Location-Adaptive Change-Point Testing for Time Series

Linlin Dai and Rui She∗
Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China.

Abstract

We propose a location-adaptive self-normalization (SN) based test for change points in time series. The SN technique has been extensively used in change-point detection for its capability to avoid direct estimation of nuisance parameters. However, we find that the power of the SN-based test is susceptible to the location of the break and may suffer from a severe power loss, especially when the change occurs at the early or late stage of the sequence. This phenomenon is essentially caused by the unbalance of the data used before and after the change point when one is building a test statistic based on the cumulative sum (CUSUM) process. Hence, we consider leaving out the samples far away from the potential locations of change points and propose an optimal data selection scheme. Based on this scheme, a new SN-based test statistic adaptive to the locations of breaks is established. The new test can significantly improve the power of the existing SN-based tests while maintaining a satisfactory size. It is a unified treatment that can be readily extended to tests for general quantities of interest, such as the median and the model parameters. The derived optimal subsample selection strategy is not specific to the SN-based tests but is applicable to any method that relies on the CUSUM process, which may provide new insights in the area for future research.

Keywords: Contrast-based method; CUSUM process; Power; Self-normalization

*Correspondence: rshe@swufe.edu.cn
1 Introduction

The assessment of structural stability for data of interest is of great significance and has received considerable attention in application areas such as climate change detection, speech and image analysis, and human activity analysis; see a comprehensive survey by [Aminikhanghahi and Cook, 2017]. From a statistical point of view, it is philosophically natural to phrase the problem of testing for structural breaks in a framework of hypothesis testing with the null hypothesis stating structural stability versus the alternative that there exists one or multiple changes in the structure. Based on this common idea, a great number of approaches were developed for detecting breaks for a sequence of independent and identically distributed (iid) random variables [Csörgő and Horváth, 1997], whereas for time series data, it becomes indispensable to incorporate the temporal dependence into the testing procedure. To this end, one may explicitly assume certain parametric model structures and the problem of change-point detection is converted to detect breaks in the associated model parameters; see literature on testing for changes in the polynomial regression [Aue et al., 2008], partially linear models [Su and White, 2010], general linear regression [Michael, Colin, and Robert, 2016], AR models [Davis, Huang, and Yao, 1995], GARCH models [Berkes et al., 2004], and spatial-temporal models [Gromenko, Kokoszka, and Reimherr, 2017]. However, these methods can be easily failed when the model is misspecified and one, instead, resorts to a model-free test statistic with the serial dependence asymptotically reflected by a long-run variance, namely, the sum of autocovariances of all orders. Then the change-point detection is to monitor the stability of certain characters of the sequence, such as the mean, variance and correlation. Nevertheless, the consistent estimators of the long-run variance usually involve a bandwidth that is typically nontrivial to be selected in practice, leading to tests with nonmonotonic power, namely, the power of the test diminishes as the magnitude of changes in the alternative increases; see [Vogelsang, 1999] and [Crainiceanu and Vogelsang, 2007].

To bypass the undesirable nonmonotonic power, [Shao and Zhang, 2010] ingeniously adopted the idea of self-normalization (SN) in [Kiefer, Vogelsang, and Bunzel, 2000] and [Lobato, 2001] to pivotalize the limiting distribution of the test statistic. They replaced the long-run variance with a normalizer that is constructed using a sequence of recursive
estimator. As the nuisance parameter is canceled out by the self-normalizer, the proposed SN-based test by [Shao and Zhang, 2010] can not only yield a monotonic power but exhibit a better size performance. Since then, various tests for change-point detection based on the SN method have been developed and attracted much attention recently; see [Zhou, 2013], [Shao, 2015], and [Zhang, 2018]. It is interesting to observe that the above-mentioned SN-based tests may suffer from serious losses in power for the situation where the changes occur at an early or late stage of the time series. To illustrate this phenomenon, we, for example, suppose that the observations \( \{X_t\}_{t=1}^n \) follow the model with one shift in the mean:

\[
X_t = \mu \times 1_{(t > [n\lambda^*])} + \varepsilon_t,
\]

where \( \mu \) is the magnitude of change, \( \varepsilon_t = 0.6\varepsilon_{t-1} + \eta_t \) with \( \{\eta_t\} \) being iid standard normal innovations, and \( \lambda^* \) is the location of the change point. For simplicity, the scenario that the change occurs at the early (or late) stage of the sequence corresponds to \( \lambda^* \) taking value 0.05 (or 0.95) as described in Table 1. It is seen apparently from Table 1 that the

| \( \mu \) | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| \( \lambda^* = 0.05 \) | 0.074 | 0.086 | 0.090 | 0.096 | 0.164 | 0.224 | 0.302 | 0.416 | 0.566 | 0.608 |
| \( \lambda^* = 0.50 \) | 0.590 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| \( \lambda^* = 0.95 \) | 0.058 | 0.068 | 0.080 | 0.116 | 0.146 | 0.240 | 0.354 | 0.460 | 0.550 | 0.634 |

SN-based test in [Shao and Zhang, 2010], though displays a monotonic power under all three settings, is powerless seriously in the cases with \( \lambda^* = 0.05 \) and 0.95, compared with the case that the change point occurs at the middle stage of the sequence (e.g., \( \lambda^* = 0.50 \)). Hence, we conjecture that the locations of breaks would heavily affect the power of the SN-based tests.

It is worth noting that the cases that changes arise at the early or late stage of a time series are frequently encountered in both the “offline” and “online” change-point detection mechanisms. The offline methods consider the entire historical data set at once, and recognize the occurrence of changes retrospectively. And thus, the cases that the occurrence of the change point is close to the starting or termination of the historical
data might be easily observed when there is no priori knowledge on the change-point location. In addition, these cases are often inevitable when one is testing for data that arrive consecutively or observed sequentially, which can be referred to as the online detection scheme. A typical example is the sequential monitoring that has been widely applied in engineering, finance, medicine and meteorology for change-point detection as shown in [Horváth et al., 2020], [Dette and Gösmann, 2020] and [Chan, Ng, and Yau, 2021]. Basically speaking, the methodology of sequential monitoring advocates to test for a change by starting with a set of stable observations, say $X_1, \ldots, X_m$, and then conducting the testing procedure successively for the augmented dataset, say $X_1, \ldots, X_{m+k}$ with $k \geq 1$. In this case, as the fresh sample $\{X_t\}_{t=m+1,m+2,\ldots}$ is added steadily to the stable sequence $\{X_t\}_{t=1}^m$, the actual change point will first occur at the late stage of the current sequence. Therefore, detecting a change point as soon as possible after it occurs can help the practitioners to make decisions in a timely manner. In general, the cases with the occurrence of changes close to one or two ends of the sequence are commonly observed. However, compared to the situation where one is safely away from the end-points, it is rather tough to work with change points near end-points, due to the unbalanced data partition before and after the break. Therefore, testing approaches that can accurately and promptly detect the structural break in time series are urgently needed.

In this article, we propose a location-adaptive SN-based testing method to address the problem of power losses in traditional SN-based tests. As above, the power loss problem appears when the change point occurs near the boundaries of a time series. Indeed, we further find that this problem is basically attributed to the unbalance of the samples used before and after the location of the change point, when one is constructing a cumulative-sum-type statistic as shown in Section 2. So we first propose an optimal data selection strategy to balance the data partitions before and after the change point. This optimal scheme suggests that we do not need to involve the entire samples to construct the test statistic when the suspected location of the change is at the early or late stage of the sequence. Instead, to achieve the best power, one needs to ignore the samples that get farther away from the change point such that the sample size on one side of the location is twice of the other, leading to a relatively balanced data partition. On the other hand, when
the location of a shift is safely away from the end-points, we still take the entire samples into account. Then, on the basis of the optimal scheme, a new SN-based test adaptive to the locations of break points is developed. In contrast, the new test can substantially improve the power performance of the traditional SN-based test as well as maintain a satisfactory level of size. Though the proposed test seems to be tailored for the cases that changes arise at the early or late stage of a time series, it is a unified treatment that can be applied to any locations of change points. Moreover, it is straightforward to extend the proposed test to detect breaks in quantities of interest, including the median, the variance, and model parameters. On top of that, a location-adaptive-type change point estimate is proposed, which completes the framework of the change-point detection and is of practical appealing. The asymptotic theories of the resulting test and estimator are also established. It is worth noting that the proposed optimal data selection way can also work for tests based on the cumulative sum (CUSUM) process and thus is not restrictive to the SN-based test. At last, as a byproduct, we provide a theoretical justification for the power loss problem of the SN-based test [Shao and Zhang, 2010] given a wrong number of change points. This complements the seminal work by [Shao and Zhang, 2010] and [Zhang and Lavitas, 2018].

The rest of the article is organized as follows. Section 2 studies the change-point testing in the mean, reasons the power loss problem in the traditional SN-based test, and introduces the optimal data selection scheme. Then, we propose the location-adaptive SN-based test for single change-point alternative and its multiple version. An associated estimator for change-point location is proposed. The extensions to tests for a break in quantities other than the mean are shown in Section 3. Simulation studies and a real data analysis are conducted in Section 4. Some conclusions are depicted in Section 5. The theoretical justification for the power issue caused by a wrong number of change points is given in the Appendix. All technical proofs are deferred to the Supplementary Materials.

2 Testing for Change Points in the Mean

To illustrate our basic idea, we first consider the problem of detecting a change point in the mean. To be precise, for an observed univariate time series $X_1, \ldots, X_n$, we intend to
test the null hypothesis of no change point in the mean, namely,

\[ H_0 : E(X_1) = \cdots = E(X_n), \]

versus the alternative

\[ H_1 : E(X_1) = \cdots = E(X_{k^*}) \neq E(X_{k^*+1}) = \cdots = E(X_n), \]

where \( k^* \) is the true location of the change point and typically unknown in practice. In this article, distinguishing from the concerns in the existing literature, we instead restrict our attention to the case that the change occurs relatively early or late, namely, \( k^* \) is close to 1 or \( n \).

To facilitate our discussion, let \( \bar{X}_{j,k} = (k - j + 1)^{-1} \sum_{i=j}^{k} X_i \) be the sample mean of \( X_j, \ldots, X_k \). In the literature on change-point detection, a commonly-used strategy to recursively compare the partial mean \( \bar{X}_{1,k} \) with the overall mean \( \bar{X}_{1,n} \) yields the CUSUM process:

\[ Z_n([nt]) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (X_i - \bar{X}_{1,n}), \]

where \([nt]\) is the integer part of \( nt \) for \( t \in [0,1] \). To figure out the asymptotic property of \( Z_n([nt]) \), we need the mild assumption that, as \( n \to \infty \),

\[ S_n([nt]) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \{X_i - E(X_i)\} \Rightarrow \sigma B(t), \tag{2.1} \]

where "\( \Rightarrow \)" signifies weak convergence on the Skorohod space [Billingsley, 1999], \( B(t) \) is the standard Brownian motion and \( \sigma^2 \) is the long-run variance, or equivalently, the spectral density function at zero frequency. Then, it follows immediately from assumption (2.1) and the continuous mapping theorem that under the null hypothesis,

\[ Z_n([nt]) = S_n([nt]) - \frac{[nt]}{n} S_n(n) \Rightarrow \sigma \{B(t) - tB(1)\}. \tag{2.2} \]

Nevertheless, the asymptotic distribution on the right-hand side of (2.2) depends on the unknown parameter \( \sigma \). A straightforward way to circumvent this problem is to plug in (2.2) a consistent estimator of \( \sigma \), which usually involves the selection of a smoothing bandwidth. A data-driven bandwidth is preferred for its adaptation to the underlying structure of data,
whereas it would lead to the power of associated test decreasing as the alternative deviates from the null; see [Vogelsang, 1999] for examples.

