Statistica Sinica Preprint No: SS-2023-0238	
Title	Testing for Variance Changes under Varying Mean and Serial Correlation
Manuscript ID	SS-2023-0238
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0238
Complete List of Authors	Cheuk Wai Leung and Kin Wai Chan
Corresponding Authors	Kin Wai Chan
E-mails	kinwaichan@cuhk.edu.hk

TESTING FOR VARIANCE CHANGES UNDER VARYING MEAN AND SERIAL CORRELATION

Cheuk Wai Dominic Leung and Kin Wai Chan

Department of Statistics, The Chinese University of Hong Kong

Abstract: Detection of variance change points is statistically difficult when the data exhibit a varying mean structure and autocorrelation. Existing variance change point tests either require the assumption of mean constancy or sacrifice testing power due to serial dependence. This article addresses these problems by proposing a trend-robust and autocorrelation-efficient variance change point test via a differencing approach. This approach removes the mean effect without fitting the mean function. It also allows the test to retrieve the reduced power due to serial dependence. We prove that the optimal difference-based test should minimize the long-run coefficient of variation of the sample second moment of the noises instead of the long-run variance in the presence of serial dependence. The optimal solution can be efficiently computed by fractional quadratic programming. The asymptotic relative efficiency under a local alternative hypothesis is derived. A rate-optimal long-run variance estimator is also proposed. It is proven to be doubly robust against varying mean and variance change points.

Key words and phrases: change point, cumulative sum, difference sequence, long-run variance, non-linear time series.

1. Introduction

Abrupt variance changes provide insights that cannot be explained by the variability of the means. Applications cover various fields, e.g., finance (Inclan and Tiao, 1994), environmental science (Gerstenberger *et al.*, 2020), medical science (Gao *et al.*, 2019), etc; see also Hsu *et al.* (1974), Lee and Park (2001), Lee *et al.* (2003), and Aue *et al.* (2009). In this article, we assume that the observed time series $X_1, \ldots, X_N$ are generated as follows:

$$X_i = \mu_i + \sigma_i Z_i \qquad (i = 1, \ldots, N), \tag{1.1}$$

where $\mu_1, \ldots, \mu_N$ are possibly non-constant deterministic signals, $\sigma_1, \ldots, \sigma_N$ are deterministic marginal standard deviations, and $(Z_i)_{i\in\mathbb{Z}}$ is a zero-mean unit-variance strictly stationary noise time series. Our goal is to test $H_0 : \sigma_1 = \cdots = \sigma_N$ against

$$H_1 : \sigma_0 \equiv \sigma_1 = \cdots = \sigma_{k^\star} \neq \sigma_{k^\star+1} = \cdots = \sigma_N \equiv \sigma_0 e^{\Delta} \tag{1.2}$$

for some $1 < k^\star < N$ and $\Delta \neq 0$, where $k^\star$ is the change point and Δ parametrizes the change. This problem is non-trivial because of (i) non-constant means and (ii) non-independent noises.

For (i), mean-constancy is a crucial assumption for the traditional cumulative sum variance change point tests. Many researchers overcome it by centering the data in advance. The first approach is differencing, which has been widely used for variance estimation in nonparametric regression; see, e.g., Chen and Gupta (1997),

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

Davis *et al.* (2006), Wu and Zhao (2007), and Chapman *et al.* (2020). However, the effect on the power was not discussed. Indeed, differencing is also recently used for specification test and trend test in time series; see Bai and Wu (2024) and To and Chan (2023). The second approach is smoothing. Typically, consistent estimators $(\hat{\mu}_i)_{i=1}^N$ of the means $(\mu_i)_{i=1}^N$ are used to construct the residuals $(X_i - \hat{\mu}_i)_{i=1}^N$. For example, Lee *et al.* (2003) used Nadawaya–Watson kernel estimator, whereas Gao *et al.* (2019) used a weighted spline regression estimator. The former work was proved to be valid for time series, while the latter one was proved for independent Gaussian data. Unfortunately, local de-trending may not lead to the highest power as we shall show in Section 6.

For (ii), Lee *et al.* (2003) showed that the cumulative sum test statistic is identical to that in the case with independent data, except that the normalizer is switched from the marginal variance to the long-run variance. The resulting test statistic has the same limiting distribution under certain weak dependence conditions, e.g., mixing (Phillips, 1987) and stability (Wu, 2011). However, Lee *et al.* (2003) indicated that the power declines as the serial dependence gets stronger. So, serial dependence was regarded as a harmful structure that hurts the power. To the best of our knowledge, the power loss issue remains unsolved no matter whether $(\mu_i)_{i=1}^N$ are constant or not.

In this article, we resolve the problems by a unified approach. Let d =

$(d_0, d_1, \ldots, d_m)^\intercal \in \mathbb{R}^{m+1}$ for $m \in \mathbb{N}$. We call d a difference sequence if $\sum_{j=0}^m d_j = 0$ and $\sum_{j=0}^m d_j^2 = 1$. If only $\sum_{j=0}^m d_j^2 = 1$ is satisfied, we call d a weight sequence. When $m = 0$, we set $d = (1)$. We also write d as $d^{(m)} = (d_0^{(m)}, \ldots, d_m^{(m)})^\intercal$ to emphasize the order m. Using d, we transform X_i to

$$D_i = \sum_{j=0}^{m} d_j X_{i+m-j} \qquad (i = 1, \ldots, n), \tag{1.3}$$

where $n = N - m$. We also denote $D_i = \phi(L; d)X_{i+m}$, where L is the lag operator, and $\phi(L; d) = d_0 + d_1 L + \cdots + d_m L^m$ is a difference operator. Since m is assumed finite, $N \to \infty$ and $n \to \infty$ are equivalent. All asymptotic results are derived as $n \to \infty$.

The principle is that $(D_i)_{i=1}^n$ is a good building block for inferring the variance because the mean is approximately removed. Our goal is to derive the optimal difference sequence $d^\star$ so that the resulting $(D_i)_{i=1}^n$ utilizes the dependent data as antithetic variates to improve the asymptotic efficiency of the test statistic, which consequently boosts up the power.

Our major contributions are listed as follows. First, when the data are serially dependent, we proved that one should minimize the long-run coefficient of variation instead of long-run variance for constructing the optimal test. It enables us to solve the optimization problem by fractional quadratic programming. Consequently, the optimal difference sequence can be numerically found in a computationally feasible manner. Second, we derived a close-form expression of the asymptotic relative

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

efficiency. It supports the optimality claim. Third, we proposed a doubly robust estimation scheme for the long-run variance under (i) varying mean and (ii) variance change points in serially dependent data. It consistently standardizes our proposed test statistics and leads to a size-accurate tests.

The theoretic parts of this article were motivated by the differencing ideas presented in Hall *et al.* (1990) and Chan (2022a). These ideas were originally developed for removing time-varying structures in the first-order moment (i.e., the mean). We extend their concepts to handle time-varying structures in both the first-order moment and the second-order moment (i.e., the variance). In our approach, the mean function is assumed to be Lipschitz continuous, while the variance function may contain change points. In particular, we demonstrate that the optimal difference sequence proposed in Hall *et al.* (1990) is no longer optimal for addressing variance change points. We establish that the optimal difference sequence depends on the serial dependence structure in our problem. Additionally, we generalize the difference-based long-run variance estimator introduced in Chan (2022a) to achieve double robustness against changes in both the mean and the variance.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

2. Background

2.1 Mathematical setup and assumptions

Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Define $\sum_{i=i_0}^{i_1} y_i = 0$ if $i_0 > i_1$. For a random variable V and $p \geqslant 1$, denote $\|V\|_p = (E|V|^p)^{1/p}$. The autocovariance of a stationary time series $(V_i)_{i\in\mathbb{Z}}$ is $\gamma_k^V = \text{cov}(V_0, V_k)$, $k \in \mathbb{Z}$. Given $V_{1:n} = (V_1, \ldots, V_n)^{\top}$, denote $\bar{V}_n = \sum_{i=1}^n V_i/n$ and $\hat{\gamma}_k^V = \sum_{i=|k|+1}^n (V_i - \bar{V}_n)(V_{i-|k|} - \bar{V}_n)/n$ for $|k| < n$. Define $\nabla V_i = (V_{i+1} - V_i)/2^{1/2}$. Denote weak convergence by "$\Rightarrow$" and convergence in probability by "$\xrightarrow{\text{pr}}$". Under H_0, we set $k^\star = \Delta = 0$.

Let $(\epsilon_i, \epsilon_i')_{i\in\mathbb{Z}}$ be independent and identically distributed (i.i.d.) random variables. So, ϵ_i' is said to be an i.i.d. copy of ϵ_i for each $i \in \mathbb{Z}$. Assume $Z_i = g(\mathscr{F}_i)$ and $Z_i' = g(\mathscr{F}_i')$, where g is measurable, $\mathscr{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i)$, and $\mathscr{F}_i' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_i)$. Let $\xi_p(i) = \|Z_i - Z_i'\|_p$ be the physical dependence measure (Wu, 2005) for $p \geqslant 1$.

