INFERENCE FOR NON-STATIONARY TIME SERIES REGRESSION WITH/WITHOUT INEQUALITY CONSTRAINTS

ZHOU ZHOU [1]

University of Toronto

September 15, 2014

Abstract

We consider statistical inference for time series linear regression where the response and predictor processes may experience general forms of abrupt and smooth non-stationary behaviors over time. Meanwhile, the regression parameters may be subject to linear inequality constraints. A simple and unified procedure for structural stability check and parameter inference is proposed. In the case where the regression parameters are constrained, the proposed methodology is shown to be consistent whether or not the true regression parameters are on the boundary of the restricted parameter space via utilizing an asymptotically invariant geometric property of polyhedral cones.

1 Introduction

Consider the following time series linear regression

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \epsilon_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

where $\boldsymbol{\beta}_0 = (\beta_{1,0}, \beta_{2,0}, \cdots, \beta_{p,0})^\top$ is the p-dimensional vector of regression coefficients, $\{\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^\top\}_{i=1}^n$ are $p \times 1$ dimensional non-stationary time series of predictors, $\{\varepsilon_i\}_{i=1}^n$ are non-stationary error series satisfying $\mathbb{E}(\varepsilon_i|\mathbf{x}_i) = 0$ almost surely. Here $^\top$ denotes matrix or vector transpose.

There is an increasing need for non-stationary time series regression in statistics and various related fields(Starica and Granger 2005, Dahlhaus and Subba Rao 2006, Fan et al. 2003, Mikosch and Starica 2004, Mercurio and Spokoiny 2004, Koo and Linton 2012 and Vogt 2012, among others). Meanwhile, due to prior knowledge or mathematical requirement, in various applications the regression coefficient $\boldsymbol{\beta}_0$ is known to belong to a

[1] Corresponding author.

AMS 2000 subject classifications: Primary 62F30; secondary 62M10

Key words and phrases. Piecewise locally stationary time series, robust bootstrap, inequality constraints, structural stability, subsampling.

subset of $\mathbb{R}^p$. For instance, in econometric applications the sign of some $\beta_{i,0}$ may be predetermined by certain economic theory. Another example is the auto-regressive conditional heteroscedastic (ARCH) models where the model parameters are restricted to be nonnegative due to mathematical requirements. The purpose of the paper is to propose a simple bootstrap methodology for both structural stability checking and parameter inference of non-stationary time series regression with or without linear inequality constraints. The major contributions of the paper lie in the following three aspects:

First, we propose a new structural stability testing procedure for time series linear regression that is robust to general forms of abrupt and smooth non-stationary behaviors in the error and predictor processes. Since the seminal paper of Brown et al. (1975), there has been a large amount of work on issues related to regression stability tests in statistics and various applied fields. See for instance Andrews (1993), Kim and Siegmund (1989), Loader (1996) and Bai and Perron (1998) for some representative works. The classic idea to perform the latter tests is to normalize or pivotalize a certain norm (e.g. $\mathcal{L}^2$ or $\mathcal{L}^\infty$ norm) of a sequence of discrepancy statistics. An important observation is that, in order to realize the pivotalization, the error and/or the predictor processes are required to be stationary. On the other hand, however, for non-stationary time series regressions considered in this paper, we demonstrate that (see Theorem 1 in Section 3) in general the structural stability test statistics cannot be pivotalized. As a consequence the classic tests will generally result in biased conclusions when applied to non-stationary time series regression. In this paper we shall utilize an interesting finding that, for a wide class of non-stationary errors and predictors, the probabilistic behavior of the process of partial sample least squares estimators can be consistently mimicked by a progressive convolution of block least squares estimators and i.i.d. standard normal random variables. With the latter finding we device a simple bootstrap procedure that is consistent for structural stability testing for a wide class of non-stationary time series linear regression. It seems that the only related work in the literature is Zhou (2013) who considered testing structural changes in non-stationary time series via a convolution of the partial sum process of the series and i.i.d. normal random variables. However, the bootstrap process therein is tailored to the problem of testing changes in marginal distributions and cannot be applied to the general regression setting.

Second, we propose a new structural stability test for constrained linear regression. To our knowledge, the issue of structural stability checking for constrained regression has not yet been discussed in the literature. The major difficulties, in our opinion, lie in the following two aspects. First, the *p*artial *s*ample *c*onstrained *l*east *s*quares *e*stimators (PSCLSE) are difficult to normalize, even for stationary time series regression. Second, the probabilistic behaviors of the PSCLSE's depend heavily on the location of $\boldsymbol{\beta}_0$. The proposed bootstrap procedure in the previous paragraph successfully handles the first issue. Indeed, we do not seek to normalize the test statistic in our bootstrap. To deal with the second issue, we utilize an invariant geometric property of polyhedral cones that, for any $\boldsymbol{\beta}$ belonging to the cone and any fixed $\delta > 0$, the δ-neighborhoods of $a\boldsymbol{\beta}$ and $b\boldsymbol{\beta}$ in the cone are identical for sufficiently large $a \in \mathbb{R}$ and $b \in \mathbb{R}$. Note that the latter property holds whether or not $\boldsymbol{\beta}$ is on the boundary of the cone. With the latter invariant geometric property, we are able to device a bootstrap procedure for structural stability test of constrained non-stationary time series regression. The bootstrap is simple and is consistent no matter $\boldsymbol{\beta}_0$ in (1) is on boundary or in the interior of the restricted parameter space. However, note that $\boldsymbol{\beta}_0$ and p in (1) are fixed and are not varying with sample size n. In particular, we are not considering the cases where $\boldsymbol{\beta}_0$ is allowed to drift and converge to the boundary as n grows (See Remark 2 in Section 4.3 for a detailed discussion) or the number of predictors grows with the sample size.

Third, we demonstrate the bootstrap procedures proposed above as simple and unified methodologies for parameter inference of non-stationary time series regression with or without inequality constraints. It is well known that the nonparametric bootstrap (Efron, 1979) is inconsistent for constrained linear regression; see for instance Andrews (2000). Alternatively, Andrews (2000) listed three resampling methods which are consistent for a constrained simple linear regression of i.i.d. normal errors. Moreover, many resampling methods have been proposed for stationary time series linear regression. However, most, if not all, of the latter methods heavily depended on the fact the data generating mechanism does not change over time to produce repeated sub- or pseudo- samples. As a result the performances of those resampling methods are unclear for non-stationary time series regression. It is impossible for us to investigate the performance of each of the aforementioned resampling methods under non-stationarity in this paper; instead, we shall choose one popular and representative method, the subsampling (Politis et al. 1999), and investigate its behavior for constrained non-stationary linear regression. It is found that the subsampling fails to be consistent under our setting due to the changing distributional information in the subsamples over time. Alternatively, we show that the bootstrap procedures proposed in this paper serve as simple and robust methods for the inference of $\boldsymbol{\beta}_0$ when $\{\mathbf{x}_i\}$ and $\{\epsilon_i\}$ are non-stationary.

In many time series regression problems, one may find the traditional stationarity assumption on the regressors and errors restrictive. Much recent effort in time series analysis has been put in developing new models and methods for non-stationary processes. It seems that many of those models focused on time-varying spectral analysis. See for instance Priestley (1988), Dahlhaus (1997), Nason et al. (2000), Ombao et al. (2005) and Van Bellegem and Dahlhaus (2006), among others. For constrained time series regression in (1), however, non-stationary models from the spectral domain do not seem to be directly useful for an asymptotic theory. In this paper, we adopt the time domain modeling of non-stationary time series in Zhou (2013). The latter time domain approach provides a natural framework for an asymptotic theory under the current setting. Furthermore, the piecewise locally stationary (PLS) framework in Zhou (2013) allows very general forms of abrupt and smooth non-stationary temporal behaviors which could be flexible and realistic in many applications.

Statistical inference for inequality constrained regression problems are non-standard when $\boldsymbol{\beta}_0$ is on the boundary of the restricted parameter space. Generally speaking, in such cases the estimated $\hat{\boldsymbol{\beta}}$ fails to be asymptotically normal and the nonparametric bootstrap is inconsistent for the inference. In the literature, there have been many discussions on inequality constrained regressions. However, it seems that most of those results are for independent or normally distributed data. See for instance Liew (1976), Gouriéroux et al. (1982), Wolak (1987), Shapiro (1988) and Andrews (2000), among others. We also refer to the monographs of Robertson et al. (1988) and Silvapulle and Sen (2005) for more discussions and references.

The rest of the paper is organized as follows. Section 2 introduces and discusses the concepts of partial sample restricted estimators, metric projections and PLS time series models which are important for the subsequent analysis of the paper. In Section 3 we introduce structural stability tests for non-stationary time series regression with/without inequality constraints as well as the robust bootstrap for the tests. In Section 4, confidence region construction and testing for $\boldsymbol{\beta}_0$ will be discussed. In particular, the inconsistency of the subsampling for constrained non-stationary time series regression will be investigated in Section 4.2. The issue of tuning parameter selection will be discussed in Section 4.4. Moderate sample Monte Carlo experiments on the accuracy of the robust bootstrap will be conducted in Section 5. An ARCH model based analysis on 10-minute mark-dollar exchange rates will be performed in Section 6. Finally, the proofs of the theorems are put in Section 7. Auxiliary lemmas and the propositions are proved in the supplemental document Zhou (2014).

2 Preliminaries

In model (1), suppose that $A\boldsymbol{\beta}_0 \geq r$, where A is a $q \times p$ dimensional full rank matrix with $0 \leq q \leq p$. With an appropriate transformation of variables, the latter inequality constraint can be simplified into

$$\boldsymbol{\beta}_0^{(1)} \geq 0, \text{ where } \boldsymbol{\beta}_0 = ([\boldsymbol{\beta}_0^{(1)}]^\top, [\boldsymbol{\beta}_0^{(2)}]^\top)^\top \text{ and } \boldsymbol{\beta}_0^{(1)} \text{ is } q\text{-dimensional,}$$

or equivalently,

$$\boldsymbol{\beta}_0 \in Q, \text{ where } Q = \{\mathbf{x} = (x_1, \cdots, x_p)^\top : x_i \geq 0, i = 1, \cdots, q\}. \tag{2}$$

Without loss of generality, inequality constraint (2) is assumed throughout the paper. Note that the classic unconstrained regression (i.e. no constraints are put on the regression coefficient $\boldsymbol{\beta}_0$) corresponds to the case where $q = 0$ and $Q = \mathbb{R}^p$ in (2). As a further note, we point out here that the theory and methodology of the paper are applicable when the regression parameters are confined to any convex polyhedral cones. For the sake of presentational simplicity and clarity we shall stick to the cone Q in this article.

Statistical inference of (1) relies heavily on the assumption that the regression coefficients remain constant over the considered time span. Therefore testing structural stability is important in the analysis of time series regression. Consider the alternative model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_i + \epsilon_i, \quad i = 1, 2, \cdots, n, \tag{3}$$

where the regression coefficients $\boldsymbol{\beta}_i \in Q$ and may not stay constant over time. In this paper, we are interested in the following structural change testing problem

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_n \quad \longleftrightarrow \quad H_a : \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_{\lfloor kn \rfloor} \neq \boldsymbol{\beta}_{\lfloor kn \rfloor + 1} = \cdots = \boldsymbol{\beta}_n \tag{4}$$

for some unknown change point $k \in (0, 1)$, where $\lfloor x \rfloor$ denotes the integer part of x.

2.1 Partial sample restricted least squares estimator

To test structural stability of model (1) with constrained regression parameters, utilizing the constraints (2) is important as it reflects our prior knowledge of the regression coefficients and is part of the model. Additionally, making use of prior knowledge of $\boldsymbol{\beta}_0$ reduces the parameter space and hence could lead to a more sensitive detection of possible changes in the regression parameters. We observe that if the null model (1) is true, then the same linear regression relationship holds in every subinterval of the considered time span. The latter observation has been implemented widely in various structural stability tests. See for instance Ploberger et al. (1989) for change point test of unconstrained linear regression with i.i.d. errors and Andrews (1993) for structural stability test of regression models driven by Generalized Method of Moments (GMM). In this paper, we shall utilize the latter idea to test for structural changes in constrained linear regression. First we define partial sample restricted least squares estimators as follows.

