent and may correlate with each other. In the literature, time varying quantile regression models are considered by Honda (2004), Kim (2007) when errors are independent, and by Cai and Xu (2008) when errors are strictly stationary and α–mixing. Note that when errors are independent and multiple realizations are available at one time, that is, in the longitudinal or functional data settings, time-varying quantile models have been applied in the analysis of reference growth data by Cole (1994), Wei et al. (2006), and Wei and He (2006), among others.

The major contributions of the paper lie in the following three aspects. First, we develop two inferential tools, the simultaneous confidence tube (SCT) and the integrated squared difference test (ISDT) for uniform nonparametric inferences of regression coefficient function $\theta_\tau(t)$ in model (1). More specifically, let $\mathbf{C}$ be a fixed $p \times s$ matrix with rank $s \leq p$, and $\theta_{\mathbf{C},\tau}(\cdot) = \mathbf{C}'\theta_\tau(\cdot)$ be a linear combination of the regression coefficients $\theta_{j,\tau}(\cdot)$, $j = 0, 1, ..., p$. For a pre-assigned coverage probability $1 - \alpha$, we shall construct a $100(1-\alpha)\%$ asymptotic SCT $\{\zeta_\alpha(t), t \in (0,1)\}$ for $\theta_{\mathbf{C},\tau}(\cdot)$ defined on $\mathbb{R}^s \times (0,1)$ such that

$$\lim_{n\to\infty} [\mathbb{P}\{\theta_{\mathbf{C},\tau}(t) \in \zeta_\alpha(t), 0 < t < 1\}] = 1 - \alpha. \tag{3}$$

Meanwhile, for a pre-specified function $\theta^0_\tau(t)$, the ISDT tests the null hypothesis $\theta_{\mathbf{C},\tau}(\cdot) = \theta^0_\tau(\cdot)$ via

$$\int_0^1 |(\mathbf{C}'\hat{\theta}_\tau(t) - \theta^0_\tau(t))|^2 \pi(t)\, dt, \tag{4}$$

where $\hat{\theta}_\tau(t)$ is some nonparametric estimate of $\theta_\tau(t)$, $\pi(t)$ is a user-chosen positive weight function and $|\cdot|$ denotes the Euclidean norm of a vector. A large value of the ISDT statistic indicates violation of the null hypothesis. Observe that the time-varying temporal dependence and cross correlation in the covariates and errors make the construction of SCT and ISDT a difficult problem. In this paper, we utilize conditional empirical process and martingale decomposition techniques for the theoretical investigation of the latter statistics. We show that the proposed SCT and ISDT asymptotically achieve the correct coverage probability and type I error rates. Furthermore, we demonstrate that the ISDT can detect

[2]a special class of non-stationary process

local alternatives with the minimax rate in the sense of Ingster (1993). Observe that the ISDT is a nonparametric specification test. In the literature, numerous nonparametric specification tests have been developed and investigated. For (conditional) mean or density functions, see for example Hall and Titterington (1988), Härdle and Mammen (1993), Zheng (1996), Stute (1997), Xia (1998), Horowitz and Spokoiny (2001), Fan, Zhang and Zhang (2001), Fan and Jiang (2007), Zhou and Wu(2010), Zhang and Wu (2012) among others. For nonparametric inference of conditional quantiles of independent data, see for example Zheng (1998), Rosenkrantz (2000), Horowitz and Spokoiny (2002), He and Zhu (2003), Kim (2007) and Wang (2007, 2008) among others. Note that, recently Zhou (2010) obtained asymptotically correct simultaneous confidence bands for quantile curves of locally stationary and dependent data. However, the techniques there are tailored to trend estimation and cannot be applied to the regression setting with stochastic and endogenous covariates. To our knowledge, there has been no previous work on asymptotically correct uniform nonparametric inference of quantile regression with random designs.

Our second major contribution is that, we develope a consistent model selection criterion for time varying quantile regression model (2) under non-stationarity and temporal dependence. The asymptotical correctness and the empirical effectiveness of our criterion are shown and verified via both mathematical proof and simulation. Variable selection plays an important role in the model building process, since in its initial stage, a large number of candidate predictor variables are usually included to remove the potential modeling bias (Fan and Li (2001)). In the literature of quantile regression, several techniques are developed and applied widely in practice, most of which are based on the penalization of parametric regression quantiles. See for example, Koenker (2004), Li and Zhu (2005), Wang, Li and Jiang (2007), the generalized version of the SCAD (Fan and Li 2001) and the adaptive LASSO (Zou 2006) discussed in Wu and Liu (2009), Ando and Tsay (2011) among others. In the literature of nonparametric mean regression, Fan et al. (2003), Abramovich et al. (2007) considered the problem of predictor selection in time-varying models with stationary errors, and Zhang and Wu (2012) proposed a model selection rule under parameter instability and temporal dependence. Recently, there are a few works regarding variable selection based on non-parametric regression quantiles. For example, Tang, Wang and Zhu (2013) considered a penalization method to select variables for varying coefficient quantile models with longitudinal data which are independent across subjects. Our selection procedure is a complement to the aforementioned popular selection methods in the sense that our procedure is more robust to possible violations of the tightly specified parametric modelling as well as the stringent independent and/or stationary assumptions.

Our third major contribution is that, we propose a non-stationarity and dependence robust bootstrap method to improve the finite sample performance of SCT and ISDT. When the sample size is moderate, direct implementation of the SCT and ISDT usually does not bear good performances due to their slow convergence rates (see Theorem 2). In such situations, a bootstrap method shall be adopted to enhance the finite sample performance, see Härdle and Marron (1991), Hall (1991), Neuman and Kreiss (1998), Zhou and Wu (2010) among others. We show that the bootstrap implementation of the two methods converges faster than the naive method of direct implementation, and examine their accuracy and power via simulations.

The rest of the paper is structured as follows. Section 2 introduces the local linear quantile estimates, the dependence measures, and the general assumptions of the paper. Section 3 presents the asymptotic results of the SCT and the nonparametric specification tests. Section 3.1 establishes a uniform Bahadur representation, based on which Section 3.2 and 3.3 establish the asymptotic results for the SCT and ISDT, respectively. The variable selection criterion is proposed in Section 3.4. Implementations including the construction of robust bootstrap tests, the bandwidth selection, and the estimation of covariance matrices are listed in Section 4. Section 5 presents simulation studies on the the accuracy and the sensitivity of the SCT and ISDT. Section 6 shows an application to the study of the potential time-varying dynamic structures in different quantiles of the quarterly US unemployment rate since 1940s. Proofs are given in the Appendix.

2 Model Assumptions

We first introduce some notation. Let $I(\cdot)$ be the usual indicator function. For an interval $\mathcal{I} \in \mathbb{R}$, denote $\mathcal{C}^d$, $d \in \mathbb{N}$ as the collection of functions that have d_{th} order continuous derivatives on $\mathcal{I}$. For a vector $\mathbf{v} = (v_1, v_2, ..., v_p) \in \mathbb{R}^p$, let $|\mathbf{v}| = (\sum_{i=1}^p v_i^2)^{1/2}$. For a $p \times p$ matrix A, define $|A| = \sqrt{trace(A'A)}$. For a random vector (matrix) $\mathbf{V}$, write $\mathbf{V} \in \mathcal{L}_q$ if $\|\mathbf{V}\|_q := [\mathbb{E}(|\mathbf{V}|^q)]^{1/q} < \infty$. Write $\|\mathbf{V}\| = \|\mathbf{V}\|_2$. Denote by $\Rightarrow$ the weak convergence. For $x \in \mathbb{R}$, define $x^+ = \max(x, 0)$. Let M denotes a finite generic constant which may vary from line to line. Let $K(\cdot) \in \mathcal{C}^2[-1,1]$ be a symmetric kernel function with support $[-1,1]$ and $\int_{-\infty}^{\infty} K(x)dx = 1$. Let $\phi_l = \phi_{l,K} = \int_{\mathbb{R}} x^l K^2(x)dx$, $l = 0, 1, ...$ For a semi-definite matrix Σ with eigen-decomposition $\Sigma = ABA'$, A is orthonormal and B is a diagonal matrix, define $\Sigma^{1/2} = AB^{1/2}A$, where $B^{1/2}$ is a diagonal matrix with each element is the square root of the corresponding element of the diagonal matrix B.

We estimate $(\theta_\tau(t), \dot{\theta}_\tau(t))$ by the local linear quantile estimates, which are defined as

$$(\hat{\theta}_{b_n,\tau}(t), \hat{\dot{\theta}}_{b_n,\tau}(t)) = \operatorname*{argmin}_{b_0,b_1} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i' b_0 - \mathbf{x}_i' b_1(i/n - t)) K_{b_n}(i/n - t), \tag{5}$$

where $\rho_\tau(x) = \tau x^+ + (1-\tau)(-x)^+$. Let $\psi_\tau(x) = \tau - I(x \leq 0)$ be the left derivative of $\rho_\tau(x)$. Define filtration $\mathcal{F}_i = (\eta_{-\infty}, ..., \eta_i)$, $\mathcal{G}_i = (\varepsilon_{-\infty}, ..., \varepsilon_i)$, where $\{\eta_i\}_{i=-\infty}^{\infty}$ and $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ are independent. Both $\mathbf{x}_i$, $e_{i,\tau}$ are assumed to be locally stationary processes (Draghicescu *et al.*, 2009 and Zhou and Wu 2009):

$$\mathbf{x}_i = \mathbf{H}(i/n, \mathcal{G}_i), \;\; e_{i,\tau} = G_\tau(i/n, \mathcal{F}_i, \mathcal{G}_i). \tag{6}$$

where $\mathbf{H} := (H_1, ..., H_p)^T$. Both $\mathbf{H}(\cdot,\cdot)$ and $G_\tau(\cdot,\cdot,\cdot)$ are measurable functions such that $\mathbf{H}(t, \mathcal{F}_i)$ and $G_\tau(t, \mathcal{F}_i, \mathcal{G}_i)$ are well defined for each $t \in [0,1]$. Note that, generally, $\mathbf{x}_i$ and $e_{i,\tau}$ are dependent as they are both functions of $\mathcal{G}_i$. Hence the above setting allows the covariates and errors to be cross-correlated in a very flexible way. For instance, when $G_\tau(t, \mathcal{F}_i, \mathcal{G}_i) = G_\tau(t, \mathcal{F}_i)$, then $\mathbf{x}_i$ and $e_{i,\tau}$ are independent and we have an exogenous model. When $G_\tau(t, \mathcal{F}_i, \mathcal{G}_i) = A_\tau(t, \mathcal{G}_i) B_\tau(t, \mathcal{F}_i)$ for some measurable and well defined functions $A_\tau(\cdot, \mathcal{G}_i)$ and $B_\tau(\cdot, \mathcal{F}_i)$, we have a special heterogenous model with multiplicative heterogeneity.

To precisely describe the temporal dependence structures of the errors and the covariates, we shall introduce the following time series dependence measures:

Definition 1. *Let $\mathcal{F}_i$ be the filtration generated by $\{\eta_j\}_{j=-\infty}^{i}$, where $\{\eta_j\}_{j\in\mathbb{Z}}$ are i.i.d. Let $\{\eta_j'\}_{j\in\mathbb{Z}}$ be an i.i.d. copy of $\{\eta_j\}_{j\in\mathbb{Z}}$. Assume that for all $t \in [0,1]$, $\mathbf{G}(t, \mathcal{F}_i) \in \mathcal{L}_q$, $q > 0$. Define the physical dependence measure for the stochastic system $\mathbf{G}(t, \mathcal{F}_i)$ in $\mathcal{L}_q$ norm as*

$$\delta_q(\mathbf{G}, i) = \sup_{t\in[0,1]} \{\|\mathbf{G}(t, \mathcal{F}_i) - \mathbf{G}(t, \mathcal{F}_i^*)\|_q\}, \tag{7}$$

where $\mathcal{F}_i^* = (\mathcal{F}_{-1}, \eta_0', \eta_1, ..., \eta_{i-1}, \eta_i)$.

By definition, $\delta_q(\mathbf{G}, i)$ measures the functional dependence of output $\mathbf{G}(t, \mathcal{F}_i)$ on the input η_0. The dependence measure of a large class of locally stationary linear or non-linear processes can be calculated directly, see for example Zhou and Wu (2009). We say a locally stationary process is short range dependent (in $\mathcal{L}_q$ norm) if $\sum_{i=1}^{\infty} \delta_q(\mathbf{G}, i) < \infty$. And the dependence measures of $\mathbf{G}(t, \mathcal{F}_i)$ are geometrically decaying (in $\mathcal{L}_q$ norm) if $\delta_q(\mathbf{G}, i) = O(\chi^i)$ for some constant $\chi \in (0, 1)$.

Define $l_n(t) = \max(\lfloor nt - nb_n, 1 \rfloor)$, $s_n(t) = \min(\lfloor nt + nb_n, n \rfloor)$, and $\mathcal{N}_n(t) = \{i \in \mathbb{Z} : l_n(t) \leq i \leq s_n(t)\}$, $\mathcal{N}_n(t-) = \mathcal{N}_n(t) \cap (0, nt]$, and $\mathcal{N}_n(t+) = \mathcal{N}_n(t) \cap (nt, n]$. Let $z_i(t) = y_i - \mathbf{x}_i'\theta_\tau(t) - \mathbf{x}_i'\dot{\theta}_\tau(t)(i/n - t)$, and $\mathbf{w}_{in}(t) = (\mathbf{x}_i', \mathbf{x}_i'(i/n - t)/b_n)' = (w_{in,1}(t), ..., w_{in,2p}(t))'$ where $\mathbf{w}_{in}(t)$ is a $(2p \times 1)$ vector. Write $t_i = i/n$. We make the following assumptions:

A0 For every $\tau \in (0, 1)$, $\theta_\tau(\cdot) \in \mathcal{C}^2[0, 1]$ and $\ddot{\theta}_\tau(\cdot)$ is Lipschitz continuous on $[0, 1]$.

A1 $Q_\tau(G_\tau(t, \mathcal{F}_i, \mathcal{G}_i)|\mathcal{G}_i) =: \inf\{x : \mathbb{P}(G_\tau(t, \mathcal{F}_i, \mathcal{G}_i) \geq x|\mathcal{G}_i) \geq \tau\} = 0$ for $t \in (0, 1)$. $\sup_{t \in (0,1)} \|G_\tau(t, \mathcal{F}_i^*, \mathcal{G}_i^*) - G_\tau(t, \mathcal{F}_i, \mathcal{G}_i)\|_1 = O(\chi^{|i|})$ for some constant $\chi \in (0, 1)$. For $t, s \in (0, 1)$, $\|G_\tau(t, \mathcal{F}_i, \mathcal{G}_i) - G_\tau(s, \mathcal{F}_i, \mathcal{G}_i)\|_v \leq M|t - s|$ for some $v > 1$.

A2 There exists $t_x > 0$, such that $\max_{1 \leq i \leq n} \mathbb{E}(\exp(t_x|\mathbf{x}_i|)) \leq M < \infty$. In addition, $\sup_{t \in (0,1)} \|\mathbf{H}(t, \mathcal{G}_i^*) - \mathbf{H}(t, \mathcal{G}_i)\|_1 = O(\chi^{|i|})$ for some constant $\chi \in (0, 1)$, and for $t, s \in (0, 1)$, $\|\mathbf{H}(t, \mathcal{G}_i) - \mathbf{H}(s, \mathcal{G}_i)\| \leq M|t - s|$.

A3 Denote $F^{(q)}(t, x, \tau|\mathcal{F}_{i-1}, \mathcal{G}_i) = \frac{\partial^q}{\partial x^q}\mathbb{P}(G_\tau(t, \mathcal{F}_i, \mathcal{G}_i) \leq x|\mathcal{F}_{i-1}, \mathcal{G}_i)$. Then for $0 \leq q \leq 2p + 1$, define for any integer s,

$$\delta_s(i - 1, \tau) := \sup_{t \in (0,1), x \in \mathbb{R}} \|F^{(q)}(t, x, \tau|\mathcal{F}_{i-1}, \mathcal{G}_i) - F^{(q)}(t, x, \tau|\mathcal{F}_{i-1}^*, \mathcal{G}_i)\|_s$$

We assume that $\delta_1(i - 1, \tau) = O(\chi^{i-1})$ for some positive constant $\chi < 1$. In addition, $F^{(q)}(t, x, \tau|\mathcal{F}_{i-1}, \mathcal{G}_i)$ are bounded for $t \in (0, 1), x \in \mathbb{R}$. Furthermore, we denote $f(t, x, \tau|\mathcal{F}_{i-1}, \mathcal{G}_i) := F^{(1)}(t, x, \tau|\mathcal{F}_{i-1}, \mathcal{G}_i)$ as the conditional density of $G_\tau(t, \mathcal{F}_i, \mathcal{G}_i)$.

A4 Write $f(t, x, \tau|\mathcal{G}_i) = \frac{\partial}{\partial x}\mathbb{P}(e_{i,\tau}(t) \leq x|\mathcal{G}_i)$. Let $\lambda(t)$ be the smallest eigenvalue of $\Sigma(t) := \mathbb{E}\{f(t, 0, \tau|\mathcal{G}_i)\mathbf{H}(t, \mathcal{G}_i)\mathbf{H}'(t, \mathcal{G}_i)\}$. Then i) $\sup_{t \in (0,1)} \|f(t, 0, \tau|\mathcal{G}_i) - f(t, 0, \tau|\mathcal{G}_i^*)\| = O(\chi^{|i|})$ for some constant $\chi \in (0, 1)$, ii) $\inf_{t \in (0,1)} \lambda(t) \geq \eta > 0$ for some positive constant η. iii)$\sup_{t\in(0,1)} \|\dot{f}(t,0,\tau|\mathcal{G}_i) - \dot{f}(t,0,\tau|\mathcal{G}_i^*)\| = O(\chi^{|i|})$ for some constant $\chi \in (0,1)$, and $\dot{f}(t,0,\tau|\mathcal{G}_0)$ is bounded.

Write $\mathbf{U}_\tau(t,\mathcal{F}_i,\mathcal{G}_i) = \psi_\tau(G(t,\mathcal{F}_i,\mathcal{G}_i))\mathbf{H}(t,\mathcal{G}_i)$. Then $\mathbf{U}_\tau(t_i,\mathcal{F}_i,\mathcal{G}_i) = \psi_\tau(e_i)\mathbf{x}_i$. Let

$$\nu_\tau^2(t) = \sum_{j=-\infty}^{\infty} cov(\mathbf{U}_\tau(t,\mathcal{F}_0,\mathcal{G}_0), \mathbf{U}_\tau(t,\mathcal{F}_j,\mathcal{G}_j)). \tag{8}$$

We have the next assumption:

A5 The smallest eigenvalue of $\nu_\tau^2(t)$ is bounded away from 0 on $[0,1]$.

Condition A0) makes assumptions on the smoothness of the conditional quantile function. A1) and A2) require that the covariates and error processes are stochastic Lipschitz continuous with geometrically decaying dependence measures. A3) will be satisfied if the dependence measures of the derivatives of the errors' conditional densities are also geometrically decaying. Hence (A3) is easy to verify for a large class of locally stationary processes. We refer to Zhou and Wu (2009) for some representative examples for the purpose of brevity. Similar conditions are made in the literature of quantile regression with stationary errors. See, for instance, Wu (2007). For condition A4), i) and iii) assumes that conditioning on the information sets of covariates ($\mathcal{G}_i$), the conditional densities (and their first derivatives w.r.t. time t) of the errors have geometrically decaying dependence measures. Consequently, it is easy to check i), iii) of [A4] for most heteroscedastic models. The ii) of [A4] means that the time-varying quantile design matrix $\Sigma(t)$ is non-degenerate on [0,1], which is rather mild. Condition [A5] is also quite mild. It means that the time-varying long run covariance matrices of the gradient vectors are non-degenerate on [0,1].

3 Theories and Methodologies

In this section, we first establish a uniform Bahadur representation for the local linear quantile estimates. Then we construct SCT for time varying coeffients based on the uniform Bahadur representation. We then further propose our ISDT for the time-varying coefficients as well as the variable selection criterion for regression quantiles under non-stationarity and temporal dependence.