To avoid the estimation of the nuisance parameter, [Shao and Zhang, 2010] proposed to use the self-normalization technique to cancel out $\sigma$ and under the null, obtained the pivotalized asymptotic distribution of

$$G_n = \max_{k=1,\ldots,n-1} \frac{Z_n^2(k)}{V_n(k)} \xrightarrow{d} G = \sup_{t \in [0,1]} \left\{ B(t) - tB(1) \right\}^2 / V(t),$$  \hspace{1cm} (2.3)

where “$\xrightarrow{d}$” denotes the convergence in distribution,

$$V_n(k) = \frac{1}{n^2} \left[ \sum_{j=1}^{k} \left\{ \sum_{i=1}^{j} (X_i - \bar{X}_{1,k}) \right\}^2 + \sum_{j=k+1}^{n} \left\{ \sum_{i=j}^{n} (X_i - \bar{X}_{k+1,n}) \right\}^2 \right],$$

and

$$V(t) = \int_{0}^{t} \left\{ B(s) - \frac{s}{t} B(t) \right\}^2 ds + \int_{t}^{1} \left\{ B(1) - B(s) - \frac{1 - s}{1 - t} [B(1) - B(t)] \right\}^2 ds.$$  

Under the alternative, it is shown that $G_n$ converges in probability to infinity, leading to the power of the test approaching to one. However, their sophisticated method may suffer from a serious power loss when the location $k^*$ is rather small or large as shown in Table 1. To illustrate the main reason, we suppose that $k^* = [n\lambda^*]$ for $\lambda^* \in (0,1)$. The situation that the location parameter $\lambda^*$ close to 0 or 1 corresponds to the occurrence of the change near the starting or terminal point of the time series. Notice that under the alternative, $Z_n(k^*)$ can be decomposed as

$$Z_n(k^*) = \left\{ S_n(k^*) - \frac{k^*}{n} S_n(n) \right\} - \frac{k^*}{\sqrt{n}} \left( 1 - \frac{k^*}{n} \right) \Delta_n, \hspace{1cm} (2.4)$$

where $\Delta_n = EX_n - EX_1 \neq 0$ is the magnitude of the change in the mean. When $|\Delta_n| \sqrt{n} \to \infty$, the assumption (2.1) further implies

$$Z_n(k^*) \sim \sqrt{n} \lambda^* (1 - \lambda^*) \Delta_n, \hspace{1cm} (2.5)$$

where “$\sim$” means the asymptotically equivalent. It can be seen that the performance of $G_n$ is only determined by the magnitude of change, $\Delta_n$, and the term $\lambda^* (1 - \lambda^*)$, since the denominator part $V_n(k^*)$ is of order $O_p(1)$ and does not rely on $\Delta_n$. As a result, under the
alternative, for any $\lambda^*$ close to 0 or 1, the CUSUM term shall be much smaller, leading to the low power phenomenon in Table 1, whereas a $\lambda^*$ not near 0 and 1 presumably has more chances to generate a CUSUM process converging rapidly to infinity. In one word, heuristically, the severe loss of power in [Shao and Zhang, 2010] for the scenario that the change occurs very early or late is attributed to the unbalanced partition of data before and after the change point in the CUSUM process.

2.1 The Location-Adaptive SN Approach

To alleviate the foregoing less power problem, we shall here propose a new test statistic that does not always involve the entire samples in the CUSUM process yet still leads to a monotonic power immune to the particular location of the change. Inspired by the contrast-based idea in [Zhang and Lavitas, 2018], we directly compare the means before and after the hypothetical change point $j_2$ and define

$$D_n(j_1, j_2, j_3) = \frac{(j_2 - j_1 + 1)(j_3 - j_2)}{(j_3 - j_1 + 1)^{3/2}}(\bar{X}_{j_1:j_2} - \bar{X}_{j_2+1:j_3}).$$  \hspace{1cm} (2.6)

Then $Z_n(k^*)$ in (2.4) can be rewritten as $D_n(1, k^*, n)$. We next induce $\lambda_1 \in (0, \lambda^*]$ and $\lambda_2 \in (0, 1 - \lambda^*]$ to balance the data proportions before and after the true location $k^*$. In particular, similar to (2.5), one can show that under $H_1$,

$$D_n(k^* - \lfloor n\lambda_1 \rfloor, k^*, k^* + \lfloor n\lambda_2 \rfloor) \sim \sqrt{n\lambda_1\lambda_2(\lambda_1 + \lambda_2)^{-3/2}} \Delta_n,$$

of which the main term $\lambda_1\lambda_2(\lambda_1 + \lambda_2)^{-3/2}$ on the right-hand side can be maximized over $\lambda_1$ and $\lambda_2$. The optimization procedure yields

$$\begin{cases} 
\text{(a) when } \lambda^* \in (0, 1/3), & \lambda_1 = \lambda^* \text{ and } \lambda_2 = 2\lambda^*; \\
\text{(b) when } \lambda^* \in [1/3, 2/3], & \lambda_1 = \lambda^* \text{ and } \lambda_2 = 1 - \lambda^*; \\
\text{(c) when } \lambda^* \in (2/3, 1), & \lambda_1 = 2(1 - \lambda^*) \text{ and } \lambda_2 = 1 - \lambda^*.
\end{cases}$$

Remark 1. By replacing the $\lambda_1$ and $\lambda_2$ in $D_n(k^* - \lfloor n\lambda_1 \rfloor, k^*, k^* + \lfloor n\lambda_2 \rfloor)$ with the associated substitutes in (a)–(c), we see that for $\lambda^* \in [1/3, 2/3]$, namely, $k^*$ is located around the middle of the series, $D_n(k^* - \lfloor n\lambda_1 \rfloor, k^*, k^* + \lfloor n\lambda_2 \rfloor)$ reduces to $Z_n(k^*)$. However, when $k^*$ is close to the starting or terminal point as in (a) and (c), it is suboptimal to compare the partial
means based on all the samples as the conventional approach in [Shao and Zhang, 2010]. Instead, we need to leave out the samples far away from \( k^* \) such that the sample size on one side of \( k^* \) is twice of the other, leading to a relatively balanced data partition. Notice that this elimination conducts at only one side of \( k^* \) for the sake of involving more data as possible. For instance, when \( k^* \) is close to 1 as in (a), we adopt all the samples before \( k^* \) but discard samples with indices after \( 3k^* \).

**Remark 2.** In essence, the problem of change-point testing involves comparing the magnitude of change between quantities of interest and more importantly, the data partition ratio before and after the break, as shown in (2.5). The optimal scheme in (a) or (c) ensures a more proper data partition by eliminating data far away from the break and thus can effectively improve the power, though the elimination of samples may lead to a less accurate estimation of the magnitude of change. Interestingly, the optimal ratio in (a) and (c) is not one-to-one, which gives the most balanced samples as our intuition suggests. This is because a one-to-one data partition would lead to a lower power as smaller number of samples are involved. Therefore, the scheme in (a)–(c) is a tradeoff between the balance of data partition and the efficiency of the testing procedure.

Figure 1 provides a more insightful illustration on the relationship between \( \lambda_1, \lambda_2 \) and \( f(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 (\lambda_1 + \lambda_2)^{-3/2} \). Then we propose the optimal subsample selection scheme by extending the results in (a)–(c) to each hypothetical change point:

\[
\Omega(\epsilon) = \{(t_1, t_2, t_3) : t_1 = 0, t_3 = 3\epsilon, \text{for} \ t_2 \in [0, \epsilon); \ t_1 = 0, t_3 = 3t_2, \text{for} \ t_2 \in [\epsilon, 1/3); \\
\quad t_1 = 0, t_3 = 1, \text{for} \ t_2 \in [1/3, 2/3); \ t_1 = 3t_2 - 2, t_3 = 1, \text{for} \ t_2 \in (2/3, 1 - \epsilon); \\
\quad t_1 = 1 - 3\epsilon, t_3 = 1, \text{for} \ t_2 \in (1 - \epsilon, 1]\},
\]

where \( \epsilon \in (0, 1/3) \) is a trimming parameter. Based on \( \Omega(\epsilon) \), a location-adaptive SN-based test statistic is established:

\[
G_{n}^{LA} = \max_{(j_1, j_2, j_3) \in \Omega_n(\epsilon)} \left\{ \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} \right\},
\]
Figure 1: The contours of the function $f(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 (\lambda_1 + \lambda_2)^{-3/2}$ and the curves of the function $g(\lambda_2) = f(\lambda^*, \lambda_2)$ with $\lambda^* = 0.1$ (left) and $\lambda^* = 0.5$ (right).

where $\Omega_n(\epsilon) = \{([nt_1] \vee 1, [nt_2] \vee 1, [nt_3]) : (t_1, t_2, t_3) \in \Omega(\epsilon)\}$ and

$$V_n(j_1, j_2, j_3) = \sum_{i=j_1}^{j_2} \frac{(i - j_1 + 1)^2 (j_2 - i)^2}{(j_3 - j_1 + 1)^2 (j_2 - j_1 + 1)^2} (\bar{X}_{j_1,i} - \bar{X}_{i+1,j_2})^2 + \sum_{i=j_2+1}^{j_3} \frac{(i - j_2)^2 (j_3 - i)^2}{(j_3 - j_1 + 1)^2 (j_3 - j_2)^2} (\bar{X}_{j_2+1,i} - \bar{X}_{i+1,j_3})^2. $$

Furthermore, for $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$, denote

$$\Delta(B, t_1, t_2, t_3) = (t_3 - t_2)\{B(t_2) - B(t_1)\} - (t_2 - t_1)\{B(t_3) - B(t_2)\}. \quad (2.8)$$

Then, the limiting distribution of $G_n^{LA}$ is presented below.
Theorem 2.1. Suppose that the condition (2.1) holds. Then (i) under $H_0$, we have
\[ G_n^{LA} \xrightarrow{d} G^{LA} := \sup_{(t_1,t_2,t_3) \in \Omega(\epsilon)} D^2(B,t_1,t_2,t_3)/V(B,t_1,t_2,t_3), \]
where
\[ D(B,t_1,t_2,t_3) = \frac{1}{(t_3 - t_1)^{3/2}} \Delta(B,t_1,t_2,t_3), \quad (2.9) \]
and (ii) under $H_1$, if $\lambda^* \in (0,1)$ and $\sqrt{n} |\Delta_n| \to \infty$, then $G_n^{LA} \xrightarrow{p} \infty$, as $n \to \infty$.

For any significant level $\alpha$, the null hypothesis is rejected when $G_n^{LA} > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$th quantile of $G^{LA}$, and its consistency is guaranteed by Theorem 2.1 (ii). The proposed location-adaptive SN-based method, though appears to be tailored for the situation where the occurrence of change is close to the boundaries of the time series, is typically general enough and its implementation does not rely on any condition concerning the true locations of change points.

Remark 3. The optimal subsample selection scheme $\Omega(\epsilon)$ adaptive to the locations of potential change points is the key ingredient that significantly increases the power for marginal change-point cases. The trimming parameter $\epsilon$ is used to guarantee that $t_3 - t_1$ in $\Omega(\epsilon)$ is uniformly bounded away from zero and to ensure that the mapping from $B(\cdot)$ to $D^2(B,\cdot\cdot\cdot)/V(B,\cdot\cdot\cdot)$ is continuous. In fact, by the law of iterated logarithm of Brownian motion, the fluctuation of $D^2(B,0,t,3t)/V(B,0,t,3t)$ is very complicated as $t \to 0$ and some additional simulation results reported in the Supplementary Materials show that the size-distortion would be serious if $\epsilon$ is removed (i.e., $\epsilon = 0$). In addition, when $t_2$ belongs to the trimming part $(0,\epsilon)$, though we are not able to choose the best subsample scheme $(0,t_2,3\epsilon)$, the scheme $(0,t_2,3\epsilon)$ used is still much better than the original $(0,t_2,1)$, as can be seen from the monotone property of $f(\lambda_1,\lambda_2)$ as shown in Figure 1 and the numerical results in Section 4.

Remark 4. An alternative idea to detect the change point near the end-points is to introduce one weight function $w(t)$ to the SN-based test and construct the weighted SN test statistic:
\[ G_n^W = \max_k w^2(k/n)Z_n^2(k)/V_n(k). \]
By downweighting the difference near the end-points, the power of the SN test can be improved when the change point is close to the boundaries. Following [Csörgö and Horváth, 1997], the most commonly used weight function is \( w(t) = [t(1-t)]^{-\kappa} \) with \( \kappa \in [0, 1/2] \). In particular, when \( \kappa = 1/2 \), it corresponds to the sup-Wald tests used in [Andrews, 1993] and is closely related to likelihood ratio statistics when the error terms are Gaussian. In the weighted SN test, the trimming parameter \( \epsilon \) is needed to guarantee the convergence of the limiting distribution and the weighted version is given by

\[
G_n^W = \max_{k=[n\epsilon], \ldots, n-[n\epsilon]} \left( \frac{k(n-k)}{n^2} \right)^{-2\kappa} \frac{Z_n^2(k)}{V_n(k)}.
\]

The \( G_n^W \) test is natural and easy to implement, which is of practical appeal. However, the power performance of the weighted SN test would depend on the chosen \( \kappa \in [0, 1/2] \). On the other hand, since the weighted SN test does not consider points in the trimmed intervals (i.e. \([0, \epsilon]\) and \([1-\epsilon, 1]\)) as the potential change point, it may suffer from serious power losses when \( \lambda^* \notin [\epsilon, 1-\epsilon] \) as shown in Table 5 in Section 4.