For two time indices $1 \le t < s \le n$, define

$$\hat{\boldsymbol{\beta}}_{t,s} = \text{argmin}_{\boldsymbol{\beta}} \sum_{i=t}^{s} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \text{ subject to } \boldsymbol{\beta} \in Q. \tag{5}$$

For a positive definite symmetric matrix Σ, denote by $\mathcal{P}_{Q,\Sigma}(\cdot)$ the metric projection onto the cone Q with respect to the inner product $\langle x, y \rangle = x^\top \Sigma y$. By the Karush-Kuhn-Tucker multiplier, the solution to the above inequality constrained optimization problem is

$$\hat{\boldsymbol{\beta}}_{t,s} = \mathcal{P}_{Q,M_{t,s}}(\tilde{\boldsymbol{\beta}}_{t,s}), \tag{6}$$

where $M_{t,s} = \sum_{j=t}^{s} \mathbf{x}_j \mathbf{x}_j^\top / (s - t + 1)$ and $\tilde{\boldsymbol{\beta}}_{t,s}$ is the classic unconstrained least squares estimator of (5); i.e. $\tilde{\boldsymbol{\beta}}_{t,s}$ is the minimizer of (5) when $Q = \mathbb{R}^p$. In practice, the projection can be easily realized with quadratic programming.

To test H_0 against H_a, we define the test statistic

$$T_n = \max_{p+1 \le i \le n-p-1} \sqrt{n} t_i (1 - t_i) |\hat{\boldsymbol{\beta}}_{1,i} - \hat{\boldsymbol{\beta}}_{i+1,n}|, \text{ where} \tag{7}$$

$t_i = i/n$ and $|\mathbf{x}| = \sqrt{\mathbf{x}^\top \mathbf{x}}$ denotes the Euclidean norm. In particular, if $\boldsymbol{\beta}_0$ is unrestricted, i.e. $Q = \mathbb{R}^p$ in (2), then T_n is equivalent to

$$\tilde{T}_n = \max_{p+1 \le i \le n-p-1} \sqrt{n} t_i (1 - t_i) |\tilde{\boldsymbol{\beta}}_{1,i} - \tilde{\boldsymbol{\beta}}_{i+1,n}| \tag{8}$$

Clearly T_n and $\tilde{T}_n$ will be large if H_a holds true for some $k \in (0, 1)$. The test T_n is related to the partial sample Wald tests of Andrews (1993). The difference is that the latter paper utilizes a normalizing matrix to pivotalize the test statistics(a classic idea of the Wald tests). However, due to non-stationarity, the differences $\{\hat{\boldsymbol{\beta}}_{1,i} - \hat{\boldsymbol{\beta}}_{i+1,n}\}$ generally cannot be pivotalized under our setting. See also Section 3 for a more detailed explanation. As a result we suggest using the un-normalized T_n and $\tilde{T}_n$ for structural stability tests for non-stationary time series regression.

On the other hand, we observe that the estimators $\hat{\boldsymbol\beta}_{1,i}$ and $\hat{\boldsymbol\beta}_{i,n}$ are unstable for indices i near 0 and n, respectively. In particular, it can be shown that under H_0 the statistic $T_n' = \max_{p+1\le i\le n-p-1}\sqrt{n}|\hat{\boldsymbol\beta}_{1,i} - \hat{\boldsymbol\beta}_{i+1,n}|$ has large variabilities at small and large i's which asymptotically dominate its probabilistic behavior on the other time indices. In (7) and (8), the factor $t_i(1-t_i)$ down-weights those unstable estimators. We also refer to Csörgö and Horváth (1997) for a related discussion of T_n versus T_n'.

2.2 Metric projection onto convex polyhedral cones

For $\mathbf{x} = (x_1, \cdots, x_p)^\top \in \mathbb{R}^p$ and $\boldsymbol\beta = (\beta_1, \cdots, \beta_p)^\top \in Q$, define

$$\Theta_\Sigma(\boldsymbol\beta, \mathbf{x}) = \lim_{n\to\infty}(\mathcal{P}_{Q,\Sigma}(n\boldsymbol\beta + \mathbf{x}) - n\boldsymbol\beta).$$

It turns out the function $\Theta_\Sigma(\cdot,\cdot)$ plays an important role in the analysis of (1). As we observe from (6), the constrained least squares estimators are closely related to metric projections onto the convex cone Q. Hence investigating geometric properties of the latter projection is very important for the analysis of constrained linear regression. The following proposition explores several important properties of the metric projection as well as the function $\Theta_\Sigma(\cdot,\cdot)$.

Proposition 1. *For any $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$ and $\boldsymbol\beta \in Q$, we have that*

(i). for every $a \ge 0$, $\mathcal{P}_{Q,\Sigma}(a\mathbf{x}) = a\mathcal{P}_{Q,\Sigma}(\mathbf{x})$.

(ii). if the minimum eigenvalue of $\Sigma > 0$, then there exists an $A(\boldsymbol\beta, \mathbf{x}) < \infty$, such that $\mathcal{P}_{Q,\Sigma}(a_1\boldsymbol\beta + \mathbf{x}) - a_1\boldsymbol\beta = \mathcal{P}_{Q,\Sigma}(a_2\boldsymbol\beta + \mathbf{x}) - a_2\boldsymbol\beta$ for $a_1, a_2 \ge A(\boldsymbol\beta, \mathbf{x})$.

(iii). $|\mathcal{P}_{Q,\Sigma}(a\boldsymbol\beta+\mathbf{x}) - a\boldsymbol\beta|_\Sigma \le |\mathbf{x}|_\Sigma$ for every $a \in \mathbb{R}^+$, where $|\mathbf{x}|_\Sigma = \sqrt{\mathbf{x}^\top\Sigma\mathbf{x}}$ for any $\mathbf{x} \in \mathbb{R}^p$.

(iv). $|\Theta_\Sigma(\boldsymbol\beta, \mathbf{x}_1) - \Theta_\Sigma(\boldsymbol\beta, \mathbf{x}_2)|_\Sigma \le |\mathbf{x}_1 - \mathbf{x}_2|_\Sigma$.

(v). if $0 < r \le \sigma_{i,\min} \le \sigma_{i,\max} \le R < \infty$, $i = 1, 2$ and $|\Sigma_1 - \Sigma_2| \le \epsilon$, where $\sigma_{i,\min}$ and $\sigma_{i,\max}$ are the minimum and maximum eigenvalues of Σ_i, respectively, then there exists a constant $C(r, R, \boldsymbol\beta, \mathbf{x}) < \infty$, such that $|\Theta_{\Sigma_1}(\boldsymbol\beta, \mathbf{x}) - \Theta_{\Sigma_2}(\boldsymbol\beta, \mathbf{x})| \le C(r, R, \boldsymbol\beta, \mathbf{x})\sqrt{\epsilon}$.

(vi). if $|\mathbf{x}| \le D$ for some finite constant D, then $A(\boldsymbol\beta, \mathbf{x})$ and $C(r, R, \boldsymbol\beta, \mathbf{x})$ in (ii) and (v) can be written as $A(\boldsymbol\beta, D)$ and $C(r, R, \boldsymbol\beta, D)$, respectively.

As a side note, we point out here that properties (i) to (vi) can be shown to hold for any convex polyhedral cone. Proposition 1 reveals several important properties of the metric projection as well as the function $\Theta_\Sigma(\cdot,\cdot)$. In particular, properties (ii) and (iii) ensures that the function $\Theta_\Sigma(\cdot,\cdot)$ exists and is finite. Furthermore, property (ii) reflects the important observation that, for polyhedral cones, the topologies of neighborhoods of $a_1\boldsymbol{\beta}$ and $a_2\boldsymbol{\beta}$ are identical for sufficiently large a_1 and a_2 no matter $\boldsymbol{\beta}$ is on the boundary or in the interior of the cones. The latter invariant geometric property of polyhedral cones is the key observation that leads to a unified treatment of constrained statistical inference problems in the current paper. Additionally, properties (iv) and (v) indicate that the function $\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x})$ is Hölder continuous in Σ and $\mathbf{x}$. However, we shall point out that generally $\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x})$ is not continuous in $\boldsymbol{\beta}$, as shown in the following example.

Example 1. *Consider the case where $\Sigma = I_{p\times p}$ and $I_{p\times p}$ is the $p\times p$ identity matrix. Then it is straightforward to derive that*

$$\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x}) = (\theta(\beta_1,x_1),\theta(\beta_2,x_2),\cdots,\theta(\beta_q,x_q),x_{q+1},\cdots,x_p)^\top,$$

where $\theta(\beta,x) = xI[\beta>0] + \max\{x,0\}I[\beta=0]$ and $I[\cdot]$ is the indictor function. Observe that $\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x}_1+\mathbf{x}_2) \neq \Theta_\Sigma(\boldsymbol{\beta},\mathbf{x}_1)+\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x}_2)$ when $\beta_i = 0$ for some $1\le i\le q$. Further note that the function $\theta(\beta,x)$ is not continuous in β at 0. Hence $\Theta_\Sigma(\boldsymbol{\beta},\mathbf{x})$ is not a continuous function of $\boldsymbol{\beta}$.

2.3 Piecewise locally stationary time series models

As we pointed out in the Introduction, many time series regressions involve non-stationary response or predictor processes. In particular, both smooth and abrupt changes in the data generating mechanisms of such processes can occur over time. In this section we shall introduce a class of piecewise locally stationary (PLS) time series models (Zhou, 2013). The PLS class provides a flexible nonparametric device to model the complex temporal dynamics of the error or predictor processes in (1).

Definition 1. *(Zhou 2013) Write $\mathbf{Z}_i = (\mathbf{x}_i^\top,\epsilon_i)^\top$. We call $\{\mathbf{Z}_i\}_{i=1}^n$ piecewise locally stationary (PLS) with r break points if there exist constants $0=b_0<b_1<\cdots<b_r<b_{r+1}=1$ and nonlinear filters $G_0,G_1,\cdots,G_r$, such that*

$$\mathbf{Z}_i = G_j(t_i,\mathcal{F}_i),\ \text{if } b_j < t_i \le b_{j+1}, \tag{9}$$

where $t_i = i/n$, $\mathcal{F}_i = (\varepsilon_i,\varepsilon_{i-1},\cdots,\varepsilon_0,\cdots)$ and ε_i's are i.i.d. random variables.

Typically the break times $b_1,\cdots,b_r$ and the number of breaks r are unknown. In (9), write $G_j(\cdot) = (G_{1,j}^\top(\cdot),G_{2,j}(\cdot))^\top$, where $G_{2,j}(\cdot)$ is real valued. Then $\mathbf{x}_i = G_{1,j}(t_i,\mathcal{F}_i)$ and

$\epsilon_i = G_{2,j}(t_i, \mathcal{F}_i)$. For any $i \in \{1, 2, \cdots, n\}$, let ζ_i be the index j such that $b_j < t_i \leq b_{j+1}$. Then (9) can be written as $\mathbf{Z}_i = G_{\zeta_i}(t_i, \mathcal{F}_i)$.

We see from the above definition that the process $\{\mathbf{Z}_i\}$ can undergo abrupt changes at the break points $b_1, \cdots, b_r$. On the other hand, if the functions $G_j(t, \cdot)$ $j = 0, 1, \cdots, r$ are smooth in t, then the data generating mechanism of $\{\mathbf{Z}_i\}$ evolves smoothly between adjacent break points. As a result the PLS class constitutes a relatively large class of non-stationary time series models which allows the data generating mechanism to change flexibly over time, both smoothly and abruptly. In particular, if $r = 0$, then model (9) is reduced to a locally stationary process in the sense of Zhou and Wu (2009). On the other hand, if all $G_j(t, \cdot)$ $j = 0, 1, \cdots, r$ in (9) are independent of t, then we obtain piecewise stationary time series models which was considered in Davis et al. (2006), among others. We observe that the PLS class combines the latter two types of non-stationarity in time series and can be flexible and realistic in many applications.

Appropriate dependence measures for the PLS class are necessary for an asymptotic theory. Here we use the idea of Wu (2005) and measure the temporal dependence of a PLS time series by quantifying the differences in the filters' outputs when past inputs of the system are changed to i.i.d. copies. Specifically, if $\max_i \|\mathbf{Z}_i\|_h < \infty$ for some $h \geq 1$, then for every integer $k \geq 0$, we define the dependence measures

$$\delta_{\mathbf{Z}}(k, h) = \max_{0 \leq i \leq r} \sup_{b_i \leq t \leq b_{i+1}} \|G_i(t, \mathcal{F}_k) - G_i(t, \{\varepsilon_k, \varepsilon_{k-1}, \cdots, \varepsilon_0^*, \mathcal{F}_{-1}\})\|_h, \tag{10}$$

where ε_0^* is independent of $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ and is identically distributed as ε_0. Let $\delta_{\mathbf{Z}}(k, h) = 0$ if $k < 0$. In (10), $\|\cdot\|_h$ denotes the $\mathcal{L}^h$ norm of a random variable or vector; i.e. $\|\cdot\|_h = (\mathbb{E}[|\cdot|^h])^{1/h}$. Write $\|\cdot\| := \|\cdot\|_2$. We see from the above definition that $\delta_{\mathbf{Z}}(k, h)$ measures the maximum change in the system's outputs when the inputs of the system k steps before are changed to an i.i.d. copy. When $\delta_{\mathbf{Z}}(k, h)$ decays fast to 0 as k goes large, we have short range dependence of the PLS time series $\{\mathbf{Z}_i\}$. We refer to Zhou (2013) for more discussions and examples of the PLS class and the associated dependence measures.

