A General Framework for Constructing Locally Self-Normalized Multiple-Change-Point Tests

Cheuk Hin Cheng and Kin Wai Chan

Department of Statistics, The Chinese University of Hong Kong, Hong Kong

ABSTRACT

We propose a general framework to construct self-normalized multiple-change-point tests with time series data. The only building block is a user-specified single-change-detecting statistic, which covers a large class of popular methods, including the cumulative sum process, outlier-robust rank statistics, and order statistics. The proposed test statistic does not require robust and consistent estimation of nuisance parameters, selection of bandwidth parameters, nor pre-specification of the number of change points. The finite-sample performance shows that the proposed test is size-accurate, robust against misspecification of the alternative hypothesis, and more powerful than existing methods. Case studies of the Shanghai-Hong Kong Stock Connect turnover are provided.

1. Introduction

Testing structural stability is critical in statistical inference. There is a considerable amount of literature and practical interest in testing for the existence of change points. Csörgö and Horváth (1988, 1997) provided comprehensive solutions to the at-most-one-change (AMOC) problem using both classical parametric and nonparametric approaches. One method assumes a fixed number of change points \(n\) for the multiple-change-point problem. The test statistics are constructed by dividing the data into segments and maximizing a sum of AMOC statistics computed within each segment; see Antoch and Jarušková (2013) and Shao and Zhang (2010). However, such approaches are vulnerable to misspecification of the true number of change points \(M\) and the computational cost often increases exponentially with \(m\). Another method uses sequential estimation (Vostrikova 1981; Bai and Perron 1998). This class of methods sequentially performs AMOC tests and produces estimates if the no-change-point hypothesis is rejected. Because the structural difference of data before and after a particular change point may be less apparent in the presence of other change points, the AMOC-type tests may lose their power under the multiple-change-point setting.

Localization can be used to improve power. Bauer and Hackl (1980) introduced the moving sum approach. Chu, Hornik, and Kaun (1995) and Eichinger and Kirch (2018) later applied the approach to change-point detection. This method selects a fixed window size and performs inference on each subsample. Fryzlewicz (2014) proposed wild binary segmentation, which draws subsamples without fixing the window size, to avoid window size selection. Both algorithms attempt to find a subsample containing only one change point to boost power without specifying \(m\). If the standard CUSUM process is used, the methods require consistent and change-point-robust estimation of the long-run variance, which can be nontrivial and may require tuning a bandwidth parameter (Andrews 1991). Worse still, different bandwidths can noticeably influence the tests (Kiefer and Vogelsang 2005). To avoid the challenging estimation, Lobato (2001) first proposed using a self-normalizer that converges to a pivotal distribution proportional to the nuisance parameter. Shao and Zhang (2010) later applied self-normalization to the AMOC problem. In the multiple-change-point problem, Zhang and Lavitas (2018) proposed a self-normalized test that intuitively scans for the first and last change points. However, it may compromise power when larger changes occur in the middle.

In this article, our contributions are as follows: (a) We propose a novel localized self-normalization framework to transform AMOC statistics into powerful and size-accurate multiple-change-point tests, which apply to a broad class of time series models and a large class of AMOC statistics, including the CUSUM process and outlier-robust change-detecting statistics, instead of any specific process as constructed in existing methods. (b) The methods support multiple-change-point detection in general model parameters. The resulting tests achieve promising power. (c) Our test is also shown to be locally powerful asymptotically. (d) The novel self-normalizers can be recursively computed. (e) The test allows an intuitive extension to estimation with good theoretical properties.

The rest of the article is structured as follows. Section 2 reviews the self-normalization method proposed by Shao and Zhang (2010) under the AMOC problem. Section 3.1 presents the proposed framework, and of Sections 3.2–3.4 explore applications of some popular change-detecting processes and extensions to estimation. In Section 4, the differences between various self-normalized approaches are comprehensively discussed. Some implementation issues are also discussed.
Simulation results are presented in Section 5, which show good size accuracy and a substantial improvement in power. The article ends with a real-data analysis of the Shanghai–Hong Kong Stock Connect turnover in Section 6. Proofs of the main results are presented in the Appendix. A supplement containing additional simulation results, recursive formulas, algorithms, and an extension to long-range dependent time series is also included. An R-package “SDnwt” is available online.

2. Introduction of Self-Normalization

Consider the signal-plus-noise model: $X_i = \mu_i + Z_i$, where $\mu_i = \mathbb{E}(X_i)$ for $i = 1, \ldots, n$ and $\{Z_i : i \in \mathbb{Z}\}$ is a stationary noise sequence. The AMOC testing problem is formulated as follows:

$H_0 : \mu_1 = \cdots = \mu_n,$ \hspace{1cm} (2.1)

$H_1 : \exists 1 < k_1 < n, \mu_1 = \cdots = \mu_{k_1} \neq \mu_{k_1+1} = \cdots = \mu_n.$ \hspace{1cm} (2.2)

Let $\xi_{k_1,k_2} = \sum_{i=k_1}^{k_2} X_i$ if $1 \leq k_1 \leq k_2 \leq n$; $\xi_{k_1,k_2} = 0$ otherwise. A popular statistic for detecting a single change point is the CUSUM process, defined as

$$C_n(|nt|) = n^{-1/2} \sum_{i=1}^{|nt|} (X_i - n^{-1}\xi_{1,n}),$$ \hspace{1cm} (2.3)

where $t \in [0,1]$ and $|nt|$ is the largest integer part of $nt$. The limiting distribution of the functional in (2.3) is established by the following assumption.

**Assumption 2.1.** As $n \to \infty$, $n^{-1/2} \sum_{i=1}^{|nt|} Z_i : t \in [0,1] \Rightarrow \{\sigma \varpi(t) : t \in [0,1]\}$, where $\sigma^2 = \lim_{n \to \infty} \text{var}(\xi_{1,n})/n \in (0, \infty)$ is the long-run variance, $\varpi(\cdot)$ is the standard Brownian motion and “$\Rightarrow$” denotes convergence in distribution in the Skorokhod space (Billingsley 1999).

Assumption 2.1 is known as the functional central limit theorem or Donsker’s invariance principle. Under regularity conditions, Assumption 2.1 is satisfied. For example, Herrndorf (1984) proved the functional central limit theorem for the dependent data under some mixing conditions; Wu (2007) later proved the strong convergence for stationary processes using physical and predictive dependence measures. Under the null hypothesis $H_0$, by the continuous mapping theorem, we have $\{C_n(|nt|) : t \in [0,1]\} \Rightarrow \{\sigma \varpi(t) - \sigma \varpi(1) : t \in [0,1]\}$.

Classically, the celebrated Kolmogorov–Smirnov test statistic, defined as $K_n(\cdot) = \sup_{k=1,\ldots,n} |C_n(k)/\sigma|$, can be used for the AMOC problem. Because $\sigma$ is typically unknown, we estimate it by an estimator $\hat{\sigma}$ that is consistent under $H_0$ and $H_1$; see Chan (2022a, 2022b) for some possible estimators. Hence, $K_n(\hat{\sigma})$ converges to $\sup_{e \in (0,1)} |\varpi(t) - \varpi(1)|$, which is known as the Kolmogorov distribution. Shao and Zhang (2010) proposed to bypass the estimation of $\sigma$ by normalizing $C_n(k)$ by a nondegenerate standardizing random process called a self-normalizer:

$$V_n(k) = n^{-2} \left\{ \sum_{i=1}^{k} \left( \xi_{1,i} - \frac{i}{k} \xi_{1,k} \right)^2 + \sum_{i=k+1}^{n} \left( \xi_{i,n} - \frac{n - i + 1}{n - k} \xi_{k+1,n} \right)^2 \right\},$$

where $k = 1, \ldots, n - 1$. The resulting self-normalized test statistic for the AMOC problem is $S_n^{(1)}$, where

$$S_n^{(1)} = \sup_{k=1,\ldots,n-1} S_n^{(1)}(k) \text{ and } (2.4)$$

$$S_n^{(1)}(k) = C_n^2(k)/V_n(k).$$

Under Assumption 2.1 and $H_0$, the limiting distribution of $S_n^{(1)}$ is nondegenerate and pivotal. The nuisance parameter $\sigma^2$ is asymptotically cancelled out in the numerator $C_n^2(k)$ and the denominator $V_n(k)$ on the right-hand side of (2.4). Moreover, because there is no change point in the intervals $[1, k_1]$ and $[k_1 + 1, n]$ under $H_1$, the self-normalizer at the true change point, $V_n(k_1)$, is invariant to the change, and therefore their proposed test does not suffer from the well-known non-monotonic power problem; see, for example, Vogelsang (1999).

However, this appealing feature no longer exists in the multiple-change-point setting. Moreover, their proposed self-normalizer is specifically designed for normalizing the CUSUM process, which may not be the best choice if the data have a heavy-tailed distribution. In this section, we propose a framework for extending user-specified AMOC statistics to a multiple-change-point self-normalized test using the localization idea.

3. General Framework

3.1. Locally Self-Normalized Test Statistic

We consider the multiple-change-point testing problem, that is, to test the $H_0$ in (2.1) against

$$H_{\geq 1} : \mu_1 = \cdots = \mu_{k_1} \neq \mu_{k_1+1} = \cdots = \mu_{k_2} \neq \cdots \neq \mu_{k_M+1} = \cdots = \mu_n,$$

for unknown times $1 < k_1 < \cdots < k_M < n$, and an unknown number of change points $M \geq 1$. In this section, combining the advantages of localization and self-normalization, we lay out the underlying principles to define a general class of statistics for testing the existence of multiple change points.

