STRUCTURAL CHANGE DETECTION FOR REGRESSION QUANTILES UNDER TIME SERIES NON-STATIONARITY

WEICHI WU[1] AND ZHOU ZHOU

University of Toronto

September 15, 2014

Abstract

We consider quantile structural change testing for linear models with random designs and a wide class of non-stationary regressors and errors. New uniform Bahadur representations are established with nearly optimal approximation rates. Two cusum-type test statistics, one based on the regression coefficients and the other based on the gradient vectors are considered. Two of the most frequently used change point testing procedures, pivotalization and independent wild bootstrap, are shown to be inconsistent for non-stationary time series quantile regression. In this paper, simple bootstrap methods are proposed and are proved to be consistent for regression quantile structural change detection under both abrupt and smooth non-stationarity and temporal dependence. Our bootstrap procedures are shown to have certain asymptotically optimal properties in terms of accuracy and power. Our methodology is applied to the USA real GDP series, and asymmetry of structural changes in different quantiles are found.

1 Introduction

Since the seminal work of Koenker and Bassett (1978), there has been an enormous interest in statistics and econometrics on quantile regression and its applications. We refer to Koenker (2005) and the references therein for a comprehensive account of the topic. Consider the following parametric time series quantile regression model:

$$y_i = \mathbf{x}_i'\beta(\alpha) + e_i(\alpha), \tag{1}$$

[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. Quantile regression, piece-wise locally stationary, bootstrap, CUSUM, structural change

where $\{\mathbf{x}_i\}_{i=1}^n$ and $\{e_i(\alpha)\}_{i=1}^n$ are the p-dimensional predictor time series and error series, respectively; $0 < \alpha < 1$ and the αth conditional quantile $Q_\alpha(e_i(\alpha)|\mathbf{x}_i) = 0$. Due to its simplicity and interpretability, parametric model (1) is frequently used in practice compared to its nonparametric counterparts.

To justify the use of a parametric model, certain lack-of-fit or specification tests are essential diagnostics. Despite the large amount of work on parameter estimation and inference of quantile regression, there is much less work on lack-of-fit tests for regression quantiles, especially under the time series framework. For i.i.d. samples, Zheng (1998) and Dette et al. (2011), among others, proposed specification tests based on discrepancies between parametric and nonparametric quantile regression estimates. He and Zhu (2003) proposed a lack-of-fit test of regression quantiles based on the cusum process of the gradient vector under the parametric null hypothesis. The advantage of the cusum test is that it does not require nonparametric fitting under the alternative and it can detect local alternatives with the $1/\sqrt{n}$ parametric rate. For other contributions of quantile specification tests for independent data, see Koenker and Machado (1999), Horowitz and Spokoiny (2002) and Wang (2008), among others. For regression quantiles of dependent data, among others, Qu (2008) evaluated cusum tests based the gradient vectors and regression coefficients for nearly stationary processes with martingale difference dependence structure. Su and Xiao (2008) presented a Wald-type test of parameter stability for stationary, ergodic data.

The purpose of the paper is to diagnose or test whether the parameter $\beta(\alpha)$ stays unchanged over time in model (1). We investigate two types of test statistics, one based on the cusum process of the gradient vectors and the other based on the cusum process of the regression coefficients. The most significant contributions of the paper lie in the following two aspects. First, we investigate the behaviors of regression quantiles and their residual processes under a general nonlinear and non-stationary time series framework and discover that traditional inferential methods for quantile regression fail under such complex temporal dynamics. Specifically, following Zhou (2013), we allow the regressors $\{\mathbf{x}_i\}$ and the errors $\{e_i(\alpha)\}$ to experience both smooth and sudden nonlinear changes in their marginal distributions and dependence structures over time. Such nonlinear and non-stationary modelling of the regressors and errors could be realistic and flexible in many time series applications; see for instance the USA real GDP series analyzed in Section 6. Under the above settings, we establish a uniform Bahadur representation of the partial sample quantile estimates with nearly optimal approximation rates and derive the limiting behaviors of the above two tests. Traditionally when dealing with stationary data, the regression coefficient cusum test is shown to be asymptotically pivotal (Qu 2008) and the gradient cusum test is advocated over the regression coefficient test as it is asymptotically free of the densities of the errors (He and Zhu (2003), Qu (2008)). However, as we discover in this paper, those properties no longer hold for non-stationary time series quantile regression. Consequently, we discover in this paper that both the classic way of structural change testing by checking the quantiles of the maxima of certain pivotal Gaussian processes (Qu 2008) and the independent wild bootstrap procedure in He and Zhu (2003) lead to biased testing results for non-stationary time series quantile regression.

Second, we propose in this paper a bootstrap procedure which is consistent for structural change tests of time series quantile regression with both abruptly and smoothly time-varying temporal dynamics. To our knowledge, there have been no results on structural change tests for time series quantile regression with non-stationary covariates and errors in the literature. For change point tests of the mean, Zhou (2013) proposed a bootstrap procedure which is robust to general forms of non-stationarity in the time series. However, it is highly non-trivial to extend such bootstrap procedures to regression quantiles. In particular, a naive extension of Zhou (2013) by progressively convoluting the partial sample quantile regression estimates and i.i.d. standard normals will not yield a consistent test. In this paper, we propose a bootstrap procedure by combining an extension of the Powell's sandwich estimates (Powell 1991) and a progressive convolution of the block sums of the estimated gradient vectors with i.i.d. standard normal auxiliary random variables. The bootstrap procedure is shown to be consistent with Type I errors approaching the nominal no slower than the nearly optimal approximation rate of the Bahadur representations. Meanwhile, we prove that our bootstrap can detect local alternatives with the optimal $1/\sqrt{n}$ parametric rate.

There is a large amount of related work in testing structural stability of parameters in least squares regressions and various other scenarios. Among them, Brown et al. (1975), McCabe and Harrison (1980) developed CUSUM tests with *i.i.d.* normal errors. Ploberger and Krämer (1992) extended such tests to stationary and ergodic errors. Andrews (1993) established Wald-type, LM, LR-like tests based on partial-sample GMM estimators with strong mixing assumptions. These test statistics are constructed through coefficients estimated by different portions of data. On the other hand, there are also a class of tests which heavily depend on the residuals of the least squares regression. For example, Bai (1996) obtained asymptotically distribution free test statistics associate with *i.i.d.* errors; see also Bai and Perron (1998) for tests of multiple structural changes. We also refer to the recent review of Aue and Horvath (2013) for more discussions and references.

The rest of the paper the organized as follows. In Section 2 we shall investigate quantile regression under non-stationary and nonlinear dependence and establish the Bahadur representation and related asymptotic results. Section 3 proposes the structural change tests and the bootstrap and investigates their asymptotic Type I error and power behaviors. In Section 4, we extend our structural change tests and bootstrap procedures to testing structural stability of finite many different regression quantiles. In Section 5, we perform moderate sample Monte Carlo experiments to study the finite sample behaviors of the tests and compare our bootstrap with classic testing procedures. Section 6 contains a empirical illustration with the USA GDP series.

2 Quantile regression under time series non-stationarity.

We first introduce some notation. Define $X_n \geq_p Y_n$ as that $\mathbb{P}(X_n \geq Y_n) \to 1$ as $n \to \infty$. Similarly we define "$\leq_p$". For a p-dimensional vector v, define $|v| = \sqrt{\sum_{i=1}^p v_i^2}$. For an $m \times n$ matrix A, define $|A| = \sqrt{trace(AA^T)}$. For random variable X, let $\|X\|_q$ be its L_q norm. For any semi-positive definite matrix Σ, let $\lambda_1(\Sigma)$ be its smallest eigenvalue. For a p-dimensional random vector v, define $\|v\|_q = \||v|\|_q$. For $m \times n$ random matrix A, define $\|A\|_q = \||A|\|_q$. We omit subscript q of $\| \cdot \|_q$ if $q = 2$ when there is no confusion caused. For filtration $\mathcal{F}_i = (..., \eta_{i-1}, \eta_i)$, write $\mathcal{F}_i^{(j)} = (..., \eta_{j-1}, \eta_j', \eta_j, ..., \eta_i)$ for $j \leq i$, where $(\{\eta_i\}_{i=-\infty}^{\infty}, \eta_j')$ are *i.i.d* random variables. Write $\mathcal{F}_i^*$ for $\mathcal{F}_i^{(0)}$, $t_i = i/n$, and write $N = \lfloor \frac{n}{\log n} \rfloor$ for short. Then we introduce the piecewise locally stationary (PLS) processes (Zhou 2013).

Definition 1. *For $k < \infty$, we say that $\{e_i\}_{i=1}^n$ is PLS w.r.t. filtrations $\mathcal{F}_{1i}$, $\mathcal{F}_{2i}$,..., $\mathcal{F}_{ki}$ with r breaks (PLS(r, $\mathcal{F}_{1i}$, $\mathcal{F}_{2i}$,..., $\mathcal{F}_{ki}$)) if there exist constants $0 = b_0 < b_1 < .... < b_r < b_{r+1} = 1$ and nonlinear filters $G_0, G_1, ..., G_r$, such that*

$$e_i = G_j(t_i, \mathcal{F}_{1i}, ..., \mathcal{F}_{ki}), \text{ if } b_j < t_i \leq b_{j+1}, \tag{2}$$

where $t_i = i/n$, $\mathcal{F}_{li} = \{..., \varepsilon_{l0}, \varepsilon_{l1}, ..., \varepsilon_{li}\}$ *for* $1 \le l \le k$*. For each* l, $\{\varepsilon_{li}\}_{i=-\infty}^{\infty}$ *are i.i.d r.v's. For* $l \neq s$, $\{\varepsilon_{li}\}_{i=-\infty}^{\infty}$ *and* $\{\varepsilon_{si}\}_{i=-\infty}^{\infty}$ *are independent.*

Note that in the definition, the functions G_0,...,G_r and the break points $b_1, ..., b_r$ are unknown nuisance parameters. If $G_j(t, \cdot)$ is a smooth function in t, then e_i changes smoothly on (b_j, b_{j+1}), $j = 0, ..., r$. The smooth change is interrupted at break points $b_1, ..., b_r$ where the time series can experience abrupt changes in its data generating mechanism. To quantify the temporal dependence of PLS processes, we shall introduce the following physical dependence measures:

Definition 2. *Consider the* $PLS(r, \mathcal{F}_{1i}, ..., \mathcal{F}_{ki})$ *process* $\{e_i\}_{i=-\infty}^{\infty}$ *defined in (2). Assume that* $\max_{1\le i\le n} \|e_i\|_p < \infty$ *for some* $p > 0$*. The* l_{th} *dependence measure for* $\{e_i\}_{i=-\infty}^{\infty}$ *in* L_p *norm,* $\Delta_p(l)$*, is defined as*

$$\Delta_p(l) = \max_{0\le i\le r} \sup_{b_i<t\le b_{i+1}} \|G_i(t, \mathcal{F}_{1l}, ..., \mathcal{F}_{kl}) - G_i(t, \mathcal{F}^*_{1l}, ..., \mathcal{F}^*_{kl})\|_p. \tag{3}$$

If we view e_i as the output of a physical system which is driven by innovations $\{\varepsilon_{si}\}_{i=-\infty}^{\infty}$, $s = 1, ..., k$, then $\Delta_p(l)$ measures the contribution of the innovations l steps ahead, via replacing them by *i.i.d* copies and measuring the magnitude of changes in the outputs of the system. The measure $\Delta_p(l)$ for a broad class of classic time series can be calculated, *e.g.*, invertible ARMA process; (G)arch models (Engle 1982; Bollerslev 1986), threshold models (Tong 1990), etc. We refer to Zhou (2013) for more details about PLS models and their physical dependence measures.

For a pre-specified quantile $\alpha \in (0, 1)$, consider model (1). When $e_i(\alpha)$'s are *i.i.d* random variables with common CDF $F(\cdot)$, Koenker and Basset (1978) first developed a LAD estimator of $\beta(\alpha)$:

$$\hat{\beta}(\alpha) = argmin_{\beta} \sum_{i=1}^{n} \rho_\alpha(y_i - \mathbf{x}_i'\beta), \tag{4}$$

where $\rho_\alpha(x) = \alpha(x)^+ + (1 - \alpha)(-x)^+$ is the checking function, which has left derivative $\psi_\alpha(x) = \alpha - \mathbf{1}(x \le 0)$. The asymptotic behavior of the LAD estimator of β in model (1) is investigated by numerous researchers, among them, for one sample i.i.d. error model, Bahadur (1966) approximated $\sqrt{n}(\hat{\beta}(\alpha) - \beta(\alpha))$ via linear forms. The celebrated Bahadur representation (Bahadur (1966)) shows that the remaining term of the approximation is of order $O_p(n^{-1/4}(\log\log n)^{3/4})$. Babu (1989) obtained asymptotic results for strong mixing errors. Portnoy (1991) acquired asymptotic approximations of $\sqrt{n}(\hat{\beta}(\alpha)-\beta(\alpha))$ when the errors are "m-decomposable". Wu (2007) obtained Bahadur representation for models with fixed design and stationary errors. The first contribution of this paper is that, we obtain a Bahadur representation with nearly optimal rate (except a multiplicative logarithm factor) for model (1) with PLS errors and regressors, under certain mild conditions which can be checked easily; see Theorem 1 below.

In addition, we also allow dependence between the errors $\{e_i(\alpha)\}_{i=1}^n$ and the regressors $\{\mathbf{x}_i\}$. Specifically, we assume that the errors $\{e_i(\alpha)\}_{i=1}^n$ is $PLS(r,\mathcal{F}_i,\mathcal{G}_i)$ with break points $b_1,...,b_r$, while the covariates $\mathbf{x}_i$ is $PLS(s,\mathcal{G}_i)$ with break point $d_1,...,d_s$. The filtration $\mathcal{G}_i=(...,\eta_{i-1},\eta_i)$ and $\mathcal{F}_i=(...,\varepsilon_{i-1},\varepsilon_i)$, where $\{\eta_i\}_{i=-\infty}^{\infty}$ and $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ are independent. Define $w(i)=j$ if $b_j<i/n\le b_{j+1}$. To simplify our notation, define

$$e_i(t,\mathcal{F}_k,\mathcal{G}_k,\alpha)=G_{w(i),\alpha}(t,\mathcal{F}_k,\mathcal{G}_k),\ t\in(b_{w(i)},b_{w(i)+1}], \tag{5}$$

where $k\in\mathbb{Z}$, and $e_i(\alpha)=e_i(i/n,\mathcal{F}_i,\mathcal{G}_i,\alpha)=G_{w(i),\alpha}(i/n,\mathcal{F}_i,\mathcal{G}_i)$. Write $f_{w(i)}(t,x,\alpha|\mathcal{G}_k)=\frac{\partial}{\partial x}\mathbb{P}\{e_i(t,\mathcal{F}_k,\mathcal{G}_k,\alpha)\le x|\mathcal{G}_k\}$ for $t\in(b_{w(i)},b_{w(i)+1}]$.
The following regularity conditions are needed:

S0 The PLS error $e_i(\alpha)=G_{j,\alpha}(i/n,\mathcal{F}_i,\mathcal{G}_i)$, $t\in(b_j,b_{j+1}]$, $0=b_0<b_1...<b_r<b_{r+1}=1$, satisfies that for all $j\in[0,r]$ and all $t,s\in(b_j,b_{j+1}]$, $t\ne s$, we have for some constant C, and some constant $v>1$, $\|\frac{G_{j,\alpha}(t,\mathcal{F}_0,\mathcal{G}_0)-G_{j,\alpha}(s,\mathcal{F}_0,\mathcal{G}_0)}{|t-s|}\|_v\le C$. The L_1 dependence measure of $e_i(\alpha)$, defined as $\Delta_{1,\alpha}(k)$, satisfies $\Delta_{1,\alpha}(k)\le M_0\chi_0^k$ for some finite constant M_0 and $\chi_0\in[0,1)$.