3.1 Uniform Bahadur Representation

Bahadur representation is a useful and popular tool for the asymptotic analysis of quantile estimators since it approximates the estimators by mathematically tractable linear forms. See, for instance, He and Shao (1996), Koenker (2005), Wu (2007), Zhou and Wu (2009) etc. Among them, Zhou and Wu (2009) obtained a uniform Bahadur representation of time varying quantile curves. A major restriction of the technique they used is that it is tailored to the problem of quantile trend estimation. Consequently, the technique they developed cannot be used to derive a Bahadur representation for models with random designs where the covariates are stochastic. Furthermore, it is not suitable for the aim of variable selection in the context of quantile regression with random designs. Before stating the theoretical results, we first introduce the following notation:

Define a $2p \times 2p$ matrix $\Sigma_1(t) = \begin{pmatrix} \Sigma(t) & 0 \\ 0 & \mu_2 \Sigma(t) \end{pmatrix}$, where the $p \times p$ matrix $\Sigma(t)$ is defined in [A4]. Here $\mu_s := \int |x|^s K(x) dx$ for positive integer s. Let

$$\tilde{\beta}_{b_n,\tau}(t) := (\hat{\beta}_{b_n,\tau}(t), \hat{\dot{\beta}}_{b_n,\tau}(t)) = \underset{c_0, c_1}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(z_i(t) - \mathbf{x}_i' c_0 - \mathbf{x}_i' c_1 (i/n - t)/b_n) K_{b_n}(i/n - t). \tag{9}$$

Compare with equation (5), it is easy to see that

$$\hat{\beta}_{b_n,\tau}(t) = \hat{\theta}_{b_n,\tau}(t) - \theta_\tau(t), \;\; \hat{\dot{\beta}}_{b_n,\tau}(t) = b_n(\hat{\dot{\theta}}_{b_n,\tau}(t) - \dot{\theta}_\tau(t)). \tag{10}$$

i.e, $\hat{\beta}_{b_n,\tau}(t)$ is the deviation of the estimated coefficient $\hat{\theta}_{b_n,\tau}(t)$ from the true value $\theta_\tau(t)$. Similar interpretation is applicable to $\hat{\dot{\beta}}_{b_n,\tau}(t)$. In the literature of non-parametric quantile regression with *i.i.d* data, a local Bahadur representation is given by Chaudhuri (1991). The following theorem establishes a global Bahadur representation of our local linear quantile regression with locally stationary covariates and errors under short range dependence:

Theorem 1. *Suppose [A0]-[A4],* $b_n \to 0$, $nb_n^4/\log^8 n \to \infty$, *and* $n^c b_n \to 0$ *for some positive* *constant c. Let* $\mathfrak{T}_n = [b_n, 1 - b_n]$,

$$\sup_{t\in\mathfrak{T}_n} |\Sigma_1(t)\tilde{\beta}_{b_n,\tau}(t) - \frac{\sum_{i=1}^n \psi_\tau(z_i(t))\mathbf{w}'_{in}(t)K_{b_n}(i/n - t)}{nb_n}| \\ = O_p(b_n^3 \log^3 n + \frac{b_n \log^6 n}{\sqrt{nb_n}} + (nb_n)^{-\frac{3}{4}} \log^3 n). \quad (11)$$

Theorem 1 derives the uniform approximation to $\tilde{\beta}_{b_n,\tau}(t)$ in the interval $\mathfrak{T}_n$. The approximation rate established in (11) is sharp enough for most uniform nonparametric inference of $\theta_\tau(\cdot)$. Since $\psi_\tau(z_i(t_i))$ is also a locally stationary process with short range dependence, it can be shown to be uniformly well approximated by certain Gaussian process (Wu and Zhou 2011). Thus uniform non-parametric inference can be carried out by studying the maximal deviation or the $\mathcal{L}^2$ norm of the corresponding Gaussian process which approximates $\frac{\sum_{i=1}^n \psi_\tau(z_i(t))\mathbf{w}'_{in}(t)K_{b_n}(i/n-t)}{nb_n}$.

3.2 Simultaneous Confidence Tube.

Theorem 2 below derives the asymptotic theory for the maximal absolute deviation of $\hat{\theta}_{b_n,\tau}(t)$ from $\theta_\tau(t)$ on (0,1).

Theorem 2. *Suppose that the conditions [A0]-[A5] hold,* $b_n \to 0$, $nb_n^4/\log^8 n \to \infty$, *and* $n^c b_n \to 0$ *for some positive constant c. Then as* $n \to \infty$, *for any* $p \times s (s \le p)$ *full rank matrix* $\mathbf{C}$, *we have*

$$\lim_{n\to\infty} \mathbb{P}\Big[\frac{\sqrt{nb_n}}{\sqrt{\phi_0}} \sup_{t\in\mathfrak{T}_n} |(M_{\mathbf{C}}(t))^{-1} \left\{ \hat{\theta}_{\mathbf{C},b_n,\tau}(t) - \theta_{\mathbf{C},\tau}(t) - \frac{\mu_2 b_n^2 \ddot{\theta}_{\mathbf{C},\tau}(t)}{2} \right\} | - B_K(m^*) \le \\ \frac{u}{\sqrt{2\log m^*}}\Big] = \exp(-2\exp(-u)), \quad (12)$$

where $m^* = 1/b_n$, $M_{\mathbf{C}}(t) = [(\mathbf{C}'\Sigma^{-1}(t))\nu_\tau^2(t)(\mathbf{C}'\Sigma^{-1}(t))']^{1/2}$ *and*

$$B_K(m^*) = \sqrt{2\log m^*} + \frac{\log(C_K) + (s/2 - 1/2)\log\log(m^*) - \log 2}{\sqrt{2\log m^*}}, \quad (13)$$

with

$$C_K = \frac{\left\{\int_{-1}^{1} |K'(u)|^2 du/\phi_0 \pi\right\}^{1/2}}{\Gamma(s/2)}.$$

where $\theta_{\mathbf{C},\tau}(t) := \mathbf{C}'\theta_\tau(t)$, $\ddot{\theta}_{\mathbf{C},\tau}(t) := \mathbf{C}'\ddot{\theta}_\tau(t)$, $\hat{\theta}_{\mathbf{C},b_n,\tau}(t) := \mathbf{C}'\hat{\theta}_{b_n,\tau}(t)$, *respectively.*

Consequently, an asymptotically correct SCT of $\theta_{\mathbf{C},\tau}(t)$ can be constructed as follows. Let $\hat{\ddot{\theta}}_{\mathbf{C},\tau}(t)$ and $\hat{M}_{\mathbf{C}}(t)$ be consistent estimators of $\ddot{\theta}_{\mathbf{C},\tau}(t)$ and $M_{\mathbf{C}}(t)$, respectively. Then we construct an SCT of $\theta_{\mathbf{C},\tau}(t)$ with asymptotic coverage probability $1-\alpha$, $\alpha \in (0,1)$ as follows:

$$\hat{\theta}_{\mathbf{C},b_n,\tau}(t) - \frac{\mu_2 b_n^2 \hat{\ddot{\theta}}_{\mathbf{C},\tau}(t)}{2} + \frac{\sqrt{\phi_0}}{\sqrt{nb_n}}\left[B_K(\frac{1}{b_n}) - \frac{\log[\log(1-\alpha)^{-1/2}]}{\sqrt{2\log(\frac{1}{b_n})}}\right]\hat{M}_{\mathbf{C}}(t)\mathcal{B}_s, \tag{14}$$

where $\mathcal{B}_s = \{\mathbf{u} \in \mathbb{R}^s : |\mathbf{u}| \leq 1\}$ is the $s-$ dimensional unit ball.

For a given function $\theta^o_{\mathbf{C},\tau}(\cdot)$[3], suppose that we are interested in testing

$$H_0 : \theta_{\mathbf{C},\tau}(t) = \theta^o_{\mathbf{C},\tau}(t) \ \forall t \in (0,1), \quad \text{VS} \quad H_1 : \theta_{\mathbf{C},\tau}(t) \neq \theta^o_{\mathbf{C},\tau}(t) \text{ for some } t \in (0,1). \tag{15}$$

We investigate problem (15) by checking whether $\theta^o_{\mathbf{C},\tau}(\cdot)$ can be fully embedded into the SCT. The next corollary considers the local power of the test, which is an instant result of the triangle inequality:

Corollary 1. *Suppose the conditions of Theorem 2 hold. Suppose that* $\theta_\tau(t) = \theta^o_\tau(t) + \gamma_n\eta(t) + o(\gamma_n)$, *where* $\theta^o_\tau(t)$, $\eta(t) \in \mathcal{C}[0,1]$, $o(\gamma_n)$ *is uniform in* t *on [0,1], and* $\gamma_n = o(1/\sqrt{nb_n})$. *Suppose* $\gamma_n = \frac{c_n}{\sqrt{-2nb_n \log b_n}}$ *for some positive real series* c_n, $m^* = 1/b_n$, ${\theta^o}_{\mathbf{C},\tau}(t) = \mathbf{C}'\theta^o_\tau(t)$, *and* $\Upsilon = \inf_{t\in(0,1)} |(M_{\mathbf{C}}^{-1}(t))\mathbf{C}'\eta(t)|$, *define*

$$\Delta_n := \mathbb{P}\Big[\frac{\sqrt{nb_n}}{\sqrt{\phi_0}} \sup_{t\in\mathfrak{T}_n} |(M_{\mathbf{C}}(t))^{-1}\left\{\hat{\theta}_{\mathbf{C},b_n,\tau}(t) - {\theta^o}_{\mathbf{C},\tau}(t) - \frac{\mu_2 b_n^2 \ddot{\theta}_{\mathbf{C},\tau}(t)}{2}\right\}| - B_K(m^*) \leq \frac{u}{\sqrt{2\log m^*}}\Big],$$

[3]Defined in Corollary 1.

Then

$$\lim_{n\to\infty}\left\{\Delta_n - \exp(-2\exp(-u-\frac{c_n\Upsilon}{\sqrt{\phi_0}})) + \exp(-2\exp(u-\frac{c_n\Upsilon}{\sqrt{\phi_0}}))\right\} \leq 0,$$

Consequently $\Delta_n \to_{n\to\infty} 0$ *if* $c_n = \gamma_n\sqrt{-2nb_n \log b_n} \to \infty$.

The corollary provides that our SCT can detect local alternatives at rate $\gamma_n = 1/\sqrt{-2nb_n \log b_n}$. For the $100(1-\alpha)\%$ SCT, the asymptotic local power of the test has lower bound:

$$1-(1-\alpha)^{\exp(-c_n\Upsilon/\sqrt{\phi_0})} + \exp(\frac{4\exp(-c_n\Upsilon/\sqrt{\phi_0})}{\log(1-\alpha)}). \tag{16}$$

By Corollary 1 and the lower bound (16), we see that the asymptotic power of the SCT goes to 1 as $c_n = \gamma_n\sqrt{-2nb_n \log b_n} \to \infty$ and $\gamma_n = o(1/\sqrt{nb_n})$.

3.3 ISDT.

Recall expression (15), the hypothesis we are interested in testing. By Theorem 2, the local linear estimates introduce a bias term which involves $\ddot{\theta}_\tau(\cdot)$. However it is generally highly non-trivial to estimate $\ddot{\theta}_\tau(\cdot)$. Most automatic bandwidth selectors select the bandwidth $b_n \sim cn^{-1/5}$ for some $c > 0$, thus make the bias term non-negligible. Consequently, Neumann and Polzehl (1998) suggested that we can choose undersmoothing bandwidth $b_n = o(n^{-1/5})$, however it is generally unclear how to choose such bandwidth. As a result, we use the following jacknife bias-corrected estimator of $\theta_\tau(t)$:

$$\tilde{\theta}_{b_n,\tau}(t) = 2\hat{\theta}_{b_n/\sqrt{2},\tau}(t) - \hat{\theta}_{b_n,\tau}(t), \tag{17}$$

which is equivalent to use the second order kernel $K^*(x) := 2\sqrt{2}K(\sqrt{2}x) - K(x)$. Then we have,

Theorem 3. *Suppose the conditions of Theorem 2 hold. Suppose* $\theta_\tau(t) = \theta^0_\tau(t) + \varrho_n\eta(t) + o(\varrho_n)$, $\varrho_n = n^{-1/2}b_n^{-1/4}$, *and* $o(\varrho_n)$ *is uniform in* t *on* $[0,1]$. *Assume* $nb_n^3/\log^{12} n \to \infty$, $nb_n^{13/2}\log^8 = o(1)$. *Let* $\pi(t)$ *be a weighting function such that* $\pi(t)$ *is non-negative and Lipschitz continuous on* $t \in [0,1]$. *Let* $T_n = \int_{\mathfrak{T}_n}[M_{\mathbf{C}}^{-1}(t)\mathbf{C}'(\tilde{\theta}_{b_n,\tau}(t) - \theta^0_\tau(t))]'[M_{\mathbf{C}}^{-1}(t)\mathbf{C}'(\tilde{\theta}_{b_n,\tau}(t) -$

$\theta^0_\tau(t))]\pi(t)dt$. *Then*

$$nb_n^{1/2}T_n - \frac{s}{\sqrt{b_n}}K^* \star K^*(0)\int_{\mathfrak{T}_n}\pi(t)dt - \int_{\mathfrak{T}_n}(M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))'(M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))\pi(t)dt \Rightarrow N(0, \Xi^*), \tag{18}$$

where $\Xi^* = 2s\int_{\mathbb{R}}[K^* \star K^*(t)]^2dt\int_0^1 \pi^2(t)dt$, $\star$ *denotes the convolution operator.*

When $\eta \equiv 0$, Theorem 3 establishes the asymptotic null distribution of T_n^*. Theorem 3 also shows that our ISDT test can detect alternatives with the rate $n^{-1/2}b_n^{-1/4}$. Simple calculation shows that the asymptotic power of the ISDT test with level α equals

$$\Phi(\frac{\int_{\mathfrak{T}_n}(M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))'(M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))dt}{\sqrt{\Xi^*}} - z_{1-\alpha}). \tag{19}$$

It is obvious that if $b_n = O(n^{-2/9})$, then T_n can detect alternatives with departure rate $n^{-4/9}$. This rate is optimal in the sense of Ingster (1993) and Lepski and Spokoiny (1999). Since $1/\sqrt{-2nb_n \log b_n} >> \rho_n$, ISDT test is asymptotically more powerful then SCT test. Consequently, in theory, the SCT is more suitable for exploring the overall patten of the regression quantile coefficients, and the ISDT is more appropriate for specification tests of the nonparametric quantile curves. However, note that in practice with moderate sample sizes, by the properties of the $\mathcal{L}^\infty$ and $\mathcal{L}^2$ norms, we expect that the SCT is more powerful when the regression coefficient functions differ from the null in a bumpy way, while ISDT is better when the difference is systematic and even. The accuracy and power of the two tests will be examined by simulations in Section 5.

3.4 Variable Selection

In this section we shall propose an information criterion for time-varying quantile regression models under temporal dependence and non-stationarity. Recall that $\mathbf{x}_i = \{x_{i,1}, ..., x_{i,p}\}$ is a p-dimensional vector. Write $\hat\theta_{b_n,\tau}$ as $\hat\theta$ for short. Define $\mathbf{x}_{i,D} = \{x_{i,1}I(1 \in D), ..., x_{i,p}I(p \in D)\}'$ for any set $D \subset \{1, ..., p\}$. Define the quantile regression variable selection criterion

$$QVC(D) = \log\sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}'_{i,D}\hat\theta_D(i/n)) + \chi_n|D|, \tag{20}$$

where $|D|$ is the cardinality of the set D, $\hat{\theta}_D(\cdot) = (\hat{\theta}_1(\cdot)I(1 \in D), ..., \hat{\theta}_p(\cdot)I(p \in D))'$. and $\chi_n \to 0$ is the penalization parameter which determines the penalization of over fitting. To balance the goodness-of-fit and model complexity, we select a subset D that minimizes $QVC(D)$. Large χ_n leads to less predictors, and vice versa. The following theorem states that our procedure consistently identifies the true set of predictors:

Theorem 4. *Suppose the conditions of Theorem 1 hold. In addition, there exist constant* M_0, η_0, *such that* $\inf_{t\in(0,1),|x|\le M_0} f(t, x|\mathcal{F}_1, \mathcal{G}_0) \ge \eta_0 > 0$. *Suppose* $\{\mathbf{x}\}_{i=1}^n$ *have strictly positive densities. Then for* $\chi_n \to 0$, $(\varphi_n b_n + \frac{\pi_n}{\sqrt{nb_n}})^{-1}\chi_n \to \infty$, *where* $\pi_n = b_n \log^6 n + (nb_n)^{-1/4} \log^3 n + b_n^3\sqrt{nb_n} \log^3 n$, $\varphi_n = \frac{\log^4 n}{\sqrt{nb_n}} + b_n^2 \log n$. *We have, for any* $D \neq D_0$, $\lim_{n\to\infty} \mathbb{P}\{QVC(D) > QVC(D_0)\} = 1$.

Theorem 4 shows that, with appropriately chosen b_n and χ_n, the probability that our procedure select the correct explanatory variables goes to 1 as the sample size increases. The selection of b_n and χ_n will be discussed in Section 4.1. In addition, note that our procedure requires only one realization at one time. In the literature of quantile regression, Lasso type penalization methods have been borrowed from the context of least squares linear regression to select significant covariates for underlying parametric model, see Koenker (2004), Li and Zhu (2005), Wang, Li and Jiang (2007), Wu and Liu (2009) among others. Our flexible, nonparametric procedure is a good complement to the aforemention Lasso type procedures: our nonparametric quantile variable selection (QVC) criterion may provide a quality pilot set of significant control variables in the sense that it reduces the bias brought by misidentifying the model, then our ISDT and SCT test will further provide useful inference of possible candidate parametric models.

4 Implementation

To apply Theorems 2 and 3, we need to consistently estimate the quantity $M_{\mathbf{C}}(t)$ over time t. Recall the definition $M_{\mathbf{C}}(t) = [(\mathbf{C}'\Sigma^{-1}(t))\nu_\tau^2(t)(\mathbf{C}'\Sigma^{-1}(t))']^{1/2}$. We can first estimate $\nu^2(t)$ and $\Sigma(t)$, and then use $\hat{M}_{\mathbf{C}}(t) = [(\mathbf{C}'\hat{\Sigma}^{-1}(t))\hat{\nu}_\tau^2(t)(\mathbf{C}'\hat{\Sigma}^{-1}(t))']^{1/2}$ as an estimator of $M_{\mathbf{C}}(t)$. Recall that $\hat{e}_i(t) = y_i - \mathbf{x}_i'\theta_\tau(i/n) - \mathbf{x}_i'\dot{\theta}_\tau(i/n)(i/n - t)$. Hence $\hat{e}_i = \hat{e}_i(t_i) = y_i - \mathbf{x}_i'\theta_\tau(i/n)$. To estimate $\nu^2(t)$, write $\mathbf{Q}_i = \sum_{j=-m}^m \psi_\tau(e_{i+j}(t_{i+j}))\mathbf{x}_{i+j}$, $\hat{\mathbf{Q}}_i = \sum_{j=-m}^m \psi_\tau(\hat{e}_{i+j}(t_{i+j}))\mathbf{x}_{i+j}$, $\Delta_i = \mathbf{Q}_i\mathbf{Q}_i'/(2m+1)$, $\hat{\Delta}_i = \hat{\mathbf{Q}}_i\hat{\mathbf{Q}}_i'/(2m+1)$. Define $w(t,i) =$

$K_{b_n}(t_i - t)/nb_n$, $\tilde{\nu}_\tau^2(t) = \sum_{i=1}^n w(t,i)\Delta_i$, $\hat{\nu}_\tau^2(t) = \sum_{i=1}^n w(t,i)\hat{\Delta}_i$. The estimation of the long run covariance function for locally stationary, temporary dependent time series has been considered in Zhou and Wu (2010), Zhang and Wu (2012). By the fact that $\mathbf{Q}_i$ are locally stationary and short range dependent, $\hat{e}_i's$ are close to the true $e_i's$, we have

Theorem 5. *Suppose the conditions of Theorem 2 hold. For $\mathfrak{I} = [\gamma_n, 1-\gamma_n] \subset (0,1)$, where $\gamma_n = b_n + (m+1)/n$, $m = O(n^{1/3})$, let $\varphi_n = \frac{\log^4 n}{\sqrt{nb_n}} + b_n^2 \log n$, $m\varphi_n \log n \to 0$, we have*

$$\sup_{t\in\mathfrak{I}} |\hat{\nu}_\tau^2(t) - \nu_\tau^2(t)| = O_p\Big(\sqrt{\frac{m}{nb_n^2}} + \frac{1}{m} + (\frac{m}{n})^{1/4}\log n + b_n^2 + \varphi_n^{1/2} b_n^{-1/2}\log n\Big). \tag{21}$$

When $b_n \sim n^{-1/5}$, $m \sim n^{1/3}$, Theorem 5 theoretically provides the consistency of our proposed estimator at the rate $n^{-1/10}\log^3 n$.