The trimming parameter \( \epsilon \) directly determines the limiting distribution and critical values, and thus is different from the smoothing parameters required by the spectral density estimation [Liu and Wu, 2010], the dependent bootstrap [Shao, 2010] and subsampling methods [Politis, Romano, and Wolf, 1999]. In the current paper, since we mainly study the marginal change-point case, the values of \( \epsilon \) such as 0.05, 0.1 are used, which is also recommended by [Zhou and Shao, 2013] and [Huang, Volgushev, and Shao, 2015].

When the null hypothesis is rejected, we can naturally define

\[
(j_1^*, j_2^*, j_3^*) = \arg\max_{(j_1,j_2,j_3) \in \Omega_n(\epsilon)} \left\{ \frac{D_n^2(j_1,j_2,j_3)}{V_n(j_1,j_2,j_3)} \right\},
\]

then the change point is estimated by \( \hat{\lambda}^* = n^{-1}j_2^* \). The next theorem gives its consistency.

**Theorem 2.2.** Under \( H_1 \), suppose that the condition \( (2.1) \) holds, and \( \sqrt{n}|\Delta_n| \to \infty \). Then, for any \( \eta > 0 \), it follows that

\[
P(|\hat{\lambda}^* - \lambda^*| < \eta) \longrightarrow 1.
\]

Theorem 2.2 shows that the estimated change-point location \( \hat{\lambda}^* \) is consistent to the true location \( \lambda^* \), in a similar to the existing results in [Jiang, Zhao, and Shao, 2021] and
Dette, Kokot, and Volgushev, 2020. And the proof procedure for Theorem 2.2 is novel for a location-adaptive-type change-point estimator; see the Supplementary Materials for technical details.

2.2 Testing for Multiple Change Points

The location-adaptive statistic \( G_{LA}^n \) is devised for the alternative that there exists only one change point in the time series. A direct application of the single SN-based change-point statistic to test for possibly more than one change points is not feasible, because the existence or the number of change points is typically unknown in practice. Given a wrong priori information on the number of change points, we rigorously show that applying the conventional SN-based statistic \( G^n \) in [Shao and Zhang, 2010] to test for multiple changes may lead to a power decreasing to zero; see Theorem A.1 in the Appendix. Thus, we consider extending the proposed location-adaptive approach to test with the alternative that does not require any priori information on the number of change points, namely, \( H_m: \) there exist \( m \geq 2 \) break points \( 1 < k_1^* < \cdots < k_m^* < n \) such that

\[
EX_i \neq EX_{i+1}, \quad i \in \{k_1^*, \cdots, k_m^*\},
\]

and \( EX_i = EX_{i+1} \) otherwise, where \( k_i^* = \lfloor n\lambda_i^* \rfloor \) for \( \lambda_i^* \in (0, 1) \).

We intend to search for any subsequence that encompasses only one unique change point and then the problem is reduced to test for the single change point. Specifically, we recursively study the subsamples \( \{X_1, \cdots, X_{nr}\} \) and \( \{X_{[ns]+1}, \cdots, X_n\} \) to seek for the first and last change points, respectively. Similar to the optimal selection set \( \Omega(\epsilon) \) with respect to the overall samples, the forward optimal set for \( \{X_1, \cdots, X_{[nr]}\} \) can be readily derived:

\[
\Omega(\epsilon; 0, r) = \{(t_1, t_2, t_3) : t_1 = 0, t_3 = 3t_2, \text{ for } t_2 \in [\epsilon, r/3); t_1 = 0, t_3 = r, \text{ for } t_2 \in [r/3, 2r/3); t_1 = 3t_2 - 2r, t_3 = r, \text{ for } t_2 \in (2r/3, r - \epsilon]\},
\]

where \( r \in [2\epsilon, 1] \). The backward optimal set for \( \{X_{[ns]+1}, \cdots, X_n\} \) is

\[
\Omega(\epsilon; s, 1) = \{(t_1, t_2, t_3) : t_1 = s, t_3 = 3t_2 - 2s, \text{ for } t_2 \in [s + \epsilon, 2s/3 + 1/3); t_1 = s, t_3 = 1, \text{ for } t_2 \in [2s/3 + 1/3, s/3 + 2/3]; t_1 = 3t_2 - 2, t_3 = 1, \text{ for } t_2 \in (s/3 + 2/3, 1 - \epsilon]\},
\]
where \( s \in [0, 1 - 2\epsilon] \). Let
\[
\Omega_n(\epsilon; 0, r) = \{(\lfloor nt_1 \rfloor \vee 1, \lfloor nt_2 \rfloor, \lfloor nt_3 \rfloor) : (t_1, t_2, t_3) \in \Omega(\epsilon; 0, r)\},
\]
\[
\Omega_n(\epsilon; s, 1) = \{(\lfloor nt_1 \rfloor \vee 1, \lfloor nt_2 \rfloor, \lfloor nt_3 \rfloor) : (t_1, t_2, t_3) \in \Omega(\epsilon; s, 1)\}.
\]
Then, the proposed multiple change-point test is defined as
\[
Q_n^{LA} = \sup_{r \in [2\epsilon, 1]} \sup_{(j_1, j_2, j_3) \in \Omega_n(\epsilon; 0, r)} \{D_n^2(j_1, j_2, j_3)/V_n(j_1, j_2, j_3)\}
+ \sup_{s \in [0, 1 - 2\epsilon]} \sup_{(j_1, j_2, j_3) \in \Omega_n(\epsilon; s, 1)} \{D_n^2(j_1, j_2, j_3)/V_n(j_1, j_2, j_3)\}.
\]
The following theorem provides the asymptotic distribution of \( Q_n^{LA} \) under both the null and the alternative.

**Theorem 2.3.** Suppose that the condition \((2.1)\) holds. (i) Under \( H_0 \), we have
\[
Q_n^{LA} \overset{d}{\rightarrow} Q^{LA} := \sup_{r \in [2\epsilon, 1]} \sup_{(t_1, t_2, t_3) \in \Omega(\epsilon; 0, r)} \{D^2(B, t_1, t_2, t_3)/V(B, t_1, t_2, t_3)\}
+ \sup_{s \in [0, 1 - 2\epsilon]} \sup_{(t_1, t_2, t_3) \in \Omega(\epsilon; s, 1)} \{D^2(B, t_1, t_2, t_3)/V(B, t_1, t_2, t_3)\},
\]
as \( n \to \infty \). (ii) Under \( H_m \), suppose that \( \min_{0 \leq i \leq m} (\lambda_{i+1}^* - \lambda_i^*) \geq \epsilon \) holds with \( \lambda_0^* = 0 \) and \( \lambda_{m+1}^* = 1 \). If \( \min_{1 \leq i \leq m} \sqrt{n}|EX_{k_i^* + 1} - EX_{k_i^*}| \to \infty \), then \( Q_n^{LA} \overset{p}{\rightarrow} \infty \).

Since the location-adaptive technique is incorporated into the extended statistic \( Q_n^{LA} \), the new test based on \( Q_n^{LA} \) is immune to some special locations of change points, such as the first (last) change point is near the starting (end) point or some consecutive changes are up-close.

### 3 Extensions to Other Quantities

By replacing the mean in the CUSUM process with the corresponding counterpart in a more general setting, the proposed location-adaptive SN approach can be readily extended from the mean case to other quantities. Let \( \theta_t \in \mathbb{R}^d \) be the quantity of interest, such as the numerical character of \( Y_t = (X_t, X_{t+1}, \cdots, X_{t+l-1})' \), for example, the mean, quantile and correlation. In contrast with the common settings in many existing methods [Zhang and Lavitas, 2018, Dette and Gösmann, 2020], \( \theta_t \in \mathbb{R}^d \) is not restricted to be a functional of the distribution of \( Y_t \). It can also be the parameters in time series models.
including the ARMA, GARCH and TAR models, among others. We consider the problem of testing for the null:

\[ H_0 : \theta_1 = \cdots = \theta_N = \theta_0^* , \]

versus

\[ H_1 : \theta_1 = \cdots = \theta_{k^*} = \theta_0^* \neq \theta_1^* = \cdots = \theta_N , \]

where \( \theta_0^* \) and \( \theta_1^* \) are two vectors and \( N = n - l + 1 \) with \( l \in \mathbb{N}^* \), and \( k^* = \lceil N \lambda^* \rceil \) with \( \lambda^* \in (0, 1) \).

Recall that the new SN-based test \( G_n^{LA} \) for detecting the mean change compares the difference between the forward sample mean \( \bar{X}_{j_1,j_2} \) and the backward \( \bar{X}_{j_2+1,j_3} \). Similarly, under the assumption that \( \theta_{l_1} = \cdots = \theta_{l_2} \), one can construct some estimator \( \hat{\theta}_{l_1,l_2} \) on the basis of the samples \( \{ Y_{l_1}, \cdots, Y_{l_2} \} \) by approaches including the maximum likelihood estimation, M-estimation, least-squares method or generalized moment method. Then, it is natural to define the counterpart of (2.6) as

\[ D_N(\hat{j}_1, \hat{j}_2, \hat{j}_3) = \frac{(j_2 - j_1 + 1)(j_3 - j_2)}{(j_3 - j_1 + 1)^{3/2}} (\hat{\theta}_{j_1,j_2} - \hat{\theta}_{j_2+1,j_3}). \]

Meanwhile, for \( 0 \leq s \leq t \leq 1 \), define

\[ L_N(s,t; \theta^*_0) = \frac{\lceil Nt \rceil \lor 1 - \lceilNs \rceil \lor 1 + 1}{\sqrt{N}} (\hat{\theta}_{\lceil Ns \rceil \lor 1, \lceil Nt \rceil \lor 1} - \theta^*_0). \]

Let \( l^\infty(\mathbb{D}_\delta) \) be the space of bounded real-valued functions on \( \mathbb{D}_\delta = \{(s,t): t - s \geq \delta\} \) for \( \delta \geq 0 \), and “ \( \rightsquigarrow \) ” denote the weak convergence in the sense of Hoffmann-Jørgensen [van der Vaart and Wellner, 1996]. We further need the following assumption to guarantee the corresponding asymptotic properties.

**Assumption 3.1.** (i) Under \( H_0 \), there exists one continuous random element \( L_0(s,t) \) in the space \( l^\infty(\mathbb{D}_{\delta_0}) \) for some \( \delta_0 \geq 0 \), such that for any \( \delta > \delta_0 \),

\[ L_N(s,t; \theta^*_0) \rightsquigarrow L_0(s,t) \text{ in the space } l^\infty(\mathbb{D}_{\delta}); \quad (3.1) \]

(ii) under \( H_1 \), there exists the other continuous random element \( L_1(s,t) \) such that

\[ L_N(s,t; \theta^*_0) \rightsquigarrow L_0(s,t) \text{ in the space } l^\infty(\mathbb{D}_{\delta} \cap \{(s,t): t \leq \lambda^*\}); \quad (3.2) \]

\[ L_N(s,t; \theta^*_1) \rightsquigarrow L_1(s,t) \text{ in the space } l^\infty(\mathbb{D}_{\delta} \cap \{(s,t): s \geq \lambda^*\}). \quad (3.3) \]
Assumption 3.1 can be regarded as a sequential version of invariance principle. However, it can not be implied by the standard invariance principle (i.e., \(s\) is fixed to be zero) as the estimator \(\hat{\theta}_{t_1,t_2}\) may be non-linear with respect to \(l_1\) and \(l_2\), such as quantile and correlation. Notice that weak convergence of \(L_N(s,t;\theta^*_i)\) in the functional space \(l^\infty(\mathcal{D}_\delta)\) as (3.1)–(3.3) is slightly weaker than that in the conventional \(l^\infty(\mathcal{D}_0)\) in [Shao and Zhang, 2010] and [Zhang and Lavitas, 2018]. This allows for a more general form of \(\hat{\theta}_{t_1,t_2}\) other than the functionals of distribution, which is of more practical relevance. We present two examples to demonstrate this point as below.