S1 $Q_\alpha(e_i(\alpha)|\mathcal{G}_i)=0$ for all $i=1,2,...,n$, where α is a pre-specified quantile.

S2 Let the PLS covariates $\mathbf{x}_i=\mathbf{H}_k(i/n,\mathcal{G}_i):=(H_{k,1}(i/n,\mathcal{G}_i),...,H_{k,p}(i/n,\mathcal{G}_i))$ for $d_k<i/n\le d_{k+1}$, where $d_0=0<d_1<..<d_s<d_{s+1}=1$ are break points. The L_1 dependence measure of $\{\mathbf{x}_i\}_{i=1}^n$, $\max_{0\le k\le s}\sup_{t\in(d_k,d_{k+1}]}\|\mathbf{H}_k(t,\mathcal{G}_i)-\mathbf{H}_k(t,\mathcal{G}_i^*)\|_1$, is $O(\chi_g^{|i|})$ for some $\chi_g\in(0,1)$. In addition, there exists a finite constant C_0, strictly positive constant t_x, such that $\max_{0\le k\le s}\sup_{t\in(d_k,d_{k+1}]}\mathbb{E}(\exp(t_x|\mathbf{H}_k(t,\mathcal{G}_0)|))\le C_0$, and

for all $k \in [0,s]$ and all $t_1, t_2 \in [d_k, d_{k+1}]$, $t_1 \neq t_2$, we have for some constant C, $\|\frac{\mathbf{H}_k(t_1,\mathcal{F}_0,\mathcal{G}_0)-\mathbf{H}_k(t_2,\mathcal{F}_0,\mathcal{G}_0)}{|t_1-t_2|}\|_v \le C$ for constant $v > 1$ defined in [S0] .

Remark 1. *Conditions [S0] and [S2] make assumptions on the dependence measure and smoothness of the error process $e_i(\alpha)$ and the covariate process $\mathbf{x}_i$, respectively. The assumption that $v > 1$ in [S0] guarantees that $\iota > 1/4$ in Proposition 6. Assumption [S1] is necessary for the consistency of $\hat\beta(\alpha)$. Furthermore, Condition [S2] assumes that $\mathbf{x}_i$ has exponentially decaying tail. Write $\mathbf{x}_i(t) = H_k(t, \mathcal{G}_i)$ for $d_k < i/n \le d_{k+1}$, $d_k < t \le d_{k+1}$. An instant fact of [S2] is that, $\max_{0\le k\le s}\sup_{t\in(d_k,d_{k+1}]}\max_{nd_k<i\le nd_{k+1}, i\in\mathbb{Z}}\|\mathbf{x}_i(t)\|_l \le \tilde{C}l$ for some large constant $\tilde{C}$.*

Remark 2. *Note that a special case of [S2] is the following heteroscedastic error model:*

$$y_i = \mathbf{x}_i'\beta(\alpha) + s(\mathbf{x}_i)e_i(\alpha), \tag{6}$$

where $\{e_i(\alpha)\}_{i=1}^n$ is independent of $\{\mathbf{x}_i\}_{i=1}^n$. $s(\cdot)$ is a smooth function, and $\{e_i(\alpha)\}$ is a PLS process. Lack of fit tests in regression quantiles of the above heteroscedastic error model with $\{e_i(\alpha)\}_{i=1}^n$ i.i.d. is investigated in He and Zhu (2003).

To obtain the Bahadur representation, we need the following assumptions:

A0 i) Let $\underline{\lambda}_n^a$ be the minimal eigenvalue of $\mathbb{E}\{\sum_{i=1}^{\lfloor an\rfloor} f_{w(i)}(i/n, 0, \alpha|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'/a\}$ for any $a \in (0,1)$. In addition, suppose

$$\max_{0\le j\le r}\sup_{t\in(b_j,b_{j+1}]}\|f_j(t,0,\alpha|\mathcal{G}_i) - f_j(t,0,\alpha|\mathcal{G}_i^*)\|_1 = O(\chi^{|i|}), \tag{7}$$

for some constant $\chi \in (0,1)$. Assume that i) $\liminf_{n\to\infty}\underline{\lambda}_n^1/n > 0$. ii)$\forall s \in [\frac{1}{\log n}, 1]$, $\liminf_{n\to\infty}\underline{\lambda}_n^s/n > 0$. In addition, $f(t, 0|\mathcal{G}_i)$ is stochasticaly lipschitz continuous for $t \in (0,1)$.

Denote $\mathbb{E}^{(q)}\{\psi_\alpha(e_i(\alpha)+x)|\mathcal{F}_k\} := \frac{\partial^q}{\partial x^q}\mathbb{E}\{\psi_\alpha(e_i(\alpha)+x)|\mathcal{F}_k\}$. for $0 \le q \le p$, define

$$F_{w(i)}^{(q)}(t,x,\alpha|\mathcal{F}_{k-1},\mathcal{G}_k) = \alpha - \mathbb{E}^{(q)}\{\psi_\alpha(e_i(t,\mathcal{F}_k,\mathcal{G}_k,\alpha)-x)|\mathcal{F}_{k-1},\mathcal{G}_k\}, t \in (b_{w(i)}, b_{w(i)+1}],$$
$$F_{w(i)}^{(q)}(t,x,\alpha|\mathcal{F}_{k-1}^*,\mathcal{G}_k) = \alpha - \mathbb{E}^{(q)}\{\psi_\alpha(e_i(t,\mathcal{F}_k^*,\mathcal{G}_k,\alpha)-x)|\mathcal{F}_{k-1}^*,\mathcal{G}_k\}, t \in (b_{w(i)}, b_{w(i)+1}].$$

and

$$\tilde{\Delta}_{s,\alpha}(k,x,q) = \max_{0\le j\le r} \sup_{b_j<t\le b_{j+1}} \|F_j^{(q)}(t,x,\alpha|\mathcal{F}^*_{k-1},\mathcal{G}_k) - F_j^{(q)}(t,x,\alpha|\mathcal{F}_{k-1},\mathcal{G}_k)\|_s$$

Write $\tilde{\Delta}_\alpha(k,x,q)$ for $\tilde{\Delta}_{1,\alpha}(k,x,q)$.

A1 For $0\le j\le r$, $t\in(b_j,b_{j+1}]$, $x\in\mathbb{R}$, and $k\in\mathbb{Z}$, $0\le q\le \max\{3,p\}$, we have $F_j^{(q)}(t,x,\alpha|\mathcal{F}_{k-1},\mathcal{G}_k)$ is bounded by some finite constant M_0, and $\sup_{|x|\in\mathbb{R}}\tilde{\Delta}_\alpha(k,x,q)\le K_0\chi^k$ for some $\chi\in[0,1)$.

Remark 3. *Condition [A0] guarantees the consistency of* $\{\hat{\beta}_j(\alpha)\}_{j=N}^n$ *where* $\hat{\beta}_j(\alpha)$ *is the quantile regression coefficient using* $(\mathbf{x}_1,y_1),...(\mathbf{x}_j,y_j)$. *Recall that* $N=\lfloor\frac{n}{\log n}\rfloor$. *The requirement that* $\liminf_{n\to\infty}\underline{\lambda}_n^s/n>0$ *for* $s\in(\frac{1}{\log n},1)$ *is quite mild. Suppose that*

$$\min_{0\le k\le r}\inf_{t\in(b_k,b_{k+1}]} f_k(t,0,\alpha|\mathcal{G}_i)\ge\eta>0.$$

By Weyl inequality, if $\exists\,\epsilon>0$, *such that for* $0\le k\le s$, $t\in(d_k,d_{k+1}]$, $\lambda_1(\mathbb{E}\{H_k(t,\mathcal{G}_i)H_k'(t,\mathcal{G}_i\})\ge\epsilon$, *which forces that* $\lambda_1(\mathbb{E}\mathbf{x}_i\mathbf{x}_i')\ge\epsilon$, *then the requirement is fulfilled. In other words, we only require that, for* $t\in(0,1)$, *the* $s+1$ *time-dependent matrices* $\mathbb{E}\{H_k(t,\mathcal{G}_i)H_k'(t,\mathcal{G}_i)\}$, $0\le k\le s$, $t\in(d_k,d_{k+1}]$, *are not degenerate. The equation (7) is also mild. It makes assumptions on the* L_1 *dependence measure of the PLS process* $\{f_{w(i)}(t_i,0,\alpha|\mathcal{G}_i)\}$, *which can be checked in various cases; see Section 3.4. For condition [A1], by (5),*

$$\sup_{x\in\mathbb{R}}\max_{0\le i\le r}\sup_{b_i<t\le b_{i+1}}\|F_i^{(q)}(t,x,\alpha|\mathcal{F}^*_{k-1},\mathcal{G}_k)-F_i^{(q)}(t,x,\alpha|\mathcal{F}_{k-1},\mathcal{G}_k)\|_s$$

quantifies the dependence (in L_s *norm) of of predictive distribution, density and derivatives of densities of* $r+1$ *locally stationary processes* $\{G_{h,\alpha}(t,\mathcal{F}_i,\mathcal{G}_i)|\mathcal{F}_{i-1},\mathcal{G}_i\}$, $h=0,1,...r$, $b_h<t\le b_{h+1}$. *[A1] assumes that such dependence is geometrically decaying. The boundedness assumption of the (derivatives of) conditional density are also mild. It implies* $|\mathbb{E}(\psi_\alpha(e_i(\alpha)-x)-\psi_\alpha(e_i(\alpha)-y)|\mathcal{F}_{i-1},\mathcal{G}_i)|\le M_0|x-y|$ *for* $i=1,..,n$ *and some constant* M_0.

Lemma 1. *Suppose [S0]-[S2], [A0] i), [A1] hold, then i)* $|\hat{\beta}_n(\alpha)-\beta(\alpha)|\le_p n^{-1/2}\log n$. *In addition, if [A0] ii) holds, then ii)* $\max_{N\le j\le n}|\hat{\beta}_j(\alpha)-\beta(\alpha)|\le_p(n^{-1/2}\log^4 n)$.

i) shows that $\hat{\beta}_n(\alpha)$ is weakly consistent. Result ii) establishes the uniform consistency of $\hat{\beta}_j(\alpha)$ estimated in different sub-samples with at least N observations. The consistency results are needed for the lack of fit test in Section 3. We have the following Bahadur representation:

Theorem 1. *Write* $\acute{\Lambda}(j,\alpha) = \sum_{i=1}^{j} f_{w(i)}(i/n, 0, \alpha|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'/n$. *Under Assumptions [S0]-[S2], [A0], [A1], considering model (1), we have i)*

$$\sqrt{n}(\hat{\beta}_n(\alpha) - \beta(\alpha)) - (\acute{\Lambda}(n,\alpha))^{-1}\sum_{i=1}^{n}\psi_\alpha(e_i(\alpha))\mathbf{x}_i/\sqrt{n} = O_p(n^{-1/4}\log^{3/2} n). \tag{8}$$

ii)

$$\max_{N\le j\le n}|\sqrt{n}(\hat{\beta}_j(\alpha) - \beta(\alpha)) - (\acute{\Lambda}(j,\alpha))^{-1}\sum_{i=1}^{j}\psi_\alpha(e_i(\alpha))\mathbf{x}_i/\sqrt{n}| = O_p(n^{-1/4}\log^{3} n). \tag{9}$$

i) establishes the Bahadur representation of $\hat{\beta}_n(\alpha)$ in non-stationary time series quantile regression, and ii) establishes the uniform Bahadur representation of $\{\hat{\beta}_j(\alpha) - \beta(\alpha)\}_{N\le j\le n}$. Both results almost achieve the optimal order $n^{-1/4}(\log\log n)^{3/4}$ except a factor of multiplicative logarithms. Observe that, due to the non-stationarity, the approximating processes depend on $f_{w(i)}(i/n, 0, \alpha|\mathcal{G}_i)$, $N \le i \le n$, which is the conditional densities of the errors $e_i(\alpha)$ from N to n at their αth quantile conditioning on $\mathcal{G}_i$. Portnoy (1991) also provides a similar form of Bahadur representation with non-stationary errors.

3 Structural stability tests.

3.1 Test statistics.

Consider the alternative nonparametric quantile regression model

$$y_i = \mathbf{x}_i'\beta_i(\alpha) + e_i(\alpha), \quad i = 1, 2, \cdots, n.$$

We consider testing whether $\beta_i(\alpha)$ remains constant over time. That is, we test

$$H_0 : \beta_1(\alpha) = \beta_2(\alpha) = ..., \beta_n(\alpha) = \beta(\alpha) \leftrightarrow H_A : \beta_i(\alpha) \neq \beta_j(\alpha) \text{ for some } 1 \leq i < j \leq n.$$

for some unknown $\beta(\alpha)$. Consider the following two test statistics:

$$T_{n1} = \max_{N \leq j \leq n} \sqrt{n} |\hat{\beta}_j(\alpha) - \hat{\beta}_n(\alpha)|, \quad T_{n2} = \max_{N \leq j \leq n} |\frac{\sum_{i=1}^{j} \psi(\hat{e}_{in}(\alpha)) \mathbf{x}_i}{\sqrt{n}}|, \tag{10}$$

where $\hat{e}_{in}(\alpha) = y_i - \mathbf{x}_i' \hat{\beta}_n(\alpha)$ are the residuals. Recall that $\hat{\beta}_j(\alpha)$ is the quantile regression coefficients using $(\mathbf{x}_1, y_1), ...(\mathbf{x}_j, y_j)$. The test statistic T_{n1} is the cusum test based on partial sample quantile regression coefficients and T_{n2} is the cusum statistic of the estimated gradient vectors of the regression. If H_0 is violated, then both T_{n1} and T_{n2} will be large. Due to the unknown non-stationary structure of covarites and errors, it is impossible that the test statistics of (10) or its normalized version is asymptotically pivotal. Hence the inference based on (10) under non-stationarity differs drastically from that under stationarity. We shall further investigate this in section 3.2.