Assumption 1 (Short-range dependence)**.** *The noise sequence $(Z_i)_{i\in\mathbb{Z}}$ admits a causal representation $Z_i = g(\mathscr{F}_i)$, and satisfies that (a) $E(Z_i^{8+e}) < \infty$ for some $e > 0$; (b) (Z_i) is strictly stationary; and (c) (Z_i) is 8-stable, i.e., $\Xi_8 = \sum_{i=0}^\infty \xi_8(i) < \infty$.*

Assumption 2 (Signal smoothness)**.** *The deterministic signal $(\mu_i)_{i=1}^N$ admits the form $\mu_i = f(i/N)$ for $i = 1, \ldots, N$, where $f : [0, 1] \to \mathbb{R}$ is Lipschitz continuous, i.e., there is a $C \in (0, \infty)$ such that $|f(x_1) - f(x_2)| \leqslant C|x_1 - x_2|$ for all $x_1, x_2 \in [0, 1]$.*

Assumption 3. *Unless otherwise stated, one of the following collections of assumptions is satisfied: (1) Assumption 1 holds, $(\mu_i)_{i=1}^N$ satisfy Assumption 2, and d is a difference sequence. (2) Assumption 1 holds, $\mu_1 = \cdots = \mu_N$, and d is a weight sequence.*

2.2 Difference-based test

Our idea is to approximate the mean-zero time series $\phi(L;d)(X_{i+m} - \mu_{i+m})$ with D_i; see Section A.1 of the supplementary note for a detailed explanation of this approximation. A suitable d allows us to achieve a trend-robust and autocorrelation-efficient test.

Let $Q_i = D_i^2$ for $i = 1, \ldots, n$. Define the forward average and backward averages of Q_i's as $\Lambda_n^F(r) = \sum_{i=1}^{\lfloor nr \rfloor} Q_i/\lfloor nr \rfloor$ and $\Lambda_n^B(r) = \sum_{i=\lfloor nr \rfloor+1}^{n} Q_i/(n - \lfloor nr \rfloor)$, respectively, for $r \in [0,1]$, where $0/0 = 0$ by convention. Then the difference-based cumulative sum process is

$$C_n(r;d) = n^{1/2} r(1-r)\left\{\Lambda_n^F(r) - \Lambda_n^B(r)\right\}, \quad r \in [0,1]. \tag{2.1}$$

We propose an mth order differenced Kolmogorov–Smirnov test statistic:

$$T_n(d) = \frac{1}{\hat{v}_m^{1/2}} \sup_{r\in[0,1]} |C_n(r;d)|, \tag{2.2}$$

where the normalizer $\hat{v}_m$ is a weakly consistent estimator of the long-run variance

$$v_m = \lim_{n\to\infty} n\mathrm{var}\left(\bar{Q}_n\right), \quad \text{where} \quad \bar{Q}_n = \frac{1}{n}\sum_{i=1}^{n} Q_i. \tag{2.3}$$

Consistent estimation of v_m under H_0 and H_1 is crucial to size accuracy and powerfulness of the test, respectively. Estimation of v_m is highly non-trivial. Our proposed estimator of v_m is presented in Section 5. From (2.2), a variance change point estimator is $\hat{k} = \lfloor (m+1)/2 \rfloor + \arg\max_{1 \leqslant k \leqslant n} |C_n(k/n; d)|$. Theorem 1 below states the asymptotic properties of $T_n(d)$.

Theorem 1 (Validity and consistency)**.** *Let* $m \in \mathbb{N}$. *Suppose Assumption 3 is satisfied. Assume that* $\hat{v}_m \xrightarrow{\text{pr}} v_m$. *(1) Under* H_0,

$$T_n(d) \Rightarrow \mathbb{K} \equiv \sup_{r \in [0,1]} |\mathbb{B}(r) - r\mathbb{B}(1)|, \tag{2.4}$$

where $\mathbb{B}$ *is the standard Brownian motion. (2) Under* H_1, *if, in addition,* $k^\star/n \to \tau \in (0,1)$, *then* $(\hat{k} - k^\star)/n \xrightarrow{\text{pr}} 0$ *and*

$$\frac{T_n(d)}{n^{1/2}} \xrightarrow{\text{pr}} (1-\tau)\tau \left(\frac{\sum_{|k| \leqslant m} \sum_{j=|k|}^{m} d_j d_{j-|k|} \gamma_k^Z}{v_m^{1/2}} \right) \sigma_0^2 \left| 1 - e^{2\Delta} \right|. \tag{2.5}$$

We emphasize that Theorem 1 remains valid even if $(\mu_i)_{i=1}^N$ are not constant. In (2.4), $\mathbb{K}$ is the Kolmogorov distribution, whose quantiles are available in standard software packages. So, performing a test based on $T_n(d)$ is straightforward. From (2.5), $T_n(d) \to \infty$ in probability under H_1. Hence, the test has power one asymptotically. Some remarks concerning applicability and implementation are listed below.

Remark 2.1. The proposed test is applicable to test for multiple change points, i.e.,

$$H_1' : \sigma_1 = \cdots = \sigma_{k_1^\star} \neq \sigma_{k_1^\star+1} = \cdots = \sigma_{k_2^\star} \neq \cdots \neq \sigma_{k_M^\star+1} = \cdots = \sigma_N, \tag{2.6}$$

for unknown $M \in \mathbb{N}$ and $1 < k_1^\star < \cdots < k_M^\star < n$. Under H_1, denote the initial variance by σ_0 and set $\sigma_{k_\iota^\star+1} = \sigma_0 e^{\Delta_\iota}$ for $1 \leqslant \iota \leqslant M$ and $\Delta_0 = 0$. Let $k_\iota^\star/n \to \tau_\iota$ and $\theta_\iota = \tau_{\iota+1} - \tau_\iota \neq 0$, where $\tau_0 = 0$, $\tau_{M+1} = 1$ and $\theta_0 = \tau_1$. The analogy of Theorem 1 is deferred to Section A.2 of the supplementary note.

Remark 2.2. We recommend doing a variance stabilizing transformation to improve the finite-sample performance. Define $C_n^L(r;d) = n^{1/2} r(1-r)\left\{\log \Lambda_n^F(r) - \log \Lambda_n^B(r)\right\}$, where $\pm\infty \times 0 = 0$. The stabilized version of $T_n(d)$ is

$$T_n^{(L)}(d) = \frac{1}{\hat{v}_{m,L}^{1/2}} \sup_{r \in [0,1]} \left| C_n^L(r;d) \right|, \tag{2.7}$$

where $\hat{v}_{m,L}$ is an estimator of $v_{m,L} = \lim_{n\to\infty} n\text{var}(\log \bar{Q}_n)$. Our proposal of $\hat{v}_{m,L}$ is given in Section 5. The analogy of Theorem 1 and related theoretical properties are deferred to Section A.3 of the supplementary note.

Remark 2.3. In the literature, there is a class of self-normalized tests that do not require estimation of the long-run variance; see, e.g., Shao (2010), Shao and Zhang (2010), Zhang and Lavitas (2018), and Cheng and Chan (2024). Although self-normalization bypasses estimation of v_m and achieves better size, it comes with the price of lower power, as mentioned in Lobato (2001). Since the goal of the current article is to derive a variance change point test that has high power, we focus on the non-self-normalized approach. Besides, we would like to emphasize that constructing a self-normalizer for a variance change point test in our setting

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

is not straightforward because the squared differenced data $Q_1, \ldots, Q_n$ exhibits change points in both mean and variance under H_1. So, the existing self-normalized approaches that are tailor-made for handling change point in mean or change point in variance with a constant mean may need non-trivial modification. We believe that it is an interesting topic and leave it for future research.

3. Methodology and main results

3.1 Optimal difference sequence

We propose utilizing neighbor observations as antithetic variates to maximally boost up the power of the test. In this section, we detail how the weights $(d_0, \ldots, d_m)$ of the antithetic variates $(X_i, \ldots, X_{i+m})$ are optimally designed to construct $D_i = \phi(L; d)X_{i+m}$.

In view of (2.5), the optimal difference sequence shall be chosen to minimize

$$w_m = \frac{v_m}{\left(\sum_{|k|\leqslant m} \delta_k \gamma_k^Z\right)^2}, \tag{3.1}$$

where $\delta_k = \sum_{j=|k|}^{m} d_j d_{j-|k|}$ for $|k| \leqslant m$. We will further elaborate that optimality holds under H_0 and H_1 in Theorem 2. Another interpretation of maximizing (3.1) is provided by using Asymptotic Relative Efficiency (ARE) in Theorem 4.

Notice that the denominator is reduced to $(\gamma_0^Z)^2$ when Z_i, which means $\mathrm{var}(Z_i) = \mathrm{var}\{\phi(L; d)Z_i\}$ when Z_i is uncorrelated. We highlight this finding because Hall *et al.*

(1990) focuses on minimizing v_m if the underlying noise is i.i.d. We added the denominator and generalize it to the time series optimization.

Directly optimizing w_m with respect to $(d_0, \ldots, d_m)$ is non-standard. Instead, we proceed in two steps as follows: (i) optimizing w_m with respect to $(\delta_0, \ldots, \delta_m)$, and (ii) converting $(\delta_0, \ldots, \delta_m)$ back to $(d_0, \ldots, d_m)$. In the rest of this subsection, we prove that step (i) is a fractional quadratic programming problem, and step (ii) can be solved iteratively by, e.g., the innovation algorithm.

To begin with, we define the following notation. Recall that $\hat{\gamma}_k^Z$ is the sample autocovariance of $Z_{1:n}$ at lag k. For $k, k' \in \{0, \ldots, m\}$, the long-run covariance of $\hat{\gamma}_k^Z$ and $\hat{\gamma}_{k'}^Z$ is

$$\Upsilon_{k,k'}^Z = \lim_{n\to\infty} n\mathrm{cov}(\hat{\gamma}_k^Z, \hat{\gamma}_{k'}^Z) = \sum_{j=-\infty}^{\infty} \zeta_{k,k'}^Z(j), \tag{3.2}$$

where $\zeta_{k,k'}^Z(j) = \zeta_{k',k}^Z(-j) = E(Z_0 Z_{0+k} Z_j Z_{j+k'}) - \gamma_k^Z \gamma_{k'}^Z$ for $j \geqslant 0$. Let $\gamma^Z = (\gamma_k^Z)_k$ and $\Upsilon^Z = (\Upsilon_{k,k'}^Z)_{k,k'}$. The limits in (3.2) exist by Assumption 1.