Let $\phi(\cdot)$ be the density function of standard normal. Let c_n be the bandwidth such that $c_n \to 0$, $nc_n \to \infty$. Define $\hat{e}_i(t) = y_i - \mathbf{x}_i'\hat{\theta}(t)$, and

$$\hat{\Sigma}(t) = \frac{1}{nb_n c_n}\sum_{i=1}^n \phi(\frac{\hat{e}_i(t)}{c_n})\mathbf{x}_i\mathbf{x}_i' K_{b_n}(i/n - t). \tag{22}$$

By considering the progressive sandwich-type estimator [4] $\hat{\Sigma}(t)$, we have

Theorem 6. *Suppose the conditions of Theorem 2 hold, and $c_n \to 0$, $nc_n \to \infty$, then*

$$\sup_{t\in\mathfrak{T}_n} |\hat{\Sigma}(t) - \Sigma(t)| = O_p(c_n^2 b_n^{-1}\log^2 n + b_n \log^4 n + \frac{\log^7 n}{\sqrt{nc_n}b_n}) \tag{23}$$

If $b_n \sim n^{-1/5}$, $c_n \sim n^{-1/5}$, then Theorem 4 shows that our proposed estimator is consistent with the convergence rate $n^{-1/5}\log^7 n$.

4.1 Bandwidth Selection.

To perform the foregoing procedures, we need to choose appropriate smoothing parameters b_n, m, and c_n. By the proof of Theorem 2, let $\mathbf{1}$ be the length p identity vector, the minimal

[4]For traditional sandwich estimator of quantile design matrix, uniform kernel is used, for example, see Powell (1991), Cai and Xu (2008). The smoother Gaussian kernel we used here avoids most singularity in sparse regions, and guarantees a satisfactory convergence rate.

asymptotic mean integrated squared error bandwidth for estimating $\theta_\tau(\cdot)$ is

$$b^*_{n,\tau} = \left(\frac{\phi_0 \int_0^1 trace\{M_{\mathbf{1}}(t)\}dt}{\mu_2^2 \int_0^1 |\ddot{\theta}_\tau(t)|^2 dt} \right)^{1/5} n^{-1/5}. \tag{24}$$

Consequently a plug-in method could be applied to evaluating b^*_n by estimate $\ddot{\theta}_\tau(\cdot)$ and $trace(M_{\mathbf{1}}(\cdot))$. However, in practice, the quantity $\ddot{\theta}_\tau(\cdot)$ is hard to estimate. To avoid this difficulty, we adopt another selector, the corrected generalized cross-validation (GCV) method (Craven and Wabba, 1979). In the literature of local linear mean regression, if $\hat{Y} = \mathbf{V}(b)\mathbf{Y}$, and $\mathbf{\Gamma}_n = (\mathbb{E}(e_i e_j))_{1\le i,j\le n}$, GCV method select

$$\hat{b}_{n,mean} = \underset{b}{\operatorname{argmin}}\{GCV(b)\}, \quad GCV(b) = \frac{n^{-1}\{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}'\hat{\Gamma}_n^{-1}\{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}}{(1 - trace(\mathbf{V}(b)/n))^2}. \tag{25}$$

The GCV method works reasonably well for selecting appropriate bandwidths of local linear mean regression, see Zhou and Wu (2010), Zhang and Wu (2012) among others. Note that for local linear mean regression, the bandwidth which minimizes the asymptotic mean integrated squared error is

$$b^*_{n,mean} = \left(\frac{\phi_0 \int_0^1 trace\{\tilde{M}(t)\}dt}{\mu_2^2 \int_0^1 |\ddot{\theta}_{mean}(t)|^2 dt} \right)^{1/5} n^{-1/5}, \tag{26}$$

where $\tilde{M}(t) = \tilde{\Sigma}^{-1}(t)\tilde{\Lambda}(t)\tilde{\Sigma}^{-1}(t)$, $\tilde{\Lambda}(t) = \sum_{-\infty}^{\infty} cov(\mathbf{H}(t,\mathcal{G}_0)e_m(t,\mathcal{F}_0,\mathcal{G}_0), \mathbf{H}(t,\mathcal{G}_j)e_m(t,\mathcal{F}_j,\mathcal{G}_j))$, $e_m(t,\mathcal{F}_i,\mathcal{G}_i)$ is the errors of local linear mean regression, and $\tilde{\Sigma}(t) = \mathbb{E}\{\mathbf{H}(t,\mathcal{G}_0)\mathbf{H}(t,\mathcal{G}_0)'\}$. Similarly to Yu and Jones (1998), which corrected the bandwidth of local linear mean regression for local linear quantile regression with $i.i.d$ data, we assume that $\ddot{\theta}_{mean}(t) \approx \ddot{\theta}_\tau(t)$. Together with (24), (26), we obtain the following relationship:

$$\hat{b}_{n,\tau} = \left(\frac{\int_0^1 trace\{M_{\mathbf{1}}(t)\}dt}{\int_0^1 trace\{\tilde{M}(t)\}dt} \right)^{1/5} \hat{b}_{n,mean}. \tag{27}$$

Note that $M_{\mathbf{1}}(t)$ depends on quantile τ. We refer to Zhou and Wu (2010) for the estimation of $\tilde{M}(t)$. We recommend a banding technique in Wu and Pourahmadi (2009) to estimate Γ_n^{-1}. We select $\hat{b}_{n,mean}$, $\lfloor n^{1/3} \rfloor$, and $\tilde{c}_n$ as pilot bandwidths of $b_{n,\tau}$, m, c_n, respectively, to estimate the quantity $M_{\mathbf{1}}(t)$, where $\tilde{c}_n$ is the minimal volatility (explained below) estimator of c_n by using $\hat{b}_{n,mean}$ for $b_{n,\tau}$ and $\lfloor n^{1/3} \rfloor$ for m. For refinements, we can use $\hat{b}_{n,mean}$ as the pilot bandwidth to get $\hat{b}^0_{n,\tau}$, obtain $\hat{c}_n$, $\hat{m}$ by the minimal volatility method, and then update the estimates of $M_{\mathbf{1}}(t)$ in (27) by using $\hat{b}^0_{n,\tau}$, $\hat{c}_n$, $\hat{m}$ to get finer $\hat{b}_{n,\tau}$ until convergence. Since the Jackknife technique in Section 3.3 reduces the bias, we recommend $\tilde{b}_{n,\tau} = 2\hat{b}_{n,\tau}$ as the bandwidth selected for the bias-corrected estimator. For ISDT, we recommend $\tilde{b}_{n,\tau,ISDT} = \tilde{b}_{n,\tau,SCT} \times n^{-1/45}$ and $\tilde{b}_{n,\tau,SCT} = \tilde{b}_{n,\tau}$, by the similar argument of Zhou (2010).

For block size m, which is needed for estimating the long run variance of the gradient vectors, as an easy implementation, we could choose $m = \lfloor n^{1/3} \rfloor$. For refinements, we recommend the extended minimum volatility method suggested by Zhou and Wu (2010). We also apply the minimum volatility method to obtain c_n. More specifically, let the grid of possible bandwidths be $\{c_1, .., c_k\}$. Together with $\hat{b}_{n,\tau}$ and $\hat{m}$, we calculate the estimated matrices $\hat{M}_1(t)$,..., $\hat{M}_k(t)$, respectively. For positive integer s, say $s = 7$, define

$$ise[\{\hat{M}^{a,b}_l(t)\}^s_{l=1}] = \{\frac{1}{s-1} \sum_{l=1}^{s} |\hat{M}^{a,b}_l(t) - \sum_{l=1}^{s} \hat{M}^{a,b}_l(t)/s|^2\}^{1/2}. \qquad (28)$$

where $\hat{M}^{a,b}_l(t)$ represents the $(a, b)_{th}$ entry of the matrix $\hat{M}_l(t)$. Define

$$ise[\{\hat{M}_l\}^s_{l=1}] = \int_0^1 \{\sum_{a=1}^{n} \sum_{b=1}^{n} ise[\{\hat{M}^{a,b}_l(t)\}^s_{l=1}]^2\}^{1/2} dt.$$

Then we choose $c_{l+\lfloor s/2 \rfloor}$ as our bandwidth $\hat{c}_n$ if $ise\{\hat{M}_l\}^s_{l=1} \le ise[\{\hat{M}_j\}^s_{j=1}]$ for $1 \le j \le k - s + 1$. Simulation shows that the performance of our procedure is reasonably well and is not sensitive to (m, c_n), as long as the pair is not very different from what is chosen by the minimum volatility method.

Now we discuss the parameter tuning in the procedure of variable selection. For χ_n, we simply choose $\hat{\chi}_n = n^{-2/5}$ which is recommended as a rule of thumb in Zhang and Wu (2012) when $b_{n,\tau}$ is of the order $n^{-1/5}$. The choice performs reasonably well in our simulation studies. Due to the model uncertainty, the choice of $b_{n,\tau}$ becomes further complicated. As an easy implementation, similar to the choice of bandwidth in (27), we can choose $b^*_{n,\tau}$ by the corrected GCV method with all possible candidate predictors as our bandwidth. For refinements, we then choose the bandwidth by further applying the corrected GCV method to the predictors selected by the $b^*_{n,\tau}$.

4.2 The bootstrap

It can be shown that the convergence rates of (13) and (18) are very slow. To circumvent this problem, we propose the following bootstrap method which substantially improve the finite sample performance, see Härdle and Marron (1991); Hall (1991); Neumann and Kreiss (1998). By the Gaussian approximation to the gradient vectors, we have the following proposition:

Proposition 1. *Assume the conditions of Theorem 3 hold and* $nb_n^6 = o(1)$, $nb_n^4/\log^8 n \to \infty$*. Then on a richer probability space, there exist i.i.d.* $\mathbf{V}_1, \cdots, \mathbf{V}_n \sim N(0, Id_s)$*, such that*

$$\sup_{t\in\mathfrak{T}_n} |\tilde{\theta}_{\mathbf{C},b_n,\tau}(t) - \theta_{\mathbf{C},\tau}(t) - M_{\mathbf{C}}(t)\sum_{i=1}^n \mathbf{V}_i K^*_{b_n}(t_i - t)/(nb_n)| = O_p(\frac{\log^2 n}{n^{3/4}b_n}), \quad (29)$$

$$T_n - \int (\sum_{i=1}^n \mathbf{V}_i K^*_{b_n}(t_i - t)/(nb_n))'(\sum_{i=1}^n \mathbf{V}_i K^*_{b_n}(t_i - t)/(nb_n))\pi(t)dt = O_p(\frac{\log^2 n}{n^{5/4}b_n^{3/2}}), \quad (30)$$

where T_n *is defined in Theorem 3.*

The proposition shows that we can well approximate $\sup_{t\in\mathfrak{T}_n} |M^{-1}_{\mathbf{C}}(t)(\tilde{\theta}_{\mathbf{C},b_n,\tau}(t) - \theta_{\mathbf{C},\tau})|$ and T_n by $\sup_{t\in\mathfrak{T}_n} |\sum_{i=1}^n \mathbf{V}_i K^*_{b_n}(t_i - t)/(nb_n)|$ and its integrated squared forms. The distribution of the latter quantities can be further bootstrapped by generating a large number of i.i.d. copies of $\sum_{i=1}^n \mathbf{V}^\sharp_i K^*_{b_n}(t_i - t)/(nb_n)$, where $\{\mathbf{V}^\sharp_i\}_{i\in\mathbb{Z}}$ are *i.i.d.* $N(0, Id_s)$. Based on the discussion of the Proposition 1, we summarize our simulation based test procedures as follows:

a Using methods in Section 4.1, find appropriate $\hat{b}_{n,\tau}$ to estimate $\theta_{\mathbf{C},\tau}(\cdot)$ for the pre-specified quantile τ. Let $\tilde{b}_{n,\tau} = 2\hat{b}_{n,\tau}$ for SCT, and $\tilde{b}_{n,\tau,ISDT} = \tilde{b}_{n,\tau,SCT} \times n^{-1/45}$ for ISDT. Calculate $\tilde{\theta}_{\mathbf{C},\tilde{b}_{n,\tau},\tau}(t)$ by (17) for SCT, ISDT via the associate bandwidths $\tilde{b}_{n,\tau}$, $\tilde{b}_{n,\tau,ISDT}$, respectively.

b Calculate $\hat{M}_{\mathbf{C}}(t)$ by Theorem 5 and Theorem 6 with the bandwidths chosen by the methods discussed in Section 4.1.

c Generate i.i.d Gaussian vectors $N(0, Id_s)$ $\mathbf{V}_1$,...,$\mathbf{V}_n$, write $\tilde{b}^*_{n,\tau} = \tilde{b}_{n,\tau,ISDT}$

$$A := \sup_{t\in\mathfrak{T}_n} |\sum_{i=1}^n \mathbf{V}_i K^*_{\tilde{b}_{n,\tau}}(i/n - t)/(n\tilde{b}_{n,\tau})|,$$

$$B := |\int_{t\in\mathfrak{T}_n} (\sum_{i=1}^n \mathbf{V}_i K^*_{\tilde{b}^*_{n,\tau}}(i/n - t)/(n\tilde{b}^*_{n,\tau}))'(\sum_{i=1}^n \mathbf{V}_i K^*_{\tilde{b}^*_{n,\tau}}(i/n - t)/(n\tilde{b}^*_{n,\tau}))\pi(t)dt|,$$

for given weighting function $\pi(t)$.

d For given test level α, repeat step (c) for 2000 (say) times, to obtain the estimated $(1-\alpha)_{th}$ quantiles $\hat{q}_{A,1-\alpha}$, $\hat{q}_{B,1-\alpha}$ of A, B, respectively.

e Construct the $(1-\alpha)_{th}$ SCT of $\theta_{\mathbf{C},\tau}(t)$ as $\tilde{\theta}_{\mathbf{C},\tilde{b}_{n,\tau},\tau} + \hat{M}_{\mathbf{C}}(t)\hat{q}_{A,1-\alpha}\mathcal{B}_s$. The simulated level α critical value for ISDT with weighting function $\pi(t)$ could be constructed as $\hat{q}_{B,1-\alpha}$.

5 Simulations.

5.1 Type I error rates.

Let $a(t) = 1/2 - (t - 1/2)^2$, $b(t) = 1/2 - t/2$, $c(t) = 1/4 + t/2$. Define $e_i = e(t_i)$, where $e(t) = \sum_{j=0}^\infty a(t)^j \zeta_{i-j}/4$, and $x_{i,1} = x_{i,1}(t_i)$, $x_{i,2} = x_{i,2}(t_i)$, where $x_{i,1}(t) = \sum_{j=0}^\infty b^j(t)\epsilon_{i-j}$, $x_{i,2}(t) = \sum_{j=0}^\infty c^j(t)\eta_{i-j}$, and $\epsilon_i = (\eta_i + \varepsilon_i)/\sqrt{2}$, $\{\zeta_i\}_{i=-\infty}^\infty$, $\{\eta_i\}_{i=-\infty}^\infty$ and $\{\varepsilon_i\}_{i=-\infty}^\infty$ are *i.i.d* standard normals. Write $\theta_0(i/n) = \sin(\frac{2\pi i}{n})$, $\theta_1(i/n) \equiv 0.5$, $\theta_2(i/n) = 2\log(1 + 2i/n)$ and $\theta_3(i/n) = exp(-(i/n - 0.5)^2)$. Let $\mathcal{G}_i = (\eta_{-\infty}, \varepsilon_{-\infty}, ..., \eta_i, \varepsilon_i)$, $\mathcal{F}_i = (\zeta_{-\infty}, ..., \zeta_i)$, and $e_{i,\tau}$ be the τ_{th} quantile of e_i. By our construction, $e_{i,\tau} = \tilde{G}(i/n)$ for some smooth function $\tilde{G}(\cdot)$. Consider the following two models:

I $y_i = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2} + e_i$. In this model, the covariates and the errors are independent. The τ_{th} conditional quantile has the form $Q_\tau(y_i|\mathbf{x}_i) = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2} + e_{i,\tau}$.

II Consider the heteroscedastic model

$$y_i = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2} + \sqrt{1 + x_{i,1}^2 + x_{i,2}^2}(e_i - e_{i,\tau})/\sqrt{3}.$$

We are interested in the τ_{th} conditional quantile, which has the form $Q_\tau(y_i|\mathbf{x}_i) = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2}$. We generate 3 sets of data for quantile 0.5, 0.7, 0.8 of model II.

We first simulate the type I error rates under two null hypotheses, $H_0 : \theta_1(t) \equiv 2$ and $H_0 : \theta_2(t) = 2\log(1+2t)$. We choose the weight function $\pi(t) \equiv 1$. For both of model I and model II, we perform the SCT and ISDT tests for three different quantiles 0.5, 07, 0.8, respectively.

Table 1 and Table 2 are the simulated type I error rates of testing $\theta_1(\cdot)$, $\theta_2(\cdot)$ via SCT, in % with nominal levels α=5%, 10% for models I and II, respectively. b is the bandwidth b_n chosen by procedures listed in Section 4.1. The sample size is 500. In our simulation, $b = 0.134$, 0.136, 0.149, for 0.5, 0.7, 0.8 quantiles of model I, and $b = 0.117$, 0.116, 0.117, for 0.5, 0.7, 0.8 quantiles of model II, respectively.

Table 3 and Table 4 are the simulated type I error rates of $\theta_1(\cdot)$, $\theta_2(\cdot)$, in % for model I and II with nominal levels α=5%, 10% for the ISDT test, respectively. The sample size is 500. b is the bandwidth b_n chosen by the relationship $\tilde{b}_{n,\tau,ISDT} = \tilde{b}_{n,\tau,SCT} \times 500^{-1/45}$ as mentioned in Section 4.1.

The simulation results show that, under the null hypotheses, the empirical type I error rates of the SCT and ISDT tests for different regression quantiles in models I and II are fairly close to the nominal levels (90%, 95%). In addition, the results are not sensitive to the choice of bandwidths as long as the bandwidth used is not very different from the bandwidth chosen by the procedure in Section 4.1.

5.2 Power

We examine the power of the two tests via simulations under two pairs of hypotheses, which correspond to the drift and the dispersion, respectively. The sample size of our simulation is 500. The data generating mechanism is

$$y_i = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2} + \sqrt{1 + x_{i,1}^2 + x_{i,2}^2}(e_i - e_{i,\tau})/\sqrt{3}, \qquad (31)$$

where $\theta_0(i/n)$, $\theta_2(i/n)$ are defined in Section 5.1. The quantile which we examine in the simulations is 0.5. Let δ be a positive constant. The first pair of hypothesis (case I), which is related to the drift, are $H_0 : \theta_1(t) \equiv 0.5$ vs $H_1 : \theta_1(t) = 0.5 + \delta$. The second pair of hypothesis (case II), which is related to the dispersion, are $H_0 : \theta_1(t) \equiv 1$ vs $H_1 : \theta_1(t) = 0.5 + 0.5\exp(-\delta(t-0.5)^2)$. The power of the two tests should increase with δ. We plot the unadjusted power as a function of δ for case I and case II in Figure 1 and Figure 2, respectively. The plots show that in both cases, the power goes to 1 as δ increases. As expected, the ISDT is more powerful in case I (which is related to the drift) while SCT is better in case II (which is related to characteristics of the norms).

5.3 Performance of QVC

In this section we shall carry out a simulation study to examine the performance of our information criterion QVC. Let $\{x_{i,3}\}_{i=-\infty}^{\infty}$ be *i.i.d.* $\chi^2(3)/3$, $\{x_{i,4}\}_{i=-\infty}^{\infty}$ be *i.i.d.* $\chi^2(4)/4$, and $x_{i,5}(t) = \sum_{j=0}^{\infty} d(t)^j R_{i-j}$, where $\{R_i\}_{i=-\infty}^{\infty}$ are *i.i.d.* Rademacher random variables, $d(t) = 0.5(5t^3 - 3t)$. Let $x_{i,5} = x_{i,5}(t_i)$. In Table 5 and Table 6, we generate data from models I and II with sample size 200 and 500, respectively, and use QVC to estimate the set of relevant predictors. For any $1 \le j \le 5$, write $\mathbf{x}_j = \{x_{i,j}\}_{i=1}^{n}$. The whole set of potential predictors are $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_4$, $\mathbf{x}_5$. In Table 7 we consider the following model III, also with sample size 200 and 500:

III Consider the heteroscedastic model

$$y_i = \theta_0(i/n) + \theta_1(i/n)x_{i,1} + \theta_2(i/n)x_{i,2} + \theta_3(i/n)x_{i,3} + \sqrt{1 + x_{i,1}^2 + x_{i,2}^2}(e_i - e_{i,\tau})/\sqrt{3}.$$

The τ_{th} conditional quantile has the form $Q_\tau(y_i|\mathbf{x}_i) = \theta_0(i/n)+\theta_1(i/n)x_{i,1}+\theta_2(i/n)x_{i,2}+\theta_3(i/n)x_{i,3}$, where $\theta_1(\cdot)$, $\theta_2(\cdot)$, $\theta_3(\cdot)$ are defined in Section 5.1.

For model III, the whole set of the potential predictors are $\{\mathbf{x}_j\}_{1\le j\le 5}$.

In Tables 5-7, we categorize the simulation results into correct, under-fitting [5] and over-fitting [6]. We can see from Tables 5-7 that the information criterion QVC performs quite well and the performances are not sensitive to the choice of the bandwidths. As expected, the performance is better when the quantile considered is less extreme or the sample size is larger.