Under conditions [S0]-[S2], $\psi_\alpha(e_i(\alpha))\mathbf{x}_i$ can be viewed as a realization from a PLS process with r_1 break points $c_1, ..., c_{r_1}$, namely, $\tilde{G}_{v(i),\alpha}(t, \mathcal{F}_i, \mathcal{G}_i)$, where $v(i) = k$ for $c_k < i/n \leq c_{k+1}$. We set $c_0 = 0$ and $c_{r_1+1} = 1$. Then $\psi_\alpha(e_i(\alpha))\mathbf{x}_i = \tilde{G}_{v(i),\alpha}(i/n, \mathcal{F}_i, \mathcal{G}_i)$. Define the long-run variance:

$$\Sigma_\alpha^2(t) = \sum_{k=-\infty}^{\infty} Cov(\tilde{G}_{i,\alpha}(t, \mathcal{F}_0, \mathcal{G}_0), \tilde{G}_{i,\alpha}(t, \mathcal{F}_k, \mathcal{G}_k)), \ t \in (c_i, c_{i+1}]. \tag{11}$$

Let $\Sigma_\alpha^2(0) = \lim_{t \downarrow 0} \Sigma_\alpha^2(t)$. In order to investigate the limiting behaviors of T_{n1} and T_{n2}, We shall further introduce the following assumptions:

A2 The smallest eigenvalue of $\Sigma_\alpha^2(t)$ is bounded away from 0 on $[0, 1]$.

It is shown in Proposition 6 that the dependence of $\{\psi_\alpha(e_i)\mathbf{x}_i\}_{i=1}^n$ decays exponentially fast to 0. Meanwhile, condition [A2] assures that the long run variance of $\psi_\alpha(e_i)\mathbf{x}_i$ is not degenerate over time. As a result, we have the following proposition, which is useful in the study of the process $\{\psi_\alpha(e_i)\mathbf{x}_i\}_{i=1}^n$:

Proposition 1. *Assume [S0] [S1],[S2]. Then on a possibly richer probability space, there exists a p-dimensional zero-mean Gaussian process* $U_\alpha(t)$*, with covariance function* $\gamma(t,s) = \int_0^{\min(t,s)} \Sigma^2_\alpha(r)dr$*,such that*

$$\max_{1\le j\le n} |\frac{1}{\sqrt{n}}\sum_{i=1}^{j}\psi_\alpha(e_i)\mathbf{x}_i - \sum_{i=1}^{j} U_\alpha(i/n)| = o_p(n^{-1/4}\log^2 n). \tag{12}$$

Write $\Lambda(t,\alpha) = \lim_{n\to\infty}\mathbb{E}\{\sum_{i=1}^{\lfloor nt\rfloor} f_{w(i)}(i/n,0,\alpha|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'\}/n$. Without loss of generality, suppose that the covariate and the error have the same break points, i.e, $\{b_1,...,b_r\} = \{d_1,...,d_s\}$. Then by the stochastically lipschitz countinuity of $f_j(t,0|\mathcal{G}_0)$, $0\le j\le r$,

$$\Lambda(t,\alpha) = \sum_{l=0}^{j-1}\int_{b_l}^{b_{l+1}}\mathbb{E}\{f_l(t,0,\alpha|\mathcal{G}_0)\mathbf{H}_l(t,\mathcal{G}_0)'\mathbf{H}_l(t,\mathcal{G}_0)\}dt + \int_{b_j}^{t}\mathbb{E}\{f_j(t,0,\alpha|\mathcal{G}_0)\mathbf{H}_j(t,\mathcal{G}_0)'\mathbf{H}_j(t,\mathcal{G}_0)\}dt,\ t\in(b_j,b_{j+1}]. \tag{13}$$

In the Proposition 1 of the supplementary material of the paper, we show that

$$\max_{0\le s\le 1}|\frac{1}{n}\sum_{i=1}^{\lfloor sn\rfloor}(f_{w(i)}(i/n,0|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i' - \mathbb{E}(f_{w(i)}(i/n,0|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'))| = O_p(\frac{1}{\sqrt{n}}\log^{\frac{7}{2}} n), \tag{14}$$

which implies that $\acute{\Lambda}(\lfloor nt\rfloor,\alpha)\to\Lambda(t,\alpha)$ uniformly in $t\in(0,1)$. The following theorem establishes the limiting behaviors of T_{n1} and T_{n2} for non-stationary time series quantile regression:

Theorem 2. *Suppose assumptions [S0]-[S2], [A0]-[A2] hold, then under the null hypothesis, we have the joint weak convergence*

$$T_{n1}\Rightarrow_{n\to\infty}\sup_{t\in(0,1]}|G_1(t)| := \sup_{t\in(0,1]}|\Lambda^{-1}(t,\alpha)U_\alpha(t) - \Lambda^{-1}(1,\alpha)U_\alpha(1)|, \tag{15}$$

$$T_{n2}\Rightarrow_{n\to\infty}\sup_{t\in(0,1]}|G_2(t)| := \sup_{t\in(0,1]}|U_\alpha(t) - \Lambda(t,\alpha)\Lambda^{-1}(1,\alpha)U_\alpha(1)|. \tag{16}$$

where $U_\alpha(t)$ *is defined in Proposition 1 and "* $\Rightarrow$*" denotes convergence in distribution.*

Theorem 2 establishes that both tests converge to the maximum of certain centered Gaussian processes. Two important observations should be made. 1). The Gaussian process $U_\alpha(t)$ is not pivotal and it has a complex covariance structure $\gamma(t,s) = \int_0^{\min(t,s)} \Sigma^2_\alpha(r)dr$. In particular, $\Sigma^2_\alpha(s)$ can change both smoothly and abruptly on [0,1] and hence it is inappropriate to perform T_{n1} and T_{n2} by checking quantile tables of certain pivotal Gaussian processes (such as the Brownian bridge). 2). Due to the non-stationarity, $\Lambda(t,\alpha)\Lambda^{-1}(1,\alpha)$ no longer equals tI_p as in the stationary case, where I_p is the $p \times p$ identity matrix. In particular, the gradient cusum test T_{n2} is no longer asymptotically free of the density functions of $\{e_i(\alpha)\}$ and the ratio $\Lambda(t,\alpha)\Lambda^{-1}(1,\alpha)$ should be estimated when performing the gradient cusum test for non-stationary time series quantile regression. Consequently, the indepdent wild bootstrap procedure in He and Zhu (2003) will in general yield inconsistent testing results under non-stationarity.

The following theorem studies the asymptotic power behavior of the tests for non-stationary time series quantile regression. For any bounded lipschitz continuous vector function $g(\cdot)$, write

$$\begin{aligned}\Lambda(t,\alpha,g(\cdot)) &= \lim_{n\to\infty} \mathbb{E}\{\sum_{i=1}^{\lfloor nt \rfloor} f_{w(i)}(i/n,0,\alpha|\mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i' g(i/n)/n\} \\ &= \sum_{l=0}^{j-1} \int_{b_l}^{b_{l+1}} \mathbb{E}\{f_l(t,0,\alpha|\mathcal{G}_0)\mathbf{H}_l(t,\mathcal{G}_0)'\mathbf{H}_l(t,\mathcal{G}_0)\}g(t)dt + \\ &\quad \int_{b_j}^{t} \mathbb{E}\{f_j(t,0,\alpha|\mathcal{G}_0)\mathbf{H}_j(t,\mathcal{G}_0)'\mathbf{H}_j(t,\mathcal{G}_0)\}g(t)dt,\ t \in (b_j, b_{j+1}]. \qquad (17)\end{aligned}$$

Define $H_1(t,g(\cdot)) = \Lambda(t,\alpha)^{-1}\Lambda(t,\alpha,g(\cdot)) - \Lambda(1,\alpha)^{-1}\Lambda(1,\alpha,g(\cdot))$, $H_2(t,g(\cdot)) = \Lambda(t,\alpha,g(\cdot)) - \Lambda(t,\alpha)\Lambda(1,\alpha)^{-1}\Lambda(1,\alpha,g(\cdot))$.

Theorem 3. *Consider the alternative hypothesis* $H_A : \beta_i(\alpha) = \beta(\alpha) + L_n g(i/n)$. *Suppose* $\exists \varepsilon > 0$, *s.t.* $\inf_{|x|\le\varepsilon} \min_{0\le j\le r} \inf_{t\in(b_j,b_{j+1}]} f_j(t,x,\alpha|\mathcal{G}_0) \ge \varrho > 0$ *for some positive constant* ϱ. *Assume [S0]-[S2], [A0]-[A2] hold, and* $g(\cdot)$ *is a bounded non-constant lipschitz continuous vector function defined in* $(0,1)$. *Then we have,* *i) If $L_n = n^{-1/2}$,*

$$T_{n1} \Rightarrow \sup_{0<t\le 1} |G_1(t) + H_1(t, g(\cdot))|, \; T_{n2} \Rightarrow \sup_{0<t\le 1} |G_2(t) + H_2(t, g(\cdot))|, \tag{18}$$

where $G_1(t)$, $G_2(t)$ is defined in Theorem 3.2.
ii) If the deterministic sequence L_n satisfies $L_n \log^{4p+7} n = o(1)$, $\sqrt{n}L_n \to \infty$, and $H_1(t, g(\cdot))$, $H_2(t, g(\cdot))$ are not constant over $t \in (0, 1)$, then $T_{n1} \to_p \infty$, $T_{n2} \to_p \infty$ at the rate $\sqrt{n}L_n$.

The theorem shows that the powers of the two tests converge to 1 if $L_n\sqrt{n} \to \infty$ and $L_n \log^{4p+7} n = o(1)$, which implies that both tests can detect local alternatives with the same rate $n^{-1/2}$ as in the classic stationary case.

3.2 The bootstrap.

As we observe from Theorem 2, the key to accurate tests under non-stationarity is to consistently mimic the behaviors of the processes $\{\Lambda(t, \alpha)\}$ and $\{U_\alpha(t)\}$. A direct but naive approach is to estimate the conditional densities $f_{w(i)}(t, \cdot, \alpha|\mathcal{G}_i)$ and long-run covariances $\Sigma^2_\alpha(t)$ over time t and use those estimates to generate the limiting distributions in Theorem 2. However, this approach is not operational in practice for the following two reasons. First, the estimation of the density and the long-run covariance at a fixed time t require a total of four bandwidth parameters. The large amount of tuning parameters are difficult to choose in practice and can cause inaccurate testing results for moderate samples. Second, the nonparametric estimates of $f_{w(i)}(t, \cdot, \alpha|\mathcal{G}_i)$ and $\Sigma^2_\alpha(t)$ are inconsistent near the break points of the PLS errors and covariates. Hence it is unclear whether those plug-in procedures asymptotically achieve the nominal size. In this section we shall propose a bootstrap procedure which avoids directly estimating the densities and long-run covariances and requires only two tuning parameters. The proposed bootstrap procedure combines the advantages of moving block bootstrap (Lahiri 2003) and subsampling (Politis et al. 1999) by progressively convoluting block sums of partial sums of the estimated gradient vectors and auxiliary standard normals in order to preserve the temporal dependence structure and to mimic the pattern of the non-stationarity over time. Furthermore, in our bootstrap, we make use of an extension the "Powell Sandwich" (Powell 1991) to optimally estimate $\{\Lambda(t, \alpha)\}$. In the following we shall discuss the estimation of $\{\Lambda(t, \alpha)\}$

and $\{U_\alpha(t)\}$ separately.

Recall the definition of $\hat{e}_{in}(\alpha)$ in Theorem 2. Let $\phi(\cdot)$ be the density of standard normal. Define $\hat{\Lambda}_{c_n}(t,\alpha) = \hat{\lambda}_{c_n}(\lfloor nt \rfloor, \alpha)$, where

$$\hat{\lambda}_{c_n}(j,\alpha) = \sum_{i=1}^{j} \frac{\phi(\hat{e}_{in}(\alpha)/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}, \tag{19}$$

The following theorem states that, we can use $\{\hat{\Lambda}_{c_n}(t,\alpha)\}_{t\in(0,1)}$ to approximate $\{\Lambda(t,\alpha)\}_{t\in(0,1)}$.

Theorem 4. *Under condition of [S0]-[S2], [A0]-[A1], $c_n \to 0$, $nc_n^3 \to \infty$, then*

$$\sup_{t\in(0,1)} |\hat{\Lambda}_{c_n}(t,\alpha) - \Lambda(t,\alpha)| = O_p(n^{-1/2}\log^7 n + \frac{\log^{10} n}{nc_n^3} + \frac{\log^4 n}{\sqrt{nc_n}} + c_n^2 \log^4 n).$$

Observe that $\hat{\lambda}_{c_n}(j,\alpha)$ is an extension of the Powell's Sandwich and it can be viewed as a progressive local constant kernel estimation of integrated conditional density. Theorem 4 shows that $\{\hat{\Lambda}_{c_n}(t,\alpha)\}_{t\in(0,1)}$ are uniformly consistent estimators of $\{\Lambda(t,\alpha)\}_{t\in(0,1)}$. Elementary calculations show that, even with PLS errors, the optimal bandwidth c_n for Theorem 4 is almost in the order of $n^{-1/5}$. Therefore the convergent rate of Theorem 4 is still almost at the order of $n^{-2/5}$ except a factor of multiplicative logarithms, where the order $n^{-2/5}$ is the well known optimal approximate rate of the Powell's sandwich estimates for i.i.d. data. Note that the nearly $n^{-2/5}$ rate above is faster than $n^{-1/4}\log^3 n$, the nearly optimal approximation rate of the Bahadur representation in (8).

The remaining task for evaluating the critical values is to simulate the data-driven non-stationary Gaussian Process $U_\alpha(t)$. The covariance structure of $U_\alpha(t)$ could be quite complex, in particular, it does not necessarily have stationary increments. We propose the following gradient-based process $\tilde{\Psi}_{m,n}(t)$ to bootstrap $U_\alpha(t)$:

$$\tilde{\Psi}_{m,n}(t) = \Psi_{t_* n,m} + n(t - t_*)(\Psi_{t^* n,m} - \Psi_{t_* n,m}), \tag{20}$$

$$\Psi_{i,m} = \sum_{j=1}^{i} \frac{1}{\sqrt{m(n-m+1)}}(\hat{\varpi}_{j,m} - \frac{m}{n}\hat{\varpi}_n)R_j, i = 1, ..., n-m+1, \tag{21}$$

where $\hat{\varpi}_{j,m} = \sum_{r=j}^{j+m-1} \psi_\alpha(\hat{e}_{rn}((\alpha)))\mathbf{x}_r$, $\hat{\varpi}_n = \hat{\varpi}_{1,n}$, and $(R_i)_{i=1}^n$ are i.i.d standard normals which are independent of $\{\mathcal{F}_i\}_{i=-\infty}^{\infty}, \{\mathcal{G}_i\}_{i=-\infty}^{\infty}$. The consistency of $\{\tilde{\Psi}_{m,n}(t)\}$ as an estimate of $\{U_\alpha(t)\}$ is provided by the following theorem:

Theorem 5. *Suppose [S0]-[S2], [A0]-[A2]. The bandwidth* $m = m(n)$ *satisfies* $m(n) \to \infty$, $m(n) \log^7 n/\sqrt{n} \to 0$. *we have,*
conditioning on $\mathcal{F}_n$, $\mathcal{G}_n$, $\tilde{\Psi}_{m,n}(t) \Rightarrow U_\alpha(t)$ *on* $\mathcal{C}(0,1)$ *with the uniform topology.*

By the proof of Theorem 5, conditioning on $\mathcal{F}_n, \mathcal{G}_n$, the covariance function of $\tilde{\Psi}_{m,n}(t)$ converges uniformly to that of $U_\alpha(t)$ at the rate $n^{-1/4}\log^{3/2} n$, which is also faster than $n^{-1/4}\log^3 n$, the nearly optimal approximation rate of the Bahadur representation in (8). Therefore Theorem 4 and Theorem 5 suggest that the type I error rate of our bootstrap methodology approaches the nominal level with an asymptotically nearly optimal rate.