Note that $D_i \approx \phi(L; d)(\sigma_{i+m} Z_{i+m})$. So, to study the behavior of w_m, we let $Y_i = \sigma_0 \phi(L; d) Z_{i+m}$ be a version of D_i defined by the underlying stationary noises (Z_i) with a constant scale σ_0. Denote $\delta = (\delta_1, \ldots, \delta_m)^{\mathsf{T}}$. We may also write δ as $\delta^{(m)} = (\delta_1^{(m)}, \ldots, \delta_m^{(m)})^{\mathsf{T}}$ to emphasize the order m. Let

$$U_m^{(1)}(\delta; \gamma^Z) = \left(\sum_{k=-m}^{m} \delta_k \gamma_k^Z\right)^2 \qquad \text{and} \qquad U_m^{(2)}(\delta; \Upsilon^Z) = \sum\sum_{k,k'=-m}^{m} \delta_k \delta_{k'} \Upsilon_{k,k'}^Z. \tag{3.3}$$

The following theorem shows that w_m is proportional to $U_m^{(2)}(\delta; \gamma^Z)/U_m^{(1)}(\delta; \Upsilon^Z)$, which depends on the dependence structure of (Z_i) and is a ratio of the two quadratic functions of δ.

Theorem 2 (Representation)**.** *Let $m \in \mathbb{N}$ and $\bar{Q}_{n,Y} = \sum_{i=1}^{n} Y_i^2/n$. Suppose Assumption 3 is satisfied. Then the following results hold. (1)*

$$\lim_{n\to\infty} E\left(\bar{Q}_{n,Y}\right) = \left\{\sigma_0^4 U_m^{(1)}(\delta;\gamma^Z)\right\}^{1/2} \quad \textit{and} \quad \lim_{n\to\infty} n\mathrm{var}\left(\bar{Q}_{n,Y}\right) = \sigma_0^4 U_m^{(2)}(\delta;\Upsilon^Z).$$

(2) If $k^\star/n \to \tau$, then

$$w_m = \sigma_0^4\left\{\tau + (1-\tau)e^{4\Delta}\right\}\varpi_m, \quad \textit{where} \quad \varpi_m = \frac{U_m^{(2)}(\delta;\Upsilon^Z)}{U_m^{(1)}(\delta;\gamma^Z)}. \tag{3.4}$$

There are two implications of Theorem 2. First, minimizing w_m can be statistically interpreted as minimizing ϖ_m, which is the square of the long-run coefficient of variation of $\bar{Q}_{n,Y}$. The intuition is that the limit of $E(\bar{Q}_{n,Y})$ is not free of δ, so one should minimize $\mathrm{var}(\bar{Q}_{n,Y})/\{E(\bar{Q}_{n,Y})\}^2$ instead of $\mathrm{var}(\bar{Q}_{n,Y})$. Second, by the constraint $\delta_0 = 1$ and the symmetry of δ_k, γ_k^Z and $\Upsilon_{k,k'}^Z$, we can express (3.3) in matrix forms: $U_m^{(r)} = a_r + 4c_r^{\mathrm{T}}\delta + 4\delta^{\mathrm{T}}H_r\delta$ for $r = 1, 2$, where $a_1 = 1$, $a_2 = \Upsilon_{0,0}^Z$,

$$c_1 = \begin{bmatrix}\gamma_1^Z\\ \vdots \\ \gamma_m^Z\end{bmatrix}, \quad c_2 = \begin{bmatrix}\Upsilon_{0,1}^Z\\ \vdots \\ \Upsilon_{0,m}^Z\end{bmatrix}, \quad H_1 = c_1c_1^{\mathrm{T}}, \quad H_2 = \begin{bmatrix}\Upsilon_{1,1}^Z & \cdots & \Upsilon_{1,m}^Z\\ \vdots & \ddots & \vdots \\ \Upsilon_{m,1}^Z & \cdots & \Upsilon_{m,m}^Z\end{bmatrix}.$$

Therefore, the minimization of ϖ_m in (3.4) is a fractional quadratic programming problem:

$$\delta^\star \equiv \delta^\star(\mathscr{D}_m;\gamma^Z,\Upsilon^Z) = \operatorname*{arg\,min}_{\delta\in\mathscr{D}_m} \varpi_m = \operatorname*{arg\,min}_{\delta\in\mathscr{D}_m} \frac{a_2 + 4c_2^{\mathrm{T}}\delta + 4\delta^{\mathrm{T}}H_2\delta}{a_1 + 4c_1^{\mathrm{T}}\delta + 4\delta^{\mathrm{T}}H_1\delta}, \tag{3.5}$$

where $\mathscr{D}_m$ is a feasible set of δ. Fractional quadratic programming has been extensively studied; see, e.g., Dinkelbach (1967) and Schaible and Ibaraki (1983). The solution $\delta^\star$ can be solved easily by standard software. Possible forms of $\mathscr{D}_m$ are stated below.

Theorem 3 (Feasible sets)**.** *Denote $\delta = F_m(d)$ whenever it is well defined. Let $m \in \mathbb{N}$ and the sets of all possible difference sequences and weight sequences be*

$$\bar{\mathscr{S}}_m = \left\{ d \in \mathbb{R}^{m+1} : \sum_{j=0}^{m} d_j^2 = 1 \right\} \quad \textit{and} \quad \tilde{\mathscr{S}}_m = \left\{ d \in \bar{\mathscr{S}}_m : \sum_{j=0}^{m} d_j = 0 \right\},$$

respectively. Also let

$$\bar{\mathscr{D}}_m = \left\{ \delta \in \mathbb{R}^m : \min_{\lambda \in [0,\pi]} \sum_{j=1}^{m} \delta_j \cos(j\lambda) \geqslant -\frac{1}{2} \right\} \quad \textit{and} \quad \tilde{\mathscr{D}}_m = \left\{ \delta \in \bar{\mathscr{D}}_m : \sum_{j=1}^{m} \delta_j = -\frac{1}{2} \right\}.$$

Then, the function F_m is surjective if F_m is defined as $F_m : \bar{\mathscr{S}}_m \to \bar{\mathscr{D}}_m$ or $F_m : \tilde{\mathscr{S}}_m \to \tilde{\mathscr{D}}_m$.

According to Theorem 3, we should set $\mathscr{D}_m = \tilde{\mathscr{D}}_m$ under Assumption 3 (1), and set $\mathscr{D}_m = \bar{\mathscr{D}}_m$ under Assumption 3 (2). Note that we use a tilde (resp. a bar) over a variable to denote varying mean assumption (resp. constant mean assumption). Besides, the function F_m is not injective in either case. To see it, one may note that $F_m(d) = F_m(-d)$. Indeed, any d satisfying $\delta = F_m(d)$ gives an equally good test $T_n(d)$.

After obtaining $\delta^\star$ in (3.5), the optimal difference sequence $d^\star = F_m^{-1}(\delta^\star)$ can be found by, e.g., Durbin–Levinson algorithm and the innovation algorithm (Propositions 5.2.1–5.2.2 of Brockwell and Davis (1991)), where $F_m^{-1}(\cdot)$ denotes any element of the inverse map of F_m defined in Theorem 3. Using $\mathscr{D}_m = \bar{\mathscr{D}}_m$ and $\mathscr{D}_m = \tilde{\mathscr{D}}_m$, we obtain the optimal weight sequence and the optimal difference sequence:

$$\bar{d}^{\star(m)} = F_m^{-1}(\bar{\delta}^{\star(m)}) \qquad \text{and} \qquad \tilde{d}^{\star(m)} = F_m^{-1}(\tilde{\delta}^{\star(m)}), \tag{3.6}$$

respectively, where $\bar{\delta}^{\star(m)} = \delta^\star(\bar{\mathscr{D}}_m; \gamma^Z, \Upsilon^Z)$ and $\tilde{\delta}^{\star(m)} = \delta^\star(\tilde{\mathscr{D}}_m; \gamma^Z, \Upsilon^Z)$. We also call them the autocorrelation efficient sequence and trend-robust autocorrelation efficient sequence, respectively. The optimal tests are denoted as $\bar{T}_n^{\star(m)} = T_n(\bar{d}^{\star(m)})$ and $\tilde{T}_n^{\star(m)} = T_n(\tilde{d}^{\star(m)})$. Unless otherwise stated, $\delta^{\star(m)}$, $d^{\star(m)}$ and $T_n^{\star(m)}$ refer to $\tilde{\delta}^{\star(m)}$, $\tilde{d}^{\star(m)}$ and $\tilde{T}_n^{\star(m)}$, respectively. We may also omit the superscript $^{(m)}$ to $\delta^\star$, $d^\star$ and $T_n^\star$ when the order m is clear in the context. The algorithm is presented in Section A.7 of the supplement. We conclude this subsection with two examples.

Example 3.1. Let $Z_{1:n}$ be generated from an AR(1) model: $Z_i = \psi_1 Z_{i-1} + \epsilon_i$, where $\epsilon_i \sim \mathrm{N}(0, s^2)$ independently and s is chosen such that $\mathrm{var}(Z_i) = 1$. The optimal difference sequences $d^{\star(m)}$ under $\psi_1 \in \{0, \pm 0.5\}$ and $m \in \{2, 3\}$ are tabulated in Table 1. When $\psi_1 = 0$, it reduces to Hall *et al.* (1990)'s optimal sequence; see Section A.4 of the supplementary note for further discussion.

