# HETEROSCEDASTICITY AND AUTOCORRELATION ROBUST STRUCTURAL CHANGE DETECTION

By Zhou Zhou[1]

*University of Toronto*

February 1, 2013

**Abstract**

The assumption of (weak) stationarity is crucial for the validity of most of the conventional tests of structure change in time series. Under complicated non-stationary temporal dynamics, we argue that traditional testing procedures result in mixed structural change signals of the first and second order and hence could lead to biased testing results. The paper proposes a simple and unified bootstrap testing procedure which provides consistent testing results under general forms of smooth and abrupt changes in the temporal dynamics of the time series. Monte Carlo experiments are performed to compare our testing procedure to various traditional tests. Our robust bootstrap test is applied to testing changes in an environmental and a financial time series and our procedure is shown to provide more reliable results than the conventional tests.

## 1 Introduction

Structural stability over time is important in many scientific endeavors. For most of the frequently used statistical tests of structural change, the assumption of (weak) stationarity under the null hypothesis is crucial for their validity. However, the stationarity assumption has become restrictive for many contemporary structural change analysis. To simplify discussion, let us consider the test of structural change in mean where we observe time

[1]Corresponding author. Department of Statistics, 100 St. George Street, Toronto, Ontario, M5S 3G3 Canada.

*E-mail: zhou@utstat.toronto.edu*

*Key words and phrases.* Piecewise locally stationary time series, structural change detection, bootstrap, CUSUM, climate change

series $\{X_i\}_{i=1}^n$ with $\mathbb{E}[X_i] = \mu_i$ and we are interested in testing whether $\mu_i$ remains constant over time; namely testing

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu, \quad \longleftrightarrow \quad H_a : \mu_i \neq \mu_j \text{ for some } 1 \leq i < j \leq n. \tag{1}$$

For most of the conventional tests of $H_0$, the covariance structure of $\{X_i\}$ should remain unchanged over time. In other words, the latter tests are applicable to time series of the form $X_i = \mu_i + e_i$, where $\{e_i\}$ is a zero-mean weakly stationary sequence. Nevertheless, the stationarity assumption is unrealistic in many important applications. For instance, in climatology, there have been numerous empirical and simulated evidences indicating that climate variability is changing. Among others, Elsner et al. (2009) finds that the strongest tropical cyclones are getting stronger globally over the last three decades while the mean cyclone strength remains constant. Karl et al. (1995) and Collins et al. (2000) have identified a general decline in the intra-monthly temperature variability over the globe. Räisänen (2002) identified an increase in monthly mean precipitation variability in most areas in 19 simulated experiments. In the contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change (2007), Chapter 10 projected that '*Future changes in anthropogenic forcing will result not only in changes in the mean climate state but also in the variability of climate*'. Another prominent example lies in econometrics where investigating consistent procedures under heteroscedasticity and autocorrelation is one of the important research areas. In the context of parameter estimation of various parametric models, the classes of heteroscedasticity consistent (HC) and heteroscedasticity and autocorrelation consistent (HAC) covariance estimators (see White (1980), Newey and West (1987) and Andrews (1991) among others) have had a major impact.

When the covariance structure of the time series is varying, it is shown in this paper that many conventional tests of $H_0$ are inconsistent and can lead to biased testing results. To understand this, think of the classic cumulative sum (CUSUM) test

$$T_n = \max_{1 \leq i \leq n} |S_i - t_i S_n|, \text{ where } S_i = \sum_{j=1}^{i} X_j \text{ and } t_i = i/n. \tag{2}$$

The classic idea to perform this test as well as most other tests of structural change is normalization. More specifically, one normalizes $T_n$ by a consistent or inconsistent estimator of $\mathrm{Cov}(S_n)/n$ to make the test asymptotically pivotal. Critical values of the test can then be obtained accordingly. Nevertheless, when $\{X_i\}$ is second order non-stationary, it is found in this paper that the behavior of $T_n$ under $H_0$ is determined by a centered Gaussian process with very complex covariance structure. As a consequence it is generally impossible to make $T_n$ pivotal by normalizing it with one or even a sequence of covariance estimators. The complicated non-stationary dynamics in the second order structure in time series has posted new challenges to the classic problem of structural change detection. To date, little progress has been made toward structural change tests that are robust to heteroscedasticity and autocorrelation of general forms.

The purpose of the paper is to perform robust structural change detection for second and higher order non-stationary time series. We discard the traditional idea of normalization; instead, as a major contribution we propose a simple bootstrap procedure that is shown to be consistent under general forms of abrupt and smooth changes in the temporal dynamics of the time series. More specifically, in this paper we discover and utilize a somewhat surprising observation that, for a wide class of non-stationary times series, progressive convolutions of their block sums and i.i.d. standard normal random variables consistently mimic the complex joint probabilistic behavior of their partial sum processes. Hence structural change tests for non-stationary time series can be easily performed by generating large samples of the latter convolutions. While remaining consistent for a much larger class of time series, the proposed bootstrap procedure is shown to have the same rate of accuracy (in terms of estimating the true covariance structure) and can detect local alternatives with the same $\sqrt{n}$ parametric rate as the conventional tests. The above theoretical findings are supported by our finite sample Monte Carlo experiments in which it is found that our bootstrap enjoys similar accuracy and power to the conventional tests when the time series is second order stationary. However, for second order non-stationary time series, our Monte Carlo simulations show that the robust bootstrap remains accurate while the conventional tests are invalid, as indicated by our theoretical findings.

The proposed robust bootstrap goes beyond testing changes in mean. It can be easily extended to testing structural stability of multidimensional parameters of higher order non-stationary time series, such as auto and cross covariances and product moments. The robust bootstrap procedure can also adapt to various testing ideas, such as the CUSUM test, the Lagrange multiplier test and the Cramer-von-Mises test. In that sense the proposed robust procedure provides a unified framework for structural change detection of non-stationary time series.

Recently, there has been an increasing interest in non-stationary time series analysis in statistics. It seems that two distinct classes of non-stationary time series models are of major interest. One class is the locally stationary time series models in which the data generating mechanism is assumed to be smoothly changing over time. See for instance Priestley (1988), Dahlhaus (1997), Nason et al. (2000), Ombao et al. (2005) and Zhou and Wu (2009). Note that the locally stationary class does not allow abrupt changes. The other class is the piecewise stationary time series models where the time series is divided into several segments and in each segment the process is assumed to be stationary. See for instance Keogh et al. (2001) and Davis et al. (2006). Note that the time series can experience abrupt changes at boundaries of segments. However, the piecewise stationary class does not admit smooth changes. It seems that current theory and methodology for the two classes of non-stationary time series are separate and do not apply to each other. One contribution of the paper is to unify the two classes of non-stationary time series into a class called the piecewise locally stationary (PLS) time series models in which the system can experience abrupt changes at several break points over time and the data generating mechanism changes smoothly between adjacent break points. The PLS class is more flexible and realistic in many applications. All theory and methodology of the paper are developed for the PLS class. Hence our robust bootstrap can be viewed as a unified approach which applies to a general class of non-stationary temporal dynamics.

There is a long-standing literature in statistics discussing time series structural change detection. It is impossible to have a complete list here and we will only mention some representative works. See for instance Page (1954) for testing changes in mean by the CUSUM statistic, Picard (1985) for likelihood ratio test and tests for changes in the spectrum, Inclán and Tiao (1994) and Lee and Park (2001) for change point detection in marginal variance, Berkes, Gombay, and Horvath (2009) and Galeano and Pena (2007) for change detection in the autocovariance function, Brown et al. (1975), Davis, Huang, and Yao (1995), Kokoszka and Leipus (2000) and Aue et al. (2008) for structural stability tests for various time series regression models. Other contributions include Wu et al. (2001), Siegmund and Yakir (2008) and Shao and Zhang (2010), among others. We also refer to the monographs of Csörgő and Horváth (1997) and Brodsky and Darkhovsky (2003) for more discussions and references.

The rest of the paper is organized as follows. In Section 2 we will take a detour to introduce the piecewise locally stationary (PLS) models under which the theory and methodology of the paper are developed. In Section 3 we will investigate the theoretical behavior of the conventional structural change tests under PLS. The robust bootstrap and its implementation are presented in Section 4. Extensions of the robust bootstrap to general multivariate parameters and other testing methods are discussed in Section 5. In Section 6, we will perform finite sample Monte Carlo experiments to study the accuracy and power of the proposed robust bootstrap and compare them to those of the conventional testing procedures. One environmental an one financial time series are analyzed in Section 7 and finally the theoretical results are proved in Section 8.

# 2 Piecewise locally stationary time series models

When testing possible changes in the mean of time series, it is of importance to propose time series models which allow the data generating mechanism of the system to experience general forms of structural changes, both smoothly and abruptly. For the latter purpose, we propose the following family of piecewise locally stationary time series.

**Definition 1.** *We say $\{e_i\}_{i=1}^n$ is piecewise locally stationary with $r$ break points (PLS($r$)) 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*

$$e_i = G_j(t_i, \mathcal{F}_i), \text{ if } b_j < t_i \le b_{j+1}, \tag{3}$$

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

If for each $j = 0, 1, \cdots, r$, $G_j(t, \cdot)$ is a smooth function of $t$, then we observe from (3) that the data generating mechanism of $\{e_i\}$ changes smoothly on $(b_j, b_{j+1})$, $j = 0, \cdots, r$. However, the system may undergo abrupt changes at break points $b_1, b_2, \cdots, b_r$. Note that the number of breaks $r$, the break points $b_1, \cdots, b_r$ and the nonlinear filters $G_0, \cdots, G_r$ are unknown nuisance parameters. Observe that if $r = 0$, then $\{e_i\}$ is a locally stationary times series in the sense of Zhou and Wu (2009). Additionally, if $G_j(t, \mathcal{F}_i)$ does not depend on $t$, $j = 0, \cdots, r$, then $\{e_i\}$ reduces to a piece-wise stationary time series. Associated with

Definition 1, we propose the following dependence measures to quantify the strength of temporal dependence of PLS time series.

**Definition 2** (Physical dependence measures for PLS time series)**.** *Let* $\{\varepsilon_i'\}$ *be an i.i.d. copy of* $\{\varepsilon_i\}$*. Consider the PLS(r) time series* $\{e_i\}$ *defined in* (3)*. Assume that* $\max_{1\le i\le n}\|e_i\|_p < \infty$ *for some positive* $p$*, where* $\|\cdot\|_p = [\mathbb{E}|\cdot|^p]^{1/p}$ *is the* $\mathcal{L}_p$ *norm of a random variable. Write* $\|\cdot\| := \|\cdot\|_2$*. For* $k \ge 0$*, define the* $k$*th physical dependence measure*

$$\delta_p(k) = \max_{0\le i\le r} \sup_{b_i\le t\le b_{i+1}} \|G_i(t,\mathcal{F}_k) - G_i(t,(\mathcal{F}_{-1},\varepsilon_0',\varepsilon_1,\cdots,\varepsilon_k))\|_p. \tag{4}$$

*Define* $\delta_p(k) = 0$ *if* $k < 0$*.*

If we view (3) as a time-varying physical system with $\{\varepsilon_i\}$ being the inputs and $\{e_i\}$ being the outputs, then $\delta_p(k)$ measures the magnitude of change in the system's output when the input of the system $k$ steps ahead is replaced by an i.i.d. copy. Strong temporal dependence of the time series is marked by large values of $\delta_p(k)$. To help understand the generality of the formulation (3) and the dependence measure (4), we list two examples as follows.