We have the following proposition on the power performances of $\{\hat{\Lambda}_{c_n}(t,\alpha)\}_{t\in(0,1)}$ and $\tilde{\Psi}_{m,n}(t)$ under the local alternative hypotheses:

Proposition 2. *Suppose the conditions of Theorem 3 hold. Assume that* $m(n)$ *is the order of* $n^{1/3}$ *except a factor of multiplicative logarithm,* c_n *is the order of* $n^{-1/5}$ *except a fact of multiplicative logarithm, then we have, under* H_A *which is defined in Theorem 3, i) if* $L_n = n^{-1/2}$,

$$\sup_{t\in(0,1)} |\hat{\Lambda}_{c_n}(t,\alpha) - \Lambda(t,\alpha)| \to 0 \text{ in probability,} \tag{22}$$

$$\tilde{\Psi}_{m,n}(t) \Rightarrow U_\alpha(t) \quad \text{on } \mathcal{C}(0,1) \text{ conditioned on } \mathcal{G}_n,\ \mathcal{F}_n \text{ with the uniform topology.} \tag{23}$$

ii) If $L_n n^{1/2} \to \infty$ *but* $L_n \log^{4p+7} n = o(1)$, *then (22) still holds, while conditioning on* $\mathcal{G}_n$, $\mathcal{F}_n$, $\sup_{t\in(0,1)} \frac{1}{\sqrt{m}L_n \log^{9/2} n}|\tilde{\Psi}_{m,n}(t) - U_\alpha(t)| = O_p(1)$. *In addition, if* $m_n L_n = o(1)$, *then conditioning on* $\mathcal{G}_n$, $\mathcal{F}_n$, $\tilde{\Psi}_{m,n}(t) \Rightarrow U_\alpha(t)$ *still holds.*

Recall that in Theorem 3, we show that under H_A, if $L_n \log^{4p+7} n \to 0$ and $\sqrt{n}L_n \to \infty$, both test statistics go to infinity at the rate $\sqrt{n}L_n$, which is faster than $\sqrt{m}L_n \log^{9/2} n$, the fastest possible rate at which $\tilde{\Psi}_{m,n}(t)$ can go to infinity. Hence together with Theorem 3, Proposition 2 shows that our bootstrap method has asymptotic power 1 under the considered local alternatives in ii). In particular, our bootstrap can detect local alternatives with the optimal $1/\sqrt{n}$ parametric rate. Combining Theorem 4 and Theorem 5, we have the following step-by-step implementation procedures for performing structural change tests for non-stationary time series quantile regression:

Theorem 6. *Under conditions [S0]-[S2], [A0]-[A2], the following procedure generates consistent estimator of the level α critical values for (15) and (16):*

i *By section 3.4, select appropriate m, c_n.*

ii *Apply Theorem 4 to get $\hat{\lambda}_{c_n}(j, \alpha)$, $j = 1, ...n$. Use Theorem 5 to generate B (say 2000) conditional iid copies $\{\Psi_{i,m}^{(r)}\}_{i=1}^{n-m+1}$, $r = 1, ..., B$.*

iii *Calculate $E_i^{(r)} = \hat{\lambda}_{c_n}^{-1}(i, \alpha)\Psi_{i,m}^{(r)} - \hat{\lambda}_{c_n}^{-1}(n - m + 1, \alpha)\Psi_{n-m+1,m}^{(r)}$, and $F_i^{(r)} = \Psi_{i,m}^{(r)} - \hat{\lambda}_{c_n}(i, \alpha)\hat{\lambda}_{c_n}^{-1}(n - m + 1, \alpha)\Psi_{n-m+1,m}^{(r)}$ for $r = 1, ..., B$, $i = N, ..., n - m + 1$.*

iv *Let $E_r = \sup_{N \le i \le n-m+1} |E_i^{(r)}|$, and $F_r = \sup_{N \le i \le n-m+1} |F_i^{(r)}|$. Let $E_{(1)} \le E_{(2)} ... \le E_{(B)}$ and $F_{(1)} \le F_{(2)} ... \le F_{(B)}$ be the order statistics of E_r, F_r, respectively. Then $E_{\lfloor (1-\alpha)B \rfloor}$ and $F_{\lfloor (1-\alpha)B \rfloor}$ are the level α critical values for coefficient based method (15) and gradient vector based method (16), respectively.*

3.3 Examples

The purpose of this section is to provide examples of two general classes of non-stationary linear and non-stationary nonlinear time series models and to verify that our regularity conditions hold for those two classes. Suppose the mild conditions [S1]-[S2], [A0] except (7) and [A2] hold. For general PLS processes $e_i(\alpha)$, $\mathbf{x}_i$, in order to apply our lack of fit tests, we need to check condition [S0], (7) of [A0] and [A1]. The following two propositions show that the three conditions hold for the two general classes of non-stationary time series models, which implies the wide applications of our methodology.

3.3.1 PLS linear processes

Suppose we have the following model: let $0 = b_0 < b_1 < ... < b_r < b_{r+1} = 1$,

$$G_k(t, \mathcal{F}_i, \mathcal{G}_i) = \sum_{j=0}^{\infty} a_{k,j}(t)\varepsilon_{i-j} f_k(t, \mathcal{G}_{i-j}) \quad b_k < t \le b_{k+1}, \tag{24}$$

where $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are $i.i.d$ r.v's with $\mathbb{E}|\varepsilon_1|^{vl} < \infty$ for some $v > 1$, $l > 1$, $\sup_{x\in\mathbb{R}} |f_{\varepsilon_1}^{(q)}(x)| \le C < \infty$ for some constant C and $0 \le q \le \max\{3, p\}$, $\sum_{j=0}^{\infty} \max_{0\le k\le r} \sup_{t\in(b_k, b_{k+1}]} |\frac{\partial}{\partial t} a_{k,j}(t)| < \infty$. In addition, $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are independent of $\{\mathcal{G}_i\}_{i\in\mathbb{Z}}$.

Proposition 3. *Suppose there exists $\eta > 0$, a large enough constant M, such that*

$$\min_{0 \le k \le r} \inf_{t \in (b_k, b_{k+1}]} |a_{k,0}(t)| \ge \eta > 0, \tag{25}$$

$$\max_{0 \le k \le r} \sup_{t \in (b_k, b_{k+1}]} |f_k(t, \mathcal{G}_0)^{-1}| \le M. \tag{26}$$

and a constant χ which lies between $(0,1)$,

$$\max_{0 \le k \le r} \sup_{t \in (b_k, b_{k+1}]} |a_{k,i}(t)| \le M\chi^{|i|}, \tag{27}$$

Furthermore, suppose $f_k(t, \mathcal{G}_i)$ is PLS processes with its dependence measure decays geometrically in $L_{\frac{v}{v-1}}$ norm and for a large constant M,

$$\max_{0 \le k \le r} \sup_{t \in (b_k, b_{k+1}]} \|f_k(t, \mathcal{G}_0)\|_{\max\{v, \frac{vl}{v-1}\}} \le M, \ \max_{0 \le k \le r} \sup_{t \in (b_k, b_{k+1}]} \|\frac{\partial}{\partial t} f_k(t, \mathcal{G}_0)\|_{\frac{vl}{v-1}} \le M. \tag{28}$$

Then for model (1) with [S1]-[S2] holding, we have [S0], (7) of [A0] ,[A1] hold if we have (24) as error process.

Observe that if $f_k(t, \mathcal{G}_j)$ is independent of $\mathcal{G}_j$, then (24) behaves as the piecewise time-varying MA(∞) model.

3.3.2 Piecewise time-varying autoregressive process

For $k = 0, 1, .., r$, let

$$y_{k,i} = a_{k,1}(i/n) y_{k,i-1} +, ..., + a_{k,l}(i/n) y_{k,i-l} + e_{k,i}, \tag{29}$$

where $e_{k,i} = A_k(i/n, \mathcal{F}_i)$ is a locally stationary process satisfying [S0]. Let $\{z_i\}_{i=1}^n$, satisfying

$$z_i = y_{k,i}, \ b_k < i/n \le b_{k+1}, \ k = 0, ..., r. \tag{30}$$

Then z_i is a piecewise time-varying autoregressive process (piecewise tvAR(l) process). By similar argument of Proposition 4.2 of Zhang and Wu (2012), we can show that if i)

$(y_1, y_2, ..., y_p) \in \mathcal{L}^v$, ii) for each $k \in [0, r]$, $j = 1, ..., p$, $a_{k,j}(\cdot)$ is Lipschitz continuous, iii) for $k \in [0, r]$, $\sum_{j=1}^{p} a_{k,j}(t)z^j \neq 1$ for all $|z| \leq 1 + c$ with $c > 0$ uniformly in $t \in (b_k, b_{k+1}]$. Write $w(i) = j$ for $b_j < i/n \leq b_{j+1}$, then there exists a PLS process $G_k(t, \mathcal{F}_i)$, such that

$$\max_{1 \leq i \leq n} \|z_i - G_{w(i)}(i/n, \mathcal{F}_i)\|_v \leq Cn^{-1}, \tag{31}$$

where the corresponding approximating PLS process $G_k(t, \mathcal{F}_i)$:

$$G_k(t, \mathcal{F}_i) = a_{k,1}(t)G_k(t, \mathcal{F}_{i-1}) +, ..., +a_{k,l}(t)G_k(t, \mathcal{F}_{i-l}) + A_k(t, \mathcal{F}_i)$$

for $t \in (b_k, b_{k+1}]$. By similar argument of proof of Lemma 1 in the appendix, $\| \max_{1 \leq i \leq n} |z_i - G_{w(i)}(i/n, \mathcal{F}_i)| \|_v = O(n^{\frac{1}{v}-1})$. Suppose $v > 3$. Then by the similar argument of proof of Proposition 6 in the supplemental material, we have that

$$\max_{1 \leq j \leq n} |\frac{1}{\sqrt{n}} \sum_{i=1}^{j} [\psi(z_i) - \psi(G_{w(i)}(i/n, \mathcal{F}_i))] \mathbf{x}_i \mathbf{x}_i'| = o_p(1). \tag{32}$$

Then by careful check of the proof of asymptotic results, we have that, if the PLS process $G_k(t, \mathcal{F}_{i-1})$ satisfy [S0]-[S2], [A0]-[A2], then the proposed methodology of testing for lack of fit is still valid if we model our error as a piecewise time-varying autoregressive process, i.e, the critical value generated by Theorem 6 is still consistent for T_{n1}, T_{n2} of Theorem 2 under both H_0 and H_A in Theorem 3.

3.3.3 Non-Linear PLS

Suppose our error is generated from the following system:

$$G_k(t, \mathcal{F}_i, \mathcal{G}_i) = R_k(t, G_k(t, \mathcal{F}_{i-1}, \mathcal{G}_{i-1}), \varepsilon_i, \eta_i) \tag{33}$$

for $b_k < t \leq b_{k+1}$ where b_k, $k = 1, 2, \cdots, r$ are break points. Let

$$\chi = \max_{0 \leq k \leq r} \sup_{x \neq y, t \in (b_k, b_{k+1}]} \frac{\|R_k(t, x, \varepsilon_0, \eta_0) - R_k(t, y, \varepsilon_0, \eta_0)\|_v}{|x - y|}$$

for some constant $v > 1$. Write $F_k(t, x, s, u) = \mathbb{P}(R_k(t, s, \varepsilon_i, u) \leq x)$. Then

Proposition 4. *Suppose we have model (1) with [S2] standing,* $G_k(t, \mathcal{F}_i, \mathcal{G}_i)$ *satisfies [S1]* $\forall k$*. Assume i)* $0 < \chi < 1$ *ii) Define* $C = \max_{0\le k\le r} \sup_{t\in(b_k,b_{k+1}]} \|M(G_k(t, \mathcal{F}_0, \mathcal{G}_0))\|_v < \infty$*, where*

$$M(x) = \max_{0\le k\le r} \sup_{t,s\in(b_k,b_{k+1}]} \frac{\|R_k(t, x, \varepsilon_0, \eta_0) - R_k(s, x, \varepsilon_0, \eta_0)\|_v}{|t - s|}.$$

iii) For some large constant M*,* $0 \le q \le \max\{3, p\}$*,* $w = \frac{v}{v-1}$

$$\max_{0\le k\le r} \sup_{t\in(b_k,b_{k+1}],x,s\in\mathbb{R}} \|\frac{\partial^q}{\partial x^q}\frac{\partial}{\partial s} F_k(t, x, s, \eta_0)\|_w \le M. \tag{34}$$

Then (33) admits a unique stationary solution for each integer $k \in [0, r]$*, and the associate* $t \in (b_k, b_{k+1}]$*. If the solution is the error process of model (1), then [S0], (7) of [A0], [A1], hold.*

Observe that if (33) is independent of the filtration $\{\mathcal{G}_i\}_{i=-\infty}^{\infty}$, then (33) is a more familiar PLS nonlinear process which includes PLS (G)ARCH models (Engle 1982; Bollerslev 1986), PLS threshold models (Tong 1990), PLS bilinear models as special cases. Proposition 4 admits the information $\mathcal{G}_i$ in the error processes to make our error structure dependent on the regressors.