3 Stability tests with/without inequality constraints

3.1 Asymptotic distributions of the tests

The following Condition (A) is needed to establish the asymptotic theory of the paper.

(A1) Let $M(t) = \mathbb{E}[G_{1,j}(t, \mathcal{F}_0)G_{1,j}^\top(t, \mathcal{F}_0)]$ if $b_j < t \leq b_{j+1}$ and $M(0) = \mathbb{E}[G_{1,0}(0, \mathcal{F}_0)G_{1,0}^\top(0, \mathcal{F}_0)]$. Assume that the minimum eigenvalue of $M(t)$ is bounded away from 0 on $[0, 1]$.

(A2) $\|G_i(t,\mathcal{F}_0)\|_4 < \infty$ for all $i \in [0,r]$ and $t \in [b_i, b_{i+1}]$. The process $\{\mathbf{Z}_i\}$ is piecewise stochastic Lipschiz continuous. Namely for all $i \in \{0,1,\cdots,r\}$ and all $t,s \in [b_i,b_{i+1}]$, $t \neq s$, we have

$$\|G_i(t,\mathcal{F}_0) - G_i(s,\mathcal{F}_0)\|_4/|t-s| \le C$$

for some finite constant C.

(A3) $\delta_{\mathbf{W}}(k,8) = O([(k+1)\log(2+k)]^{-2})$, where $\mathbf{W}(t,\mathcal{F}_i) = G_{1,j}(t,\mathcal{F}_i)G_{2,j}(t,\mathcal{F}_i)$ if $t \in (b_j, b_{j+1}]$, $j = 0,\cdots,r$.

(A4) $\delta_{\mathbf{x}}(k,8) = O([(k+1)\log(2+k)]^{-2})$.

(A5) Define the long-run covariance function of $\{\mathbf{W}_i := \mathbf{x}_i\varepsilon_i\}$

$$\Psi(t) = \sum_{i=-\infty}^{\infty} \mathrm{Cov}[G_{1,j}(t,\mathcal{F}_0)G_{2,j}(t,\mathcal{F}_0), G_{1,j}(t,\mathcal{F}_i)G_{2,j}(t,\mathcal{F}_i)]$$

if $t \in (b_i, b_{i+1}]$. Let $\Psi(0) = \lim_{t\downarrow 0}\Psi(t)$. Assume that the minimum eigenvalue of $\Psi(t)$ is bounded away from 0 on $[0,1]$.

A few remarks on the regularity conditions are in order. Condition (A1) requires that the design matrix is non-singular. The Lipschiz continuity condition in (A2) grantees that the data generating mechanism of $\{\mathbf{Z}_i\}$ changes smoothly between adjacent break points. Conditions (A3) and (A4) require the series $\{\mathbf{x}_i\epsilon_i\}$ and $\{\mathbf{x}_i\}$ to be short range dependent, respectively. Finally, condition (A5) is mild and it means that the long-run covariance function of $\{\mathbf{x}_i\epsilon_i\}$ is non-singular on $[0,1]$. The following theorem states the limiting null distribution of the structural stability tests T_n and $\tilde{T}_n$.

Theorem 1. *For $t \in [0,1]$, define $M^*(t) = \int_0^t M(s)\,ds/t$ and $M^o(t) = \int_t^1 M(s)\,ds/(1-t)$, $0 \le t \le 1$. Then under condition (A) and the null hypothesis H_0, we have*

$$\begin{aligned} T_n &\Rightarrow \sup_{0\le s\le 1} |\Theta_{M^*(s)}(\boldsymbol{\beta}_0, (1-s)[M^*(s)]^{-1}U(s)) - \Theta_{M^o(s)}(\boldsymbol{\beta}_0, s[M^o(s)]^{-1}[U(1)-U(s)]|, \\ \tilde{T}_n &\Rightarrow \sup_{0\le s\le 1} |(1-s)[M^*(s)]^{-1}U(s) - s[M^o(s)]^{-1}[U(1)-U(s)]|, \end{aligned} \tag{11}$$

where $U(s)$ is a p-dimensional zero-mean Gaussian process with covariance function $\gamma(t,s) = \int_0^{\min(t,s)} \Psi(r)\,dr$ and '$\Rightarrow$' denotes convergence in distribution.

When $\boldsymbol{\beta}_0$ is unconstrained, we observe from Theorem 1 that the null distribution of $\tilde{T}_n$ is determined by that of a Gaussian process $\{U(s)\}_{s=0}^1$ with a very complicated covariance function. In particular, the covariance matrices $\Psi(s)$ can change smoothly and flexibly between adjacent break points and can jump at the points $b_1, \cdots, b_r$. As a consequence the classic idea of normalizing or pivotalizing the test statistic is not operational for the non-stationary time series considered in the current paper. More specifically, Let $\hat{W}_i = \hat{\epsilon}_i \mathbf{x}_i$, where $\hat{\epsilon}_i = y_i - \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}_n$ are the residuals of the regression. Define the lag window estimator of long-run covariance matrix

$$\hat{\Sigma}_m = \frac{m}{n-m+1} \sum_{j=1}^{n-m+1} \Big(\frac{1}{m} \sum_{i=j}^{j+m-1} \hat{W}_i \Big) \Big(\frac{1}{m} \sum_{i=j}^{j+m-1} \hat{W}_i \Big)^\top,$$

where m is a user-chosen window size satisfying $m \to \infty$ with $m/n \to 0$. The classic way to perform the test $\tilde{T}_n$ is to normalize it with $[\hat{M}^*(1)]^{-1} \hat{\Sigma}_m^{1/2}$, where $\hat{M}^*(1) = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top / n$. Specifically, let

$$\tilde{T}_n' = \max_{p+1 \le i \le n-p-1} \sqrt{n} t_i (1-t_i) \Big| \hat{\Sigma}_m^{-1/2} \hat{M}^*(1) (\tilde{\boldsymbol{\beta}}_{1,i} - \tilde{\boldsymbol{\beta}}_{i+1,n}) \Big|.$$

If the process $\{\mathbf{Z}_i\} = \{(\mathbf{x}_i^\top, \epsilon_i)^\top\}$ is stationary, then under H_0 and Condition (A)

$$\tilde{T}_n' \Rightarrow \sup_{0 \le s \le 1} \Big| (1-s) B(s) - s[B(1) - B(s)] \Big| = \sup_{0 \le s \le 1} \Big| B(s) - sB(1) \Big|,$$

where $\{B(s)\}_{s=0}^1$ is the standard p-dimensional Brownian motion on $[0,1]$. And we observe that the normalized test statistic is asymptotically pivotal. However, when the regressors or the errors are non-stationary as considered in this paper, we have by a similar argument as that of Lemma 2 in the supplemental document Zhou (2014) that, under H_0, $[\hat{M}^*(1)]^{-1} \hat{\Sigma}_m^{1/2} \to [M^*(1)]^{-1} [\int_0^1 \Psi(s)\, ds]^{1/2}$ in probability. Hence, by Theorem 1,

$$\tilde{T}_n' \Rightarrow \sup_{0 \le s \le 1} \Big| [\int_0^1 \Psi(t)\, dt]^{-1/2} M^*(1) \{ (1-s) [M^*(s)]^{-1} U(s) - s [M^o(s)]^{-1} [U(1) - U(s)] \} \Big|$$

is no longer asymptotically pivotal due to the fact that the functions $M(t)$ and $\Psi(t)$ are time-varying. Generally speaking, we observe from Theorem 1 that the time-varying second order structures in the error and predictor processes post new challenges for change point tests of non-stationary time series linear regression.

When $\boldsymbol{\beta}_0$ is constrained, we observe from Theorem 1 that the null distribution of T_n depends critically on the location of $\boldsymbol{\beta}_0$. More specifically, if $\boldsymbol{\beta}_0$ sits in the interior of the convex cone Q, then the null distribution for T_n is the same as that of $\tilde{T}_n$, which is the maximum norm of a Gaussian process with complex covariance structure. On the other hand, however, if $\boldsymbol{\beta}_0$ is on the boundary of Q, then for any $s \in (0,1)$, the distribution of $\Theta_{M^*(s)}(\boldsymbol{\beta}_0, (1-s)[M^*(s)]^{-1} U(s))$ in (11) will have non-zero mass on the latter boundary. As a result statistical inference in such a case is non-standard and will be more delicate.

3.2 The robust bootstrap without inequality constraints

It might be tempting to use the asymptotic distributions in (11) to perform the structural stability test T_n or $\tilde{T}_n$. However, two major factors impede the practical applicability of the latter idea. First, the asymptotic behavior of T_n depends critically on the location of the unknown regression parameter $\boldsymbol{\beta}_0$ under H_0. In practice, it may be possible to test and determine the location of $\boldsymbol{\beta}_0$ under H_0. However, note that $\Theta_\Sigma(\boldsymbol{\beta}, \mathbf{x})$ is not continuous in $\boldsymbol{\beta}$. Consequently the latter procedure will typically produce extra uncertainty and complication to the structural stability test and hence lead to inaccurate testing results for finite samples. Second, the null distributions of the tests involve the Gaussian process $\{U(s)\}_{s=0}^1$ with very complicated covariance structure. In particular, the covariance matrices $\Psi(s)$ near the unknown break points $b_1, \cdots, b_r$ cannot be accurately estimated by standard nonparametric methods. Therefore in practice it is difficult to directly simulate the Gaussian process $U(s)$.

Summarizing the above discussion, to test H_0, it is desirable to have a simple procedure which works uniformly regardless of the location of $\boldsymbol{\beta}_0$ and avoids direct estimation of the complex covariance matrices $\Psi(s)$ of $U(s)$. In this and the next subsection we shall propose a robust bootstrap methodology for the purpose of testing H_0 with/without inequality constraints. Here the meaning of robustness is two fold. First, the robustness refers to consistency of the our methodology under general smooth and abrupt non-stationary behaviors of the predictor and error processes; second, the robustness refers to consistency of the bootstrap whether or not the underlying regression coefficients are on the boundary of the cone Q.

In this subsection we first consider testing $\tilde{T}_n$, the structural stability test without inequality constraints. There are two main reasons for considering $\tilde{T}_n$ separately. First, note that $\hat{\boldsymbol{\beta}}_{i,j}$ is a metric projection of the unconstrained least squares estimator $\tilde{\boldsymbol{\beta}}_{i,j}$. Hence to mimic the joint probabilistic behaviors of the partial sample constrained estimators $\{\hat{\boldsymbol{\beta}}_{1,i}, \hat{\boldsymbol{\beta}}_{i+1,n}\}_{i=1}^n$, it is helpful to first bootstrap the process of the unconstrained estimators $\{\tilde{\boldsymbol{\beta}}_{1,i}, \tilde{\boldsymbol{\beta}}_{i+1,n}\}_{i=p+1}^{n-p-1}$ and then mimic the projection operator. Second, as we discussed in the Introduction, for most commonly used structural change tests for time series regression, the predictor and/or the error processes are required to be stationary and the latter tests typically fail under non-stationarity. In view of the large amount of unconstrained non-stationary time series regressions performed in various fields, the test $\tilde{T}_n$ can be of separate interest. The following Theorem 2 is a key theorem which summarizes the robust bootstrap for the unconstrained test of structural change.

For a fixed positive integer m and $1 \le j \le m^*$ with $m^* = n - m + 1$, define

$$\Upsilon_{m,n}(j) = \sqrt{m} \sum_{k=1}^{j} M_{k,k+m-1}(\tilde{\boldsymbol{\beta}}_{k,k+m-1} - \tilde{\boldsymbol{\beta}}_{1,n}) V_k / \sqrt{m^*}, \tag{12}$$

where V_k's are i.i.d. standard normal random variables and are independent of $\{\mathbf{Z}_i\}_{i=1}^n$. Recall the definition of $M_{t,s}$ in (6).