Generally, for the AMOC problem, one may consider any specific global change-detecting process, $\{D_n(|nt|)\}_{0 \leq t \leq 1}$. For some positive definite matrix $\sigma_D$ and some empirical process $\mathbb{D}(\cdot)$ with a pivot distribution, $\{D_n(|nt|)\}_{0 \leq t \leq 1}$ converges weakly to $\sigma_D \mathbb{D}(\cdot)$ in the Skorokhod space. Typically, $D_n(k)$ is designed to detect whether $k$ is a change point for $k = 1, \ldots, n$. We generalize the input global change-detecting process into a localized change-detecting statistic, defined as

$$L_n^{(D)}(k \mid s, e) = \left( \frac{n}{e - s + 1} \right)^{1/2} \left[ D_n(k) - D_n(s - 1) - \frac{k - s + 1}{e - s + 1} \{D_n(e) - D_n(s - 1)\} \right],$$

where $1 \leq s \leq k \leq e \leq n$. By the continuous mapping theorem, $L_n^{(D)}(\cdot \mid s, e)$ converges in distribution to a random variable proportional to a nuisance parameter $\sigma_D$. A self-normalizer is
constructed by a function of the global change-detecting process to cancel $\sigma_D$, defined as

$$V_n^{(D)}(k \mid s, e) = \frac{k - s + 1}{(e - s + 1)^2} \sum_{j=s}^{k} T_n^{(D)}(j \mid s, k) A \otimes 2$$

$$+ \frac{e - k}{(e - s + 1)^2} \sum_{j=k+1}^{e} T_n^{(D)}(j \mid k + 1, e) A \otimes 2,$$

where $A \otimes 2 = AA^\top$ for any matrix $A$. For $1 \leq s \leq k \leq e \leq n$, define

$$T_n^{(D)}(k \mid s, e) = L_n^{(D)}(k \mid s, e) T_n^{(D)}(k \mid s, e) - 1 L_n^{(D)}(k \mid s, e),$$

and set $T_n^{(D)}(k \mid s, e) = 0$ otherwise. By the continuous mapping theorem, it is not difficult to see that $T_n^{(D)}(\cdot \mid \cdot)$ is asymptotically pivotal. Consequently, $T_n^{(D)}(k \mid s, e)$ is a generalized locally self-normalized (LSN) statistic for detecting whether $k$ is a change point over the local window $[s, e]$. To infer whether $k$ is a change point, we aggregate $T_n^{(D)}(k \mid \cdot)$ over all of the symmetric local windows and propose the score function, defined as

$$T_n^{(D)}(k) = \sup_{0 < e < 1/2} T_n^{(D)}(k \mid k - d, k + 1 + d),$$

where $0 < e < 1/2$ is a fixed local tuning parameter, which is similarly used in, for example, Huang, Volgushev, and Shao (2015) and Zhang and Lavitas (2018); see Remark 3.1 for the rationale for considering only the symmetric local windows. In particular, they suggested choosing $e = 0.1$, which is used throughout this article. Essentially, (3.5) compares the local subsamples of length $d + 1$ before and after time $k$, that is, $S_{\text{before}} = \{X_{k-d}, \ldots, X_k\}$ and $S_{\text{after}} = \{X_{k+1}, \ldots, X_{k+1+d}\}$, for each possible width $d$ that is not too small. The statistic $T_n^{(D)}(k)$ records the change corresponding to the most discernible $d$; see Section 3.2 for more discussion on the change-detecting ability of the LSN statistic. Finally, to capture the effects of all potential change points, we aggregate the scores $T_n^{(D)}(k)$ across $k$ and obtain the proposed test statistic,

$$T_n^{(D)} = \frac{1}{n - 2[\epsilon n] - 1} \sum_{k=\lfloor \epsilon n \rfloor + 1}^{n - [\epsilon n] - 1} T_n^{(D)}(k).$$

Our framework constructs multiple-change-point tests through a two-layer aggregation process. The first aggregation across $d$ is to let the data choose the window sizes. We remark that the performance of the moving sum (MOSUM) statistic, which is a special case of the localized CUSUM with a fixed window size, depends critically on the window size (Eichinger and Kirch 2018). To handle window size selection, Cho and Kirch (2022) proposed localized pruning by first taking the union of fixed bandwidth change-point estimators produced by the MOSUM method or other multiscale method over a specific grid of bandwidths, and then pruning down the size of the first-step estimator. Although the procedure may work for point estimation, it is unclear whether it can be extended to construct powerful tests. The proposed framework employs an automatic window selection, based on the strength of discrepancy of $S_{\text{before}}$ and $S_{\text{after}}$ according to the LSN statistic, prior to the estimation stage to avoid additional pruning. Nevertheless, both authors suggested considering multiple bandwidths for robustness.

The purpose of the second layer is to gather all information on all possible change points. Generally, users may choose their preferred aggregation operators, for example, median, trimmed mean, etc, for aggregation. Formally, for any finite set of indices $S \subset Z$ having a size $|S|$, we say that $P$ is an aggregation operator if $P_{i} \in S$ $a_{i}$ maps $\{a_{i} \in \mathbb{R} : i \in S\}$ to a real number. For example, if $P = \max$, $P_{i} \in S$ gives the maximum; and if $P = \text{mean}$, $P_{i} \in S = \sum_{i \in S} a_{i} / |S|$ gives the sample mean. Hence, the most general form of our proposed test statistic is

$$T_n^{(D)} = T_n^{(D)}[P^{(1)}, P^{(2)}]$$

$$= P^{(2)}_{[\epsilon n], ck \in [\epsilon n]} \left[ P^{(1)}_{[\epsilon n] \leq d \leq n} T_n^{(D)}(k \mid k - d, k + 1 + d) \right],$$

where $P^{(1)}$ and $P^{(2)}$ are some aggregation operators. The optimal choices of aggregation operators depend on the number of change points and their respective sizes. By construction, the scores (3.5) remain large over the neighborhoods of the true change points. Intuitively, the average score captures evidence from all change points and their neighborhoods, whereas the maximum score can only capture evidence from a single change point corresponding to the highest score. According to our experience and the simulation results provided in Section C.12 of the supplement, the gain in power is noticeable when $P^{(2)} = \text{mean}$ and all the changes are of the same size. However, the gain diminishes as the portions of smaller changes increase relative to those of larger changes. Nevertheless, $T_n^{(D)}[\text{max, mean}]$ still has greater power in the cases considered. Therefore, we recommend selecting $P^{(2)} = \text{mean}$, as used in the rest of this article.

Our approach is more flexible as it supports a general change-detecting process $D_n(\cdot)$ and general aggregation operators $P^{(1)}$ and $P^{(2)}$. Moreover, the locally self-normalized statistic is a function of the supplied global change-detecting process. Therefore, recursive computation of the locally self-normalized statistics, regardless of the input change-detecting process, is possible for our approach. Briefly, our framework is an automatic procedure for generalizing any single-change-detecting statistic $D_n(\cdot)$ to a multiple-change-detecting statistic $T_n^{(D)}$. We demonstrate this feature through various examples. Section 3.2 presents the theoretical details of the proposed framework under the popular CUSUM process. An application of outlier robust statistic is demonstrated in Section 3.3. Extension to testing change in general parameters is discussed in Section 3.4.

Remark 3.1 (Symmetric windows). Let the score function based on nonsymmetric windows be $T_n^{(D)}(k) = \sup_{[\epsilon n] \leq d \leq n} T_n^{(D)}(k \mid k - d, k + 1 + d)$. The corresponding test statistic is $T_n^{(D)} = \sum_{k=\lfloor \epsilon n \rfloor + 1}^{n - [\epsilon n] - 1} T_n^{(D)}(k) / (n - 2[\epsilon n] - 1)$. According to our experience, the gain in power from nonsymmetric windows is marginal, but the computational cost increases exponentially; see the simulation results provided in Section C.10 of the supplement. Therefore, symmetric windows are recommended to balance cost and power.
3.2. CUSUM Statistic

To improve the power of the CUSUM process under the multiple-change-point setting, we generalize it to a CUSUM LSN statistic according to our proposed framework. In particular, the localized CUSUM statistic can be rewritten as

$$L_n^{(C)}(k | s, e) = \left( \frac{n}{e - s + 1} \right)^{1/2} \left[ C_n(k) - C_n(s - 1) - \frac{k - s + 1}{e - s + 1} \right],$$

where $1 \leq s \leq k \leq e \leq n$. Therefore, the localized CUSUM statistic compares the sample means of data from time $s$ to $k$ and $k + 1$ to $e$. Because (3.7) is not normalized, according to (3.3)–(3.4), the localized self-normalizer and localized CUSUM statistic are respectively

$$V_n^{(C)}(k | s, e) = \frac{k - s + 1}{(e - s + 1)^2} \sum_{j=s}^{k} L_n^{(D)}(j | s, k)^2 + \frac{e - k}{(e - s + 1)^2} \sum_{j=k+1}^{e} L_n^{(D)}(j | k + 1, e)^2,$$

$$T_n^{(C)}(k | s, e) = \frac{L_n^{(C)}(k | s, e)^2}{V_n^{(C)}(k | s, e)},$$

To infer how likely $k$ is a potential change-point location, we consider all symmetric local windows that centered at $k$ and define the score function according to (3.5) as

$$T_n^{(C)}(k) = \sup_{|\epsilon| \leq d \leq n} T_n^{(C)}(k | k - d, k + 1 + d),$$

where $\epsilon = 0.1$. In Figure 1, the proposed score function (3.8) is compared with Shao and Zhang (2010) method. Clearly, $T_n^{(C)}(\cdot)$ achieves the local maxima near all true change-point locations, while $S_n^{(1)}(\cdot)$ does not. Because the $S_n^{(1)}$ self-normalizers are not localized, they include data containing change points and are, therefore, inflated by other changes. Other possible reasons include reduced change-point detection capacity of the global CUSUM because of other changes and nonmonotonic change-point configuration. As a remark, the under-performance of $S_n^{(1)}$ comes with the benefit of smaller computational complexity; for more discussion, see Section 4.1. Finally, we obtain the proposed test statistic by aggregating the scores $T_n^{(C)}(k)$ across $k$ as follows:

$$T_n^{(C)} = \frac{1}{n - 2|\epsilon n| - 1} \sum_{k=|\epsilon n| + 1}^{n} T_n^{(C)}(k),$$

which captures the effects of all potential change points. For $j = 1, \ldots, M$, define $\pi_j = k_j/n$ and $\Delta_j = \mu_{k_j+1} - \mu_{k_j}$ be the $j$th relative change-point time and the corresponding change, respectively. Theorem 3.1 states the limiting distribution and the consistency of $T_n^{(C)}$.