3.4 Bandwidth Selection.

Due to our complex data structure, a robust bandwidth selection method which does not depend on specific forms of the data generating mechanisms is desired. To this end, for selecting proper m of Theorem 5, we apply the method of minimum volatility (MV) suggested by Zhou (2013) to $\tilde{\Psi}_{m,n}(t)$ in (20). The procedures are quite similar except that we replace unknown $\psi(e_i(\alpha))\mathbf{x}_i$ with estimated $\psi(\hat{e}_{in}(\alpha))\mathbf{x}_i$. Thus we omit the detailed description of selecting m here. For more discussions about the "MV" method, see Politis, Romano, and Wolf (1999). We also apply the MV method to selecting the bandwidth c_n. Our procedure of selecting c_n is as follows:

i Choose suitable end point $a_1 < a_2$, such that the optimal $c_n \in I = [a_1, a_2]$.

ii Divide interval I into $\bar{m}$, say 99 pieces. Specifically, Let $h_1 = a_1$, $h_{100} = a_2$, and $h_k = a_1 + (k - 1)(a_2 - a_1)/99$.

iii For each h_i, use it as bandwidth to calculate the estimating quantity $\{\hat{\Lambda}_{h_i}(t_j,\alpha)\}_{j=N}^{n}$. Let $C_1(i)$, $C_2(i)$ be the maximal values of RHS porcess of equation (15), (16) in $t \in [\frac{1}{\log n}, 1]$ obtained by replacing $\{U_\alpha(t), t \in [\frac{1}{\log n}, 1]\}$ with $\{\sum_{i=1}^{\lfloor nt \rfloor} \frac{\psi(\hat{e}_{in}(\alpha))}{\sqrt{n}}, t \in [\frac{1}{\log n}, 1]\}$ and replacing $\{\Lambda(t,\alpha), t \in [\frac{1}{\log n}, 1]\}$ with $\{\hat{\Lambda}_{h_i}(t,\alpha), t \in [\frac{1}{\log n}, 1]\}$, respectively.

iv For some $k > 0$, Define $D_1(i) = \frac{1}{2k}\{\sum_{j=i-k}^{i+k}[C_1(j) - \frac{1}{2k+1}\sum_{j=i-k}^{i+k} C_1(j)]^2\}^{1/2}$, $D_2(i) = \frac{1}{2k}\{\sum_{j=i-k}^{i+k}[C_2(j) - \frac{1}{2k+1}\sum_{j=i-k}^{i+k} C_2(j)]^2\}^{1/2}$. Let i_1, i_2 be the minimizer of $D_1(i)$, $D_2(i)$, respectively. Then for (15), we select h_{i_1} for c_n, and for (16), we select h_{i_2} for c_n, respectively.

4 Extension to finite many conditional quantiles.

In this section we extend our bootstrap to testing whether there are structural changes in at least one of the s conditional quantiles: τ_1, τ_2,...,τ_s. Assume that

$$Q_{\tau_i}(y|\mathbf{X}) = \mathbf{X}'\beta(\tau_i), \quad i = 1, 2, \cdots, s. \tag{35}$$

For the τ_jth quantile, $j = 1, 2, \cdots, s$, recall that $\hat{e}_i(\tau_j) = y_i - \mathbf{x}_i'\hat{\beta}_n(\tau_j)$, $e_i(\tau_j) = y_i - \mathbf{x}_i'\beta(\tau_j)$, where $\hat{\beta}_n(\tau_j)$ is the quantile estimator of $\beta(\tau_j)$ using $\mathbf{x}_1, ..., \mathbf{x}_n$. Suppose for $l = 1, 2, ..., s$, $e_i(\tau_l) = G_j(i/n, \mathcal{G}_i, \mathcal{F}_i, \tau_l)$ for $0 \le j \le r$, $b_j < i/n \le b_{j+1}$ are PLS processes. To simplify our notation and without loss of generality, we assume that $e_i(\tau)$ and $\mathbf{x}_i$ have same break points for $\tau = \tau_1, ..., \tau_s$. For $0 \le j \le r$, let

$$\mathbf{W}_j(t, \mathcal{F}_i, \mathcal{G}_i) = (\psi_{\tau_1}(e_j(t, \mathcal{G}_i, \mathcal{F}_i, \tau_1))\mathbf{H}_j(t, \mathcal{G}_i), ..., \psi_{\tau_s}(e_j(t, \mathcal{G}_i, \mathcal{F}_i, \tau_s))\mathbf{H}_j(t, \mathcal{G}_i)),\ b_j < t \le b_{j+1}$$

be a $1 \times sp$ vector. Define $\tilde{\Sigma}^2(t) = \sum_{k=-\infty}^{\infty} cov(\mathbf{W}_i(t, \mathcal{F}_0, \mathcal{G}_0), \mathbf{W}_i(t, \mathcal{F}_k, \mathcal{G}_k))$ if $t \in (b_i, b_{i+1}]$. Assume [A2*]: the smallest eigenvalue of $\tilde{\Sigma}^2(t)$ is bounded away from 0 on (0,1]. We have the following theorem:

Theorem 7. *Suppose the condition [S0]-[S2], [A0]-[A1] hold with $e_i(\alpha)$ replaced by $e_i(\tau_j)$,*

$j=1,2,...s$. *Assume [A2*]. Then we have,*

$$\max_{1\le i\le s}\sqrt{n}(\max_{N\le j\le n}|\hat\beta_j(\tau_i)-\hat\beta_n(\tau_i)|)\Rightarrow_{n\to\infty}$$
$$\max_{1\le i\le s}(\sup_{t\in(0,1]}|\Lambda^{-1}(t,\tau_i)U(t,\tau_i)-\Lambda^{-1}(1,\tau_i)U(1,\tau_i)|), \qquad (36)$$

where $U(t,\tau_i)=(U_{(i-1)p+1}(t),...,U_{ip}(t))$ *is a* p*-dimensional vector formed by* $[(i-1)p+1]_{th}$,...,$[ip]_{th}$ *entry of* $U(t)$*, where* $U(t)=(U_1(t),...,U_{sp}(t))^T$ *is a zero mean Gaussian process with covariance function* $\gamma(s,t)=\int_0^{\min(s,t)}\tilde\Sigma^2(r)dr$*. Similarly,*

$$\max_{1\le i\le s}(\max_{N\le j\le n}|\frac{\sum_{l=1}^{j}\psi(\hat e_l(\tau_i))\mathbf{x}_l}{\sqrt{n}}|)\Rightarrow_{n\to\infty}$$
$$\max_{1\le i\le s}(\sup_{t\in(0,1]}|U(t,\tau_i)-\Lambda(t,\tau_i)\Lambda^{-1}(1,\tau_i)U(1,\tau_i)|). \qquad (37)$$

Define sp-dimensional vectors $\psi^\diamond(\hat e_r)=(\psi(\hat e_r(\tau_1))\mathbf{x}_r,...,\psi(\hat e_r(\tau_s))\mathbf{x}_r)^T$,

$$\hat\varpi^\diamond_{j,m}=\sum_{r=j}^{j+m-1}\psi^\diamond(\hat e_r),\ \hat\varpi^\diamond_n=\hat\varpi^\diamond_{1,n},\ \tilde\Psi^\diamond_{m,n}(t)=\Psi^\diamond_{t_*n,m}+n(t-t_*)(\Psi^\diamond_{t^*n,m}-\Psi^\diamond_{t_*n,m}), \qquad (38)$$

$$\Psi^\diamond_{i,m}=\sum_{j=1}^{i}\frac{1}{\sqrt{m(n-m+1)}}(\hat\varpi^\diamond_{j,m}-\frac{m}{n}\hat\varpi^\diamond_n)R_j,\ i=1,...,n-m+1, \qquad (39)$$

where $\{R_j\}_{j=1}^n$ are $i.i.d$ standard normal r.v's independent of $\{\mathcal{F}_i\}_{i=-\infty}^{\infty}$, $\{\mathcal{G}_i\}_{-\infty}^{\infty}$. Similar to the proofs of Theorems 4 and 5, we have: if c_n and m are of the same orders as in Theorem 4 and Theorem 5, respectively, then

i) $\max_{1\le l\le s}\sup_{t\in(0,1]}|\Lambda(t,\tau_l)-\sum_{i=1}^{\lfloor nt\rfloor}\frac{\phi(\hat e_i(\tau_l)/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}|=o_p(1)$,

ii) Conditioning on $\mathcal{F}_n$, $\mathcal{G}_n$, $\tilde\Psi^\diamond_{m,n}(t)\Rightarrow U(t)$ on $\mathcal{C}(0,1)$ with the uniform topology.

Hence our robust bootstrap can be applied in the same way to test multiple conditional quantiles. The detailed implementation procedures are very similar to the single quantile case and are omitted here.

5 Simulation studies

5.1 Type I error.

In this section we shall compare our testing procedure with existing tests for structural change in quantile regression (Qu 2008) via Monte Carlo experiments. Throughout our simulations the number of bootstrap sample $B = 2000$. To estimate $\{\Lambda(t, \alpha), t \in (0, 1]\}$, we choose bandwidth from 100 equally spaced points in a certain range. In each iteration we select bandwidth by the method we proposed in Section 3.4. The following heteroscedastic linear quantile regression model is considered:

$$y_i = 1 + x_i + (1 + \gamma x_i)(e_i - Q_\alpha(e_i)). \tag{40}$$

for $i = 1, ..., n$, $\gamma = 0.2$, and a pre-specified quantile $\alpha \in (0, 1)$. In our simulations, x_i are i.i.d chi-square random variables with degrees of freedom 3. We shall consider the following models for $\{e_i\}_{i=-\infty}^{\infty}$: (The filtration $\mathcal{F}_i$ is generated by $\{\varepsilon_s\}_{s=-\infty}^{i}$ in the following all models)

I Consider

$$e_i = 0.75 \cos(2i\pi/n)e_{i-1} + \varepsilon_i, \tag{41}$$

where ε_i are *i.i.d* N(0,1). This is a tvAR(1) model, and the corresponding approximating PLS process is locally stationary since its AR(1) coefficient $0.75 \cos(2\pi t)$ changes smoothly over $[0, 1]$.

II Consider $e_i = z_{1,i}\mathbf{1}(0 < i/n \leq 0.8) + z_{2,i}\mathbf{1}(0.8 < i/n \leq 1)$, where

$$z_{1,i} = 0.75 \cos(2i\pi/n)z_{1,i-1} + \varepsilon_i, \;\; z_{2,i} = (0.5 - i/n)z_{2,i-1} + \varepsilon_i. \tag{42}$$

where $\varepsilon_i' s$ are *i.i.d* N(0,1). This is piecewise tvAR(1) model. The AR(1) coefficient of the corresponding approximating PLS process changes smoothly before and after $t = 0.8$, with an abrupt change on $t = 0.8$.

III An usual AR(1) model: $e_i = 0.5e_{i-1} + \varepsilon_i$ for $\varepsilon_i' s$ are *i.i.d* N(0,1). This model is

Table 1: Simulated type I error rate in % for Gradient Method with nominal level γ=5%,10% under model I,II,III, IV, in quantiles α=0.5,0.75,0.9 with sample size n=300, 600.

	$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 0.9$			
	$n = 300$		$n = 600$		$n = 300$		$n = 600$		$n = 300$		$n = 600$	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
I	4.25	10.85	4.6	10.45	4.0	10.1	4.55	10.9	3.6	8.1	3.6	9.8
II	3.35	8.35	4.9	9.7	4.7	9.95	4.1	8.95	4.25	9.65	4.0	9.1
III	3.5	8.6	5.0	10.2	3.5	8.6	4.2	8.75	3.5	8.7	3.95	8.95
IV	3.6	9.1	4.3	9.15	4.75	9.4	4.4	9.75	4.1	9.65	4.25	9.55

stationary.

IV *i.i.d* standard normal.

We report the simulated type I errors in Tables 1-3 by using test based on gradient vectors (16), based on coefficients (15) and SQ method (Qu 2008), respectively. (The SQ method is documented to be superior to the SW method in Qu (2008), so we focus on SQ method in our paper). We exam the 0.5th, 0.75th and 0.9th quantiles with sample sizes 300 and 600 and two nominal levels, 5% and 10%. For the two methods we proposed, the simulated Type I errors are quite close to nominal level. As expected, the increase of sample size from 300 to 600 significantly improves the simulation results, and the two tests perform better when the quantile is not extreme. The Monte Carlo experiments also show the inadequacy of the SQ method when e_i shows dependence and stationarity (III) or dependence and approximately (piecewise) locally stationarity (I, II). The SQ method works well for the *i.i.d.* error model IV, which is consistent with the results reported in Qu (2008).

5.2 Simulated Power.

We consider the alternative model that

$$y_i = 1 + x_i(1 + \delta \mathbf{1}(i \geq \lfloor n/2 \rfloor)) + (1 + \gamma x_i)(e_i - Q_\alpha(e_i)). \tag{43}$$

Table 2: Simulated type I error rate in % for Coefficient Method with nominal level γ=5%,10% under model I,II,III, IV in quantiles α=0.5,0.75,0.9 with sample size n=300, 600.

	$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 0.9$			
	$n = 300$		$n = 600$		$n = 300$		$n = 600$		$n = 300$		$n = 600$	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
I	6.15	9.05	6.4	8.75	6.1	10.2	6.0	10.7	6.45	9.55	5.75	10.0
II	5.9	8.55	6.0	8.9	6.2	9.3	5.65	10.3	6.1	10.7	5.95	9.65
III	5.1	8.95	5.0	9.1	5.4	8.5	4.65	8.95	5.0	9.5	5.0	10.2
IV	5.2	10.45	4.6	10.1	4.85	10.0	4.55	9.05	6.4	11.2	4.6	9.05

Table 3: Simulated type I error rate in % for SQ method with nominal level γ=5%,10% under model I,II,III in quantiles α=0.5,0.75,0.9 with sample size n=300, 600.

	$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 0.9$			
	$n = 300$		$n = 600$		$n = 300$		$n = 600$		$n = 300$		$n = 600$	
α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
I	13.8	22.45	14.35	23.75	12.55	21.7	14.1	24.3	11.7	17.9	12.35	20.2
II	8.0	14.5	8.45	15.8	7.35	13.75	8.45	15.4	7.25	12.4	7.9	14.55
III	21.15	34.35	23.55	34.7	19.15	30.6	20.9	32.15	12.35	20.4	15.4	24.65
IV	4.7	9.5	5.0	9.25	4.75	9.2	5.15	10.7	2.95	6.65	4.7	9.4

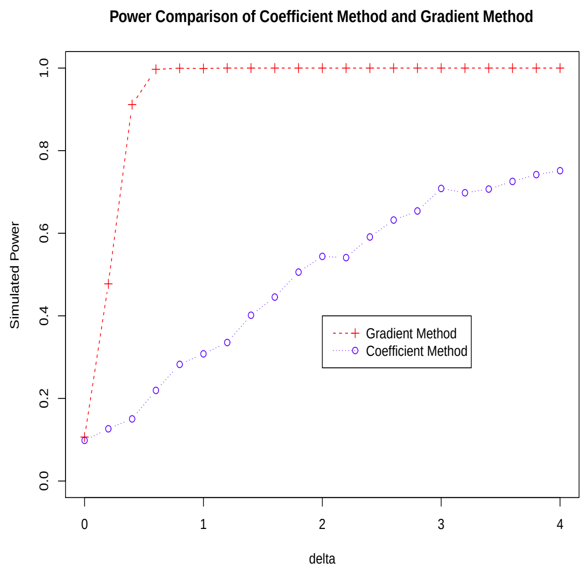

Figure 1: Simulated power with error e_i following Model I for coefficient method and gradient method, respectively.