Theorem 2. *For $\alpha \in (0,1)$, let $c_{m,\alpha}$ be the $(1-\alpha)$th quantile of*

$$\Xi_{m,n} := \max_{1 \le j \le n-2m} \left| (1 - t_{j+m-1}) M_{1,j+m-1}^{-1} \Upsilon_{m,n}(j) - t_{j+m} M_{j+m,n}^{-1} [\Upsilon_{m,n}(m^*) - \Upsilon_{m,n}(j)] \right|$$

conditional on $\{\mathbf{Z}_i\}_{i=1}^n$. Assume that condition (A) holds and that $m \to \infty$ with $m/n \to 0$.
(i) : Under H_0, we have

$$\mathbb{P}(\tilde{T}_n \ge c_{m,\alpha}) \to \alpha \text{ as } n \to \infty. \tag{13}$$

(ii): Under the conditions that H_a holds with $k \in (0,1)$ and that the break size $|\boldsymbol{\beta}_{\lfloor kn \rfloor + 1} - \boldsymbol{\beta}_{\lfloor kn \rfloor}| = n^{-1/2} L_n$ with $|L_n| \to \infty$, we have

$$\mathbb{P}(\tilde{T}_n \ge c_{m,\alpha}) \to 1 \text{ as } n \to \infty. \tag{14}$$

Part (i) of Theorem 2 shows that the conditional distribution of $\Xi_{m,n}$ coincides with the law of $\tilde{T}_n$ asymptotically. Hence one could test the null hypothesis H_0 by generating a large (say of size 1500) sample of conditionally i.i.d. copies of $\{\Xi_{m,n}\}$ and obtain the critical values of the test via the empirical distribution of the latter sample. The associated process $\{\Upsilon_{m,n}(j)\}$ is simple and it involves only one tuning parameter m. In particular, generating $\{\Upsilon_{m,n}(j)\}$ does not require estimating the very complex covariance function $\{\Psi(t)\}_{t=0}^1$. The following are the detailed steps of implementation in practice.

(1). Select block size m by the minimum volatility method stated in Section 4.4.

(2). Generate i.i.d. standard normal random variables and obtain B (say 1500) conditionally i.i.d. copies $\{\Upsilon_{m,n}^{(i)}(j)\}_{j=1}^{m^*}$ of $\{\Upsilon_{m,n}(j)\}_{j=1}^{m^*}$, $i = 1, 2, \cdots, B$.

(3). Calculate $\Xi_{m,n,i} := \max_{1 \le j \le n-2m} \left| (1 - t_{j+m-1}) M_{1,j+m-1}^{-1} \Upsilon_{m,n}^{(i)}(j) - t_{j+m} M_{j+m,n}^{-1} [\Upsilon_{m,n}^{(i)}(m^*) - \Upsilon_{m,n}^{(i)}(j)] \right|$, $i = 1, \cdots, B$. Let $\Xi_{(1)} \le \cdots \le \Xi_{(B)}$ be the ordered statistics of $\{\Xi_{m,n,i}\}$.

(4). Reject the null hypothesis H_0 at level α when $\tilde{T}_n > \Xi_{(\lfloor (1-\alpha)B \rfloor)}$. The p-value of test can be approximated by $1 - i/B$, where i is the largest index such that $\Xi_{(i)} \le \tilde{T}_n$.

As we observe from Theorem 1, to perform $\tilde{T}_n$ under non-stationarity the major difficulty lies in mimicking the Gaussian process $\{U(t)\}$ with very complicated covariance structure. By Lemma 4 in the supplemental document Zhou (2014), we have

$$\Upsilon_m(t) \Rightarrow U(t) \text{ on } \mathcal{C}[0,1], \tag{15}$$

where $\Upsilon_{m,n}(t)$ is the linear interpolation of the sequence $\{\Upsilon_{m,n}(j)\}_{j=1}^{m^*}$. Recall that '$\Rightarrow$' denotes weak convergence. The functional convergence in (15) is an important result that leads to the validity of the robust bootstrap in part (i) of Theorem 2. The surprising finding in (15) is that a simple progressive convolution of the unconstrained partial sample estimators $\tilde{\boldsymbol{\beta}}_{k,k+m-1}$ and i.i.d. standard normal random variables can consistently mimic the complex behavior of the process $\{U(s)\}_{s=0}^1$ with both abruptly and slowly changing covariance structure. On the other hand, part (ii) of Theorem 2 investigates the power behavior of the robust bootstrap. We claim that the robust bootstrap achieves asymptotic power 1 as long as the magnitude the break is greater than $n^{-1/2}$. It is well-known in the literature that the $n^{-1/2}$ rate is the optimal rate of local alternative detection for many parametric and nonparametric change-point tests. Therefore (ii) of Theorem 2 states that the robust bootstrap procedure is rate-optimal in terms of local alternative detection.

3.3 The robust bootstrap with inequality constraints

In this subsection we shall implement T_n, the structural stability test with inequality constraints. As we observed from Theorem 1, T_n has complicated behavior on the boundary of Q when one or more of the inequality constraints are binding. To overcome this difficulty, we observe from (6) that the constrained least squares estimators $\hat{\boldsymbol{\beta}}_{i,j}$ are metric projections of the unconstrained estimators $\tilde{\boldsymbol{\beta}}_{i,j}$. On the other hand, by Theorem 2 in the previous subsection, we have that the joint probabilistic behavior of the process of the unconstrained estimators $\{\sqrt{n}t_i(1-t_i)(\tilde{\boldsymbol{\beta}}_{1,i}-\boldsymbol{\beta}_0, \tilde{\boldsymbol{\beta}}_{i+1,n}-\boldsymbol{\beta}_0)\}$ can be consistently mimicked by the process

$$\{\Xi_{m,n}(j) = (\Xi_{m,n,1}(j), \Xi_{m,n,2}(j)) := ((1-t_j)M_{1,j}^{-1}\Upsilon_{m,n}(j), t_j M_{j+1,n}^{-1}[\Upsilon_{m,n}(m^*) - \Upsilon_{m,n}(j)])\}.$$

Hence to perform the test T_n, one only has to consistently mimic the projection operators. Now we observe that, for sufficiently large n,

$$\begin{aligned} \sqrt{n}t_i(1-t_i)\hat{\boldsymbol{\beta}}_{1,i} &= \mathcal{P}_{Q,M_{1,i}}(\sqrt{n}t_i(1-t_i)\boldsymbol{\beta}_0 + (1-t_i)M_{1,i}^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{i}\mathbf{W}_j) \\ &\approx \mathcal{P}_{Q,M_{1,i}}(\sqrt{n}t_i(1-t_i)\boldsymbol{\beta}_0 + \Xi_{m,n,1}(i)). \end{aligned} \tag{16}$$

Similar equality holds for $\hat{\boldsymbol{\beta}}_{i+1,n}$. At first glance one could simply replace $\boldsymbol{\beta}_0$ in (16) by a consistent estimator under H_0, say $\hat{\boldsymbol{\beta}}_{1,n}$, to mimic the projection. However, the latter naive operation is inconsistent. The reason is that, under H_0,

$$\sqrt{n}\hat{\boldsymbol{\beta}}_{1,n} = \sqrt{n}\boldsymbol{\beta}_0 + O_{\mathbb{P}}(1) \tag{17}$$

and the $O_{\mathbb{P}}(1)$ error term is not negligible. To solve this problem, the invariant geometric property (ii) in Proposition 1 plays a key roll. The latter property insures that, no matter where $\boldsymbol{\beta}_0$ is in the cone Q, the function $\mathcal{P}_{Q,\Sigma}(a\boldsymbol{\beta}_0 + \mathbf{x}) - a\boldsymbol{\beta}_0$ keeps the same value for sufficiently large a. Hence instead of simply replacing $\boldsymbol{\beta}_0$ by $\hat{\boldsymbol{\beta}}_{1,n}$, one could replace $\boldsymbol{\beta}_0$ in (16) by $\hat{\boldsymbol{\beta}}_{1,n}/l_n$ for some $l_n \to \infty$ with $l_n/\sqrt{n} \to 0$. Now the magnitude of the error reduces to $O_{\mathbb{P}}(1/l_n)$ which converges to 0 in probability and the bootstrap is consistent.

The choice of l_n involves balancing two competing forces. On one hand, the error introduced by $\hat{\boldsymbol{\beta}}_{1,n}/l_n$ in (16) is of the order $1/l_n$. This error is directly related to the accuracy of the test when $\boldsymbol{\beta}_0$ is on the boundary of the cone. On the other hand, if $\boldsymbol{\beta}_0 \neq 0$, then the signal to noise ratio in (16) after the replacement is of the order $\sqrt{n}/l_n$. The latter ratio decides the accuracy of the robust bootstrap when $\boldsymbol{\beta}_0$ is in the interior of Q. Since there is no prior knowledge on where $\boldsymbol{\beta}_0$ is, to balance those two forces, l_n could be chosen as $n^{1/4}$. All the above discussions lead to the following robust bootstrap process $\{\Lambda_{m,n}(j) = (\Lambda^*_{m,n}(j), \Lambda^o_{m,n}(j))\}_{t_j \in (0,1)}$ in Theorem 3.

Theorem 3. *For $\alpha \in (0,1)$, let $c^R_{m,\alpha}$ be the $(1-\alpha)$th quantile of $\max_{1\le j\le n-2m} |\Lambda^*_{m,n}(j) - \Lambda^o_{m,n}(j)|$ conditional on $\{\mathbf{Z}_i\}_{i=1}^n$, where*

$$\begin{aligned}
\Lambda^*_{m,n}(j) &= \mathcal{P}_{Q,M_{1,j+m-1}}(n^{1/4} t_{j+m-1}(1-t_{j+m-1})\hat{\boldsymbol{\beta}}_{1,n} + (1-t_{j+m-1})M^{-1}_{1,j+m-1}\Upsilon_{m,n}(j)),\\
\Lambda^o_{m,n}(j) &= \mathcal{P}_{Q,M_{j+m,n}}(n^{1/4} t_{j+m-1}(1-t_{j+m-1})\hat{\boldsymbol{\beta}}_{1,n} + t_{j+m-1}M^{-1}_{j+m,n}[\Upsilon_{m,n}(m^*) - \Upsilon_{m,n}(j)]).
\end{aligned}$$

Assume that the limiting distribution of T_n in (11) is continuous at its $(1-\alpha)$th quantile. Then under the conditions of Theorem 2, we have (i) and (ii) of the latter theorem hold with $\tilde{T}_n$ and $c_{m,\alpha}$ therein replaced by T_n and $c^R_{m,\alpha}$, respectively.

Theorem 3 claims that the process $\{\Lambda_{m,n}(j) = (\Lambda^*_{m,n}(j), \Lambda^o_{m,n}(j))\}$ consistently mimics the probabilistic behavior of $\{(\hat{\boldsymbol{\beta}}_{1,i}, \hat{\boldsymbol{\beta}}_{i+1,n})\}$. Hence in practice one could generate a large sample of i.i.d. copies of $\{\Lambda_{m,n}(j)\}$ and compute $\max_{1\le j\le n-2m} |\Lambda^*_{m,n}(j) - \Lambda^o_{m,n}(j)|$. Critical values of the test T_n can then be obtained by the empirical quantiles of the generated sample. The detailed implementation is very similar to steps (1)-(4) in the previous subsection and hence is omitted here. An important observation here is that one does not need to know the location of the underlying $\boldsymbol{\beta}_0$ nor the complicated covariance function $\Psi(t)$ to generate the bootstrap process $\{\Lambda_{m,n}(j)\}$. Instead, one just simply repeatedly generate i.i.d. normal random variables, convolute them with the partial sample least squares estimators and then project onto the cone Q. Finally, (ii) of Theorem 3 insures that the proposed bootstrap procedure is rate-optimal in terms of local alternative detection.

Remark 1. Due to the projection operator, the distributions of $\Lambda^*_{m,n}(j)$ and $\Lambda^o_{m,n}(j)$ are discontinuous at 0. In Theorem 3 we require that the limiting distribution of T_n is continuous at its $(1-\alpha)$th quantile. Note that there are at most countably many discontinuous points for a distribution. When verifying the latter continuity condition in practice, one can simply plot the empirical cumulative distribution function (CDF) of $\max_{1\le j\le n-2m} |\Lambda^*_{m,n}(j) - \Lambda^o_{m,n}(j)|$ from the bootstrap samples and check whether the latter empirical CDF is continuous at its $(1-\alpha)$th quantile.