**Theorem 3.1 (Limiting distribution and consistency).** Suppose Assumption 2.1 is satisfied. (i) Under $H_0$, $T_n^{(C)} \overset{\mathcal{D}}{\rightarrow} T$ for any $0 < \epsilon < 1/2$, where

$$T = (1 - 2\epsilon)^{-1} \int_{\epsilon}^{1-\epsilon} \sup_{\delta > \epsilon} \mathbb{L}(t | t - \delta, t + \delta)^2 dt,$$

$$\mathbb{L}(t | \tau_1, \tau_2) = \frac{1}{(\tau_2 - \tau_1)^{1/2}} \left[ \mathbb{B}(t) - \mathbb{B}(\tau_1) - \frac{t - \tau_1}{\tau_2 - \tau_1} \mathbb{B}(\tau_2) - \mathbb{B}(\tau_1) \right],$$

$$\mathbb{V}(t | t - \delta, t + \delta) = \frac{1}{4\delta} \left( \int_{t - \delta}^{t + \delta} \mathbb{L}(\tau | t - \delta, t + \delta)^2 d\tau + \int_{t - \delta}^{t + \delta} \mathbb{L}(\tau | t, t + \delta)^2 d\tau \right).$$

(ii) Under $H_{1,j}$, if there exist $j \in \{1, \ldots, M\}$ and $C > 0$ such that the $j$th change-point time satisfies $\epsilon < \min(\pi_j - \pi_{j-1}, \pi_{j+1} - \pi_j)$ and the corresponding $j$th change magnitude satisfies $|\Delta_j| = |\mu_{k_j+1} - \mu_{k_j}| > C$ as $n \rightarrow \infty$, then $\lim_{n \rightarrow \infty} T_n^{(C)} = \infty$ in probability.

The test statistic $T_n^{(C)}$ has a pivotal limiting distribution under $H_0$, the quantiles of which can be found by simulation; see Table 1. See also Section 4 for more discussion on the critical

| Table 1. Simulated quantiles of $T$, that is, the limiting distribution of $T_n^{(C)}$. |
|------------------|-----|-----|-----|
| $P(T \leq t)$    | 90% | 95% | 99% |
| $t$              | 17.6| 19.6| 23.7|

![Figure 1](image-url) (a) A realization of data (black solid line) with 5 change-points (blue dashed lines) and its mean function (red solid line). (b) Shao and Zhang's (2010) score function $S_n^{(3)}(k)$. (c) The proposed score function $T_n^{(C)}(k)$. 

value adjustment procedure for handling strong serial correlation when the sample size is small. Under $H_{z,1}$, the test is of power 1 asymptotically if the change magnitude is at least a nonzero constant. **Theorem 3.2** concerns a local alternative hypothesis that contains change points with diminishing change magnitudes.

**Theorem 3.2 (Localization power).** Under $H_{z,1}$, parameterize the $j$th change magnitude $|\Delta_j(n)| = |\mu_{k+1} - \mu_k|$ as a function of $n$. If there exist $j \in \{1, \ldots, M\}$ and $0.5 < k \leq 1$ such that the $j$th change point satisfies $\varepsilon < \min(\pi_j - \pi_{j-1}, \pi_{j+1} - \pi_j)$ and $|\Delta_j(n)| \asymp n^{\kappa - 1}$ as $n \to \infty$, then $\lim_{n \to \infty} T_n^{(C)} = \infty$ in probability.

Motivated by Figure 1(c), a natural change-point set estimator $\widehat{\Pi}$ can be intuitively obtained from the points corresponding to the local maxima of the general score functions $T_n^{(D)}(\cdot)$ in (3.5). Specifically, the estimator is defined as

$$
\widehat{\Pi}^{(D)} = \left\{ \frac{k}{n} \in \left\{ \frac{n}{n}, \ldots, \frac{n - |\pi|}{n} \right\} : T_n^{(D)}(\cdot) \right\} = \max_{k - |\pi| < n} \left( \frac{T_n^{(D)}(\cdot)}{\Pi^{(D)}(\cdot)} \right),
$$

where $\Pi$ is a decision threshold and $D$ is the global change-detecting process. In particular, if the CUSUM process is used, that is, $D = C$, **Lemma 3.3** and **Theorem 3.4** show the consistency of the proposed point estimator, $\widehat{\Pi}^{(C)}$.

**Lemma 3.3.** Let $\pi_j$ be the $j$th relative change point, $k_j$ be the $j$th change-point time such that $k_j/n \to \pi_j$ as $n \to \infty$ for $j = 1, \ldots, M$, $k_0 = 0$, and $k_{M+1} = n$. Let $h$ be any non-change-point time such that $|h/n - \pi_j| \to c > 0$ as $n \to \infty$ for all $j$. Define $E(X_i) = \mu + \sum_{j=1}^{M} \Delta_j \mathbb{I}(i/n \geq \pi_j)$. If $\pi_j - \pi_{j-1} > \varepsilon$ and $\Delta_j \asymp n^{\kappa - 1}$ as $n \to \infty$, where $0.5 < \kappa \leq 1$ for all $j$, then for $\varepsilon_{nj} = o(n^{2\kappa - 1})$,

$$
P \left( T_n^{(C)}(\cdot) > \varepsilon_{nj} \right) \to 1 \quad \text{and} \quad P \left( T_n^{(C)}(h) > \varepsilon_{nj} \right) \to 0.
$$

**Theorem 3.4.** Assume that for all $j = 1, \ldots, M$, where $M$ is a finite constant, $\pi_j - \pi_{j-1} > \varepsilon$ and $\Delta_j \asymp n^{\kappa - 1}$ as $n \to \infty$, where $0.5 < \kappa \leq 1$. Let $\widehat{\pi}_1^{(C)} < \widehat{\pi}_2^{(C)} < \cdots < \widehat{\pi}_M^{(C)}$ be the elements of $\widehat{\Pi}^{(C)}$. If $\xi = o(n^{2\kappa - 1})$ where $\kappa = \min_{j=1, \ldots, M} \kappa_j$, then for any small enough $\eta > 0$ such that $\eta < \varepsilon$,

$$
P \left( \min_{f=1, \ldots, \ell} |\pi_j - \widehat{\pi}_j^{(C)}| > \eta \right) \to 0.
$$

By **Lemma 3.3**, the optimal threshold, $\varepsilon$, depends on the unknown change magnitudes. If the sizes of the changes are further assumed to be $O(1)$, then we can set $\varepsilon = cn^{\eta'}$ with $0 < \eta < 1$. Theoretically, it is challenging to derive an optimal choice of $c$ and $\eta$ without a specific cost function that trades off the accuracy and size of $\widehat{\Pi}^{(C)}$. It is possible to select $(c, \eta)$ across a class of change-point models through a large-scale simulation. As point estimation is beyond the scope of this article, we leave it as an open question for further study. From our experience, choosing $\varepsilon = \sqrt{n}$ yields a reasonable result. Because the score functions are computed during the computation of the proposed test statistics $T_n^{(C)}$, the point estimator can be directly computed.

**Remark 3.2 (Improving point estimation).** The change-point estimation may be improved by further integrating other screening algorithms, for example, the screening and ranking algorithm (Niu and Zhang 2012) and the scanning procedure proposed by Yau and Zhao (2016), to perform model selection in order to avoid overestimating the number of change points. Alternatively, our proposed test can be applied to stepwise estimation, for example, binary segmentation (Vostrikova 1981), wild binary segmentation (Fryzlewicz 2014), and SNCP (Zhao, Jiang, and Shao 2021); see Section G of the supplement for information on how to integrate our method with existing screening methods.

### 3.3. Outlier-Robust Statistics

An outlier-robust nonparametric change-point test can be constructed using the Wilcoxon two-sample statistic. The corresponding $D_n(\cdot)$ is

$$
W_n([nt]) = \frac{1}{n^{3/2}} \sum_{i=0}^{[nt] - 1} \sum_{j=0}^{[nt] - 1} \left( \mathbb{I}(X_i \leq X_j) - \frac{1}{2} \right)
$$

$$
= n^{-3/2} \left( \sum_{i=0}^{[nt]} R_i - \frac{[nt]}{n} \sum_{i=1}^{[nt]} R_i \right),
$$

for $t \in [0, 1]$, where $R_i = \sum_{j=0}^{[nt]} \mathbb{I}(X_i \leq X_j)$ is the rank of $X_i$ when there is no tie in $X_1, \ldots, X_n$. It is assumed that the data are weakly dependent and can be represented as a functional of an absolutely regular process. Under some mixing conditions, we have $[W_n([nt])] \Rightarrow \sigma_n \mathbb{E}(\mathbb{I}(t - \mathbb{E}(1)))$, where $\sigma_n^2 = \sum_{k \in \mathbb{Z}} \text{cov}(F(X_0), F(X_1))$ and $F(t)$ is the distribution function of $X_1$ with a bounded density; see Theorem 3.1 in Dehling et al. (2015). From the principles in Section 3.1, we may use (3.14) to construct a self-normalized multiple-change-point Wilcoxon test and eliminate the nuisance parameter $\sigma_n$ by using the test statistic $T_n^{(W)} = T_n^{(W)}[\text{max, mean}]$; see the corollary below.