Example 3.2. Let $Z_{1:N}$ be generated from an ARMA(1, 1) model: $Z_i - \psi_1 Z_{i-1} = \epsilon_i + \theta\epsilon_{i-1}$, where $\epsilon_i \sim \mathrm{N}(0, s^2)$ independently and s is chosen such that $\mathrm{var}(Z_i) = 1$. We pick $\psi_1 \in \{0.2, 0.5, 0.8\}$ and $\theta = 0.5$ for positive dependence and $\psi_1 \in$

Table 1: The optimal difference sequences under the AR(1) model.

Differencing order m	AR-parameter ψ_1	Optimal difference sequence $(\tilde{d}_0^\star, \ldots, \tilde{d}_m^\star)$
2	-0.5	$(0.802, -0.267, -0.534)$
	0	$(0.809, -0.500, -0.3090)$
	$+0.5$	$(0.759, -0.639, -0.120)$
3	-0.5	$(0.814, -0.039, -0.520, -0.254)$
	0	$(0.858, -0.383, -0.281, -0.194)$
	$+0.5$	$(0.782, -0.612, -0.074, -0.095)$

$\{-0.2, -0.5, -0.8\}$ and $\theta = -0.5$ for negative dependence. Let $m = 5$, $N = 400$, $\mu_i = f(i/N)$ as stated in (6.1), $\sigma_0 = 1$, $k^\star = N/2$, and $\Delta \in [0, 0.6]$ in (1.2). The power curves of the test $T_n(d)$ with the optimal difference sequence $d^{\star(m)}$ and Hall *et al.* (1990)'s difference sequence are shown in in Figure 1. It shows that the proposed test is significantly more powerful than the naive difference-based test. The improvement is more obvious when the dependence is stronger.

By default, we recommend users to employ $\tilde{d}^{\star(m)}$ since it works for time series with non-constant means; see Assumption 3. If users are sure that the time series has a constant mean, $\bar{d}^{\star(m)}$ can be used instead. In other words, the default choice $\tilde{d}^{\star(m)}$ requires a weaker assumption on the mean structure. Therefore, it protects us against making an incorrect assumption about the mean structure.

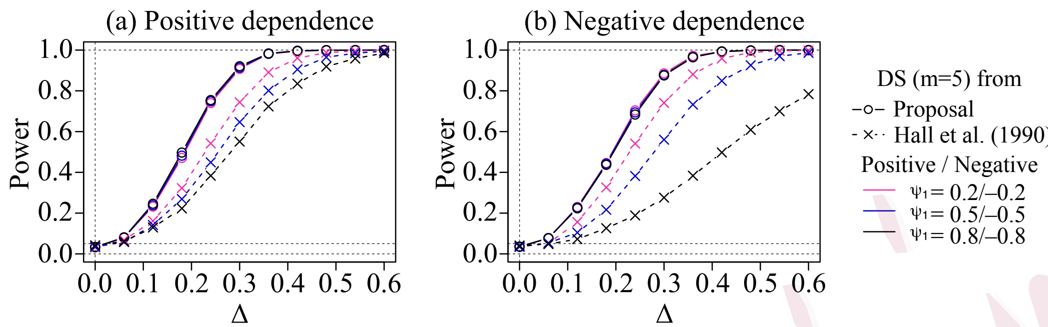

Figure 1: The power curves of the test $T_n(d)$ with two difference sequences (DS) d: the proposed optimal DS (denoted by —○—) and Hall *et al.* (1990)'s DS (denoted by – –×– –); see Example 3.1. Plots (a) and (b) show the results for positive and negative dependence, respectively. For each d, three lines (from top to bottom) indicate different strengths of autocorrelation ($|\psi_1| = 0.2, 0.5, 0.8$, respectively). The number of replications is 2^{12}.

3.2 Asymptotic relative efficiency

This section provides analytical comparisons through asymptotic relative efficiency (ARE); see, e.g., Dehling *et al.* (2017) and Chapter 14.3 of van der Vaart (1998) for details. We consider an alternative hypotheses indexed by Δ:

$$H(\Delta, \tau): \quad \sigma_i = \sigma_0 \exp\left\{\Delta \mathbb{1}_{(i > \lfloor N\tau \rfloor)}\right\} \quad \text{for} \quad i = 1, \ldots, N, \tag{3.7}$$

where $\tau \in (0,1)$. Let $n(d)$ be the number of differenced observations needed for the test $T_n(d)$ to have size α and power β under $H(\Delta, \tau)$. Hence, $n(d)$ is a function of α, β, Δ, and τ.

Suppose that the change Δ satisfies $|\Delta| \downarrow 0$ so that we can construct a sequence of local alternative hypotheses. Then we can compare the tests $T_n(d)$ and $T_n(d')$ via the asymptotic relative efficiency, which is defined as

$$\mathsf{ARE}\left(T_n(d), T_n(d')\right) = \lim_{|\Delta|\downarrow 0} \frac{n(d')}{n(d)}, \tag{3.8}$$

where d and d' are difference sequences with orders $m, m' \in \mathbb{N}$. If $\mathsf{ARE}\,(T_n(d), T_n(d')) > 1$, it signifies that $T_n(d)$ requires fewer observations to detect the variance change at the prescribed size and power. In this case, we say that the test $T_n(d)$ is more efficient than $T_n(d')$. The value of $\mathsf{ARE}\,(T_n(d), T_n(d'))$ depends on α, β, τ, and the rate how $|\Delta| \downarrow 0$. In particular, we consider $\Delta = \bar{\Delta}/N^{1/2}$ with $\bar{\Delta} \neq 0$.

Theorem 4 (Asymptotic relative efficiency)**.** *Let d and d' be two difference sequences of order $m, m' \in \mathbb{N}$. Denote $\delta = F_m(d)$ and $\delta' = F_{m'}(d')$. Also denote $\varpi_m = U_m^{(2)}(\delta; \Upsilon^Z)/U_m^{(1)}(\delta; \gamma^Z)$ and $\varpi'_{m'} = U_{m'}^{(2)}(\delta'; \Upsilon^Z)/U_{m'}^{(1)}(\delta'; \gamma^Z)$. Suppose Assumption 3 holds. If $\Delta = \bar{\Delta}/N^{1/2}$ with $\bar{\Delta} \neq 0$, then*

$$\mathsf{ARE}(T_n(d), T_n(d')) = \frac{U^{(2)}(\delta'; \Upsilon^Z)/U^{(1)}(\delta'; \gamma^Z)}{U^{(2)}(\delta; \Upsilon^Z)/U^{(1)}(\delta; \gamma^Z)} = \frac{\varpi'_{m'}}{\varpi_m}.$$

From Theorem 4, the asymptotic relative efficiency depends on ϖ_m and $\varpi'_{m'}$. It gives ϖ_m an additional statistical interpretation apart from the squared long-run coefficient of variation stated in Theorem 2. See Section A.5 of the supplementary note for further discussion and comparison among different choices of d in terms of the asymptotic relative efficiency. Also, see Example A.1 for a graphical illustration.

4. Estimation of the optimal difference sequence

4.1 Representation of the optimal difference sequence

The optimal difference sequence $d^\star = \tilde{d}^{\star(m)}$ in (3.6) depends on unknown γ^Z and Υ^Z. Estimating them is non-trivial as (Z_i) is masked by varying means (μ_i) and shifting variances (σ_i).

We propose estimate $d^\star$ in two steps: (i) relating the optimal difference sequence $d^\star$ for (X_i) with the optimal average sequence for (∇X_i), and (ii) "inconsistently" estimating the unknowns in the optimal average sequence for (∇X_i). These two steps handle the two aforementioned nuisance structures, respectively. Although the estimators in step (ii) are inconsistent, we show that they still lead to a consistent estimator of $d^\star$ without handling the variance structure.

Suppose $d^\star$ is the optimal difference sequence. Since $\sum_{j=0}^{m} d_j^\star = 0$, we can factorize

$$\phi(L; d^\star) = (c^\star)^{-1/2}\phi(L; a^\star)\phi(L; d^{\star(1)}), \quad \text{where} \quad c^\star = 1 - \sum_{j=0}^{m-2} a_j^\star a_{j-1}^\star, \tag{4.1}$$

$d^{\star(1)} = (1/2^{1/2}, -1/2^{1/2})^\top$ is the first order difference sequence, and $a^\star = (a_0^\star, \ldots, a_{m-1}^\star)^\top$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

is a weight sequence. So, applying the difference operator $\phi(L; d^\star)$ is equivalent to applying the operators $\phi(L; d^{\star(1)})$ and $(c^\star)^{-1/2}\phi(L; a^\star)$ sequentially (at any order).

Let $\nabla Z_i = (Z_{i+1} - Z_i)/2^{1/2}$ for $i \in \mathbb{Z}$. Denote $\gamma_k^{\nabla Z}$ and $\hat{\gamma}_k^{\nabla Z}$ as the autocovariance and the sample autocovariance of $(\nabla Z_i)_{i=1}^{N-1}$ at lag k. Similar to $\Upsilon^Z_{k,k'}$ in (3.2), the long-run covariance of $\hat{\gamma}_k^{\nabla Z}$ and $\hat{\gamma}_{k'}^{\nabla Z}$ is

$$\Upsilon^{\nabla Z}_{k,k'} = \lim_{n\to\infty} n\mathrm{cov}(\hat{\gamma}_k^{\nabla Z}, \hat{\gamma}_{k'}^{\nabla Z}) = \sum_{j=-\infty}^{\infty} \zeta^{\nabla Z}_{k,k'}(j), \tag{4.2}$$

for $k, k' \in \{0, \ldots, m\}$, where $\zeta^{\nabla Z}_{k,k'}(j) = \zeta^{\nabla Z}_{k',k}(-j) = E(\nabla Z_0 \nabla Z_k \nabla Z_j \nabla Z_{j+k'}) - \gamma_k^{\nabla Z}\gamma_{k'}^{\nabla Z}$. The following lemma shows that $a^\star$ is the optimal weight sequence for (∇Z_i).