(i) (PLS linear processes)

$$G_k(t,\mathcal{F}_i) = \sum_{j=0}^{\infty} a_{k,j}(t)\varepsilon_{i-j} \quad b_k < t \le b_{k+1}, \tag{5}$$

where $a_{k,j}(\cdot)$'s are Lipschitz continuous functions. We observe that on each $(b_k, b_{k+1})$, the system is linear with smoothly varying coefficients. The coefficient functions may break at $b_1,\cdots,b_r$, which leads to abrupt changes in the covariance structure of the time series at the latter points. Additionally, straightforward calculations show that $\delta_p(k) = O(\max_{0\le i\le r}\sup_{0\le t\le 1}|a_{i,k}(t)|)$ if $\|\varepsilon_0\|_p < \infty$.

(ii) (PLS nonlinear processes)

$$G_k(t,\mathcal{F}_i) = R_k(t, G_k(t,\mathcal{F}_{i-1}),\varepsilon_i), \quad b_k < t \le b_{k+1}, \tag{6}$$

where $R_k(t,\cdot,\cdot)$ is a smooth function of $t$. Many stationary nonlinear time series models, including the (G)ARCH models (Engle 1982, Bollerslev 1986), threshold models (Tong 1990) and bilinear models, can be written in the form $X_i = R(X_{i-1}, \varepsilon_i)$. It is easily seen that, at each interval $(b_k, b_{k+1})$, (6) naturally extends the latter stationary models into the non-stationary domain by allowing the data generating mechanism, $R_k$, to vary smoothly with time. Additionally, the nonlinear filters are allowed to experience jumps or change from one to another at break points $b_1, \cdots, b_r$, which grants great flexibility in modeling complex dynamics of the time series.

Note that, at each interval $(b_k, b_{k+1})$, the time series model (6) is locally stationary in the sense of Zhou and Wu (2009). The physical dependence measures $\delta_p(k)$ in (4) can be easily calculated from Proposition 2 and Theorem 6 of the latter paper.

# 3 Limiting behavior of CUSUM test for PLS time series

To test the null hypothesis $H_0$, we assume the observed time series

$$X_i = \mu_i + e_i, \text{ where } \{e_i\} \in \text{PLS}(r) \text{ with } \mathbb{E}e_i = 0 \tag{7}$$

and unknown $r \geq 0$. As discussed in the previous section, model (7) allows very general forms of structural changes in the temporal dependence structure of the error process $\{e_i\}$, which facilitates an in-depth investigation of the behavior of various testing procedures under complex temporal dynamics.

Consider the testing problem (1). We will consider in this paper the general alternative

$$H_a^{PS} : \mu_i = \mu(t_i) \text{ for some piecewise Lipschitz continuous nonconstant function } \mu(\cdot), \tag{8}$$

where the number of discontinuities in $\mu(\cdot)$ is assumed to be bounded throughout the paper and $t_i = i/n$. Here the superscript $PS$ stands for 'piecewise smooth'. The piecewise smooth alternatives include the popular smooth and break point alternatives as special cases, where

$$\begin{aligned} H_a^{smooth} &: \mu_i = \mu(t_i) \text{ for some smooth nonconstant function } \mu(\cdot), \\ H_a^{break} &: \mu_1 = \cdots = \mu_{k^*} \neq \mu_{k^*+1} = \cdots = \mu_n \text{ for some } k^* \in (1, n). \end{aligned}$$

To test $H_0$, we shall mainly focus on the classic cumulative sum (CUSUM) test statistic $T_n$ in (2) though other tests such as the Lagrange multiplier test and the Cramer-von-Mises test will also be discussed in Section 5.2. It is easy to see that when $H_0$ is violated, then one is expected to observe large values of $T_n$. Before we state the results, we need the following regularity conditions.

(A1) The process $\{e_i\}$ is piecewise stochastic Lipschiz continuous. Namely for all $i \in [0, r]$ and all $t, s \in [b_i, b_{i+1}]$, $t \neq s$, we have

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

for some finite constant $C$.

(A2) $\|G_i(t, \mathcal{F}_0)\|_4 < \infty$ for all $i \in [0, r]$ and $t \in [b_i, b_{i+1}]$.

(A3) $\delta_4(k) = O(\chi^k)$ for some $\chi \in (0, 1)$.

(A4) Define the long-run variance function

$$\sigma^2(t) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(G_i(t, \mathcal{F}_0), G_i(t, \mathcal{F}_k)) \text{ if } t \in (b_i, b_{i+1}].$$

Let $\sigma^2(0) = \lim_{t \downarrow 0} \sigma^2(t)$. Assume that $\inf_{t \in [0,1]} \sigma^2(t) > 0$.

A few remarks on the regularity conditions are in order. Condition (A1) asserts that the filter $G_i(t, \cdot)$ is smoothly varying in time on the interval $(b_i, b_{i+1}]$, $i = 0, 1, \cdots, r$. Condition (A2) requires the time series to have fourth moment. Condition (A3) requires that the temporal dependence of $\{e_i\}$ decays exponentially fast to 0. As we mentioned in Examples (i) and (ii), (A3) can be verified easily for a large class of PLS linear and nonlinear models. Additionally, the theoretical results of the paper can be established with $\delta_4(k)$ decaying algebraically. However, the technical details are much more involved and we shall stick to condition (A3) for presentational simplicity. Condition (A4) is very mild. It means that the long run variance (or spectral density at frequency 0) of the time series $\{e_i\}$ is non-degenerate over time.

**Theorem 1.** *Under condition (A) and the null hypothesis $H_0$, we have*

$$T_n / \sqrt{n} \Rightarrow \sup_{0 \leq t \leq 1} |U(t) - tU(1)|, \tag{9}$$

*where $\Rightarrow$ stands for convergence in distribution and $U(t)$ is a zero-mean Gaussian process with covariance function $\gamma(t, s) = \int_0^{\min(t,s)} \sigma^2(r)\, dr$.*

Theorem 1 establishes the limiting null distribution of the CUSUM test when the time series is PLS. Observe that if the long-run variances do not change over time; namely $\sigma^2(t) = \sigma^2$ for all $t$, then it is clear that $U(t) = \sigma B(t)$, where $B(t)$ is a standard Brownian motion on $[0, 1]$. As a consequence the classic change point detection technique using the Brownian bridge limit is asymptotically correct under $H_0$. However, due to the abrupt and smooth changes in the dynamics of PLS time series, $\sigma(t)$ could change very flexibly over time. Consequently the covariance structure of $U(t)$ can be very complicated.

**Theorem 2.** *[Local power of CUSUM test for PLS time series] Assume that $\mu(\cdot) = \mu + L_n f(\cdot)$, where $f(\cdot)$ is a nonconstant piecewise Lipschitz continuous function. Further assume condition (A) holds. Then (i): if $L_n = n^{-1/2}$, we have*

$$T_n/\sqrt{n} \Rightarrow \sup_{0\le t\le 1} |U(t) - tU(1) + \int_0^t f(s)\,ds - t\int_0^1 f(s)\,ds|.$$

*(ii): If $L_n\sqrt{n} \to \infty$, then*

$$T_n/\sqrt{n} \to \infty \text{ in probability.}$$

We observe from Theorem 2 that, under PLS, the CUSUM test can detect local alternatives with the same rate $n^{-1/2}$ as the classic stationary case. Therefore the CUSUM test continue to serve as a powerful tool for structural change detection in the PLS case. The asymptotic local power of the test under case (i) can be calculated directly when combining results of Theorems 1 and 2.

## 3.1 Behavior of traditional testing procedures under PLS

The classic way to perform the CUSUM test (2) utilizes consistent estimators of the long-run variance of the series. If the time series is weakly stationary under the null hypothesis, then $T_n$ divided by the latter consistent estimator will have a Brownian bridge limit. There is a rich literature in statistics and econometrics discussing consistent long-run variance estimators, most of which belong to a class of estimators called heteroscedasticity and autocorrelation consistent covariance matrix estimators. See for instance Andrews (1991) and Newey and West (1987). A popular representative in the class is the classic lagged window estimate, which is defined as

$$\hat{\sigma}^2_{LW} = \frac{m}{n-m+1} \sum_{j=1}^{n-m+1} \Big( \frac{1}{m} \sum_{i=j}^{j+m-1} X_i - \frac{1}{n} S_n \Big)^2. \qquad (10)$$

**Proposition 1.** *Under $H_0$, condition (A) and the assumption that $m \to \infty$ with $m/n \to 0$, we have*

$$\hat{\sigma}^2_{LW} \to \int_0^1 \sigma^2(t)\, dt \ \text{ in probability.}$$

*Consequently*

$$T_n/(\hat{\sigma}_{LW}\sqrt{n}) \Rightarrow \sup_{0 \le t \le 1} |U(t) - tU(1)| / \sqrt{\int_0^1 \sigma^2(t)\, dt}, \qquad (11)$$

*where $U(\cdot)$ was defined in Theorem 1.*

Proposition 1 follows from (29) in Section 8. Proposition 1 establishes that the lagged window estimator converges to the integration of the long-run variance function when the time series is PLS. We observe from (11) that the null distribution of the classic test is no longer pivotal under PLS due to the fact that the long-run variance $\sigma^2(t)$ is time varying. Due to the complex nature of $U(\cdot)$, it is easy to show along the lines of the proof of Proposition 1 that $T_n$ cannot be made pivotal even when normalized by a sequence of covariance estimates. Specifically, generally the statistic

$$T_n' = \max_i \{ |S_i - t_i S_n| / \sigma_{n,i} \}$$

is not pivotal under PLS no matter how one chooses the normalizing constants $\sigma_{n,i}$. Therefore, generally the classic idea of normalization in the PLS case yields in a mixed structural change signal of the first and second order, which can be biased when one's interest is in testing $H_0$.

Recently, There are several interesting results on structural break detection using functionals of inconsistently normalized CUSUM statistics; see for instance Shao and Zhang (2010) and Sayginsoy and Vogelsang (2011). The latter type of tests are shown to have better finite sample Type I error rates. Following Shao and Zhang (2010), define

$$V_n(k) = n^{-1} [\sum_{t=1}^{k} \{S_{1,t} - \frac{t}{k} S_{1,k}\}^2 + \sum_{t=k+1}^{n} \{S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n}\}^2],$$

$k = 1, \cdots, n-1$. Shao and Zhang (2010) proposed the test statistic

$$G_n = \max_{k=1,\cdots,n-1} T_n^2(k)/V_n(k), \text{ where } T_n(k) = S_k - t_k S_n. \tag{12}$$

By Proposition 5 in Section 8, we have the following

**Proposition 2.** *Under $H_0$ and condition (A), we have*

$$G_n \Rightarrow \sup_{s\in[0,1]} \{U(s) - sU(1)\}^2/V(s), \tag{13}$$

*where* $V(s) = \int_0^s \{U(t)-(t/s)U(s)\}^2\, dt + \int_s^1 [U(1)-U(t)-(1-t)/(1-s)\{U(1)-U(t)\}]^2\, dt$.