**Example 3.1.** Suppose that one is interested in the quantity \(\theta_t = \phi(F_t)\), where \(\phi\) is a functional and \(F_t\) is the distribution function of \(Y_t\). The corresponding estimator of \(\theta_t\) is \(\hat{\theta}_{t_1,t_2} = \phi(\hat{F}_{t_1,t_2})\), where \(\hat{F}_{t_1,t_2}\) is the empirical distribution of \((Y_{t_1}, \ldots, Y_{t_2})\). Under some dependent and smoothing conditions, (3.1) can be easily verified with any \(\delta \geq 0\); for example, see [Volgushev and Shao, 2014], [Shao, 2015] and [Dette and Gösmann, 2020]. In this case, \(\delta\) can be chosen to be 0.

**Example 3.2.** Suppose that \(X_t = Z_t\theta_t + \varepsilon_t\), where \(Z_t = (f_1(t/n), \ldots, f_p(t/n), Y_{t-1})\)' and \(f_i(s)\) is one non-random trend function. Consider the least squares estimator \(\hat{\theta}_{t_1,t_2} = (\sum_{t=t_1}^{t_2} Z_tZ_t')^{-1}(\sum_{t=t_1}^{t_2} Z_tX_t)\). To ensure that \(\hat{\theta}_{t_1,t_2}\) is well-defined, the length \(l_2 - l_1 + 1\) must be larger than \(p + 1\), resulting in \(\delta > 0\); see [Rho and Shao, 2015] for more technical details.

Based on Assumption 3.1, the self-normalizer should be modified as

\[
V_N(j_1, j_2, j_3) = \sum_{i=j_1+[N\delta]}^{j_2-[N\delta]} \frac{j_2 - j_1 + 1}{(j_3 - j_1 + 1)^2} D_N(j_1, i, j_2) D_N'(j_1, i, j_2) + \sum_{i=j_2+[N\delta]+1}^{j_3-[N\delta]} \frac{j_3 - j_2}{(j_3 - j_1 + 1)^2} D_N(j_2 + 1, i, j_3) D_N'(j_2 + 1, i, j_3).
\]

By Assumption 3.1, it is straightforward to show that \(V_N(j_1, k^*, j_3)\) is of the order \(O_p(1)\). Moreover, this assumption further implies that,

\[
D_N(k^* - [N\lambda_1], k^*, k^* + [N\lambda_2]) \sim \sqrt{N\lambda_1\lambda_2}\lambda_1 + \lambda_2)^{-3/2}(\theta^*_0 - \theta^*_1).
\]

Therefore, the optimal selection scheme in the mean case is still feasible to the general setting and notice that the potential \(t_2\) must belong to \([2\delta, 1-2\delta]\), leading to the following
test statistic

\[
\tilde{G}_{N}^{LA} = \max_{(j_1,j_2,j_3)} \left\{ D_N(j_1,j_2,j_3) V_N^{-1}(j_1,j_2,j_3) D_N(j_1,j_2,j_3) \right\},
\]

(3.4)

where \( \Omega_N(\epsilon, \delta) = \{(\lfloor N t_1 \rfloor \lor 1, \lfloor N t_2 \rfloor \lor 1, \lfloor N t_3 \rfloor) : (t_1, t_2, t_3) \in \Omega(\epsilon, \delta) \} \) with \( \Omega(\epsilon, \delta) = \{(t_1, t_2, t_3) : (t_1, t_2, t_3) \in \Omega(\epsilon) \) and \( t_2 \in [2\delta, 1 - 2\delta] \). The counterparts of (2.8)–(2.10) are given by

\[
\begin{align*}
\Delta(L_0, t_1, t_2, t_3) &= (t_3 - t_2)L_0(t_1, t_2) - (t_2 - t_1)L_0(t_2, t_3) \\
D(L_0, t_1, t_2, t_3) &= \frac{1}{(t_3 - t_1)^3/2} \Delta(L_0, t_1, t_2, t_3), \\
V(L_0, t_1, t_2, t_3) &= \frac{1}{(t_3 - t_1)^2} \left\{ \int_{t_1 + \delta}^{t_2 - \delta} \frac{\Delta(L_0, t_1, s, t_2) \Delta'(L_0, t_1, s, t_2)}{(t_2 - t_1)^2} ds, \\
&\quad + \int_{t_2 + \delta}^{t_3 - \delta} \frac{\Delta(L_0, t_2, s, t_3) \Delta'(L_0, t_2, s, t_3)}{(t_3 - t_2)^2} ds \right\}.
\end{align*}
\]

Then, the asymptotic property of \( \tilde{G}_{N}^{LA} \) is stated as follows.

**Theorem 3.1.** Suppose that Assumption 3.1 holds with \( \delta < \epsilon/2 \) and \( V(L_i, t_1, t_2, t_3) \) is invertible almost surely. (i) Under \( H_0 \), we have

\[
\tilde{G}_{N}^{LA} \xrightarrow{d} \tilde{G}^{LA} := \sup_{(t_1, t_2, t_3) \in \Omega(\epsilon, \delta)} D'(L_0, t_1, t_2, t_3) V^{-1}(L_0, t_1, t_2, t_3) D(L_0, t_1, t_2, t_3);
\]

(ii) under \( H_1 \), if \( \lambda^* \in [2\delta, 1 - 2\delta] \) and \( \sqrt{n} |\theta_1^* - \theta_0^*| \to \infty \), then we have

\[
\tilde{G}_{N}^{LA} \xrightarrow{p} \infty, \text{ as } n \to \infty.
\]

In addition, if \( \delta \) can be released to 0, the above results will hold under one more assumption that \( \sup_t |L_i(t, t)| = 0 \) almost surely, where \( \Omega(\epsilon, \delta) \) is reduced to \( \Omega(\epsilon) \). In a special case, when \( L_i(s, t) = \Sigma_i [B(t) - B(s)] \) holds with some invertible matrix \( \Sigma_i \) and a standard \( d \)-dimensional Brownian motion \( B(t) \), \( \Sigma_i \) is eliminated and the limiting distribution \( \tilde{G}^{LA} \) can be simulated for different \( \epsilon \) and \( \delta \); see also [Jiang, Zhao, and Shao, 2021] for the other pivotal case. Nevertheless, the limiting distribution \( \tilde{G}^{LA} \) is usually unknown and may depend on nuisance parameters [Rho and Shao, 2015]. One possible method to approximate its critical values is the bootstrap as shown in [Xu, 2012] and [Zhao and Li, 2013]. However, an efficient bootstrap procedure often relies on the concrete models or temporal dependence. For instance, in the classical unit-root testing problem, [Zhang and Chan, 2020] has shown
that the sieve wild bootstrap method proposed by [Cavaliere, Georgiev, and Taylor, 2018] is useless when the innovation is generated from iid to GARCH structure. Thus, whether there exists a unified bootstrap selection procedure under certain assumptions on $L(s, t; \theta^*_t)$ is an interesting but challenging problem, which is left for future research.

**Remark 5.** Similar to the mean case, we can estimate the change point $\lambda^*$ for quantities of interest by $\hat{\lambda}^* = N^{-1} j^*_2$, where $j^*_2$ is defined as

$$(j^*_1, j^*_2, j^*_3) = \max_{(j_1, j_2, j_3) \in \Omega_N(\epsilon, \delta)} \{ D_N(j_1, j_2, j_3) V^{-1}_N(j_1, j_2, j_3) D_N(j_1, j_2, j_3) \}.$$  

Though the simulation results on the median case presented in Section 4 demonstrate the consistency of $\lambda^*$ in an empirical manner, it seems not sufficient to derive the consistency of $\hat{\lambda}^*$ only under Assumption 3.1. Actually, the Assumption 3.1 states that the original data can be divided into two stationary segments before and after $k^*$, but the properties of $\hat{\theta}_{j_2+1, j_3}$ with $j_2 < k^*$ and $j_3 > k^*$ are not provided. Therefore, whether the consistency of $\lambda^*$ holds or not relies on the specific conditions for $\hat{\theta}_{l_1, l_2}$. In particular, when the estimator $\hat{\theta}_{l_1, l_2}$ is linear with respect to $l_1$ and $l_2$, the consistency of $\lambda^*$ follows from a similar proof with Theorem 2.2. Additionally, the consistency is possible to hold if $\hat{\theta}_{l_1, l_2}$ comes from a linear model as that in Example 3.2; see [Jiang, Zhao, and Shao, 2021] and [Dette, Kokot, and Volgushev, 2020] for more details.

## 4 Empirical Studies

### 4.1 Simulation

In this section, we investigate the finite sample properties of the proposed location-adaptive SN-based (LASN) test for detecting a change in both the mean and the median. The notations LASN$_{0.05}$ and LASN$_{0.1}$ correspond to the trimming parameter taking values 0.05 and 0.1, respectively. A comparison with other detection schemes based on the traditional SN-type statistic in (2.3) and the weighted SN statistic in (2.11) with $\kappa = 1/2$, which we separately abbreviate as SN and WSN, is also provided in terms of the size and power performance. All the results presented here are based on 1000 replications with sample size
\(n = 500\). The simple-location model is considered:

\[X_t = \mu 1_{(t > \lfloor n\lambda^* \rfloor)} + \varepsilon_t,\]

where \(\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t\) with \(\{\eta_t\}\) being a sequence of iid \(N(0, 1)\) and \(\rho \in \{0, \pm 0.3, \pm 0.6\}\).

We first examine the performance of all the testing methods under the null, and the empirical acceptance rates at nominal levels 90% and 95% are provided in Table 2. The proposed test exhibits a stable acceptance rate as the SN test, which implies that the proposed method carries on the size advantage from self-normalization. In addition, as can be seen from Table 2, the size performance of the proposed test is basically immune to the trimming parameter. While the SN-based test and the weighted SN-based test exhibit a satisfactory size performance, they may have a serious power issue under the alternative, when the change point occurs near the end-points.

On the other hand, to investigate the power performance of the proposed LASN, we set the location of the change point \(\lambda^*\) to be 0.90, 0.95 and 0.98, respectively, for \(\rho \in \{0, 0.3, 0.6\}\). The empirical powers under different magnitude of the change are reported in Tables 3–5, from which one can observe that the SN-based test has severe power losses when the change occurs near the end-points. As expected, in the situation where the change occurs relatively late (e.g. \(\lambda^* = 0.95\) or 0.98), the power performance of the proposed test has a substantial improvement compared with the traditional SN test. The LASN also delivers a monotonic power, which is one of the appealing features of the SN-type methodology. Moreover, when the true break is not trimmed by \(\epsilon\) (i.e. \(\lambda^* \in [\epsilon, 1 - \epsilon]\)), the power performance of the LASN test is comparable to the weighted SN test, seeing from Table 4 with \(\epsilon = 0.05\). In the case that the true break appeared in the trimmed data (i.e. \(\lambda^* \notin [\epsilon, 1 - \epsilon]\)), such as \(\lambda^* = 0.98\) with \(\epsilon = 0.05\) or 0.1, the weighted SN-based test failed to detect the change point in the median, while the proposed test is still feasible and delivers a reasonable power as shown in Table 5. This is because that different from the weighted SN, the LASN test also considers the trimmed points as the potential change point and thus might be robust to test for a change in the non-linear transformation of data, such as the median. Furthermore, we observe that the effect of trimming on the power performance is relatively little for the proposed LASN if \(\lambda^* \in [\epsilon, 1 - \epsilon]\), as shown in Table 3. When one adopt a trimming parameter \(\epsilon\) such that \(\lambda^* \not\in [\epsilon, 1 - \epsilon]\), the proposed test is still much more
powerful than the SN test. In addition, the mean absolute errors $|\hat{\lambda}^* - \lambda^*|$ or $|\tilde{\lambda}^* - \lambda^*|$ for $\lambda^* = 0.95$ and $\rho = 0.3$ are plotted in Figure 2, which shows that the proposed change-point location estimator $\hat{\lambda}^*$ or $\tilde{\lambda}^*$ is consistent for both the change in the mean and median.

In general, the proposed location-adaptive detection scheme not only substantially improves the power performance in detecting a shift in the mean or median, but has a reasonable size level for the scenarios under consideration. Moreover, the new method is stable in the sense that the effect of trimming is acceptable for both the size and power performance.