**Corollary 3.5 (Limiting distribution of $T_n^{(W)}$).** Under the regularity conditions in Theorem 3.1 of Dehling et al. (2015) and $H_0$, we have $T_n^{(W)} \Rightarrow T$.

The Hodges–Lehmann statistic is another popular alternative to the CUSUM statistic in (2.3). Its global change-detecting process is

$$
H_n([nt]) = n^{-3/2} \left( [nt] (n - [nt]) \right) \quad \text{median} \left\{ (X_i - X_j) : 1 \leq i \leq [nt] < j \leq n \right\},
$$

for $t \in [0, 1]$, where median$(S)$ denotes the sample median of a set $S$. It outperforms the CUSUM test under skewed and heavy-tailed distributions. Under the regularity conditions in Dehling, Fried, and Wendler (2020), we have $H_n([nt]) \Rightarrow \sigma_n \mathbb{E}(\mathbb{I}(t - \mathbb{E}(1)))$, where $\mu(x) = \int_0^x f(y)(s + y) \, dy$, $\sigma_n^2 = \sum_{k \in \mathbb{Z}} \text{cov}(F(X_0), F(X_1))$, $d$ is the density of $X_1$, and $F$ is the distribution function of $X_1$. Similarly, we can apply the Hodges–Lehmann statistic to test for the existence of multiple change points. The corresponding test statistic is $T_n^{(H)} = T_n^{(H)}[\text{max, mean}]$. Its limiting distribution is presented in Corollary 3.6. To our knowledge, there is no existing self-normalized
tests use the Hodges–Lehmann statistic. It is included to demonstrate the generality of the proposed self-normalization framework and to detect change points in heavy-tailed data.

Corollary 3.6 (Limiting distribution of $T_n^{(H)}$). Under the regularity conditions in Theorem 1 of Dehling, Fried, and Wendler (2020) and $H_0$, we have $T_n^{(H)} \overset{d}{\rightarrow} T$.

3.4. Extension for Testing Change in General Parameters

Instead of testing changes in mean, one may be interested in other quantities, for example, variances, quantiles and model parameters. Let $\theta_i = P(F_i^{(h)}) \in \mathbb{R}^q$, where $h \in \mathbb{N}$, $P(\cdot)$ is a functional, $F_i^{(h)}$ is the joint distribution function of $Y_i = (X_{i1}, \ldots, X_{ih})'$ for $i = 1, \ldots, n - h + 1$ and $q \in \mathbb{N}$.

For example, for $h = 1$, $\mu_i = \int_{-\infty}^{\infty} x \, df_1^{(1)}(x)$ and $\sigma_i^2 = \int_{-\infty}^{\infty} x^2 \, df_1^{(1)}(x) - (\int_{-\infty}^{\infty} x \, df_1^{(1)}(x))^2$ are the marginal mean and variance at time $i$, respectively. For $h = 2$, the lag-1 autocovariance at time $i$ is $\gamma_i(1) = \int_{-\infty}^{\infty} (x - \mu_i)(x_i - \mu_{i-1}) \, df_2^{(1)}(x_i - x_{i-1})$. The hypotheses in (2.1) and (3.1) are redefined by replacing $\mu_i$'s by $\theta_i$'s. A possible global change-detecting process is

$$G_n(\{nt\}) = \frac{\{nt\}(n - [nt])}{n^{3/2}} \hat{\theta}_{[nt]} - \hat{\theta}_{[nt] + 1,n},$$

for $t \in \{0, 1\}$; see, for example, Shao (2010), Shao and Zhang (2010), and Shao (2015). If $\hat{\theta}_{i,j} = (j - i + 1)^{-1} \sum_{s=i}^{j} y_{st}$, then $G_n(\cdot) = C_n(\cdot)$. Therefore, (3.16) generalizes (2.3) from testing changes in $\mu_i$'s to testing changes in $\theta_i$'s. The final test statistic is $T_n^{(G)} = T_n^{(C)}[\text{max, mean}]$.

The limiting distribution of (3.16) requires standard regularity conditions in handling statistical functions. Based on the sample $Y_1, \ldots, Y_n$, define the empirical distribution of $Y_i$'s by $F_{i,e}(\cdot) = (e - s + 1)^{-1} \sum_{s=i}^{e} y_{st}$, where $\delta_y$ is a point mass at $y \in \mathbb{R}^q$. Assume that $\hat{\theta}_{i,e} = P(F_{i,e})$ is asymptotically linear in the following sense, as $e \rightarrow \infty$:

$$P(F_{i,e}) = P(F_{i,e}^{(h)}) + (e - s + 1)^{-1} \sum_{s=i}^{e} \text{IF}(Y_i \mid P, P_{i,e}) + R_{s,e},$$

where $R_{s,e}$ is a remainder term, $F_{i,e}^{(h)} \equiv F_{i,e}^{(h)} \equiv \cdots \equiv F_{e}^{(h)}$, and $\text{IF}(y \mid P, P_{i,e}) = \lim_{e \rightarrow 0} (P(1 - \epsilon)P_{i,e} + \epsilon \delta_y - P(P_{i,e})/\epsilon)$ is the influence function; see Wasserman (2006).

Corollary 3.7 (Limiting distribution of $T_n^{(G)}$). For $t \in \{0, 1\}$ and $h \in \mathbb{N}$, define $I_n(\{nt\}) = n^{-1/2} \sum_{s=i}^{e} \text{IF}(Y_i \mid P, P_{i,e})$, where $F_{i,e}^{(h)} = F_{i,e}^{(h)} = \cdots = F_{e}^{(h)}$. Then $T_n^{(G)} \overset{d}{\rightarrow} T^{(q)}$, where

$$T^{(q)} = (1 - 2e)^{-1} \int_{-\infty}^{1-e} \sup_{\delta \in \mathbb{R}^q} \text{IF}(y \mid P, P_{i,e})^2 \, dy$$

where $L(\cdot)$ is defined as $L$ in (3.11) but with $B(\cdot)$ being replaced by $\mathbb{B}(\cdot)$; and $V(\cdot)$ is defined as $V$ in (3.11) but with $L(\cdot, \cdot)^2$ being replaced by $L^{(q)}(\cdot, \cdot)^{\otimes 2}$. In particular, $T^{(1)} = T$.

4. Discussion and Implementation

4.1. Comparison with Existing Methods

Shao and Zhang (2010) extended the self-normalized one-change-point test (2.4) to a supervised multiple-change-point test tailored for testing $m$ change-points, where $m$ is pre-specified. The test statistic is

$$S_{n}^{(m)}(k_1 | 1, k_2) = \sup_{(k_1, k_2) \in \Omega_{n}^{(m)}(\epsilon)} \sum_{j=1}^{m} T_n^{(C)}(k_1 | k_{j-1}, k_{j+1}),$$

where $\Omega_{n}^{(m)}(\epsilon) = \{k_1, \ldots, k_m : j \in [1, \ldots, m], |k_{j+1} - k_j| \leq \min(k_j - k_{j+1}, k_{j+1} - k_j)\}$, $k_0 = 1$, and $k_{m+1} = n$. The trimmed region $\Omega_{n}^{(m)}(\epsilon)$ prevents estimates from being computed with too few observations. Later, Zhang and Lavitas (2018) proposed an unsupervised self-normalized multiple-change-point test that bypasses specifying $m$ and is defined as,

$$Z_n = \sup_{(k_1, k_2) \in \Omega_{n}^{(2)}(\epsilon)} T_n^{(C)}(k_2 | k_1, n).$$

Although $S_{n}^{(m)}$, $Z_n$ and $T_n^{(C)}$ use LSN CUSUM statistics, that is, $T_n^{(C)}(k \mid s, e)$, as building blocks, they have different restrictions on the local windows and aggregate the LSN statistics in different ways. The Shao and Zhang (2010) multiple-change-point test has strict control over the $m$ local windows because the boundaries of a window relate to preceding and subsequent windows. If the number of change points is misspecified, some windows may not contain only one change point. If $m > M$, the self-normalizers are not robust to changes. If $m > M$, some degrees of freedom are lost trying to detect change points that do not exist. Both cases may lead to a significant loss of power. Moreover, their computational cost increases exponentially with $m$. The Zhang and Lavitas’s (2018) approach sets the left end of the window to 1 in the forward scan, and the right end to $n$ in the backward scan. Therefore, their approach tends to scan for the first change point $k_1 \in [1, e]$ and the last change point $k_M \in [s, n]$ for some $e$ and $s$, which may lead to a loss of power; see Section 5.2. In contrast, our approach scans for all possible change points because windows can start and end at any time with the same computation complexity; see Section 2 of the supplement. The score function at each time takes $O(n)$ steps to compute and $O(1)$ spaces to store using recursive computation. Because the test statistic is computed as a function of $O(n)$ scores, it has a computational cost of $O(n^2)$ and a memory requirement of $O(n)$. Table 2 summarizes the comparison. As a remark, the self-normalized segmentation algorithms proposed by Jiang, Zhao, and Shao (2021) and Zhao, Jiang, and Shao (2021) are also based on localized CUSUM contrast-type change-detecting statistics. In the testing stage, a nested window
set is introduced to control the computational cost by considering a finite set of windows. In contrast, we develop recursive LSN statistics that consider all symmetric windows with sizes greater than $\epsilon$. Therefore, our method considers all possible windows with size greater than $|\epsilon n|$ and possibly increases power.

### 4.2. Finite-n Adjusted Critical Values

In the time series context, the accuracy of the invariance principle, that is, Assumption 2.1, may deteriorate when the sample size is small (Kiefer and Vogelsang 2005). Therefore, the asymptotic theory in, for example, Theorem 3.1, may not apply when serial dependency is strong and the sample size is small. It may lead to severe size distortion, observed in both existing self-normalized and non-self-normalized methods, including the proposed tests; see Section 5 and Section C.13 of the supplement for simulation evidences. For the multiple-change-point problem (3.1) in mean, a finite-sample adjusted critical value procedure based on the strength of serial dependency is proposed to mitigate this problem.