We observe from Proposition 2 that, under weak stationarity, we have $U(t) = \sigma B(t)$ and the limiting distribution in (13) is pivotal. The latter observation facilitates the testing procedure in Shao and Zhang (2010). However, as we see from the above proposition, the pivotality is sensitive to the change of the long-run variance. Consequently, the test can be biased under PLS.

# 4 The robust bootstrap procedure

From discussions in Section 3, we observe that the key to accurate tests of structural change under PLS is to account for the time-varying covariance structure. One direct but naive approach is to estimate the long-run variance function $\sigma(\cdot)$ on $[0,1]$ and then use Theorem 1 to find the critical values of the CUSUM test. However, two important factors hamper the practical applicability of the latter idea. First, estimating $\sigma(\cdot)$ requires correct estimation of the break points $b_1, \cdots, b_r$ since most nonparametric estimators become inconsistent near the break points. However, estimating the number and locations of the break points is an extremely difficult task for PLS time series. Furthermore, bias and variance associated with estimating the break points could seriously affect the accuracy of the CUSUM test for finite samples. Second, for time spans where $\sigma(\cdot)$ is smooth, generally one needs two tuning parameters to consistently estimate $\sigma(\cdot)$ at time $t$. Specifically, one needs one time bandwidth to constrain the estimation in a neighborhood of $t$ and another window size to cut-off the auto correlation which is similar to the roll of $m$ in (10). It is generally desirable to have as few nuisance parameters as possible to avoid loss of finite sample accuracy of the test caused by selection bias of the nuisance parameters.

In this Section, we shall propose a bootstrap procedure for structural change detection of PLS time series that can be easily implemented in practice. To begin with, for a fixed window size $m$, define the process

$$\Phi_{i,m} = \sum_{j=1}^{i} \frac{1}{\sqrt{m(n-m+1)}}(S_{j,m} - \frac{m}{n}S_n)R_j, \qquad i = 1, \cdots, n-m+1,$$

where $S_{j,m} = \sum_{r=j}^{j+m-1} X_r$ and $(R_i)_{i=1}^n$ are i.i.d. standard normal and are independent of $(X_i)_{i=1}^n$. On $[1/n, (n-m+1)/n]$, define the associated linear interpolation

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

where $t_* = \lfloor tn \rfloor / n$ and $t^* = t_* + 1/n$. Our bootstrap methodology is based on the following key theorem:

**Theorem 3.** *Assume that condition (A) holds. Further assume that $H_0$ holds and $m \to \infty$ with $m/n \to 0$. Then conditional on $(X_i)_{i=1}^n$, we have*

$$\tilde{\Phi}_{m,n}(t) \Rightarrow U(t) \text{ on } \mathcal{C}(0,1) \text{ with the uniform topology.} \tag{14}$$

*Consequently, conditional on $(X_i)_{i=1}^n$,*

$$\max_{m+1 \le i \le n-m+1} \left| \Phi_{i,m} - \frac{i}{n-m+1} \Phi_{n-m+1,m} \right| \Rightarrow \sup_{0 \le t \le 1} |U(t) - tU(1)|. \tag{15}$$

Though the covariance structure of $\{U(t)\}_{t=0}^1$ can be very complex with both abrupt and smooth changes, Theorem 3 reveals that the probabilistic behavior of $\{U(t)\}_{t=0}^1$ can be fully characterized by that of a simple process $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ under the null hypothesis. The process $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ do not involve estimation of the long-run variance function $\sigma(\cdot)$ and it contains only one tuning parameter $m$. Based on Theorem 3, one can generate a large (say 2000) sample of conditionally i.i.d. copies of $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ and use the sample distribution to consistently approximate the null distribution of the CUSUM test. The latter task is easy and fast to accomplish in practice since generating $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ only involves the partial sums of the original time series and i.i.d. Gaussian random variables. Specifically, the following are the detailed steps of our robust bootstrap:

1 Select the window size $m$ according to the rules in Section 4.1.

2 Generate $B$ (say 2000) conditionally i.i.d. copies $\{\Phi_{i,m}^{(r)}\}_{i=1}^{n-m+1}$, $r = 1, 2, \cdots, B$. Let $M_r = \max_{m+1 \le i \le n-m+1} |\Phi_{i,m}^{(r)} - \frac{i}{n-m+1}\Phi_{n-m+1,m}^{(r)}|$.

3 Let $M_{(1)} \le M_{(2)} \le \cdots \le M_{(B)}$ be the ordered statistics of $M_r$ $r = 1, 2, \cdots B$. Reject $H_0$ at level $\alpha$ if $T_n/\sqrt{n} > M_{(\lfloor B(1-\alpha) \rfloor)}$, where $\lfloor x \rfloor$ denotes the largest integer smaller or equal to $x$. Let $B^* = \max\{r : M_{(r)} \le T_n/\sqrt{n}\}$. The $p$-value of the test can be obtained by $1 - B^*/B$.

**Remark 1.** We now discuss the connections and differences between the proposed robust bootstrap and various resampling methods in the literature. The proposed bootstrap is an extension of the class of wild bootstrap in Wu (1986). Additionally, The robust bootstrap borrows the ideas of moving block bootstrap (Lahiri 2003) and subsampling (Politis et al. 1999) by utilizing block sums $S_{i,m}$ with the auxiliary variables to account for temporal dependence. Indeed, note that $\Phi_{n-m+1,m} = \sigma_{LW} R^*$, where $R^*$ is a standard normal random variable and $\sigma_{LW}$ is defined in (10). Hence $\Phi_{n-m+1,m}$ is equivalent to the classic lag-window or subsampling long-run variance estimator.

On the other hand, there are two major distinctions between the proposed robust bootstrap and aforementioned resampling methods. First, the robust bootstrap focuses on uniformly mimicking the behavior of the whole partial sum process $\{S_j\}_{j=1}^n$ while conventional resampling methods typically focus on inferencing functionals of $S_n$. Note that the probabilistic structure of $\{S_j\}_{j=1}^n$ is much more complicated than that of $S_n$. As a result methods that are perfectly suitable for the inference of $S_n$ can be inconsistent for the partial sum process. Second and most importantly, the robust bootstrap is proposed as a unified methodology for piecewise locally stationary processes while most conventional resampling methods are targeted at stationary processes. As we already demonstrated in Section 3.1, resampling methods which work well in the stationary case generally do not carry over to the non-stationary domain. In general, resampling under complex temporal dynamics needs to be carefully designed based on a thorough understanding of what is to be mimicked. $\diamondsuit$

The following proposition studies the power performance of the proposed robust bootstrap procedure:

**Proposition 3.** *Assume that condition (A) holds and that $m \to \infty$ with $m/n \to 0$. Then conditional on $(X_i)_{i=1}^n$, we have i): under the fixed alternative $H_a^{PS}$,*

$$\max_{1 \le i \le n-m+1} \left| \Phi_{i,m} - \frac{i}{n-m+1} \Phi_{n-m+1,m} \right| / \sqrt{m} \Rightarrow \sup_{0 \le t \le 1} |U^*(t) - tU^*(1)|,$$

*where $U^*(t)$ is a zero mean Gaussian process with covariance function $\gamma^*(t,s) = \int_0^t [\mu(r) - \int_0^1 \mu(u)\, du]^2\, dr$, $0 \le t \le s \le 1$. ii): Under the local alternative that $\mu(\cdot) = n^{-1/2} f(\cdot)$, where $f(\cdot)$ is a nonconstant piecewise Lipschitz continuous function, we have* (15).

Part i) of the above proposition implies that the bootstrap statistic converges at the rate $\sqrt{m}$ under the fixed alternative $H_a^{PS}$. Note that the CUSUM statistic $T_n/\sqrt{n}$ converges at the rate $\sqrt{n}$ under $H_a^{PS}$ and $m \ll n$. Hence i) implies that the robust bootstrap procedure has asymptotic power 1 under the fixed alternative. On the other hand, ii) of the above proposition establishes that the bootstrap statistic converges to the limiting null distribution of the CUSUM statistic $T_n/\sqrt{n}$ under the local alternative. Consequently the robust bootstrap enjoys the same asymptotic local power as that in Theorem 2 (i). In particular, the robust bootstrap procedure can detect local alternatives with the $\sqrt{n}$ parametric rate, which is the same as the conventional long-run variance normalization method. We conclude from the above discussion that the robust bootstrap serves as a powerful tool for structural change detection for PLS time series.

## 4.1 Window size selection

### 4.1.1 Theoretically optimal window size

Note that the probabilistic structure of a Gaussian process is fully determined by its covariance function. By the proof of Theorem 3, we observe that the linear interpolation of $(\Phi_{i,m})$ converges weakly to $U(t)$ on $\mathcal{C}(0,1)$. Hence for each pair $(r,s)$, $m < r \le s \le n-m+1$,

$$\hat{\gamma}_m(r/n, s/n) := \mathrm{Cov}(\Phi_{r,m}, \Phi_{s,m}) = \sum_{j=1}^{r} (S_{j,m} - mS_n/n)^2 / (m(n-m+1))$$

is a consistent estimate of the covariance $\gamma(r/n, s/n)$ of $U(\cdot)$. Here the covariance is conditional on the observations $(X_i)$. The quality of the bootstrap based test is determined by the accuracy of the latter estimation. We propose to use

$$L(m) := \max_{1\le r\le s\le n-m+1} \left\|\hat{\gamma}_m(r/n, s/n) - \gamma(r/n, s/n)\right\| = \max_{1\le r\le n-m+1} \left\|\hat{\gamma}_m(r/n, r/n) - \gamma(r/n, r/n)\right\| \quad (16)$$

to quantify the accuracy of the estimation. Note that, for any fixed pair $(r, s)$, the quantity $\|\hat{\gamma}_m(r/n, s/n) - \gamma(r/n, s/n)\|^2$ is the mean squared error (MSE) of $\hat{\gamma}_m(r/n, s/n)$. We shall call $L^2(m)$ the uniform mean squared error (UMSE) of $\hat{\gamma}_m(r/n, s/n)$.

To implement the bootstrap test, one selects $\hat{m}$ which minimizes the UMSE $L^2(m)$. By Lemmas 3 and 4 in Section 8, we have the following

**Theorem 4.** *Under regularity conditions of Theorem 3 and the null hypothesis of no structural change, we have*

$$L(m) = O(\sqrt{\frac{m}{n}} + \frac{1}{m}). \quad (17)$$

*Consequently,* $\hat{m} = O(n^{1/3})$ *with the optimal root UMSE* $L(\hat{m}) = O(n^{-1/3})$.

In Theorem 4, the terms $\sqrt{m/n}$, $1/m$ correspond to the standard deviation of $\hat{\gamma}_m(\cdot, \cdot)$ and the bias of the latter estimate, respectively. It is well known that the lag window estimator of long-run variance converges at the rate $n^{-1/3}$. Hence our robust bootstrap enjoys the same convergence rate as the classic method of structural change. The above theoretical finding is reflected in our simulation results that the robust bootstrap and the long-run variance normalization method are similarly accurate for second order stationary time series.