Table 2: Empirical acceptance rates of the tests for a change in the mean and median

| $\rho$ | SN 90% | SN 95% | LASN$_{0.05}$ 90% | LASN$_{0.05}$ 95% | WSN$_{0.05}$ 90% | WSN$_{0.05}$ 95% | LASN$_{0.1}$ 90% | LASN$_{0.1}$ 95% | WSN$_{0.1}$ 90% | WSN$_{0.1}$ 95% |
|--------|--------|--------|------------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0      | 0.906  | 0.957  | 0.911 0.949      | 0.911 0.946      | 0.903 0.944    | 0.920 0.954    | 0.901 0.944    | 0.908 0.960    | 0.914 0.964    | 0.924 0.962    |
| 0.3    | 0.897  | 0.953  | 0.907 0.954      | 0.905 0.950      | 0.904 0.953    | 0.903 0.950    | 0.901 0.944    | 0.908 0.960    | 0.914 0.964    | 0.924 0.962    |
| 0.6    | 0.900  | 0.953  | 0.885 0.942      | 0.905 0.965      | 0.901 0.944    | 0.908 0.960    | 0.910 0.958    | 0.914 0.964    | 0.914 0.964    | 0.914 0.964    |
| -0.3   | 0.903  | 0.946  | 0.922 0.965      | 0.918 0.962      | 0.910 0.958    | 0.914 0.964    | 0.914 0.958    | 0.914 0.964    | 0.914 0.964    | 0.914 0.964    |
| -0.6   | 0.910  | 0.955  | 0.936 0.969      | 0.919 0.960      | 0.928 0.964    | 0.924 0.962    | 0.928 0.964    | 0.924 0.962    | 0.924 0.962    | 0.924 0.962    |

Results for the mean

| $\rho$ | 0      | 0.3    | 0.6    | -0.3   | -0.6   |
|--------|--------|--------|--------|--------|--------|
| 0      | 0.857  | 0.870  | 0.867  | 0.890  | 0.846  |
| 0.3    | 0.852  | 0.879  | 0.855  | 0.894  | 0.849  |
| 0.6    | 0.865  | 0.864  | 0.867  | 0.895  | 0.872  |
| -0.3   | 0.857  | 0.864  | 0.867  | 0.895  | 0.872  |
| -0.6   | 0.850  | 0.886  | 0.860  | 0.896  | 0.850  |

Results for the median

Figure 2: Mean absolute errors of $\hat{\lambda}^*$ (left) and $\tilde{\lambda}^*$ (right) for $\lambda^* = 0.95$ and $\rho = 0.3$. 
Table 3: Empirical powers of the tests for a change in the mean and median with $\lambda^* = 0.90$

| $\rho$ | $\mu$ | Mean |   | Median |   |
|-------|-------|------|---|--------|---|
|       |       | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ |
| 0.0   | 0.3   | 0.105 | 0.168 | 0.167 | 0.170 | 0.149 | 0.107 | 0.142 | 0.134 | 0.161 | 0.133 |
| 0.3   | 0.3   | 0.080 | 0.097 | 0.107 | 0.108 | 0.103 | 0.096 | 0.130 | 0.113 | 0.135 | 0.109 |
| 0.6   | 0.3   | 0.085 | 0.161 | 0.108 | 0.146 | 0.103 | 0.102 | 0.176 | 0.130 | 0.174 | 0.121 |
| 0.9   | 0.3   | 0.153 | 0.288 | 0.285 | 0.323 | 0.271 | 0.144 | 0.251 | 0.237 | 0.268 | 0.219 |
| 1.2   | 0.3   | 0.302 | 0.547 | 0.552 | 0.551 | 0.511 | 0.280 | 0.488 | 0.473 | 0.507 | 0.440 |
| 1.5   | 0.3   | 0.502 | 0.824 | 0.789 | 0.803 | 0.748 | 0.429 | 0.697 | 0.701 | 0.709 | 0.681 |
| 1.8   | 0.3   | 0.704 | 0.917 | 0.906 | 0.908 | 0.878 | 0.592 | 0.855 | 0.826 | 0.855 | 0.802 |
| 2.0   | 0.3   | 0.885 | 0.985 | 0.977 | 0.988 | 0.964 | 0.829 | 0.973 | 0.946 | 0.971 | 0.939 |

Table 4: Empirical powers of the tests for a change in the mean and median with $\lambda^* = 0.95$

| $\rho$ | $\mu$ | Mean |   | Median |   |
|-------|-------|------|---|--------|---|
|       |       | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ |
| 0.0   | 0.5   | 0.052 | 0.146 | 0.152 | 0.094 | 0.093 | 0.075 | 0.169 | 0.152 | 0.111 | 0.090 |
| 0.3   | 0.5   | 0.055 | 0.120 | 0.085 | 0.080 | 0.059 | 0.076 | 0.138 | 0.104 | 0.109 | 0.080 |
| 0.6   | 0.5   | 0.099 | 0.348 | 0.313 | 0.162 | 0.149 | 0.101 | 0.313 | 0.263 | 0.185 | 0.148 |
| 0.9   | 0.5   | 0.189 | 0.665 | 0.596 | 0.413 | 0.296 | 0.129 | 0.549 | 0.526 | 0.312 | 0.225 |
| 1.2   | 0.5   | 0.331 | 0.851 | 0.808 | 0.642 | 0.467 | 0.251 | 0.767 | 0.718 | 0.506 | 0.357 |
| 1.5   | 0.5   | 0.536 | 0.952 | 0.925 | 0.800 | 0.652 | 0.478 | 0.915 | 0.868 | 0.711 | 0.500 |
| 1.8   | 0.5   | 0.687 | 0.983 | 0.967 | 0.903 | 0.783 | 0.634 | 0.958 | 0.948 | 0.844 | 0.627 |
| 2.0   | 0.5   | 0.873 | 1.000 | 0.995 | 0.977 | 0.919 | 0.846 | 0.995 | 0.995 | 0.995 | 0.934 |
| 0.3   | 1.0   | 0.049 | 0.097 | 0.056 | 0.067 | 0.046 | 0.069 | 0.116 | 0.068 | 0.098 | 0.064 |
| 0.6   | 1.0   | 0.069 | 0.188 | 0.116 | 0.117 | 0.078 | 0.082 | 0.211 | 0.136 | 0.142 | 0.092 |
| 0.9   | 1.0   | 0.080 | 0.322 | 0.234 | 0.176 | 0.108 | 0.083 | 0.354 | 0.206 | 0.184 | 0.118 |
| 1.2   | 1.0   | 0.095 | 0.497 | 0.375 | 0.254 | 0.179 | 0.089 | 0.448 | 0.347 | 0.235 | 0.153 |
| 1.5   | 1.0   | 0.155 | 0.668 | 0.552 | 0.432 | 0.265 | 0.155 | 0.637 | 0.522 | 0.369 | 0.230 |
| 1.8   | 1.0   | 0.202 | 0.789 | 0.705 | 0.543 | 0.337 | 0.233 | 0.752 | 0.682 | 0.495 | 0.325 |
| 2.0   | 1.0   | 0.423 | 0.952 | 0.860 | 0.747 | 0.545 | 0.446 | 0.906 | 0.866 | 0.675 | 0.441 |
Table 5: Empirical powers of the tests for a change in the mean and median with $\lambda^* = 0.98$

| $\rho$ | $\mu$ | Mean | Median |
|-------|-------|------|--------|
|       |       | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ | SN  | LASN$_{0.05}$ | WSN$_{0.05}$ | LASN$_{0.1}$ | WSN$_{0.1}$ |
| 0.0   | 0.5   | 0.054 | 0.044 | 0.058 | 0.043 | 0.056 | 0.063 | 0.083 | 0.079 | 0.077 | 0.063 |
| 1.0   | 0.044 | 0.084 | 0.088 | 0.056 | 0.057 | 0.056 | 0.124 | 0.091 | 0.082 | 0.063 |
| 2.0   | 0.067 | 0.379 | 0.354 | 0.179 | 0.124 | 0.061 | 0.266 | 0.191 | 0.152 | 0.054 |
| 3.0   | 0.209 | 0.717 | 0.634 | 0.480 | 0.237 | 0.135 | 0.563 | 0.257 | 0.357 | 0.032 |
| 5.0   | 0.652 | 0.964 | 0.946 | 0.883 | 0.510 | 0.628 | 0.921 | 0.196 | 0.834 | 0.001 |
| 7.0   | 0.870 | 1.000 | 0.997 | 0.976 | 0.760 | 0.910 | 0.985 | 0.122 | 0.974 | 0.000 |
| 9.0   | 0.965 | 1.000 | 1.000 | 0.996 | 0.917 | 0.981 | 1.000 | 0.046 | 0.996 | 0.000 |
| 0.3   | 0.056 | 0.055 | 0.063 | 0.054 | 0.067 | 0.060 | 0.087 | 0.062 | 0.083 | 0.052 |
| 1.0   | 0.045 | 0.070 | 0.079 | 0.054 | 0.050 | 0.055 | 0.103 | 0.074 | 0.075 | 0.058 |
| 2.0   | 0.058 | 0.177 | 0.182 | 0.091 | 0.087 | 0.062 | 0.177 | 0.139 | 0.103 | 0.073 |
| 3.0   | 0.078 | 0.432 | 0.382 | 0.184 | 0.142 | 0.065 | 0.344 | 0.158 | 0.186 | 0.038 |
| 5.0   | 0.307 | 0.832 | 0.765 | 0.624 | 0.303 | 0.386 | 0.762 | 0.152 | 0.632 | 0.013 |
| 7.0   | 0.614 | 0.973 | 0.927 | 0.872 | 0.479 | 0.730 | 0.943 | 0.078 | 0.890 | 0.000 |
| 9.0   | 0.806 | 0.999 | 0.979 | 0.953 | 0.689 | 0.934 | 0.984 | 0.038 | 0.967 | 0.000 |

4.2 Real Example

We shall here apply the new location-adaptive testing procedure to detect potential change points in the financial market. In particular, we consider the close price of the NASDAQ Composite Index from 2018-01-02 to 2020-06-30 with totally 628 observations. All data were collected from finance.yahoo.com on December 11, 2020. To begin with, we calculate the log-returns of the close prices, which are denoted by \{r_t\}$_{t=1}^{628}$, and investigate the possible changes in the variance. The log-returns are shown in Figure 3. Let $\hat{\theta}_{t_1,t_2}$ be the sample variance of \{r_{t_1}, \ldots, r_{t_2}\} and then the corresponding test statistic $\tilde{G}_{t_1}^{LA}$ is defined in (3.4), where we set $\epsilon = 0.05$ and $\delta = 0$.

In practice, since the actual number of change points is unknown in advance, we adopt the following testing procedure similar to the mechanism of the sequential monitoring:

Step 1. Test for a change in the variance based on the initial $m$ samples \{r_t\}$_{t=1}^{m}$, where $m$ is chosen to be 191 corresponding to approximately three quarters (i.e., $255 \times 0.75$);
Step 2. If the null hypothesis is rejected in Step 1, we have

$$(j_1^*, j_2^*, j_3^*) = \arg \max_{(j_1, j_2, j_3) \in \Omega_m(\epsilon)} \{D_m'(j_1, j_2, j_3)V_m^{-1}(j_1, j_2, j_3)D_m(j_1, j_2, j_3)\},$$

where $j_2^*$ is the estimated change-point location, and then we go back to Step 1 and do the test for the new sequence $\{r_t\}_{j_2^* + m}^{t+1}$; otherwise, we do Step 3.

Step 3. If the null hypothesis is not rejected in Step 1, we repeat Step 1 and test for the augmented sequence $\{r_t\}_{t=1}^{m+\Delta}$, where $\Delta$ is chosen to be 10 corresponding to approximately two weeks.

The above testing procedure illustrates the application of the proposed test in the sequential monitoring field. Table 6 shows the results of this procedure and the estimated change-point locations are displayed in Figure 3. The first row in Table 6, for instance, is obtained as follows. We start by implementing the LASN test to the first 191 observations $\{r_t\}_{t=1}^{191}$ and the associated $p$-value is larger than the nominal value 0.05. So we do not reject the null and run Step 1 again by testing for the augmented observations $\{r_t\}_{t=1}^{201}$. When the samples $\{r_t\}$ range from 2018-01-02 ($t = 1$) to 2018-10-17 ($t = 201$), the $p$-value of the LASN test is less than $10^{-3}$ for the first time. Then we reject the null hypothesis and find the first change point at the date 2018-10-05 ($t = 193$) with the location $\lambda^* = 193/201 \approx 0.960$. After that, the new searching procedure starts from 2018-10-08 ($t = 194$) and then we can similarly get the rest of the results in Table 6. For a comparison, the $p$-values of the traditional SN-based test as (2.3) are also provided in Table 6. It is interesting to observe that when the locations of change points are found to be close to the end of the current sequence as described in the first and third rows of Table 6, the proposed test suggests rejecting the null while the SN test advocates the acceptance. In the scenario that the changes are observed around the middle of the sequence, both the two methods consider rejecting the null. As seen from Figure 3, our proposed test might well capture the patterns of the NASDAQ Composite Index data.