[5]Under-fitting: at least one relevant predictor is missing.

[6]Over-fitting: at least one irrelevant predictor is contained without under-fitting.

Table 1: SCT: Simulated type I error rates of $\theta_1(\cdot)$, $\theta_2(\cdot)$, in % for model I with nominal level α=5%, 10%. b is the bandwidth b_n chosen by procedures listed in Section 4.1, and is present in Section 5.1.

	$\tau = 0.5$				$\tau = 0.7$				$\tau = 0.8$			
	θ_1		θ_2		θ_1		θ_2		θ_1		θ_2	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$b/1.25$	4.6	9.55	5.25	10.15	5.45	10.3	5.9	10.55	7.75	14.65	7.7	13.1
b	6.6	14.15	6.6	14	5.45	9.65	5.85	9.6	6.95	12.15	6.85	11.65
$1.25b$	6.75	13.3	6.7	12.55	4.35	9.45	5.4	9.85	6.35	12.2	7.75	12.25

Table 2: SCT: Simulated type I error rates of $\theta_1(\cdot)$, $\theta_2(\cdot)$, in % for model II with nominal level α=5%, 10%. b is the bandwidth b_n chosen by procedures listed in Section 4.1, and is present in Section 5.1.

	$\tau = 0.5$				$\tau = 0.7$				$\tau = 0.8$			
	θ_1		θ_2		θ_1		θ_2		θ_1		θ_2	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$b/1.25$	4.1	8.65	4.85	8.75	5.05	9.9	4.35	8.3	6.05	10.85	6.05	9.85
b	7.3	14.7	6.4	13.25	6.2	13.9	7.7	13.7	7.95	14.05	6.3	12.9
$1.25b$	3.95	9.45	5.2	10.25	5.6	11.05	5.55	10.25	6.75	12.05	5.55	12.9

Table 3: ISDT: Simulated type I error rate of $\theta_1(\cdot)$, $\theta_2(\cdot)$, in % for model I with nominal level α=5%, 10%. b is the bandwidth b_n chosen by procedures listed in Section 4.1, and is present in Section 5.1.

	$\tau = 0.5$				$\tau = 0.7$				$\tau = 0.8$			
	θ_1		θ_2		θ_1		θ_2		θ_1		θ_2	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$b/1.25$	4.75	10.3	5.6	12.0	4.75	9.9	5.0	10.25	5.0	11.45	5.85	11.8
b	3.5	9.4	4.6	10.2	4.0	11.55	4.7	11.75	6.3	14.0	5.45	12.3
$1.25b$	4.8	12.75	4.5	11.85	5.55	10.95	6.4	12.65	5.45	12.1	6.15	13.05

Table 4: ISDT: Simulated type I error rate of $\theta_1(\cdot)$, $\theta_2(\cdot)$, in % for model II with nominal level α=5%, 10%. b is the bandwidth b_n chosen by procedures listed in Section 4.1, and is present in Section 5.1.

	$\tau = 0.5$				$\tau = 0.7$				$\tau = 0.8$			
	θ_1		θ_2		θ_1		θ_2		θ_1		θ_2	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$b/1.25$	6.15	12.5	5.55	11.5	6.0	12.1	5.2	10.85	6.8	14.15	6.5	12.55
b	5.3	12.0	4.75	10.5	5.2	11.05	4.85	10.55	5.4	11.65	5.95	11.95
$1.25b$	5.35	11.65	4.35	10.55	4.75	10.55	3.95	10.95	3.8	11.2	4.75	10.15

Table 5: Percentage of under-fitted, correctly fitted and over-fitted models selected by QVC of Section 3.4 for Model I and II, sample size $n = 200$ and quantiles 0.5, 0.7, 0.9. b is the bandwidth b_n chosen by procedures listed in Section 4.1. $b = 0.210, 0.215, 0.233$ for 0.5, 0.7, 0.9 quantiles of Model I, and $b = 0.186, 0.196, 0.197$ for 0.5, 0.7, 0.9 quantiles of Model II.

	$\tau = 0.5$			$\tau = 0.7$			$\tau = 0.9$		
Model I	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	100	0.00	0.00	100	0.00	0.00	95.55	4.45
b	0.00	100	0.00	0.00	100	0.00	0.00	98.3	1.7
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	98.5	1.5
Model II	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	99.9	0.10	0.00	99.7	0.30	0.00	92.75	7.15
b	0.00	100	0.00	0.00	100	0.00	0.00	97.7	2.3
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	98.7	1.3

Table 6: Percentage of under-fitted, correctly fitted and over-fitted models selected by QVC of Section 3.4 for Model I and II, sample size $n = 500$ and quantiles 0.5, 0.7, 0.9. b is the bandwidth b_n chosen by procedures listed in Section 4.1. $b = 0.162, 0.164, 0.180$ for 0.5, 0.7, 0.9 quantiles of Model I, and $b = 0.137, 0.139, 0.154$ for 0.5, 0.7, 0.9 quantiles of Model II.

	$\tau = 0.5$			$\tau = 0.7$			$\tau = 0.9$		
Model I	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
b	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
Model II	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	100	0.00	0.00	100	0.00	0.00	99.95	0.05
b	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00

Table 7: Percentage of under-fitted, correctly fitted and over-fitted models selected by QVC of Section 3.4 for Model III, sample sizes $n = 200$, $n = 500$ and quantiles 0.5, 0.7, 0.9. b is the bandwidth b_n chosen by procedures listed in Section 4.1. When $n = 200$, $b = 0.201, 0.201, 0.216$ for 0.5, 0.7, 0.9 quantiles. When $n = 500$, $b = 0.139, 0.151, 0.160$ for 0.5, 0.7, 0.9 quantiles.

	$\tau = 0.5$			$\tau = 0.7$			$\tau = 0.9$		
n=200	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	100	0.00	0.00	99.65	0.35	0.00	86.8	13.2
b	0.00	100	0.00	0.00	100	0.00	0.00	93.9	6.1
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	97.45	2.55
n=500	under	correct	over	under	correct	over	under	correct	over
$b/1.25$	0.00	100	0.00	0.00	100	0.00	0.00	99.25	0.75
b	0.00	100	0.00	0.00	100	0.00	0.00	99.95	0.05
$1.25b$	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00

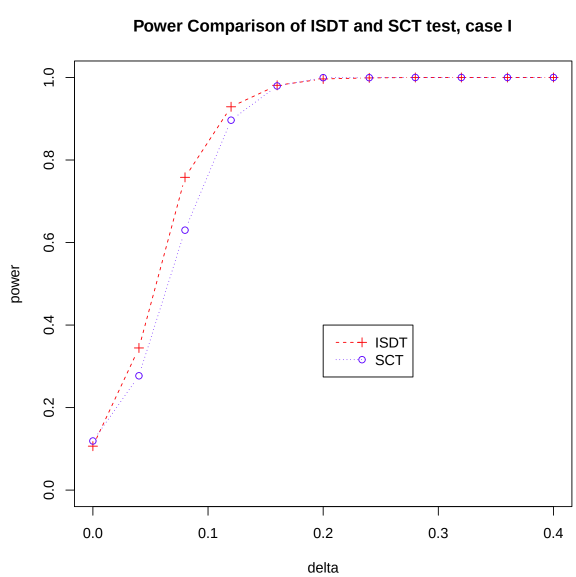

Figure 1: Power Comparison for SCT and ISDT, case I

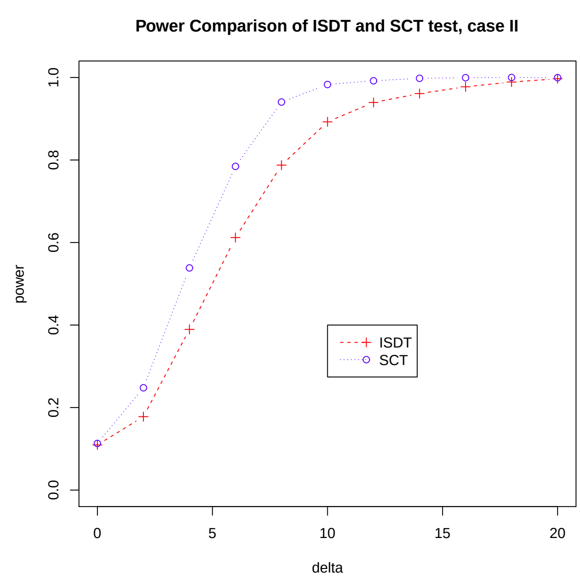

Figure 2: Power Comparison for SCT and ISDT, case II

6 Data Analysis.

In this section, we apply our test to the analysis of an unemployment rate time series. The response of the unemployment rate to the expansionary or contractionary shocks are asymmetric, which has important implications in economic policy. See Koenker and Xiao (2006).

We consider seasonally adjusted quarterly U.S. unemployment rate (1948:1 to 2014:2). The data can be downloaded from http://research.stlouisfed.org/. Three quantiles are considered: 0.2, 0.5, 0.8. We model the three conditional quantiles of the time series with time varying quantile AR models. We use Epanechnikov Kernel in our analysis. Note that for the quarterly unemployment rate data, Koenker and Xiao (2006) shows that there is no evidence of unit root for the three conditional quantiles considered in our example.

We first apply the information criterion QVC from Section 3.4 to select the appropriate lags of the regressions. Let x_t be quarter $t's$ unemployment rate. The set of potential predictors is taken to be $\{x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}\}$. The bandwidth parameter $\hat{b}_n = 0.214, 0.232, 0.246$ for $\tau = 0.2, 0.5, 0.8$, and $\chi_n = 0.108$ respectively. We pick $\{x_{t-1}, x_{t-2}\}$ as the regressors for all three quantiles 0.2, 0.5, 0.8, which coincides with Koenker and Xiao (2006).

We perform our ISDT and SCT tests to check whether the coefficients of x_{t-2} for the three quantiles are zero. For $\tau = 0.2$, the test statistics of ISDT and SCT are $0.78,\ 1.92$, respectively, while the associate 5% nominal critical values of the two tests are $0.029,\ 0.45$, respectively. For $\tau = 0.5$, the ISDT and SCT test statistics are 0.76, 3.39, respectively, and the the associate 5% nominal critical values of the two tests are $0.022,\ 0.43$. For $\tau = 0.8$, the ISDT and SCT test statistics are 1.25, 3.33, respectively, and the the associate 5% nominal critical values of the two tests are $0.030,\ 0.41$. Hence for the three quantiles considered, the coefficients of x_{t-2} are statistically non-zero.

We then examine whether a time-invariant linear model is appropriate for the three quantiles. This amounts to testing whether all coefficients of the model are constant over time. The test results are summarized in Table 8. By Table 8, for the 0.2 and 0.5 quantiles, the time-invariant linear model is rejected by both the ISDT and SCT tests, while for $\tau = 0.8$, both tests fail to reject the null hypothesis. Koenker and Xiao (2006) applies time invariant quantile auto-regression model to studying the unemployment rate.

Table 8: Tests of parameter constancy, quarterly unemployment rate

	test statistics		95% C.V.		p-value	
	ISDT	SCT	ISDT	SCT c.v.	ISDT	SCT
$\tau = 0.2$	0.077	0.73	0.065	0.57	0.0138	< 0.001
$\tau = 0.5$	0.094	1.12	0.049	0.54	< 0.001	< 0.001
$\tau = 0.8$	0.020	0.39	0.047	0.52	0.602	0.506

They suggest that the response of the unemployment rate to different shocks is asymmetric in the sense that the AR coefficients are not constant over different quantiles, while our analysis implies that the AR coefficients for low and mid quantiles may not only vary among different quantiles, but also vary with time. In addition, our result also suggests that the dynamics of unemployment rate at high quantile can be captured by time-invariant linear models, which is very different from the dynamic structures of unemployment rate when quantiles are low and mid. In other words, neither time-varying conditional mean models nor time-invariant conditional quantile models can be used to model the dynamics of all conditional quantiles of the unemployment rate solely. Such asymmetric behavior that different quantiles need to be fitted by different models cannot be detected by the procedures proposed in Koenker and Xiao (2006).

7 Appendix: Proofs.

In the proofs, the symbol C represents a finite constant which may vary from place to place. We shall omit the subscript τ if there is no confusion caused. Let $\beta = (\beta_0', \beta_1')' = (\beta_{01}, ..., \beta_{0p}, \beta_{11}, ..., \beta_{1p})'$ be a $(2p\times 1)$ vector. Write $\eta_i(t, \beta) = \psi(z_i(t) - \mathbf{w}_{in}'(t)\beta)K_{b_n}(i/n - t)$, $S_n(t, \beta) = \sum_{i=1}^n \eta_i(t, \beta)\mathbf{w}_{in}(t)$. Write

$$\tilde{K}_n(t, \beta) := S_n(t, \beta) - \mathbb{E}(S_n(t, \beta)|\mathcal{G}_n) := M_n(t, \beta) + N_n(t, \beta), \tag{32}$$

$$M_n(t, \beta) = \sum_{i=1}^n \eta_i(t, \beta)\mathbf{w}_{in}(t) - \mathbb{E}(\eta_i(t, \beta)|\mathcal{F}_{i-1}, \mathcal{G}_n)\mathbf{w}_{in}(t), \tag{33}$$

$$N_n(t, \beta) = \sum_{i=1}^n \mathbb{E}(\eta_i(t, \beta)|\mathcal{F}_{i-1}, \mathcal{G}_n)\mathbf{w}_{in}(t) - \mathbb{E}(\eta_i(t, \beta)|\mathcal{G}_n)\mathbf{w}_{in}(t). \tag{34}$$

Let $\lceil a \rceil = \min\{k \in \mathbb{Z} : k \geq a\}$ and $\lfloor a \rfloor = \max\{k \in \mathbb{Z} : k \leq a\}$, $a \in \mathbb{R}$, be the usual ceiling and floor functions, and the projection operator $\mathcal{P}_{j,n}(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_j, \mathcal{G}_n) - \mathbb{E}(\cdot|\mathcal{F}_{j-1}, \mathcal{G}_n)$. For any filtration $\mathcal{G}_i$ generated by some innovations $\{\eta_i\}_{i\in\mathbb{Z}}$, write $\mathcal{G}_i^j = (\eta_{-\infty}, ..., \eta_{j-1}, \eta_j', \eta_{j+1}, ..., \eta_i)$, where $\{\eta_i'\}_{i\in\mathbb{Z}}$ are independent copies of $\{\eta_i\}_{i\in\mathbb{Z}}$.

Proposition 2. *Let $\Upsilon_n(t)$ be a sequence of random variables and be once differentiable in t, $t \in [0,1]$. Let p be a positive constant such that $p \geq 1$. Then if for any $t \in [0,1], \|\Upsilon_n(t)\|_p = O(m_n), \|\dot{\Upsilon}_n(t)\|_p = O(l_n)$, where m_n, l_n are sequences of real numbers, $m_n = O(l_n)$, then $\|\sup_{t\in[0,1]} |\Upsilon_n(t)|\| = O(m_n(\frac{m_n}{l_n})^{-\frac{1}{p}})$. In particular, if $p = 2$, we have $\|\sup_{t\in[0,1]} |\Upsilon_n(t)|\| = O(\sqrt{m_n l_n})$.*

Proof. Define b_n as a number sequence, let $\tau_i = ib_n$, $i = 1, 2, ..., \tilde{b}_n$ and $\tau_i = 1$ for i=$\tilde{b}_n + 1$, where $\tilde{b}_n = \lfloor {b_n}^{-1} \rfloor$. Then by triangle inequality, we have

$$\sup_{t\in(0,1)} |\Upsilon_n(t)| \leq \max_{0\leq i\leq \tilde{b}_n+1} |\Upsilon_n(t_i)| + \max_{1\leq i\leq \tilde{b}_n+1} Z_{in},$$

where $Z_{in} = \sup_{\tau_i - b_n < t < \tau_i} |\Upsilon_n(t) - \Upsilon_n(\tau_i)|$. By $\|Z_{in}\|_p \leq \int_{\tau_i - b_n}^{\tau_i} \|\dot{\Upsilon}_n(t)\|_p dt = O(b_n l_n)$ and $\max_{1\leq i\leq \tilde{b}_n+1} Z_{in}^p \leq \sum_{i=1}^{\tilde{b}_n+1} Z_{in}^p$, we have

$$\|\max_{1\leq i\leq \tilde{b}_n+1} Z_{in}\|_p = O((l_n^p b_n^{(p-1)})^{1/p}) = O(l_n b_n^{(p-1)/p}).$$

Similarly, we have

$$\|\max_{0\leq i\leq \tilde{b}_n+1} |\Upsilon_n(t_i)|\| = O_p(b_n^{-1/p} m_n).$$

We just pick $b_n = m_n/l_n$ to finish the proof. □

Lemma 1. *Assume [A0],[A1],[A2],[A3], $\gamma_n \to 0$, $b_n \to 0$, $nb_n^2 \to \infty$, then* $\sup_{|\beta|\leq\gamma_n} |M_n(t,\beta) - M_n(t,0)| = O_p(\sqrt{nb_n\gamma_n} \log n + n^{-3})$.

Proof. Let $w_{in,l}(t)$ be the l_{th} component of $\mathbf{w}_{in}(t)$. Write

$$M_{nl}(t,\beta) = \sum_{i=1}^{n}[\eta_i(t,\beta) - \mathbb{E}(\eta_i(t,\beta)|\mathcal{F}_{i-1},\mathcal{G}_n)]w_{in,l}(t), 1 \le l \le p,$$

$$M_{nl1}(t,\beta) = \sum_{i\in\mathcal{N}_n(t+)}^{n}[\eta_i(t,\beta) - \mathbb{E}(\eta_i(t,\beta)|\mathcal{F}_{i-1},\mathcal{G}_n)]w_{in,l}(t), p+1 \le l \le 2p,$$

$$M_{nl2}(t,\beta) = \sum_{i\in\mathcal{N}_n(t-)}^{n}[\eta_i(t,\beta) - \mathbb{E}(\eta_i(t,\beta)|\mathcal{F}_{i-1},\mathcal{G}_n)]w_{in,l}(t), p+1 \le l \le 2p.$$

It suffices to show Lemma 1 with $M_n(t,\beta)$ therein replaced by $M_{nl}(t,\beta)$ for $1 \le l \le p$ and $M_{nl1}(t,\beta)$, $M_{nl2}(t,\beta)$ for $p+1 \le l \le 2p$. We consider the case of M_{nl1}, $l = p+1$ for presentation clarity. Let g_n be a real sequence which goes to infinity arbitrarily slow with $g_n \ge 3$ for all n, and $u_n = (nb_n\gamma_n)^{1/2}g_n/\log g_n$, $\phi_n = (nb_n\gamma_n)^{1/2}g_n\log n$, $l = p+1$,

$$A_n(t) = \max_{i\in\mathcal{N}_n(t+)}\sup_{|\beta|\le\gamma_n}|(\eta_i(t,\beta) - \eta_i(t,0))w_{in,l}(t)|,$$

$$U_n(t) = \sum_{i\in\mathcal{N}_n(t+)}\mathbb{E}\{(\psi(z_i(t) + |\mathbf{w}_{in}(t)||\gamma_n|)$$

$$-\psi(z_i(t) - |\mathbf{w}_{in}(t)||\gamma_n|)^2|\mathcal{G}_n,\mathcal{F}_{i-1}\}w_{in,l}(t)^2K_{b_n}(i/n - t).$$

By monotonicity of $\psi(\cdot)$, we have

$$\sum_{i\in\mathcal{N}_n(t+)}\sup_{|\beta|\le\gamma_n}\mathbb{E}\{(\eta_i(t,\beta) - \eta_i(t,0) - \mathbb{E}(\eta_i(t,\beta) - \eta_i(t,0)|\mathcal{F}_{i-1},\mathcal{G}_n))^2w_{in,l}(t)^2|\mathcal{F}_{i-1},\mathcal{G}_n\} \le U_n(t). \tag{35}$$

Then by assumption [A3], we can get

$$\mathbb{E}[\sup_{t\in(0,1)}U_n(t)] \le \sup_{t\in(0,1),x\in\mathbb{R}}f(t,x|\mathcal{F}_{i-1},\mathcal{G}_i)|\gamma_n|$$

$$\sup_{t\in(0,1)}\sum_{i\in\mathcal{N}_n(t+)}\mathbb{E}[|\mathbf{w}_{i,n}(t)|^3]K_{b_n}(i/n - t) \le Cnb_n\gamma_n.$$

So we have

$$\mathbb{P}[\sup_{t\in(0,1)} U_n(t) \geq u_n^2] = O((\log g_n/g_n)^2) = o(1). \tag{36}$$

