Submitted to the Annals of Statistics

GRADIENT-BASED STRUCTURAL CHANGE DETECTION FOR NON-STATIONARY TIME SERIES M-ESTIMATION

WEICHI WU[1] AND ZHOU ZHOU

University College London and University of Toronto
February 10, 2017

We consider structural change testing for a wide class of time series M-estimation with non-stationary predictors and errors. Flexible predictor-error relationships, including exogenous, state-heteroscedastic and autoregressive regressions and their mixtures, are allowed. New uniform Bahadur representations are established with nearly optimal approximation rates. A CUSUM-type test statistic based on the gradient vectors of the regression is considered. In this paper, a simple bootstrap method is proposed and is proved to be consistent for M-estimation structural change detection under both abrupt and smooth non-stationarity and temporal dependence. Our bootstrap procedure is shown to have certain asymptotically optimal properties in terms of accuracy and power. A public health time series dataset is used to illustrate our methodology, and asymmetry of structural changes in high and low quantiles is found.

1. Introduction. Consider the following stochastic linear regression:

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

where $\{\mathbf{x}_i\}_{i=1}^n$ and $\{e_i\}_{i=1}^n$ are the p-dimensional predictor time series and error series, respectively. We estimate the unknown parameter vector β by an M-estimator $\hat{\beta}_n$:

$$\hat{\beta}_n = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho(y_i - \mathbf{x}_i'\beta), \tag{2}$$

where $\rho(\cdot)$ is a convex loss function with left derivative $\psi(\cdot)$. By choosing different loss functions ρ, (1) contains a wide class of frequently used regression models. For instance, for a pre-specified $\tau \in (0,1)$, $\hat{\beta}_n$ is the estimate of the τ_{th} quantile regression coefficient if we set $\rho(x) = \tau x^+ + (1-\tau)(-x)^+$ with the left derivative $\psi(x) = \tau - \mathbf{1}(x \le 0)$. Other important examples

[1]Corresponding author. Department of Statistics, University College, Gower Street, WC1E 6BT, London, UK.
E-mail: *w.wu@ucl.ac.uk*

Key words and phrases. M-estimation, piece-wise local stationarity, bootstrap, structural change

include expectile regression with $\rho(x) = |\mathbf{1}(x \le 0) - \alpha|x^2$, $0 < \alpha < 1$, robust $\mathcal{L}_q$ regression with $\rho(x) = |x|^q$, $1 < q < 2$, the Huber's estimate with $\rho(x) = x^2\mathbf{1}(|x| \le \varsigma)/2 + (\varsigma|x| - \varsigma^2/2)\mathbf{1}(|x| > \varsigma)$, $\varsigma > 0$ and the least squares estimate with $\rho(x) = x^2$.

The purpose of this paper is to provide a theoretical foundation as well as a unified methodological tool for the inference of (1) with a wide class of non-stationary predictor and error processes. For brevity and clarity, we will focus on the structural change detection problem for model (1) throughout this article. Various other results such as confidence region construction and goodness of fit tests can be easily established as corollaries of those provided in this paper. The most significant contributions of the paper lie in the following two aspects. First, we investigate the behaviors of a wide class of stochastic M-estimators and their residual processes under a general nonlinear and non-stationary time series framework with a very flexible modeling of the relationship between the regressors and errors. On one hand, following [32], we allow the regressors $\{\mathbf{x}_i\}$ and the errors $\{e_i\}$ to experience both smooth and abrupt nonlinear changes in their marginal distributions as well as dependence structures over time. Such nonlinear and non-stationary modeling of the regressors and errors could be realistic and flexible in many time series applications; see for instance the Hong Kong public health time series analyzed in Section 5. On the other hand, by carefully choosing the filtration (information) that generates the predictor and error processes, we are able to provide a unified treatment for a wide class of predictor-error relationships, including exogenous, state-heteroscedastic and autoregressive regressions and their mixtures. Here "state-heteroscedasticity" refers to probabilistic dependence between the errors and covariates. Equivalently, it represents that, conditional on the covariates, the distribution of the error at any fixed time changes with respect to different levels of the covariates. Under the aforementioned settings, we are able to establish a uniform Bahadur representation of the partial sample M-estimators with nearly optimal approximation rates and derive the limiting behaviors of a gradient-based structural change test. Our theoretical development depends heavily on investigating the conditional empirical processes of M-estimators of dependent and heteroscedastic data. In particular, both martingale and conditional chaining techniques are used to investigate the maximum stochastic oscillations of the conditional gradient processes. Then the maximum stochastic oscillations of the unconditional empirical processes are recovered by certain integration techniques. To our knowledge, this paper provides the first theoretical investigation into general stochastic M-estimations under time series non-stationarity.

Second, we propose in this paper a unified bootstrap methodology which is consistent for structural change tests of a wide class of M-estimations under both abruptly and smoothly time-varying temporal dynamics and predictor-error dependence. To our knowledge, there have been no methodological results on structural change tests for time series M-estimation with non-stationary covariates and errors in the literature. For change point tests of the mean, [32] 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 gradient change point tests for M-estimations. In particular, a naive extension of [32] by progressively convoluting the block gradient vectors and i.i.d. standard normals will not yield a consistent test. Specifically, note that the key to a successful bootstrap is to mimic the behavior of the estimated gradient cumulative sum (CUSUM) process

$$\left\{\sum_{i=1}^{j} \psi(\hat{e}_i)\mathbf{x}_i/\sqrt{n}\right\}_{j=1}^{n}, \text{ where } \hat{e}_i = y_i - \mathbf{x}_i'\hat{\beta}_n \tag{3}$$

are the residuals of the M-estimation. The bootstrap in [32] mimics the latter process by $\{\sum_{i=1}^{j}[\sum_{k=i}^{i+m} \psi(\hat{e}_k)\mathbf{x}_k/\sqrt{nm}]V_i\}_{j=1}^{n-m}$ where m is a user-chosen block size and V_i's are standard normals that are independent of the the data. If the true errors $\{e_i\}$ were known and the residuals in (3) are replaced by the true errors, then it can be shown that the above bootstrap consistently mimics (3). However, due to the non-negligible differences between the true errors and the residuals, it is shown that the limiting behaviors of (3) and the above bootstrap process differ and hence the bootstrap is inconsistent.

In this paper, we extend and modify the procedure in [32] and propose an easy-to-implement bootstrap methodology by combining an extension of the Powell's sandwich estimates ([24]) and a progressive convolution of the block sums of the estimated gradient vectors with i.i.d. standard normal auxiliary random variables. The bootstrap is shown to be consistent for a wide class of M-estimation under non-stationary temporal dependence and predictor-error dependence. The bootstrap procedure enjoys the asymptotic optimal property that it approaches the covariance structure of the target limiting Gaussian process no slower than the nearly optimal approximation rate of the Bahadur representation in various important cases such as the quantile regression. Meanwhile, we prove that our bootstrap can detect local alternatives with the optimal $1/\sqrt{n}$ parametric rate. Furthermore, our simulation studies indicate that the gradient-based method has a superior finite sample power performance than regression-coefficient-based structural stability tests under time series non-stationarity.

There is a large amount of work in testing structural stability of parameters for general M-estimation and special cases such as least squares and quantile regressions. It is impossible to gather a complete list here and we shall only mention some representative works. For least squares regression, [9], [20] developed CUSUM tests with i.i.d. normal errors. [21] extended such tests to stationary and ergodic errors. [1] 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. There are also a class of tests which heavily depend on the residuals of the least squares regression. For example, [6] obtained asymptotically distribution free test statistics associate with i.i.d. errors; see also [7] for tests of multiple structural changes. Recently, [15] and [27] investigated structural change detection for least squares regression when covariates and errors are non-stationary. For quantile regression, traditionally when dealing with stationary data, the regression coefficient CUSUM test is shown to be asymptotically pivotal ([26]) and the gradient CUSUM test is advocated over the regression coefficient test as it is asymptotically free of the densities of the errors ([16], [26]). For general M-estimation, [25] investigated the gradient-based change point tests for stationary predictors and errors with predictor-error independence. We also refer to the recent review of [2] for more discussions and references. A testing strategy that prevails throughout most of the above mentioned papers is to pivotalize the test statistics. As a result, functionals of some well investigated processes, such as the Brownian bridge, can be used to approximate the large sample behaviors of the tests. However, due to the non-stationarity of the predictors and errors considered in this paper, it is shown that the gradient CUSUM process behaves complexly over time and it cannot be pivotalized in general. As a result, in general, the classic testing procedures based on the idea of pivotalization are not consistent for structural change tests under non-stationarity. We also refer to [32] for a detailed discussion in the case of testing structural changes in mean.

The rest of the paper is organized as follows. In Section 2 we will introduce the non-stationary time series models for the predictors and errors with multiple illustrative examples. A uniform Bahadur representation and related asymptotic results are established for general M-estimation under non-stationarity and temporal dependence. Section 3 proposes the structural change tests and the bootstrap and investigates their asymptotic Type I error and power behaviors. In Section 4, we perform Monte Carlo experiments to study the finite sample behaviors of the gradient-based test. Section 5 contains an empirical illustration using a public health time series.

All technical proofs are relegated to the online appendix.

2. M-estimation Under Time Series Non-stationarity. We first introduce some notations. Define $X_n \geq_p Y_n$ if $\mathbb{P}(X_n \geq Y_n) \to 1$ as $n \to \infty$. Similarly define "$\leq_p$". For a p-dimensional vector $\mathbf{v} = (v_1, ..., v_p)'$, let $|\mathbf{v}| = \sqrt{\sum_{i=1}^p v_i^2}$. For an $m \times n$ matrix A, define $|A| = \sqrt{trace(AA')}$. For a random variable X, let $\|X\|_q := (\mathbb{E}|X|^q)^{1/q}$ be its $\mathcal{L}_q$ norm. For a semi-positive definite matrix Σ, let $\lambda_1(\Sigma)$ be its smallest eigenvalue. For a p-dimensional random vector $\mathbf{v}$, let $\|\mathbf{v}\|_q = \||\mathbf{v}|\|_q$. Write $\mathbf{v} \in \mathcal{L}_q$ if $\|\mathbf{v}\|_q < \infty$. For an $m \times n$ random matrix A, define $\|A\|_q = \||A|\|_q$. Write $\|\cdot\| := \|\cdot\|_2$. Let $\mathcal{F}_i = (..., \eta_{i-1}, \eta_i)$ and $\mathcal{F}_i^{(j)} = (..., \eta_{j-1}, \eta_j', \eta_{j+1}, ..., \eta_i)$ for $j \leq i$, where $(\{\eta_i\}_{i=-\infty}^{\infty}, \{\eta_j'\}_{j=-\infty}^{\infty})$ are i.i.d. random variables. Write $\mathcal{F}_i^*$ for $\mathcal{F}_i^{(0)}$. For $x \in \mathbb{R}$, let $\lfloor x \rfloor = \max\{k \in \mathbb{Z}, k \leq x\}$, and $\lceil x \rceil = \min\{k \in \mathbb{Z} : k \geq x\}$. Write $N = \lfloor \frac{n}{\log n} \rfloor$ for short. Let "$\Rightarrow$" denote the convergence in distribution. Throughout the paper, let $\chi \in (0, 1)$ be a constant which may vary from case to case, and M be a sufficiently large constant which may vary from line to line. Let $\mathbf{1}(\cdot)$ be the usual indicator function. Let $\psi(u; \epsilon) = |\psi(u + \epsilon)| + |\psi(u - \epsilon)|$, where $\psi(\cdot)$ is the left derivative of $\rho(\cdot)$, the loss function of the corresponding M-estimator. For any function $f(x), x \in \mathbb{R}$ and an open interval I, we write $f(x) \in \mathcal{C}^l(I)$ if the l_{th} derivative of f is continuous on I. Write $a \vee b$ for $\max\{a, b\}$, and $a \wedge b$ for $\min\{a, b\}$.