5 Conclusion

In this article, we propose a new SN-based testing procedure for detecting structural changes in time series. The proposed location-adaptive testing method is specifically tai-
Table 6: Rejection dates (End date), \( p \)-values of tests under consideration, and the corresponding estimates of the change-point locations (\( \lambda^* \)) for the NASDAQ Composite Index data.

| Start date | Change-point | End date | \( P \)-value of \( \text{LASN}_{0.05} \) | \( P \)-value of SN | \( \lambda^* \) |
|------------|--------------|----------|-----------------|-----------------|------------|
| 2018-01-02 \((t = 1)\) | 2018-10-05 \((t = 193)\) | 2018-10-17 \((t = 201)\) | \(< 10^{-3} \) | 0.390 | 0.960 |
| 2018-10-08 \((t = 194)\) | 2019-01-11 \((t = 259)\) | 2019-07-12 \((t = 384)\) | \(< 10^{-3} \) | \(< 10^{-3} \) | 0.346 |
| 2019-01-14 \((t = 260)\) | 2020-01-23 \((t = 518)\) | 2020-02-25 \((t = 540)\) | 0.005 | 0.999 | 0.922 |

explored for the change-point testing problem, especially when the locations of changes are close to the starting or terminal point of a time series. To be precise, when the suspected location is close to the starting or end point of the sequence, we suggest ignoring the samples far away from this location such that the sample size on one side of the location is twice of the other, leading to a relatively balanced data partition. As expected, the proposed location-adaptive SN-based test inherits the monotonic power and stable size features in the traditional SN-based methods. Moreover, compared to many existing approaches, the new test can significantly improve the power performance and meanwhile, it maintains a reasonable level of size. In addition, it can be widely applied to tests for quantities of interest, such as the vector parameter of a time series regression model, as long as the Assumption 3.1 is satisfied. On the other hand, a new estimator for the location of change point is proposed based on the core idea of the location-adaptive scheme. Furthermore, as a byproduct of importance in the Appendix, we justify theoretically the reason of the phenomenon that the classical SN-based method would suffer from serious losses in power when one is given an incorrect number of change points. This further complements the seminal work by [Shao and Zhang, 2010] and [Zhang and Lavitas, 2018].

One direction for future research is to develop the related method in the sequential monitoring change-point detection as considered by [Dette and Gösmann, 2020] and [Chan, Ng, and Yau, 2021], where a stable phase is available and data arrive consecutively. As our method is powerful for the cases that change points occur near the end-points, we anticipate that the proposed location-adaptive SN scheme might be able to efficiently improve the existing SN tests for the sequential monitoring. The other interesting but challenging
problem is to find a data-driven trimming parameter $\epsilon$ or $\delta$ as there is no appropriate strategy in the literature so far as we know. On the other hand, the method of verifying Assumption 3.1 for parameter estimation in concrete time series models is also needed to be considered in future research. In particular, it is possible to extend the asymptotic normality of a class of robust self-weighted LAD estimators [Zhu and Ling, 2015, Yang and Ling, 2017] to its sequential version and then solve the change-point detection problem in heavy-tailed time series. Further research along these directions is well underway.

**Appendix: A Power Loss Problem**

We shall here show a theorem that might provide more insights into the power loss problems in the SN-based methods, especially when one is given a wrong pre-specified number of structural breaks. For simplicity and readability, we only show the results associated with tests for the mean. The following two situations on power losses are studied:

1. First, [Shao and Zhang, 2010] has attempted a direct application of the SN idea
in [Lobato, 2001] under the change-point setting and proposed one naive SN test

$$NG_n = \max_{k=1,\ldots,n-1} Z^2_n(k)/V_n(n),$$

where $V_n(n) = n^{-1} \sum_{k=1}^n Z^2_n(k)$. As shown in Figure 1 of [Shao and Zhang, 2010], if there exists one change point (i.e., $H_1$), then the test $NG_n$ is completely powerless.

(2) Moreover, [Zhang and Lavitas, 2018] has further pointed out that the SN-based test $G_n$ in (2.3) will suffer from serious power losses when there exist two change points (i.e., $H_2$).

However, to our knowledge, there is no theoretical justification to account for these two phenomena. We now present the following theorem.

**Theorem A.1.** Suppose that the condition (2.1) holds. It follows that

(i) under $H_1$, suppose $\mu_1 = EX_1$ and $\mu_2 = EX_{k^*+1}$ are two different constants, and then we have

$$NG_n \xrightarrow{p} 3,$$

as $n \to \infty$;

(ii) under $H_2$, suppose $\mu_1 = EX_1$, $\mu_2 = EX_{k^*_1+1}$ and $\mu_3 = EX_{k^*_2+1}$ are three fixed constants with $\mu_1 \neq \mu_2$ and $\mu_2 \neq \mu_3$, and then we have

$$G_n \xrightarrow{p} \sup_{r \in [0,1]} C(r),$$

as $n \to \infty$, where $C(r)$ is a continuous positive function only relying on $\lambda^*_1$, $\lambda^*_2$ and $d = (\mu_3 - \mu_2)/(\mu_1 - \mu_2)$. The concrete form of $C(r)$ is given in the Supplementary.

The proof of Theorem A.1 is presented in the Supplementary Materials. The asymptotic results in (i) and (ii) of Theorem A.1 correspond to the statistics in the cases (1) and (2). When there are two change points, a direct application of the test $G_n$ would lead to a severe power loss because the statistic $G_n$ under the alternative converges to a constant $\sup_{r \in [0,1]} C(r)$ as in (ii) of Theorem A.1. In other words, if the constant is smaller than the critical value, its power shall decrease to zero. Intuitively, the emergence of the asymptotic constant is due to the testing scheme that involves all samples $\{X_1, \ldots, X_n\}$ containing two change points. As a consequence, for any $k$, both of the numerator $Z^2_n(k)$ and the denominator $V_n(k)$ are very large. For the naive test $NG_n$, the reason of the power loss is similar.
Supplementary Materials

Supplementary materials present technical proofs of theoretical results and some additional numerical results.

Acknowledgements

Linlin Dai’s research was supported by the Fundamental Research Funds for the Central Universities (Nos. JBK1806002) of China and the National Natural Science Foundation of China (Nos. 12001441). Rui She’s research was supported by Fundamental Research Funds for the Central Universities (Nos. JBK1806002 and JBK190905) of China.

References

[Aminikhanghahi and Cook, 2017] Aminikhanghahi, S., and Cook, D. J. (2017), “A Survey of Methods for Time Series Change Point Detection,” Knowledge and Information Systems, 51, 339–367.

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

[Aue et al., 2008] Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2008), “Testing for Changes in Polynomial Regression,” Bernoulli, 14, 637–660.

[Berkes et al., 2004] Berkes, I., Gombay, E., Horváth, L., and Kokoszka, P. (2004), “Sequential Change-Point Detection in GARCH(p,q) Models,” Econometric Theory, 20, 1140–1167.

[Billingsley, 1999] Billingsley, P. (1999), Convergence of Probability Measures, 2nd Ed, New York: Wiley.

[Cavaliere, Georgiev, and Taylor, 2018] Cavaliere, G., Georgiev, I., and Taylor, A. M. R. (2018), “Unit Root Inference for Non-Stationary Linear Processes Driven by Infinite Variance Innovations,” Econometric Theory, 34, 302–348.
[Chan, Ng, and Yau, 2021] Chan, N. H., Ng, W. L., and Yau, C. Y. (2021), “A Self-Normalized Approach to Sequential Change-Point Detection for Time Series,” *Statistica Sinica*, 31, 491–517.

[Crainiceanu and Vogelsang, 2007] Crainiceanu, C. M., and Vogelsang, T. J. (2007), “Spectral Density Bandwidth Choice: Source of Nonmonotonic Power for Tests of a Mean Shift in a Time Series,” *Journal of Statistical Computation and Simulation*, 77, 457–476.

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

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

[Dette and Gösmann, 2020] Dette, H., and Gösmann, J. (2020), “A Likelihood Ratio Approach to Sequential Change Point Detection for a General Class of Parameters,” *Journal of the American Statistical Association*, 115, 1361–1377.

[Dette, Kokot, and Volgushev, 2020] Dette, H., Kokot, K., and Volgushev, S. (2020), “Testing Relevant Hypotheses in Functional Time Series via Self-Normalization,” *Journal of the Royal Statistical Society, Series B*, 82, 629–660.

[Gromenko, Kokoszka, and Reimherr, 2017] Gromenko, O., Kokoszka, P., and Reimherr, M. (2017), “Detection of Change in the Spatiotemporal Mean Function,” *Journal of the Royal Statistical Society, Series B*, 79, 29–50.

[Horváth et al., 2020] Horváth, L., Liu, Z., Rice, G., and Wang, S. (2020), “Sequential Monitoring for Changes from Stationarity to Mild Non-Stationarity,” *Journal of Econometrics*, 215, 209–238.

[Huang, Volgushev, and Shao, 2015] Huang, Y., Volgushev, S., and Shao, X. (2015), “On Self-Normalization for Censored Dependent Data,” *Journal of Time Series Analysis*, 36, 109–153.
[Jiang, Zhao, and Shao, 2021] Jiang, F., Zhao, Z., and Shao, X. (2021), “Time Series Analysis of COVID-19 Infection Curve: A Change-Point Perspective,” *Journal of Econometrics*, in press, 2021.

[Kiefer, Vogelsang, and Bunzel, 2000] Kiefer, N. M., Vogelsang, T. J., and Bunzel, H. (2000), “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68, 695–714.

[Liu and Wu, 2010] Liu, W., and Wu, W. B. (2010), “Asymptotics of Spectral Density Estimates,” *Econometric Theory*, 26, 1218–1245.

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

[Michael, Colin, and Robert, 2016] Michael, W. R., Colin, M. G., and Robert, B. L. (2016), “A General Regression Changepoint Test for Time Series Data,” *Journal of the American Statistical Association*, 111, 670–683.

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

[Rho and Shao, 2015] Rho, Y., and Shao, X. (2015), “Inference for Time Series Regression Models With Weakly Dependent and Heteroscedastic Errors,” *Journal of Business and Economic Statistics*, 33, 444–457.

[Shao, 2010] Shao, X. F. (2010), “The Dependent Wild Bootstrap,” *Journal of the American Statistical Association*, 105, 218–235.

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

[Shao and Zhang, 2010] Shao, X. F., and Zhang, X. Y. (2010), “Testing for Change Points in Time Series,” *Journal of the American Statistical Association*, 105, 1228–1240.

[Su and White, 2010] Su, L., and White, H. (2010), “Testing Structural Change in Partially Linear Models,” *Econometric Theory*, 26, 1761–1806.
van der Vaart and Wellner, 1996] van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer Verlag.

[Vogelsang, 1999] 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.

[Volgushev and Shao, 2014] Volgushev, S., and Shao, X. (2014), “A General Approach to the Joint Asymptotic Analysis of Statistics from Sub-Samples,” *Electronic Journal of Statistics*, 8, 390–431.

[Xu, 2012] Xu, K.-L. (2012), “Robustifying Multivariate Trend Tests to Nonstationary Volatility,” *Journal of Econometrics*, 169, 147–154.

[Yang and Ling, 2017] Yang, Y., and Ling, S. (2017), “Self-Weighted LAD-Based Inference for Heavy-Tailed Threshold Autoregressive Models,” *Journal of Econometrics*, 197, 368–381.

[Zhang and Chan, 2020] Zhang, R. M., and Chan, N. H. (2020), “Nonstationary Linear Processes with Infinite Variance GARCH Errors,” *Econometric Theory*, 0, 1–34.

[Zhang, 2018] Zhang, T. (2018), “A Thresholding-Based Prewhitened Long-Run Variance Estimator and Its Dependence-Oracle Property,” *Statistica Sinica*, 28, 319–338.

[Zhang and Lavitas, 2018] Zhang, T., and Lavitas, L. (2018), “Unsupervised Self-Normalized Change-Point Testing for Time Series,” *Journal of the American Statistical Association*, 113, 637–648.