On the other hand, note that $A_n^2(t) \leq \sum_{i\in\mathcal{N}_n(t+)} \sup_{|\beta|\leq\gamma_n} |(\eta_i(t,\beta) - \eta_i(t,0))w_{in,l}(t)|^2$, we have $\mathbb{E}(\sup_{t\in(0,1)} A_n(t)^2) \leq Cnb_n\gamma_n$. Consequently,

$$\mathbb{P}[\sup_{t\in(0,1)} A_n(t) \geq u_n] \leq u_n^{-2}\mathbb{E}(\sup_{t\in(0,1)} A_n(t)^2) = O((\log g_n/g_n)^2) = o(1). \tag{37}$$

Let $\prod_p = \{-1,+1\}^p$. For $i \in \mathbb{N}$, define $D_{\mathbf{x}}(i) = (2 \times I(x_{i1} \geq 0) - 1, ..., 2 \times I(x_{ip} \geq 0) - 1)$. Define $M_{nl1,\mathbf{d}}(t,\beta) = \sum_{i\in\mathcal{N}_n(t+)}^{n} [\eta_i(t,\beta) - \mathbb{E}(\eta_i(t,\beta)|\mathcal{F}_{i-1},\mathcal{G}_n)]w_{in,l}I(D_{\mathbf{x}}(i) = \mathbf{d})$. By the fact that $M_{nl1}(t,\beta) = \sum_{\mathbf{d}\in\prod_p} M_{nl1,\mathbf{d}}(t,\beta)$, it suffices to show the theorem holds with $M_{nl1,\mathbf{d}}$ for all $\mathbf{d} \in \prod_p$. For presentation clarity we consider $\mathbf{d} = (-1,-1,1,1,...,1)$. Write $\zeta_{i,\mathbf{d},l}(t,\beta) = (\eta_i(t,\beta) - \eta_i(t,0))w_{in,l}I(D_{\mathbf{x}}(i) = \mathbf{d})$, $B_n(t,\beta) = \sum_{i\in\mathcal{N}_n(t+)} \mathbb{E}\{\zeta_{i,\mathbf{d},l}(t,\beta)I(|\zeta_{i,\mathbf{d},l}(t,\beta)| \geq u_n)|\mathcal{G}_n,\mathcal{F}_{i-1}\}$. Hence for large n, by our choice of u_n, ϕ_n,

$$\mathbb{P}(\sup_{t\in(0,1)} |B_n(t,\beta)| \geq \phi_n, \sup_{t\in(0,1)} U_n(t) \leq u_n^2) \leq \mathbb{P}(\sup_{t\in(0,1)} U_n \geq \phi_n u_n, \sup_{t\in(0,1)} U_n \leq u_n^2) = 0. \tag{38}$$

Let $q = n^{10}$ and $\mathfrak{G}_q = \{(k_1/q, ..., k_{2p}/q) : k_i \in \mathbb{Z}, |k_i| \leq n^{10}\} \cap [-\gamma_n,\gamma_n]^{2p}$, $\mathfrak{H}_q = \{k_{p+1}/q : 0 \leq k_{p+1} \leq n^{10}, k_{p+1} \in \mathbb{Z}\}$. Then there are at most in total $(2n^{10}+1)^{2p} \times (n^{10}+1)$ points in $\mathfrak{G}_q \times \mathfrak{H}_q$. Combining (36)(37)(38), the similar argument of Wu (2007) and Freedman's (1975) exponential inequality for martingale differences, we have that

$$\begin{aligned}\mathbb{P}\{&\sup_{\beta\in\mathfrak{G}_q,t\in\mathfrak{H}_q} |M_{nl1,\mathbf{d}}(t,\beta) - M_{nl1,\mathbf{d}}(t,0)| \geq 2\phi_n, \sup_{t\in(0,1)} A_n(t) \leq u_n, \sup_{t\in(0,1)} U_n(t) \leq u_n^2\} \\ &\leq (2n^{10}+1)^{2p} \times (n^{10}+1)O(\exp\{-\phi_n^2/(4u_n\phi_n + 2u_n^2)\}).\end{aligned} \tag{39}$$

Consequently, by (36)(37)(39) and the definition of u_n, ϕ_n, we have

$$\lim_{n\to\infty} \mathbb{P}\{\sup_{\beta\in\mathfrak{G}_q,t\in\mathfrak{H}_q} |M_{nl1,\mathbf{d}}(t,\beta) - M_{nl1,\mathbf{d}}(t,0)| \geq 2\phi_n\} = 0. \tag{40}$$

For any $a \in \mathbb{R}$, define $\langle a\rangle_{q,-1} = \lceil a\rceil_q = \lceil aq\rceil/q$, $\langle a\rangle_{q,1} = \lfloor a\rfloor_q = \lfloor aq\rfloor/q$. Write $\langle\beta\rangle_{q,\mathbf{d}} = (\lceil\beta_{01}\rceil_q, \lceil\beta_{02}\rceil_q, \lfloor\beta_{03}\rfloor_q, ..., \lfloor\beta_{0p}\rfloor_q, \lceil\beta_{11}\rceil_q, \lceil\beta_{12}\rceil_q, \lfloor\beta_{13}\rfloor_q, ..., \lfloor\beta_{1p}\rfloor_q)$ and $\langle\beta\rangle_{q,-\mathbf{d}} = (\lfloor\beta_{01}\rfloor_q, \lfloor\beta_{02}\rfloor_q,$

$\lceil\beta_{03}\rceil_q, ..., \lceil\beta_{0p}\rceil_q, \lfloor\beta_{11}\rfloor_q, \lfloor\beta_{12}\rfloor_q, \lceil\beta_{13}\rceil_q, ..., \lceil\beta_{1p}\rceil_q)$. By the monotonicity of $\psi(\cdot)$, for $i \in \mathcal{N}_n(t+)$, $t \in \mathfrak{H}_q$, $\beta \in \mathfrak{G}_q$, $l = p+1$,

$$\zeta_{i,\mathbf{d},l}(t, \langle\beta\rangle_{q,\mathbf{d}}) \leq \zeta_{i,\mathbf{d},l}(t,\beta) \leq \zeta_{i,\mathbf{d},l}(t, \langle\beta\rangle_{q,-\mathbf{d}}). \tag{41}$$

Note that by [A1]

$$\sup_{|\beta|\leq\gamma_n} \sum_{i\in\mathcal{N}_n(t+)} |\mathbb{E}[(\zeta_{i,\mathbf{d},l}(t,\beta) - (\zeta_{i,\mathbf{d},l}(t, \langle\beta\rangle_{q,\mathbf{d}})|\mathcal{F}_{i-1}, \mathcal{G}_n]| \leq C \sum_{i\in\mathcal{N}_n(t+)} |\mathbf{w}_{in}(t)|^2/q. \tag{42}$$

Similar inequality to (42) holds for $\langle\beta\rangle_{q,-\mathbf{d}}$. Note that by [A2], $\sum_{i\in\mathcal{N}_n(t+)} |\mathbf{w}_{in}(t)|^2 \leq \mathbb{E}(\sum_{i=1}^n |\mathbf{x}_i|^2) \leq Cn$. By similar chaining argument in Lemma 5 of Zhou and Wu (2009), and the fact that $g_n \to \infty$ arbitrarily slow, we can show that

$$\sup_{t\in\mathfrak{H}_l, |\beta|\leq\gamma_n} |M_{nl1,\mathbf{d}}(t,\beta) - M_{nl1,\mathbf{d}}(t,0)| = O_p((nb_n\gamma_n)^{1/2}\log n + n^{-4}). \tag{43}$$

Let $t_k = k/q$, $k = 0, ..., q$. By similar chaining argument, we can get

$$\max_{0\leq k\leq q-1} \sup_{0<t-t_k<q^{-1}, |\beta|\leq\gamma_n} |M_{nl1,\mathbf{d}}(t,\beta) - M_{nl1,\mathbf{d}}(t_k,\beta)| = o_p(n^{-3}). \tag{44}$$

Then by the triangle inequality and (43)(44), the lemma holds. □

Lemma 2. *Suppose the conditions of Lemma 1 hold. Then*

$$\sup_{t\in(0,1), |\beta|\leq\gamma_n} |N_n(t,\beta) - N_n(t,0)| = O_p((nb_n)^{1/2}\gamma_n \log^{2p+7/2} n).$$

Proof. Let $J = \{\alpha_1, ..., \alpha_q\} \subseteq \{1, 2, ..., 2p\}$ be a nonempty set and $1 \leq \alpha_1 < ... < \alpha_q$. For a $2p$-dimension vector $\mathbf{u} = (u_1, ..., u_{2p})$, let $\mathbf{u}_J = (u_1 I(1 \in J), ..., u_{2p} I(2p \in J))$. Then

$$\sup_{t\in(0,1), |\beta|\leq\gamma_n} \int_0^{\beta_J} |\frac{\partial^q N_n(t, \mathbf{u}_J)}{\partial \mathbf{u}_J}| d\mathbf{u}_J \leq \int_{-\gamma_n}^{\gamma_n}, ..., \int_{-\gamma_n}^{\gamma_n} \sup_{t\in(0,1)} |\frac{\partial^q N_n(t, \mathbf{u}_J)}{\partial \mathbf{u}_J}| d\mathbf{u}_J. \tag{45}$$

Observe that

$$|\frac{\partial^q N_n(t,\mathbf{u}_J)}{\partial \mathbf{u}_J}| = |\sum_{i=1}^n \sum_{k=1}^\infty \mathcal{P}_{i-k,n}[\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}(\eta_i(t,\mathbf{u}_j)|\mathcal{F}_{i-1},\mathcal{G}_n)]\mathbf{w}_{in}(t)|. \quad (46)$$

Note that by Burkholder inequality, for $s > 1$,

$$\begin{aligned}&\|\sum_{i=1}^n \mathcal{P}_{i-k,n}[\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}(\eta_i(t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)]\mathbf{w}_{in}(t)\|_s^2 \\ &\le s\sum_{i=1}^n \|\mathcal{P}_{i-k,n}[\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}(\eta_i(t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)]\mathbf{w}_{in}(t)\|_s^2. \end{aligned} \quad (47)$$

Elementary calculation shows that $z_i(t) = e_i + \mathbf{x}_i'(\theta(i/n) - \theta(t) - \dot{\theta}(t)(i/n - t))$. Then we have for $1 \le q \le 2p$,

$$\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}[\eta_i(t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n]\mathbf{w}_{in}(t) = -f^{(q-1)}(i/n,\Delta(i,t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)\mathbf{W}_{in,J}(t)K_{b_n}(i/n - t), \quad (48)$$

where $\mathbf{W}_{in,J}(t) = \mathbf{w}_{in}(t)w_{in,\alpha_1}(t)...w_{in,\alpha_q}(t)$, and $\Delta(i,t,\mathbf{u}_J) = \mathbf{w}_{in}'(t)\mathbf{u}_J - \mathbf{x}_i'(\theta(i/n) - \theta(t) - \dot{\theta}(t)(i/n - t))$. Note that $\mathbb{P}(G(t,\mathcal{F}_i,\mathcal{G}_i) \le x|\mathcal{F}_{i-1},\mathcal{G}_i) = \mathbb{P}(G(t,\mathcal{F}_i,\mathcal{G}_i) \le x|\mathcal{F}_{i-1},\mathcal{G}_n)$, by using the similar argument of Lemma 1 in Wu (2007) and condition [A3], we shall have

$$\|\mathcal{P}_{i-k,n}f^{(q-1)}(i/n,\Delta(i,t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)\mathbf{W}_{in,J}(t)\|_s \le \sup_{t\in(0,1)}\|\mathbf{W}_{in,J}(t)\|_{2s}\delta_{2s}(k-1). \quad (49)$$

By Cauchy-Schwarz inequality, Proposition 7 and [A2], we see that $\sup_{t\in(0,1)}\max_{1\le i\le n}\|\mathbf{W}_{in,J}(t)\|_{2s} \le Cs^{2p+1}$. Consequently, by (46)(47)(48)(49), Proposition 6 and the triangle inequality, for some large constant C,

$$\sup_{|\mathbf{u}_J|\le C\gamma_n}\|\frac{\partial^q N_n(t,\mathbf{u}_J)}{\partial \mathbf{u}_J}\|_s \le C\sqrt{nb_n}s^{2p+3/2}\frac{1}{1-\chi^{\frac{1}{2s}}} = O(\sqrt{nb_n}s^{2p+5/2}). \quad (50)$$

Consider $\mathcal{P}_{i-k,n}[\frac{\partial}{\partial t}\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}(\eta_i(t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)]\mathbf{w}_{in}(t)$. Then elementary calculation shows that

$$\begin{aligned}&\frac{\partial}{\partial t}\frac{\partial^q}{\partial \mathbf{u}_J}\mathbb{E}(\eta_i(t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)]\mathbf{w}_{in}(t)\\ =&f^{(q-1)}(i/n,\Delta(i,t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)\mathbf{R}_1(i,n,J,t)+f^{(q)}(i/n,\Delta(i,t,\mathbf{u}_J)|\mathcal{F}_{i-1},\mathcal{G}_n)\mathbf{R}_2(i,n,J,t),\end{aligned} \tag{51}$$

where $\mathbf{R}_1(i,n,J,t)=-\frac{\partial}{\partial t}[\mathbf{W}_{in,J}(t)K_{b_n}(i/n-t)]$, $\mathbf{R}_2(i,n,J,t)=-\frac{\partial}{\partial t}[\Delta(i,t,\mathbf{u}_J)]\times\mathbf{W}_{in,J}(t)$ $K_{b_n}(i/n-t)$. Then similarly to (50), by elementary calculation, we can get

$$\sup_{|\mathbf{u}_J|\le\gamma_n}\|\frac{\partial}{\partial t}\frac{\partial^q N_n(t,\mathbf{u}_J)}{\partial \mathbf{u}_J}\|_s\le C\sqrt{nb_n}s^{2p+7/2}/b_n. \tag{52}$$

Hence by proposition 2, we have that

$$\sup_{|\mathbf{u}_J|\le\gamma_n}\|\sup_{t\in(0,1)}|\frac{\partial^q N_n(t,\mathbf{u}_J)}{\partial \mathbf{u}_J}|\|_s=O(\sqrt{nb_n}b_n^{-1/s}s^{2p+7/2}). \tag{53}$$

Combining with (45), by taking $s=\lfloor\log\frac{1}{b_n}\rfloor$, the lemma is valid in view of

$$N_n(t,\beta)-N_n(t,0)=\sum_{J\subseteq 1,\dots,2p}\int_0^{\beta_J}\frac{\partial^{|J|}N_n(t,\mathbf{u}_J)}{\partial \mathbf{u}_J}d\mathbf{u}_J. \tag{54}$$

□

Recall the definition of $\tilde\beta_{b_n,\tau}(t)$ in equation (9). The following lemma deals with the consistency of $\tilde\beta_{b_n,\tau}(t)$. We omit the subscript b_n and τ for brevity. Let X_n, Y_n be random series. Write $X_n\le_p Y_n$ if $\lim_{n\to\infty}\mathbb{P}(X_n\le Y_n)=1$, and vice versa. We also write $\tilde\beta_{b_n,\tau}(t)$ as $\tilde\beta_n(t)$ for short.

Lemma 3. *Suppose [A0]-[A4], $b_n\to 0$, $nb_n^4/\log^7 n\to\infty$. Then we have,* $\sup_{t\in(0,1)}|\tilde\beta_n(t)|\le_p (nb_n)^{-1/2}\log^4 n+b_n^2\log n$*.*

Proof. Let $g_n=(nb_n)^{-1/2}\log^4 n+b_n^2\log n$. Write $\Theta_n(\beta,t)=\sum_{i=1}^n[\rho(z_i(t)-\mathbf{w}'_{in}(t)\beta)-\rho(z_i(t))]K_{b_n}(i/n-t)$, and

$$A_n(t,\beta)=-\sum_{i=1}^n\int_0^1\mathbb{E}[\psi(z_i(t)-\mathbf{w}'_{in}(t)\beta s)|\mathcal{G}_n]\mathbf{w}'_{in}(t)\beta K_{b_n}(i/n-t)ds. \tag{55}$$

By the fact that $\rho(z_i(t)) - \rho(z_i(t) - \mathbf{w}'_{in}(t)\beta) = \int_0^1 \psi(z_i(t) - \mathbf{w}'_{in}(t)\beta s)\mathbf{w}'_{in}(t)\beta ds$, we have

$$\sup_{|\beta|\le g_n, t\in(0,1)} |\Theta_n(t,\beta) - A_n(t,\beta)| = \sup_{|\beta|\le g_n, t\in(0,1)} |\int_0^1 \tilde{K}_n(t,\beta s)\beta ds|, \tag{56}$$

where $\tilde{K}_n(t,\beta)$ are defined in (32). Via the similar argument of Lemma 1 and Lemma 2, $\sup_{t\in(0,1)} |\tilde{K}_n(t,0)| = O_p(\sqrt{nb_n}\log^{7/2} n)$. Combining with (56), Lemma 1 and Lemma 2, we can get

$$\sup_{|\beta|\le g_n, t\in(0,1)} |\Theta_n(t,\beta) - A_n(t,\beta)| = O_p(\sqrt{nb_n} g_n \log^{7/2} n). \tag{57}$$

On the other hand, since $z_i(t) = e_i + \mathbf{x}'_i(\theta(i/n) - \theta(t) - (i/n - t)\dot{\theta}(t)) := e_i - \Lambda(t,i)$,

$$\begin{aligned}\inf_{t\in(0,1),|\beta|=g_n} A_n(t,\beta) = \inf_{t\in(0,1),|\beta|=g_n} \int_0^1 \sum_{i=1}^n \{f(i/n,0|\mathcal{G}_i)[\Lambda(t,i) + \mathbf{w}'_{in}(t)\beta s] + \\ f^{(1)}(i/n,\theta_i^*(s)|\mathcal{G}_i)[\Lambda(t,i) + \mathbf{w}'_{in}(t)\beta s]^2/2\}\mathbf{w}'_{in}(t)\beta K_{b_n}(i/n - t)ds,\end{aligned} \tag{58}$$

where $\theta_i^*(s)$ is a number between $\Lambda(t,i) + \mathbf{w}'_{in}(t)\beta s$ and 0, and elementary calculation shows that $\Lambda(t,i) = -\mathbf{x}'_i\ddot{\theta}(t_i^*)\frac{(i/n-t)^2}{2}$ for some t_i^* between t and i/n. Note that by our choice of g_n, $g_n/b_n^2 \to \infty$ as $n \to \infty$. By Lemma 12 and elementary calculation, there exists some constant $\iota > 0$ such that

$$\inf_{t\in(0,1),|\beta|=g_n} A_n(t,\beta) \ge_p \iota n b_n g_n^2. \tag{59}$$

then

$$\begin{aligned}\inf_{|\beta|=g_n, t\in(0,1)} \Theta_n(t,\beta) \ge \inf_{|\beta|=g_n, t\in(0,1)} A_n(t,\beta) - \sup_{|\beta|\le g_n, t\in(0,1)} |\Theta_n(t,\beta) - A_n(t,\beta)| \\ \ge_p \iota n b_n g_n^2 - \sqrt{nb_n}\log^{7/2} n g_n.\end{aligned} \tag{60}$$

Hence $\inf_{|\beta|=g_n, t\in(0,1)} \Theta_n(t,\beta) \geq_p \epsilon \log^8 n$ for some small positive ϵ. By convexity of $\Theta_n(t,\cdot)$,

$$\{\inf_{t\in(0,1),|\beta|=g_n} \Theta_n(t,\beta)\} = \inf_{t\in(0,1)}\{\inf_{|\beta|=g_n} \Theta_n(t,\beta)\}$$
$$= \inf_{t\in(0,1)}\{\inf_{|\beta|\geq g_n} \Theta_n(t,\beta)\} = \{\inf_{t\in(0,1),|\beta|\geq g_n} \Theta_n(t,\beta)\}. \tag{61}$$

Since $\tilde{\beta}_n(t)$ is the minimizer of $\Theta_n(t,\beta)$, we have that $\sup_{t\in(0,1)} |\tilde{\beta}_n(t)| \leq_p (nb_n)^{-1/2}\log^4 n + b_n^2 \log n$. □

Lemma 4. *Assume [A1]-[A4]. Then* $\sup_{t\in(0,1)} S_n(t,\tilde{\beta}_n(t)) = O_p(\log n)$.