2.1. *Non-stationary Time Series Models.* In order to model the complex temporal dynamics of the covariate and error processes, we introduce the following class of piecewise locally stationary (PLS) time series ([32]):

DEFINITION 2.1. *For $k < \infty$, we say that $\{e_i\}_{i=1}^n$ is a PLS process generated by filtrations $\mathcal{F}_{1,i}$, $\mathcal{F}_{2,i}$,..., $\mathcal{F}_{k,i}$ with r breaks (PLS(r, $\mathcal{F}_{1,i}$, $\mathcal{F}_{2,i}$,..., $\mathcal{F}_{k,i}$)) 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*

$$e_i = G_j(t_i, \mathcal{F}_{1,i}, ..., \mathcal{F}_{k,i}), \text{ if } b_j < t_i \leq b_{j+1}, \tag{4}$$

where $t_i = i/n$, $\mathcal{F}_{l,i} = \{..., \varepsilon_{l,0}, \varepsilon_{l,1}, ..., \varepsilon_{l,i}\}$ for $1 \leq l \leq k$. For each l, $\{\varepsilon_{l,i}\}_{i=-\infty}^{\infty}$ are i.i.d. r.v's. For $l \neq s$, $\{\varepsilon_{l,i}\}_{i=-\infty}^{\infty}$ and $\{\varepsilon_{s,i}\}_{i=-\infty}^{\infty}$ are independent.

In Definition 2.1, 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. The PLS class is appropriate to describe stochastic temporal systems which experience occasional structural breaks and otherwise evolve smoothly over time. In this paper, we model both the covariate and error processes as PLS series to capture their complexly time-varying behaviors. To quantify the temporal dependence of PLS processes, we shall introduce the following dependence measure:

DEFINITION 2.2. *Consider the $PLS(r, \mathcal{F}_{1,i}, ..., \mathcal{F}_{k,i})$ process $\{e_i\}_{i=-\infty}^{\infty}$ defined in (4). Assume $\max_{1\le i\le n} \|e_i\|_p < \infty$ for some $p > 0$. Then we define $\Delta_p(G, l)$, the l_{th} dependence measure for $\{e_i\}_{i=-\infty}^{\infty}$ in $\mathcal{L}_p$ norm as*

$$\Delta_p(G, l) := \max_{0\le j\le r} \sup_{b_j<t\le b_{j+1}} \|G_j(t, \mathcal{F}_{1,l}, ..., \mathcal{F}_{k,l}) - G_j(t, \mathcal{F}^*_{1,l}, ..., \mathcal{F}^*_{k,l})\|_p.$$

Note that $\Delta_p(G, l) = 0$ for $l < 0$. If we view e_i as the output of a physical system which is driven by innovations $\{\varepsilon_{s,i}\}_{i=-\infty}^{\infty}$, $s = 1, ..., k$, then $\Delta_p(G, l)$ measures the contribution of the innovations l steps ahead in generating the current observation of the system, via replacing them with i.i.d. copies and measuring the magnitude of change in the output of the system. The measure $\Delta_p(G, l)$ for a broad class of classic time series models can be calculated, *e.g.*, invertible ARMA process, (G)arch models ([12], [8]) and threshold models ([28]). We refer to [32] for more details about PLS models and their dependence measures.

Throughout this paper, we assume that

$$e_i = G_k(i/n, \mathcal{F}_i, \mathcal{G}_i)$$

if $b_k < i/n \le b_{k+1}$ with break points $0 = b_0 < b_1 < \cdots < b_r < b_{r+1} = 1$. Define $w(i) = j$ if $b_j < i/n \le b_{j+1}$. We formulate the covariates as (where we fix $H_{k,1} \equiv 1$ below for the intercept)

$$\text{(5)}\ \mathbf{x}_i = \mathbf{H}_k(i/n, \mathcal{F}_{i-1}, \mathcal{G}_i) := (H_{k,1}(i/n, \mathcal{F}_{i-1}, \mathcal{G}_i), ..., H_{k,p}(i/n, \mathcal{F}_{i-1}, \mathcal{G}_i))'$$

for $d_k < i/n \le d_{k+1}$, where $d_0 = 0 < d_1 < .. < d_s < d_{s+1} = 1$ are break points of $\{\mathbf{x}_i\}$. Here the filtrations $\mathcal{G}_i$ and $\mathcal{F}_i$ are generated by $(..., \eta_{i-1}, \eta_i)$ and $(..., \varepsilon_{i-1}, \varepsilon_i)$, respectively, where $\{\eta_i\}_{i=-\infty}^{\infty}$ and $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ are independent. Observe that the errors $\{e_i\}_{i=1}^{n}$ and the covariates $\{\mathbf{x}_i\}_{i=1}^{n}$ are allowed to be dependent as they are generated by common filtrations $\{\mathcal{G}_i\}_{i\in\mathbb{Z}}$ and $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$.

The above formulation of the error and covariate processes contains a wide range of state-heteroscedastic, exogenous and autoregressive linear regression models used in practice. The introduction of $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ in the covariates and errors is to accommodate autoregressive-type models where the covariates at time i contain response information up to time $i-1$. On the other hand, we introduce filtrations $\mathcal{G}_i$ in the definitions of both e_i and $\mathbf{x}_i$ to accommodate any extra information which could influence the covariates or errors. In particular, if the generating mechanisms of $\{e_i\}_{i=1}^n$ and $\mathbf{x}_i$ are functionally independent of $\mathcal{G}_i$ and $\mathcal{F}_{i-1}$ respectively, then we obtain a purely exogenous model where the covariates and errors are independent. Below we list two other frequently used subclasses of the above formulation.

EXAMPLE 1. *Consider the following heteroscedastic error model:*

$$y_i = \mathbf{x}_i'\beta + s(\mathbf{x}_i)\eta_i,$$

where $s(\cdot)$ *is a piecewise smooth function,* $\{\eta_i\}_{i=1}^n$ *is PLS*$(r, \mathcal{F}_i)$ *and* $\{\mathbf{x}_i\}_{i=1}^n$ *is PLS*$(s, \mathcal{G}_i)$*. Furthermore the filtrations* $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$ *and* $\{\mathcal{G}_i\}_{i\in\mathbb{Z}}$ *are independent. Note that* $e_i = s(\mathbf{x}_i)\eta_i$ *can be written as a PLS process generated by* $(\mathcal{F}_i, \mathcal{G}_i)$*. Lack of fit tests in regression quantiles of the above heteroscedastic error model with* $\{\eta_i\}_{i=1}^n$ *i.i.d. are investigated in [16], among others.*

EXAMPLE 2. *Consider the following autoregressive model:*

$$y_i = \mathbf{x}_i'\beta + e_i,$$

where $\mathbf{x}_i = (y_{i-1}, ..., y_{i-p})'$, $\beta = (\beta_1, ..., \beta_p)'$, $\sum_{j=1}^p \beta_j z^j \neq 1$ *for all* $|z| \leq 1+c$ *with some constant* $c > 0$ *and* $\{e_i\}$ *is PLS*$(r, \mathcal{F}_i)$*. Note that* $\mathbf{x}_i$ *is a PLS process generated by* $\mathcal{F}_{i-1}$*.*

2.2. *Asymptotic Theory for M-estimation.* Consider model (1). In this paper, we focus on robust loss functions in the sense that

$$|\psi(x) - \psi(y)| \leq M_1 + M_2|x-y| \tag{6}$$

for all $x, y \in \mathbb{R}$ and some positive constants M_1 and M_2. It is easy to check that the left derivatives of the loss functions of quantile, expectile, $\mathcal{L}_q$ for $1 < q < 2$, least squares and the Huber regressions all satisfy (6).

The asymptotic behavior of the M-estimator $\hat{\beta}_n$ in (2) was investigated by numerous researchers. Among them, for quantile regression with i.i.d. error, [5] approximated $\sqrt{n}(\hat{\beta}_n - \beta)$ by linear forms and [18] showed that the remaining term of the approximation is of order $O_{a.s}(n^{-1/4}(\log\log n)^{3/4})$. [4] obtained asymptotic results for strong mixing errors. [23] acquired asymptotic approximations of $\sqrt{n}(\hat{\beta}_n - \beta)$ when the errors are "m-decomposable". With physical dependence measures, [29] obtained Bahadur representation

for models with fixed design and stationary errors. As a first contribution of this paper, we obtain a Bahadur representation with nearly optimal rate (except a multiplicative logarithm factor) for model (1) with a wide class of PLS errors and regressors. The imposed conditions are mild and can be checked easily; see Proposition 2.1 and below.
For $q \geq 0$, $j = 0, 1, \cdots, r$, define for $t \in (b_j, b_{j+1}]$,

$$\begin{aligned}
&\Xi_j^{(q)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = \frac{\partial^q}{\partial x^q}\mathbb{E}\{\psi(G_j(t, \mathcal{F}_k, \mathcal{G}_k) + x)|\mathcal{F}_{k-1}, \mathcal{G}_k\},\\
&\bar{\kappa}_j(t, x, \mathbf{x}_i) = \frac{\partial}{\partial x}\mathbb{E}(\psi(G_j(t, \mathcal{F}_i, \mathcal{G}_i) + x)|\mathbf{x}_i),\\
&F_j^{(q)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = \frac{\partial^q}{\partial x^q}\mathbb{E}\{\mathbf{1}(G_j(t, \mathcal{F}_k, \mathcal{G}_k) \leq x)|\mathcal{F}_{k-1}, \mathcal{G}_k\},\\
&F_j^{(q)}(t, x|\mathbf{x}_k) = \frac{\partial^q}{\partial x^q}\mathbb{E}\{\mathbf{1}(G_j(t, \mathcal{F}_k, \mathcal{G}_k) \leq x)|\mathbf{x}_k\},\\
&f_j(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = F_j^{(1)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k), f_j(t, x|\mathbf{x}_k) = F_j^{(1)}(t, x|\mathbf{x}_k).
\end{aligned}$$

Note that for integer $q \geq 0$, $f_j^{(q)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = F_j^{(q+1)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k)$. For τ_{th} quantile regression, $F_j^{(q)}(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = \tau - \Xi_j^{(q)}(t, -x|\mathcal{F}_{k-1}, \mathcal{G}_k)$. Also by (5), $\bar{\kappa}_j(t, x, \mathbf{x}_i) = \mathbb{E}(\Xi_j^{(1)}(t, x|\mathcal{F}_{i-1}, \mathcal{G}_i)|\mathbf{x}_i)$. Omit the superscript q if $q = 0$. The following regularity conditions are needed for the covariate and error processes:

(S0) Assume that $\max_{0 \leq i \leq r} \sup_{b_i < s < t \leq b_{i+1}} \|\frac{G_i(t, \mathcal{F}_0, \mathcal{G}_0) - G_i(s, \mathcal{F}_0, \mathcal{G}_0)}{t-s}\|_v \leq M$ for some constant $v > 1$. The dependence measure of e_i in $\mathcal{L}_v$ norm, $\Delta_v(G, k)$, satisfies $\Delta_v(G, k) = O(\chi^k)$. Assume $v = 4(p+1) \vee 20$ unless otherwise specified.