Lemma 1. *Let $a^\star$ be defined in (4.1). Also let $\gamma^{\nabla Z} = (\gamma_k^{\nabla Z})_k$ and $\Upsilon^{\nabla Z} = (\Upsilon^{\nabla Z}_{k,k'})_{k,k'}$. Then*

$$a^\star = F_{m-1}^{-1}(\alpha^\star), \quad \textit{where} \quad \alpha^\star = \delta^\star(\bar{\mathscr{D}}_{m-1}; \gamma^{\nabla Z}, \Upsilon^{\nabla Z}). \tag{4.3}$$

Lemma 1 is remarkable. Originally, we need to estimate γ^Z and Υ^Z using varying mean and shifting variance data $(X_i)_{i=1}^N$. Now, with Lemma 1, we only need to estimate $\gamma^{\nabla Z}$ and $\Upsilon^{\nabla Z}$ using $(\nabla X_i)_{i=1}^{N-1}$ as an approximate for $(\nabla Z_i)_{i=1}^{N-1}$. Although $(\nabla X_i)_{i=1}^{N-1}$ differs from $(\nabla Z_i)_{i=1}^{N-1}$ by the scaling parameter sequence (σ_i), which might contains change points as well, the structure of the data is substantially simpler as $(\nabla X_i)_{i=1}^{N-1}$ has a nearly constant mean. The next subsection presents estimators that address this issue.

4.2 Consistent estimation of the optimal difference sequence

To solve for $a^\star$, our approach is to construct inconsistent estimators $\hat{\gamma}_k^{\nabla X}$ and $\hat{\Upsilon}_{k,k'}^{\nabla X}$ (to be defined below) of $\gamma_k^{\nabla Z}$ and $\Upsilon_{k,k'}^{\nabla Z}$ so that they are biased by some controllable constant multipliers, i.e., there exist $\kappa_1, \kappa_2 > 0$ such that for all $k, k' \in \{0, \pm 1, \ldots, \pm(m-1)\}$

$$\hat{\gamma}_k^{\nabla X} \xrightarrow{\text{pr}} \kappa_1 \gamma_k^{\nabla Z} \qquad \text{and} \qquad \hat{\Upsilon}_{k,k'}^{\nabla X} \xrightarrow{\text{pr}} \kappa_2 \Upsilon_{k,k'}^{\nabla Z}. \tag{4.4}$$

The constant multipliers κ_1 and κ_2 are expected to be dependent on σ_0^2, τ and Δ under H_1. For this reason, we call $\hat{\gamma}_k^{\nabla Z}$ and $\hat{\Upsilon}_{k,k'}^{\nabla Z}$ partial estimators instead of estimators. This strategy relies on the fact that the minimizer of (3.5) is invariant to scaling factors that are multiplied to the objective function. This approach bypasses estimating or handling the potential variance change points. We highlight that this approach is valid even under multiple change points.

First, we construct $\hat{\gamma}_k^{\nabla X}$. By Assumption 2, we have $\nabla X_i \approx \sigma_i \nabla Z_i$ for most i. So, $E(\nabla X_i) \approx 0$ is a constant approximately. Consequently, the sample auto-covariance of $(\nabla X_i)_{i=1}^{N-1}$ is a potential partial estimator of $\gamma_k^{\nabla Z} = \text{cov}(\nabla Z_0, \nabla Z_k)$, i.e., for $k = 0, \pm 1, \ldots, \pm(m-1)$,

$$\hat{\gamma}_k^{\nabla X} = \frac{1}{N-1} \sum_{i=1+|k|}^{N-1} (\nabla X_i - \hat{\mu}^{\nabla X})(\nabla X_{i-|k|} - \hat{\mu}^{\nabla X}), \tag{4.5}$$

and $\hat{\mu}^{\nabla X} = \sum_{i=1}^{N-1} \nabla X_i/(N-1)$. Denote $\hat{\gamma}^{\nabla X} = (\hat{\gamma}_k^{\nabla X})_k$.

Second, we construct $\hat{\Upsilon}^{\nabla X}_{k,k'}$ for each k and k'. We define a sequence of random m-vectors

$$V_i = (\nabla X_i \nabla X_i,\ \nabla X_i \nabla X_{i-1},\ \ldots,\ \nabla X_i \nabla X_{i-m+1})^{\mathrm{T}} \tag{4.6}$$

as a proxy of the stationary time series $\sigma_0^2(\nabla Z_i \nabla Z_i,\ \nabla Z_i \nabla Z_{i-1},\ \ldots,\ \nabla Z_i \nabla Z_{i-m+1})^{\mathrm{T}}$. One is tempted to use the sample autocovariance of (V_i) in order to construct a kernel estimator for the long-run covariance $\Upsilon^{\nabla Z}_{k,k'}$. If the mean of V_i is constant, then Andrews (1991) provides a consistent estimator for $\Upsilon^{\nabla Z}_{k,k'}$. However, it does not work in our case because $E(V_i) \approx \sigma_i^2(\gamma_0^{\nabla Z}, \ldots, \gamma_{m-1}^{\nabla Z})^{\mathrm{T}}$ is not a constant across i.

We propose to apply a difference operator on V_i to get rid of the potential change point effect. Then our proposed partial estimator of $\Upsilon^{\nabla Z}_{k,k'}$ is $\hat{\Upsilon}^{\nabla X}_{k,k'}$, which is the $(k+1, k'+1)$ entry of

$$\hat{\Upsilon}^{\nabla X} = \sum_{|k| \leqslant B_n} K(k/B_n)\hat{\zeta}_k, \tag{4.7}$$

where $K(\cdot)$ is a kernel function, B_n is a bandwidth, $h_n = 2B_n$ is a lag parameter, and

$$\hat{\zeta}_k = \hat{\zeta}_{-k}^{\mathrm{T}} = \frac{1}{2(N-m)} \sum_{i=m+k}^{N-1} (V_i - V_{i-h_n})(V_{i-k} - V_{i-k-h_n})^{\mathrm{T}}, \qquad k \geqslant 0.$$

For example, the Bartlett kernel $K(x) = (1-|x|)\mathbb{1}_{(|x| \leqslant 1)}$ can be used. The estimator $\hat{\Upsilon}^{\nabla X}$ in (4.7) admits a similar form as in Chan (2022b), however the asymptotic theories concerning $\hat{\Upsilon}^{\nabla X}$ are substantially more challenging because (X_i) potentially has varying mean and varying variance. Alternative estimators can be found in Casini and Perron (2024), Casini (2023) and Chan (2022a). The probability limits of $\hat{\gamma}^{\nabla X}$ and $\hat{\Upsilon}^{\nabla X}$ are stated below.

Theorem 5 (Consistency). *Suppose Assumption 3 holds, and* $u_q \equiv \sum_{i\in\mathbb{Z}} |i|^q |\xi_4(i) < \infty$ *for some* $q \in \mathbb{N}$*. Define* $(\tau_\iota)_{\iota=0}^{M+1}$ *and* $(\Delta_\iota)_{\iota=0}^{M}$ *as in Remark 2.1. If* $1/B_n + B_n/n = o(1)$*, then under* H_0*,* H_1 *or* H_1'*, (4.4) holds for each* $k, k' = 0, \pm 1, \ldots, \pm(m-1)$*, where* $\kappa_1 = \sigma_0^2 \sum_{\iota=0}^{M} (\tau_{\iota+1} - \tau_\iota) e^{2\Delta_\iota}$ *and* $\kappa_2 = \sigma_0^4 \sum_{\iota=0}^{M} (\tau_{\iota+1} - \tau_\iota) e^{4\Delta_\iota}$*. Also,*

$$\delta^\star(\bar{\mathscr{D}}_{m-1}; \hat{\gamma}^{\nabla X}, \hat{\Upsilon}^{\nabla X}) \xrightarrow{\text{pr}} \delta^\star(\bar{\mathscr{D}}_{m-1}; \gamma^{\nabla Z}, \Upsilon^{\nabla Z}).$$

The assumption $u_q < \infty$ in Theorem 5 is used to characterize the strength of serial dependence of (Z_i). The larger the value of q, the weaker the serial dependence of (Z_i). The mean squared error-optimal B_n is stated below.

Theorem 6 (Optimal bandwidth). *Assume the conditions in Theorem 5. The* $\mathscr{L}^2$*-optimal bandwidth for* $\hat{\Upsilon}_{k,k'}^{\nabla X}$ *is* $B_n = O\{n^{1/(1+2q)}\}$ *for each* $k, k' = 0, \pm 1, \ldots, \pm(m-1)$.

Without prior knowledge, we assume $q = 1$, which corresponds to the weakest assumption among $q \in \mathbb{N}$. The algorithms for computing the optimal difference sequence $\tilde{d}^\star$ and the optimal average sequence $\bar{d}^\star$ can be found in Algorithms A.1 and A.2 of the supplement.