Proof. Similarly to Lemma 8 of Zhou and Wu (2009), define $\breve{\theta}_n(t) = (\hat{\theta}_n(t)', \hat{\dot{\theta}}_n(t)')'$,

$$\sup_{t\in(0,1)} \sum_{i=1}^n S_n(t,\tilde{\beta}_n(t)) \leq \sup_{t\in(0,1)} \sum_{i=1}^n |\mathbf{w}'_{in}(t)| I(y_i = \mathbf{w}'_{in}(t)\breve{\theta}_n(t))$$
$$\leq (\sup_{t\in(0,1)} |\mathbf{w}'_{in}(t)|)(\sup_{t\in(0,1)} \sum_{i=1}^n I(y_i = \mathbf{w}'_{in}(t)\breve{\theta}_n(t)). \tag{62}$$

Note that $\sup_{t\in(0,1)} |\mathbf{w}'_{in}(t)| = O_p(\log n)$, by Babu (1989),

$$\sup_{t\in(0,1)} \sum_{i=1}^n I(y_i = \mathbf{w}'_{in}(t)\breve{\theta}_n(t)) \leq \sup_{\beta_1\in\mathbb{R},\beta_2\in\mathbb{R}} \sum_{i=1}^n I(y_i = \beta_1 + \frac{i}{n}\beta_2)$$
$$\leq \sup_{\beta_1\in\mathbb{R},\beta_2\in\mathbb{R}} \sum_{i=1}^n I(e_i = \beta_1 + \frac{i}{n}\beta_2) = O_p(1). \tag{63}$$

Therefore the lemma follows from (62)(63). □

Write

$$\Lambda_n(t) = \frac{1}{nb_n}\sum_{i=1}^n f(t,0|\mathcal{G}_i)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)K_{b_n}(i/n-t). \tag{64}$$

Proposition 3. *Under condition [A0]-[A4], suppose* $\frac{nb_n^4}{\log^7 n} \to \infty$*. Then*

$$\sup_{t\in\mathfrak{T}_n} |\Lambda_n(t) - \Sigma_1(t)| = O_p(b_n). \tag{65}$$

Proof. By elementary calculation, we can show that

$$\sup_{t\in\mathfrak{T}_n}|\mathbb{E}(\Lambda_n(t))-\Sigma_1(t)|=O(b_n+\frac{1}{nb_n})=O_p(b_n). \tag{66}$$

Write $\lambda(i,t,\mathcal{G}_i)=\frac{f(t,0|\mathcal{G}_i)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)K_{b_n}(i/n-t)}{nb_n}$, and $\mathcal{P}_i(\cdot)=\mathbb{E}(\cdot|\mathcal{G}_i)-\mathbb{E}(\cdot|\mathcal{G}_{i-1})$. By definition, $\mathbf{w}_{in}(t)=(\mathbf{H}'(i/n,\mathcal{G}_i),(\frac{i/n-t}{b_n})\mathbf{H}'(i/n,\mathcal{G}_i))'$. Then

$$\Lambda_n(t)-\mathbb{E}(\Lambda_n(t))=\sum_{i=1}^n\sum_{k=0}^\infty\mathcal{P}_{i-k}\lambda(i,t,\mathcal{G}_i). \tag{67}$$

Define

$$\mathbf{w}_{in}^{(i-k)}(t)=(\mathbf{H}'(i/n,\mathcal{G}_i^{(i-k)}),(\frac{i/n-t}{b_n})\mathbf{H}'(i/n,\mathcal{G}_i^{(i-k)})')'.$$

Note that for $v\geq 1$, by similar argument of Lemma 2 of Wu (2007) and triangle inequality,

$$\begin{aligned}\|\mathcal{P}_{i-k}\lambda(i,t,\mathcal{G}_i)\|_v\leq\|\lambda(i,t,\mathcal{G}_i)-\lambda(i,t,\mathcal{G}_i^{(i-k)})\|_v\\ \leq\frac{K_{b_n}(i/n-t)}{nb_n}\{\|f(t,0|\mathcal{G}_i)[\mathbf{w}_{in}^{(i-k)}(t)(\mathbf{w}_i^{(i-k)}n(t))'-\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)]\|_v+\\ \|\mathbf{w}_{in}^{(i-k)}(t)(\mathbf{w}_i^{(i-k)}n(t))'[f(t,0|\mathcal{G}_i)-f(t,0|\mathcal{G}_i^{(i-k)})]\|_v\}.\end{aligned} \tag{68}$$

By [A2] [A4], elementary calculation with Cauchy-Schwarz inequality and the triangle inequality, we have

$$\|\lambda(i,t,\mathcal{G}_i)-\lambda(i,t,\mathcal{G}_i^{(i-k)})\|_v\leq C\frac{K_{b_n}(i/n-t)}{nb_n}v^2\tilde{\chi}^{\frac{k}{2v}} \tag{69}$$

for some large constant C and $\tilde{\chi}\in(0,1)$. Combining with (67), by Burkholder inequality and similar argument of Lemma 2, we have

$$\|\Lambda_n(t)-\mathbb{E}(\Lambda_n(t))\|_v=O_p(\frac{v^2\sqrt{v}}{\sqrt{nb_n}(1-\tilde{\chi}^{\frac{1}{2v}})}). \tag{70}$$

Similarly,

$$\|\frac{\partial}{\partial t}(\Lambda_n(t)-\mathbb{E}(\Lambda_n(t)))\|_v=O_p(\frac{v^2\sqrt{v}}{\sqrt{nb_n}b_n(1-\tilde{\chi}^{\frac{1}{2v}})}). \tag{71}$$

Take $v = \lfloor \log \frac{1}{b_n} \rfloor$. Then the Lemma follows by Proposition 2. $\square$

Proof of Theorem 1. Combining Lemma 1, Lemma 2, Lemma 3 and Lemma 4, we could easily get

$$\sup_{t\in\mathfrak{T}_n} |\sum_{i=1}^{n} \psi(z_i(t))\mathbf{w}_{in}(t)K_{b_b}(i/n-t) - \sum_{i=1}^{n} \mathbb{E}[\psi(z_i(t))- \\ \psi(z_i(t)-\mathbf{w}'_{in}(t)\tilde{\beta}_n(t))|\mathcal{G}_n]\mathbf{w}_{in}(t)K_{b_n}(i/n-t)| = O_p((nb_n)^{1/4}\log^3 n + n^{1/2}b_n^{3/2}\log n). \tag{72}$$

Write $f_{z_i(t)}(x|\mathcal{G}_n) = \frac{\partial}{\partial x}\mathbb{P}(z_i(t) \le x|\mathcal{G}_n)$. By elementary calculation, $f_{z_i(t)}(x|\mathcal{G}_n) = f(i/n, x+\Lambda(t,i)|\mathcal{G}_n)$, where $\Lambda(t,i)$ is defined in Lemma 3. It is also obvious that $\sup_{t\in(0,1)x\in\mathbb{R}} \frac{\partial}{\partial x} f_{z_i(t)}(x|\mathcal{G}_n)$ $\le C_0 < \infty$ for some constant C_0 and for all $i = 1, ..., n$. Consider $A_n = \{\sup_{t\in(0,1)} |\tilde{\beta}_n(t)| \le \frac{\log^4 n}{\sqrt{nb_n}} + b_n^2 \log n\}$, by Lemma 3, $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$. By Taylor expansion and similar argument of Lemma 2,

$$\sup_{t\in\mathfrak{T}_n} |\sum_{i=1}^{n} \mathbb{E}[\psi(z_i(t)) - \psi(z_i(t)-\mathbf{w}'_{in}(t)\tilde{\beta}_n(t))|\mathcal{G}_i]\mathbf{w}_{in}(t)K_{b_n}(i/n-t)I(A_n) \\ -\sum_{i=1}^{n} f_{z_i(t)}(0|\mathcal{G}_n)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)\tilde{\beta}_n(t)K_{b_n}(i/n-t)I(A_n)| = O_p(\log^{11} n + nb_n^6 \log^5 n). \tag{73}$$

Similarly,

$$\sup_{t\in\mathfrak{T}_n} |\sum_{i=1}^{n} f_{z_i(t)}(0|\mathcal{G}_n)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)\tilde{\beta}_n(t)K_{b_n}(i/n-t)I(A_n) \\ -\sum_{i=1}^{n} f(i/n, 0|\mathcal{G}_n)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)\tilde{\beta}_n(t)K_{b_n}(i/n-t)I(A_n)| = O_p(\sqrt{nb_n}b_n^2 \log^6 n + nb_n^5 \log^3 n). \tag{74}$$

Via the stochastic Lipchitz continuity of $f(\cdot, x|\mathcal{G}_n)$ on $(0,1)$, we have that

$$\sup_{t\in\mathfrak{T}_n}|\sum_{i=1}^{n} f(i/n,0|\mathcal{G}_n)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)\tilde{\beta}_n(t)K_{b_n}(i/n-t)I(A_n)-$$

$$\sum_{i=1}^{n} f(t,0|\mathcal{G}_n)\mathbf{w}_{in}(t)\mathbf{w}'_{in}(t)\tilde{\beta}_n(t)K_{b_n}(i/n-t)I(A_n)| = O_p(\sqrt{nb_n}b_n\log^6 n + nb_n^4\log^3 n). \tag{75}$$

By (72) (73) (74) (75) and Proposition 3,

$$\sup_{t\in\mathfrak{T}_n}|\Lambda_n(t)\tilde{\beta}_n(t) - \frac{\sum_{i=1}^n \psi(z_i(t))\mathbf{w}'_{in}(t)K_{b_n}(i/n-t)}{nb_n}| = O_p(b_n^3\log^3 n + \frac{b_n\log^6 n}{\sqrt{nb_n}} + (nb_n)^{-\frac{3}{4}}\log^3 n). \tag{76}$$

Then the theorem follows. □

Lemma 5. *Suppose the conditions of Theorem 1 hold. Then*

$$\sup_{t\in\mathfrak{T}_n}\frac{|\sum_{i=1}^n[(\psi(z_i(t))-\psi(e_i))\mathbf{w}'_{in}(t) - \mathbb{E}\{\psi(z_i(t))\mathbf{w}'_{in}(t)|\mathcal{G}_n\}]K_{b_n}(i/n-t)|}{nb_n} = O_p(\frac{b_n\log n}{\sqrt{nb_n}}). \tag{77}$$

Proof. Write $\eta(i,t) = [\psi(e_i)-\psi(z_i(t))]\mathbf{w}_{in}(t)$. Define $M_n(t) = \sum_{i=1}^n[\eta(i,t)-\mathbb{E}(\eta(i,t)|\mathcal{F}_{i-1},\mathcal{G}_n)]$, and $N_n(t) = \sum_{i=1}^n[\mathbb{E}(\eta(i,t)|\mathcal{F}_{i-1},\mathcal{G}_n) - \mathbb{E}(\eta(i,t)|\mathcal{G}_n)]$. By the similar argument of Lemma 1 and Lemma 2, we can show that

$$\sup_{t\in\mathfrak{T}_n}|M_n(t)| = O_p(\sqrt{nb_n}b_n\log n),\ \sup_{t\in\mathfrak{T}_n}|N_n(t)| = O_p(\sqrt{nb_n}b_n^{3/2}). \tag{78}$$

Then the lemma follows by elementary calculation.

Lemma 6. *Under the conditions of Theorem 1,*

$$\sup_{t\in\mathfrak{T}_n}|\sum_{i=1}^n\frac{\mathbb{E}\{\psi(z_i(t))\mathbf{w}'_{in}(t)|\mathcal{G}_n\}]K_{b_n}(i/n-t)}{nb_n} - \frac{b_n^2\tilde{\Sigma}(t)\ddot{\Theta}(t)}{2}| = O_p(b_n^3\log^2 n), \tag{79}$$

where $\ddot{\Theta}'(t) = (\ddot{\theta}'(t), \mathbf{0}_{1\times p})$ *is a* $1\times 2p$ *vector,* $\tilde{\Sigma}(t) = \begin{pmatrix}\mu_2\Sigma(t) & 0\\ 0 & \mu_4 b_n^2\Sigma(t)\end{pmatrix}$ *is* $2p\times 2p$ *matrix.*

Proof. Elementary calculation shows that

$$\sup_{t\in\mathfrak{T}_n}\frac{1}{nb_n}|\sum_{i=1}^{n}[\mathbb{E}\{\psi(z_i(t))|\mathcal{G}_i\}-f(i/n,0|\mathcal{G}_i)\mathbf{x}_i'\ddot{\theta}(t)(i/n-t)^2/2]\mathbf{w}_{in}(t)K_{b_n}(i/n-t)|$$
$$= O_p(b_n^3\log^2 n). \quad (80)$$

Note that

$$\frac{1}{nb_n}\sum_{i=1}^{n}[f(i/n,0|\mathcal{G}_i)\mathbf{x}_i'\ddot{\theta}(t)(i/n-t)^2]\mathbf{w}_{in}(t)K_{b_n}(i/n-t)$$
$$=\left\{\frac{1}{nb_n}\sum_{i=1}^{n}[f(i/n,0|\mathcal{G}_i)(i/n-t)^2]\mathbf{w}_{in}(t)\mathbf{w}_{in}'(t)K_{b_n}(i/n-t)\right\}\ddot{\Theta}(t):=\tilde{\Lambda}_n(t)\ddot{\Theta}(t). \quad (81)$$

Similar to the argument of Proposition 3, we can show that

$$\sup_{t\in\mathfrak{T}_n}|\tilde{\Lambda}_n(t)-b_n^2\tilde{\Sigma}(t)|=O_p(b_n^3), \quad (82)$$

which completes the proof. □

Lemma 7. *Under conditions of Theorem 1, we have*

$$\sup_{t\in\mathfrak{T}_n}|\sqrt{nb_n}(\hat{\theta}_n(t)-\theta(t)-\frac{\mu_2 b_n^2}{2}\ddot{\theta}(t))-\Sigma^{-1}(t)\frac{\sum_{i=1}^{n}\psi(e_i)\mathbf{x}_iK_{b_n}(i/n-t)}{\sqrt{nb_n}}|=O_p(\pi_n), \quad (83)$$

where $\pi_n=b_n\log^6 n+(nb_n)^{-1/4}\log^3 n+b_n^3\sqrt{nb_n}\log^3 n$.

Proof. This lemma is an immediate consequent result of Theorem 1, Lemma 5 and Lemma 6. □

The following proposition is an immediate result of Proposition 5 and Wu and Zhou (2011):

Proposition 4. *Suppose that the conditions of Theorem 2 hold. Then for any* $s\times p$, $s\le p$ *full rank matrix* $\mathbf{C}$, *on a possibly richer probability space, there exist i.i.d* $N(0,Id_s)$ $V_1,\ldots,V_n$, *such that*

$$\max_{1\le j\le n}|\sum_{i=1}^{j}\mathbf{C}\psi(e_i)\mathbf{x}_i-\sum_{i=1}^{j}\nu_{\mathbf{C}}(t_j)V_j|=o_p(n^{1/4}\log^2 n), \quad (84)$$

where $\nu_{\mathbf{C}}(t) = (\mathbf{C}\nu^2(t)\mathbf{C}')^{1/2}$.

Proof of Theorem 2. By Lemma 7, we have

$$\sup_{t\in\mathfrak{T}_n} |\mathbf{C}'(\hat{\theta}_n(t) - \theta(t) - \frac{\mu_2 b_n^2}{2}\ddot{\theta}(t)) - \mathbf{C}'\Sigma^{-1}(t)\frac{\sum_{i=1}^n \psi(e_i)\mathbf{x}_i K_{b_n}(i/n - t)}{nb_n}| = O_p(\pi_n/\sqrt{nb_n}). \tag{85}$$

By Proposition 4 and the summation by parts formula, there exist *i.i.d* $N(0, Id_s)$ $\{V_i\}_{i\in\mathbb{Z}}$, such that

$$\sup_{t\in\mathfrak{T}_n} |\mathbf{C}'\Sigma^{-1}(t)\sum_{i=1}^n \psi(e_i)\mathbf{x}_i K_{b_n}(i/n - t) - \sum_{i=1}^n M_{\mathbf{C}}(t_i)V_i K_{b_n}(i/n - t)| = o_p(n^{1/4}\log^2 n). \tag{86}$$

Write $S_n(t) = \sum_{i=1}^n (M_{\mathbf{C}}(t_i) - M_{\mathbf{C}}(t))V_i K_{b_n}(i/n - t)$. By the lipchitz continuity of $\nu(t)$, $\Sigma(t)$, and Proposition 2, one can show that

$$\sup_{t\in\mathfrak{T}_n} |S_n(t)| = O_p(\sqrt{nb_n}b_n \log n). \tag{87}$$

Hence by (86) (87), we have

$$\sup_{t\in\mathfrak{T}_n} |\frac{\mathbf{C}'\Sigma^{-1}(t)}{nb_n}\sum_{i=1}^n \psi(e_i)\mathbf{x}_i K_{b_n}(i/n - t) - M_{\mathbf{C}}(t)\sum_{i=1}^n V_i K_{b_n}(i/n - t)| = O_p(\frac{n^{1/4}\log^2 n}{nb_n} + \frac{b_n \log n}{\sqrt{nb_n}}). \tag{88}$$

Now the theorem follows from Lemma 1 of Zhou and Wu (2010). $\square$

Proof of Theorem 3 . Write $\Theta_n = \frac{n^{1/4}\log^2 n}{nb_n} + \frac{\pi_n}{\sqrt{nb_n}}$. By carefully checking Theorem 1 and Theorem 2, we shall have that, on a possibly richer probability space, there exist i.i.d $N(0, Id_s)$ $V_1, ..., V_n$, such that

$$\sup_{t\in\mathfrak{T}_n} |\mathbf{C}'(\tilde{\theta}_n(t) - \theta(t)) - M_{\mathbf{C}}(t)\frac{\sum_{i=1}^n V_i K_{b_n}^*(i/n - t)}{nb_n}| = O_p(\Theta_n). \tag{89}$$

Then

$$\sup_{t\in\mathfrak{T}_n} |M_{\mathbf{C}}^{-1}(t)\mathbf{C}'(\tilde{\theta}_n(t)-\theta(t)) - \frac{\sum_{i=1}^n V_i K_{b_n}^*(i/n-t)}{nb_n}| = O_p(\Theta_n). \tag{90}$$

Define the $s\times 1$ vector $\Upsilon_n(t) = \frac{\sum_{i=1}^n V_i K_{b_n}^*(i/n-t)}{nb_n}$, with j_{th} entry $\Upsilon_{nj}(t) = \sum_{i=1}^n V_{ij} K_{b_n}^*(i/n-t)/(nb_n)$, $j = 1, ..., s$. Then

$$\Upsilon_n(t)'\Upsilon_n(t) = \frac{1}{(nb_n)^2}\sum_{j=1}^s(\sum_{i=1}^n V_{ij}K_{b_n}^*(i/n-t))^2 = \sum_{j=1}^s \Upsilon_{nj}^2(t). \tag{91}$$

By Lemma 4 of Zhou (2010), we have, for $1 \le j \le s$,

$$n\sqrt{b_n}\int_{\mathfrak{T}_n}(\Upsilon_{nj}(t))^2\pi(t)dt - \frac{1}{\sqrt{b_n}}K^*\star K^*(0)\int_0^1\pi(t)dt \Rightarrow N(0,\frac{\Xi^*}{s}). \tag{92}$$

Since $\Upsilon_{ni}(t)\perp\Upsilon_{nj}(t)$ for $i\neq j$, we have

$$\sum_{j=1}^s(n\sqrt{b_n}\int_{\mathfrak{T}_n}(\Upsilon_{nj}(t))^2\pi(t)dt - \frac{1}{\sqrt{b_n}}K^*\star K^*(0)\int_0^1\pi(t)dt) \Rightarrow N(0,\Xi^*), \tag{93}$$

which further implies that

$$n\sqrt{b_n}\int_{\mathfrak{T}_n}\Upsilon_n(t)'\Upsilon_n(t)\pi(t)dt - \frac{s}{\sqrt{b_n}}K^*\star K^*(0)\int_0^1\pi(t)dt \Rightarrow N(0,\Xi^*). \tag{94}$$

By (90), (94) and the definition of T_n, we get

$$T_n - \int_{\mathfrak{T}_n}(\Upsilon_n(t)+\varrho_n M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))'(\Upsilon_n(t)+\varrho_n M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))\pi(t)dt = O_p(\frac{\Theta_n\log n}{\sqrt{nb_n^{1/2}}}). \tag{95}$$

It is easy to see that

$$\int_{\mathfrak{T}_n}\Upsilon_n(t)'\varrho_n M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t)\pi(t)dt = O_p(\frac{\varrho_n}{\sqrt{n}}). \tag{96}$$

Combining with (95), we have

$$nb_n^{1/2}T_n - nb_n^{1/2}\int_{\mathfrak{T}_n} \Upsilon(t)'\Upsilon(t)\pi(t)dt - \int_0^1 (M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))'(M_{\mathbf{C}}^{-1}(t)\mathbf{C}'\eta(t))\pi(t)dt$$
$$= O_p(\sqrt{nb_n}\varrho_n + \sqrt{n}b_n^{1/4}\Theta_n \log n). \qquad (97)$$

By (94) and the fact that $\sqrt{nb_n}\varrho_n + \sqrt{n}b_n^{1/4}\Theta_n \log n = o(1)$ we get proof. □

In the following proofs, we omit subscripts b_n, τ, n if there is no confusion caused.