(S1) Define $\bar{\Xi}^{(q)}(x|\mathbf{x}_i) = \frac{\partial^q}{\partial x^q}\mathbb{E}(\psi(e_i + x)|\mathbf{x}_i)$. We require that for all $i = 1, 2, ..., n$ and any p-dimensional vector g,

$$\begin{aligned}
&\mathbb{E}(\psi(e_i)|\mathbf{x}_i) = \bar{\Xi}(0|\mathbf{x}_i) = 0 \ a.s.,\\
&\mathbb{E}(\bar{\Xi}(\mathbf{x}_i'\delta|\mathbf{x}_i)\mathbf{x}_i'g) = \mathbb{E}(\bar{\Xi}^{(1)}(0|\mathbf{x}_i)\mathbf{x}_i'\delta\mathbf{x}_i'g) + O(|\delta|^2) \qquad (7)
\end{aligned}$$

and $\bar{\Xi}^{(1)}(x|\mathbf{x}_i) > 0$ a.s. for $|x| \leq \varepsilon$ for some $\varepsilon > 0$.
Define

$$\nu_i(\delta) = \mathbb{E}\{[\psi(e_i + |\mathbf{x}_i||\delta|) - \psi(e_i - |\mathbf{x}_i||\delta|)]^2|\mathbf{x}_i|^2\}.$$

Assume that $\nu_i(t)$ is continuous at $t = 0$ and that for $0 \leq j \leq r$,

$$\begin{aligned}
&\sup_{t \in (b_j, b_{j+1}], x, y \in \mathbb{R}} |\Xi_j(t, x|\mathcal{F}_{-1}, \mathcal{G}_0) - \Xi_j(t, y|\mathcal{F}_{-1}, \mathcal{G}_0)|\\
&\leq C_{j,1}|x - y| + C_{j,2}|x^2 - y^2| + C_{j,3}|x - y|^2, \qquad (8)
\end{aligned}$$

where $C_{j,1}, C_{j,2}, C_{j,3} \in \mathcal{L}_4$ are $(\mathcal{F}_{-1}, \mathcal{G}_0)$ measurable random variables. We also require that for $0 \le j \le r$, $\Xi_j^{(1)}(t, 0|\mathcal{F}_{-1}, \mathcal{G}_0)$ is stochastically Lipschitz continuous for $t \in (b_j, b_{j+1}]$, i.e., $\exists M < \infty$, s.t. $\forall\ t_1, t_2 \in (b_j, b_{j+1}]$, $0 \le j \le r$,

$$\|\Xi_j^{(1)}(t_1, 0|\mathcal{F}_{-1}, \mathcal{G}_0) - \Xi_j^{(1)}(t_2, 0|\mathcal{F}_{-1}, \mathcal{G}_0)\| \le M|t_1 - t_2|. \tag{9}$$

(S2) For covariates process, assume that $\Delta_v(\mathbf{H}, k) = O(\chi^k)$ and $\max_{1 \le i \le n} \|\mathbf{x}_i\|_{5p+10} \le M$. For all $0 \le k \le s$ and all $t_1, t_2 \in (d_k, d_{k+1}]$, assume that $\|\mathbf{H}_k(t_1, \mathcal{F}_{-1}, \mathcal{G}_0) - \mathbf{H}_k(t_2, \mathcal{F}_{-1}, \mathcal{G}_0)\|_v \le M|t_1 - t_2|$ for v defined in (S0).

A few comments on the above regularity conditions are in order. Condition (S0) (resp. (S2)) requires that the process e_i (resp. $\mathbf{x}_i$) to be short range dependent with exponentially decaying dependence measures. Furthermore, (S0) (resp. (S2)) requires that the data generating mechanisms of e_i (resp. $\mathbf{x}_i$) to be smooth between adjacent break points by posting certain piecewise stochastically Lipschitz continuous constraints. The assumption that $v = 4(p + 1) \vee 20$ in (S0) guarantees that the process $\{\psi(e_i)\mathbf{x}_i\}_{i=1}^n$ is a stochastically Hölder continuous PLS process with order higher than $1/4$. We point out that when $\psi(\cdot)$ is bounded or light-tailed, the moment requirements in (S0) and (S2) can be significantly relaxed. Our simulation results also show that our method works well under less restrictive moment conditions. However, for simplicity of presentation, we will omit the separate discussions and use (S0) and (S2) throughout this paper. (S2) also implies that $\max_{1 \le i \le n} |\mathbf{x}_i| = O_p(n^{1/(5p+10)})$.

Assumption (S1) is necessary for the consistency of $\hat{\beta}_n$. Since $\psi(\cdot)$ is monotone, by Cauchy's inequality, the dominated convergence theorem and (6), the continuity of $\nu_i(t)$ at $t = 0$ is satisfied whenever $e_i, 1 \le i \le n$ have continuous distribution functions. (8) holds if $\sup_u |\psi^{(1)}(u)| < \infty$, thus it holds for least squares regression. For quantile regression, (8) holds if $\max_{0 \le j \le r} \sup_{t \in (b_j, b_{j+1}], x \in \mathbb{R}} |f_j(t, x|\mathcal{F}_{-1}, \mathcal{G}_0)| < \infty$. In general, a sufficient condition for (8) is (6) with condition (A1) below, which we show in Proposition A.3 of the supplemental material.

In addition, (9) in (S1) is required for the existence of quantity $\Lambda(s)$; see equation (21) of the paper. A sufficient condition for (9) is similarly

$$\max_{0 \le j \le r} \sup_{t \in (b_j, b_{j+1}], x \in \mathbb{R}} \int \left\| \frac{\partial}{\partial t} f_j^{(1)}(t, x|\mathcal{F}_{-1}, \mathcal{G}_0) \right\| |\psi(x)| dx < \infty.$$

Finally, (7) is implied by condition (10) in (A1), (S2) and (6).

For $t \in (b_j, b_{j+1}]$, $0 \le j \le r$, $\epsilon \in \mathbb{R}$, define

$$v_j^{(q)}(t,\epsilon) = \int (\psi(x;\epsilon)+1)\|f_j^{(q)}(t,x|\mathcal{F}_{-1},\mathcal{G}_0)\|_4 dx,$$

$$w_j^{(q)}(t,k,\epsilon) = \int (\psi(x;\epsilon)+1)\|f_j^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k) - f_j^{(q)}(t,x|\mathcal{F}^*_{k-1},\mathcal{G}^*_k)\|_4 dx.$$

We need following additional conditions for the Bahadur representations:

(A0) Let $\underline{\lambda}_n^a$ be the smallest eigenvalue of $\mathbb{E}\{\sum_{i=1}^{\lfloor an \rfloor} \Xi^{(1)}(0|\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'/a\}$ for any $a \in (0,1]$. Assume that $\forall s \in (0,1]$, $\liminf_{n\to\infty} \underline{\lambda}_n^s/n > 0$, and $\liminf_{n\to\infty, s\to 0^+, ns\to\infty} \underline{\lambda}_n^s/n \in (0,+\infty)$.

(A1) There exist constants ϵ_0 and M such that for $0 \le q \le p+1$, $k \in \mathbb{N}$,

$$\max_{0\le j\le r} \sup_{t\in(b_j,b_{j+1}],|\epsilon|\le\epsilon_0} v_j^{(q)}(t,\epsilon) \le M, \tag{10}$$

$$\max_{0\le j\le r} \sup_{t\in(b_j,b_{j+1}],|\epsilon|\le\epsilon_0} w_j^{(q)}(t,k,\epsilon) = O(\chi^k). \tag{11}$$

Furthermore, for quantile regression where $\psi(x) = \tau - \mathbf{1}(x \le 0)$, we assume the following condition (A1*) instead of (A1):

(A1*) There exists some constant ϵ_0 s.t. for $0 \le q \le p$, $k \in \mathbb{N}$, $0 \le j \le s$,

$$\begin{aligned}\sup_{t\in(d_j,d_{j+1}],|u|\le\epsilon_0} &\|F^{(q)}_{\delta(t)}(t,\mathbf{H}_j(t,\mathcal{F}_{k-1},\mathcal{G}_k)'u|\mathcal{F}_{k-1},\mathcal{G}_k) - \\ &F^{(q)}_{\delta(t)}(t,\mathbf{H}_j(t,\mathcal{F}^*_{k-1},\mathcal{G}^*_k)'u|\mathcal{F}^*_{k-1},\mathcal{G}^*_k)\|_4 = O(\chi^k),\end{aligned} \tag{12}$$

where $\delta(t) = l$ if $b_l < t \le b_{l+1}$, $0 \le l \le r$.

By definition, the right hand sides of (11) and (12) will be excatly 0 when $k < 0$. Condition (A0) guarantees the consistency of $\{\hat\beta_j\}_{j=N}^n$ where $\hat\beta_j$ is the M-estimation coefficient using $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_j, y_j)$. It is actually quite mild. By condition (S1) and Weyl inequality, if there exists an $\epsilon > 0$ such that

$$\min_{0\le k\le s} \inf_{t\in(d_k,d_{k+1}]} \lambda_1(\mathbb{E}\{\mathbf{H}_k(t,\mathcal{F}_{-1},\mathcal{G}_0)\mathbf{H}_k'(t,\mathcal{F}_{-1},\mathcal{G}_0)\}) \ge \epsilon,$$

then (A0) is fulfilled. In other words, we only require that the matrices $\mathbb{E}\{\mathbf{H}_k(t,\mathcal{F}_{-1},\mathcal{G}_0)\mathbf{H}_k'(t,\mathcal{F}_{-1},\mathcal{G}_0)\}$, $0 \le k \le s$, $t \in (d_k, d_{k+1}]$ are not degenerate. For (A1) and (A1*), (10) requires the (differentiated) conditional densities $f_j^{(q)}(t,x|\mathcal{F}_{-1},\mathcal{G}_0)$ to be sufficiently light-tailed with respect to x. Meanwhile, (11) and (12) are short-range-dependent conditions for the processes

$\{F_j^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)\}_{k=1}^n$. The following four examples show that conditions (A1) and (A1*) can be verified for a wide range of non-stationary time series M-estimation.

EXAMPLE 3. *(Quantile regression for PLS linear processes) Suppose we have the following PLS linear time series:*

$$(13)\ G_k(t,\mathcal{F}_i,\mathcal{G}_i)=\sum_{j=0}^{\infty}a_{k,j}(t)\varepsilon_{i-j}h_k(t,\mathcal{G}_{i-j}),\quad b_k<t\le b_{k+1},\ 0\le k\le r$$

where $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ *are i.i.d. mean 0 r.v's with* $\mathbb{E}|\varepsilon_0|^{4u}<\infty$ *for some* $u>1$. *In addition,* $h_k(t,\cdot)$ *and* $a_{j,k}(\cdot)$ *are piecewise (stochastically) Lipschitz continuous functions. Without loss of generality, let* $a_{k,0}(t)\equiv 1$. *Assume that* $\sup_{x\in\mathbb{R}}|f_\varepsilon^{(l)}(x)|\le M<\infty$ *for* $0\le l\le p$ *where* $f_\varepsilon(x)$ *is the density function of* ε_0. *Write* $w=u/(u-1)$. *We have the following proposition.*

PROPOSITION 2.1. *Assume that i): there exists an* $\eta>0$ *such that* $|h_k(t,\cdot)|\ge\eta$, *ii): for* $0\le k\le r$, $t\in(b_k,b_{k+1}]$, $\|h_k(t,\mathcal{G}_0)\|_{4w}\le C$, $\|h_k(t,\mathcal{G}_i)-h_k(t,\mathcal{G}_i^*)\|_{4w}=O(\chi^i)$, $a_{k,j}(t)=O(\chi^j)$ *and iii):*

$$\max_{0\le j\le s}\sup_{t\in(d_j,d_{j+1}]}\|\mathbf{H}_k(t,\mathcal{F}_{-1},\mathcal{G}_0)\|_{4u}<\infty,\Delta_4(\mathbf{H},i)=O(\chi^i).$$

Then (A1) holds.*

As a side note, for PLS linear model (13), (S0) holdes if we further assume

$$\mathbb{E}|\varepsilon_0|^{vu}<\infty,\sum_{j=0}^{\infty}\left(\max_{0\le k\le r}\sup_{t\in(b_k,b_{k+1}]}|\dot{a}_{k,j}(t)|\right)<\infty,$$
$$\max_{0\le k\le r}\sup_{t,s\in(b_k,b_{k+1}]}\|h_k(t,\mathcal{G}_0)-h_k(s,\mathcal{G}_0)\|_{vw}\le M|t-s|.$$

EXAMPLE 4. *(General M-estimation for PLS linear processes) Consider general M-estimation with errors following (13). Write* $\theta_\gamma(du)=(1+|u|)^\gamma du$ *and* $\Delta_\gamma=\int\psi^{4/3}(x)\theta_{-\gamma}(dx)$. *Then we have the following proposition.*

PROPOSITION 2.2. *Assume i): for* $0\le k\le r$, $t\in(b_k,b_{k+1}]$, $\|h_k(t,\mathcal{G}_i)-h_k(t,\mathcal{G}_i^*)\|_8=O(\chi^i)$, $h_k(t,\mathcal{G}_0)\ge\eta>0$ *and that* $a_{k,j}(t)=O(\chi^j)$, *ii): there exists a pair of positive numbers* v_1,v_2, $v_1^{-1}+v_2^{-1}=1$, *such that* $\varepsilon_0\in\mathcal{L}_{v_1(3\gamma\vee 4)}$, $\max_{0\le k\le r}\sup_{b_k<t\le b_{k+1}}\|h_k(t,\mathcal{G}_0)\|_{v_2(3\gamma\vee 4)}<\infty$, *iii): there exists*

$\gamma > 1$ s.t. $\Delta_\gamma < \infty$ and iv):

$$\sum_{q=0}^{p+2} \int (f_\varepsilon^{(q)}(u))^4 \theta_{3\gamma}(du) < \infty, \ \sum_{q=0}^{p+2} \int (f_\varepsilon^{(q)}(u)u)^4 \theta_{3\gamma}(du) < \infty, \tag{14}$$

where the second inequality of (14) can be removed if for $0 \le j \le r$, $b_j < t \le b_{j+1}$, $h_j(t, \mathcal{G}_0) \equiv 1$. Then we have that condition (A1) holds.