4 Parameter inference

After the linear model (1) is validated, statistical inference such as confidence region construction and hypothesis testing can be performed on the regression parameters $\boldsymbol{\beta}_0$. When $\boldsymbol{\beta}_0$ sits on the boundary of the parameter space Q, it is well known that classic asymptotic theory for statistical inference fails and the nonparametric bootstrap is inconsistent for the inference of $\boldsymbol{\beta}_0$ (e.g. Andrews, 2000). A popular alternative to the bootstrap in the latter cases is subsampling, which is shown to be consistent when the sample is i.i.d. or stationary; see for instance Politis et al. (1999). In this section we shall investigate the asymptotic behaviors of $\tilde{\boldsymbol{\beta}}_{1,n}$ and $\hat{\boldsymbol{\beta}}_{1,n}$ for non-stationary time series. In particular, we shall show that the subsampling is not robust to piecewise local stationarity in the predictors and errors. On the other hand, however, the robust bootstrap proposed in Section 3.2 provides a simple and unified methodology for robust parameter inference of $\boldsymbol{\beta}_0$ under non-stationarity and temporal dependence.

4.1 Asymptotic distributions

For $\mathbf{x} = (x_1, \cdots, x_p)^\top \in \mathbb{R}^p$ and a set $A = \{a_1, \cdots, a_r\} \subset \{1, 2, \cdots, p\}$ with $a_1 < \cdots < a_r$, write $\mathbf{x}^{(A)} = (x_{a_1}, x_{a_2}, \cdots, x_{a_r})^\top$. In many applications, one may want to test the null hypothesis

$$H_0^A : \boldsymbol{\beta}_0^{(A)} = 0 \text{ against the alternative } H_a^A : \boldsymbol{\beta}_0 \in Q \text{ but } \boldsymbol{\beta}_0^{(A)} \neq 0.$$

We emphasize that, under H_0^A, the parameter space is $Q \bigcap Q^{(A)}$ with $Q^{(A)} = \{\mathbf{x} : \mathbf{x}^{(A)} = 0\}$ since the inequality constraint $\boldsymbol{\beta}_0 \in Q$ is assumed to be true throughout the paper. In particular, in contrary to most conventional testing problems, a subset of the regression parameters may be subject to inequality constraints under H_0^A. A classic way to test the latter hypothesis is by analyzing the residual sum of squares (RSS). Specifically we define the RSS test statistic

$$T_1^A = \text{RSS}_0 - \text{RSS}_a \tag{18}$$

where RSS_0 and RSS_a are the RSS under H_0^A and H_a^A, respectively. A large T_1^A indicates violation of H_0^A.

Theorem 4. *Write $\tilde{\boldsymbol{\beta}}_n = \tilde{\boldsymbol{\beta}}_{1,n}$ and $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_{1,n}$. Under Conditions (A1) to (A4) and model* (1)*, we have (i):*

$$\sqrt{n}[\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0] \Rightarrow [M^*(1)]^{-1}U \text{ and } \sqrt{n}[\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0] \Rightarrow \Theta_{M^*(1)}(\boldsymbol{\beta}_0, [M^*(1)]^{-1}U), \tag{19}$$

where U is normally distributed with mean 0 and covariance $\int_0^1 \Psi(t)\, dt$.

Define $M_1(t) = \mathbb{E}\{G_{1,j}^{(A^c)}(t, \mathcal{F}_0)[G_{1,j}^{(A^c)}(t, \mathcal{F}_0)]^\top\}$ if $b_j < t \le b_{j+1}$ and $M_1 = \int_0^1 M_1(t)\, dt$. Then under H_0^A we have (ii) :

$$nT_1^A \Rightarrow \tau(\Theta^1_{M_1}(\boldsymbol{\beta}_0^{(A^c)}, [M_1]^{-1}U^{(A^c)}), M_1, U^{(A^c)}) - \tau(\Theta_{M^*(1)}(\boldsymbol{\beta}_0, [M^*(1)]^{-1}U), M^*(1), U), \tag{20}$$

where $\tau(x, y, z) = x^\top y x - 2z^\top x$, $\Theta^1_\Sigma(\boldsymbol{\beta}, \mathbf{x}) = \lim_{n\to\infty}(\mathcal{P}_{Q\bigcap Q^{(A)},\Sigma}(n\boldsymbol{\beta} + \mathbf{x}) - n\boldsymbol{\beta})$ and A^c is the complement of A.

Theorem 4 reveals the asymptotic distributions of $\tilde{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_n$ as well as the test T_1^A and it generalizes former results on i.i.d. or normally distributed data. In the special case where $\{\mathbf{Z}_i\}$'s are i.i.d., ε_i is independent of $\mathbf{x}_i$ and $A = \{1, 2 \cdots, q\}$, simple algebra shows that (20) boils down to the classic mixture of chi-squared distributions. However, to construct confidence regions for $\boldsymbol{\beta}_0$ or to test the relatively complex null $H_0^{(A)}$ for non-stationary time series, we observe from Theorem 4 that the limiting distributions for both $\hat{\boldsymbol{\beta}}_n$ and $T^{(A)}$ involve the complex covariance matrix of U as well as the unknown regression parameter $\boldsymbol{\beta}_0$. Therefore directly using the asymptotic distributions in (19) and (20) does not seem operational in practice. In such situations one generally resorts to certain resampling methods for the inference. In the next subsection we shall investigate the behavior of subsampling for constrained regression of non-stationary time series as a representative example on how traditional resampling methods perform under non-stationarity.

4.2 Inconsistency of the subsampling

The inconsistency of the nonparametric bootstrap for inequality constrained statistical problems has been illustrated in several previous works. See for instance Andrews (2000). Alternatively, when the sample is stationary, Politis et al. (1999) shows that the subsampling is consistent under very weak conditions. In particular those conditions are satisfied by most constrained linear regression for stationary time series. For any $x \geq 0$, write

$$L_{m,n}(x) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} I\{\sqrt{m}|\hat{\boldsymbol{\beta}}_{j,j+m-1} - \hat{\boldsymbol{\beta}}_n| \leq x\}. \quad (21)$$

Then the subsampling uses $L_{m,n}(x)$ to approximate the probability of $\sqrt{n}|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}| \leq x$. The following proposition investigates the behavior of subsampling for constrained non-stationary time series regression.

Proposition 2. *Assume Conditions (A1)-(A3), (A5) and (A4)': $\sum_{k=0}^{\infty} \delta_{\mathbf{x}}(k, 8) < \infty$. Then under model* (1) *and the assumption that $m \to \infty$ with $m/n \to 0$, we have for any $x > 0$*

$$L_{m,n}(x) \to \int_0^1 \mathbb{P}[|\Theta_{M(s)}(\boldsymbol{\beta}_0, [M(s)]^{-1}R(s))| \leq x]\, ds \text{ in probability}, \quad (22)$$

where $R(s)$ is normally distributed with mean 0 and covariance $\Psi(s)$.

Proposition 2 shows that the subsampling distribution converges to the continuous mixture distribution of $|\Theta_{M(s)}(\boldsymbol{\beta}_0, [M(s)]^{-1}R(s))|$, $s \in [0,1]$. Note that the target distribution is $|\Theta_{M^*(1)}(\boldsymbol{\beta}_0, [M^*(1)]^{-1}U)|$ in (19). When $\{\mathbf{Z}_i\}$ is stationary, we have $M(s) = M^*(1)$ and $R(s) \stackrel{\mathcal{D}}{=} U$ for every $s \in [0,1]$. Hence $L_{m,n}(x) \to \mathbb{P}[|\Theta_{M^*(1)}(\boldsymbol{\beta}_0, [M^*(1)]^{-1}U)| \leq x]$ and the subsampling is consistent. However, in the general case when $\{\mathbf{Z}_i\}$ is non-stationary, the mixture distribution in (22) and the target distribution differ and therefore the subsampling provides inconsistent results for parameter inference. Note that, in the literature, Politis et al. (1997), among others, showed that the subsampling is consistent for some non-stationary processes. However, the non-stationarity allowed there is different from the PLS class. In particular, it can be shown that item (ii) in Assumptions (A) and (B) of the latter paper is not satisfied for constrained and unconstrained least squares estimators when $\{\mathbf{x}_i\}$ and/or $\{\varepsilon_i\}$ are PLS.

4.3 The robust bootstrap

In this section we demonstrate the robust bootstrap proposed in Section 3 as a simple and unified methodology for parameter inference of non-stationary time series regression with or without inequality constraints. Note that, by Theorems 2 and 3, the probabilistic behavior of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ can be consistently mimicked by $\bar{\Upsilon}_{m,n} := M_{1,n}^{-1}\Upsilon_{m,n}(m^*)$ and

$$\bar{\Lambda}_{m,n} := \mathcal{P}_{Q,M_{1,n}/n}(n^{1/4}\tilde{\boldsymbol{\beta}}_{1,n} + \bar{\Upsilon}_{m,n}) - n^{1/4}\tilde{\boldsymbol{\beta}}_n$$

respectively. Further note that the RSS statistic T_1^A is asymptotically equivalent to a continuous function of $\hat{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_n$ and RSS test without inequality constraints is a special case of T_1^A with $Q = \mathbb{R}^p$. Therefore both confidence region construction and RSS testing with/without inequality constraints can be carried out using the robust bootstrap. The following proposition summaries the above discussion.

Proposition 3. [2] *Under the assumptions of Theorem 4, we have (i):*

$$\bar{\Upsilon}_{m,n} \Rightarrow [M^*(1)]^{-1}U \text{ and } \bar{\Lambda}_{m,n} \Rightarrow \Theta_{M^*(1)}(\boldsymbol{\beta}_0, [M^*(1)]^{-1}U) \tag{23}$$

conditional on $\{\mathbf{Z}_i\}_{i=1}^n$. *Recall that '$\Rightarrow$' denotes weak convergence. Define*

$$H_{m,n} = \tau(\bar{\Lambda}_{m,n}, M_{1,n}, \Upsilon_{m,n}(m^*)), \quad H_{m,n}^* = \tau(\bar{\Lambda}_{m,n}^*, M_{1,n}^{(A^c)}, \Upsilon_{m,n}^{(A^c)}(m^*)),$$

where $M_{1,n}^{(A^c)} = \sum_{j=1}^n \mathbf{x}_j^{(A^c)}[\mathbf{x}_j^{(A^c)}]^\top/n$ *and*

$$\bar{\Lambda}_{m,n}^* = \mathcal{P}_{Q\bigcap Q^{(A)},M_{1,n}^{(A^c)}}(n^{1/4}\tilde{\boldsymbol{\beta}}_{1,n}^{(A^c)} + [M_{1,n}^{(A^c)}]^{-1}\Upsilon_{m,n}^{(A^c)}(m^*)) - n^{1/4}\tilde{\boldsymbol{\beta}}_{1,n}^{(A^c)}.$$

For any $\alpha \in (0,1)$, *let* $d_{m,\alpha}$ *be the* $(1-\alpha)$*th quantile of* $H_{m,n}^* - H_{m,n}$ *conditional on* $\{\mathbf{Z}_i\}_{i=1}^n$. *Then we have (ii) : under* H_0^A,

$$\mathbb{P}(T_1^A > d_{m,\alpha}) \to \alpha \text{ as } n \to \infty. \tag{24}$$

Proposition 3 follows from the proofs of Theorems 2 and 3. Note that the right hand sides of (23) are the limiting distributions of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$, respectively. Further observe that $\bar{\Lambda}_{m,n} = \bar{\Upsilon}_{m,n}$ and $\bar{\Lambda}_{m,n}^* = \bar{\Upsilon}_{m,n}^{(A^c)}$ when $\boldsymbol{\beta}_0$ is unconstrained; i.e. when $Q = \mathbb{R}^p$. Proposition 3 reveals that one can generate large (conditionally) i.i.d. samples of $\bar{\Upsilon}_{m,n}$, $\bar{\Lambda}_{m,n}$ and $H_{m,n}^* - H_{m,n}$ to consistently mimic the probabilistic behaviors of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ and T_1^A, respectively. Confidence regions for the regression coefficients and critical values of the RSS test can be determined accordingly. The detailed implementation procedures are very similar to steps (1)-(4) listed in Section 3.2 and are omitted here.

[2]There are a few typos in the journal version of Proposition 3. Specifically, $\bar{\Upsilon}_{m,n}$ and $\bar{\Upsilon}_{m,n}^{(A^c)}$ in the definitions of $H_{m,n}$ and $H_{m,n}^*$ there should be changed to $\Upsilon_{m,n}(m^*)$ and $\Upsilon_{m,n}^{(A^c)}(m^*)$, respectively. $\bar{\Upsilon}_{m,n}^{(A^c)}$ in the definition of $\bar{\Lambda}_{m,n}^*$ there should be changed to $[M_{1,n}^{(A^c)}]^{-1}\Upsilon_{m,n}^{(A^c)}(m^*)$.