#### 4.1.2 Practical implementation

For structure change detection problems in general, the constants for the optimal window size are difficult to derive or estimate. To overcome the difficulty, one popular alternative is to pretend that the time series follow a simple parametric model (such as an AR(1) process) and use the optimal window size derived under the latter model; see for instance Andrews (1991). However, the aforementioned window size selection rule can perform unsatisfactorily when the parametric model does not adequately describe the second order structure of the time series. Note that, in this paper, the second order structure of series is allowed to be very complex and hence generally the latter parametric window size selection rule is not recommended under our setting. In our Monte Carlo experiments and real data analysis, we find that the minimum volatility (MV) method as advocated in Politis et. al (1999) performs reasonably well when the dependence structure of the time series is complex. The advantage of the MV methods is that it does not depend on the specific form of the underlying time series dependence structure and hence is robust to misspecification of the latter structure. See Chapter 9 of Politis et. al (1999) for more discussions. Specifically, the MV method utilizes the fact that the the estimator $\hat{\gamma}_m(t, s)$ becomes stable when the block size $m$ is in an appropriate range. Therefore one could first propose a grid of possible window sizes $m_1 < m_2 < \cdots < m_M$ and obtain $\{\hat{\gamma}_{m_j}(r/n, r/n)\}_{r=1}^{n-m_j+1}$, $j = 1, \cdots, M$. For each $m_j$, calculate

$$\max_{1\le r\le n-m_M+1} se\Big(\{\hat{\gamma}_{m_{j+k}}(r/n, r/n)\}_{k=-3}^{3}\Big), \tag{18}$$

where $se$ denotes standard error, namely

$$se\Big(\{\hat{\gamma}_{m_j}(r/n, r/n)\}_{j=1}^{k}\Big) = \Big[\frac{1}{k-1}\sum_{j=1}^{k} |\hat{\gamma}_{m_j}(r/n, r/n) - \bar{\hat{\gamma}}(r/n, r/n)|^2\Big]^{1/2}$$

with $\bar{\hat{\gamma}}(r/n, r/n) = \sum_{j=1}^{k} \hat{\gamma}_{m_j}(r/n, r/n)/k$. Clearly (18) measures the uniform variability of $\hat{\gamma}_m(\cdot, \cdot)$ as a function of $m$. Then one chooses $\hat{m}_{MV}$ which minimizes the above maximized standard errors.

# 5 Extensions

## 5.1 Extensions to multidimensional statistics of PLS time series

In this section, we shall extend the robust bootstrap procedure to general multidimensional statistics of PLS time series. For the latter purpose, we first define a $d$-dimensional PLS time series $\{\mathbf{X}_i\}_{i=1}^n$ the same way as in (3) except that now $G_j(\cdot)$, $j = 0, \cdots, r$ are maps from $\mathbb{R}^\infty$ to $\mathbb{R}^d$. Here we shall consider testing structure changes of

$$\boldsymbol{\eta}_i = \mathbb{E}\phi(\mathbf{X}_i, \mathbf{X}_{i-1}, \cdots, \mathbf{X}_{i-q}), \text{ where } \{\mathbf{X}_i\} \text{ is a } d\text{-dimensional PLS}(r) \text{ time series} \tag{19}$$

and $\phi : \mathbb{R}^{qd} \to \mathbb{R}^p$. Structure change problems covered in (19) is large. Prominent examples include i) : testing structure changes in general moments and product moments of a time series where $\boldsymbol{\eta}_i = \mathbb{E}|X_i|^r$, $r \in \mathbb{R}$ or $\boldsymbol{\eta}_i = \mathbb{E}[X_{i-d_1} \cdots X_{i-d_q}]$, $0 \le d_1 \le d_2 \le \cdots \le d_q$ are integers and $\{X_i\}$ is a univariate PLS time series; ii) : testing changes in auto- and cross-covariance matrices of multivariate PLS time series where $\boldsymbol{\eta}_i = \mathbb{E}\mathbf{X}_{1,i}\mathbf{X}_{2,i-k}^\top$ and $\{\mathbf{X}_{1,i}\}$ and $\{\mathbf{X}_{2,i}\}$ are $d_1$- and $d_2$-dimensional zero-mean PLS time series, respectively; iii) : detecting structure changes in characteristic functions of a multivariate PLS time series where $\boldsymbol{\eta}_i = \mathbb{E}\exp(\sqrt{-1}\langle s, \mathbf{X}_i\rangle)$, $s \in \mathbb{R}^d$, $\langle .,. \rangle$ denotes inner product and $\{\mathbf{X}_i\}$ is a $d$-dimensional PLS time series. Note that a complex number can be written as a 2-dimensional real valued vector of its real and imaginary parts. Hence case iii) belongs to (19).

Generally, to test the null hypothesis that $\boldsymbol{\eta}_i$ remains constant over time, one could use the CUSUM test statistic

$$T_n(\phi) = \max_{1+q \le i \le n} |\mathbf{S}_i(\phi) - t_i \mathbf{S}_n(\phi)|, \text{ where } \mathbf{S}_i(\phi) = \sum_{j=q+1}^{i} \phi(\mathbf{X}_j, \mathbf{X}_{j-1}, \cdots, \mathbf{X}_{j-q}). \quad (20)$$

Write $\mathbf{Y}_i = \phi(\mathbf{X}_i, \mathbf{X}_{i-1}, \cdots, \mathbf{X}_{i-q})$, $i = q+1, q+2, \cdots, n$. Then $\{\mathbf{Y}_i\}$ is a $p$-dimensional PLS time series. Specifically, similar to (3), there exist constants $0 = b_0 < b_1 < \cdots < b_r < b_{r+1} = 1$ and $\mathbb{R}^p$-valued zero-mean nonlinear filters $\mathbf{W}_0, \mathbf{W}_1, \cdots, \mathbf{W}_r$, such that

$$\mathbf{Y}_i = \boldsymbol{\eta}_i + \mathbf{W}_j[t_i, (\cdots, \varepsilon_{i-1}, \varepsilon_i)], \text{ if } b_j < t_i \le b_{j+1}, \quad (21)$$

where $\varepsilon_i$'s are i.i.d. random variables. Define physical dependence measures $\delta_p(k)$ of $\{\mathbf{Y}_i\}$ similar to (4). The following theorem extends the robust bootstrap to general multidimensional statistics of PLS time series:

**Theorem 5.** *Suppose that conditions (A1)-(A3) hold with $G_j$ therein replaced by $\mathbf{W}_j$. Assume that (A4)$^*$ : the smallest eigenvalue of $\Sigma^2(t)$ is bounded away from 0 on $(0,1]$, where $\Sigma^2(t) = \sum_{k=-\infty}^{\infty} Cov(\mathbf{W}_i(t, \mathcal{F}_0), \mathbf{W}_i(t, \mathcal{F}_k))$ if $t \in (b_i, b_{i+1}]$. We have i) : under the null hypothesis that $\boldsymbol{\eta}_i$ stays constant over time,*

$$T_n(\phi)/\sqrt{n} \Rightarrow \sup_{0 \le t \le 1} |\mathbf{U}(t) - t\mathbf{U}(1)|,$$

*where $\mathbf{U}(t)$ is a $p$-dimensional zero-mean Gaussian process with covariance function $\gamma(t,s) = \int_0^{\min(t,s)} \Sigma^2(r)\, dr$. ii) : Let*

$$\Phi_{i,m}(\phi) = \sum_{j=q+1}^{i} \frac{1}{\sqrt{m(n-m-q+1)}} (\mathbf{S}_{j,m}(\phi) - \frac{m}{n}\mathbf{S}_n(\phi)) R_j,$$

*$i = q+1, \cdots, n-m+1$, where $\mathbf{S}_{j,m}(\phi) = \sum_{r=j}^{j+m-1} \mathbf{Y}_r$ and $(R_i)_{i=1}^n$ are i.i.d. univariate standard normal random variables and are independent of $(\mathbf{X}_i)_{i=1}^n$. Assume that $\boldsymbol{\eta}_i$ stays constant and $m \to \infty$ with $m/n \to 0$. Then conditional on $(\mathbf{X}_i)_{i=1}^n$, we have*

$$\max_{q+1 \le i \le n-m+1} \left| \Phi_{i,m}(\phi) - \frac{i}{n-m+1} \Phi_{n-m+1,m}(\phi) \right| \Rightarrow \sup_{0 \le t \le 1} |\mathbf{U}(t) - t\mathbf{U}(1)|.$$

Theorem 5 concludes that the simple robust bootstrap procedure applies to a wide range of multidimensional structural change detection problems when the higher order temporal dynamics of the time series is complex with both abrupt and smooth changes. Conditions (A1) - (A3) for $\{\mathbf{W}_i\}$ can often be verified easily from smoothness and short range dependence properties of $\{\mathbf{X}_i\}$. The following proposition provides sufficient conditions for (A1) - (A3) when $\phi(\cdot)$ is Lipschitz continuous or of product moment type.

**Proposition 4.** *Assume either (a): $\phi(\cdot)$ is Lipschitz continuous and $\{\mathbf{X}_i\}$ satisfies Conditions (A1)-(A3); or (b): $\phi(\cdot) = X_{i-d_1} \cdots X_{i-d_q}$, $0 \le d_1 \le d_2 \le \cdots \le d_q$ and $\{X_i\}$ satisfies Condition (A1) with $\|\cdot\|$ therein replaced by $\|\cdot\|_{2q}$ and Conditions (A2)-(A3) with $\|\cdot\|_4$ therein replaced by $\|\cdot\|_{4q}$. Then we have that Conditions (A1)-(A3) hold for $\mathbf{W}_j$.*

## 5.2 Extensions to other tests of structural change

Many existing tests of structural change in mean are in the form of (continuous) functionals of the partial sum process $\{S_i\}_{i=1}^n$. Based on (14) in Theorem 3, the probabilistic behavior of $\{S_i\}_{i=1}^n$ can be fully captured by $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ under $H_0$. Therefore by the celebrated continuous mapping theorem, the behavior of the latter tests of structural change under PLS can be consistently mimicked from the sampling distribution of the corresponding functionals of $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$. Hence the idea of our robust bootstrap can be naturally extended to the latter tests of structural change. In the following, we shall give two specific examples.