Proof of Theorem 4. Recall that $\hat{\beta}(t) = \hat{\theta}(t) - \theta(t)$, let $\varphi_n = \frac{\log^4 n}{\sqrt{nb_n}} + b_n^2 \log n$. Define $A_n = \{\sup_{t\in(0,1)} |\hat{\beta}(t)| \leq \varphi_n\}$. Then by Lemma 3, $\lim_{n\to} \mathbb{P}(A_n) = 1$. Let D_0 be the true subset. Recall that $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$, $\mathbf{x}_{i,D} = (x_{i1}I(1 \in D), ..., x_{ip}I(p \in D))'$, and $\hat{\theta}_D(\cdot) = (\hat{\theta}_1(\cdot)I(1 \in D), ..., \hat{\theta}_p(\cdot)I(p \in D))'$. Similarly we define θ_D, β_D and $\hat{\beta}_D$. If $D_0 \subseteq D$, then

$$A(D) := \sum_{i=1}^n \rho(y_i - \mathbf{x}_{i,D}'\hat{\theta}_D(i/n)) = \sum_{i=1}^n \rho(e_i + \mathbf{x}_{i,D}'\theta_D(i/n) - \mathbf{x}_{i,D}'\hat{\theta}_D(i/n))$$
$$= \sum_{i=1}^n \rho(e_i) - \int_0^1 \sum_{i=1}^n \psi(e_i - \mathbf{x}_{i,D}'\hat{\beta}_D(i/n)t)\mathbf{x}_{i,D}'\hat{\beta}_D(i/n)dt. \qquad (98)$$

Write $\mathbf{x}_{i,D}'\hat{\beta}_D(i/n) = z_{D,i}$, then we have

$$\sup_{|z_{D,i}|\leq\delta} |\sum_{i=1}^n (\psi(e_i - z_{D,i}) - \psi(e_i))z_{D,i}| \leq \sum_{i=1}^n (\psi(e_i + |\delta|) - \psi(e_i - |\delta|))|\delta|. \qquad (99)$$

Since $|z_{D,i}| \leq_p M\varphi_n \log n$ for some large constant M, by using similar martingale decomposition technique and chaining argument in Lemma 1 and Lemma 2, we have

$$\sup_{|z_{D,i}|\leq\delta} |\sum_{i=1}^n (\psi(e_i - z_{D,i}) - \psi(e_i))z_{D,i} - \mathbb{E}[(\psi(e_i - z_{D,i}) - \psi(e_i))z_{D,i}|\mathcal{G}_n]|$$
$$= O_p(\sqrt{n\varphi_n \log n}\varphi_n \log^2 n), \qquad (100)$$

which implies that

$$\sum_{i=1}^{n}(\psi(e_i - \mathbf{x}'_{i,D}\hat{\beta}_D(i/n)) - \psi(e_i))\mathbf{x}'_{i,D}\hat{\beta}_D(i/n) - \mathbb{E}[(\psi(e_i - \mathbf{x}'_{i,D}\hat{\beta}_D(i/n)))\mathbf{x}'_{i,D}\hat{\beta}_D(i/n)|\mathcal{G}_n]|$$
$$= O_p(\sqrt{n\varphi_n \log n}\varphi_n \log^2 n). \tag{101}$$

On the other hand, by equation (85), one have

$$\sum_{i=1}^{n}\psi(e_i)\mathbf{x}'_{i,D}\hat{\beta}_D(i/n) = \frac{1}{nb_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi(e_i)\psi(e_j)\mathbf{x}'_{i,D}\Sigma^{-1}(i/n)\mathbf{x}_{D,j}K_{b_n}(i/n - j/n)$$
$$+O_p(nb_n\varphi_n + \sqrt{n}b_n^2 + \pi_n\sqrt{n}/\sqrt{b_n}). \tag{102}$$

Via using Lemma A.2 in Zhang and Wu (2012), Lemma 5 in Zhou and Wu (2010), and the similar argument in Lemma A.6 of Zhang and Wu (2012), we can show that

$$\frac{1}{nb_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi(e_i)\psi(e_j)\mathbf{x}'_{i,D}\Sigma^{-1}(i/n)\mathbf{x}_{D,j}K_{b_n}(i/n - j/n) = O_p(b_n^{-1}). \tag{103}$$

Combining (101), (102), (103), we have

$$A(D) = \sum_{i=1}^{n}\rho(e_i) + O_p(n\varphi_n b_n + \pi_n\sqrt{n/b_n}). \tag{104}$$

Note that $\theta_{\bar{D}_0}(\cdot) = 0$, for the case that $D_0 \not\subseteq D$, which implies

$$\hat{e}_{D,i} = e_i - \mathbf{x}'_{i,D}\hat{\beta}_D(i/n) + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n), \tag{105}$$

we have

$$A(D) = \sum_{i=1}^{n}\rho(e_i) - \int_0^1 \sum_{i=1}^{n}\psi(e_i - (\mathbf{x}'_{i,D}\hat{\beta}_D(i/n)$$
$$- \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n))t)(\mathbf{x}'_{i,D}\hat{\beta}_D(i/n) - \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n))dt. \tag{106}$$

Similarly to the argument of the case that $D_0 \subseteq D$, we have

$$\int_0^1 \sum_{i=1}^n \psi(e_i - (\mathbf{x}'_{i,D}\hat{\beta}_D(i/n) - \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n))t)\mathbf{x}'_{i,D}\hat{\beta}_D(i/n)dt = O_p(n\varphi_n \log n). \tag{107}$$

Similarly to the argument of (98)–(103) and the references therein, we can show that

$$\sup_{t\in(0,1)} |\sum_{i=1}^n \psi(e_i - (\mathbf{x}'_{D,i}\hat{\beta}_D(i/n) - \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n))t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n) - \mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n]| = O_p(\sqrt{n}\log n + n\varphi_n \log n). \tag{108}$$

Recall the assumption that there exist constant M_0, η_0, such that $\inf_{t\in(0,1),|x|\le M_0} f(t,x|\mathcal{F}_1,\mathcal{G}_0) \ge \eta_0 > 0$. Note that $\{\inf_{t\in(0,1),|x|\le M_0} f(t,x|\mathcal{F}_i,\mathcal{G}_{i-1})\}_{i=1}^n$ share the same distribution. Since

$$\mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n] > 0, \tag{109}$$

we have that

$$\begin{aligned}&\mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n] > \\ &\mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n]I(|\mathbf{x}_i| \le M_0/c) \\ &\ge \eta_0(\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}})^2 I(|\mathbf{x}_i| \le c_1),\end{aligned} \tag{110}$$

where $c = \sup_{t\in(0,1)} |\theta(t)|$, $c_1 = M_0/c$. So

$$\mathbb{E}(\mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n]) \ge \eta_0\mathbb{E}\{(\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}})^2 I(|\mathbf{x}_i| \le c_1)\}t. \tag{111}$$

On the other hand, similar argument of Proposition 3 leads to

$$\sup_{t\in(0,1)} \sum_{i=1}^n \{\mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)|\mathcal{G}_n] - \mathbb{E}[\psi(e_i + \mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)t)\mathbf{x}'_{i,\bar{D}}\theta_{\bar{D}}(i/n)]\} = O_p(n^{1/2}). \tag{112}$$

Since $|\mathbf{x}_i|$ are continuous random variables with positive densities on $(0, +\infty)$, by integrating t on $(0,1)$, elementary calculation shows that

$$A(D) \geq \sum_{i=1}^{n} \rho(e_i) + \eta n + O_p(n\varphi_n^2 \log n + n^{1/2} \log n) \tag{113}$$

for some positive number $\eta > 0$. Combining (98), (113), by the fact that $\chi_n = o(1)$ and $\varphi_n b_n + \frac{\pi_n}{\sqrt{nb_n}} = o(\chi_n)$, we get proof. $\square$

Proposition 5. *Under conditions of Theorem 1, we have: a)* $\sup_{t\in[0,1]} \|U(t, \mathcal{F}_0, \mathcal{G}_0)\|_4 < \infty$. *b)* $\sup_{t\in(0,1)} \|U(t, \mathcal{F}_i, \mathcal{G}_i) - U(t, \mathcal{F}_i^*, \mathcal{G}_i^*)\|_4 = O(\bar{\chi}^{|i|})$ *for some* $\bar{\chi} \in (0,1)$. *c) There exists some* $\nu > 1/4$*, such that for* $t, s \in (0,1)$*,* $\|U(t, \mathcal{F}_0, \mathcal{G}_0) - U(s, \mathcal{F}_0, \mathcal{G}_0)\| \leq M|t-s|^\nu$ *for some large enough constant* M*.*

Proof a) follows by the boundedness of $\psi(\cdot)$ and condition [A2]. For b), by the boundedness of $\psi(\cdot)$, condition [A2], Cauchy-Schwarz inequality and triangle inequality, write $H(t, \mathcal{G}_i, \mathcal{G}_i^*) = H(t, \mathcal{G}_i) - H(t, \mathcal{G}_i^*)$ for short,

$$\begin{aligned}
&\sup_{t\in(0,1)} \|U(t, \mathcal{F}_i, \mathcal{G}_i) - U(t, \mathcal{F}_i^*, \mathcal{G}_i^*)\|_4 \leq \qquad (114)\\
&M \sup_{t\in[0,1]} \|\psi(G(t, \mathcal{F}_i^*, \mathcal{G}_i^*) - \psi(G(t, \mathcal{F}_i, \mathcal{G}_i)\|_5 + \sup_{t\in[0,1]} M\|\mathbf{H}(t, \mathcal{G}_i^*) - \mathbf{H}(t, \mathcal{G}_i)\|_4 := A + B.\\
&A \leq \sup_{t\in(0,1)} M\|I(|G(t, \mathcal{F}_i^*, \mathcal{G}_i^*) - G(t, \mathcal{F}_i, \mathcal{G}_i)| \geq \epsilon)\|_5 + \sup_{t\in(0,1)} M\|I(|G(t, \mathcal{F}_i, \mathcal{G}_i)| \leq \epsilon)\|_5\\
&\leq M \left[\frac{\|G(t, \mathcal{F}_i, \mathcal{G}_i) - G(t, \mathcal{F}_i^*, \mathcal{G}_i^*)\|_1}{\epsilon}\right]^{\frac{1}{5}} + M\epsilon^{\frac{1}{5}} = O(\chi^{\frac{|i|}{10}}). \qquad (115)\\
&B \leq \sup_{t\in(0,1)} M\|\mathbf{H}(t, \mathcal{G}_i, \mathcal{G}_i^*)\|_1^{\frac{1}{8}} \|\mathbf{H}(t, \mathcal{G}_i, \mathcal{G}_i^*)\|_7^{\frac{7}{8}} = O(\chi^{\frac{|i|}{8}}). \qquad (116)
\end{aligned}$$

For (c), by triangle inequality, $\|U(t, \mathcal{F}_0, \mathcal{G}_0) - U(t, \mathcal{F}_0, \mathcal{G}_0)\| \leq C + D$ where

$$\begin{aligned}
C &= \|\mathbf{H}(t, \mathcal{G}_i)(\psi(G(t, \mathcal{F}_0, \mathcal{G}_0)) - \psi(G(s, \mathcal{F}_0, \mathcal{G}_0))\|,\\
D &= \|\psi(G(t, \mathcal{F}_i, \mathcal{G}_i))(\mathbf{H}(t, \mathcal{G}_0) - \mathbf{H}(s, \mathcal{G}_0))\|. \qquad (117)
\end{aligned}$$

By the boundedness of $\psi(\cdot)$ and condition [A2], $D < |t-s|$. For C, let $v' \in (2, \frac{4v}{v+1})$, where

v is defined in condition [A1],

$$\begin{aligned}|C| &\leq M\|\psi(G(t,\mathcal{F}_0,\mathcal{G}_0)) - \psi(G(s,\mathcal{F}_0,\mathcal{G}_0)\|_{v'} \\ &\leq M\|I(|G(t,\mathcal{F}_0,\mathcal{G}_0) - G(s,\mathcal{F}_0,\mathcal{G}_0)| \geq \epsilon)\|_{v'} + M\|I(|G(t,\mathcal{F}_0,\mathcal{G}_0)| \leq \epsilon)\|_{v'} \\ &\leq M\left[\frac{|t-s|^{\frac{v}{v'}}}{\epsilon^{\frac{v}{v'}}} + \epsilon^{\frac{1}{v'}}\right] = O(|t-s|^{\frac{v}{v'(v+1)}}). \end{aligned} \quad (118)$$

Then (c) follows from the fact that $\frac{v}{v'(v+1)} > \frac{1}{4}$, which completes the proof. □

Proof of Theorem 5. The theorem follows from Lemma 8 and Lemma 9. □

Define $\tilde{\nu}^2(t) = \sum_{i=1}^n w(t,i)\Delta_i$, where Δ_i is defined in Section 4.

Lemma 8. *Let $\mathfrak{I} = [\gamma_n, 1-\gamma_n] \subset (0,1)$, $\gamma_n = b_n + (m+1)/n$, $m = O(n^{1/3})$. Suppose the conditions of Theorem 5 hold, then*

$$\|\sup_{t\in\mathfrak{I}}|\tilde{\nu}^2(t) - \nu^2(t)|\| = O(\sqrt{\frac{m}{nb_n^2}} + \frac{1}{m} + (\frac{m}{n})^{1/4}\log n + b_n^2). \quad (119)$$

Proof. Write $\mathbf{Q}_i^\diamond = \sum_{j=-m}^m U(t_i,\mathcal{F}_{i+j})$, $\Delta_i^\diamond = \mathbf{Q}_i^\diamond(\mathbf{Q}_i^\diamond)'/(2m+1)$, $\bar{\nu}^2(t) = \sum_{i=1}^n \Delta_i^\diamond w(t,i)$. Then by similar argument of Lemma 3 in Zhou and Wu (2010) and Proposition 2, we have

$$\|\sup_{t\in\mathfrak{I}}|\tilde{\nu}^2(t) - \mathbb{E}(\tilde{\nu}^2(t))|\| = O(\sqrt{\frac{m}{nb_n^2}}). \quad (120)$$

Also by similar argument of Lemma 5 of Zhou and Wu (2010), we can get

$$\sup_{t\in\mathfrak{I}}|\nu^2(t) - \mathbb{E}(\bar{\nu}^2(t))| = O(\frac{1}{m} + b_n^2). \quad (121)$$

Write $\mathcal{P}_i(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_i) - \mathbb{E}(\cdot|\mathcal{F}_{i-1})$, by Proposition 5, note that

$$\|\mathbf{Q}_i - \mathbf{Q}_i^\diamond\| \le \sum_{k=0}^{\infty} \|\sum_{j=-m}^{m} \mathcal{P}_{i+j-k}[U(t_{i+j}, \mathcal{F}_{i+j}, \mathcal{G}_{i+j}) - U(t_i, \mathcal{F}_{i+j}, \mathcal{G}_{i+j})]\|$$
$$= \sum_{k=0}^{\infty} O(\sqrt{m}) \min\{(\frac{m}{n})^{1/4}, \chi^{|k|}\}. \quad (122)$$

By similar argument of Lemma 4 in Zhou and Wu (2010), we have

$$\sup_{t\in\mathfrak{I}} |\mathbb{E}(\bar{\nu}^2) - \mathbb{E}(\tilde{\nu}^2)| = O((\frac{m}{n})^{1/4} \log n). \quad (123)$$

The lemma follows by (120) (121) (123).

Lemma 9. *Under the conditions of Lemma 8, and* $\frac{m \log^5 n}{\sqrt{nb_n}} + mb_n^2 \log^2 n \to 0$*, then*

$$\sup_{t\in\mathfrak{I}} |\tilde{\nu}^2(t) - \hat{\nu}^2(t)| = O_p((nb_n)^{-1/4} b_n^{-1/2} \log^3 n + b_n^{1/2} \log^{3/2} n). \quad (124)$$

Proof. For some large enough constant M, write $W_{1n} = \{\sup_{t\in(0,1)} |\hat{\theta}(t) - \theta(t)| \le M((nb_n)^{-1/2} \log^4 n + b_n^2 \log n)\}$, $W_{2n} = \{\max_{1\le i\le n} |\mathbf{x}_i| \le M \log n\}$. Write

$$M_i(\delta) = \sum_{j=-m}^{m} [\psi(e_{i+j}) - \psi(e_{i+j} - |\mathbf{x}_{i+j}||\delta|)]|\mathbf{x}_{i+j}|.$$

Similarly to Lemma 1, let $g_n \to \infty$ be a real sequence which is allowed to go to infinity at arbitrarily slow rate. Define

$$\delta_n = \frac{\log^4 n}{\sqrt{nb_n}} + b_n^2 \log n,\ \phi_{m,n} = 2g_n\sqrt{m\delta_n}\log n,\ t_{m,n} = \frac{g_n\sqrt{m\delta_n}}{\log g_n},\ u_{m,n} = t_{m,n}^2,$$
$$\eta_s(\theta) = [\psi(e_s - |\mathbf{x}_s'||\theta|) - \psi(e_s)]|\mathbf{x}_s|,\ T_{i,m} = \max_{i-m\le s\le i+m} \sup_{|\theta|\le\delta_n} |\eta_s(\theta)|,$$
$$U_{i,m} = \sum_{j=-m}^{m} \mathbb{E}\{[\psi(e_{i+j} + |\mathbf{x}_{i+j}||\delta|) - \psi(e_{i+j} - |\mathbf{x}_{i+j}||\delta|]^2|\mathcal{F}_{i-1}, \mathcal{G}_n\}|\mathbf{x}_{i+j}|^2.$$

Let $l = n^8$, $\mathfrak{G}_l = \{k/l : k \in \mathbb{Z}, k \le n^9\}$. Then by Freedman (1975), for any $\zeta > 1$,

$$\mathbb{P}[\sup_{\delta \in \mathfrak{G}_l} |M_i(\delta) - M_i(0)| \ge 2\phi_{m,n}, T_{i,m} \le t_{m,n}, U_{i,m} \le u_{m,n}] = O(n^{-\zeta p}). \tag{125}$$

Let $A^m_{n,i} = \{\sup_{\delta \in \mathfrak{G}_l} |M_i(\delta) - M_i(0)| \ge 2\phi_{m,n}\}$, $A_n = \cup_{m+1 \le i \le n-m} A^m_{n,i}$, then $\lim_{n\to\infty} \mathbb{P}(\bar{A}_n) = 1$. Let $W_n = W_{1n} \cap W_{2n} \cap \bar{A}_n$. Then $\lim_{n\to\infty} \mathbb{P}(W_n) = 1$. By Proposition 8, it's enough to consider

$$\begin{aligned} \sup_{t \in \mathfrak{I}} |(\tilde{\nu}^2(t) - \hat{\nu}^2(t)) I(W_n)| &= \sup_{t \in \mathfrak{I}} |\sum_{i=1}^n w(t,i)(\Delta_i - \hat{\Delta}_i) I(W_n)| \\ &= \sup_{t \in \mathfrak{I}} |\sum_{i=1}^n w(t,i)[(\mathbf{Q}_i \mathbf{Q}_i' - \hat{\mathbf{Q}}_i \hat{\mathbf{Q}}_i') I(W_n)/(2m+1)]|. \end{aligned} \tag{126}$$

Note that

$$\|\mathbf{Q}_i(\mathbf{Q}_i' - \hat{\mathbf{Q}}_i') I(W_n)\| \le \|\mathbf{Q}_i\|_4 \|(\mathbf{Q}_i' - \hat{\mathbf{Q}}_i') I(W_n)\|_4. \tag{127}$$

Using the argument of Theorem 1 in Wu (2007), we have $\sup_i \|\mathbf{Q}_i\|_4 = O(\sqrt{m})$. Observe that for $i = m+1, ..., n-m$,

$$|(\mathbf{Q}_i - \hat{\mathbf{Q}}_i) I(W_n)| \le \sup_{|\delta| \le \delta_n} M_i(\delta) I(W_{2n} \cap \bar{A}_n). \tag{128}$$

By using martingale decomposition technique in Lemma 1 and Lemma 2,

$$\|(\mathbf{Q}_i' - \hat{\mathbf{Q}}_i') I(W_n)\|_4 = O(m^{1/2} \delta_n^{1/2} \log n), \tag{129}$$

So by (127),

$$\|\mathbf{Q}_i(\mathbf{Q}_i' - \hat{\mathbf{Q}}_i') I(W_n)\|/(2m+1) = O(\delta_n^{1/2} \log n). \tag{130}$$

Similarly,

$$\|[\mathbf{Q}_i - \hat{\mathbf{Q}}_i] \hat{\mathbf{Q}}_i' I(W_n)\|/(2m+1) = O(\delta_n^{1/2} \log n). \tag{131}$$

So we have

$$\|(\tilde{\nu}^2(t) - \hat{\nu}^2(t))I(W_n)\| = O(\delta_n^{1/2}\log n). \tag{132}$$

Similar argument yields that

$$\|\frac{\partial}{\partial t}(\tilde{\nu}^2(t) - \hat{\nu}^2(t))I(W_n)\| = O(\delta_n^{1/2}b_n^{-1}\log n). \tag{133}$$

Then the lemma follows from Proposition 2. □

Proof of theorem 6. The theorem follows from Lemma 10 and Lemma 11. □

Write $\Lambda(t) = \frac{1}{nb_n}\sum_{i=1}^n f(i/n, 0|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i' K_{b_n}(i/n - t)$, $\tilde{\Sigma}(t) = \frac{1}{nb_nc_n}\sum_{i=1}^n \phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i' K_{b_n}(i/n - t)$.