We propose to compute a critical value $c_\alpha(n, \rho)$ by matching the autocorrelation function (ACF) at lag one $\rho$ and $n$ for various specified levels of significance $\alpha \in (0, 1)$. The values of $c_\alpha(n, \rho)$ are tabulated in Section D of the supplement for $\alpha \in \{0.01, 0.05, 0.1\}$, $n \in \{100, 200, \ldots, 1000, 2000, \ldots, 10000\}$, and $\rho \in \{0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4, \pm 0.9\}$. The testing procedure is outlined as follows.

1. Compute the sample lag-1 ACF $\tilde{\rho}$ of $\{X_{i+b_n} - X_i\}_{i=1}^{n-b_n}$, where $b_n = \lfloor n^{1/3} \rfloor$ is the differencing parameter.
2. Obtain the critical value $c_\alpha(n, \tilde{\rho})$ by matching $n$ and $\rho$. Use linear interpolation if $n$ and/or $\tilde{\rho}$ do not lie in the grids.
3. Reject the null if $T_n^{(N)} > c_\alpha(n, \tilde{\rho})$, where $N \in \{C, W, H, G\}$.

The adjustment procedure borrows from the idea of bootstrap-based critical value simulation procedures; see for example, Zhou (2013) and Pešta and Wendler (2020). The limiting distribution is pivotal by Theorem 3.1 and Corollaries 3.5–3.7. Therefore, the critical value simulation procedure can be performed prior to the computation of the test statistic. Because $\rho$ depends on the differenced data with a suitable lag, the change points (if any) have negligible effect on $\tilde{\rho}$.

The consistency of $\tilde{\rho}$ is developed on the following framework: Let $X_i = \mu_i + Z_i$, where $\mu_i$'s are deterministic and $Z_i$'s are zero-mean stationary noises. Define $Z_i = g(\ldots, \epsilon_{i-1}, \epsilon_i)$, where $\epsilon_i$'s are independent and identically distributed (i.i.d) random variables and $g$ is some measurable function. Let $\epsilon_i^*$ be an i.i.d copy of $\epsilon_i$ and $Z_i^* = g(\ldots, \epsilon_{i-1}^*, \epsilon_i^*, \ldots, \epsilon_i)$. For $p > 1$ and $\|Z_i\|_p < \infty$, the physical dependence measure (Wu 2011) is defined as $\lambda_p(i) = \|Z_i - Z_i^*\|_p$, where $\|\cdot\|_p = (\mathbb{E} | \cdot |^p)^{1/p}$.

**Theorem 4.1.** Assume that $\mathbb{E}(Z_i^4) < \infty$ and $\sum_{i=1}^{\infty} \lambda_4(i) < \infty$. Define $\Delta_n = \sup_{1 \leq j \leq M} \mathbb{E}|J_{i,j}|$. Let $b_n$ be an $\mathbb{N}$-valued sequence. If $b_n/n + 1/b_n + (b_n/M)\Delta_n^2 \to 0$ as $n \to \infty$, then $\tilde{\rho} \to \rho$.

Although it is possible to match the ACF at a higher lag, the AR(1) model is considered to balance the accuracy and the computational cost; see Section C.13 of the supplement. The sensitivity analysis of the differencing parameter $b_n$ in Section C.14 of the supplement also verifies that the choice has a minimal effect on the performance of the tests. The simulation result shows that the tests have a remarkably accurate size even when the true model is not the AR(1) model; see Section 5 and Section C of the supplement. It indicates that the selected AR(1) model has certain explanatory power. If another approximation model is more appropriate for a specific application, then the AR(1) model should be replaced. Users may choose between the standard testing procedure using Table 1 or the proposed adjustment, if it is appropriate.

## 5. Simulation Experiments

### 5.1. Setting and Overview

Throughout Section 5, the experiments are designed as follows. The time series is generated from a signal-plus-noise model: $X_i = \mu_i + Z_i$. The values of $\mu_i$'s will be specified in Sections 5.2 and 5.3. The zero-mean noises $Z_i$'s are simulated from a stationary bilinear autoregressive (BAR) model:

$$Z_i = (\sigma + \vartheta \epsilon_i)Z_{i-1} + \epsilon_i, \quad i = 1, \ldots, n$$  \hspace{2cm} (5.1)

where $\epsilon_i$'s are independent standard normal random variables, and $\sigma, \vartheta \in \mathbb{R}$ such that $\sigma^2 + \vartheta^2 < 1$. The BAR model is a
Table 3. Null rejection rates \( \hat{\alpha} \) at nominal size \( \alpha = 5\% \) under BAR model and mean function (5.2).

| \( \theta \) | \( \sigma \) | \( n = 200 \) | \( n = 400 \) |
|---------|---------|--------|--------|
|         | \( R_0 \) | \( R_0 \) | \( R_0 \) |
| 0.8     | 0.5     | 14.4   | 11.3   |
| 0.3     | 13.5    | 5.6    | 5.3    |
| 0.0     | 15.4    | 4.7    | 3.9    |
| -0.3    | 23.9    | 4.6    | 3.9    |
| -0.5    | 31.2    | 3.9    | 3.1    |
| 0.5     | 0.8     | 31.8   | 31.8   |
| 0.3     | 13.7    | 6.5    | 6.3    |
| 0.0     | 15.4    | 4.7    | 3.9    |
| -0.3    | 23.9    | 4.6    | 3.9    |
| -0.5    | 31.2    | 3.9    | 3.1    |

NOTE: The sample RMSE\(_n\) of \( \hat{\alpha} \) over all cases of \( \theta \) and \( \sigma \) is computed for each test and each \( n \). The values of \( \hat{\alpha} \) and RMSE are expressed in percentage. The effectiveness of the test in controlling the Type-I error increases with a decreasing RMSE\(_n\).

class of nonlinear time series models that have been extensively studied in the literature; see, for example, Granger and Anderson (1978), Rao (1981), and Rao and Gabr (1984). The bilinear time series model is widely used in the fields of control theory (Bruni, DiPillo, and Koch 1974), econometrics (Weiss 1986), and seismology (Dargahi-Noubary, Laycock, and Rao 1978), which the phenomenon of occasional sharp spikes occurs in sample paths. Similar findings are obtained under other noise models, including the autoregressive-moving-average model, threshold AR model, and absolute value nonlinear AR model. Because of space constraints, these results are presented in the supplement. The critical value is chosen according to the adjustment procedure described in Section 4.2.

5.2. Size and Power

In this section, we examine the size and power of different tests when there exist various numbers of change points. Following the suggestion of Huang, Volgoshev, and Shao (2015) and Zhang and Lavitas (2018), we choose \( \epsilon = 0.1 \) in \( s_{n}^{(1)}, s_{n}^{(2)}, s_{n}^{(3)}, Z_{n} \) and in our proposed tests for a fair comparison. Suppose that the change points are evenly spaced and the mean change directions are alternating (increasing or decreasing):

\[
\mu_{i} = \Delta \sum_{j=1}^{M} (-1)^{j+1} 1_{\{|i-n_{j}/(M+1)|\}},
\]

where \( M \in \{1, \ldots, 6\} \) denotes the number of change points, and \( \Delta \in \mathbb{R} \) controls the magnitude of the mean changes. Then, \( \mu_{i} \equiv 0 \) under \( H_{0} \).

All tests are computed at the nominal size \( \alpha = 5\% \). The null rejection rates \( \hat{\alpha} \), for sample sizes \( n \in \{200, 400\} \), are presented in Table 3. To summarize the result, we further report the sample root mean squared error (RMSE) of \( \hat{\alpha} \) over all cases of \( \theta \) and \( \sigma \) and \( \sigma \) = \( -0.8, -0.5, 0.5, 0.8 \) and \( \sigma \) in (5.2). The self-normalized tests generally have a more accurate size than the non-self-normalized test KS\(_n\). This finding is consistent with that of Shao (2015). The self-normalized approach is a special case of the fixed bandwidth approach with the largest possible bandwidth and thus achieves the smallest size distortion, as observed by Kiefer and Vogelsang (2002). In comparison, \( T_{n}^{(4)} \) and \( T_{n}^{(H)} \) control the size most accurately among the self-normalized tests. In particular, our proposed tests have the least severe under-size problem, observed in existing self-normalized tests (i.e., \( s_{n}^{(1)}, s_{n}^{(2)}, s_{n}^{(3)}, Z_{n} \)) when \( \sigma < 0 \). Moreover, the test proposed by Shao and Zhang (2010), \( S_{n}^{(m)} \), suffers from an increasing size distortion as \( m \) increases. This finding is consistent with that of Zhang and Lavitas (2018). The fact that \( T_{n}^{(W)} \) and \( T_{n}^{(H)} \) outperform \( T_{n}^{(C)} \) can be attributed to the outlier robustness of rank and order statistics. This advantage is particularly noticeable in bilinear time series, as this model is well known for producing sudden high amplitude oscillations that mimic structures in, for example, explosion and earthquake data in seismology; see Section 5.2 of Rao and Gabr (1984).
However, in the more standard ARMA models, $T_n^{(C)}$ performs as well as $T_n^{(W)}$ and $T_n^{(H)}$; see the supplement for the detailed results.

The power curve against $\Delta$ is computed at 5% nominal size using $210^6$ replications with $n = 200$. Figure 2 displays the size-adjusted power when $\sigma = \vartheta = 0.5$. The (unadjusted) power is presented in the supplement. The results with various values of $\sigma$ and $\vartheta$ are similar, thus, they are presented in the supplement. Generally, $T_n^{(W)}$ and $T_n^{(H)}$ have the highest power. They outperform $T_n^{(C)}$ because of their robustness against outliers. When only the CUSUM-type tests are considered, the non-self-normalized test $K_{S_n}$ performs better than self-normalized tests when $M = 1$. However, when $M > 1$, $K_{S_n}$ significantly underperforms because it is tailor-made for the AMOC alternative (2.2). Although $S_n^{(1)}$ has the highest power among the self-normalized CUSUM tests when $M = 1$, it suffers from the notorious non-monotonic power problem when $M > 1$ because its self-normalizer is not robust to multiple change points. Thus, $S_n^{(1)}$ is not a consistent test when $M > 1$.