The first example is the Cramer-von-Mises test statistic

$$CM_n = \frac{1}{n} \sum_{i=1}^{n} [(S_i - t_i S_n)/\sqrt{n}]^2, \tag{22}$$

which corresponds to the $\mathcal{L}^2$ norm of the sequence $(S_i - t_i S_n)_{i=1}^n$. Note that the CUSUM test statistic $T_n$ is the $\mathcal{L}^\infty$ norm of the latter sequence. The second example we consider here is the Lagrange Multiplier or Rao's score type test (Rao (1948), Silvey (1959), Andrews (1993)):

$$LM_n = \max_{\lfloor \lambda n \rfloor \le i \le \lfloor (1-\lambda)n \rfloor} \left[ \frac{S_i^2}{i} + \frac{S_{i+1,n}^2}{n-i} - \frac{S_n^2}{n} \right], \tag{23}$$

where $\lambda > 0$ is a user chosen constant. When the null $H_0$ is violated, one is expected to observe large values of $CM_n$ and $LM_n$. The following theorem extends the robust bootstrap to the latter two tests:

**Theorem 6.** *Under $H_0$ and conditions $(A1)$ to $(A4)$, we have*

$$CM_n \Rightarrow \int_0^1 [U(t) - tU(1)]^2\, dt, \text{ and } LM_n \Rightarrow \sup_{\lambda \le t \le 1-\lambda} \left\{ \frac{U(t)^2}{t} + \frac{[U(1)-U(t)]^2}{1-t} - U(1)^2 \right\}, \tag{24}$$

*where $U(\cdot)$ is defined in Theorem 1. Additionally, if $m \to \infty$ with $m/n \to 0$, then conditional on $(X_i)_{i=1}^n$, we have*

$$\sum_{i=1}^{n-m+1} \left( \Phi_{i,m} - \frac{i}{n-m+1} \Phi_{n-m+1,m} \right)^2 / n \quad \Rightarrow \quad \int_0^1 [U(t) - tU(1)]^2\, dt,$$

$$\max_{\lfloor \lambda n \rfloor \le i \le \lfloor (1-\lambda)n \rfloor} \left[ \frac{\Phi_{i,m}^2}{i/n} + \frac{(\Phi_{n-m+1,m} - \Phi_{i+1,m})^2}{(n-i)/n} - \Phi_{n-m+1,m}^2 \right] \quad \Rightarrow \quad \sup_{\lambda \le t \le 1-\lambda} Q(t), \tag{25}$$

*where $Q(t) = \frac{U(t)^2}{t} + \frac{[U(1)-U(t)]^2}{1-t} - U(1)^2$.*

Theorem 6 follows easily from Theorems 1, 3 and the continuous mapping theorem. We see from (24) that the limiting behaviors of the two tests are complicated under PLS and traditional procedures using the long-run variance normalization fail to produce consistent testing results. Based on the above theorem, if $LM_n$ or $CM_n$ is used to test $H_0$ under PLS, then one can generate a large sample of conditionally i.i.d. copies of $\{\Phi_{i,m}\}_{i=1}^{n-m+1}$ and use the sample distribution of the corresponding statistic on the left hand side of (25) to perform the tests. The detailed procedure is very similar to those listed in Section 4 and is omitted here.

# 6 Simulations

In this section, we shall design Monte Carlo experiments to study the finite sample accuracy and sensitivity of the proposed robust bootstrap procedure and compare them with those of the existing tests of structural change. Throughout the section we will test the null hypothesis $H_0$ of no structural change in mean via the CUSUM test. The number of replications is fixed at 5000 and the number of bootstrap samples $B = 2000$.

## 6.1 Accuracy of the robust bootstrap

In this section, for comparative purpose, we introduce the AR(1) window size selection rule introduced in Andrews (1991). The performance of the latter rule will be compared with that of our MV method. Following Andrews (1991), one could assume that the observed time series $(X_i)$ follows an AR(1) process; i.e. $(X_{i+1} - \mu) = a(X_i - \mu) + \varepsilon_{i+1}$, $|a| < 1$. After elementary but tedious algebra, it can be shown that the optimal $m$ which minimizes the UMSE (16) coincides with that of Andrews (1991), namely

$$\hat{m}_{AR1} = \lfloor 1.1447 \Big[\frac{4a^2 n}{1 - a^2}\Big]^{1/3} \rfloor. \tag{26}$$

The AR coefficient $a$ can be estimated efficiently by the lag-1 sample auto correlation.

Now the finite sample Type I error rates of the following five methods are compared in this section:

(i) LW1 : the classic long-run variance normalizing method when the window size of the lag window estimator (10) is selected by the $AR(1)$ method in (26).

(ii) LW2 : the classic long-run variance normalizing method when the window size of the lag window estimator (10) is selected by the minimum volatility method described in Section 4.1.2.

(iii) SN : the self-normalizing method of Shao and Zhang (2010).

(iv) RB1 : the robust bootstrap when the window size is selected by the $AR(1)$ method described in (26).

(v) RB2 : the robust bootstrap when the window size is selected by the minimum volatility method described in Section 4.1.2.

Our first experiment aims at Type I error comparison when the time series is weakly stationary. In this case, the classic long-run variance normalizing method, the self-normalizing method and our robust bootstrap are all valid. We are interested in checking whether our robust bootstrap has similar performance to the existing methods under stationarity. The following three types of stationary linear models are considered: (a). AR(1) model with

AR-coefficient $a$. (b). ARMA(1,1) model with AR-coefficient $a$ and MA coefficient 0.5. (c). ARMA(1,1) model with AR-coefficient $a$ and MA coefficient $-0.6$. We select three levels of $a$, i.e. 0.2, 0.5 and 0.8 to represent weak, moderate and strong temporal dependence. The time series errors are independent standard normals. At nominal levels 5% and 10%, the simulated Type I error rates are listed in Tables 1 and 2 below for $n = 200$ and 500.

| | | $\alpha = 0.05$ | | | | | $\alpha = 0.1$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $a$ | LW1 | LW2 | SN | RB1 | RB2 | LW1 | LW2 | SN | RB1 | RB2 |
| (a) | 0.2 | 6.7 | 4.1 | 5.6 | 7.0 | 3.9 | 13.5 | 9.7 | 10.1 | 13.9 | 9.3 |
| (a) | 0.5 | 5.8 | 4.6 | 5.9 | 5.1 | 3.4 | 13.1 | 11.0 | 11.2 | 11.6 | 9.6 |
| (a) | 0.8 | 3.5 | 3.0 | 5.6 | 8.3 | 2.3 | 10.2 | 9.3 | 14.4 | 8.5 | 7.5 |
| (b) | 0.2 | 4.6 | 3.4 | 5.4 | 4.7 | 3.2 | 10.5 | 9.5 | 10.7 | 10.6 | 8.9 |
| (b) | 0.5 | 4.5 | 4.2 | 6.4 | 3.4 | 3.1 | 11.3 | 11.2 | 12.1 | 9.5 | 9.3 |
| (b) | 0.8 | 2.3 | 2.9 | 9.4 | 1.9 | 2.3 | 8.8 | 9.6 | 14.9 | 7.2 | 7.9 |
| (c) | 0.2 | 0.4 | 1.0 | 2.3 | 0.5 | 1.2 | 1.4 | 3.9 | 5.4 | 1.5 | 4.1 |
| (c) | 0.5 | 1.3 | 3.1 | 4.4 | 1.3 | 2.5 | 3.3 | 9.0 | 8.3 | 3.4 | 7.7 |
| (c) | 0.8 | 19.3 | 6.2 | 8.2 | 19.4 | 5.8 | 29.4 | 14.1 | 14.0 | 29.7 | 13.0 |

Table 1. Simulated type I error rates (in %) for the five methods with nominal levels 5% and 10% under stationary models $(a)$, $(b)$ and $(c)$. Series length $n = 200$.

| | | $\alpha = 0.05$ | | | | | $\alpha = 0.1$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $a$ | LW1 | LW2 | SN | RB1 | RB2 | LW1 | LW2 | SN | RB1 | RB2 |
| (a) | 0.2 | 6.6 | 5.8 | 5.1 | 6.5 | 5.6 | 11.8 | 11.0 | 9.4 | 12.2 | 11.2 |
| (a) | 0.5 | 6.9 | 6.5 | 6.0 | 6.1 | 5.9 | 12.9 | 12.4 | 11.0 | 12.3 | 11.8 |
| (a) | 0.8 | 5.2 | 6.0 | 6.2 | 4.5 | 5.2 | 11.4 | 12.6 | 10.9 | 10.0 | 11.3 |
| (b) | 0.2 | 4.7 | 4.7 | 4.8 | 4.4 | 4.3 | 9.8 | 9.7 | 9.4 | 9.4 | 9.3 |
| (b) | 0.5 | 5.2 | 5.0 | 5.7 | 4.6 | 4.6 | 10.9 | 10.8 | 10.8 | 10.0 | 9.8 |
| (b) | 0.8 | 4.3 | 6.3 | 6.8 | 3.3 | 5.3 | 10.5 | 12.6 | 11.6 | 9.0 | 11.6 |
| (c) | 0.2 | 0.6 | 1.5 | 4.0 | 0.5 | 1.4 | 1.8 | 4.5 | 7.6 | 1.8 | 4.6 |
| (c) | 0.5 | 1.6 | 3.8 | 5.5 | 1.7 | 3.4 | 4.2 | 9.0 | 9.7 | 4.3 | 8.6 |
| (c) | 0.8 | 18.9 | 6.8 | 6.0 | 19.0 | 6.0 | 29.6 | 13.8 | 11.2 | 29.7 | 12.9 |

Table 2. Simulated type I error rates (in %) for the five methods with nominal levels 5% and 10% under stationary models $(a)$, $(b)$ and $(c)$. Series length $n = 500$.

We observe from Tables 1 and 2 that the performance of the proposed robust bootstrap is reasonably accurate except in Case (c) with $a = 0.2$. In the latter case neither the robust bootstrap or the long-run variance normalization method yield accurate type I error rates. In fact, when there is a dominating negative MA term, several previous papers have documented decrease in the accuracy of change point testing or covariance estimation in moderate samples, see for instance Andrews (1991). The robust bootstrap has similar performances as the corresponding classic long-run variance normalization method for stationary time series. The latter two methods perform similarly to the self-normalizing method in Cases (a) and (b). However, in Case (c) where there is a negative moving average component, the self-normalizing method performs slightly better for $a = 0.2$ and 0.5. Additionally, we find that the AR(1) window size selection method is reasonably accurate in Cases (a) and (b). However, it performs unsatisfactorily in Case (c) in which the lag 1 auto correlation is no longer the dominating term in the second order structure of the time series. On the other hand, the minimum volatility method performs reasonably well in all three cases. The simulation results support our suggestion in Section 4.1.2 that the MV method is recommended when the second order structure of the time series is complex. Finally, as we expected, the performances of all three methods improve as the sample size increases.

Next, we study the accuracy of the above five methods for non-stationary time series. Specifically, we investigate the following four non-stationary models:

(I) : Consider the model

$$X_i = V(t_i)X_i^*, \qquad i = 1, 2, \cdots, n, \text{ where } V(t) = 1 + 4I\{t > 0.75\},$$

$\{X_i^*\}$ is a zero-mean AR(1) process with AR coefficient 0.5 and i.i.d. standard normal innovations and $I(\cdot)$ is the indicator function. Here the marginal variance of the time series is inflated by a factor of 25 after the break point $t = 0.75$. However, the correlation structure of the series is the same as that of an $AR(1)$ stationary series $\{X_i^*\}$. In particular, the non-stationarity purely comes from a sudden change of the marginal variance of the series.