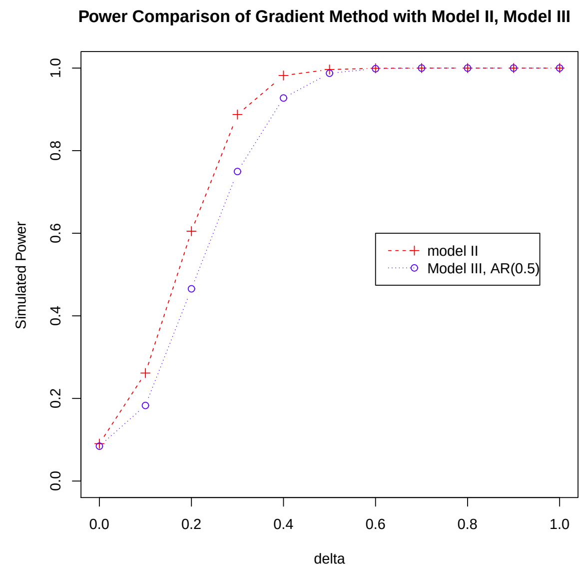

Figure 2: Simulated power for gradient methods when the error e_i follows model II and III for quantile $\alpha = 0.5$, respectively.

We shall simulate for different jump sizes δ to investigate the power performances of our testing procedures. The sample size $n = 400$ in our simulation. Figure 1 examines the simulated powers for the tests based on the gradient test (16) and the coefficient test (15). The quantile we choose is $\alpha = 0.3$ and $e_i(\alpha)$ follows model I. It can be seen that the moderate sample power of (16) is better than (15). It is likely due to the fact that estimators of the ratio $\Lambda(t, \alpha)\Lambda^{-1}(1, \alpha)$ is more stable and accurate than estimators of the sparsity matrix $\Lambda^{-1}(t, \alpha)$ under the alternative hypothesis. The inaccuracy in estimating $\Lambda^{-1}(t, \alpha)$ under H_A leads to a significant power loss in the coefficient tests. In summary, our Monte Carlo experiments suggest that the gradient test (16) performs better than the coefficient test (15) in terms of finite sample power and hence is more recommended in practice. Figure 2 is the power of the gradient test when the error $e_i(\alpha)$ follows model II and model III, where we choose $\alpha = 0.5$ here. It shows that the gradient method has decent power in both 2 cases. Figure 3 is the simulated power of the gradient method and SQ method when the error $e_i(\alpha)$ follows model IV. We also choose $\alpha = 0.5$ in this case. Since now the error is *i.i.d.*, the SQ method is valid in this case. We observe that the gradient test has moderately higher power than the SQ method in this case.

6 Data Analysis.

In this section, we apply our robust method of testing lack of fit based on gradient vectors (16) to annualized quarterly real US GDP growth data, chained in 2009 dollars. The data can be downloaded from the website of U.S. Bureau of Economic Analysis (BEA). It's a well known fact that the volatility of the US real GDP growth substantially decline, which is called the "Great Moderation". McConnell and Perez-Quiros (2000) detected a large break of residual variance of AR(1) model in the first quarter of 1984. See Figure 4. Oka and Qu (2011) performed test for parameter stability in quantile AR(2) models, and discovered the heteroscedasticity of structural change in different quantiles. In our analysis, we revisit the quarterly U.S real GDP growth rate data from 1947:2 to 2009:2, the same period analyzed by Oka and Qu (2011). Nine equally spaced quantiles, from $\tau = 0.2$ to $\tau = 0.8$, are considered to exam the central tendency and the dispersion of the conditional distribution. As in Oka and Qu (2011), BIC conservatively selects the lag $p = 2$ for all nine quantiles under consideration.

As a result, we consider the AR(2) model $y_t = \alpha_0(\tau) + \alpha_1(\tau)y_{t-1} + \alpha_2(\tau)y_{t-2} + e_t(\tau)$ where y_t is the annualized real US GDP growth rate and $e_t(\tau)$ are assumed to be PLS processes. However, after fitting the AR(2) model, the residuals of $\tau = 0.575$, for instance, shows a structural change: the dispersion of the error suddenly shrinks, which implies the non-stationarity of e_t. We also plot the PACF of the residuals after the breaking date 1984:1 (McConnell and Perez-Quiros (2000)) at $\tau = 0.575$, and find that there may exist serial dependence in $e_t(\tau)$. See Figure 5 and Figure 6. Our analysis also indicates that the covariance structure of the errors within the two periods (before and after 1984:1) is time-varying. The non-stationary error structure motivates us to apply our robust method of detecting structural break. We report our testing results in Table 5. We obtained our critical value via simulating 5000 iterations of the bootstrap. From the table, we find that there exists an asymmetry in structural change, i.e., the most of the low quantiles stays unchanged while the test shows that the high quantile, $0.65, 0.725, 0.825$ have change points at 1% significance level, which is confirmed by Oka and Qu (2011). However, our result also implies that the extreme low quantile, 0.2, also has structural change during the period considered. This coincide with Oka and Qu (2011)'s result obtained by testing the sub-sample, but conflicts with their testing result using the whole sample. By latter, Oka and Qu (2011) argued that ”the recessions have remained just as severe when they occurred”, while we doubt that the big recessions may have different severity from before. We also provide results of simultaneously testing the nine equally-spaced quantiles jointly in Table 6 and find strong evidence that there is structural change in at least one of the nine equally-spaced quantiles.

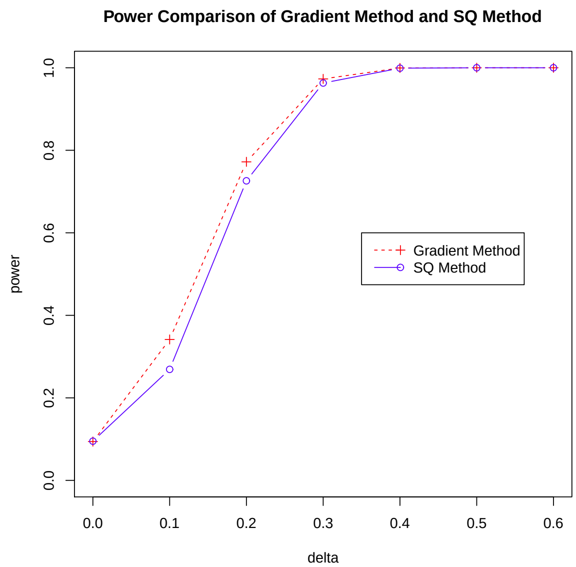

Figure 3: Simulated power for Gradient and SQ methods when errors are $i.i.d$, quantile $\alpha = 0.5$.

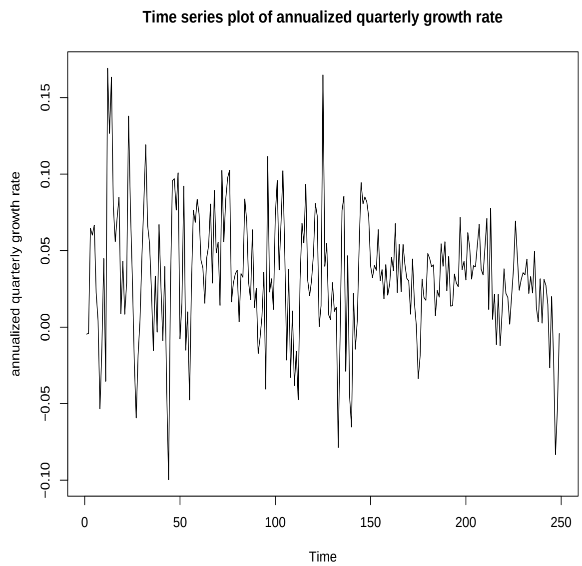

Figure 4: Annualized quarterly growth rate of US real GDP, implies non-stationary and structural break.

7 Technique Appendix.

In the following establish the proofs of theorems. More details and proof of propositions/lemmas are in the supplemental materials of the paper. Without loss of generality, we assume $\alpha = 0.5$ thus omit the (sub)script α if there is no confusion caused. We assume that under H_0 (there is no structural change), $\beta = 0$. We also omit subscript c_n if there is no confusion caused for short. To establish our result, we utilize the fact that

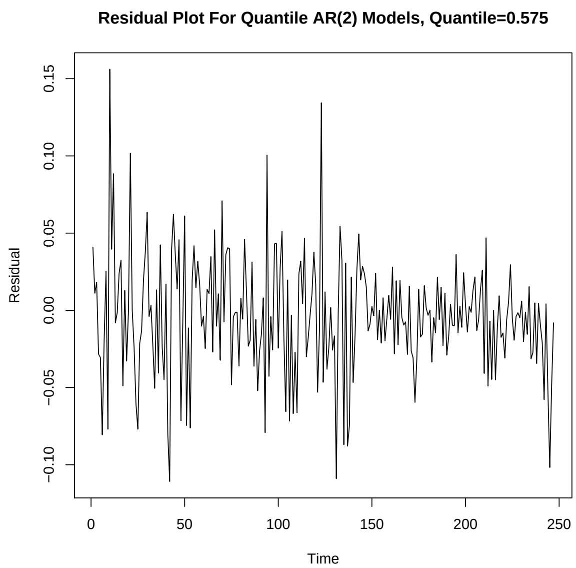

Figure 5: The residual plot of fitted quantile AR(2) model with 0.575 quantile, implies possible non-stationarity in disturbance.

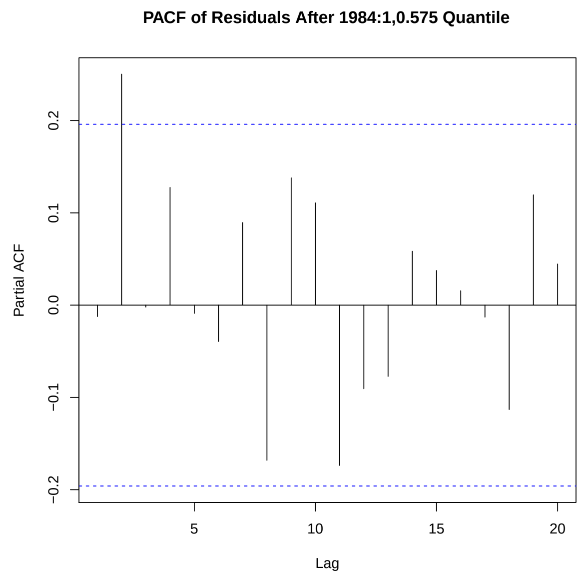

Figure 6: Partial autocorrelation function of residual $\hat{e}_i(0.575)$ after 1984:1, indicates possible existence of the serial dependence of error $e_i(0.575)$ in the period.

Table 4: Test statistics and simulated critical value for different quantile τ's of real US.GDP data in chained 2009 dollars[a]

τ	0.2	0.275	0.35	0.425	0.5	0.575	$0.65^{\triangle\triangle}$	$0.725^{\triangle}$	$0.8^{\triangle}$
Test statistics	9.42*	9.52	8.66	4.11	6.51	9.52	14.56**	12.85**	12.81**
Bandwidth c_n	.024	.038	.015	.042	.045	.018	.031	.028	.048
Bandwidth m	8	11	10	14	12	14	8	8	8
90% C.V	8.27	8.78	8.83	8.67	7.64	8.96	9.24	8.93	8.23
95% C.V	9.26	9.79	9.75	9.82	8.43	10.05	10.36	9.98	9.06
99% C.V	10.93	11.97	11.10	12.30	10.31	12.08	12.52	11.94	11.23

[a]For $\triangle$: 0.65th quantile rejects H_0 of simultaneous test of lack of fit at 1% significance level, while the 0.8th and 0.725th quantiles reject at 5% significance level, see Table 5.

Table 5: Simultaneous test statistics and simulated critical value for different quantile τ's of real US.GDP data in chained 2009 dollars

Test Stat.	Bandwidth m	90% C.V	95% C.V	99% C.V
14.56**	10	10.86	11.77	13.63

Proposition 5. *Suppose A_n are sets such that $\mathbb{P}(A_n) \to 0$ as $n \to \infty$, and $X_n \mathbf{1}(\bar{A}_n) = O_p(1)$. Then $X_n = O_p(1)$.*

The following propositions are also needed:

Proposition 6. *Assume [S0]-[S2], then $\psi(e_i(\alpha))\mathbf{x}_i$ satisfies: for all $i \in [0, r_1]$ and $t, s \in (c_i, c_{i+1}]$, and some constant $\iota > 1/4$, i) $\|\tilde{G}_{i,\alpha}(t, \mathcal{F}_0, \mathcal{G}_0) - \tilde{G}_{i,\alpha}(s, \mathcal{F}_0, \mathcal{G}_0)\| \le C|t-s|^{\iota}$. ii) $\|\tilde{G}_{i,\alpha}(t, \mathcal{F}_0, \mathcal{G}_0)\|_4 < \infty$. Define the L_4 dependence measure for PLS $\tilde{G}_{i,\alpha}(t, \mathcal{F}_i, \mathcal{G}_i)$:*

$$\tilde{\delta}_{4,\alpha}(k) = \max_{0 \le i \le r_1} \sup_{t \in (c_i, c_{i+1}]} \|\tilde{G}_{i,\alpha}(t, \mathcal{F}_k, \mathcal{G}_k) - \tilde{G}_{i,\alpha}(t, \mathcal{F}_k^*, \mathcal{G}_k^*)\|_4.$$

Then we have $\tilde{\delta}_{4,\alpha}(k) = O(\chi_1^k)$ for some $\chi_1 \in (0,1)$.

Proposition 7. *Under condition [A1], i) $\sup_{|x| \le \mathbb{R}} \tilde{\Delta}_s(k, x, q) \le 2M_0 K_0^{-1/s} \chi^{k/s}$ for $s \ge 1$, $0 \le q \le \max\{3, p\}$. Recall that constant M_0, K_0 and χ are defined in [A1].*

Write for $j = 1, ..., n$,

$$M_j(\theta) = \sum_{i=1}^{j} \psi(e_i - \mathbf{x}_i'\theta)\mathbf{x}_i - \mathbb{E}[\psi(e_i - \mathbf{x}_i'\theta)\mathbf{x}_i | \mathcal{F}_{i-1}, \mathcal{G}_n], \tag{44}$$

$$N_j(\theta) = \sum_{i=1}^{j} \mathbb{E}[\psi(e_i - \mathbf{x}_i'\theta)\mathbf{x}_i | \mathcal{F}_{i-1}, \mathcal{G}_n] - \mathbb{E}[\psi(e_i - \mathbf{x}_i'\theta) | \mathcal{G}_n]\mathbf{x}_i. \tag{45}$$

$$\tau_j(\delta) = \sum_{i=1}^{j} \mathbb{E}\{|\mathbf{x}_i|^2 [\psi(e_i + |\mathbf{x}_i \delta|) - \psi(e_i - |\mathbf{x}_i \delta|)]\}. \tag{46}$$

Note that by [S2], [A1], we have, for $\delta \to 0$,

$$\begin{aligned} \tau_n(\delta) &= \sum_{i=1}^{n} \mathbb{E}\left\{|\mathbf{x}_i|^2 \mathbb{E}\left[\psi(e_i + |\mathbf{x}_i||\delta|) - \psi(e_i - |\mathbf{x}_i||\delta|)|\mathcal{G}_n\right]\right\} \\ &= \sum_{i=1}^{n} \mathbb{E}\{2|\mathbf{x}_i|^2 f_{w(i)}(i/n, 0|\mathcal{G}_i)|\mathbf{x}_i||\delta| + O(|\mathbf{x}_i^5||\delta^3|)\} = O(n\delta). \end{aligned} \tag{47}$$

Let K_1 be the constant that $\tau_n(\delta) \le K_1 n\delta$. Let $K_j(\theta) = \Omega_j(\theta) - \mathbb{E}(\Omega_j(\theta)|\mathcal{G}_n)$, where $\Omega_j(\theta) = \sum_{i=1}^{j} \psi(e_i - \mathbf{x}_i'\theta)\mathbf{x}_i$.