EXAMPLE 5. *(Quantile regression for PLS nonlinear processes) Suppose the errors are generated from the following nonlinear system:*

$$G_k(t, \mathcal{F}_i, \mathcal{G}_i) = R_k(t, G_k(t, \mathcal{F}_{i-1}, \mathcal{G}_{i-1}), \varepsilon_i, \eta_i) \tag{15}$$

for $b_k < t \le b_{k+1}$, $0 \le k \le r$. Assume that $\max_{0\le k\le r} \|R_k(t, x_0, \varepsilon_0, \eta_0)\|_v < \infty$ for some x_0 where v is defined in (S0). Let

$$\chi_0 = \max_{0\le k\le r} \sup_{x\neq y, t\in(b_k,b_{k+1}]} \frac{\|R_k(t, x, \varepsilon_0, \eta_0) - R_k(t, y, \varepsilon_0, \eta_0)\|_v}{|x-y|}.$$

The above formulation offers natural extensions of many frequently used stationary nonlinear time series models, e.g. (G)ARCH models ([12]; [8]), threshold models ([28]) and bilinear models, into the non-stationary realm. Write $\bar{F}_k(t, x, s, u) = \mathbb{P}(R_k(t, s, \varepsilon_i, u) \le x)$. For quantile regression with errors following (15), we have the following proposition.

PROPOSITION 2.3. *Assume that*
i): $0 < \chi_0 < 1$,
ii): $\max_{0\le k\le r} \sup_{t\in(b_k,b_{k+1}]} \|M(G_k(t, \mathcal{F}_0, \mathcal{G}_0))\|_v < \infty$, where

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

iii): for $0 \le q \le p$,

$$\max_{0\le k\le r} \sup_{t\in(b_k,b_{k+1}], x,s\in\mathbb{R}} \left| \frac{\partial^q}{\partial x^q} (\frac{\partial}{\partial s} + \frac{\partial}{\partial x}) \bar{F}_k(t, x, s, \eta_0) \right| \le M \ a.s..$$

Then (15) admits a unique solution for each integer $k \in [0, r]$ and the associated $t \in (b_k, b_{k+1}]$. Furthermore, assume $\Delta_4(\mathbf{H}, i) = O(\chi^i)$. Then (S0), (A1) hold.*

EXAMPLE 6. *(General M-estimation for PLS nonlinear processes) Assume that for $t \in (b_k, b_{k+1}]$, $0 \le k \le r$, and ε_i i.i.d. with density $f_\varepsilon(x)$,*

$$G_k(t, \mathcal{F}_i, \mathcal{G}_i) = \nu_k(t, G_k(t, \mathcal{F}_{i-1}, \mathcal{G}_{i-1}), \eta_i) + \varepsilon_i. \tag{16}$$

Consider general M-estimation with errors satisfying (16). Recall the definition of Δ_γ and $\theta_\gamma(du)$ in Example 4. We have the following proposition.

PROPOSITION 2.4. *Assume i): $\|\nu_k(t, x, \eta_0) - \nu_k(t, y, \eta_0)\|_{6\gamma\vee 8} \le \chi_0 |x - y|$ for some $0 < \chi_0 < 1$ and for $t \in (b_k, b_{k+1}]$, $0 \le k \le r$; ii): there exists some x_0 such that $\max_{0\le k\le r} \sup_{t\in(b_k,b_{k+1}]} \|\nu_k(t, x_0, \eta_0)\|_{6\gamma\vee 8} < \infty$ and iii): $\Delta_\gamma < \infty$ for some $\gamma > 1$, and that*

$$\sum_{q=0}^{p+2} \int (f_\varepsilon^{(q)}(u))^4 \theta_{3\gamma}(du) < \infty.$$

Then we have that condition (A1) holds and $\Delta_{6\gamma\vee 8}(G, l) = O(\chi_0^l)$ when errors follow (16).

For the rest of the paper, we assume that for $1 \le i \le n$,

$$\nu_i(t) = O(|t|^\eta) \text{ for some } \eta > 0,$$

as it is satisfied by all quantile, expectile, least squares, the Huber and robust $\mathcal{L}_q$ regressions. Without the assumption, via the same techniques used in this paper, all the results can still be established but with more factors of logarithms involved. Write $d_{1,n} = n^{-1/2} \log n$, $d_{2,n} = n^{-1/2} (\log n)^2$ for short.

LEMMA 2.1. *Assume (S0)-(S2), (A0) and (A1) (or (A1*) for quantile regression). Then we have i): $|\hat\beta_n - \beta| \le_p d_{1,n}$, ii): $\max_{N\le j\le n} |\hat\beta_j - \beta| \le_p d_{2,n}$.*

Result i) shows that $\hat\beta_n$ is weakly consistent. Result ii) establishes the uniform consistency of $\hat\beta_j$ estimated by different sub-samples with at least N observations. The consistency results are needed for the structural stability test in Section 3. The following theorem establishes an important uniform Bahadur representation for a wide class of non-stationary time series M-estimation.

THEOREM 2.1. *Write $\acute\Lambda(j) = \mathbb{E}\{\sum_{i=1}^{j} \bar\kappa_{w(i)}(i/n, 0, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'/n\}$. Assume that $\max_{0\le j\le r} \sup_{t\in(b_j,b_{j+1}], x\in\mathbb{R}} |f_j(t, x|\mathcal{F}_{-1}, \mathcal{G}_0)| < \infty$. Then under assumptions (S0)-(S2), (A0) and (A1) (or (A1*) for quantile regression), we have* *i):*

$$\sqrt{n}(\hat{\beta}_n - \beta) - (\acute{\Lambda}(n)n^{1/2})^{-1}\sum_{i=1}^{n}\psi(e_i)\mathbf{x}_i$$

$$= O_p\left(\sqrt{\frac{\sum_{i=1}^{n}\nu_i(d_{1,n})}{n}}\log n + r_n\right), \tag{17}$$

and ii):

$$\max_{N\le j\le n}\left|\sqrt{n}(\hat{\beta}_j - \beta) - (\acute{\Lambda}(j)n^{1/2})^{-1}\sum_{i=1}^{j}\psi(e_i)\mathbf{x}_i\right| \tag{18}$$

$$= O_p\left(\sqrt{\frac{\sum_{i=1}^{n}\nu_i(d_{2,n})}{n}}\log n + r_n\right).$$

The quantity r_n equals zero if $\psi(\cdot)$ is continuous, otherwise $r_n = n^{\frac{1}{5p+10}}$.

REMARK 2.1. *(The order of the quantity $\nu_i(\delta)$)*
For quantile regression, if for $i = 1, ..., n$, the conditional densities satisfy

$$\max_{0\le k\le r}\sup_{t\in(b_k,b_{k+1}],x\in\mathbb{R}}|f_k(t,x|\mathbf{x}_i)| \le M_0 < \infty \ a.s., \tag{19}$$

then we have $\nu_i(\delta) = O(|\delta|)$, which results in that $\sqrt{\sum_{i=1}^{n}\nu_i(d_{2,n})} = O(n^{1/4}\log^2 n)$. As in the discussion of Example 1 of [29], for the Huber, $\mathcal{L}_2$ and expectile regressions, there exists u such that for $1 \le i \le n$, $\sup_{|\delta|\le u}\|\psi'(e_i + |\mathbf{x}_i||\delta|)\|_4 \le M < \infty$. Hence $\nu_i(\delta) = O(\delta^2)$, and consequently $\sqrt{\sum_{i=1}^{n}\nu_i(d_{2,n})} = O(\log^2 n)$. For robust $\mathcal{L}_q$ regression, noting that $\nu_i(t) \le 2(\nu_i^+(t) + \nu_i^-(t))$, where

$$\nu_i^+(\delta) = \mathbb{E}\{[\psi(e_i + |\mathbf{x}_i||\delta|) - \psi(e_i)]^2|\mathbf{x}_i|^2\},$$
$$\nu_i^-(\delta) = \mathbb{E}\{[\psi(e_i) - \psi(e_i - |\mathbf{x}_i||\delta|)]^2|\mathbf{x}_i|^2\}.$$

By Lemma 4 of [3], we have that for $3/2 < q < 2$,

$$\nu_i^+(\delta) \le 36\delta^2\mathbb{E}(|\mathbf{x}_i|^4|e_i|^{2q-4}).$$

Since $2q - 4 > -1$, assuming (19), we get

$$\mathbb{E}(|e_i|^{2q-4}|\mathbf{x}_i) = \mathbb{E}(\int |y|^{2q-4}f_{w(i)}(i/n, y|\mathbf{x}_i)dy) \le$$

$$\mathbb{E}(M_0\int_0^1 |y|^{2q-4}dy + \int_1^\infty f_{w(i)}(i/n, y|\mathbf{x}_i)dy) \le M.$$

By applying similar arguments to $\nu_i^-(\delta)$, we have that $\nu_i(\delta) = O(\delta^2)$. For $1 < q \le 3/2$, by Lemma 4 of [3], we obtain

$$\nu_i^+(\delta) \le 36\delta^2 \mathbb{E}(|\mathbf{x}_i|^4 |e_i|^{2q-4} \mathbf{1}(|e_i| \ge \delta)) + 2^{4-2q}|\delta|^{2q-2}\mathbb{E}(|\mathbf{x}_i|^{2q}\mathbf{1}(|e_i| \le \delta)).$$

For $q = 3/2$, by (19), we have $\mathbb{E}(|e_i|^{2q-4}\mathbf{1}(|e_i| \ge \delta)|\mathbf{x}_i) \le M\delta^{2q-3}\log(|\delta|^{-1})$ and $\mathbb{E}(\mathbf{1}(|e_i| \le \delta)|\mathbf{x}_i) \le M\delta$, which lead to that $\nu_i^+(\delta) = O(|\delta|^{2q-1}\log(|\delta|^{-1}))$. By applying similar arguments to $\nu_i^-(\delta)$, we have $\nu_i(\delta) = O(|\delta|^{2q-1}\log(|\delta|^{-1}))$. Similar but easier arguments show that $\nu_i(\delta) = O(|\delta|^{2q-1})$ for $1 < q < 3/2$.