Remark 2. As pointed out by a referee, a challenge in boundary problems is to obtain asymptotic coverage probability for $\boldsymbol{\beta}_0$ of at least $1-\alpha$ when the true parameter is allowed to drift and converge to the boundary as the sample size grows, where $1-\alpha$ is the nominal coverage probability. See e.g. Andrews and Guggenberger (2010, Section 4.1). If $\boldsymbol{\beta}_0$ approaches boundary faster than $n^{-1/2}$, then it is impossible to distinguish it from the boundary case and the robust bootstrap asymptotically covers the nominal percentage. On the other hand, if $\boldsymbol{\beta}_0$ approaches boundary at rate $n^{-\gamma}$ with $0 < \gamma < 1/2$, then one can choose $l_n = n^{(1/2-\gamma)/2}$ and the robust bootstrap asymptotically achieves the correct coverage probability, where l_n is defined on page 15 of the paper. A real challenge is the case where $\boldsymbol{\beta}_0$ approaches the boundary at the rate $n^{-1/2}$, in which case the coverage of the robust bootstrap can be asymptotically lower than the nominal. We shall leave the problem of valid (i.e. at least conservative) inference when $\boldsymbol{\beta}_0$ approaches boundary at the rate $n^{-1/2}$ in a future paper.

4.4 Window size selection

The implementation of the robust bootstrap involves the choice of one tuning parameter, the window size m. Note that by (15) the covariance function of the process $\{\Upsilon_{m,n}(i)\}$ consistently approximates that of $\{U(t)\}_{t=0}^1$. Note that the precision of the latter approximation decides the finite sample accuracy of the robust bootstrap for both structural change testing and parameter inference. Observe that the tuning parameter m controls the precision of the latter approximation. In particular, when m is too small, $\{\Upsilon_{m,n}(i)\}$ will not be able to fully take into account the dependence structure of the data and hence lead to a large bias in the approximation. On the other hand, however, when m is too large, the effective number of summands in $\{\Upsilon_{m,n}(i)\}$ will be too small and hence lead to a large variance of the approximation. The following Proposition 4 rigorously accounts for the latter bias-variance tradeoff.

Proposition 4. *Under conditions of Theorem 1, we have*

$$\begin{aligned}\max_{i,j} \|Cov(\Upsilon_{m,n}(i), \Upsilon_{m,n}(j)|\{\mathbf{Z}_i\}_{i=1}^n) - Cov(U(t_i), U(t_j))\| &= O(\sqrt{\frac{m}{n}} + \frac{1}{m}) \\ \|Cov(\Upsilon_{m,n}(n-m+1)|\{\mathbf{Z}_i\}_{i=1}^n) - Cov(U(1))\| &= O(\sqrt{\frac{m}{n}} + \frac{1}{m}). \qquad (25)\end{aligned}$$

Recall that $\|\cdot\|$ denotes the $\mathcal{L}^2$ norm of a random vector. Proposition 4 follows from the proof of Lemma 4 in the supplemental document Zhou (2014). In Proposition 4, the terms $\sqrt{m/n}$ and $1/m$ correspond to the standard error and bias of the aforementioned covariance approximation, respectively. It is clear from (25) that the optimal window size m should be chosen at the rate $n^{1/3}$ to balance the bias and variance. Note that in the stationary case the optimal rate for covariance matrix approximation is of the rate $n^{-1/3}$ (Carlstein, 1986) which is the same as that of the robust bootstrap. In this sense the robust bootstrap is expected to be accurate to apply though it is valid for a much wider class of temporal dynamics. We call the left hand side of the first equation in (25) the uniform root mean squared error (URMSE) of $\{\Upsilon_{m,n}(i)\}$. This measure should be used when choosing the window size for the structural stability test T_n. See also Zhou (2013) for a related discussion. On the other hand, one should use the second equation in (25) when choosing the window size for parameter inference. As we discussed in Section 4.3, for parameter inference we only need to focus on $\Upsilon_{m,n}(n-m+1)$ which consistently approximates the covariance structure of $U(1)$.

Remark 3. Note that the optimality in the above discussions refers to minimizing the MSE of covariance function approximation instead of optimizing the finite sample Type I or Type II error rates of the tests. As pointed out by a referee, there is a recent literature showing that the latter minimum MSE window size may be too small when the criteria is the size and power of the test. See for instance Sun and Phillips (2009). We shall leave the detailed theoretical and Monte Carlo comparisons of the latter two types of window size selection criterions for non-stationary time series regression in a rewarding future project.

In practice, the constants for the optimal window sizes are difficult to derive or estimate. To circumvent the difficulty, we suggest to use the minimum volatility (MV) method advocated in Politis et al. (1999). The MV method is a nonparametric method which utilizes the observation that the process $\{\Upsilon_{m,n}(i)\}$ becomes stable when m is in an appropriate range. Note that the MV method does not use any information on the underlying non-stationary models of the regressors and errors, which is convenient in our setting since the latter underlying models are typically difficult to approximate. The detailed implementation of the MV method for the structural stability test is listed below.

(a). Select a grid $m_1 < m_2 < \cdots < m_K$ of possible window sizes. For each m_k, calculate $\mu_{m_k,i} := \mathrm{Cov}(\Upsilon_{m_k,n}(i), \Upsilon_{m_k,n}(j)|\{\mathbf{Z}_r\}_{r=1}^n)$, $i \le j$ for $i = 1, 2, \cdots, n - m_k + 1$.

(b). For each m_i, calculate $\max_{1\le j\le n-m_K+1} se\Big(\{\mu_{m_{i+r},j}\}_{r=-3}^{3}\Big)$, where se denotes standard error.

(c). Choose window size $\hat{m}_{MV}$ which minimizes the maximized standard errors in (b).

In (a), we note that $\mathrm{Cov}(\Upsilon_{m,n}(i), \Upsilon_{m,n}(j)|\{\mathbf{Z}_i\}_{i=1}^n)$ only depends on i when $i \leq j$. When selecting window sizes for parameter inference, one just need to change the criterion function in (a) into $\mu^*_{m_k} := \mathrm{Cov}(\Upsilon_{m_k,n}(n - m_k + 1))|\{\mathbf{Z}_r\}_{r=1}^n)$ and calculate in (b) the quantity $se\Big(\{\mu^*_{m_{i+r}}\}_{r=-3}^3\Big)$.

5 Simulation studies

In this section, we shall conduct moderate sample Monte Carlo experiments to study the accuracy of the proposed robust bootstrap procedures for structural stability tests and parameter inference. The inconsistency of the subsampling for parameter inference of constrained non-stationary time series regression will also be investigated. For the latter purposes, let us consider the simple linear regression

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \quad i = 1, 2, \cdots, n$$

with both unconstrained and constrained β_2. It is assumed that $\beta_2 \geq 0$ when constrained. The following time series models are considered for the regressors and the errors:

(i). (Stationary model). Let $\{x_i\}$ and $\{\epsilon_i\}$ follow a stationary AR(1) model with AR(1) coefficient 0.5 and i.i.d. standard normal innovations. Furthermore, the processes $\{x_i\}$ and $\{\epsilon_i\}$ are independent. In this scenario we are interested in investigating the behavior of the robust bootstrap for the traditional constrained time series regression with stationary regressors and errors.

(ii) (Conditional homoscedastic non-stationary model). Consider the following model

$$x_i = L_1(t_i, (\zeta_i, \zeta_{i-1}, \cdots)) \text{ if } t_i \leq 1/2 \text{ and } x_i = L_2(t_i, (\zeta_i, \zeta_{i-1}, \cdots)) \text{ if } t_i > 1/2, \quad (26)$$

where $L_1(t, (\zeta_i, \zeta_{i-1}, \cdots)) = -0.2L_1(t, (\zeta_{i-1}, \zeta_{i-2}, \cdots)) + \zeta_i$, $L_2(t, (\zeta_i, \zeta_{i-1}, \cdots)) = 0.5L_2(t, (\zeta_{i-1}, \zeta_{i-2}, \cdots)) + \zeta_i$ and ζ_i's are i.i.d. standard normal. Here $\{x_i\}$ is a piece-wise stationary AR(1) process in the sense that the AR(1) coefficient changes from -0.2 to 0.5 at the break point $t = 0.5$. The process is stationary before and after the break point. Furthermore, consider the following model

$$\begin{aligned} e_i &= H(t_i, (\eta_i, \eta_{i-1}, \cdots)), \text{ where} \\ & H(t, (\eta_i, \eta_{i-1}, \cdots)) = 0.7\cos(2\pi t) H(t, (\eta_{i-1}, \eta_{i-2}, \cdots)) + \eta_i \end{aligned} \quad (27)$$

and η_i's are i.i.d. standard normal and are independent of ζ_i's. Here $\{e_i\}$ is locally stationary in the sense that the AR(1) coefficients $0.7\cos(2\pi t)$ changes smoothly on

[0,1]. Let $\epsilon_i = e_i$. Then the regression is homoscedastic since the regressors and the errors are independent.

(iii). (Conditional heteroscedastic non-stationary model). In this scenario we consider the same model for the regressors $\{x_i\}$ as in (26) and let $\epsilon_i = x_i e_i$, where e_i is defined in (27). Here we have a heteroscedastic non-stationary linear regression in the sense that variability of the errors depends on the regressors $\{x_i\}$.

(iv). (Non-stationary auto regressive model). Consider the following regression

$$x_{i+1} = \beta_1 + \beta_2 x_i + \epsilon_i, \tag{28}$$

where $\epsilon_i = v(t_i)\eta_i$ with $v(t) = 1/2 + 2(t - 1/2)^2$ and i.i.d. standard normal η_i's. Here the AR(1) coefficient β_2 is constrained to be non-negative. Clearly $y_i = x_{i+1}$, $i = 1, 2, \cdots, n-1$. In this scenario we have an auto-regressive model with non-stationary innovations. In particular, it can be shown that the process $\{x_i\}$ is a locally stationary with smoothly varying covariance structures.

In all simulations conducted in this section, the intercept $\beta_1 = 1$. We consider unconstrained (UC) as well as two constrained scenarios for the slope coefficient β_2; namely scenario (B) where the inequality constraint is binding with $\beta_2 = 0$ and scenario (NB) where the inequality constraint is non-binding with $\beta_2 = 1$ in models (i)-(iii) and $\beta_2 = 0.5$ in model (iv). The parameter β_2 in the (UC) case is chosen to be the same as that in the (NB) case. Note that in the (NB) case the signal to noise ratios are set to be less than 1 in all four models. In our Monte Carlo experiments we shall investigate the accuracy of the robust bootstrap for the unconstrained test $\tilde{T}_n$. Additionally, we want to check whether the robust bootstrap for T_n performs accurately and similarly in the boundary and interior cases. The simulated Type I error rates at nominal levels $\alpha = 0.1$ and 0.05 are reported in Table I below. The window sizes are selected by the MV method. Sample sizes $n = 200$ and 400 with 1000 replicates. The bootstrap sample size $\Theta = 1500$.

	$n = 200$						$n = 400$					
	$\alpha = 0.05$			$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.1$		
Model	(UC)	(B)	(NB)	(UC)	(B)	(NB)	(UC)	(B)	(NB)	(UC)	(B)	(NB)
(i)	5.8	3.3	7.0	11.4	9.5	13.2	5.2	4.3	4.3	10.2	9.9	9.7
(ii)	3.6	3.6	3.8	8.6	11.3	10.6	4.3	5.6	4.3	10.5	11.7	11.3
(iii)	3.6	3.7	6.4	8.0	10.6	13.3	3.8	4.7	5.1	9.3	9.3	12
(iv)	5.0	4.2	3.4	11.6	7.8	13.6	5.1	4.3	6	10.4	9.4	12.2

Table I. Simulated type I error rates (in %) for the structural stability test T_n with nominal levels 5% and 10% under models (i)-(iv). (UC), (B) and (NB) stand for unconstrained, binding and non-binding β_2, respectively.

We observe from Table I that the structural stability tests $\tilde{T}_n$ and T_n with the robust bootstrap perform reasonably well for all four models. With the moderate signal to noise ratios, the performances for the boundary (B) and interior point (NB) cases are similar. The performance of the robust bootstrap for the stationary case (i) and non-stationary cases (ii)-(iv) are similar. This is consistent with Proposition 4 that the robust bootstrap converges at the rate $n^{-1/3}$ for non-stationary models which is the same as the classic rate for stationary time series. The MV method's performance is acceptable for the models considered in the simulations. Furthermore, as we expected, generally the accuracy increases and the test behaves more stably as we enlarge the sample size.