Lemma 2. *Suppose conditions of Theorem 1 hold. Let $\{\delta_i\}_{i=1}^n$ be a number array such that $\delta_n \to 0$. Then there exists a set W_n such that, for $0 < t \le n^3$, M be large enough constant,*

$$\lim_{n\to\infty} \mathbb{P}(W_n) = 0, \tag{48}$$

$$\mathbb{E}\{\exp(t \sup_{|\theta|\le\delta_n} |M_n(\theta) - M_n(0)|\mathbf{1}(\bar{W}_n))\} \le M \exp(4t\sqrt{K_1 n\delta_n} \log n). \tag{49}$$

Proof. Following Wu (2007), for any positive real sequence $g_n \to \infty$, $\delta_n > 0$, define

$$\phi_n = 2g_n\sqrt{\tau_n(\delta_n)}\log n, \quad t_n = \frac{g_n\sqrt{\tau_n(\delta_n)}}{\log g_n}, \quad u_n = t_n^2.$$

$$\eta_i(\theta) = [\psi(e_i - \mathbf{x}_i'\theta) - \psi(e_i)]\mathbf{x}_i, \; T_n = \max_{i\le n} \sup_{|\theta|\le\delta_n} |\eta_i(\theta)|,$$

$$U_n = \sum_{i=1}^n \mathbb{E}\{[\psi(e_i + |\mathbf{x}_i|\delta_n) - \psi(e_i - |\mathbf{x}_i|\delta_n)]^2|\mathbf{x}_i|^2|\mathcal{F}_{i-1}, \mathcal{G}_n\}.$$

Then by monotonicity, one can get

$$\mathbb{E}(\sup_{|\theta|\le\delta_n} |\eta_i(\theta)|^2) \le \mathbb{E}\{|\mathbf{x}_i|^2[\psi(e_i + |\mathbf{x}_i|\delta_n) - \psi(e_i - |\mathbf{x}_i|\delta_n)]^2\}. \tag{50}$$

Hence we have $\mathbb{E}(T_n^2) \le \tau_n(\delta_n)$. Then by Markov's inequality,

$$\mathbb{P}(T_n > t_n) \to 0. \tag{51}$$

By our settings, $\mathbb{E}(U_n) \le \tau_n(\delta_n)$, and we have

$$\mathbb{P}(U_n > u_n) \to 0. \tag{52}$$

Let $l = n^8$ and $G_l = \{|\theta| \le \delta_n, (k_1/l, ..., k_p/l) : k \in \mathbb{Z}, |k_i| \le n^8\}$, following Wu (2007), we define $\lceil a \rceil_l = \lceil al \rceil / l$, and $\lfloor a \rfloor_l = \lfloor al \rfloor / l$. Write $\mathbf{x}_i = (x_{i,1} ..., x_{i,p}), \Pi_p = \{-1, +1\}^p$. For fixed $\mathbf{x}$, set $D_{\mathbf{x}}(i) = (2 \times \mathbf{1}(x_{i1} \ge 0) - 1, ..., 2 \times \mathbf{1}(x_{ip} \ge 0) - 1) \in \Pi_p$. For $\mathbf{d} \in \Pi_p$ and $1 \le j \le p$, define

$$M_{n,j,\mathbf{d}}(\theta) = \sum_{i=1}^{n} \{\psi(e_i - \mathbf{x}_i'\theta) - \mathbb{E}(\psi(e_i - \mathbf{x}_i'\theta)|\mathcal{F}_{i-1}, \mathcal{G}_n)\} x_{ij} \mathbf{1}(D_{\mathbf{x}}(i) = \mathbf{d}), \tag{53}$$

$$\eta_{i,j,\mathbf{d}}(\theta) = [\psi(e_i - \mathbf{x}'\theta) - \psi(e_i)] x_{ij} \mathbf{1}(D_{\mathbf{x}}(i) = \mathbf{d}). \tag{54}$$

Define $A_{n,i,\mathbf{d}} = \{\sup_{\theta \in G_l} |M_{n,i,\mathbf{d}}(\theta) - M_{n,i,\mathbf{d}}(0)| \geq 2\phi_n\}$, $B_n = \{T_n \leq t_n, U_n \leq u_n\}$, $B_n(\theta) = \sum_{i=1}^{n} \mathbb{E}[\eta_{i,j,\mathbf{d}}(\theta) \mathbf{1}(|\eta_{i,j,\mathbf{d}}(\theta)| > t_n)|\mathcal{F}_{i-1}, \mathcal{G}_n]$, by $u_n = o(t_n \phi_n)$, we have, for large n,

$$\mathbb{P}(|B_n(\theta)| \geq \phi_n, U_n \leq u_n) \leq \mathbb{P}(t_n^{-1} U_n \geq \phi_n, U_n \leq u_n) = 0. \tag{55}$$

Since $t_n \phi_n \log n = o(\phi_n^2)$, $u_n \log n = o(\phi_n^2)$, by similar argument in Lemma 4 of Wu (2007), use Proposition 2.1 of Freedman (1975), for any $\zeta > 1$,

$$\mathbb{P}(A_{n,i,\mathbf{d}} \bigcap B_n) = O(n^{-\zeta p}), \tag{56}$$

and thus

$$\mathbb{P}(\bigcup_{i=1}^{p} \bigcup_{\mathbf{d}} (A_{n,i,\mathbf{d}} \bigcap B_n)) = O(n^{-\zeta' p}) \tag{57}$$

for any $\zeta' > 1$, where $\bigcup_{\mathbf{d}}$ represents that the union of all possible $\mathbf{d}$. Define $A_n = \bigcup_{i=1}^{p} \bigcup_{\mathbf{d}} A_{n,i,\mathbf{d}}$, then

$$\mathbb{P}(A_n \bigcap B_n) = O(n^{-\zeta' p}) \tag{58}$$

Note that $\lim_{n\to\infty} \mathbb{P}(B_n) = 1$, so $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$.

On the other hand, define

$$C_n = \{\max_{1 \leq i \leq n} |\mathbf{x}_i| \geq \frac{2}{t_x} \log n\}. \tag{59}$$

By assumption [S2] and Markov's inequality, it is easy to see that $\lim_{n\to\infty} \mathbb{P}(C_n) = 0$ via Markov inequaliy. Let $W_n = A_n \cup C_n$. By $M_n = \sum_{\mathbf{d} \in \Pi_p} (M_{n,1,\mathbf{d}}, ..., M_{n,p,\mathbf{d}})$, we only need to show that the theorem holds with M_n replaced by $M_{n,j,\mathbf{d}}$ for all $\mathbf{d} \in \Pi_p$ and $1 \leq j \leq p$.

By [A1],

$$|\mathbb{E}(\psi(e_i - l) - \psi(e_i - s)|\mathcal{F}_{i-1}, \mathcal{G}_n)| \leq \max_{t,x} f_{w(i)}(t, x|\mathcal{F}_{i-1}, \mathcal{G}_i)|s - l| \leq M_0|s - l|.$$

Then by the similar chaining argument to (58) of Wu (2007), for $t > 0$,

$$\mathbb{P}\{[\exp(\mathbf{1}(\bar{W}_n)t \sup_{|\theta|\leq\delta_n} |M_{n,1,\mathbf{d}}(\theta) - M_{n,1,\mathbf{d}}(0)|) > \exp(\mathbf{1}(\bar{W}_n)(2t\phi_n + tnM_0(\frac{2}{t_x})^2 \log^2 n/l))]|\mathcal{G}_n\} = 0. \quad (60)$$

consequently, by $l = n^8$, $0 < t \leq n^3$, For n large enough, $tnM_0(\frac{2}{t_x})^2 \log^2 n/l \leq \log 2$,

$$\mathbb{E}\{\exp(t \sup_{|\theta|\leq\delta_n} |M_n(\theta) - M_n(0)|\mathbf{1}(\bar{W}_n))\} \leq \mathbb{E}\{2\exp((2t\phi_n)\mathbf{1}(\bar{W}_n))\}. \quad (61)$$

By the definition of ϕ_n and the fact that the rate at which $g_n \to \infty$ is allowed to be arbitarily slow, we get proof. □
By the Burkholder Inequality and [A1], one can show that

Lemma 3. *Suppose conditions of Theorem 1 hold, $\delta_n \to 0$,*

$$\| \sup_{|g|\leq\delta_n} |N_n(g) - N_n(0)|\|_s \leq C\delta_n(s^{1/2}\sqrt{n}\log^{2p+2} n/(1 - \chi^{1/s}) + n^2 s^{p+1} n^{\frac{1-t_x \log n}{2s}}). \quad (62)$$

where C is large constant independent of s and n.

Proof. Let $I = \{\alpha_1, ..., \alpha_q\} \subseteq \{1, ..., p\}$ be a nonempty set and $1 \leq \alpha_1 < ... < \alpha_q$. For a p-dimensional vector $\mathbf{u} = (u_1, ...u_p)$, let $\mathbf{u} = (u_1\mathbf{1}(1 \in I), ..., u_p\mathbf{1}(p \in I))$, write

$$\int_0^{g_I} \frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I} d\mathbf{u}_I = \int_0^{g\alpha_1} ... \int_0^{g\alpha_q} \frac{\partial^q N_n(\mathbf{u}_I)}{\partial u_{\alpha_1} ... \partial u_{\alpha_q}} du_{\alpha_1} ... du_{\alpha_q}. \quad (63)$$

Let $w_i = \mathbf{x_i} x_{i\alpha_1} ... x_{i\alpha_q}$, we have, for $1 \leq q \leq p$,

$$\frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I} = \sum_{i=1}^{n} (\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I))w_i|\mathcal{F}_{i-1}, \mathcal{G}_n) - \mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)w_i|\mathcal{G}_n). \quad (64)$$

By triangular inequality, we have

$$\| \sup_{|\mathbf{g}|\le \delta_n} |N_n(\mathbf{g}) - N_n(0)|\|_s \le \sum_{I\subseteq\{1,\ldots,p\}} \int_{-\delta_n}^{\delta_n} \cdots \int_{-\delta_n}^{\delta_n} \|\frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I}\|_s d\mathbf{u}_I. \tag{65}$$

Note that

$$\frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I} = (\sum_{k=0}^{\infty}\sum_{i=1}^{n} \mathcal{P}_{i-k,n}\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)w_i|\mathcal{F}_{i-1}, \mathcal{G}_n), \tag{66}$$

where $\mathcal{P}_{i,n}(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_i, \mathcal{G}_n) - \mathbb{E}(\cdot|\mathcal{F}_{i-1}, \mathcal{G}_n)$. By triangular inequality, it is easy to see that for s positive integer,

$$\|\frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I}\|_s \le \sum_{k=0}^{\infty} \|\sum_{i=1}^{n} \mathcal{P}_{i-k,n}\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)w_i|\mathcal{F}_{i-1}, \mathcal{G}_n)\|_s. \tag{67}$$

Denote $J_k = \sum_{i=1}^{n} \mathcal{P}_{i-k,n}\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)w_i|\mathcal{F}_{i-1}, \mathcal{G}_n)$, and $D_n = \{\max_{1\le i\le n} |\mathbf{x}| \le \log^2 n\}$. Note that the summands of J_k are martingale differences, D_n is $\mathcal{G}_n$ measurable, $|w_i\mathbf{1}(D_n)| \le \log^{2p+2} n$, By Burkholder inequality, and triangular inequality, for any interger $s \ge 2$,

$$\begin{aligned}
\|J_k\mathbf{1}(D_n)\|_s^2 &\le C_s^2\|(\sum_{i=1}^{n}[\mathcal{P}_{i-k}\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)\mathbf{1}(D_n)w_i|\mathcal{F}_{i-1}, \mathcal{G}_n)]^2)^{1/2}\|_s^2 \\
&\le C_s^2 \sum_{i=1}^{n} \|\mathcal{P}_{i-k}\mathbb{E}^{(q)}(\psi(e_i - \mathbf{x}_i'\mathbf{u}_I)|\mathcal{F}_{i-1}, \mathcal{G}_n)\|_s^2 \log^{4p+4} n \\
&\le C_s^2 n \sup_{|x|\le\mathbb{R}} \Delta_s(k, x, q)^2 \log^{4p+4} n. \qquad (68)
\end{aligned}$$

where C_s is a number only depend on s, and by Burkholder (1973), $C_s \le 18s^{1/2}$. The last inequality holds by [A1] and Lemma 1 of Wu (2007). As a result, apply Proposition 7, we have

$$\|\sum_{k=0}^{\infty} J_k\mathbf{1}(D_n)\|_s/\log^{2p+2} n \le 18s^{1/2}\sum_{k=0}^{\infty}\sqrt{n}\sup_{|x|\le\mathbb{R}} \Delta_s(k, x, q) \le Cs^{1/2}\sqrt{n}/(1-\chi^{1/s}). \tag{69}$$

On the other hand, by the boundedness of conditional density in condition [A1],

$$|\frac{\partial^q N_n(\mathbf{u}_I;\mathcal{G}_n)}{\partial \mathbf{u}_I}\mathbf{1}(\bar{D}_n)| = |\sum_{i=1}^n(\mathbb{E}^{(q)}(\psi(e_i-\mathbf{x}_i'\mathbf{u}_I))w_i|\mathcal{F}_{i-1},\mathcal{G}_n) - \mathbb{E}^{(q)}(\psi(e_i-\mathbf{x}_i'\mathbf{u}_I)w_i|\mathcal{G}_n)\mathbf{1}(\bar{D}_n)|$$
$$\leq 2M_0 n\mathbf{1}(\bar{D}_n)\max_{1\leq i\leq n}|w_i|. \tag{70}$$

By condition [S2], there exists constant C, such that $\max_{1\leq i\leq n}\|x_i\|_v \leq Cv$ for $v\geq 1$, then by *Hölder's* inequality, $\sum_{1\leq i\leq n}\|w_i\|_v \leq n(C(p+1)v)^{p+1}$. As a consequence, we have, for large constant C_1, C_2,

$$\begin{aligned}\|\frac{\partial^q N_n(\mathbf{u}_I;\mathcal{G}_n)}{\partial \mathbf{u}_I}\mathbf{1}(\bar{D}_n)\|_s &\leq 2M_0 n\|\max_{1\leq i\leq n}|w_i|\|_{2s}\|\mathbf{1}(\bar{D}_n)\|_{2s}\\ &\leq C_1 n^2((p+1)s)^{p+1}(\mathbb{E}\{\frac{\exp(t_x\max_{1\leq i\leq n}|\mathbf{x}_i|)}{\exp(t_x\log^2 n)}\})^{\frac{1}{2s}}\\ &\leq C_2 n^2((p+1)s)^{p+1}n^{\frac{1-t_x\log n}{2s}}.\end{aligned} \tag{71}$$

Combine inequality (69)(71), we have

$$\|\frac{\partial^q N_n(\mathbf{u}_I)}{\partial \mathbf{u}_I}\|_s \leq C_3(s^{1/2}\sqrt{n}\log^{2p+2}n/(1-\chi^{1/s}) + n^2 s^{p+1}n^{\frac{1-t_x\log n}{2s}}). \tag{72}$$

We complete the proof by combining equation (65), (72). □

Proof of Lemma 1:

Lemmas follows by Lemma 2 and Lemma 3, the convexity of checking functions and the Proposition 1 in the supplemental material. Details are also in supplemental material.□

For any sequence of random variables $\{Z_i\}_{i=1}^n$, for $p>1$, we have that $\exp[\max_{1\leq i\leq n}|Z_i|] \leq \sum_{i=1}^n \exp[|Z_i|]$ and $\max_{1\leq i\leq n}|Z_i|^p \leq \sum_{i=1}^n |Z_i|^p$. Write $K_j(\theta) = M_j(\theta)+N_j(\theta)$, by using these two facts and Lemmas 2 and 3, we shall see that

Lemma 4. *Suppose conditions of Theorem 1 hold,*

$$\max_{1\leq j\leq n}\sup_{|\theta|\leq\delta_n}|K_j(\theta)-K_j(0)| = O_p(n^{1/2}\delta_n^{1/2}\log^{2p+7/2}n), as$$

$\delta_n \to 0$. *In addition, if there exists some* $\iota>0$, *such that* $\delta_n n^{\iota}\to 0$, *then the order can be* *reduced to* $\sqrt{n\delta_n}\log n$.