[Zhao and Li, 2013] Zhao, Z., and Li, X. (2013), “Inference for Modulated Stationary Processes,” *Bernoulli*, 19, 205–227.

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

[Zhou and Shao, 2013] Zhou, Z., and Shao, X. (2013), “Inference for Linear Models with Dependent Errors,” *Journal of the Royal Statistical Society, Series B*, 75, 323–343.
[Zhu and Ling, 2015] Zhu, K., and Ling, S. (2015), “LADE-Based Inference for ARMA Models with Unspecified and Heavy-Tailed Heteroscedastic Noises,” *Journal of the American Statistical Association*, 110, 784–794.
# The Supplementary Materials to “Location-Adaptive Change-Point Testing for Time Series”

## 1 Technical Proofs for Main Results

In this section, we give the details for Theorems 2.1–3.1. Notice that the proposed tests $G^L_n$ and $Q^L_n$ can be regarded as the functionals of $D_n(j_1, j_2, j_3)$ and $V_n(j_1, j_2, j_3)$. Thus, we give the following lemma to describe their asymptotic properties.

**Lemma S.1.** Under the null hypothesis, if the condition (2.1) holds, then we have

\[
\sup_{(r,s,t) \in \mathcal{F}_0} |D_n(\lfloor nr \rfloor \vee 1, \lfloor ns \rfloor \vee 1, \lfloor nt \rfloor) - D(S_n, r, s, t)| = o_p(1), \quad (S1)
\]

and

\[
\sup_{(r,s,t) \in \mathcal{F}_0} |V_n(\lfloor nr \rfloor \vee 1, \lfloor ns \rfloor \vee 1, \lfloor nt \rfloor) - V(S_n, r, s, t)| = o_p(1), \quad (S2)
\]

where $\mathcal{F}_0 = \{(r, s, t) : r \leq s \leq t, t - r \geq \epsilon\}$ with $\epsilon > 0$, and $D(S_n, r, s, t)$ and $V(S_n, r, s, t)$ are defined similarly as $D(B, r, s, t)$ and $V(B, r, s, t)$ in Theorem 2.1.

**Proof of Lemma S.1.** First, we consider the proof for (S1). Under the null hypot...
esis, it follows that
\[
D_n([nr] \lor 1, [ns] \lor 1, [nt]) = \frac{\sqrt{n}}{([nt] - [nr] \lor 1 + 1)^{3/2}} \times \left\{ ([nt] - [ns] \lor 1) \times [S_n([ns] \lor 1) - S_n([nr] - 1 \lor 0)] \\
- ([ns] \lor 1 - [nr] \lor 1 + 1) \times [S_n([nt]) - S_n([ns] \lor 1)] \right\}. \tag{S3}
\]
Note that the condition (2.1) implies that \( \{S_n([n\tau])\} \) is tight on the Skorohod space, namely, for any \( \delta_1, \delta_2 > 0 \), there exists \( \eta > 0 \) such that
\[
\limsup_{n \to \infty} P \left( \sup_{|\tau_1 - \tau_2| < \eta} |S_n([n\tau_1]) - S_n([n\tau_2])| > \delta_1 \right) > \delta_2.
\]
Then, for any given positive sequence \( \eta_n \to 0 \), we have
\[
\limsup_{n \to \infty} P \left( \sup_{|\tau_1 - \tau_2| < \eta_n} |S_n([n\tau_1]) - S_n([n\tau_2])| > \delta_1 \right) < \delta_2.
\]
Thus, it follows from the arbitrary of \( \delta_1 \) and \( \delta_2 \) that
\[
\sup_{|\tau_1 - \tau_2| < \eta_n} |S_n([n\tau_1]) - S_n([n\tau_2])| = o_p(1). \tag{S4}
\]
Let \( \eta_n = 2/n \) and by \( t - r \geq \epsilon \) and (S3), we have
\[
D_n([nr] \lor 1, [ns] \lor 1, [nt]) = \frac{\sqrt{n}}{([nt] - [nr] \lor 1 + 1)^{3/2}} \times \left\{ ([nt] - [ns] \lor 1) \times [S_n([ns]) - S_n([nr])] \\
- ([ns] \lor 1 - [nr] \lor 1 + 1) \times [S_n([nt]) - S_n([ns])] \right\} + o_p(1), \tag{S5}
\]
where \( o_p(1) \) means the convergence in probability to zero uniformly in \( F_0 \).

Furthermore, by \( t - r \geq \epsilon \) and (S5), it is straightforward to show that
\[
\sup_{(r,s,t) \in F_0} |D_n([nr] \lor 1, [ns] \lor 1, [nt]) - D(S_n, r, s, t)| \leq Cn^{-1} \sup_{\tau \in [0,1]} |S_n([n\tau])| + o_p(1),
\]
where \( C \) is a constant only relying on \( \epsilon \). By continuous mapping theorem and condition (2.1), it is obvious that \( \sup_{\tau \in [0,1]} |S_n([n\tau])| \) converges weakly to \( \sigma \sup_{\tau \in [0,1]} |B(\tau)| \), namely, \( \sup_{\tau \in [0,1]} S_n([n\tau]) = O_p(1) \). Therefore, (S1) holds.
Now, we focus on the proof for (S2). Notice that $V_n([nr] \lor 1, [ns] \lor 1, [nt])$ can be rewritten as

$$V_n([nr] \lor 1, [ns] \lor 1, [nt]) = \sum_{i=[nr]\lor1}^{[ns]\lor1} \frac{[ns] \lor 1 - [nr] \lor 1 + 1}{([nt] - [nr] \lor 1 + 1)^2} D_n^2([nr] \lor 1, i, [ns] \lor 1)$$

$$+ \sum_{i=[ns]\lor1+1}^{[nt]} \frac{[nt] - [ns] \lor 1}{([nt] - [nr] \lor 1 + 1)^2} D_n^2([ns] \lor 1 + 1, i, [nt])$$

$$\equiv V_{n1}([nr] \lor 1, [ns] \lor 1, [nt]) + V_{n2}([nr] \lor 1, [ns] \lor 1, [nt]).$$

Thus, it is sufficient to show that

$$\sup_{(r,s,t) \in F_0} |V_{n1}([nr] \lor 1, [ns] \lor 1, [nt]) - V_1(S_n, r, s, t)| = o_p(1), \quad \text{(S6)}$$

$$\sup_{(r,s,t) \in F_0} |V_{n2}([nr] \lor 1, [ns] \lor 1, [nt]) - V_2(S_n, r, s, t)| = o_p(1), \quad \text{(S7)}$$

where

$$V_1(S_n, r, s, t) = \frac{1}{(t-r)^2} \int_r^s \frac{\Delta^2(S_n, r, u, s)}{(s-r)^2} du,$$

$$V_2(S_n, r, s, t) = \frac{1}{(t-r)^2} \int_s^t \frac{\Delta^2(S_n, s, u, t)}{(t-s)^2} du$$

are the two parts of $V(S_n, r, s, t)$. Here, we only prove that the equation (S6) holds.

For simplicity, we divide the set $F_0$ as two parts:

$$F_1 = \{(r, s, t) : s-r \geq \delta; t-r \geq \epsilon\}; \quad F_2 = \{(r, s, t) : s-r < \delta; t-r \geq \epsilon\},$$

where $\delta > 0$ is one arbitrary positive number with $\delta < \epsilon/2$. When $(r, s, t) \in F_1$, since $[ns] > 1$, it is not hard to see that

$$\sum_{i=[nr]\lor1}^{[ns]} D_n^2([nr] \lor 1, i, [ns]) = n \int_r^s D_n^2([nr] \lor 1, [nu] \lor 1, [ns]) du$$

$$+ (nr - [nr]) D_n^2([nr] \lor 1, [nr] \lor 1, [ns])$$

$$+ (1 + [ns] - ns) D_n^2([nr] \lor 1, [ns], [ns]).$$

Meanwhile, by the condition (2.1) and continuous mapping theorem, and using (S1) and $s-r \geq \delta > 0$ and $t-r \geq \epsilon$, it follows that

$$V_{n1}([nr] \lor 1, [ns], [nt]) = \frac{n([ns] - [nr] \lor 1 + 1)}{([nt] - [nr] \lor 1 + 1)^2} \int_r^s D^2(S_n, r, u, s) du + o_p(1).$$
Therefore, by the definition of $D(S_n, r, u, s)$ and the fact $t - r \geq \epsilon$, we have

$$\sup_{(r, s, t) \in \mathcal{F}_1} |V_1([nr] \vee 1, [ns] \vee 1, [nt]) - V_1(S_n, r, s, t)| = o_p(1). \quad (S9)$$

When $(r, s, t) \in \mathcal{F}_2$, by the similar procedure for $(S5)$ and Cauchy inequality, it follows

$$\sup_{(r, s, t) \in \mathcal{F}_2} |V_1(S_n, r, s, t) - S_n([nt_1]) - S_n([nt_2])|^2 \leq \frac{4n^2}{|t - r|} \sup_{|\tau_1 - \tau_2| < 2\delta} |S_n([nt_1]) - S_n([nt_2])|^2 \quad (S10)$$

and

$$\sup_{(r, s, t) \in \mathcal{F}_2} |V_1(S_n, r, s, t)| \leq \frac{4}{t - r} \sup_{|\tau_1 - \tau_2| < 2\delta} |S_n([nt_1]) - S_n([nt_2])|^2. \quad (S11)$$

Therefore, by the fact $t - r \geq \epsilon$ and the tightness of $S_n([nt])$ (see also $(S4)$) and the arbitrary of $\delta$, and using $(S9)$–$(S11)$, we have $(S6)$. This completes the whole proof of Lemma S.1.

**Proof of Theorem 2.1.** Under the null hypothesis, notice that $\Omega(\epsilon) \subset \mathcal{F}_0$, and thus Lemma S.1 holds. Since the Brownian motion is nowhere differential almost surely, it is direct to show that

$$P(\min_{(r, s, t) \in \mathcal{F}_0} V(B, r, s, t) > 0) = 1. \quad (S12)$$

Then, by the continuous mapping theorem and Lemma S.1, we have $G_n^{LA} \overset{d}{\rightarrow} G^{LA}$. On the other hand, under the alternative, since $\lambda^* \in (0, 1)$, there exists $(t_1, \lambda^*, t_3) \in \Omega(\epsilon)$ with $(\lambda^* - t_1) \wedge (t_3 - \lambda^*) \geq \delta$ for some $\delta > 0$. It follows from the assumption (2.1) that

$$D_n([nt_1] \vee 1, [n\lambda^*], [nt_3]) = \frac{([n\lambda^*] - [nt_1] \vee 1 + 1)([nt_3] - [n\lambda^*])}{([nt_3] - [nt_1] \vee 1 + 1)^{3/2}} \Delta_n + O_p(1). \quad (S13)$$

Recall that $V_n([nt_1] \vee 1, [n\lambda^*], [nt_3])$ is independent with $\Delta_n$. In a similar to the proof procedure for the null hypothesis, we have

$$V_n([nt_1] \vee 1, [n\lambda^*], [nt_3]) \overset{d}{\rightarrow} V(B, t_1, \lambda^*, t_3). \quad (S14)$$
As a result, by \( \sqrt{n}|\Delta_n| \to \infty \) and (S13)-(S14), we have

\[
G_n^{LA} \geq D_n^2([n\lambda^* \lor 1, |nt_3]) / V_n([n\lambda^* \lor 1, |nt_3]) \overset{p}{\to} \infty.
\]

This completes the proof of Theorem 2.1.

**Proof of Theorem 2.2.** Without loss of generality, we focus on the case \( \lambda^* \in [\epsilon, 1 - \epsilon] \). Firstly, for \( \eta > 0 \), we can divide the optimal subsample selection set \( \Omega(\epsilon) \) as follows:

\[
\Omega_1(\epsilon) = \left\{ (t_1, t_2, t_3) : t_2 \leq \lambda^* - \eta, (t_1, t_2, t_3) \in \Omega(\epsilon) \right\}
\]

\[
\Omega_2(\epsilon) = \left\{ (t_1, t_2, t_3) : t_2 \geq \lambda^* + \eta, (t_1, t_2, t_3) \in \Omega(\epsilon) \right\},
\]

and the associated sets \( \Omega_{n1}(\epsilon) \) and \( \Omega_{n2}(\epsilon) \) can be similarly defined as that in \( \Omega_n(\epsilon) \).