In Theorem 2.1, i) establishes a Bahadur representation of $\hat{\beta}_n$ for non-stationary time series M-estimation and ii) establishes a uniform Bahadur representation of $\{\hat{\beta}_j, N \le j \le n\}$. When $\mathcal{L}_1$ loss is applied, both results almost achieve the optimal order $n^{-1/4}(\log\log n)^{3/4}$ except a factor of multiplicative logarithms. For Huber, $\mathcal{L}_q, 3/2 < q \le 2$ and expectile regressions, according to Remark 2.1, the approximation rates are $\log^2 n$ and $\log^3 n$ in i) and ii) respectively. The latter rates are again nearly optimal except a factor of multiplicative logarithms. Observe that, due to the non-stationarity, the approximating processes depend on $\{\bar{\kappa}_{w(i)}(i/n, 0, \mathbf{x}_i), 1 \le i \le n\}$, which are the conditional densities of the errors e_i in the scenario of quantile regression. [23] also provided a similar form of Bahadur representation with non-stationary errors for quantile regression.

3. Structural Stability Tests.

We are now ready to propose and investigate change point tests for general non-stationary time series M-estimation.

3.1. *Test Statistics.*

Consider a general nonparametric M-estimation model of the form

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

We are interested in testing whether β_i's remain constant over time. That is, we test

$$H_0 : \beta_1 = \beta_2 = ... = \beta_n = \beta \leftrightarrow H_A : \beta_i \ne \beta_j \text{ for some } 1 \le i < j \le n$$

for some unknown β. Consider the following test statistic:

$$T_n = \max_{1 \le j \le n} \left| \frac{\sum_{i=1}^{j} \psi(\hat{e}_{i,n})\mathbf{x}_i}{\sqrt{n}} \right|,$$

where $\hat{e}_{i,n} = y_i - \mathbf{x}_i'\hat{\beta}_n$ are the residuals. The test statistic T_n is the CUSUM statistic of the estimated gradient vectors of the regression. If H_0 is violated, then T_n tends to be large. In the following, we shall investigate the asymptotic null distribution and power behavior of T_n in detail. Note that, under conditions (S0)-(S2), $\psi(e_i)\mathbf{x}_i$ can be viewed as a realization from a PLS process with r_1 break points $c_1, ..., c_{r_1}$, namely, $\tilde{G}_{v(i)}(t, \mathcal{F}_i, \mathcal{G}_i)$, 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 $\psi(e_i)\mathbf{x}_i = \tilde{G}_{v(i)}(i/n, \mathcal{F}_i, \mathcal{G}_i)$. Here $r_1 = |A \cup B|$, where $A = \{b_1, ..., b_r\}$ is the set of break points of the errors, and $B = \{d_1, ..., d_s\}$ is the set of break points of the covariates. Here $|\cdot|$ denotes the cardinality of a set. The detailed mathematical form of $\tilde{G}_{v(i)}(t, \mathcal{F}_i, \mathcal{G}_i)$ is complex which we treat as a nuisance parameter. Define the long-run covariance matrices:

$$\Sigma^2(t) = \sum_{h=-\infty}^{\infty} Cov(\tilde{G}_k(t, \mathcal{F}_0, \mathcal{G}_0), \tilde{G}_k(t, \mathcal{F}_h, \mathcal{G}_h)),\ t \in (c_k, c_{k+1}],\ 0 \leq k \leq r_1.$$

Let $\Sigma^2(0) = \lim_{t\downarrow 0} \Sigma^2(t)$. In order to investigate the limiting behavior of T_n, we shall further introduce the following assumption:

(A2) The smallest eigenvalue of $\Sigma^2(t)$ is bounded away from 0 for $t \in [0, 1]$.

It is shown in Proposition A.2 in the supplemental material that the dependence of $\{\psi(e_i)\mathbf{x}_i\}_{i=1}^n$ decays exponentially fast to 0. Meanwhile, condition (A2) assures that the long run variance of $\psi(e_i)\mathbf{x}_i$ is not degenerate over time, which is a mild requirement. We have the following proposition, which is useful in the asymptotic study of the process $\{\psi(e_i)\mathbf{x}_i\}_{i=1}^n$:

PROPOSITION 3.1. *Let $f_k(t,x)$ be the density of $G_k(t, \mathcal{F}_0, \mathcal{G}_0)$. Assume (A2), (S0)-(S2) with $v = 4(p+1)$. Assume that (i): there exists a sufficiently small positive ι such that*

$$\sup_{0\leq k\leq r, t\in(b_k, b_{k+1}]} \|\psi\left(G_k(t, \mathcal{F}_0, \mathcal{G}_0)\right)\|_{\frac{4(p+1)}{p}+\iota} \leq M < \infty,$$

and (ii): $\int \psi(x;1)|\frac{\partial}{\partial x} f_k(t,x)|dx$ is finite for $0 \leq k \leq r, t \in (b_k, b_{k+1}]$. Then on a possibly richer probability space, there exists a p-dimensional zero-mean Gaussian process $U(t)$ with covariance function $\gamma(t,s) = \int_0^{\min(t,s)} \Sigma^2(r)dr$, such that

$$\max_{1\leq j\leq n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \psi(e_i)\mathbf{x}_i - U(j/n) \right| = o_p(n^{-1/4}\log^2 n).$$

Write $\Lambda(t) = \lim_{n\to\infty} \mathbb{E}\{\sum_{i=1}^{\lfloor nt\rfloor} \bar{\kappa}_{w(i)}(i/n, 0, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'\}/n$. Since $\mathbf{x}_i$ is $(\mathcal{F}_{i-1}, \mathcal{G}_i)$ measurable, we have that

$$\mathbb{E}\left\{\sum_{i=1}^{\lfloor nt\rfloor} \bar{\kappa}_{w(i)}(i/n, 0, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'\right\} = \mathbb{E}\left\{\sum_{i=1}^{\lfloor nt\rfloor} \Xi_{w(i)}^{(1)}(i/n, 0|\mathcal{F}_{i-1}, \mathcal{G}_i)\mathbf{x}_i\mathbf{x}_i'\right\}.$$

Without loss of generality, suppose that the covariates and the errors have the same break points, i.e, $\{b_1, ..., b_r\} = \{d_1, ..., d_s\}$. Then by (9) in (S1), we have that for $s \in (b_j, b_{j+1}]$,

$$\Lambda(s) = \sum_{l=0}^{j-1} \int_{b_l}^{b_{l+1}} \mathbb{E}\{\Xi_l^{(1)}(t, 0|\mathcal{F}_{-1}, \mathcal{G}_0)\mathbf{H}_l(t, \mathcal{F}_{-1}, \mathcal{G}_0)\mathbf{H}_l(t, \mathcal{F}_{-1}, \mathcal{G}_0)'\}dt +$$

$$\int_{b_j}^{s} \mathbb{E}\{\Xi_j^{(1)}(t, 0|\mathcal{F}_{-1}, \mathcal{G}_0)\mathbf{H}_j(t, \mathcal{F}_{-1}, \mathcal{G}_0)\mathbf{H}_j(t, \mathcal{F}_{-1}, \mathcal{G}_0)'\}dt. \tag{21}$$

The following theorem establishes the limiting null distribution of T_n for non-stationary time series M-estimation:

THEOREM 3.1. *Assume that* $\max_{0\le j\le r} \sup_{t\in(b_j, b_{j+1},] x\in\mathbb{R}} |f_j(t, x|\mathcal{F}_{-1}, \mathcal{G}_0)| < \infty$. *Suppose (S0)-(S2), (A0)-(A2) and the conditions of Proposition 3.1 hold. Then under the null hypothesis of no structural change, we have*

$$T_n \Rightarrow \sup_{t\in(0,1]} |G(t)| := \sup_{t\in(0,1]} |U(t) - \Lambda(t)\Lambda^{-1}(1)U(1)|, \tag{22}$$

where $U(t)$ *is defined in Proposition 3.1.*

Theorem 3.1 establishes that T_n converges to the maximum of certain centered Gaussian process. Two important observations should be made. First, the Gaussian process $U(t)$ is not pivotal and it has a complex covariance structure $\gamma(t, s) = \int_0^{\min(t,s)} \Sigma^2(r)dr$. In particular, $\Sigma^2(s)$ can change both smoothly and abruptly on [0,1] and hence it is inappropriate to perform T_n by checking quantile tables of certain pivotal Gaussian processes (such as the Brownian bridge). Second, due to the non-stationarity, $\Lambda(t)\Lambda^{-1}(1)$ 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_n is no longer asymptotically free of the density functions of $\{e_i\}_{i=1}^n$ and the ratio $\Lambda(t)\Lambda^{-1}(1)$ should be estimated when performing the gradient CUSUM test for non-stationary time series M-estimation. Consequently, the independent wild bootstrap procedure in [16] will in general yield inconsistent testing results under non-stationarity.

The following theorem studies the asymptotic power behavior of the test for non-stationary time series M-estimation. For any bounded piecewise Lipschitz continuous $p \times 1$ vector function $g(\cdot)$, write

$$\Lambda(s, g(\cdot)) = \lim_{n\to\infty} \mathbb{E}\left\{ \sum_{i=1}^{\lfloor nt \rfloor} \bar{\kappa}_{w(i)}(i/n, 0, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i' g(i/n)/n \right\}.$$

By the arguments of (21), $\Lambda(s, g(\cdot))$ is well defined. Write

$$\digamma(t, g(\cdot)) = \Lambda(t, g(\cdot)) - \Lambda(t)\Lambda(1)^{-1}\Lambda(1, g(\cdot)).$$

THEOREM 3.2. *Consider the alternative model $H_A : \beta_i = \beta + L_n g(i/n)$, where $g(\cdot)$ is a bounded non-constant piecewise Lipschitz continuous $p \times 1$ vector function defined in $[0, 1]$. Suppose that (S0)-(S2), (A0)-(A2) and the conditions of Proposition 3.1 hold. Assume*

$$\max_{0\le k\le r} \sup_{t\in(b_k, b_{k+1}], x\in\mathbb{R}} |f_k^{(j)}(t, x|\mathcal{F}_{-1}, \mathcal{G}_0)| < \infty$$

for $0 \le j \le 3$. For quantile regression, assumes (A1) instead of (A1). Then we have,*
i): if $L_n = n^{-1/2}$,

$$T_n \Rightarrow \sup_{0<t\le 1} |G(t) + \digamma(t, g(\cdot))|,$$

where $G(t)$ is defined in Theorem 3.1.
ii): If the deterministic sequence L_n satisfies $L_n = o(1)$, $\frac{\sqrt{\sum_{i=1}^n \nu_i(L_n)}}{\sqrt{n}} \log n \to 0$, $\sqrt{n}L_n \to \infty$, then $T_n \to_p \infty$ at the rate $\sqrt{n}L_n$.

Theorem 3.2 shows that the power of the test converges to 1 if $\sqrt{n}L_n \to \infty$, $L_n = o(1)$ and $\frac{\sqrt{\sum_{i=1}^n \nu_i(L_n)}}{\sqrt{n}} \log n \to 0$, which implies that our test can detect local alternatives at the same rate $n^{-1/2}$ as the classic stationary case.

3.2. *The Bootstrap.* Theorem 3.1 reveals that the key to accurate tests under non-stationarity is to consistently mimic the behaviors of the processes $\{\Lambda(t)\}$ and $\{U(t)\}$. A straightforward way to generate the limiting distribution in Theorem 3.1 is to directly estimate $\{\Lambda(t)\}$ and $\{U(t)\}$, which involves the estimation of conditional densities and long-run covariances $\Sigma^2(t)$ over time t, respectively. However, this approach is not operational in practice for the following two reasons. First, the estimation of the densities and the long-run covariance at a fixed time t requires a total of four bandwidth parameters, which are difficult to choose in practice and can cause inaccurate testing results for moderate samples. Second, the nonparametric estimates of $\{\Lambda(t)\}$ and $\Sigma^2(t)$ 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 while requiring only two tuning parameters. The proposed bootstrap procedure combines the advantages of moving block bootstrap ([19]) and subsampling ([22]) by progressively convoluting block 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 of the "Powell Sandwich" ([24]) to optimally estimate $\{\Lambda(t)\}$. In the following we shall discuss the approximations of $\{\Lambda(t)\}$ and $\{U(t)\}$ separately.