We further investigate the accuracy of the robust bootstrap for parameter inference. For this purpose we consider models (i) to (iv) and investigate the coverage probabilities of the confidence intervals (CI's) and Type I error rates of the RSS tests for the inference of β_2. Both constrained and unconstrained regressions are considered. For CI construction the performances of the robust bootstrap is compared with those of the subsampling (with inequality constraints) to confirm our theoretical results in Proposition 2. With 1500 bootstrap samples and 1000 replicates, the simulation results are summarized in Table II below. Window sizes for both the robust bootstrap and the subsampling are selected by the MV method.

We observe from Table II that the subsampling is inconsistent for non-stationary models (ii) to (iv). The latter results support our theoretical findings in Proposition 2. On the other hand, we find that the subsampling performs well for stationary model (i) when β_2 is non-binding as expected. However, for moderate sample sizes the subsampling does not perform well when the inequality constraint is binding. This somewhat contradictory result may be due to a slow convergence rate of the subsampling in the boundary cases. This suspicion is supported by our further simulation studies in which the subsampling is reasonably accurate when the sample size is increased to 1000. For the considered models, the robust bootstrap with the MV method performs reasonably well for both the CI construction and RSS test. Similarly to our findings in Table I, we observe that the performance of the robust bootstrap does not decrease for the non-stationary cases.

				$n = 200$				$n = 400$			
α			Method	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
95%	CI	(UC)	RB	93.2	93.0	92.2	93.1	94.2	94.8	93.8	93.2
90%	CI	(UC)	RB	87.1	87.6	86.9	87.8	89.9	89.0	88.0	90.8
95%	CI	(B)	RB	93.6	94.2	93.2	93.9	94.9	94.6	94	94.8
90%	CI	(B)	RB	87.8	88.9	87.2	88.4	89.6	89.7	88.3	88.9
95%	CI	(B)	SUB	98.7	99.4	87.9	99.6	97.6	99.8	91.5	99.4
90%	CI	(B)	SUB	97.6	99.4	87.2	99.6	94.1	99.8	91	99.4
95%	CI	(NB)	RB	93.1	93.4	92.3	92	93.5	95	93.2	93.8
90%	CI	(NB)	RB	87.2	88	86.9	86.7	88.0	88.5	88	90.5
95%	CI	(NB)	SUB	96	97.1	87.9	100	95.7	97.7	90.4	100
90%	CI	(NB)	SUB	91.2	93.6	81	100	91	93.3	84	100
5%	RSS	(UC)	RB	7.0	6.4	6.6	6.1	5.6	5.2	6.2	4.8
10%	RSS	(UC)	RB	12	12.1	11.9	8.8	11.1	11.5	11.3	10.2
5%	RSS	(B)	RB	6.1	5.9	6.9	6.1	5.1	4.9	6.1	5.0
10%	RSS	(B)	RB	12	10.7	11.8	10.8	9.2	9	12.0	10.5

Table II. Simulated coverage probabilities and Type I error rates (in %) for the confidence intervals (CI) and RSS tests. α represents the confidence level in the CI case and α equals the test level in the RSS case. RB stands for the robust bootstrap method and SUB stands for the subsampling method. (UC), (B) and (NB) stand for unconstrained, binding and non-binding β_2, respectively.

6 A real data illustration

In this section we shall analyze the data set of percentage changes of the exchange rate between mark and dollar in 10-minute intervals. The first 240 hours of data is used in our data analysis. Figure 1 below shows the time series plot of the data. The data set was analyzed in Tsay (2005) to illustrate stationary ARCH models. Due to mathematical requirements, the ARCH coefficients are constrained to be nonnegative. Though the ARCH and GARCH models are widely applied in empirical research and practice, there are relatively less discussions on checking the structural stability of such models. Exceptions include, among others, Kokoszka and Leipus (2000) who studied change point detection of an ARCH(∞) model via a CUSUM test of marginal variances. To our knowledge, there are no results on structural change detection of ARCH models taking into account the restricted parameter space of the model coefficients. As we discussed before, considering model coefficients in a restricted parameter space may lead to a higher sensitivity to possible structural changes of the model. Let $x_1, x_2, \cdots, x_{1440}$ be the observed time series. Let us consider the following null model:

$$x_t^2 = \beta_0 + \beta_1 x_{t-1}^2 + \beta_2 x_{t-2}^2 + \beta_3 x_{t-3}^2 + \epsilon_t, \tag{29}$$

where $\beta_i \geq 0$, $i = 1, 2, 3$ and ϵ_t's are possibly non-stationary satisfying $\mathbb{E}[\epsilon_t | x_{t-1}, x_{t-2}, x_{t-3}] = 0$. The above model includes the ARCH(3) model in Tsay (2005) where $x_t = e_t h_t$, $h_t^2 = \beta_0 + \beta_1 x_{t-1}^2 + \beta_2 x_{t-2}^2 + \beta_3 x_{t-3}^2$ and $\{e_t\}$'s are i.i.d. with mean 0 and variance 1. Furthermore, compared to the ARCH(3) model, the major advantage of model (29) is that it allows non-stationary ϵ_t's and hence could result in a more robust detection of possible structural changes in the ARCH effect. Indeed, for this high-frequency exchange rate data with a relatively long period of 240 hours, the assumption of stationarity is suspectable. We first apply the robust bootstrap to test structural stability of model (29). The MV method selects window size $m = 26$ and the p-value with 1000 bootstrap replicates is about 7%. Hence there is an evidence against the null model of structural stability. The detection of changes in the ARCH structure is important as it indicates that one may need to segment the series and use a stable subseries when forecasting future volatilities.

Note that one could also utilize the structural stability test $\tilde{T}_n$ in (8) with no inequality constraints to detect changes in model (29). When we apply $\tilde{T}_n$ with the robust bootstrap to the series, the MV method selects window size $m = 26$ and the resulting p-value with 1000 bootstrap replicates is about 65%. Hence $\tilde{T}_n$ shows no evidence against (29). The large increase in the p-value for the unconstrained structural stability test is due to its relatively larger parameter space and lower sensitivity to departures from the null model. This example shows the usefulness of testing in constrained parameter spaces.

Now let us focus on the subseries of the last 24 hours of the 240 hour time span. For the structural stability test T_n, the MV method selects window size $m = 17$ and the resulting p-value is about 0.5. Hence there is no evidence indicating there is structural change in the parameters of (29) in this subseries. As a result parameter inference could be performed. First, we shall test the existence of ARCH effect. This is equivalent to testing $\beta_i = 0$, $i = 1, 2, 3$ against $\beta_i \geq 0$, $i = 1, 2, 3$. To this end, we perform the RSS test T_1^A with the robust bootstrap. The MV method selects window size $m = 17$ and the p-value is less than 0.5%. Hence there is a strong evidence indicating that ARCH effect exists in the subseries. In the literature, Silvapulle and Silvapulle (1995) tested ARCH effects via a one-sided score test. Both the one-sided score test and our robust bootstrap RSS test utilize the inequality constraints on the ARCH parameters. However, the advantage of our test is that it allows ARCH models with dependent and non-stationary innovations which could be more robust to model misspecifications. Second, since the fitted β_2 and β_3 are very small for the subseries, we shall test $\beta_2 = \beta_3 = 0$ against $\beta_i \geq 0$ $i = 1, 2, 3$. Note that the one sided score test in Silvapulle and Silvapulle (1995) cannot be used for this test since the null hypothesis contains inequality constraints. The RSS test with robust bootstrap yields a p-value ≈ 1 and it indicates that one could use a simple model $x_t^2 = \beta_0 + \beta_1 x_{t-1}^2 + \epsilon_t$ for the forecasting purpose. Finally, the robust bootstrap provides a 95% confidence interval for β_1 as $[0.19, 0.22]$.

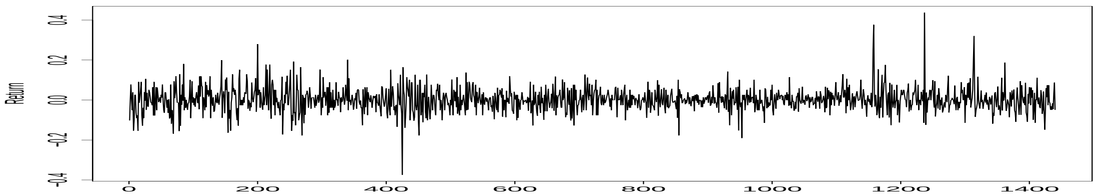

Figure 1: Time series plot of the percentage changes of the exchange rate between mark and dollar in 10-minute intervals. Time series length = 1440.

Acknowledgements. I am grateful to the editor, the associate editor and two referees for their many helpful comments which lead to a substantial improvement of the paper. The research was supported in part by NSERC of Canada.

7 Proofs

Proof of Theorem 1. By Lemma 6 (ii) of Zhou (2013) and (A4), we have

$$\| \max_{p+1 \leq i \leq n} i |M_{1,i} - \mathbb{E} M_{1,i}| \| = O(n^{1/2} \sum_{j=0}^{\infty} \delta_{\mathbf{x}}(j, 4)) = O(n^{1/2}). \tag{30}$$

Let $L(n) = [n^{-1/4}, 1 - n^{-1/4}]$ and $\bar{L}(n) = [(p+1)/n, n^{-1/4}) \bigcup (1 - n^{-1/4}, 1 - (p+1)/n]$. We have by (30) that

$$\| \max_{t_i \in L(n)} |M_{1,i} - M^*(t_i)| \| = O(n^{-1/4}) = o(1).$$

Similar equality holds for $M_{i+1,n}$. Observe that $\tilde{\boldsymbol{\beta}}_{1,i} - \boldsymbol{\beta}_0 = i M_{1,i}^{-1} \sum_{j=1}^{i} \mathbf{W}_j$. Hence we have by the supplemental Lemma 3 in Zhou (2014) and the continuous mapping theorem

$$\max_{t_i \in L(n)} \sqrt{n} t_i (1 - t_i) |\hat{\boldsymbol{\beta}}_{1,i} - \hat{\boldsymbol{\beta}}_{i+1,n}| \Rightarrow \sup_{0 < s < 1} |(1 - s)[M^*(s)]^{-1} U(s) - s[M^o(s)]^{-1}[U(1) - U(s)]|.$$

By similar arguments, we have $\max_{t_i\in\bar{L}(n)}\sqrt{n}t_i(1-t_i)|\tilde{\boldsymbol{\beta}}_{1,i}-\tilde{\boldsymbol{\beta}}_{i+1,n}|=o_{\mathbb{P}}(1)$. Hence the second equality in (11) holds.

We now prove the first weak convergence result in (11). By the supplemental Lemma 3 in Zhou (2014) and Proposition 1 (v), (vi), we have

$$\max_{t_i\in L_n}\sqrt{n}|\mathcal{P}_{Q,M_{1,i}}(\tilde{\boldsymbol{\beta}}_{1,i})-\mathcal{P}_{Q,M^*(t_i)}(\tilde{\boldsymbol{\beta}}_{1,i})|=O_{\mathbb{P}}(\sqrt{n^{-1/4}})=o_{\mathbb{P}}(1). \tag{31}$$

For $j=1,2,\cdots,n$, let $\tilde{\mathbf{W}}_j=\sum_{k=1}^{j}\mathbf{W}_k/\sqrt{n}$ and $\tilde{\mathbf{W}}_0=0$. Let $\tilde{\mathbf{W}}_n(t)$, $t\in[0,1]$, be the linear interpolation of $\{\tilde{\mathbf{W}}_j\}$. By Proposition 1 (iv), (v), (vi), (31), Lemma 3 in Zhou (2014) and the continuous mapping theorem,

$$\mathcal{P}_{Q,M_{n,1}(t)}(\sqrt{n}\boldsymbol{\beta}_0+M_{n,1}^{-1}(t)\tilde{\mathbf{W}}_n(t))-\sqrt{n}\boldsymbol{\beta}_0\Rightarrow\Theta_{M^*(t)}(\boldsymbol{\beta}_0,[M^*(t)]^{-1}U(t)) \tag{32}$$

on $L(n)$, where $M_{n,1}(t)$ is the linear interpolation of $\{M_{1,t}\}_{t=p+1}^{n-p-1}$. Similarly and jointly,

$$\mathcal{P}_{Q,M_{n,2}(t)}(\sqrt{n}\boldsymbol{\beta}_0+M_{n,2}^{-1}(t)[\tilde{\mathbf{W}}_n(1)-\tilde{\mathbf{W}}_n(t)])-\sqrt{n}\boldsymbol{\beta}_0\Rightarrow\Theta_{M^o(t)}(\boldsymbol{\beta}_0,[M^o(t)]^{-1}[U(1)-U(t)])$$

on $L(n)$, where $M_{n,2}(t)$ is the linear interpolation of $\{M_{t+1,n}\}_{t=p+1}^{n-p-1}$. At the same time, it is easy to show that $\max_{t_i\in\bar{L}(n)}\sqrt{n}t_i(1-t_i)|\hat{\boldsymbol{\beta}}_{1,i}-\hat{\boldsymbol{\beta}}_{i+1,n}|=o_{\mathbb{P}}(1)$ under H_0. Theorem 1 follows. $\diamondsuit$