Surprisingly, our proposed test $T_n^{(C)}$ outperforms the tests of Shao and Zhang (2010), $S_n^{(m)}$, even when $M$ is well-specified. This finding demonstrates that it is not advantageous to know $M$ because structural change can be identified by observing the data around it without looking at the entire segmented series. Moreover, as discussed in Section 4.1, $S_n^{(m)}$ defines the local windows restrictively. It accumulates errors if the boundaries of the local windows differ from the actual change points. It is also interesting to see that, compared with $S_n^{(1)}$, $S_n^{(2)}$, and $S_n^{(3)}$ are less sensitive to misspecification of $M$; however, they are still less powerful than our proposed tests.

### 5.3. Ability to Capture Effects of All Changes

The proposed tests are robust against multiple change points compared with existing methods. Consider the three-change-point setting in which the magnitude of changes at the first and the last change points is only half that of the second change point. Specifically, define the mean function for case 1 as (5.2) with $M = 3$ and for case 2 as

$$
\mu_i = \Delta \left\{ \frac{1}{2} \cdot \mathbb{I}_{i > \lfloor n/4 \rfloor} + \frac{1}{2} \cdot \mathbb{I}_{i > \lceil 3n/4 \rceil} \right\}.
$$

From Figure 3, $Z_n$ and $S_n^{(2)}$ lose approximately half of their power in Case 2 compared with that in Case 1. In contrast, the power of our tests only decreases by $1/3$ when $\Delta \leq 1$, while it remains roughly the same when $\Delta \geq 2$. Therefore, our tests are more powerful and can capture a larger class of structural changes than existing methods. In contrast, $S_n^{(2)}$ and $Z_n$ tend to consider the first and/or the last change points.

### 5.4. Change in General Parameters

In many applications, trends may exist in the data. Therefore, we consider the multiple-change-point problem in the mean and median trend models. Specifically, we consider the regression model: $\Phi(X_i) = \alpha_i + \beta_i(i/n)$ with $\Phi(\cdot) = \mathbb{E}(\cdot)$ and $\Phi(\cdot) = \text{median}(\cdot)$ under the following multiple-change-point model:

$$
\alpha_i = \Delta_{\alpha,0} + \sum_{j=1}^{M} \Delta_{\alpha,j} \mathbb{I}(i/n > \pi_j) \quad \text{and}
$$

$$
\beta_i = \Delta_{\beta,0} + \sum_{j=1}^{M} \Delta_{\beta,j} \mathbb{I}(i/n > \pi_j)
$$

for $i = 1, \ldots, n$, where $M \in \mathbb{N}$, $0 < \pi_1 < \cdots < \pi_M < 1$, and $\Delta_{\alpha,0}, \ldots, \Delta_{\alpha,M}, \Delta_{\beta,0}, \ldots, \Delta_{\beta,M} \in \mathbb{R}$. Note that $\Phi(\cdot) = \mathbb{E}(\cdot)$ and $\Phi(\cdot) = \text{median}(\cdot)$ refer to the mean trend and median trend models, respectively. Our goal is to test $H_0$ against $H_{\leq 1}$ with $\mu_i$’s replaced by $\beta_i$’s.
Table 4. Empirical rejection rates (%) of the mean and median trend tests, \(T_{\mu}^G(\text{mean})\) and \(T_{\mu}^G(\text{med})\), respectively, at 5% nominal size under the BAR model.

| \(n\) | \(\Delta\) | \(M\) | \(\theta = 0.5\) | \(\sigma = 0.5\) | \(\theta = -0.5\) | \(\sigma = -0.5\) |
|-------|-----------|-------|----------------|----------------|----------------|----------------|
| 1     | 0         | 0.0   | 3.0           | 3.0            | 0.0           | 0.0            |
|       | 33       | 0.33  | 6.7           | 5.5            | 0.4           | 0.3            |
|       | 67       | 0.67  | 7.1           | 5.1            | 0.4           | 0.3            |
|       | 1        | 1.0   | 10.6          | 12.3           | 0.4           | 0.3            |
|       | 2        | 2.0   | 28.3          | 51.9           | 5.4           | 4.7            |
|       | 3        | 3.0   | 51.8          | 82.1           | 14.4          | 32.6           |

To handle changes in general parameters, the global change-detecting process \(G_n(\cdot)\) in (3.16) is used. Let \(\hat{\theta}_{SC}\) be an estimator of the common value of \(\beta_0, \ldots, \beta_c\) based on observations \(X_0, \ldots, X_c\) under \(H_0\). The estimators \(\hat{\theta}_{SC}\) are obtained using ordinary least-squares regression under the mean trend model; while they are obtained using quantile regression under the median trend model and computed by the R package “quantreg”. Let \(T_{\mu}^G(\text{mean})\) and \(T_{\mu}^G(\text{med})\) be the mean trend version and median trend version of the test statistics \(T_n^G(\cdot)\) in Section 3.4, respectively. In our simulation, we generate the data as

\[
X_i = \Delta \sum_{j=0}^M \left\{ \left( \frac{1}{\eta(M+1)} \right) j \right\} 1_{\{i/nj/(M+1)\} + Z_i,}
\]

where \(Z_i\)’s are simulated from the BAR model (5.1), \(M\) \(\in\{1, \ldots, 6\}\) denotes the number of change points, and \(\Delta \in \mathbb{R}\) controls the magnitude of the changes. Briefly, the model implies an uptrend in mean and median before every odd index of change points until level \(\Delta\). After reaching \(\Delta\) at odd change points, the trend shifts downward and decreases to 0 at the next change point. The empirical rejection rates are reported in Table 4. From the results, \(T_{\mu}^G(\text{mean})\) suffers less severe under-size distortion than \(T_{\mu}^G(\text{med})\). Both tests have decent powers in the multiple-change-point setting and significantly improve from \(n = 200\) to \(n = 400\). For reference, the size-adjusted power is also reported in Section C.11 of the supplement.

6. Shanghai-Hong Kong Stock Connect turnover

We use our proposed method to perform change-point analysis of the Shanghai-Hong Kong Stock Connect Southbound Turnover index (Bloomberg code: AHHXHT index) from March 23, 2017 to March 22, 2022 (\(n = 1232\)). The data are retrieved from Bloomberg. The Stock Connect is a channel that allows mainland China and Hong Kong investors to access other stock markets jointly. The southbound line enables investors in mainland China to invest in the Hong Kong stock market. Studies have shown that the Stock Connect improves mutual market liquidity and capital market liberalization (Bai and Chow 2017; Huo and Ahmed 2017; Xu et al. 2020). Therefore, it is of practical interest, to determine whether change points exist in the Stock Connect Southbound daily turnover.

We investigate whether changes in trend exist in the log daily turnovers. Under the mean and median trend models, we consider the tests statistics \(T_{\mu}^G(\text{mean})\) and \(T_{\mu}^G(\text{med})\), specified in Section 5.4. The resulting test statistics are 47.46 and 40.64. Both \(p\)-values are less than 1%. The trend change-point location estimates are indicated by red vertical lines in Figure 4. The mean and median trend analyses agree on four estimated change points, estimated on March 7, 2018, January 24, 2019, October 2, 2019, and January 12, 2021. The median trend analysis detects an additional change point on May 25, 2020. The first change point is likely to be the China–United States trade war. After the trade war began in January 2018, the Stock Connect turnover
displayed a downtrend. An uptrend is detected after 2019 until the COVID-19 events, while it stops after the beginning of 2021.

7. Conclusion

Our method improves existing change-point tests and has several advantages: (i) it has high power and size accuracy, (ii) it does not require specification of the number of change points nor consistent estimation of nuisance parameters, (iii) a consistent estimate of change points can be naturally produced, (iv) general change-detecting statistics, for example, rank and order statistics, can be used to enhance robustness, (v) it can test a change in general parameter of interest, and (vi) it applies to a wide range of time series models. Table 2 summarizes the properties of the proposed tests. Moreover, our proposed framework is driven by intuitive principles. If a single-change-detecting statistic is provided, our framework can generalize it to a multiple-change-detecting statistic. We anticipate that future works will apply our framework to non-time-series data, for example, spatial and spatial–temporal data.

8. Proofs of Theorems

Proof of Theorems 3.1 and 3.2. (i) Under $H_0$ and by the continuous mapping theorem, Assumption 2.1 implies that $\{\mathcal{C}_n([nt]) : t \in [0,1]\} \Rightarrow \{\sigma(\mathcal{B}(t) - t\mathcal{B}(1)) : t \in [0,1]\}$. Note that $T_n^{(C)}$ is a composite function of $\mathcal{C}_n(\cdot)$ through (3.2), (3.3), (3.4), (3.5), and (3.6), each of which is a continuous and measurable map. By the continuous mapping theorem, we obtain $T_n^{(C)} \xrightarrow{d} T$ in (3.10).

The limiting distribution $T$ is well-defined because $V_n^{(C)}(k | k-d, k+1 + d)$ converges to a nonnegative and nondegenerate distribution for any $|\lfloor en \rfloor| \leq d \leq n$ and $|\lfloor en \rfloor| + 1 \leq k \leq n - |\lfloor en \rfloor| - 1$, if $\epsilon > 0$.