(II) : Consider the model $X_i = G_0(t_i, \mathcal{F}_i)$ for $t_i \le 1/3$ and $X_i = G_1(t_i, \mathcal{F}_i)$ for $t_i > 1/3$, where

$$G_0(t, \mathcal{F}_i) = 0.5G_0(t, \mathcal{F}_{i-1}) + \varepsilon_i, \qquad G_1(t, \mathcal{F}_i) = -0.5G_1(t, \mathcal{F}_{i-1}) + \varepsilon_i,$$

and $\varepsilon_i$'s are i.i.d. $N(0, 1)$. Here the series is piece-wise stationary and the non-stationarity is due to an abrupt change of the AR(1) coefficient from 0.5 to $-0.5$ at the break point $t = 1/3$. Note that the marginal variance of the time series stays the same.

(III) : Here we have $X_i = G_0(t_i, \mathcal{F}_i)$, where

$$G_0(t, \mathcal{F}_i) = 0.75 \cos(2\pi t)G_0(t, \mathcal{F}_{i-1}) + \varepsilon_i,$$

and $\varepsilon_i$'s are i.i.d. $N(0, 1)$. There are no abrupt structural changes in the above model. Instead, the model is locally stationary in the sense that the AR(1) coefficient $0.75 \cos(2\pi t)$ changes smoothly on the interval $[0, 1]$.

(IV) : Consider the model $X_i = G_0(t_i, \mathcal{F}_i)$ for $t_i \leq 0.8$ and $X_i = G_1(t_i, \mathcal{F}_i)$ for $t_i > 0.8$, where

$$G_0(t, \mathcal{F}_i) = 0.75 \cos(2\pi t)G_0(t, \mathcal{F}_{i-1}) + \varepsilon_i, \qquad G_1(t, \mathcal{F}_i) = (0.5 - t)G_1(t, \mathcal{F}_{i-1}) + \varepsilon_i \quad (27)$$

and $\varepsilon_i$'s are i.i.d. $N(0, 1)$. The time series experiences an abrupt change of the AR(1) coefficient at the break point $t = 0.8$. The covariance of the system changes smoothly before and after that.

Note that the mean of the above four non-stationary models always stays at 0. Simulated Type I error rates of the five methods for the above four non-stationary models are summarized in Table 3 below. It is clear that the classic long-run variance normalization method and self-normalizing method are biased when the second order structure of the series is time varying, which is consistent with our theoretical findings in Section 3.1. On the other hand, we observe that the robust bootstrap with window size selected by the MV method performs reasonably accurate for all four models. In particular, the simulation results further support the use of MV window size selector when the covariance structure of the process is complex. Additionally, the robust bootstrap with AR(1) window size selector is accurate for model I where the correlation structure is identical to that of an stationary AR(1) process. However, the AR(1) window size selector performs unsatisfactorily for the other three models as the covariance structures become complex.

| | $\alpha = 0.05$ | | | | | $\alpha = 0.1$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | LW1 | LW2 | SN | RB1 | RB2 | LW1 | LW2 | SN | RB1 | RB2 |
| | $n = 200$ | | | | | | | | | |
| I | 18.4 | 16.9 | 22.9 | 3.9 | 3.5 | 29.4 | 29.0 | 30.8 | 10.5 | 10.3 |
| II | 17.6 | 9.2 | 12.0 | 10.7 | 3.7 | 25.1 | 19.1 | 19.3 | 19.8 | 10.4 |
| III | 38.5 | 12.9 | 13.1 | 33.1 | 3.7 | 51.0 | 26.7 | 21.4 | 46.1 | 12.1 |
| IV | 22.0 | 16.0 | 21.6 | 14.7 | 5.9 | 30.9 | 28.7 | 30.1 | 24.9 | 14.3 |
| | $n = 500$ | | | | | | | | | |
| I | 20.8 | 21.5 | 23.0 | 6.4 | 6.5 | 29.4 | 29.9 | 30.7 | 12.1 | 12.0 |
| II | 14.6 | 9.3 | 13.0 | 9.7 | 4.2 | 23.4 | 18.4 | 20.1 | 17.8 | 10.5 |
| III | 31.4 | 13.2 | 11.9 | 21.9 | 5.9 | 43.5 | 24.0 | 19.2 | 37.9 | 12.4 |
| IV | 27.4 | 19.1 | 22.4 | 16.3 | 5.8 | 37.2 | 29.3 | 30.5 | 26.0 | 13.5 |

Table 3. Simulated type I error rates (in %) for the five methods with nominal levels 5% and 10% under non-stationary models I - IV. Series lengths $n = 200$ and 500.

## 6.2 Power of the robust bootstrap

In this section we shall study the finite sample sensitivity of the robust bootstrap. To this end we let the observed time series $X_i$ follow (7) and consider testing $H_0$ under the following two types of $\{e_i\}$'s. Case (A):

$$e_i = 0.5e_{i-1} + \varepsilon_i \text{ where } \varepsilon_i \sim \text{ i.i.d. } N(0,1).$$

Case (B) : $\{e_i\}$ follow the following PLS AR(1) model $e_i = G_0(t_i, \mathcal{F}_i)$ for $t_i \le 1/2$ and $e_i = G_1(t_i, \mathcal{F}_i)$ for $t_i > 1/2$, where

$$G_0(t, \mathcal{F}_i) = 0.75\cos(2\pi t)G_0(t, \mathcal{F}_{i-1}) + \varepsilon_i, \qquad G_1(t, \mathcal{F}_i) = (0.5 - t)G_1(t, \mathcal{F}_{i-1}) + \varepsilon_i \quad (28)$$

with i.i.d. standard normal $\varepsilon_i$'s. Cases (A) and (B) correspond to stationary and non-stationary second order structure, respectively. Consider the following piecewise linear fixed alternative:

$$\mu_i = \delta[t_i I\{0 \le t_i \le 0.5\} + (t_i - 1)I\{0.5 < t_i \le 1\}],$$

where $\delta \geq 0$ is a constant. Note that $H_0$ holds when $\delta = 0$ and the magnitude of structural change increases as $\delta$ increases. At nominal level $\alpha = 0.1$, the simulated rejection rates of the various methods listed in Section 6.1 with various choices of $\delta$ are displayed in Figure 1 below. Sample size $n = 200$. For $\delta = 0, 1, 2, 3$, the MV method chooses average window size $m = 8, 8, 8, 10$, respectively in Case (A) and $m = 7, 7, 7, 8$ respectively in Case (B). Since in Case (B) the AR(1) window size selector totally misspecified the dependence structure and selects unreasonable window sizes, rejection rates for methods LW1 and RB1 are not shown for Case (B).

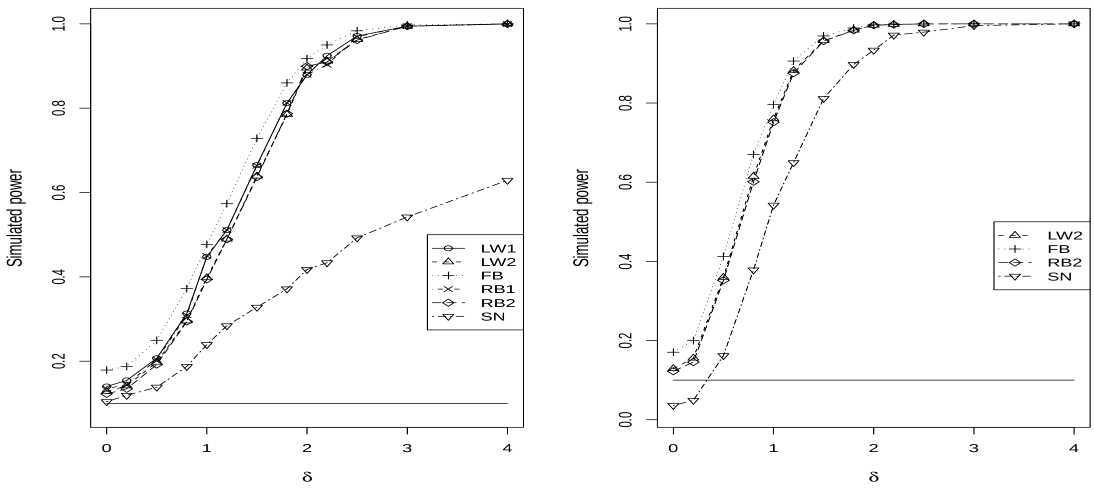


Figure 1: Left panel: simulated rejection rates for Case (A) in Section 6.2. Right panel: simulated rejection rates for Case (B) in Section 6.2. The horizontal lines represent the nominal level $y = 0.1$.

Figure 1 shows that, when window sizes are selected by the same method, the conventional long-run variance normalizing method and our robust bootstrap have very similar finite sample power performances. The above observation is consistent with the theoretical results in Proposition 3 that the robust bootstrap can detect local alternatives with the same $\sqrt{n}$ rate as the conventional tests. Both our theoretical findings and simulation results suggest that the robust bootstrap performs similarly in size and power to the conventional tests for second order stationary time series. On the other hand, for time series with complicated second order structure, our robust bootstrap achieves consistency with reasonably accurate finite sample Type I error rates (for the models considered in the simulation) while conventional tests fails to be valid in general .

We observe further from Figure 1 that in Case (A) where the second order structure is indeed AR(1), the powers for the AR(1) window size selector are moderately better than those of the MV selector. The self normalizing method seems to have relatively lower power, which is consistent with the findings in the literature.

# 7 Empirical Illustrations

In this section, we apply the robust bootstrap to two real time series datasets and compare the results with those of the conventional methods. The first series is composed of yearly rainfall in Tucumán Province, Argentina in millimeters from 1884 to 1996. See upper panel of Figure 2 for the plot of the data. The data provider, Eng Cesar Lamelas, a meteorologist in the Agricultural Experimental Station Obispo Colombres, Tucumán, believes that there is a change in the mean rainfall due to the construction of a dam near the region during 1952-1962. The series has been analyzed several times in the literature with various change point detection methods. See for instance Wu, Woodroofe and Mentz (2001) for an isotonic regression approach and Shao and Zhang (2010) for a self-normalizing approach. Recently Jandhyala et al. (2010) did a detailed change point analysis of the data using various test statistics and variance estimation methods. See also the latter paper for a more detailed data description. All previous analyses assume stationary second order structure. Nevertheless, lower panel of Figure 2 displays the estimated marginal standard deviation of the series over time. It is clear from the plot that the standard deviation experiences periodic changes and it seems that the peek of the second cycle is higher than the first. The non-stationary second order behavior may influence the accuracy of the conventional methods. Here we applied the robust bootstrap to the data. The MV method selects window size 4 and the corresponding $p$-value of the test with 10000 bootstrap samples is about 2%, which provides a strong evidence against $H_0$. On the other hand, both the conventional CUSUM test with lag window long-run variance normalization and the isotonic regression approach give $< 0.1\%$ $p$-values, which provide extremely strong evidences. And the self-normalizing method of Shao and Zhang yields a 10% $p$-value. We argue that the conventional long-run variance normalizing and isotonic regression methods are contaminated by the structural change signals in the variance of the series, which leads to over-optimistic testing results against the null hypothesis. On the other hand, the relatively large $p$-value of the self-normalizing method seems to be due to its lack of sensitivity. Based on our theoretical findings and simulation results, for this example it seems reasonable to believe that the robust bootstrap provides more reliable evidence than the conventional methods.