5. Estimation of the long-run variance

5.1 Doubly robust estimators

This section addresses the estimation of v_m and $v_{m,L}$ for the tests $T_n^{\star(m)}$ and $T_n^{\star(m,L)}$. Our goal is to derive estimators that are doubly robust to (i) varying means $\mu_1, \ldots, \mu_N$, and (ii) change points in variance $\sigma_1, \ldots, \sigma_N$. It ensures that the tests are as powerful as we claim in Section 3.2.

The classical estimators (e.g., Andrews, 1991; Liu and Wu, 2010; Chan and Yau, 2017)

$$\tilde{v}_m = \sum_{k=1-n}^{n-1} K(k/B_n)\tilde{\gamma}_k^Q \qquad \text{and} \qquad \tilde{v}_{m,L} = \frac{\tilde{v}_m}{(\bar{Q}_n)^2}, \tag{5.1}$$

are consistent for v_m and $v_{m,L}$ only under H_0 but not H_1, where $\tilde{\gamma}_k^Q = \sum_{i=|k|+1}^{n}(Q_i - \bar{Q}_n)(Q_{i-|k|} - \bar{Q}_n)/n$; see also Chan and Yau (2024) for an alternative estimator for v_m. To see it, we note that $E(Q_1), \ldots, E(Q_n)$ are not constant when H_0 is false. So, $\bar{Q}_n$ cannot be consistent for all $E(Q_i)$. Utilizing an inconsistent long-run variance estimator in a test statistic leads to a drop of power, especially under early and late change points as well as the existence of multiple change points; see, e.g., Gerstenberger *et al.* (2020).

To derive doubly robust estimators, we need the following representation. Recall from Lemma 1 that $\alpha^\star = (\alpha_0^\star, \ldots, \alpha_{m-1}^\star)^\top$, $c^\star = 1-\alpha_1^\star$, and $\alpha_k^\star = \sum_{j=|k|}^{m-1} a_j^\star a_{j-|k|}^\star$ for $|k| \leqslant m-1$. Then, in view of Theorem 2, the long-run variances can be expressed as

$$v_m = (c^\star)^{-2} U^{(2)}_{m-1}(\alpha^\star; \Upsilon^{\nabla Z}) \qquad \text{and} \qquad v_{m,L} = \frac{U^{(2)}_{m-1}(\alpha^\star; \Upsilon^{\nabla Z})}{U^{(1)}_{m-1}(\alpha^\star; \gamma^{\nabla Z})},$$

where $U^{(1)}_m$ and $U^{(2)}_m$ are defined in (3.3). Our proposed estimators for v_m and $v_{m,L}$ are

$$\hat{v}_m = (c^\star)^{-2} U^{(2)}_{m-1}(\hat{\alpha}^\star; \hat{\Upsilon}^{\nabla X}) \qquad \text{and} \qquad \hat{v}_{m,L} = \frac{U^{(2)}_{m-1}(\hat{\alpha}^\star; \hat{\Upsilon}^{\nabla X})}{U^{(1)}_{m-1}(\hat{\alpha}^\star; \hat{\gamma}^{\nabla X})}, \tag{5.2}$$

where $\hat{\alpha}^\star = \delta^\star(\bar{\mathscr{D}}_{m-1}; \hat{\gamma}^{\nabla X}, \hat{\Upsilon}^{\nabla X})$; and $\hat{\Upsilon}^{\nabla X}$ and $\hat{\gamma}^{\nabla X}$ are defined in (4.7) and (4.5), respectively. The theorem below states that the estimators $\hat{v}_m$ and $\hat{v}_{m,L}$ are doubly robust, while $\tilde{v}_m$ and $\tilde{v}_{m,L}$ are not.

Theorem 7 (Double robustness)**.** *Suppose Assumption 3 is satisfied and* $u_q = \sum_{i\in\mathbb{Z}} |i|^q |\xi_4(i) < \infty$ *for some* $q \in \mathbb{N}$*. If* $1/B_n + B_n/n = o(1)$*, then the following results hold. (1) Under* H_0*, we have* $\tilde{v}_m \overset{\text{pr}}{\to} v_m$ *and* $\tilde{v}_{m,L} \overset{\text{pr}}{\to} v_{m,L}$*. (2) Under* H_0*,* H_1 *or* H_1'*, we have* $\hat{v}_m \overset{\text{pr}}{\to} v_m$ *and* $\hat{v}_{m,L} \overset{\text{pr}}{\to} v_{m,L}$*.*

Remark 5.1. We highlight that $\hat{v}_m$ and $\hat{v}_{m,L}$ do not require any additional estimation since their building blocks $\hat{\alpha}^\star$, $\hat{\Upsilon}^{\nabla X}$ and $\hat{\gamma}^{\nabla X}$ are computed in Algorithm A.1 of the supplement. Also, this does not require any prior information or assumption on the number of change points. Incorporating information of change points can lead to better performance, and this is illustrated in the null-protected estimators in Section 5.2.

5.2 Null-protected estimators

According to Theorem 7, $\tilde{v}_m$ and $\tilde{v}_{m,L}$ are consistent only under H_0, while $\hat{v}_m$ and $\hat{v}_{m,L}$ are consistent under H_0, H_1 and H_1'. This is achieved because $\hat{v}_m$ and $\hat{v}_{m,L}$ are constructed based on difference-based estimators, which are consistent even in the presence of variance change point. However, the price for this consistency is that its efficiency is slightly lower; see Chan (2022b) for a similar phenomenon. Consequently, the estimators $\tilde{v}_m$ and $\tilde{v}_{m,L}$ work better than $\hat{v}_m$ and $\hat{v}_{m,L}$ under H_0, while $\hat{v}_m$ and $\hat{v}_{m,L}$ work better than $\tilde{v}_m$ and $\tilde{v}_{m,L}$ under H_1 and H_1'.

Using good estimators of the the long-run variances under H_0 ensures that the tests control the size (i.e., type-I error rate) well. It motivates us to combine (5.1) and (5.2) so that the resulting tests are more size-accurate. Precisely, we propose to estimate v_m and $v_{m,L}$ by

$$\hat{v}_m(\lambda) = \lambda\hat{v}_m + (1-\lambda)\tilde{v}_m \quad \text{and} \quad \hat{v}_{m,L}(\lambda) = \lambda\hat{v}_{m,L} + (1-\lambda)\tilde{v}_{m,L}, \tag{5.3}$$

where $\lambda \in [0,1]$ is a weight. Ideally, $\lambda \approx 1$ under H_1.

We propose a subsampling scheme to formulate our data-driven $\hat{\lambda}$ of λ. The data are partitioned into g_n groups each of approximately length $\ell_n = \lfloor n/g_n \rfloor$. We require that $1/\ell_n + \ell_n/n = o(1)$ as $n \to \infty$. Using a similar notation as in Carlstein (1986), the group starting on index $i+1$ with length ℓ_n is denoted as $D^i_{\ell_n} = (D_{i+1}, D_{i+2}, \ldots, D_{i+\ell_n})$. We set the non-overlapping groups to be $\{D^{i_1}_{\ell_{n,1}}, D^{i_2}_{\ell_{n,2}}, \ldots, D^{i_{g_n}}_{\ell_n,g_n}\}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

where $i_j = \sum_{j'=0}^{j-1} \ell_{n,j'}$ and $\ell_{n,j'} = \lfloor n/g_n \rfloor + \mathbb{1}_{(j' \leqslant n - g_n \lfloor n/g_n \rfloor)}$, for $j = 1, \ldots, g_n$. The sample variance of the jth group is

$$\hat{s}_j = \frac{1}{\ell_{n,j}} \sum_{i=i_j+1}^{i_j+\ell_{n,j}} (D_i - \hat{m}_j)^2, \quad \text{where} \quad \hat{m}_j = \frac{1}{\ell_{n,j}} \sum_{i=i_j+1}^{i_j+\ell_{n,j}} D_i.$$

Let $\bar{s} = \sum_{i=1}^{g_n} \hat{s}_i / g_n$. Also let

$$\hat{\varsigma}_0 = \frac{1}{g_n} \sum_{j=1}^{g_n} (\hat{s}_j - \bar{s})^2 + e_n \quad \text{and} \quad \hat{\varsigma}_1 = \frac{1}{2(g_n - 1)} \sum_{i=1}^{g_n - 1} (\hat{s}_{i+1} - \hat{s}_i)^2 + e_n$$

with a positive sequence $e_n \downarrow 0$ to avoid $\hat{\varsigma}_1/\hat{\varsigma}_0 = 0/0$. By default, $e_n = 1/n$. Our proposed $\hat{\lambda}$ is

$$\hat{\lambda} = \max\left(1 - \hat{\varsigma}_1/\hat{\varsigma}_0,\ 0\right), \tag{5.4}$$

The subsampled variance estimator $\hat{\varsigma}_0$ assumes mean constancy of all $(\hat{s}_i)$ whereas $\hat{\varsigma}_1$ is still robust under a variance change. Hence, the ratio $\hat{\varsigma}_1/\hat{\varsigma}_0$ measures the degree of constancy of (σ_i). When H_0 is false, $\hat{\varsigma}_0$ diverges and leads to a larger value of $\hat{\lambda}$.

In a nutshell, our proposed null-protected estimators of v_m and $v_{m,L}$ are $\hat{v}_m(\hat{\lambda})$ and $\hat{v}_{m,L}(\hat{\lambda})$, respectively. The following theorem states the consistency under H_0, H_1 and H_1'.