Proof of Theorem 2. We first prove (i). By the supplemental Lemma 4 in Zhou (2014) and similar arguments as those in the proof of Theorem 1, we have, under H_0 and conditional on $\{\mathbf{Z}_i\}$,

$$M_{n,1}^{-1}(t)\Delta_m(t)\Rightarrow[M^*(t)]^{-1}U(t)\text{ on }L(n). \tag{33}$$

Recall the definition of $M_{n,1}(t)$ in (32). Note that, under H_0,

$$\Upsilon_{m,n}(j)=\Delta_m(t_j)-\sqrt{m}\sum_{k=1}^{j}M_{k,k+m-1}(\tilde{\boldsymbol{\beta}}_{1,n}-\boldsymbol{\beta}_0)V_k/\sqrt{m^*}.$$

On the other hand, observe that

$$\tilde{\boldsymbol{\beta}}_{1,n}=\boldsymbol{\beta}_0+M_{1,n}^{-1}\tilde{W}_n/\sqrt{n}=\boldsymbol{\beta}_0+O_{\mathbb{P}}(n^{-1/2}). \tag{34}$$

Therefore, conditional on $\{\mathbf{Z}_i\}$,

$$\max_{t_j\in[0,1]}\|\Upsilon_{m,n}(j)-\Delta_m(t_j)\|=O_{\mathbb{P}}(\sqrt{\frac{m}{n}})=o_{\mathbb{P}}(1). \tag{35}$$

By Lemma 4 in Zhou (2014) and (33) to (35), we have, under H_0 and conditional on $\{\mathbf{Z}_i\}$

$$\begin{aligned}\max_{t_j \in L_n} \Big|(1 &- t_{j+m-1})M_{1,j+m-1}^{-1}\Upsilon_{m,n}(j) - t_{j+m}M_{j+m,n}^{-1}[\Upsilon_{m,n}(m^*) - \Upsilon_{m,n}(j)]\Big| \\ &\Rightarrow \sup_{s\in L(n)} |(1-s)[M^*(s)]^{-1}U(s) - s[M^o(s)]^{-1}[U(1)-U(s)]|.\end{aligned}$$

On the other hand, it is easy to show that

$$\max_{t_j \in \bar{L}(n)} \Big|(1 - t_{j+m-1})M_{1,j+m-1}^{-1}\Upsilon_{m,n}(j) - t_{j+m}M_{j+m,n}^{-1}[\Upsilon_{m,n}(m^*) - \Upsilon_{m,n}(j)]\Big| = o_{\mathbb{P}}(1).$$

Note that by Theorem 1 and Example 5 of Cirel'soN et al. (1976), the limiting distribution of $\tilde{T}_n$ is continuous at its $(1-\alpha)$th quantile. Hence (i) follows from the above arguments. To prove (ii), without loss of generality, assume that $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_{\lfloor kn \rfloor} = 0$ and $\boldsymbol{\beta}_{\lfloor kn \rfloor+1} = \cdots = \boldsymbol{\beta}_n = n^{-1/2}L_n$. Otherwise we just need to subtract $\boldsymbol{\beta}_i$'s by a constant vector. Note that under H_a

$$\Upsilon_{m,n}(j) = \tilde{\Delta}_{j,m} + I_j,$$

where $\tilde{\Delta}_{j,m} = \sum_{k=1}^{j}[\sqrt{m}\mathbf{W}_{k,k+m-1} - mM_{k,k+m-1}M_{1,n}^{-1}\tilde{W}_n/\sqrt{n}]V_k/\sqrt{mm^*}$, $\mathbf{W}_{k,k+m-1} = \sum_{i=k}^{k+m-1}\mathbf{W}_i/\sqrt{m}$ and

$$I_j = \sum_{k=1}^{j}[\tau_k - mM_{k,k+m-1}M_{1,n}^{-1}\tau_n^*/n]V_k/\sqrt{mm^*} \text{ with } \tau_k = \sum_{i=k}^{k+m-1}\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_i$$

and $\tau_n^* = \sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_i$. By the proof of Lemma 4 in Zhou (2014), we have $\max_j |\tilde{\Delta}_{j,m}| = O_{\mathbb{P}}(1)$ conditional on $\{\mathbf{Z}_r\}_{r=1}^n$. On the other hand, similarly to the proof of Lemma 4 in Zhou (2014),

$$\begin{aligned}\max_{1\le j\le m^*}\Big|\sum_{k=1}^{j}\{[\tau_k - \mathbb{E}\tau_k][\tau_k - \mathbb{E}\tau_k]^\top\Big|/(mm^*) &= O_{\mathbb{P}}(n^{-1}|L_n|), \\ \max_{1\le j\le m^*} m\Big|\sum_{k=1}^{j}[M_{k,k+m-1} - \mathbb{E}M_{k,k+m-1}][M_{k,k+m-1} - \mathbb{E}M_{k,k+m-1}]^\top\Big|/m^* &= O_{\mathbb{P}}(1).\end{aligned}$$

It is easy to show that

$$\max_{1\le j\le m^*}\sum_{k=1}^{j}\{|\mathbb{E}\tau_k|^2 + m^2|\mathbb{E}M_{k,k+m-1}[\mathbb{E}M_{1,n}]^{-1}\mathbb{E}[\tau_n^*]|^2/n^2\}/(mm^*) = O_{\mathbb{P}}\Big(\frac{m}{n}|L_n|^2\Big).$$

Now elementary manipulations show that $\max_j |I_j| = O_{\mathbb{P}}(1 + \sqrt{m/n}L_n)$ conditional on $\{\mathbf{Z}_i\}$. Together with $\max_j |\tilde{\Delta}_{j,m}| = O_{\mathbb{P}}(1)$ conditional on $\{\mathbf{Z}_i\}$, we have

$$c_{m,\alpha} = O_{\mathbb{P}}(1 + \sqrt{m/n}|L_n|) \text{ under } H_a. \tag{36}$$

On the other hand, observe that

$$\sqrt{n}|\tilde{\boldsymbol{\beta}}_{1,\lfloor kn\rfloor} - \tilde{\boldsymbol{\beta}}_{\lfloor kn\rfloor+1,n}| = |L_n| + O_{\mathbb{P}}(1) \text{ under } H_a.$$

Hence by (36) and the facts that $m/n \to 0$ and $|L_n| \to \infty$, we have (ii) follows. $\Diamond$

Proof of Theorem 3. We first prove (i). Note that, under H_0, $n^{1/4}\hat{\boldsymbol{\beta}}_{1,n} = n^{1/4}\boldsymbol{\beta}_0 + O_{\mathbb{P}}(n^{-1/4})$. By the proofs of Theorem 2 and Proposition 1, we have

$$\Lambda^*_{m,n}(s) - n^{1/4}s(1-s)\boldsymbol{\beta}_0 \Rightarrow \Theta_{M^*(s)}(\boldsymbol{\beta}_0, (1-s)[M^*(s)]^{-1}U(s)) \tag{37}$$

on $\mathcal{C}(0,1)$ conditional on $\{\mathbf{Z}_j\}$, where $\Lambda^*_{m,n}(s)$ is the linear interpolation of $\Lambda^*_{m,n}(j)$. Similarly, let $\Lambda^o_{m,n}(s)$ be the linear interpolation of $\Lambda^o_{m,n}(j)$. Then jointly with (37),

$$\Lambda^o_{m,n}(s) - n^{1/4}s(1-s)\boldsymbol{\beta}_0 \Rightarrow \Theta_{M^o(s)}(\boldsymbol{\beta}_0, s[M^o(s)]^{-1}[U(1) - U(s)]) \tag{38}$$

on $\mathcal{C}(0,1)$ conditional on $\{\mathbf{Z}_j\}$. Hence (i) follows by (37) and (38) and the assumption that the limit distribution of T_n is continuous at its $(1-\alpha)$th quantile. To prove (ii), note that by Proposition 1 and similar arguments as those in the proof of Theorem 2, we have

$$c^R_{m,\alpha} = O_{\mathbb{P}}(1 + \sqrt{m/n}|L_n|) \text{ under } H_a. \tag{39}$$

On the other hand, note that, by Proposition 1,

$$\begin{aligned}\sqrt{n}|\hat{\boldsymbol{\beta}}_{1,k^*} - \hat{\boldsymbol{\beta}}_{k^*+1,n}| &= \Big|[\mathcal{P}_{Q,M_{1,k^*}/n}(\sqrt{n}\tilde{\boldsymbol{\beta}}_{1,k^*}) - \sqrt{n}\boldsymbol{\beta}_{k^*}] \\ &\quad -[\mathcal{P}_{Q,M_{k^*+1,n}/n}(\sqrt{n}\tilde{\boldsymbol{\beta}}_{k^*+1}) - \sqrt{n}\boldsymbol{\beta}_{k^*+1}] + \sqrt{n}\boldsymbol{\beta}_{k^*} - \sqrt{n}\boldsymbol{\beta}_{k^*+1}\Big| \\ &= |L_n| + O_{\mathbb{P}}(1) \end{aligned} \tag{40}$$

under H_a, where $k^* = \lfloor kn \rfloor$. (ii) follows from (39) and (40). $\Diamond$

REFERENCES

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405.

Andrews, D. W. K. and Guggenberger, P. (2010) Applications of subsampling, hybrid, and size-correction methods. *Journal of Econometrics* **158** 285–305.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussions). *Journal of the Royal Statistical Society. series B* **37** 149–192.

Carlstein, E. (1986). The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Time Series. *The Annals of Statistics* **14** 1171–1179.

Cirel'soN, B. S., Ibragimov, I. A., and Sudakov, V. N. (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory* (pp. 20-41). Springer Berlin Heidelberg.

Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, Chichester.

Dahlhaus, R. (1997). Fitting time series models to non-stationary processes. *Ann. Statist.* **25** 1–37.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* **101** 223–239.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

Fan, J., Jiang, J., Zhang, C. and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statist. Sinica* **13** 965–992.

Gouriéroux, C., Holly, A. and Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* **50** 63–80.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling* **14** 517–533.

Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76** 409–423.

Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli* **6** 513–539.

Koo, B. and Linton, O. (2012). Semiparametric estimation of locally stationary diffusion models. *J. Econometrics* 170 210233.

Liew, C. K. (1976). Inequality constrained least-squares estimation. *Journal of the American Statistical Association* **71** 746–751.

Loader, C. R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics* **24** 1667–1678.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32** 577–602.

Mikosch, T. and Starica, C. (2004). Non-stationarities in financial time series, the long-range dependence, and the IGARCH effects. *Rev. Econ. Statist.* **86** 378–390.

Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society series B* **62** 271-292.

Ombao, H., von Sachs, R. and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519-531.

Ploberger, W., Krämer, W. and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics* **40** 307–318.

Politis, D. N., Romano, J. P. and Wolf, M. (1997). Subsampling for heteroskedastic time series. *Journal of Econometrics* **81** 281–317.

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

Robertson, T, Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* New York: Wiley.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56** 49–62.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference, Inequality, Order and Shape Restrictions.* Wiley, New Jersey.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternative. *Journal of the American Statistical Association* **90** 342–349.

Starica, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics.* **87** 503–522.

Sun, Y. and P.C.B. Phillips (2009). Optimal bandwidth choice for interval estimation in GMM regression. unpublished working paper, UCSD.

Tsay, R. S. (2005). *Analysis of financial time series*, second edition. Wiley-Interscience.

Van Bellegem, S. and Dahlhaus, R. (2006). Semiparametric estimation by model selection for locally stationary processes. *Journal of the Royal Statistical Society. series B* **68** 721–746.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Annals of Statistics* **40** 2601–2633.

Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* **82** 782–793.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl Acad. Sci. USA* **102** 14150-14154.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association* **108** 726–740.

Zhou, Z. (2014). Supplement for "Inference for Non-stationary Time Series Regression with/without Inequality Constraints".

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation of nonstationary time series. *The Annals of Statistics* **37** 2696-2729.