Let c_n be a bandwidth parameter. Define $\hat{\Lambda}_{c_n}(t) = \hat{\lambda}_{c_n}(\lfloor nt \rfloor)$, where

$$\hat{\lambda}_{c_n}(j) = \sum_{i=1}^{j} \frac{(\psi(\hat{e}_{i,n} + c_n) - \psi(\hat{e}_{i,n} - c_n))\mathbf{x}_i \mathbf{x}_i'}{2nc_n}. \tag{23}$$

Note that for least squares regression, $\hat{\lambda}_{c_n}(j)$ in (23) equals $\sum_{i=1}^{j} \mathbf{x}_i \mathbf{x}_i'/n$ which is independent of c_n. In addition, for quantile regression, we propose another smooth estimator for $\{\Lambda(t)\}_{t \in (0,1]}$. Define

$$\hat{\lambda}_{c_n}(j) = \sum_{i=1}^{j} \frac{K(\hat{e}_{i,n}/c_n)\mathbf{x}_i \mathbf{x}_i'}{nc_n}. \tag{24}$$

where $K(\cdot)$ is a symmetric smooth kernel function with bounded second order derivative, satisfying $\int K(x)dx = 1$, $\int K(x)x^2 dx \le M$, $\int K^2(x)dx \le M$ and $\int K'^2(x)dx \le M$. The following theorem states that $\{\hat{\Lambda}_{c_n}(t)\}_{t \in (0,1]}$ could be used to approximate $\{\Lambda(t)\}_{t \in (0,1]}$ uniformly.

Theorem 3.3. *Assume (S0)-(S2), (A0) and (A1) with* $0 \le q \le (3 \vee (p+1))$ *(or (A1*) for quantile regression). Further assume for* $0 \le s \le 3$, $\max_{0 \le k \le r} \sup_{b_k < t \le b_{k+1}, x \in \mathbb{R}} |f_k^{(s)}(t, x|\mathcal{F}_{-1}, \mathcal{G}_0)| < \infty$. *Then (i): assuming* $c_n \to 0$, $\frac{\sqrt{\sum_{i=1}^n \nu_i(c_n) \log n}}{nc_n} \to 0$, $nc_n^2/\log^2 n \to \infty$, *we have*

$$\sup_{t \in (0,1]} |\hat{\Lambda}_{c_n}(t) - \Lambda(t)| = O_p\left(\frac{(\sqrt{\sum_{i=1}^n \nu_i(c_n)} \vee 1) \log n}{nc_n} + c_n^2 + n^{-1/2} \log n \right).$$

ii): If smooth estimator (24) *is used for quantile regression, assuming* $\max_{0\le j\le r}\sup_{b_j<t\le b_{j+1},x\in\mathbb{R}}\|f_j^{(q)}(t,x|\mathcal{F}_{i-1},\mathcal{G}_i)-f_j^{(q)}(t,x|\mathcal{F}_{i-1}^*,\mathcal{G}_i^*)\|=O(\chi^i)$ *for* $q=0,1$*,* $c_n\log^2 n\to 0$ *and* $nc_n^3/\log^2 n\to\infty$*, then we have*

$$\sup_{t\in(0,1]}|\hat{\Lambda}_{c_n}(t)-\Lambda(t)|=O_p\left(n^{-1/2}c_n^{-1/2}+c_n^2+\frac{\log^2 n}{nc_n^3}\right).$$

In fact, $\hat{\lambda}_{c_n}(j)$ is an extension of the "Powell's Sandwich". Furthermore, for quantile regression, $\hat{\lambda}_{c_n}(j)$ can be viewed as a progressive local constant kernel estimation of integrated conditional densities. Theorem 3.3 shows that $\{\hat{\Lambda}_{c_n}(t)\}_{t\in(0,1]}$ are uniformly consistent estimators of $\{\Lambda(t)\}_{t\in(0,1]}$. Elementary calculations show that, even with PLS errors, the optimal bandwidth c_n for Theorem 3.3 is almost of the order of $n^{-1/5}$ for quantile regression. Therefore the convergence rate of Theorem 3.3 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 approximation rate of the Powell's sandwich estimates for i.i.d. data. Note that the nearly $n^{-2/5}$ rate above is faster than $n^{-1/4}\log^2 n$, which is the nearly optimal approximation rate of the Bahadur representation in (18). For the Huber, expectile and $\mathcal{L}_q$, $q\in(3/2,2]$ regressions, our method also achieves the almost optimal rate $n^{-1/2}\log n$ when c_n satisfies the stated bandwidth conditions and converges to zero no slower than $n^{-1/4}$ and no faster than $n^{-1/2}$. For $\mathcal{L}_{1.5}$, the convergence rate could achieve $n^{-1/2}\log^2 n$.

The remaining task for evaluating the critical values is to find a simple and data-driven way to simulate the non-stationary Gaussian process $U(t)$. The covariance structure of $U(t)$ 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 $U(t)$:

$$\tilde{\Psi}_{m,n}(t)=\Psi_{\lfloor nt\rfloor,m}+(nt-\lfloor nt\rfloor)(\Psi_{\lfloor nt\rfloor+1,m}-\Psi_{\lfloor nt\rfloor,m}),\tag{25}$$
$$\Psi_{i,m}=\sum_{j=1}^{i}\frac{1}{\sqrt{m(n-m+1)}}(\hat{\varpi}_{j,m}-\frac{m}{n}\hat{\varpi}_n)R_j, i=1,...,n-m+1,$$

where $\hat{\varpi}_{j,m}=\sum_{r=j}^{j+m-1}\psi(\hat{e}_{r,n})\mathbf{x}_r$, $\hat{\varpi}_n=\hat{\varpi}_{1,n}$ and $(R_i)_{i=1}^n$ are i.i.d. standard normals 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 $\{U(t)\}$ is provided by the following theorem:

THEOREM 3.4. *Assume that* $\max_{0\le j\le r}\sup_{t\in(b_j,b_{j+1}],x\in\mathbb{R}}|f_j(t,x|\mathcal{F}_{-1},\mathcal{G}_0)|<\infty$ *and the bandwidth* $m=m(n)$ *satisfies* $m\to\infty$*,* $m/n\to 0$*. Suppose (S0)-(S2), (A0)-(A2) and the conditions of Proposition 3.1 hold. Further assume* *that there exists some constant* ε_0*, such that for* m *and* $1 \le j \le n-m+1$, $|\sum_{r=j}^{j+m-1} \nu_r(\delta)| \le M\frac{m}{n}\sum_{i=1}^n \nu_i(\delta)$ *for all* $|\delta| \le \varepsilon_0$*. Then conditional on* $(\mathcal{F}_n, \mathcal{G}_n)$*,* $\tilde{\Psi}_{m,n}(t) \Rightarrow U(t)$ *on* $\mathcal{C}(0,1)$ *with the uniform topology.*

By the proof of Theorem 3.4, for quantile regressions, if we further require $\sqrt{n}\log n/m \to \infty$, then conditional on $(\mathcal{F}_n, \mathcal{G}_n)$, the covariance function of $\tilde{\Psi}_{m,n}(t)$ converges uniformly to that of $U(t)$ at the rate $n^{-1/4}\log^{3/2} n$, which is also of the same order of the nearly optimal approximation rate of the Bahadur representation in (17). Therefore Theorem 3.3 and Theorem 3.4 suggest that our bootstrap approaches the covariance structure of the target limiting Gaussian process no slower than the nearly optimal approximation rate of the Bahadur representation. We also note that for the Huber, $\mathcal{L}_q, 3/2 \le q \le 2$ and expectile regressions, by Remark 2.1, the optimal Bahadur representation rate is almost of the order of $\frac{1}{\sqrt{n}}$ except a fact of multiplicative of logarithms, which could not be archived by Theorem 3.4. The reason is that the optimal approximation rate of the robust bootstrap is of the order $n^{-1/3}$, see [32].

In Theorem 3.2, we show that under H_A, if $L_n \to 0$, $\frac{\sqrt{\sum_{i=1}^n \nu_i(L_n)}}{\sqrt{n}}\log n \to 0$ and $\sqrt{n}L_n \to \infty$, the test statistic goes to infinity at the rate $\sqrt{n}L_n$. Proposition B.1 in the supplemental material further discusses the property of $\{\hat{\Lambda}_{c_n}(t)\}_{t\in(0,1]}$ and $\tilde{\Psi}_{m,n}(t)$ under the local alternative hypotheses. Assumes that $\frac{m\log^8 n}{n} = o(1)$. Then the divergence rate of T_n under the local alternatives in ii) of Theorem 3.2 is $\sqrt{n}L_n$, which is faster than $\sqrt{m}(L_n \vee \frac{\log^2 n}{\sqrt{n}})\log n$, the fastest possible rate at which $\tilde{\Psi}_{m,n}(t)$ can go to infinity by Proposition B.1. Hence Theorem 3.2 together with Proposition B.1 shows that our bootstrap method has asymptotic power 1 under the considered local alternatives in ii) of Theorem 3.2. In particular, our bootstrap can detect local alternatives with the optimal $n^{-1/2}$ parametric rate.

REMARK 3.1. *The limiting distribution of the test statistic, and hence the non-local power, is hard to evaluate when* $L_n = 1$ *due to the time series non-stationarity and the possibly non-differentiable gradient function. When* $L_n \gg 1$*, [17] proposes to replace* $\hat{e}_i$ *with* $\tilde{e}_i$ *to deal with the non-local power issue, where* $\hat{e}_i$ *is the residual of the parametric linear regression under the null hypothesis, and* $\tilde{e}_i$ *is the residual of a general nonparametric regression. We only consider very general form of alternatives. Hence [17] is not directly applicable. We leave the problem of non-local power as a rewarding future work.*

Combining Theorems 3.3 and 3.4, we have the following step-by-step implementation procedures for performing structural change tests for non-stationary time series M-estimation.

ALGORITHM 3.1.
(i): By Section 3.3, select appropriate m *and* c_n.
(ii): Apply Theorem 3.3 to get $\hat{\lambda}_{c_n}(j)$, $j = 1, ...n$. *Use Theorem 3.4 to generate* B *(say 2000) conditional i.i.d. copies* $\{\Psi_{i,m}^{(r)}\}_{i=1}^{n-m+1}$, $r = 1, ..., B$.
(iii): Calculate $F_i^{(r)} = \Psi_{i,m}^{(r)} - \hat{\lambda}_{c_n}(i)\hat{\lambda}_{c_n}^{-1}(n-m+1)\Psi_{n-m+1,m}^{(r)}$ *for* $r = 1, ..., B$, $i = m, ..., n-m+1$.
(iv): Let $F_r = \sup_{m \le i \le n-m+1} |F_i^{(r)}|$. *Let* $F_{(1)} \le F_{(2)} ... \le F_{(B)}$ *be the order statistics of* F_r. *Then* $F_{\lfloor (1-\alpha)B \rfloor}$ *consistently estimates the level* α *critical value for the gradient-based structural change test* (22).

THEOREM 3.5. *Under conditions of Theorem 3.3 and 3.4, algorithm 3.1 generates consistent estimator of the level* α *critical value for the test* (22).

In the supplemental material, we also extend our method to test structural changes for multiple M-estimators simultaneously and discuss the applicability of the method to dynamic models.