Lemma 10. *Suppose the conditions of Theorem 2 hold, and $c_n \to 0$, $nc_n \to \infty$, then*

$$\sup_{t\in\mathfrak{T}_n}|\tilde{\Sigma}(t) - \Sigma(t)| = O_p(\frac{\log^2 n}{\sqrt{nc_n}b_n} + c_n^2 b_n^{-1}\log n + b_n). \tag{134}$$

Proof. Write $W_n = \{\max_i |\mathbf{x}_i| \le M\log n\}$ for some large constant M. Then $\mathbb{P}(W_n) \to 1$ as $n \to \infty$. Then we have

$$\begin{aligned}(\tilde{\Sigma}(t) - \mathbb{E}(\tilde{\Sigma}(t)|\mathcal{G}_n))I(W_n) &= \sum_{i=1}^n \left[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i' - \mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\frac{K_{b_n}(i/n-t)}{nb_nc_n}I(W_n)\\ &+ \sum_{i=1}^n \left[\mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'|\mathcal{F}_{i-1},\mathcal{G}_n] - \mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'|\mathcal{G}_n]\right]\frac{K_{b_n}(i/n-t)}{nb_nc_n}I(W_n)\\ &:= A_n(t) + B_n(t).\end{aligned} \tag{135}$$

By the property of martingale difference, and Jensen's inequality, we have

$$\|A_n(t)\|^2 \le \frac{M\log^4 n}{(nb_n)^2}\sum_{i=1}^n \|\frac{\phi(\frac{e_i}{c_n})}{c_n}\|^2 K_{b_n}^2(i/n - t) \tag{136}$$

for some large constant M. Elementary calculation shows that $\|\frac{\phi(\frac{e_i}{c_n})}{c_n}\| \le \frac{M}{\sqrt{c_n}}$. Consequently,

$$\|A_n(t)\| = O(\frac{\log^2 n}{\sqrt{nb_nc_n}}). \tag{137}$$

Similar argument yields that

$$\|\frac{\partial}{\partial t}A_n(t)\| = O(\frac{\log^2 n}{\sqrt{nb_nc_n}b_n}). \tag{138}$$

Then by Proposition 2, we have that $\|\sup_{t\in\mathfrak{T}_n}|A_n(t)|\| = O(\frac{\log^2 n}{\sqrt{nb_nc_n}b_n^{1/2}})$. On the other hand,

$$B_n(t) = \sum_{i=1}^n\sum_{k=1}^\infty \mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'I(W_n)|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\frac{K_{b_n}(i/n-t)}{nb_nc_n}. \tag{139}$$

As a consequence,

$$\begin{aligned}
\|B_n(t)\| &\le \sum_{k=1}^\infty\|\sum_{i=1}^n \mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'I(W_n)|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\frac{K_{b_n}(i/n-t)}{nb_nc_n}\| \\
&\le \sum_{k=1}^\infty\left\{\sum_{i=1}^n\|\mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})\mathbf{x}_i\mathbf{x}_i'I(W_n)|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\frac{K_{b_n}(i/n-t)}{nb_nc_n}\|^2\right\}^{1/2} \\
&\le \sum_{k=1}^\infty\left\{\sum_{i=1}^n\|\mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\|^2(\frac{K_{b_n}(i/n-t)}{nb_nc_n}\log^2 n)^2\right\}^{1/2}.
\end{aligned} \tag{140}$$

On the other hand, since $\mathbb{E}[\phi(\frac{e_i}{c_n})|\mathcal{F}_{i-1},\mathcal{G}_n] = \int\phi(\frac{x}{c_n})f(i/n,x|\mathcal{F}_{i-1},\mathcal{G}_n)dx$, elementary calculation shows that

$$\begin{aligned}
\|\mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\| &= \|c_n\int\phi(y)(\mathcal{P}_{i-k,n}f(i/n,c_ny|\mathcal{F}_{i-1},\mathcal{G}_n))dy\| \\
&\le c_n\int\phi(y)\|f(i/n,c_ny|\mathcal{F}_{i-1},\mathcal{G}_n) - f(i/n,c_ny|\mathcal{F}_{i-1}^{(i-k)},\mathcal{G}_n)\|dy,
\end{aligned} \tag{141}$$

where the last inequality follows from Lemma 1 of Wu (2007). By condition [A3], and Proposition 6,

$$\|\mathcal{P}_{i-k,n}\left[\mathbb{E}[\phi(\frac{e_i}{c_n})|\mathcal{F}_{i-1},\mathcal{G}_n]\right]\| = O(c_n\chi^{\frac{k-1}{2}}).$$

So we have $\|B_n(t)\| = O(\frac{\log^2 n}{\sqrt{nb_n}})$. Similar argument yields that $\|\frac{\partial}{\partial t}B_n(t)\| = O(\frac{\log^2 n}{\sqrt{nb_n}b_n})$. By Proposition 2, $\|\sup_{t\in(0,1)}|B_n(t)|\| = O(\frac{\log^2 n}{\sqrt{nb_n}})$. Hence we have

$$\sup_{t\in\mathfrak{T}_n}|\tilde{\Sigma}(t) - \mathbb{E}(\tilde{\Sigma}(t)|\mathcal{G}_n)| = O_p(\frac{\log^2 n}{\sqrt{nc_n}b_n}). \tag{142}$$

By elementary calculation,

$$\begin{aligned}
&|\mathbb{E}(\tilde{\Sigma}(t)|\mathcal{G}_n) - \Lambda(t)|I(W_n) \\
&\quad \le \frac{I(W_n)}{nb_n}\sum_{i=1}^{n}\left|\mathbb{E}\{\frac{\phi(\frac{e_i}{c_n})}{c_n}|\mathcal{G}_n\} - f(i/n,0|\mathcal{G}_n)\right||\mathbf{x}_i\mathbf{x}_i'|K_{b_n}(i/n-t) \\
&\quad \le Mc_n^2\log^2 n\sum_{i=1}^{n}K_{b_n}(i/n-t)/(nb_n) \le Mc_n^2\log^2 n. \qquad (143)
\end{aligned}$$

Apply similar argument to $\frac{\partial}{\partial t}(\mathbb{E}(\tilde{\Sigma}(t)|\mathcal{G}_n) - \Lambda(t))I(W_n)$ and then apply Proposition 2, we shall have

$$\sup_{t\in\mathfrak{T}_n}|\mathbb{E}(\tilde{\Sigma}(t)|\mathcal{G}_n) - \Lambda(t)| = O_p(c_n^2 b_n^{-1}\log^2 n). \tag{144}$$

Finally, the lemma follows from Proposition 3 and the fact that $\sup_{t\in(0,1)}|\mathbb{E}\Lambda(t) - \mathbb{E}\Lambda_n^*(t)| = O(b_n)$, where $\Lambda_n^*(t)$ is the $p\times p$ matrix formed by the first p rows and columns of $\Lambda_n(t)$ (which is defined in (64)) . $\square$

Lemma 11. *Suppose that the conditions of Lemma 10 hold, then*

$$\sup_{t\in\mathfrak{T}_n}|\tilde{\Sigma}(t) - \hat{\Sigma}(t)| = O_p(b_n\log^4 n + \frac{\log^7 n}{\sqrt{nc_n}b_n} + c_n^2 b_n^{-1}\log^2 n) \tag{145}$$

Proof. Note that $e_i(t) = y_i - \mathbf{x}_i'\theta(t)$. Write

$$\Sigma^{\diamondsuit}(t) = \frac{1}{nb_n c_n}\sum_{i=1}^{n}\phi(\frac{e_i(t)}{c_n})\mathbf{x}_i\mathbf{x}_i'K_{b_n}(i/n-t). \tag{146}$$

Using the factor that $\hat{e}_i(t) = e_i(t) + \mathbf{x}_i'(\theta(t) - \hat{\theta}(t))$, the martingale decomposition technique in Lemma 1 and the similar argument in Lemma 9, we can show that

$$\sup_{t\in\mathfrak{T}_n}|\Sigma^{\diamondsuit}(t)-\hat{\Sigma}(t)| = O_p(b_n\log^4 n+\frac{\log^7 n}{\sqrt{nc_n}b_n}+c_n^2b_n^{-1}\log^2 n). \tag{147}$$

On the other hand, by using Proposition 2, we can show that

$$\sup_{t\in\mathfrak{T}_n}|\Sigma^{\diamondsuit}(t)-\tilde{\Sigma}(t)| = O_p(\frac{\log^7 n}{\sqrt{nc_n}b_n}+c_n^2b_n^{-1}\log^2 n+b_n\log^3 n), \tag{148}$$

which completes the proof. □

Lemma 12. *Suppose [A3] [A4] hold. Let* $b_n\to 0$, $\frac{nb_n^4}{\log^7 n}\to\infty$. *Let* $\bar{\lambda}_n(t)$ *be the smallest eigenvalue of* $\sum_{i=1}^n f(i/n,0|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'K_{b_n}(i/n-t)$, *then* $\mathbb{P}\{\liminf_{n\to\infty}\inf_{t\in(0,1)}\frac{\bar{\lambda}_n(t)}{nb_n}\geq\eta>0\}\to 1$.

Proof. Similar to Proposition 3, we can show that

$$\sup_{t\in(0,1)}|\frac{\sum_{i=1}^n f(i/n,0|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'K_{b_n}(i/n-t)}{nb_n}-\Sigma(t)| = O_p(b_n). \tag{149}$$

Then the lemma holds in view of [A4]. □

Proposition 6. *Suppose [A3], [A4], then we have for any* $s>1$, $\delta_s(k)=O(\chi^{k/s})$.

Proof. Let M_0 be a large constant such that $\sup_{t\in(0,1),x\in\mathbb{R}}|F^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)|\leq M_0$. By the boundedness of $F^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)$, we have

$$\begin{aligned}&\mathbb{E}\{|F^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)-F^{(q)}(t,x|\mathcal{F}_{k-1}^*,\mathcal{G}_k)|^s\}\\ &\qquad\leq (2M_0)^{s-1}\mathbb{E}\{|F^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)-F^{(q)}(t,x|\mathcal{F}_{k-1}^*,\mathcal{G}_k)|\}.\end{aligned}$$

By the definition of $\delta_s(k)$ and condition [A3], we get proof. □

Proposition 7. *Suppose [A2]. Then i)* $\max_{1\leq i\leq n}\|x_i\|_k\leq Ck$ *for any positive integer* k *and some constant* C. *In addition, ii) there exists a large constant* M_0, *such that* $\lim_{n\to\infty}\mathbb{P}(\max_{1\leq i\leq n}|\mathbf{x}_i|\leq M_0\log n)=1$.

Proof. By the assumption that $\max_{1\leq i\leq n}\mathbb{E}(\exp(t_x|\mathbf{x}_i|)\leq M$, we have that for any positive integer k, $\max_{1\leq i\leq n}t_x^k\mathbb{E}(|\mathbf{x}_i|^k)/k!\leq M$. Then i) holds by $k!\leq k^k$. For ii), write

$X_n = \max_{1\le i\le n} |\mathbf{x}_i|$. Then

$$\mathbb{P}(X_n \ge M_0 \log n) \le \frac{\mathbb{E}(\exp(t_x X_n))}{\exp(M_0 t_x \log n)} \tag{150}$$

Since $\mathbb{E}(\exp(t_x X_n)) \le \sum_{i=1}^n \mathbb{E}(\exp(t_x|\mathbf{x}_i|)) \le nM$, take M_0 large such that $t_x M_0 > 1$ and we get proof. $\square$

Proposition 8. *Suppose sets A_n satisfy $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$. For any sequence of random variables X_n, if $X_n I(\bar{A}_n) = O_p(1)$, then $X_n = O_p(1)$, where $I(\cdot)$ is the usual indicator function.*

Proof. For any $\epsilon > 0$, let N be a large constant such that $\mathbb{P}(A_n) \le \epsilon/2$ for $n \ge N$, and M be a large constant such that $\mathbb{P}(|X_n|I(\bar{A}_n) \ge M/2) \le \epsilon/2$ for $n \ge N$. Then

$$\begin{aligned}\mathbb{P}(|X_n| \ge M) &\le \mathbb{P}(|X_n|I(A_n) \ge M/2) + \mathbb{P}(|X_n|I(\bar{A}_n) \ge M/2)\\ &\le \mathbb{P}(A_n) + \mathbb{P}(|X_n|I(\bar{A}_n) \ge M/2) \le \epsilon \end{aligned}\tag{151}$$

for all $n \ge N$. $\square$

REFERENCES

Abramovich, Y. I. , Spencer, N. K. , Turley, M. D. E. (2007), Order estimation and discrimination between stationary and time-varying (TVAR) autoregressive models. *IEEE Trans. Signal Process*, **55** 2861–2876.

Ando, T. and Tsay, R. S. (2011), Quantile regression models with factor–augmented predictors and information criterion, *Econometrics Journal*, **14** 1–24.

Babu, G. J. (1989). Strong representations for LAD estimators in linear models. *Probability Theory and Related Fields*, **83** 547–558.

Cai, Z. (2007) Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, **136**, 163–188.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coeffcient models. *Journal of the American Statistical Association*, **103** 1595–1608.

Chaudhuri, P. (1991), Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation. *The Annals of Statistics*, **19** 760–777.

Cole, T. J. (1994), Growth charts for both cross-sectional and longitudinal data, *Statistics in Medicine*, **13** 2477–2492.

Craven, P. and Wahba, G. (1979), Smoothing noisy data with spline functions. *Numer. Math.*, **31** 377–403.

Draghicescu, D. , Guillas, S. and Wu, W. B. (2009), Quantile curve estimation and visualization for non-stationary time series. *J. Comput. Graph. Statist.* , **18** 1–20.

Eubank, R. L. and Speckman, P. L. (1993), Confidence bands in nonparametric regression. *J. Am. Statist. Ass.*, **88** 1287–1301.

Fan, J. and Jiang, J. (2005), Nonparametric inference for additive models. *J. Amer. Statist. Assoc.*, **100** 890–907.

Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* , **96** 1348–1360.

Fan, J., Yao, Q. and Cai, Z. (2003), Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* , **65** 57–80.

Fan, J. and Zhang, W. (1999), Statistics estimation in varying coefficient models. *The Annals of Statistics*, **27** 1491–1518.

Fan, J. , Zhang, C. , Zhang, J. (2001), Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* , **29** 153–193.

Freedman, D. A. (1975), On tail probabilities for martingales. *Ann. Probab.* , **3** 100–118.

Hall, P. (1991) On the convergence rate of suprema. *Probability Theory and Related Fields*, **89** 447–455.

Hall, P. , Titterington, D. M. (1988), On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* , **27** 228–254.

Härdle, W. , Mammen, E. (1993), Comparing nonparametric versus parametric regression fits. *Ann. Statist.* , **21** 1926–1947.

Härdle, W. and Marron, J. S. (1991), Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* , **19**, 778–796.

Hastie, T. and Tibshirani, R. (1993), Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** 757–796.

He, X. and Shao, Q. (1996), A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs, *Ann. Statist.* , **24** 2608–2630.

He, X. and Zhu, L.-X. (2003), A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.* , **98** 1013–1022.

Hendricks, W. and Koenker, R. (1992), Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Amer. Statist. Assoc.* , **87** 58–68.

Horowitz, J. L. , Spokoiny, V. G. (2001), An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* **69** 599–631

Horowitz, J. L. , Spokoiny, V. G. (2002), An adaptive, rate-optimal test of linearity for median regression models. *J. Amer. Statist. Assoc.* , **97** 822–835.

Hoover, D. R. , Rice, J. A. , Wu, C. O. and Yang, L.–P. (1998), Nonparametric Smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85** 809–822.

Honda, T. (2004), Quantile regression in varying coefficient models. *J. Statist. Plann. Inference*, **121** 113–125.

Huang, J. Z., Wu, C. O. and Zhou, L. (2004), Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sin.* , **14** 763–788.

Ingster, Y. I. (1993), Asymptotically minimax hypothesis testing for nonparametric alternatives. ICIII. *Math. Methods Statist.* **2** 85–114; **3** 171–189; **4** 249–268.

Kim, M. (2007), Quantile regression with varying coefficients. *Annals of Statistics*, **35** 92–108.

Koenker, R. , (2004), Quantile regression for longitudinal data. *J. Multivariate Anal.* **91** 74–89.

Koenker, R. , (2005), *Quantile Regression*, Cambridge Books, Cambridge University Press.

Koenker, R. , and Xiao, Z. (2006), Quantile Autoregression. *Journal of the American Statistical Association*, **101** 980–1006.

Lepski, O. V. and Spokoiny, V. G. (1999), Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358.

Li, Y. and Zhu, J. (2005), L_1-norm quantile regressions. *J. Comput. Graph. Statist.* **17** 163–185.

Neumann, M. H. and Kreiss, J. P. (1998), Regression-type inference in nonparametric autoregression. *Ann. Statist.* , **26** 1570–1613.

Neumann, M. H. and Polzehl, J. (1998), Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparam. Statist.* , **9** 307–333.

Oka, T. and Qu, Z. (2011), Estimating Structural Changes in Regression Quantiles, *Journal of Econometrics*, **162** 248–267.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2005), Nonparametric estimation of time-varying parameters under shape restrictions. *Journal of Econometrics*, **126** 53–77.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2006), On the estimation and testing of time varying constraints in econometric models. *Statistica Sinica*, **16** 1313–1333.

Powell, J. L. (1991), Estimation of monotonic regression models under quantile restrictions, *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Barnett, W. A. , Powell, J. L. and Tauchen, G. E. eds.* Cambridge: Cambridge University Press.

Robinson, P. M. (1989), Nonparametric estimation of time-varying parameters. *Statistical Analysis and Forecasting of Economic Structural Change* (ed. Hackl, P.), pp. 164–253. Berlin: Springer.

Robinson, P. M. (1991), Time-varying nonlinear regression. *Economic Structure Change Analysis and Forecasting* (ed. Hackl, P. and Westland, A. H.), pp. 179–290. Berlin: Springer.

Rosemkrantz, W. (2000), Confidence bands for quantile functions: A parametric and graphic alternative for testing goodness-of-fit. *Amer. Statist.* , **54** 185–190.

Stute, W. (1997), Nonparametric model checks for regression. *Ann. Statist.* , **25** 613–641.

Tang, Y. , Wang, H. , Zhu, Z. (2013), Variable selection in quantile varying coefficient models with longitudinal data. *Computational Statistics and Data Analysis,* **57** 435–449.

Wang, L., (2007), A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model. *Statist. Probab. Lett.* , **77** 1034–1042.

Wang, L., (2008), Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canad. J. Statist.* , **36** 321–336.

Wang, H. , Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *J. Business & Economic Statistics*, **25** 347–355.

Wei, Y. , and He, X. (2006), Conditional growth charts (with discussion), *Annals of Statistics*, **34** 2069–2097.

Wei, Y, Pere, A., Koenker, R., and He, X. (2006), Quantile Regression Methods for Reference Growth Charts, *Statistics in Medicine*, **25** 1369–1382.

Wu, C. O. , Chiang, C. T. and Hoover, D. R. (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Am. Statist. Ass.* , **93** 1388–1402.

Wu, W. B. (2007), M-estimation of linear models with dependent errors. *Annals of Statistics*, **35** 495–521.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica*, **19** 1755–1768.

Wu, W. B. and Zhou, Z. (2011), Gaussian approximations for non-stationary multiple time series. *Statist. Sin*, **21** 1397–1413.

Wu, Y. and Liu, Y. (2009), Variable selection in quantile regression. *Statistica Sinica*, **19** 801–817.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Stat. Soc. Ser. B Stat. Methodol*, **60** 797–811.

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

Zhang, T and Wu, W. B. (2012), Inference of time-varying regression models, *Annals of Statistics*, **3** 1376–1402.

Zhang, W., Lee, S.–Y. and Song, X. (2002), Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, **82** 166–188.

Zheng, J. (1996), A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics*, **75** 263–289.

Zheng, J. X. (1998), A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* **14** 123–138.

Zhou, Z. (2010), Nonparametric inference of quantile curve for nonstationary time series, *Annals of Statistics*, **38** 2197–2217.

Zhou, Z. and Wu, W. B. (2009), Local Linear Quantile Estimation for Non-stationary Time Series. *Annals of Statistics*, **37** 2696–2729.

Zhou, Z. and Wu, W. B. (2010) Simultaneous inference of linear models with time varying coefficients, *J. R. Stat. Soc. Ser. B Stat. Methodol,* **72** 513–531.

Zou, H. (2006), The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* , **101** 1418–1429.