Theorem 8 (Consistency of $\hat{v}_m(\hat{\lambda})$ and $\hat{v}_{m,L}(\hat{\lambda})$)**.** *Assume the conditions in Theorem 7. If $1/\ell_n + \ell_n/n = o(1)$, then $\hat{v}_m(\hat{\lambda}) \xrightarrow{\text{pr}} v_m$ and $\hat{v}_{m,L}(\hat{\lambda}) \xrightarrow{\text{pr}} v_{m,L}$ under H_0, H_1 or H_1'.*

Theorem 8 shows that $\hat{v}_m(\hat{\lambda})$ and $\hat{v}_{m,L}(\hat{\lambda})$ continue to be doubly robust. The subsample size ℓ_n is a parameter for $\hat{\lambda}$, which does not affect the final estimators $\hat{v}_m(\hat{\lambda})$ and $\hat{v}_{m,L}(\hat{\lambda})$ significantly.

Remark 5.2. As suggested by one of the anonymous referees, in (5.3), it is possible to replace $\hat{v}_m$ and $\hat{v}_{m,L}$ with estimators that can directly estimate and remove the change points. This may possibly improve the finite-sample performance. However, since it is beyond the scope of the current article, we leave the thorough investigation in future research.

6. Simulation experiment

We evaluate the finite-sample performance of our proposed test $T_n^{\star(m,L)} = T_n^{(L)}(\tilde{d}^{\star(m)})$ in (2.7) together with the null-protected estimator $\hat{v}_{m,L}(\hat{\lambda})$ in Section 5.2 via simulation experiments. The following settings are chosen. The data are generated as $X_i = \mu_i + \sigma_i Z_i$ for $i = 1, \ldots, N$. We consider a non-constant mean function: $\mu_i = f(i/N)$, where

$$f(t) = 3\cos(5t) - 10(t - 0.7)^2 + 2(t - 0.3)^3, \qquad t \in [0, 1]. \tag{6.1}$$

Four noise models for $(Z_i)_{i=1}^N$ are considered. The innovations $(\epsilon_i)_{i\in\mathbb{Z}}$ below are independent $\mathrm{N}(0, s^2)$ with s selected such that $\mathrm{var}(Z_i) = 1$. The value of s may vary across different models.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0238

- Case 1: ARMA$(1,1)$ model. $Z_i = 0.9Z_{i-1} + 0.9\epsilon_{i-1} + \epsilon_i$ for each i.
- Case 2: AR(3) model. $Z_i = -0.3Z_{i-1} - 0.54Z_{i-2} - 0.112Z_{i-3} + \epsilon_i$ for each i.
- Case 3: Bilinear model. $Z_i = (0.3 + 0.6\epsilon_i)Z_{i-1} + \epsilon_i$ for each i.
- Case 4: ARCH model. $Z_i = \sigma_i\epsilon_i$ and $\sigma_i^2 = 0.2 + 0.8Z_{i-1}^2$ for each i.

Case 1 generates highly autocorrelated data. Case 2 produces a non-monotone autocovariance function. Case 3 is a non-linear time series. Case 4 generates conditionally heteroskedastic data. Two existing mean-robust variance change point tests are considered for comparison:

- (SPL) Gao *et al.* (2019) proposed to use smoothing splines to get rid of the mean effect, however, their theory only supports independent and identically distributed Gaussian noises.
- (KE) Lee *et al.* (2003) proposed to use a quadratic smoothing kernel to detrend the data. They suggest a flat top kernel with a bandwidth $\lfloor n^{1/4} \rfloor$ for estimating the long-run variance.

Unless otherwise specified, we set $N = 400$, $\sigma_0^2 = 1$, and the nominal size at 5%. The power curves of the proposed tests ($T_n^{\star(m,L)}$, $m = 1, 2, 3$) and the above two existing tests are simulated at different values of Δ. The number of replications of each scenario is 4096. We set $B_n = \lfloor 2n^{1/3} \rfloor$ and $\ell_n = \lfloor n^{1/2} \rfloor$ throughout the simulation experiments.

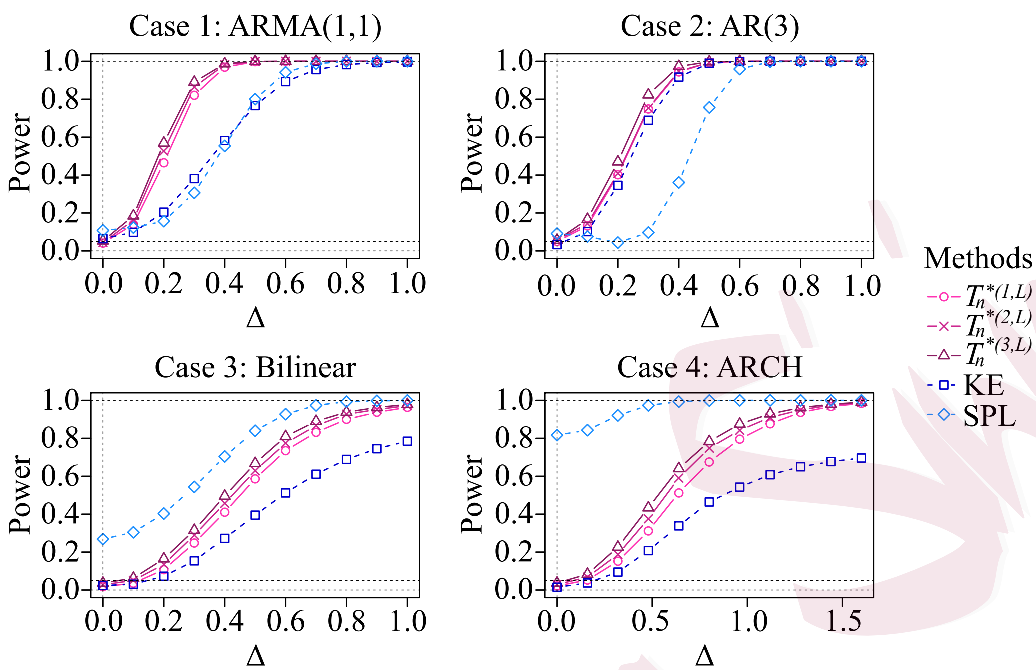

Figure 2: The power curves of $T_n^{\star(m,L)}$ (our proposals), KE and SPL are plotted under time series models in cases 1–4 stated in Section 6. Dashed horizontal lines indicate zero, 5% and one.

First, we consider the one-change point alternative hypothesis H_1 with $k^\star = \lfloor 0.5N \rfloor$. Figure 2 shows the results. Our methods generally have a higher power compared to the existing methods. The improvement is prominent. On the other hand, the size of SPL is faulty since it does not accustom to serial dependence. KE gets more penalized in all cases as the data deviates from independence.

Second, we consider the multiple-change point alternative H_1' with

$$\sigma_i = \sigma_0 \exp\left\{\Delta \mathbb{1}_{(\lfloor 0.3N \rfloor < i \leqslant \lfloor 0.7N \rfloor)} - \Delta \mathbb{1}_{(\lfloor 0.7N \rfloor < i \leqslant N)}\right\}, \tag{6.2}$$

which consists of two variance change points at $\lfloor 0.3N \rfloor$ and $\lfloor 0.7N \rfloor$ when $\Delta \neq 0$. The power curves are plotted in Figure 3. Our proposals continue to outperform the existing methods significantly, and the improvement is even more obvious than in the single change point case. Besides, the power of $T_n^{\star(m,L)}$ also increases with m as Algorithm A.1 of the supplement is still valid under H_1'.

Additional simulation experiments can be found in Section B of the supplement. These experiments include (i) comparisons across different strengths of serial dependence, (ii) comparisons between tests with or without variance stabilizing transforms, and (iii) comparisons between tests that use the TRACE sequence, the ACE sequence and the classical difference sequence proposed in Hall *et al.* (1990); see Sections B.1, B.2, and B.3 of the supplementary note, respectively.

7. Market Sentiment in the Crypto Market

We illustrate an application on detecting signals in the crypto market. The Spent Output Profit Ratio (SOPR) of Bitcoin (BTC) is a key indicator that tracks investors' behaviour on trading Bitcoin. More description of the indicator can be found on website of Glassnode. The indicator calculates the aggregated profit and loss for all coins moved on chains for a day. A SOPR greater than 1 indicates that

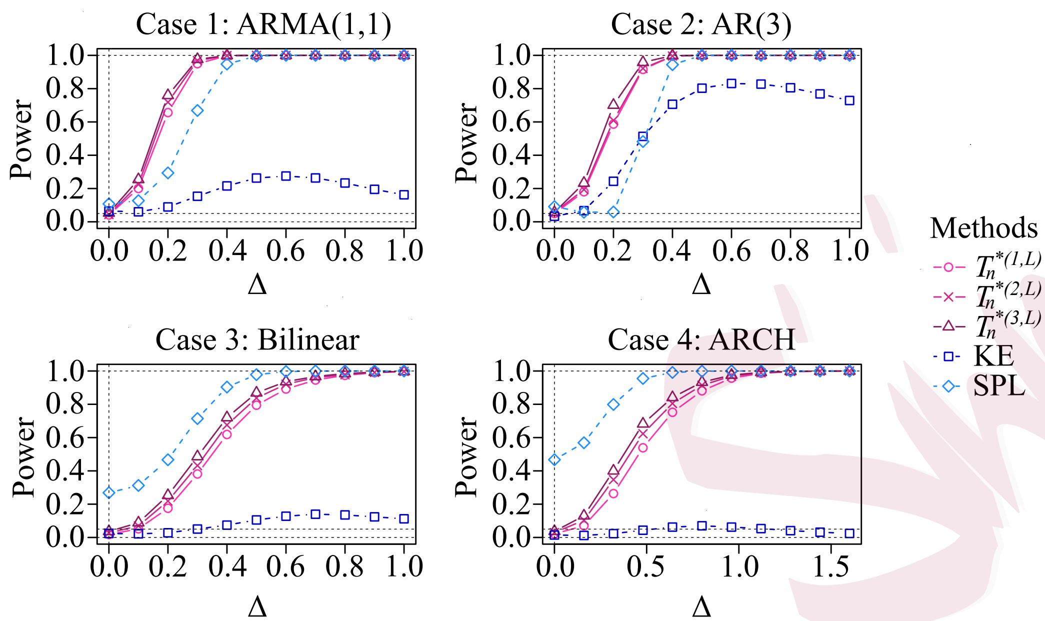

Figure 3: Simulation results under time series models in cases 1–4. The multiple-change point alternative (6.2) is used. Dashed horizontal lines indicate zero, 5% and one.

coins are selling at a profit, and vice versa.