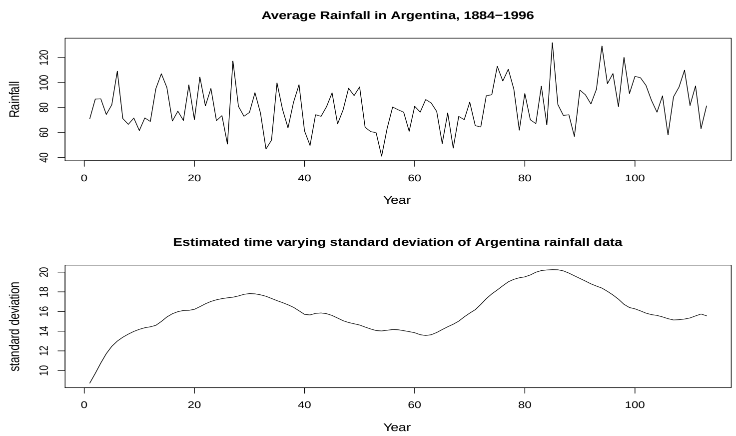


Figure 2: Upper panel: Time series plot of the Argentina rainfall data. Lower panel: Fitted marginal standard deviation of Argentina rainfall data.

The second example we consider is the U.S. 1-year Treasury constant maturity rate from January 4, 1962 to September 10, 1999. Here we have weekly change series with a total of 1966 observations. Figure 3 displays the data. The interest rate series was used in Tsay (2005) to demonstrate time series regression and stationary linear time series models. Here we want to check whether the traditional linear time series model is sufficient for the data. To this end, we will test the constancy of the first and second order structure of the series via the robust bootstrap and compare the results with those of the conventional change point methods. Note that weights of the best linear forecast of a time series (in terms of minimizing the squared risk) are decided by the mean and the covariance function of the sequence (Brockwell and Davis 2009). Therefore if one wishes to use traditional linear time series models to forecast future interest rates, it is crucial to check stability of the mean and covariance function over time. For structural change in mean, the MV method selects window size 8 and the corresponding $p$-value with 10000 bootstrap samples is 22%. Hence there is no evidence against the null hypothesis of no structural change in mean. On the hand, the conventional long-run variance normalization method yield a 12% $p$-value. The large reduction of $p$-value in the conventional long-run variance normalization method is mainly due to an abrupt increase in the variability of the time series in the 1980's (near observation 1000). The change signal in the variance is mistakenly reflected in the CUSUM test for mean when the conventional long-run variance normalization method is used. On the other hand, our bootstrap method is robust to the latter variability change and provides reasonable testing results.

To test changes in the second order structure of the data, we test constancy of the marginal variance and the first order auto covariance of the series, which appear to be the dominating terms in the covariance structure. Denote by $X_1, X_2, \cdots, X_{1966}$ the observed weekly change series. Since the mean remains constant, the latter two tests are equivalent to testing structural change in mean for $Y_i = X_i^2$ and $Z_i = X_i X_{i+1}$. Hence methods in Section 5.1 can be applied. The robust bootstrap for $\{Y_i\}$ selects bandwidth 6 and the resulting $p$-value with 10000 bootstrap samples is $< 0.1\%$, which provides a very strong evidence against the null hypothesis of constant marginal variance. At the same time, the long-run variance normalization method for $\{Y_i\}$ also gives a $< 0.1\%$ $p$-value. Consequently the traditional stationary linear time series models do not seem to be adequate for this data. Furthermore, when testing constancy of $\mathbb{E}Z_i$, the $p$-values for the robust bootstrap and the long-run variance normalization methods are 18% and 13% respectively. Based on our previous discussions, the 18% $p$-value of the robust bootstrap is more reliable and we conclude that there appears to be no significant change in the first order auto covariance and the major source of non-stationarity in the data comes from shifts of the marginal variances. Finally, note that the automatic segmentation technique in Davis et al. (2008) can be applied to detect changes in mean as well as second order structure in this series. The latter technique is good at pick up sudden variance changes such as those near observation 1000 of the series. On the other hand, however, generally it is not expected to be sensitive or accurate for smooth variance changes and our robust bootstrap can be applied in both situations.

**Acknowledgements.** We are grateful to the editor, the associate editor and two referees for their many helpful comments. The research was supported in part by NSERC of Canada.

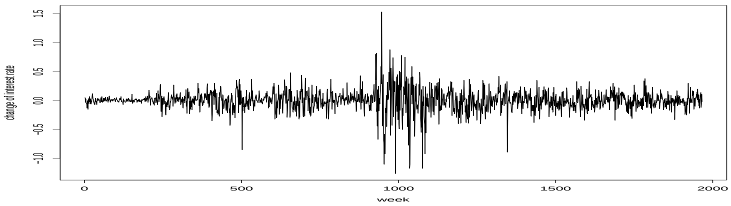


Figure 3: Change of weekly U.S. 1-year Treasury constant maturity rate from January 4, 1962 to September 10, 1999.

# 8 Proofs

Theorems 1 and 2 follow directly from the following important proposition.

**Proposition 5.** *Assume condition (A). Then on a possibly richer probability space, there exist i.i.d. standard normal random variables* $V_1, \cdots, V_n$*, such that*

$$\max_{1\le i\le n} |\sum_{j=1}^{i} e_j - \sum_{j=1}^{i} \sigma(t_j)V_j| = o_{\mathbb{P}}(n^{1/4}\log^2 n).$$

*Proof.* Note that the number of break points $r$ is bounded and is not increasing with $n$. The time series $\{e_i\}$ is composed of $r+1$ zero-mean locally stationary time series in the sense of Zhou and Wu (2009). The proof of Proposition 5 is a straightforward extension of that of Corollary 1 of Wu and Zhou (2011). ♢

The following Lemmas 1 to 4 are needed for the proof of Theorem 3.

**Lemma 1.** *For each* $i$*, let* $\zeta_i$ *be the integer which satisfies* $t_i \in (b_{\zeta_i}, b_{\zeta_i+1}]$*. For* $1 \le r \le s \le n-m+1$*, define*

$$\Lambda_{r,s} = \frac{1}{m(n-m+1)} \sum_{j=r}^{s} (S^o_{j,m})^2,$$

*where* $S^o_{i,m} = \sum_{j=i}^{i+m-1} G_{\zeta_j}(t_j, \mathcal{F}_j)$*. Then under the conditions of Theorem 3, we have*

$$\left\| \max_{1\le r\le s\le n-m+1} |\Lambda_{r,s} - \mathbb{E}\Lambda_{r,s}| \right\| = O(\sqrt{m/n}).$$

*Proof.* Define $\mathcal{F}_i = (\cdots, \varepsilon_{i-1}, \varepsilon_i)$ and $\mathcal{F}_{i,j} = (\cdots, \varepsilon_{j-1}, \varepsilon'_j, \varepsilon_{j+1}, \cdots, \varepsilon_i)$. Note that $S^o_{i,m}$ is $\mathcal{F}_{i+m-1}$ measurable and can be written as $f_i(\mathcal{F}_{i+m-1})$. Define $S^o_{i,l,m} = f_i(\mathcal{F}_{i+m-1,l})$. By Lemma 6 (i) and condition (A3), we have $\max_i \|S^o_{i,m}\|_4 = \max_i \|S^o_{i,l,m}\|_4 = O(\sqrt{m})$. Hence

$$\|(S^o_{i,m})^2 - (S^o_{i,l,m})^2\| \leq (\|S^o_{i,m}\|_4 + \|S^o_{i,l,m}\|_4)\|S^o_{i,m} - S^o_{i,l,m}\|_4 = O(\sqrt{m})(\sum_{j=l-m+1}^{l} \delta_4(j)).$$

Consider the time series $\{f_i^2(\mathcal{F}_{i+m-1})\}_{i=1}^{n-m+1}$. By Lemma 6 (ii), we have

$$\Big\| \max_{1\leq s\leq n-m+1} |\Lambda_{m+1,s} - \mathbb{E}\Lambda_{m+1,s}| \Big\| \leq \frac{\sqrt{n}O(\sqrt{m})}{m(n-m+1)} \sum_{l=0}^{\infty} \sum_{j=l-m+1}^{l} \delta_4(j) = O(\sqrt{m/n}).$$

Therefore

$$\Big\| \max_{1\leq r\leq s\leq n-m+1} |\Lambda_{r,s} - \mathbb{E}\Lambda_{r,s}| \Big\| \leq 2 \Big\| \max_{1\leq s\leq n-m+1} |\Lambda_{m+1,s} - \mathbb{E}\Lambda_{m+1,s}| \Big\| = O(\sqrt{m/n}).$$

$\diamondsuit$

Define the set

$$\mathbb{B} := \{t_i : [t_i, t_i + (m-1)/n] \text{ contains a break point}\}.$$

Note that $\mathbb{B}$ contains $O(m)$ elements. Let $\bar{\mathbb{B}}$ be the complement of $\mathbb{B}$ in $\{t_1, t_2, \cdots, t_n\}$.

**Lemma 2.** *Define* $S^\diamond_{i,m} = \sum_{j=i}^{i+m-1} G_{\zeta_i}(t_i, \mathcal{F}_j)$. *Under the conditions of Theorem 3, we have*

$$\max_{i\in\bar{\mathbb{B}}} |\mathbb{E}(S^o_{i,m})^2 - \mathbb{E}(S^\diamond_{i,m})^2| = O(\sqrt{m^3/n}).$$

*Proof.* Consider the time series $\{G_{\zeta_i}(t_j, \mathcal{F}_j) - G_{\zeta_i}(t_i, \mathcal{F}_j)\}_{j=i}^{i+m-1}$. By condition (A1), we have

$$\max_{1\leq i\leq n-m+1} \max_{i\leq j\leq i+m-1} \|\mathcal{P}_{j-k}[G_{\zeta_i}(t_j, \mathcal{F}_j) - G_{\zeta_i}(t_i, \mathcal{F}_j)]\| = O(\min\{m/n, \delta_2(k)\}),$$

where $\mathcal{P}_j(\cdot) = \mathbb{E}[\cdot|\mathcal{F}_j] - \mathbb{E}[\cdot|\mathcal{F}_{j-1}]$. Hence we obtain by Lemma 6 (i) that

$$\max_{1\leq i\leq n-m+1} \|S^o_{i,m} - S^\diamond_{i,m}\| = O(\sqrt{m}) \sum_{k=0}^{\infty} \min\{m/n, \delta_2(k)\} = O(\sqrt{m^2/n})$$

since $\delta_2(k) = O(\chi^k)$. Therefore uniformly in $i$,

$$|\mathbb{E}(S^o_{i,m})^2 - \mathbb{E}(S^\diamond_{i,m})^2| \leq (\|S^o_{i,m}\| + \|S^\diamond_{i,m}\|)\|S^o_{i,m} - S^\diamond_{i,m}\| = O(\sqrt{m^3/n}).$$

$\diamondsuit$

**Lemma 3.** *Under the conditions of Theorem 3, we have*

$$\max_{i\in\mathbb{B}} |\mathbb{E}(S^o_{i,m})^2 - \mathbb{E}(S^\diamond_{i,m})^2| = O(m).$$

*Proof.* By Lemma 6 (i),

$$\max_{i\in\mathbb{B}} \|S^o_{i,m}\| = O(\sqrt{m}) \quad \max_{i\in\mathbb{B}} \|S^\diamond_{i,m}\| = O(\sqrt{m}).$$