Proof. See supplemental material, details are ommitted. □

On the other hand, by the properties of the gradient vectors, we shall see that

Lemma 5. *Suppose condition [S2] holds, then* $\mathbb{P}(\sup_{1\le j\le n}|\Omega_j(\hat\beta_j)|\le(p+1)\max_{1\le i\le n}|\mathbf{x}_i|)=1$.

Proof. See supplemental material, details are ommitted. □

Proof of Theorem 1.

The theorem 1 follows from Lemma 4, Lemma 1, Lemma 5 and Taylor expansion.

Proof of Theorem 2.

(15) of Theorem 2 follows from Theorem 1 and Proposition 1 immediately. (16) follows from (15), consistent result of Lemma 1 and the following corollary, which is an instant result of triangle inequality:

Corollary 1. *Suppose conditions of Theorem 1 hold. Define* $\Theta=\{\theta:|\theta|\le Cn^{-1/2}\log^4 n\}$. *Then*

$$\sup_{1\le j\le n}\sup_{\theta,\theta'\in\Theta}|K_j(\theta)-K_j(\theta')|=O_p(n^{1/4}\log^3 n).\ \square \tag{73}$$

Proof of Theorem 3.

Theorem 3 follows from the similar arguments as those in the proofs of Theorem 1 and Theorem 2, details are in supplemental material. □

Proof of Theorem 4.

Write $\tilde\lambda_{c_n}(j,\alpha)=\sum_{i=1}^j\frac{\phi(e_i(\alpha)/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}$, $\lambda(j)=\sum_{i=1}^j\frac{f_{w(i)}(i/n,0|\mathcal{G}_n)\mathbf{x}_i\mathbf{x}_i'}{n}$. We show

i)$\max_{1\le j\le n}|\hat\lambda_{c_n}(j,\alpha)-\tilde\lambda_{c_n}(j)|=O_p(n^{-1/2}\log^7 n+\frac{\log^{10}n}{nc_n^3})$.

ii)$\max_{1\le j\le n}|\tilde\lambda_{c_n}(j,\alpha)-\acute\Lambda(j,\alpha)|=O_p(\frac{\log^4 n}{\sqrt{nc_n}}+c_n^2\log^4 n)$. Then the theorem follows from i), ii) and (14). Write

$$\tilde\lambda(j)-\mathbb{E}(\tilde\lambda(j)|\mathcal{G}_n):=M_j+N_j, \tag{74}$$

$$\text{where } M_j=\sum_{i=1}^j(\frac{\phi(e_i/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}-\mathbb{E}(\frac{\phi(e_i/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}|\mathcal{G}_n,\mathcal{F}_{i-1})), \tag{75}$$

$$N_j=\sum_{i=1}^j(\mathbb{E}(\frac{\phi(e_i/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}|\mathcal{G}_n,\mathcal{F}_{i-1})-\mathbb{E}(\frac{\phi(e_i/c_n)\mathbf{x}_i\mathbf{x}_i'}{nc_n}|\mathcal{G}_n)). \tag{76}$$

Then similar to Lemma 4, by properties of martingale differences, the chaining argument and condition [A1], we can get we have,

$$\sup_{1\le j\le n}|\tilde\lambda(j)-\mathbb{E}(\tilde\lambda(j)|\mathcal{G}_n)|\mathbf{1}(D_n)=O_p(n^{-1/2}\log^{\frac{11}{2}}n+\frac{\log^4 n}{\sqrt{nc_n}}), \tag{77}$$

where D_n is an event that $\lim_{n\to\infty}\mathbb{P}\{D_n\}=1$. On the other hand, a Taylor expansion argument of the conditional density $f_{w(i)}(i/n,\cdot|\mathcal{G}_n)$ leads to

$$\sup_{1\le j\le n}|\mathbf{1}(D_n)(\mathbb{E}(\tilde\lambda(j)|\mathcal{G}_n)-\acute\Lambda(j))|=O_p(c_n^2\log^4 n). \tag{78}$$

so ii) of Theorem 4 follows from (77) and (78), i) follows from similar arguments of ii).□

The next proposition is the foundation of proofing Theorem 5, which is a direct result of Zhou (2013). For $m\to\infty, m/n\to 0$, $t_*=\lfloor tn\rfloor/n$, $t^*=t_*+1/n$, define

$$\tilde\Phi_{m,n}(t)=\Phi_{t_*n,m}+n(t-t_*)(\Phi_{t^*n,m}-\Phi_{t_*n,m}), \tag{79}$$

$$\Phi_{i,m}=\sum_{j=1}^{i}\frac{1}{\sqrt{m(n-m+1)}}(\varpi_{j,m}-\frac{m}{n}\varpi_n)R_j, i=1,...,n-m+1, \tag{80}$$

where $\varpi_{j,m}=\sum_{r=j}^{j+m-1}\psi_\alpha(e_r(\alpha))\mathbf{x}_r$, $\varpi_n=\varpi_{1,n}$, and $(R_i)_{i=1}^n$ are *i.i.d* standard normals which are independent of $\{\mathcal{F}_i\}_{i=-\infty}^{\infty},\{\mathcal{G}_i\}_{i=-\infty}^{\infty}$.

Proposition 8. *Under the conditions of Theorem 5, we have,*
conditioning on $\mathcal{F}_n$, $\mathcal{G}_n$, $\tilde\Psi_{m,n}(t)\Rightarrow U_\alpha(t)$ on $\mathcal{C}(0,1)$ with the uniform topology.

Proof of Theorem 5.

We shall show $\sup_{t\in(0,1]}|\tilde\Phi_{m,n}(t)-\tilde\Psi_{m,n}(t)|=O_p(n^{-1/4}\log^{3/2}n)$, then the theorem holds by proposition 8. Write $\varpi_{j,m}(\theta)=\sum_{r=j}^{j+m-1}\psi(e_r-\mathbf{x}_r'\theta)\mathbf{x}_r$, $\varpi_n(\theta)=\varpi_{1,n}(\theta)$, Note that $\varpi_{j,m}-\hat\varpi_{j,m}=\sum_{r=j}^{j+m-1}[\psi(e_r)-\psi(\hat e_{rn})]\mathbf{x}_r=\varpi_{j,m}(0)-\varpi_{j,m}(\hat\beta_n)$. We construct a set W_n independent of $\{R_i\}_{i=-\infty}^{\infty}$, $\mathbb{P}\{W_n\}\to 0$ in the proof of Theorem 5 of Appendix. Write $H_n=\{|\hat\beta_n|\le Cn^{-1/2}\log n\}$, by Lemma 1, $\lim_{n\to\infty}\mathbb{P}(H_n)=1$. Furthermore, H_n is $(\mathcal{F}_n,\mathcal{G}_n)$ measurable. Observe that $(\Phi_{i,m}-\Psi_{i,m})\mathbf{1}(\bar W_n)\mathbf{1}(H_n)=\sum_{j=1}^{i}Y_j(\hat\beta_n)\mathbf{1}(\bar W_n)\mathbf{1}(H_n)R_j$ is a martingale with respect to $\{\mathcal{F}_n, \mathcal{G}_n, \{R_s\}_{s=1}^i\}$, where

$$Y_j(\hat{\beta}_n) = \frac{1}{\sqrt{m(n-m+1)}}(\varpi_{j,m}(0) - \frac{m}{n}\varpi_n(0) - (\varpi_{j,m}(\hat{\beta}_n) - \frac{m}{n}\varpi_n(\hat{\beta}_n))). \tag{81}$$

By the similar arguments of Theorem 1, decompose $Y_j(\theta)$ into the summation of a martingale difference part and a centralized conditional expectation part. Let C, K be a large enough constant. One can show that, for $1 \leq j \leq n-m+1$,

$$\| \sup_{|\theta| \leq Cn^{-1/2}\log n} |Y_j(\theta)| \mathbf{1}(\bar{W}_n)\|_2 \leq Kn^{-1/4}\log^{3/2} n/\sqrt{(n-m+1)}. \tag{82}$$

By Doob's inequality, $\forall\ M \in \mathbb{R}^+$,

$$\begin{aligned} \mathbb{P}\{\sup_{1\leq i\leq n-m+1} |\Phi_{i,m} - \Psi_{i,m}| \mathbf{1}(\bar{W}_n)\mathbf{1}(H_n) \geq M\} &\leq \frac{\||\Phi_{n-m+1,m} - \Psi_{n-m+1,m}|\mathbf{1}(\bar{W}_n)\mathbf{1}(H_n)\|_2^2}{M^2} \\ &\leq \sum_{j=1}^{n-m+1} \| \sup_{|\theta|\leq Cn^{-1/2}\log n} |Y_j(\theta)|\mathbf{1}(\bar{W}_n)\|_2^2/M^2. \end{aligned} \tag{83}$$

Thus $\sup_{1\leq i\leq n-m+1} |\Phi_{im} - \Psi_{im}| = O_p(n^{-1/4}\log^{3/2} n)$. Let $\Theta = \{k/n, k = 1, ..., n\}$, we have $\sup_{0\leq t\leq 1} |\tilde{\Phi}_{m,n}(t) - \tilde{\Psi}_{m,n}(t)| \leq \sup_{t\in\Theta_n} |\tilde{\Phi}_{m,n}(t) - \tilde{\Psi}_{m,n}(t)|(1 + 2\sup_{t^*\in\Theta_n}\sup_{|t-t^*|\leq\frac{1}{n}} n(t - t^*)) = O_p(n^{-1/4}\log^{3/2} n)$, which completes the proof. □

REFERENCES

Andrews, D. W. K. (1993), Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, **61**, 821–856.

Aue, A. and Horvath, L. (2013). Structural breaks in time series. Journal of Time Series Analysis **34** 1-16.

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

Bahadur, R. R. (1966), A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577-580.

Bai, J. (1996), Testing for Parameter Constancy in Linear Regressions: an Empirical Distribution Function Approach, *Econometrica*, **64**, 597-622, 1996.

Bai, J. and Perron, P. (1998) Testing for and Estimation of Multiple Structural Changes, (with Pierre Perron), *Econometrica*, **66**, 47-79, 1998.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31** 307-327.

Brown, R. L., Durbin, J. and Evans, J. M. (1975), Techniques for Testing the Constancy of Regression Relationships Over Time. *Journal of the Royal Statistical Society, B* **37**, 149–163.

Burkholder, D. L. (1973), Distribution function inequalities for martingales. *Annals of Probability* **1**, 19–42.

Dette, H. Wagener, J. and Volgushev, S. (2011). Comparing conditional quantile curves. *Scandinavian Journal of Statistics* **38** 63–88.

Diaconis, P. and Freedman, D. (1999), Iterated random functions. *SIAM Rev.* **41** 45–76.

Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica,* **50** 987–1007.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications* . Chapman and Hall, London.

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

He, X. and Zhu, L. A Lack-of-Fit Test for Quantile Regression. *Journal of the American Statistical Association*, **98** 1013–1022.

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

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

Koenker, R. and Bassett, G. (1978), Regression Quantiles. *Econometrica,* **46** 33–50.

Koenker, R., and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94** 1296–1310.

Krämer, W., Ploberger, W. , and Alt, R. (1988), Testing for Structural Change in Dynamic Models. *Econometrica*, **56**, 1355-1369.

Mccabe, B. P. M., and Harrison, M. J. (1980), Testing the Constancy of Regression Relationships Over Time Using Least Squares Residuals, *Journal of the Royal Statistical Society, C*, **29**, 142–148.

McConnell, M. M., and Perez-Quiros, G. (2000), Output Fluctuations in the United States: What Has Changed since the Early 1980's? *American Economic Review*, **90**, 1464-1476.

Lahiri, S. N. (2003), *Resampling Methods for Dependent Data*, New York: Springer.

Qu, Z. (2008), Testing for Structural Change in Regression Quantiles. *Journal of Econometrics*, **148**, 170-184.

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

Ploberger, W. and Kramer, W. (1992), The CUSUM Test with OLS Residuals, *Econometrica*, **60** 271–85

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, New York: Springer.

Portnoy, S. (1991), Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *J. Multi-variate Analysis*, **38**, 100–113.

Powell, J. L. (1991), Estimation of monotonic regression models under quantile restrictions. *Non-parametric and Semiparametric Methods in Economics.* Cambridge: Cambridge University Press.

Su, L. and Xiao, Z. (2008), Testing for Parameter Stability in Quantile Regression Models. *Statistics & Probability Letters* **78**, 2768–2775.

Tong, H. (1990), Non-linear Time Series: A Dynamical System Approach. *Oxford: Oxford University Press.*

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

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

Wu W. B., Zhou Z. (2011), Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, **21**,1397-413.

Zhang T., Wu W. B. (2012), Inference of time-varying regression models. *The Annals of Statistics*, **40**, 1376–1402.

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

Zhou, Z. and Wu, W. B. (2009), Local linear quantile estimation for nonstationary time series. *Ann. Statist,* **37** 2597–3097.

Zhou, Z. (2013) Heteroscedasticity and Autocorrelation Robust Structural Change Detection. *Journal of the American Statistical Association,* **108** 726–740.