We conjecture that market sentiment is not only determined by the level of this ratio, but also its variance, because the volatility measures the activeness of the investors in the market, which affects the common interpretation of that signal. To justify it, we employ the proposed test $T_n^{\star(3,L)}$, and use binary segmentation (Vostrikova, 1981) on the SOPR of Bitcoin to locate the change points. Notice that

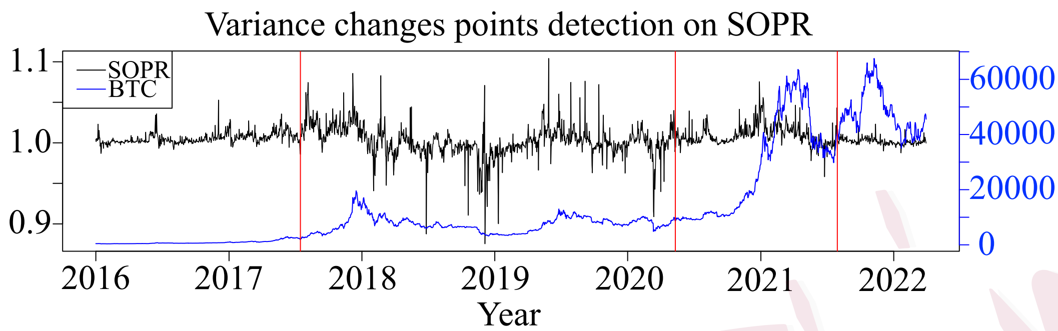

Figure 4: SOPR and BTC price from January 1, 2016 to March 31, 2022 are plotted. Data are retrieved from Glassnode. The vertical lines indicate change points detected by our proposal.

we use binary segmentation to locate the multiple change points for illustration. The benefits of more robust methods, such as wild binary segmentation (Fryzlewicz, 2014) and the two-stage procedure proposed in Cho and Kirch (2022) may possibly yield better results. Since it is beyond the scope of the current article, we leave it for future research.

The data from January 1, 2016 to March 31, 2022 ($N = 2282$) are collected. Figure 4 shows the time series plot of SOPR and BTC with the detected change points. It is observed that SOPR has a varying mean so it suggests us to use our proposed test. For inference on the mean structure, readers may refer to, e.g., Wu (2004) and Fryzlewicz (2014).

Comparing with the KE method, our method detected three change points on July 16, 2017, May 11, 2020 and July 30, 2021 with p-values 2.74×10^{-4}, 6.56×10^{-4} and 3.67×10^{-4}, respectively, while KE fails to reject H_0 with a p-value of 6.98%. After partitioning the SOPR by detected variance changes, we further investigate the significance of this variance change point detection. From Figure 4, it is observed that the price of BTC also behaves differently in those periods. The first detected change point (July 16, 2017) locates the time when BTC price starts to fluctuate in a higher level compared to the previous period. The second detected change point (May 11, 2020), which signifies a drop in the variance of SOPR, coincides with the time that BTC price starts to soar. The last change point (March 31, 2022) detected that the variance of SOPR further drops, in which BTC prices fluctuate a lot without an upward trend. All these findings obtained from inferences on variance changes are useful for market sentiment analysis.

Supplementary Material

It includes the extra theoretical results, additional simulation experiments and proofs.

Acknowledgments

The authors would like to thank the anonymous referees, an associate editor, and the editor for their constructive comments that improved the scope and presentation of the paper. Chan thanks the partial support provided by grants GRF-14304420, 14306421, and 14307922 from the Research Grants Council of HKSAR.

References

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

Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009) Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, **37**, 4046–4087.

Bai, L. and Wu, W. (2024) Difference-based covariance matrix estimate in time series nonparametric regression with applications to specification tests. *Biometrika*, **111**, 1277–1292.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. Springer, New York.

Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, **14**, 1171–1179.

Casini, A. (2023) Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *J. Econometrics*, **235**, 372–392.

Casini, A. and Perron, P. (2024) Prewhitened long-run variance estimation robust to nonstationarity. *J. Econometrics*, **242**, 105794.

Chan, K. W. (2022a) Mean-structure and autocorrelation consistent covariance matrix estimation. *Journal of Business & Economic Statistics*, **40**, 201–215.

Chan, K. W. (2022b) Optimal difference-based variance estimators in time series: A general framework. *Ann. Statist.*, **50**, 1376–1400.

Chan, K. W. and Yau, C. Y. (2017) High order corrected estimator of asymptotic variance with optimal bandwidth. *Scand. J. Statist.*, **44**, 866–898.

Chan, K. W. and Yau, C. Y. (2024) Asymptotically constant risk estimator of the time-average variance constant. *Biometrika*, **111**, 825–842.

Chapman, J.-L., Eckley, I. A. and Killick, R. (2020) A nonparametric approach to detecting changes in variance in locally stationary time series. *Environmetrics*, **30**, e2576.

Chen, J. and Gupta, A. K. (1997) Testing and locating variance changepoints with application to stock price. *J. Am. Statist. Ass.*, **92**, 739–747.

Cheng, C. H. and Chan, K. W. (2024) A general framework for constructing locally self-normalized multiple-change-point tests. *J. Bus. Econom. Statist.*, **42**, 719–731.

Cho, H. and Kirch, C. (2022) Two-stage data segmentatino permitting multiscale change points, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics*, **74**, 653–684.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006) Structural break estimation for nonstationary time series models. *J. Am. Statist. Ass.*, **101**, 223–239.

Dehling, H., Rooch, A. and Taqqu, M. S. (2017) Power of change-point tests for long-range dependent data. *Electronic Journal of Statistics*, **11**, 2168–2198.

Dinkelbach, W. (1967) On nonlinear fractional programming. *Management Science*, **13**, 492–498.

Fryzlewicz, P. (2014) Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, **42**, 2243–2281.

Gao, Z., Shang, Z., Du, P. and Robertson, J. L. (2019) Variance change point detection under a smoothly-changing mean trend with application to liver procurement. *J. Amer. Statist. Assoc.*, **114**, 773–781.

Gerstenberger, C., Vogel, D. and Wendler, M. (2020) Tests for scale changes based on pairwise differences. *J. Am. Statist. Ass.*, **115**, 1336–1348.

Hall, P., Kay, J. W. and Titterinton, D. M. (1990) Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.

Hsu, D.-A., Miller, R. B. and Wichern, D. W. (1974) On the stable paretian behavior of stock-market prices. *J. Amer. Statist. Assoc.*, **69**, 108–113.

Inclan, C. and Tiao, G. C. (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Am. Statist. Ass.*, **89**, 913–923.

Lee, S., Na, O. and Na, S. (2003) On the cusum of squares test for variance change in nonstationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics*, **55**, 467–485.

Lee, S. and Park, S. (2001) The Cusum of squares test for scale changes in infinite order moving average processes. *Scand. J. Statist.*, **28**, 625–644.

Liu, W. and Wu, W. B. (2010) Asymptomatic of spectral density estimates. *Econometric Theory*, **26**, 1218–1245.

Lobato, I. N. (2001) Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, **96**, 1066–1076.

Phillips, P. (1987) Time series regression with a unit root. *Econometrica*, **55**, 277–301.

Schaible, S. and Ibaraki, T. (1983) Fractional programming. *European Journal of Operational Research*, **12**, 325–338.

Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *J. Roy. Statist. Soc. Ser. B*, **72**, 343–366.

Shao, X. and Zhang, X. (2010) Testing for change points in time series. *J. Amer. Statist. Assoc.*, **105**, 1228–1240.

To, H. K. and Chan, K. W. (2023) Mean stationarity test in time series: A signal variance-based approach. *Bernoulli*, **30**, 1231–1256.

van der Vaart, A. W. (1998) *Asymptotic Statistics.* Cambridge University Press.

Vostrikova, L. (1981) Detection of the disorder in multidimensional random-processes. *Doklady Akademii Nauk SSSR*, **259**, 270–274.

Wu, W. B. (2004) A test for detecting changes in mean. In *Time Series Analysis and Applications to Geophysical Systems* (eds. D. R. Brillinger, E. A. Robinson and F. Schoenberg.), vol. 139, 105–122. Springer-Verlag New York.

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

Wu, W. B. (2011) Asymptotic theory for stationary processes. *Stat. Interface*, **4**, 207–226.

Wu, W. B. and Zhao, Z. (2007) Inference of trends in time series. *J. R. Statist. Soc. B*, **69**, 391–410.

Zhang, T. and Lavitas, L. (2018) Unsupervised self-normalized change-point testing for time series. *J. Amer. Statist. Assoc.*, **113**, 637–648.

Department of Statistics, The Chinese University of Hong Kong.

E-mail: `cwdleung@link.cuhk.edu.hk` and `kinwaichan@sta.cuhk.edu.hk`