3.3. *Bandwidth Selection.* To implement our testing procedure, one has to choose the tuning parameters c_n and m (except for least squares regression where only m needs to be chosen). Due to the 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 3.4, we apply the method of minimum volatility (MV) suggested by [32] to $\tilde{\Psi}_{m,n}(t)$ in (25). The procedures are quite similar except that we replace unknown $\psi(e_i)\mathbf{x}_i$ with estimated $\psi(\hat{e}_{i,n})\mathbf{x}_i$. Define

$$\hat{\gamma}_m(r/n, s/n) = \sum_{i=1}^{r \wedge s} (\hat{\varpi}_{i,m} - \frac{m}{n}\hat{\varpi}_n)^2 / (m(n-m+1)).$$

Calculate $\{\hat{\gamma}_{m_j}(r/n, r/n)\}_{r=1}^{n-m_j+1}$ for a grid of possible window sizes $m_1 \le ... \le m_M$. Write

$$se(\{\hat{\gamma}_{m_j}(r/n, r/n)\}_{j=a}^{b}) = \left(\frac{1}{b-a} \sum_{j=a}^{b} \left(\hat{\gamma}_{m_j}(r/n, r/n) - \bar{\hat{\gamma}}(r/n, r/n)\right)^2 \right)^{1/2},$$

where $\bar{\hat{\gamma}}(r/n, r/n) = \frac{1}{b-a+1}\sum_{j=a}^{b} \hat{\gamma}_{m_j}(r/n, r/n)$. Then we choose $m = m_j$ where $j = \operatorname{argmin}_{4 \le j \le 3}(\max_{1 \le r \le n-m_M+1} se(\{\hat{\gamma}_{m_{j+k}}(r/n, r/n)\}_{k=-3}^{3}))$. For more discussions about the "MV" method, see [22]. We also apply the MV method to the selection of the bandwidth c_n. Our procedure of selecting c_n is as follows:

(i): Choose suitable end points $a_1 < a_2$, such that the optimal $c_n \in I := [a_1, a_2]$.
(ii): Divide interval I into $\bar{m}$, say $\bar{m} = 99$ pieces. Specifically, let $h_1 = a_1$, $h_{100} = a_2$, and $h_{k'} = a_1 + (k'-1)(a_2 - a_1)/99$, $1 \le k' \le 100$.
(iii): For each h_i, use it as a bandwidth to calculate the estimating quantity $\{\hat{\Lambda}_{h_i}(t_j)\}_{j=1}^n$. Let $C(i)$ be the maximum of RHS process of (22) for $t \in [\frac{1}{n}, 1]$ obtained by replacing $\{U(t), t \in [\frac{1}{n}, 1]\}$ with $\{\sum_{i=1}^{\lfloor nt \rfloor} \frac{\psi(\hat{e}_{i,n})\mathbf{x}_i}{\sqrt{n}}, t \in [\frac{1}{n}, 1]\}$ and replacing $\{\Lambda(t), t \in [\frac{1}{n}, 1]\}$ with $\{\hat{\Lambda}_{h_i}(t), t \in [\frac{1}{n}, 1]\}$, respectively.
(iv): Define $D(s) = \{\frac{1}{2k}\sum_{j=s-k}^{s+k}[C(j) - \frac{1}{2k+1}\sum_{j=s-k}^{s+k} C(j)]^2\}^{1/2}$ for some $k > 0$. Let l be the minimizer of $\{D(s)\}_{s=k+1}^{100-k}$. Then we select h_l as our bandwidth c_n.

Since $\Lambda(t)$ is a $p \times p$ matrix, directly applying the MV method for selecting c_n will be time-consuming. Our proposed procedure is based on the fact that $U(t)$ is the limiting distribution of $\sum_{i=1}^{\lfloor nt \rfloor} \frac{\psi(e_{i,n})\mathbf{x}_i}{\sqrt{n}}$. This motivates us to generate the pseudo limiting distribution via replacing $U(t)$ with $\sum_{i=1}^{\lfloor nt \rfloor} \frac{\psi(\hat{e}_{i,n})\mathbf{x}_i}{\sqrt{n}}$ and choose the bandwidth as the one which stabilizes the pseudo quantity mostly. The numeric experiment shows that our bandwidth selection criteria work well under various circumstances.

4. Simulation Studies.

4.1. *Type I Error.* In this section we examine the performance of our method for $\mathcal{L}_2$ (least squares) regression, the Huber regression with $\varsigma = 1.5$, $\mathcal{L}_{1.5}$ regression and quantile regression with quantiles 0.5 and 0.8. Throughout our simulations the number of bootstrap samples $B = 2000$. For quantile regression, we also compare our results with SQ method in [26]. The description of the SQ method for τ_{th} quantile regression is as follows.

Let $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)'$. Define $H_{\lambda,n}(\hat{\beta}) = (\mathbf{X}\mathbf{X}')^{-1/2}\sum_{i=1}^{\lfloor \lambda n \rfloor} \mathbf{x}_i \psi_\tau(y_i - \mathbf{x}_i'\hat{\beta})$, where $\psi_\tau(x) = \tau - \mathbf{1}(x \le 0)$. Then the SQ test statistic is defined as

$$SQ_\tau = \sup_{\lambda \in [0,1]} \|(\tau(1-\tau))^{-1/2}[H_{\lambda,n}(\hat{\beta}) - \lambda H_{1,n}(\hat{\beta})]\|_\infty.$$

The associated critical values for the SQ_τ test are in Table 1 of [26]. To estimate $\{\Lambda(t), t \in (0, 1]\}$, we choose bandwidth from 100 equally spaced points in a certain range. In each iteration we select the bandwidths by the MV method we proposed in Section 3.3.

We consider the following heteroscedastic linear regression model

$$y_i = 1 + x_i + e_i, e_i = (1 + \gamma x_i)(\Upsilon_i - c_i)/2 \tag{26}$$

for $i = 1, ..., n$, $\gamma = 0.1$. In our simulations, x_i are i.i.d. $\chi^2(5)/5$ and $c_i = \tilde{F}_i^{-1}(0)$ where $\tilde{F}_i(x) = \mathbb{E}(\psi(\Upsilon_i - x))$. Let filtration $\mathcal{F}_i = (\varepsilon_{-\infty}, ..., \varepsilon_i)$ where $\{\varepsilon_s\}_{s=-\infty}^{\infty}$ are independent of $\{x_i\}_{i=1}^{n}$. We shall consider the following models for $\{\Upsilon_i\}_{i=-\infty}^{\infty}$.

(I): Consider

$$\Upsilon_i = G(t_i, \mathcal{F}_i),\ G(t, \mathcal{F}_i) = 0.75\cos(2\pi t)G(t, \mathcal{F}_{i-1}) + \varepsilon_i,$$

where ε_i are i.i.d. N(0,1). This is a locally stationary model since its AR(1) coefficient $0.75\cos(2\pi t)$ changes smoothly over $(0, 1]$.

(II): Consider $\Upsilon_i = G(t_i, \mathcal{F}_i)$, and

$$G(t, \mathcal{F}_i) = G_1(t, \mathcal{F}_i)\mathbf{1}(0 < t \le 0.8) + G_2(t, \mathcal{F}_i)\mathbf{1}(0.8 < t \le 1),$$

where

$$\begin{aligned} G_1(t, \mathcal{F}_i) &= 0.6\cos(2\pi t)G_1(t, \mathcal{F}_{i-1}) + \varepsilon_i, \\ G_2(t, \mathcal{F}_i) &= (0.5 - t)G_2(t, \mathcal{F}_{i-1}) + \varepsilon_i \end{aligned}$$

and $\varepsilon_i' s$ are i.i.d. N(0,1). This is a PLS model. The AR coefficient changes smoothly before and after $t = 0.8$, with an abrupt change at $t = 0.8$.

(II'): Model II' is the same as model II except we change the i.i.d. N(0,1) $\varepsilon_i' s$ to i.i.d. student t distribution with 12 degrees of freedom ($t(12)$).

(II*): Model II* is the same as model II' except we set x_i i.i.d. $1+(5/3)^{-1/2}t(5)$.

Note that in Models I, II, II' and II*, the covariates are independent and identically distributed. However, the errors are heteroscedastic with respect to the covariates. Furthermore, the errors are PLS processes which exhibit smooth and (or) abrupt changes in their data generating mechanisms over time.

We also consider the following non-stationary dynamic model III. Let $\{z_i\}_{i\in\mathbb{Z}}$ be i.i.d. $\chi^2(1)$ random variables, and

(III): $y_i = 0.3y_{i-1}/(1 + z_{i-1}) + e_i/3,\ y_0 \sim N(0, 1)$, where $e_i = G(t_i, \mathcal{F}_i)$ with

$$\begin{aligned} G(t, \mathcal{F}_i) = (1{+}\frac{1}{3}(t-1)^2)\epsilon_i\mathbf{1}(0 < t \le 0.5) \\ + (1 + 0.5\cos(2\pi t))\epsilon_i\mathbf{1}(0.5 < t \le 1), \epsilon_i' s \text{ i.i.d. } N(0, 1), \end{aligned}$$

TABLE 1
Simulated type I error rates for Huber ($\varsigma = 1.5$), $\mathcal{L}_{1.5}$ and $\mathcal{L}_2$ regression

	Huber Regression				$\mathcal{L}_{1.5}$ Regression				Least Squares Regression			
	n=300		n=600		n=300		n=600		n=300		n=600	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
I	4.25	11.55	4.35	11	5.3	13.05	4.8	11.55	5.55	13.2	5.35	12.25
II	4.45	10.35	4.55	9.95	3.85	9.8	4.25	10.4	4.8	12.05	5.2	11.6
II'	4.1	10.35	4.75	10.3	4.35	10.3	4.7	10.2	5.65	12.35	4.8	10.9
II*	4.05	9.95	5.2	10.75	3.7	9.25	4.2	10.05	4.9	11.7	4.2	10.65
III	4.3	10.55	4.95	10.85	4.3	10.85	4.65	10.3	3.95	10.6	4.45	10.15
IV	3.15	8.0	5.1	10.85	3.9	8.75	4.75	10.25	3.15	7.95	4.4	9.6

Finally we consider the following scenario IV. Let Υ_i be the PLS process defined in Model II and c_i be the corresponding quantity defined below (26). Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}, \{\eta_i\}_{i\in\mathbb{Z}}, \{\epsilon_i\}_{i\in\mathbb{Z}}$ be i.i.d. N(0,1)'s. Furthermore, $\{\varepsilon_i\}_{i\in\mathbb{Z}}, \{\eta_i\}_{i\in\mathbb{Z}}$ and $\{\epsilon_i\}_{i\in\mathbb{Z}}$ are independent of each other. Let $\upsilon_i = \frac{\eta_i+\epsilon_i}{\sqrt{2}}$. Consider

(IV): Let $x_{1,i}$ be the PLS process generated from $G_1(t, \mathcal{G}_i) = \sum_{j=0}^{\infty}(0.5 - 0.5t)^j \upsilon_{i-j}$, and $x_{2,i}$ be the PLS process generated from $G_2(t, \mathcal{G}_i) = \sum_{j=0}^{\infty}(0.25 + 0.5t)^j \epsilon_{i-j}$, where $\mathcal{G}_i = (..., \epsilon_i, \eta_i)$. Let

$$y_i = 1 + x_{1,i} + x_{2,i} + e_i, e_i = \frac{\sqrt{1 + x_{1,i} + x_{2,i}}(\Upsilon_i - c_i)}{4}.$$

TABLE 2
Simulated type I error rates for quantile regressions, gradient-based test

	0.5 Quantile				0.8 Quantile			
	n=300		n=600		n=300		n=600	
	5%	10%	5%	10%	5%	10%	5%	10%
Model I	4.8	14.05	4.1	10.9	4.8	13.45	4.55	10.75
Model II	4.05	11.35	4.25	11.1	4.95	12.55	4.35	9.95
Model II'	4.7	12.05	4.25	10.6	5.6	12.3	4.65	10.9
Model II*	3.1	10.1	3.65	9.05	3.6	10.75	3.8	9.65
Model III	4.35	9.2	5.2	10.3	3.95	10.2	4.3	10.15
Model IV	4.0	8.55	4.25	9.45	3.65	10.1	4.5	10.2
AR(0.5)	4.05	9.2	4.95	10.85	4.2	9.25	5.55	10.25

Note that in Model IV both the covariates and errors are non-stationary time series. We examine our test with sample sizes 300 and 600 at two nominal levels 5% and 10%. We report the simulated type I errors in Tables 1-2 for our proposed gradient-vector-based test (22). In Table 1, we present our simulation results for the Huber regression, $\mathcal{L}_{1.5}$ regression and least squares regression, respectively. Table 2 contains our simulation results for quantile regressions with $\tau = 0.5$ and 0.8, respectively. The simulated Type I errors are quite close to the nominal levels. Meanwhile, our simulation results show that increasing the sample size from 300 to 600 in general significantly improves the performance of our test. In addition, for quantile regression, the test performs better when the quantile is less extreme. The Monte Carlo experiments also show the inadequacy of the SQ method when e_i and (or) x_i are non-stationary. For comparison purpose, we also generated a stationary $AR(0.5)$ model: $y_i = 0.5y_{i-1} + \varepsilon_i$ where $\varepsilon_i' s$ are i.i.d. $N(0,1)$. We find that the SQ method works well for this stationary $AR(0.5)$ model, which is consistent with the results reported in [26]. Meanwhile, simulation results show that our method performs almost as well as SQ method under the stationary scenario.