(ii) For consistency, we consider the $j$th relative change-point time $\tau_j$ such that the corresponding change magnitude satisfies $|\Delta_j| \asymp n^{\kappa - 1}$, where $0.5 < \kappa \leq 1$. Let $\Delta_s = \Delta_j$ and $k_s = \lfloor n\tau_j \rfloor$. Under this assumption, there exist $c \in \mathbb{R}^+$ and $3/2 - \kappa \leq \gamma < 1$ such that $c + \lfloor cn^{\gamma} \rfloor / n < \min(\tau_j - \tau_{j-1}, \tau_{j+1} - \tau_j)$ is satisfied for a large enough $n$. Therefore, there is only one change point in the interval $k_s \pm \lfloor cn^{\gamma} \rfloor$. It suffices to consider $d = \lfloor cn \rfloor$. For all $k$ such that $|k - k_s| \leq \lfloor bn^{\gamma} \rfloor$ where $0 < b < c$, we can decompose

$$L_n^{(C)}(k | k - d, k + d + 1) = \frac{d + 1}{8} (\bar{X}_{(k-d):k} - \bar{X}_{k+1} / (k+d+1))^2 = \frac{d + 1}{8} \left( \left( \frac{\bar{Z}_{(k-d):k} - \bar{Z}_{k+1} / (k+d+1)}{\sqrt{d+1}} \right) - \left\{ \frac{|k_s - k|}{d+1} \right\} \Delta_s \right)^2 \geq \left\{ O_p(1) + \Xi_{1,n} \right\}^2,$$

where $|\Xi_{1,n}| \asymp n^{\kappa - 1/2}$. Equation (8.1) holds uniformly in $k$ for $|k - k_s| \leq \lfloor bn^{\gamma} \rfloor$. Next, for the self-normalizer $V_n^{(C)}(k | k-d, k+1 + d)$, we consider two cases: $k \geq k_s$ and $k < k_s$. If $k \geq k_s$, there is only one change point in the interval $[k-d, k]$. Following similar calculations as in (8.1), we have $V_n^{(C)}(j | k-d, k)^2 \leq L_n^{(C)}(k_s | k-d, k) / n$ for all $j = k-d, \ldots, k$, if $n$ is large enough. Also, because there is no change point in the interval $[k+1, k + 1 + d]$, $V_n^{(C)}(j | k+1, k+1 + d)^2 = O_p(1)$ for all $j = k+1, \ldots, k+1 + d$. So,

$$V_n^{(C)}(k | k-d, k+1 + d) = \frac{1}{4(d+1)} \left\{ \sum_{j=k-d}^{k+1} L_n^{(C)}(j | k-d, d) + \sum_{j=k+1}^{k+1+d} L_n^{(C)}(j | k+1, k+1 + d) \right\} \geq \frac{1}{4} V_n^{(C)}(k_s | k-d, k)^2 + O_p(1) = \frac{(d+1 - \lfloor bn^{\gamma} \rfloor)^2 / (2d+1)^2}{2(2d+1)^3} \left( \bar{Z}_{(k-d):k_s} - \bar{Z}_{k_s} \right)^2 + O_p(1) \leq \left\{ O_p(1) + \Xi_{2,n} \right\}^2,$$

where $|\Xi_{2,n}| \asymp n^{\gamma + \kappa - 3/2}$. For the case in which $k < k_s$, the analysis is similar and $V_n^{(C)}$ are of the same order. Therefore, there exists a constant $c' \neq 0$ such that

$$T_n^{(C)}(k) \geq \left\{ O_p(1) + \Xi_{1,n} \right\}^2 \approx \left\{ O_p(1) + \Xi_{2,n} \right\}^2 = \left\{ O_p(1) + c' n^{\gamma - 1} \right\}^2.$$

Figure 4. Estimated (a) mean trend and (b) median trend change points are indicated by red vertical lines. The orange lines indicate the fitted regression lines within each region separated by the estimated change points. Data are retrieved from Bloomberg.
for all \( k = k_a - \lfloor bn^\gamma \rfloor, \ldots, k_a + \lfloor bn^\gamma \rfloor \). Because \( T^{(C)}_n(k) \geq 0 \) for all \( k \), we have

\[
T^{(C)}_n \geq n^{-1} \sum_{k=k_a - \lfloor bn^\gamma \rfloor}^{k_a + \lfloor bn^\gamma \rfloor} T^{(C)}_n(k) \\
\geq \frac{2\lfloor bn^\gamma \rfloor + 1}{n} \left( \epsilon_0 + c' \sqrt{n-1} \right)^2, \\
= c' \sqrt{n-1} \left( \epsilon_0 + 1 \right)^2,
\]

for a large enough \( n \), where \( c' > 0 \) is a constant. Consequently, \( T^{(C)}_n \rightarrow \infty \) in probability as \( n \rightarrow \infty \) because \( Y < 1 \). \( \square \)

**Proof of Lemma 3.3.** By the assumption stated in the lemma, there exists an \( \epsilon \) such that \( \tau_i \) is the unique relative change-point time in the interval \([\tau_i - \epsilon, \tau_i + \epsilon]\). Let \( d = \lfloor \alpha \rfloor \) and \( k_j = \lfloor \alpha \tau_i \rfloor \). Let \( Z_{j,\alpha} = (e - s + 1)^{-1} \sum_{i=j}^{\alpha} Z_i \), where \( Z_i = X_i - E(X_i) \). Then,

\[
Z^{(C)}_n(k_j | k_j - d, k_j + 1 + d) = \frac{1}{8} \left\{ \sqrt{d + 1}(Z_{(k_j - d); k_j - d}) - Z_{(k_j + 1); k_j + 1 + d}) - \sqrt{d + 1} \Delta_j \right\}^2
= \epsilon_0 + \epsilon_0 n^{-1/2}(n^{-1})^2 + O(n^{-2\gamma - 1}).
\]

Because there is no change point in the intervals \([k_j - d, k_j]\) and \((k_j + 1, k_j + 1 + d],\) the self-normalizer \( V^{(C)}_n(k_j | k_j - d, k_j + 1 + d) \) is free of \( \Delta_j \) and therefore is of \( O(p) \). Then, let \( \zeta_{n,j} = \gamma n^{2\gamma - 1} \), as \( n \rightarrow \infty \),

\[
\zeta^{-1} n_{n,j} T^{(C)}_n(k_j) \geq \zeta^{-1} n_{n,j} V^{(C)}_n(k_j | k_j - d, k_j + 1 + d) \rightarrow \infty.
\]

Therefore, \( P[T_{n,j}^{(C)}(k_j) > \zeta_{n,j}] \rightarrow 1 \). For any time \( i \), there are \( M_i^- \geq 0 \) change point(s) in the interval \([0, j] \), \( M_i^+ \geq 0 \) change point(s) in the interval \([i + 1, n] \). Let \( k_{i,0} = i \) and

\[
k_{i,a} = \min(k \in \{k_i, \ldots, k_M, n\} : k > k_{i,a-1}),
\]

\[
\text{if } a = 1, 2, \ldots, \\
\max(k \in \{0, k_1, \ldots, k_M, n\} : k < k_{i,a+1}),
\]

\[
\text{if } a = -1, -2, \ldots.
\]

Therefore, \( k_{i,a} \) denotes the time of \( |a| \)th change point after time \( i \) if \( a > 0 \); before time \( i \) if \( a < 0 \). Likewise, let \( \Delta_i \) be the size of change corresponding to \( k_{i,a} \) for \( |a| > 0 \), and \( \Delta_i \) is \( 0 \). Let \( \Delta^{(C)}_i = \sup_{\bf w_\tau = \ldots, \tau_\epsilon} \{ |E(X_{i+1}) - E(X_i)| \} \) and a set of time indices in \( (s, \epsilon) \) that are asymptotically away from all change points be,

\[
\Omega_{s,\epsilon} = \{ i \in [s, \epsilon] \cap \mathbb{Z} : \lim_{n \rightarrow \infty} |i/n - \pi_j| > 0 \text{ for all } j = 1, \ldots, M \},
\]

Consider non-change-point time \( h \in \Omega_{s,\epsilon} \). For all \( j = 1, \ldots, M \), \( d \geq \lfloor \alpha \rfloor \), and time \( i \in \Omega_{(h-d); (h+1+d),} \), we have \( M_i^- + M_i^+ \geq 1 \) and

\[
L_n^{(C)}(i | h - d, h + 1 + d) = \frac{1}{8} \left\{ \sqrt{d + 1}(Z_{(h-d); h-d}) - Z_{(h+1); h+1+d}) - \sqrt{d + 1} \Delta_i \right\}^2
\]

\[
= \frac{1}{4(d + 1)} \left\{ \sum_{j \in \Omega_{(h-d); (h+1+d)}} L_n^{(C)}(j | h - d, h) + \sum_{j \in \Omega_{(h-d); (h+1+d)}} L_n^{(C)}(j | h + 1, h + 1 + d) \right\}^2
\]

\[
= O_p[(\Delta^{(C)}_{h-d}; (h+1+d))]^2.
\]

Using the result, the self-normalizer is at least of the same order because for some \( \epsilon > 0 \),

\[
V_n^{(C)}(h | h - d, h + 1 + d) \geq \frac{1}{4(d + 1)} \left\{ \sum_{j \in \Omega_{(h-d); (h+1+d)}} L_n^{(C)}(j | h - d, h) + \sum_{j \in \Omega_{(h-d); (h+1+d)}} L_n^{(C)}(j | h + 1, h + 1 + d) \right\}^2
\]

\[
\text{Therefore, } T^{(C)}_n(h) = O_p (\epsilon), \text{ implying } P(\sup_{s \in [\pi_j - \epsilon, \pi_j + \epsilon]} |\hat{\pi}_j - \pi_j| > \eta) \rightarrow 0. \square
\]