The Lemma follows. $\diamondsuit$

**Lemma 4.** *Under condition (A), we have*

$$\max_{m+1\le i\le n-m+1} |\mathbb{E}(S^\diamond_{i,m})^2 - m\sigma^2(t_i)| = O(1).$$

*Proof.* Note that $G_{\zeta_i}(t_i, \mathcal{F}_j)_{j=-\infty}^{\infty}$ is a stationary time series. Let $\Gamma_i(k)$ be the $k$th auto covariance of the latter series. By Lemma 5 of Zhou and Wu (2010), we have

$$|\Gamma_i(k)| = O(\chi^{|k|}) \text{ uniformly in } i$$

since $\delta_2(k) = O(\chi^k)$. Note that $\sigma^2(t_i) = \sum_{k=-\infty}^{\infty} \Gamma_i(k)$. Hence

$$|\mathbb{E}(S^\diamond_{i,m})^2 - m\sigma^2(t_i)| \le 2\sum_{j=0}^{m-1} j|\Gamma_i(j)| + 2m\sum_{j\ge m} |\Gamma_i(j)| = O(1)$$

uniformly in $i$. $\diamondsuit$

**Lemma 5.** *Define process* $\tilde{\Phi}^o_{m,n}(t) = \Phi^o_{t_*n,m} + n(t-t_*)(\Phi^o_{t^*n,m} - \Phi^o_{t_*n,m})$, $t \in [1/n, (n-m+1)/n]$, *where*

$$\Phi^o_{i,m} = \sum_{j=1}^{i} \frac{1}{\sqrt{m(n-m+1)}}(S_{j,m} - \mathbb{E}S_{j,m})R_j, \qquad i = 1, \cdots, n-m+1,$$

*Then under the null hypothesis, conditional on* $(X_i)_{i=1}^n$, *we have*

$$\tilde{\Phi}^o_{m,n}(t) \Rightarrow U(t) \text{ on } \mathcal{C}(0,1) \text{ with the uniform topology.}$$

*Proof.* Note that $\tilde{\Phi}^o_{m,n}(t)$ is the linear interpolation of $(\Phi^o_{i,m})$. By Lemmas 1 to 4, we have

$$\left\| \max_{1\le r\le s\le n-m+1} |\Lambda_{r,s} - \int_{r/n}^{s/n} \sigma^2(t)\,dt| \right\| = o(1). \tag{29}$$

The finite dimensional convergence follows easily from (29) and the Cramer-Wold device. The tightness of $\tilde{\Phi}^o_{m,n}(t)$ also follows easily from (29), the piecewise continuity of $\sigma^2(\cdot)$ and Theorem 7.3 of Billingsley (1999). $\diamondsuit$

*Proof of Theorem 3.* Observe that, under the null hypothesis, $\Phi_{i,m} = \Phi^o_{i,m} - II_{i,m}$, where

$$II_{i,m} = \sum_{j=1}^{i} \frac{1}{\sqrt{m(n-m+1)}} (\frac{m}{n} S_n - \mathbb{E}[\frac{m}{n} S_n]) R_j.$$

It is straightforward to show that, conditional on $(X_i)$,

$$\max_{1 \le i \le n-m+1} |II_{i,m}| = O_{\mathbb{P}}(\sqrt{m/n}) = o_{\mathbb{P}}(1). \tag{30}$$

Hence Theorem 3 follows from (30), Lemma 5 and the continuous mapping theorem. ♢

*Proof of Proposition 3.* Under the alternative hypothesis, write $\Phi_{i,m} = \Phi^o_{i,m} - II_{i,m} + III_{i,m}$, where

$$III_{i,m} = \sum_{j=1}^{i} \frac{1}{\sqrt{m(n-m+1)}} (\mathbb{E}[S_{j,m}] - \mathbb{E}[\frac{m}{n} S_n]) R_j.$$

On $[1/n, (n-m+1)/n]$, define process

$$III_{m,n}(t) = III_{t_* n,m} + n(t - t_*)(III_{t^* n,m} - III_{t_* n,m}),$$

where $t_* = \lfloor tn \rfloor / n$ and $t^* = t_* + 1/n$. Now note that both $\mathbb{E}S_{j,m}$ and $\mathbb{E}S_n$ are non-random in $III_{i,m}$. By the piecewise smoothness of $\mu(\cdot)$, it is easy to show that

$$III_{m,n}(t)/\sqrt{m} \Rightarrow U^*(t) \text{ on } \mathcal{C}(0,1) \tag{31}$$

under $H_a^{PS}$ and the condition that $m \to \infty$ with $m/n \to 0$. Similarly,

$$\sup_{0<t<1} |III_{m,n}(t)| = O_{\mathbb{P}}(\sqrt{m/n}) = o_{\mathbb{P}}(1) \tag{32}$$

under the local alternative. Proposition 3 follows from the proof of Theorem 3, (31) and (32). ♢

*Proof of Theorem 5.* Theorem 5 is a multivariate extension of Theorems 1 and 3. The proof of Theorem 5 can be easily carried out from those of Theorems 1 and 3 without any essential difficulties. Details are omitted. ♢

*Proof of Proposition 4.* For presentational simplicity, we will only consider in case (a) $Y_i = \phi(X_i)$, where $X_i$ and $Y_i$ are univariate; and in case (b) $Y_i = X_i^2$. General cases follow by similar arguments. In case (a), we have

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

Further note that $\mathbb{E}|\phi(G_i(t,\mathcal{F}_0)) - \phi(G_i(s,\mathcal{F}_0))| \le \|\phi(G_i(t,\mathcal{F}_0)) - \phi(G_i(s,\mathcal{F}_0))\|$. Hence we have (A1). Condition (A3) follows by similar arguments. Now we consider case (b). By the Cauchy-Schwarz inequality,

$$\|G_i^2(t,\mathcal{F}_0)) - G_i^2(s,\mathcal{F}_0)\| \le \|G_i(t,\mathcal{F}_0) - G_i(s,\mathcal{F}_0)\|_4 \|G_i(t,\mathcal{F}_0) + G_i(s,\mathcal{F}_0)\|_4 \le C|t-s|.$$

Again, $\mathbb{E}|G_i^2(t,\mathcal{F}_0)) - G_i^2(s,\mathcal{F}_0)| \le \|G_i^2(t,\mathcal{F}_0)) - G_i^2(s,\mathcal{F}_0)\|$. Therefore (A1) follows in this case. Similarly we have (A3). ♢

**Lemma 6.** *Suppose $X_i = H_i(\mathcal{F}_i)$ satisfies $\mathbb{E}[X_i] = 0$ and $\max_i \mathbb{E}|X_i|^q < \infty$ for some $q > 1$. For $k \ge 0$, define $\delta_X(k) = \max_{1\le i\le n} \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,i-k})\|_q$. Let $\delta_X(k) = 0$ if $k < 0$. Write $\gamma_k = \sum_{i=0}^k \delta_X(i)$. Let $S_i = \sum_{j=1}^i X_j$. Then (i)*

$$\|S_n\|_q^{q'} \le C_q \sum_{i=-n}^{\infty} (\gamma_{i+n} - \gamma_i)^{q'},$$

*where $q' = \min(2,q)$. If additionally $\Delta := \sum_{j=0}^{\infty} \delta_X(j) < \infty$. Then (ii)*

$$\| \max_{1\le i\le n} |S_i| \|_q \le C_q n^{1/q'} \Delta.$$

*In (i) and (ii), $C_q$ are generic finite constants which only depend on $q$ and can vary from place to place.*

Lemma 6 is essentially an extension of Theorem 1 (i) and (iii) of Wu (2007) into non-stationary time series. A careful check of the proof of the latter Theorem shows that the arguments are also valid for non-stationary processes in the current setting.

## REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858.

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

Aue, A., Horváth, L., Hu**v**sková, M., and Kokoszka, P. (2008). Testing for changes in polynomial regression. *Bernoulli* **14** 637–660.

Barnett, T. P., Pierce, D. W. and Schnur, R. (2001). Detection of anthropogenic climate change in the world's oceans. *Science* **292** no. 5515 270–274.

Berkes, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference* **139** 2044–2063.

Billingsley, P. (1999). *Convergence of Probability Measures. Second edition.* Wiley.

Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods, second edition.* Springer, New York.

Brodsky, B. E., and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications*, Vol. 243, Dordrecht: Kluwer.

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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327.

Collins, D. A., P. M. Della-Marta, N. Plummer and B. C. Trewin (2000). Trends in annual frequencies of extreme temperature events in Australia. *Australian Meteorological Magazine* **49** 277–292.

*Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change.* (2007) Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Csörgő, M., and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* New York: Wiley.

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

Davis, R. A., Huang, D., and Yao, Y.-C. (1995), Testing for a Change in the Parameter Values and Order of an Autoregressive Model, *The Annals of Statistics* **23** 282–304.

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

Davis, R. A., Lee, T. C. M and Rodriguez-Yam, G. A. (2008). Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis* **29** 834–867.

Elsner, J. B., Kossin, J. P. and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455** 9295 doi:10.1038/nature07234.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50** 987–1007.

Inclán, C. and Tiao, G. C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Change of Variance. *Journal of the American Statistical Association* **89** 913–923.

Jandhyala, V. K., Fotopoulos, S. B. and You, J. (2010). Change-point analysis of mean annual rainfall data from Tucumán, Argentina. *Environmetrics* **21** 687–697.

Karl, T.R., Knight, R.W. and Plummer, N. (1995). Trends in high-frequency climate variability in the twentieth century. *Nature* **377** 217–220

Keogh, E., Chu, S., Hart, D. and Pazzani, M. (2001). An online algorithm for segmenting time series. *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on* 289–296.

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

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer, New York.

Lee, S., and Park, S. (2001), The Cusum of Squares Test for Scale Changes in Infinite Order Moving Average Processes. *Scandinavian Journal of Statistics* **28** 625–644.

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.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708.

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

Page, E. S. (1954). Continuous inspection schemes. *Biometrika.* **41** 100–114.

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

Priestley, M. B. (1988). *Nonlinear and nonstationary time series analysis.* Academic Press, London.

Räisänen, J. (2002) CO2-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *Journal of Climate* **15** 2395–2411.

Rao, C. R. (1948). Large Sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philos. Soc.* **44** 50–57.

Silvey, S. D. (1959). The Lagrangian multiplier test. *Annals of Mathematical Statistics* **30** 389–407.

Sayginsoy, Ö and Vogelsang, T. J. (2011). Testing for a shift in trend at an unknown date: a fixed-b analysis of heteroskedasticity autocorrelation robust OLS-based tests. *Econometric Theory* **27** 992–1025.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association.* **105** 1228–1240.

Siegmund, D. O. and Yakir, B. (2008). Minimax optimality of the Shiryayev-Roberts change-point detection rule. *Journal of Statistical Planning and Inference* **138** 2815–2825.

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

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

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics* **14** 1261–1350.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Annals of Probability*, **35**, 2294-2320.

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

Wu, W. B., Woodroofe, M., and Mentz, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika* **88** 793–804.

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