Table 3
Simulated type I error rates for quantile regressions, SQ method

	0.5 Quantile				0.8 Quantile			
	n=300		n=600		n=300		n=600	
	5%	10%	5%	10%	5%	10%	5%	10%
I	15	24.8	15.65	24.85	11.3	19.1	14.9	24.2
II	5.55	10.55	5.55	11.45	5.15	10.8	6.55	12.7
II'	5.7	10.25	6	10.8	5.2	10.8	6.05	12.2
II*	4.8	9.3	5.9	11.75	6.75	12.15	7.1	13.3
III	2.25	4.7	3.1	6.15	5.7	10.8	6.85	13.05
IV	9.15	18.75	9.5	19.65	9.05	17.9	9.9	20.2
AR(0.5)	5.4	10.45	4.65	10.1	3.85	7.55	4.85	9.45

4.2. *Simulated Power.* We consider the alternative model that

$$y_i = 1 + x_i(1 + \delta\mathbf{1}(i \geq \lfloor n/2 \rfloor)) + e_i, e_i = (1 + \gamma x_i)(\Upsilon_i - c_i)/2,$$

where Υ_i follows model II, which is PLS. We shall simulate different jump sizes δ to investigate the power performances of our testing procedures. The sample size and significance level are 300 and 10% in our simulation. The left panel of Figure 1 examines the simulated powers for the Huber regression with $\varsigma = 1.5$, $\mathcal{L}_{1.5}$ regression, $\mathcal{L}_2$ (least squares) regression and median qunatile regression. The results show that our testing procedure has decent power for general M-estimation with moderate sample size. As expected, the regression with more robust loss function tends to have less power. The significance level is 10%. We also construct and examine a regression-coefficient-based CUSUM test and find that it has a significantly inferior power performance than the gradient-based test. See right panel of Figure 1. Additional empirical results show that under stationarity, our method is a little less powerful than the SQ method in [26] when investigating quantile regression but more powerful than the SCB method in [31] when investigating least squares regression. The detailed results and explanations are relegated to Section 5 of the online supplemental material.

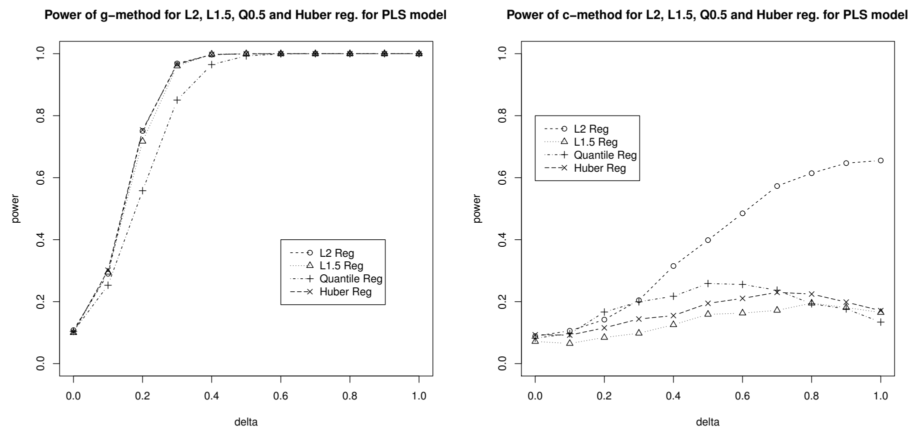

Fig 1: Simulated power with Υ_i following Model II for gradient-based method (left) and for coefficient-based method (right)

5. Data Analysis. In this section, we apply our method to the Hong Kong circulatory and respiratory data. It consists of daily measurements of pollutants and daily hospital admissions in Hong Kong between January 1, 1994 and December 31, 1995. This dataset has been analyzed under i.i.d. assumptions in [13], [14] and [10] among others. It has also been studied under locally stationary assumptions; see for example [31] and [30]. The aim of this data analysis is to capture the relationship between the daily total number of hospital admissions of circulation/respiration and the levels of pollutants such as sulphur dioxide (SO_2) (in micrograms per cubic metre), nitrogen dioxide (NO_2) (in micrograms per cubic metre) and dust (in micrograms per cubic metre). By fitting time-varying linear regression models, the results of [31] indicated the existence of change points in the least squares regression coefficients between January 1st, 1994, and December 31st, 1995. By carefully observing the patterns of regression coefficients and their simultaneous confidence band plotted in Figure 2 of their paper, it is difficult to tell whether there is a change point in the mean regression relationship between January 1st, 1995, and December 31st, 1995. To justify the necessity to use the PLS formulation of the errors and covariates, we first perform the tests in [32] and [11] for change points in the mean and autocovariances. The $p - values$ for no change point in mean for NO_2, SO_2 and dust are 0.5%, 5.15% and 0.025%, respectively. For NO_2 and dust, the $p - values$ for no change points in lag-1 autocovariance are 9.9% and 8%, respectively, while the $p - values$ for no change points in lag-2 autocovariance are 4.7% and 3.3%, respectively. These results show strong evidence that the data we considered are non-stationary. Thus, we consider the following model:

$$y_i = \beta_0 + \sum_{l=1}^{3} \beta_l x_{i,l} + \epsilon_i,$$

where y_i is the daily number of hospital admissions, and $\{x_{i,1}, 1 \le i \le n\}$ is the level of SO_2, $\{x_{i,2}, 1 \le i \le n\}$ is the level of NO_2, $\{x_{i,3}, 1 \le i \le n\}$ is the level of dust, and ϵ_i is a PLS noise. We test the null hypothesis that $\beta := (\beta_0, ..., \beta_3)'$ remains constant from January 1st, 1995 to December 31st, 1995 for both least squares regression and quantile regression. For quantile regression, we consider 7 different quantiles $0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$. For the least squares regression, we choose bandwidth $m = 20$. It turns out that the 90% critical value is 10532.89, and 95% critical value is 11973.6, while the test statistic is 10464.35. Our result shows that there is no evidence indicating that the relationship in mean between hospital admissions and pollutants changed in the year of 1995 under the PLS assumptions.

We also summarize the quantile regression results as follows:

Table 4
Structural Change Test for 0.2,0.3,0.4,0.5,0.6,0.7,0.8 quantiles. The null hypothesis of no structural change in the relationship between daily hospital admissions and pollutants levels is rejected for 0.2,0.3,0.4 quantiles at 5% significance level, rejected for 0.5 quantile at 10% significance level, and is not rejected for 0.6,0.7,0.8 quantiles.

Quantile	0.2**	0.3**	0.4**	0.5*	0.6	0.7	0.8
test statistics	61.27	79.74	85.53	81.84	74.52	78	70.27
m	13	12	7	12	12	12	17
h	0.84	0.67	0.16	0.36	0.26	0.16	0.27
90%	48.65	68.34	74.75	75.16	96.39	104.7	102.83
95%	53.83	78.23	85.33	82.92	112.42	120.74	115.41
SQ method	5.38	5.86	6.19	5.85	5.32	6.28	6.8

Our results show that for mid and low quantiles there are structural changes in the regression coefficients; while for high quantiles there are none. For 0.5 quantile, our results show that the p-value is between 5% and 10%. This is a potentially interesting finding which shows the influence of pollutants on low hospital admissions has changed while there is no change in the relationship on the high end. Consequently, it is appropriate to use a parametric model to fit the high quantile regression while nonparametric dynamic models are more appropriate to model the low quantiles. Such asymmetric behavior across different quantiles cannot be identified by the hypothesis testing procedures in mean regression proposed in [30]. The last line of the table lists the test statistics generated via SQ method. The 95% and 99% critical values of SQ test are 1.569 and 1.795, respectively. By Table 4, SQ method strongly rejects the null hypothesis at all quantiles considered in this paper. This is likely due to the violation of the strict stationarity assumptions in [26] for this data set, which makes SQ test overact to the spurious patterns of change points in regression coefficients caused by the non-stationary errors and covariates. As a result, it seems that the SQ method in [26] cannot detect the asymmetric behavior across different quantiles and it yields too significant testing results with too small p-values due to the non-stationarity of the errors and covariates of the regression.

6. Technical Appendix.

The proofs of the theoretical results have been moved to the supplemental material.

References.

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

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

[3] Arcones, M. A. (1996). The Bahadur-Kiefer representation of L_p regression estimators. *Econometric Theory*, 12:257–283.

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

[5] Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist*, 37:577–580.

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

[7] Bai, J. and Perron, P. (1998). Testing for and Estimation of Multiple Structural Changes. *Econometrica*, 66:47–79.

[8] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

[9] 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.

[10] Cai, Z., Fan, J. and Li, R. Z. (2000). Efficient Estimation and Inferences for Varying-coefficient Models. *Journal of the American Statistical Association*, 95:888–902.

[11] Dette, H. Wu, W. and Zhou, Z. (2015). Change Point Analysis of Second Order Characteristics in Non-stationary Time Series. arXiv:1503.08610.

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

[13] Fan, J. and Zhang, W. (1999). Statistical Estimation in Varying Coefficient Models. *Annals of Statistics*, 27: 1491–1518.

[14] Fan, J. and Zhang, W. (1999). Simultaneous Confidence Bands and Hypothesis Testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27: 715–731.

[15] Hansen, B. E. (2000). Testing for Structural Change in Conditional Models. *Journal of Econometrics*, 97:93–115.

[16] He, X. and Zhu, L. (2003). A Lack-of-Fit Test for Quantile Regression. *Journal of the American Statistical Association*, 98: 1013–1022.

[17] Juhl, T., Xiao, Z. (2009). Tests for Changing Mean with Monotonic Power. *Journal of Econometrics*,148(1): 14–24.

[18] Kiefer, J. (1967). On Bahadurs Representation of Sample Quantiles. *Ann. Math. Statist,* 38: 1323-1342.

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

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

[21] Ploberger, W. and Kramer, W. (1992). The CUSUM Test with OLS Residuals. *Econometrica*, 60: 271–285.

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

[23] Portnoy, S. (1991). Asymptotic Behavior of Regression Quantiles in Nonstationary, Dependent Cases. *J. Multi-variate Analysis*, 38: 100–113.

[24] Powell, J. L. (1991). Nonparametric and Semiparametric Methods in Economics: Estimation of Monotonic Regression Models under Quantile Restrictions *Cambridge: Cambridge University Press.*

[25] Proskova, Z., and Chochola, O. (2014). M-procedures for Detection of a Change under Weak Dependence. *Journal of Statistical Planning and Inference*, 149: 60–76.

[26] Qu, Z. (2008). Testing for Structural Change in Regression Quantiles. *Journal of Econometrics*, 148: 170–184.

[27] Su, L., and White, H. (2010). Testing Structural Change in Partially Linear Models. *S Econometric Theory*, 26: 1761–1806.

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

[29] Wu, W. B. (2007). M-estimation of Linear Models with Dependent Errors. *Annals of Statistics*, 35: 495–521.

[30] Zhang T., Wu W. B. (2012). Inference of Time-varying Regression Models. *The Annals of Statistics*, 40: 1376–1402.

[31] Zhou, Z. and Wu, W. B. (2010). Simultaneous Inference of Linear Models with Time-Varying Coefficients. *Journal of the Royal Statistical Society, series B.* , 72: 513–531.

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