**Proof of Theorem 3.4.** For any small enough \( \eta > 0 \) such that \( \eta < \epsilon \), let \( A_f := \{ \min_{1 \leq j \leq M} |\hat{\pi}_j - \pi_j| > \eta \} \) be the event that the absolute difference between the \( f \) th true change point and the closest estimator in \( \hat{\Pi}^{(C)} \) is greater than \( \eta \). Let \( \Psi_f(\epsilon, \eta) = [\pi_j - \epsilon, \pi_j + \epsilon] \setminus [\pi_j - \eta, \pi_j + \eta] \). By the definition of \( \hat{\Pi}^{(C)} \) in (3.13) with \( \Delta = \epsilon \), we can represent \( A_f \) as

\[
A_f = \left\{ \sup_{i \in [\pi_j - \eta, \pi_j + \eta]} T^{(C)}_n(|nt|) \leq \sup_{s \in \Psi_f(\epsilon, \eta)} T^{(C)}_n(|ns|) \right\} \cup \left\{ \sup_{i \in [\pi_j - \eta, \pi_j + \eta]} T^{(C)}_n(|nt|) \leq \varepsilon \right\}
\]

\[
\subseteq \left\{ T^{(C)}_n(|\pi_f n|) \leq \sup_{s \in \Psi_f(\epsilon, \eta)} T^{(C)}_n(|ns|) \right\} \cup \left\{ T^{(C)}_n(|\pi_f n|) \leq \varepsilon \right\} .
\]

Using Lemma 3.3, we have, for all \( j = 1, \ldots, M \) as \( n \rightarrow \infty \),

\[
P(A_f) \leq P \left\{ T^{(C)}_n(|\pi_f n|) < \sup_{s \in \Psi_f(\epsilon, \eta)} T^{(C)}_n(|ns|) \right\} + \left\{ T^{(C)}_n(|\pi_f n|) \leq \varepsilon \right\} \rightarrow 0.
\]

Then, \( P(\max_{j=1,\ldots, M} \min_{i=1,\ldots, M} |\pi_j - \pi_j| > \eta) = P(\cup_{j=1}^{M} A_f) \leq \sum_{j=1}^{M} P(A_f) \rightarrow 0. \square
\]

**Supplementary Materials**

The supplementary note contains additional simulation results, finite-\( n \) adjusted critical values, recursive formulas, algorithms, and proofs of corollaries. Extension to long-range dependent time series is discussed. Implementation code is provided in the R library **SNmct**.

**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 article.


Funding

This research was partially supported by grants GRF-14304420, 14306421, and 14307922 provided by the Research Grants Council of HKSAR.

Disclosure Statement

The authors report there are no competing interests to declare.

ORCID

Cheuk Hin Cheng http://orcid.org/0000-0003-2388-2897
Kim Wai Chan http://orcid.org/0000-0003-2022-547X

References

Andrews, D. W. (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” Econometrica, 59, 817–858. [719]

Antoch, J., and Jarušková, D. (2013), “Testing for Multiple Change Points,” Computational Statistics, 28, 2161–2183. [719]

Bai, J., and Perron, P. (1998), “Estimating and Testing Linear Models with Multiple Structural Changes,” Econometrica, 66, 47–78. [719]

Bai, Y., and Chow, D. Y. P. (2017), “Shanghai-Hong Kong Stock Connect: An Analysis of Chinese Partial Stock Market Liberalization Impact on the Local and Foreign Markets,” Journal of International Financial Markets, Institutions and Money, 50, 182–203. [728]

Bauer, P., and Hacid, P. (1980), “An Extension of the MOSUM Technique for Quality Control,” Technometrics, 22, 1–7. [719]

Billingsley, P. (1999), Convergence of Probability Measures. Wiley Series in Probability and Statistics (2nd ed.), New York: Wiley. [720]

Bruni, C., DiPillo, G., and Koch, G. (1974), “Bilinear Systems: An Appealing Class of ‘Nearly Linear’ Systems in Theory and Applications,” IEEE Transactions on Automatic Control, 19, 334–348. [726]

Chan, K. W. (2022a), “Mean-Structure and Autocorrelation Consistent Covariance Matrix Estimation,” Journal of Business & Economic Statistics, 40, 201–215. [720]

(2022b), “Optimal Difference-based Variance Estimators in Time Series: A General Framework,” Annals of Statistics, 50, 1376–1400. [720]

Cho, H., and Kirch, C. (2022), “Two-Stage Data Segmentation Permitting Multiscale Change Points, Heavy Tails and Dependence,” Annals of the Institute of Statistical Mathematics, 74, 653–684. [721]

Chu, C.-S. J., Hornik, K., and Kauf, C.-M. (1995), “MOSUM Tests for Parameter Constancy,” Biometrika, 82, 603–617. [719]

Csörgö, M., and Horváth, L. (1988), “20 Nonparametric Methods for Change-point Problems,” Handbook of Statistics, 7, 403–425. [719]

(1997), Limit Theorems in Change-point Analysis, Wiley Series in Probability and Statistics, New York: John Wiley. [719]

Dargahi-Noubary, G., Laycock, P., and Rao, T. S. (1978), “Non-Linear Stochastic Models for Seismic Events with Applications in Event Identification,” Geophysical Journal International, 55, 655–668. [726]

Dehling, H., Fried, R., Garcia, I., and Wendler, M. (2015), “Change-Point Detection Under Dependence Based on Two-Sample U-Statistics,” in Asymptotic Laws and Methods in Stochastics, eds. D. Dawson, R. Kulik, M. Ould Haye, B. Szyszkowicz and Y. Zhao), pp. 195–220. New York: Springer. [723]

Dehling, H., Fried, R., and Wendler, M. (2020), “A Robust Method for Shift Detection in Time Series,” Biometrika, 107, 647–660. [723, 724]

Eichinger, B., and Kirch, C. (2018), “A MOSUM Procedure for the Estimation of Multiple Random Change Points,” Bernoulli, 24, 526–564. [719,721]

Fryzlewicz, P. (2014), “Wild Binary Segmentation for Multiple Change-Point Detection,” Annals of Statistics, 42, 2243–2281. [719,723]

Granger, C. W. J., and Anderson, A. P. (1978), An Introduction to Bilinear Time Series Models, Gottingen: Vandenhoek & Ruprecht. [726]

Herrndorf, N. (1984), “A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables,” Annals of Probability, 12, 141–153. [720]

Huang, Y., Volgushev, S., and Shao, X. (2015), “On Self-Normalization for Censored Dependent Data,” Journal of Time Series Analysis, 36, 109–124. [721,726]

Huo, R., and Ahmed, A. D. (2017), “Return and Volatility Spillovers Effects: Evaluating the Impact of Shanghai-Hong Kong Stock Connect,” Economic Modelling, 61, 260–272. [728]

Jiang, F., Zhao, Z., and Shao, X. (2021), “Modelling the Covid-19 Infection Trajectory: A Piecewise Linear Quanlitative Trend Model,” Journal of the Royal Statistical Society, Series B, 84, 1589–1607. [724]

Kiefer, N. M., and Vogelsang, T. J. (2002), “Heteroskedasticity-Autocorrelation Robust Standard Errors using the Bartlett Kernel Without Truncation,” Econometrica, 70, 2093–2095. [726]

(2005), “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” Econometric Theory, 21, 1130–1164. [719,725]

Lobato, I. N. (2001), “Testing that a Dependent Process is Uncorrelated,” Journal of the American Statistical Association, 96, 1066–1076. [719]

Liu, Y. S., and Zhang, H. (2012), “The Screening and Ranking Algorithm to Detect DNA Copy Number Variations,” The Annals of Applied Statistics, 6, 1306–1326. [723]

Peña, M., and Wendler, M. (2020), “Nuisance-Parameter-Free Changepoint Detection in Non-Stationary Series,” Test, 29, 379–408. [725]

Rao, T. S. (1981), “On the Theory of Bilinear Time Series Models,” Journal of the Royal Statistical Society, Series B, 43, 244–255. [726]

Rao, T. S., and Gabr, M. M. (1984), An Introduction to Bispacial Analysis and Bilinear Time Series Models (Vol. 24, 1st ed.), New York: Springer-Verlag. [726]

Shao, X. (2010), “A Self-Normalized Approach to Confidence Interval Construction in Time Series,” Journal of the Royal Statistical Society, Series B, 72, 343–366. [724]

(2015), “Self-Normalization for Time Series: A Review of Recent Developments,” Journal of the American Statistical Association, 110, 1797–1817. [724,726]

Shao, X., and Zhang, X. (2010), “Testing for Change Points in Time Series,” Journal of the American Statistical Association, 105, 1228–1240. [719,720,722,724,726,727]

Vogelsang, T. J. (1999), “Sources of Nonmonotonic Power When Testing for a Shift in Mean of a Dynamic Time Series,” Journal of Econometrics, 88, 283–299. [720]

Vostrikova, L. Y. (1981), “Detection Disorder in Multidimensional Random Processes,” Soviet Mathematics - Doklady, 259, 55–59. [719,723]

Wasserman, L. (2006), All of Nonparametric Statistics (1st ed.), New York: Springer-Verlag. [724]

Weiss, A. A. (1986), “Arch and Bilinear Time Series Models: Comparison and Combination,” Journal of Business & Economic Statistics, 4, 59–70. [726]

Wu, W. B. (2007), “Strong Invariance Principles for Dependent Random Variables,” Annals of Probability, 35, 2294–2320. [720]

(2011), “Asymptotic Theory for Stationary Processes,” Statistics and Its Interface, 4, 207–226. [725]

Xu, K., Zheng, X., Pan, D., Xing, L., and Zhang, X. (2020), “Stock Market Openness and Market Quality: Evidence from the Shanghai-Hong Kong Stock Connect Program,” Journal of Financial Research, 43, 373–406. [728]

Yau, C. Y., and Zhao, Z. (2016), “Inference for Multiple Change Points in Time Series via Likelihood Ratio Scan Statistics,” Journal of the Royal Statistical Society, Series B, 87, 43–73. [726]

Zhou, Z. (2013), “Heteroscedasticity and Autocorrelation Robust Structural Change Detection,” Journal of the American Statistical Association, 108, 726–740. [725]