Therefore, to prove the conclusion, it is sufficient to show that,

\[
P \left( \max_{(j_1,j_2,j_3) \in \Omega_n(\epsilon)} \frac{D_n^2(j_1,j_2,j_3)}{V_n(j_1,j_2,j_3)} < \max_{(j_1,j_2,j_3) \in \Omega_n(\epsilon)} \frac{D_n^2(j_1,j_2,j_3)}{V_n(j_1,j_2,j_3)} \right) \to 1,
\]

as \( n \to \infty \), where \( i = 1, 2 \). Since their proof are close, we only show the details for \( i = 1 \). Notice that

\[
D_n(j_1,j_2,j_3) = \frac{(j_2 - j_1 + 1)(j_3 - j_2)}{(j_3 - j_1 + 1)^{3/2}} \{(X - EX)_{j_1,j_2} - (X - EX)_{j_2+1,j_3}\}
\]

\[
+ \frac{(j_2 - j_1 + 1)(j_3 - j_2)}{(j_3 - j_1 + 1)^{3/2}} (EX_{j_1,j_2} - EX_{j_2+1,j_3})
\]

\[
\equiv D_{n0}(j_1,j_2,j_3) + D_{n1}(j_1,j_2,j_3),
\]

where \( (X - EX)_{j_1,j_2} \) and \( EX_{j_1,j_2} \) is the sample mean of \( \{X_t - EX, t = j_1, \cdots, j_2\} \) and \( \{EX, t = j_1, \cdots, j_2\} \), respectively. By assumption (2.1), we can see that Lemma S.1 still holds for \( D_{n0}(j_1,j_2,j_3) \), namely,

\[
\sup_{(r,s,t) \in F_0} |D_{n0}([nr] \lor 1, [ns] \lor 1, [nt]) - D(S_n,r,s,t)| = o_p(1).
\]

So, \( D_{n1}(j_1,j_2,j_3) \) is the unique possible nonstationary part in \( D_n(j_1,j_2,j_3) \). Then, it is necessary to consider the following three different cases in \( \Omega_1(\epsilon) \):

\[
\begin{align*}
\Omega_{1,1}(\epsilon) &= \{(t_1, t_2, t_3) : t_3 \leq \lambda^*, (t_1, t_2, t_3) \in \Omega_1(\epsilon)\} \\
\Omega_{1,2}(\epsilon) &= \{(t_1, t_2, t_3) : \lambda^* < t_3 \leq \lambda^* + \delta_n, (t_1, t_2, t_3) \in \Omega_1(\epsilon)\} \\
\Omega_{1,3}(\epsilon) &= \{(t_1, t_2, t_3) : t_3 > \lambda^* + \delta_n, (t_1, t_2, t_3) \in \Omega_1(\epsilon)\},
\end{align*}
\]
where \( \delta_n = o(1) \) is a positive sequence satisfying the condition that

\[
n\delta_n \to \infty \text{ and } \sqrt{n|\Delta_n|\delta_n} \to \infty. \quad (S18)
\]

Now, we further discuss the magnitude of \( D_n^2(j_1, j_2, j_3)/V_n(j_1, j_2, j_3) \) in the corresponding sets \( \Omega_{n1,i}(\epsilon) = \{([nt_1] \lor 1, [nt_2] \lor 1, [nt_3]) : (t_1, t_2, t_3) \in \Omega_{1,i}(\epsilon) \} \) with \( i = 1, 2, 3 \).

When \((j_1, j_2, j_3) \in \Omega_{n1,1}(\epsilon)\), since \( j_3 \leq k^* \), it is obvious to see that

\[
D_n(j_1, j_2, j_3) = D_{n0}(j_1, j_2, j_3)
\]

\[
V_n(j_1, j_2, j_3) = \sum_{i=j_1}^{j_2} \frac{(j_2 - j_1 + 1)}{(j_3 - j_1 + 1)^2} D_{n0}^2(i, j_2) + \sum_{i=j_2+1}^{j_3} \frac{(j_3 - j_2)}{(j_3 - j_1 + 1)^2} D_{n0}^2(j_2 + 1, i, j_3).
\]

Then, by Lemma S.1, we can show that

\[
\max_{(j_1, j_2, j_3) \in \Omega_{n1,1}(\epsilon)} \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} \overset{d}{\rightarrow} \sup_{(t_1, t_2, t_3) \in \Omega_{1,1}(\epsilon)} \frac{D^2(B, t_1, t_2, t_3)}{V(B, t_1, t_2, t_3)}. \quad (S19)
\]

On the other hand, when \((t_1, t_2, t_3) \in \Omega_{1,2}(\epsilon) \cup \Omega_{1,3}(\epsilon) \) and \( t_2 \leq \epsilon \), by the definition of \( \Omega_1(\epsilon) \), there must exist a fixed positive number \( \delta \), we uniformly have \( \lambda^* - t_2 \geq \eta \) and \( 3\epsilon - \lambda^* \geq \delta \) (i.e., \( t_3 = 3\epsilon \)). Then, the consistency is easily derived by the classical method in Jiang et al. (2020) and Dette et al. (2020). Therefore, we restrict our attention on the case \( t_2 > \epsilon \).

When \((j_1, j_2, j_3) \in \Omega_{n1,2}(\epsilon)\) with \( t_2 > \epsilon \), since \( j_3 \) is larger than \( k^* \), it follows that

\[
D_n(j_1, j_2, j_3) = D_{n0}(j_1, j_2, j_3) - \frac{(j_2 - j_1 + 1)(j_3 - k^*)}{(j_3 - j_1 + 1)^{3/2}} \Delta_n. \quad (S20)
\]

By the Cauchy inequality and (S20) and \( j_2 < k^* \), it follows that

\[
D_n^2(j_1, j_2, j_3) \leq 2D_{n0}^2(j_1, j_2, j_3) + \frac{2(j_2 - j_1 + 1)^2(j_3 - k^*)^2}{(j_3 - j_1 + 1)^3} \Delta_n^2
\]

\[
V_n(j_1, j_2, j_3) \geq \sum_{i=j_1}^{j_2} \frac{(j_2 - j_1 + 1)}{(j_3 - j_1 + 1)^2} D_{n0}^2(i, j_2),
\]

\[6\]
Furthermore, by $t_2 - t_1 \geq \epsilon$, and using the above two inequalities and (S17)-(S18), we have

$$\max_{(j_1, j_2, j_3) \in \Omega_{n, 2}(\epsilon)} \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} = O_p(n \Delta_n^2 \delta_n^2). \quad (S21)$$

When $(j_1, j_2, j_3) \in \Omega_{n, 3}(\epsilon)$, since $j_3$ is still larger than $k^*$, then (S20) holds. For $V_n(j_1, j_2, j_3)$, it follows that

$$V_n(j_1, j_2, j_3) \geq \sum_{i=j_2+1}^{k^*} \frac{(j_3 - j_2)}{(j_3 - j_1 + 1)^2} D_n^2(j_2 + 1, i, j_3). \quad (S22)$$

Similar to (S20), we have

$$D_n(j_2 + 1, i, j_3) = D_n^0(j_2 + 1, i, j_3) - \frac{(i - j_2)(j_3 - k^*)}{(j_3 - j_2)^{3/2}} \Delta_n. \quad (S23)$$

Then, we have

$$\sum_{i=j_2+1}^{k^*} D_n^2(j_2 + 1, i, j_3) = \sum_{i=j_2+1}^{k^*} D_n^2(j_2 + 1, i, j_3) + \sum_{i=j_2+1}^{k^*} \frac{(i - j_2)^2(j_3 - k^*)^2}{(j_3 - j_2)^3} \Delta_n$$

$$- 2 \sum_{i=j_2+1}^{k^*} \frac{(i - j_2)(j_3 - k^*)}{(j_3 - j_2)^{3/2}} D_n^0(j_2 + 1, i, j_3) \Delta_n. \quad (S24)$$

Then, using the fact $t_2 \leq \lambda^* - \eta$ and (S17) and (S18), we can further show that

$$\sum_{i=j_2+1}^{k^*} \frac{(j_3 - j_2)}{(j_3 - j_1 + 1)^2} D_n^2(j_2 + 1, i, j_3) = (1 + o_p(1)) \frac{(j_3 - k^*)^2(k^* - j_3)^3}{3(j_3 - j_1)^2(j_3 - j_2)^2} \Delta_n. \quad (S25)$$

where $o_p(1)$ means convergence to 0 in probability uniformly in $\Omega_{n, 3}(\epsilon)$. Moreover, (S20) implies that

$$D_n^2(j_1, j_2, j_3) = (1 + o_p(1)) \frac{(j_2 - j_1)^2(j_3 - k^*)^2}{(j_3 - j_1)^3} \Delta_n. \quad (S26)$$

Then, since $t_2 - t_1 \geq \epsilon$ when $(t_1, t_2, t_3) \in \Omega_{n, 3}(\epsilon)$, and by (S25)-(S26) and (S22), it follows that, with probability to one,

$$\max_{(j_1, j_2, j_3) \in \Omega_{n, 3}(\epsilon)} \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} \leq \max_{(j_1, j_2, j_3) \in \Omega_{n, 3}(\epsilon)} \frac{3(j_3 - j_2)(j_2 - j_1)^2}{(k^* - j_2)^3(j_3 - j_1)} \leq C_0. \quad (S27)$$
where $C_0$ is a constant only relying on $\epsilon$ and $\eta$. Then, using the results (S19), (S21) and (S27) and (S18), we can imply that
\[
\max_{(j_1, j_2, j_3) \in \Omega_{n}(\epsilon)} \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} = O_p(n \Delta_n^2 \delta_n^2).
\] (S28)

On the other hand, from the proof for Theorem 2.1 (see (S13)-(S14)), it is obvious that
\[
\max_{(j_1, j_2, j_3) \in \Omega_{n}(\epsilon)} \frac{D_n^2(j_1, j_2, j_3)}{V_n(j_1, j_2, j_3)} \geq O_p(n \Delta_n).
\] (S29)

Then, (S15) holds as $\delta_n = o(1)$. This completes the whole proof.

**Proof of Theorem 2.3.** For Theorem 2.3, we only need to pay attention to the following key points: (1) Under the null, both $\cup_{r \in [2\epsilon, 1]} \Omega(\epsilon; 0, r)$ and $\cup_{s \in [0, 1-2\epsilon]} \Omega(\epsilon; s, 1)$ are the compact subsets of $F_0$; (2) under the alternative, the condition
\[
\min_{0 \leq i \leq m} (\lambda_{i+1}^* - \lambda_i^*) \geq \epsilon
\]
implies that there exists one $r$ such that $\lambda_r^*$ is the unique change-point in the location interval $[0, r]$. Hence, Theorem 2.3 follows by a similar argument for Theorem 2.1. □

**Proof of Theorem 3.1.** Under $H_0$, it follows from Assumption 3.1 that
\[
D_N([Nr] \lor 1, [Ns] \lor 1, [Nt]) = \left\{ \frac{\sqrt{N}([Nt] - [Ns] \lor 1)}{([Nt] - [Nr] \lor 1 + 1)^{3/2}} L_N(r, s; \theta_0^*) \right. \\
- \frac{\sqrt{N}([Ns] \lor 1 - [Nr] \lor 1 + 1)}{([Nt] - [Nr] \lor 1 + 1)^{3/2}} \\
\left. \times L_N(\frac{[Ns] \lor 1 + 1}{N}, t; \theta_0^*) \right\}.
\] (S30)

Then, according to tightness of $L_N(s, t; \theta_0^*)$, it is not hard to show that
\[
D_N([Nr] \lor 1, [Ns] \lor 1, [Nt]) = \left\{ \frac{\sqrt{N}([Nt] - [Ns] \lor 1)}{([Nt] - [Nr] \lor 1 + 1)^{3/2}} L_N(r, s; \theta_0^*) \right. \\
- \frac{\sqrt{N}([Ns] \lor 1 - [Nr] \lor 1 + 1)}{([Nt] - [Nr] \lor 1 + 1)^{3/2}} L_N(s, t; \theta_0^*) \\
+ o_p(1),
\]
where $\mathcal{F}_4 = \{(r, s, t) : t - r \geq \epsilon, (t - s) \land (s - r) \geq \delta\}$ and $o_p(1)$ means the convergence in probability to zero uniformly in $\mathcal{F}_4$. Furthermore, it follows that

$$
\sup_{(r, s, t) \in \mathcal{F}_4} |D_N([Nr] \lor 1, [Ns] \lor 1, [Nt]) - D(L_N(\cdot; \theta^*_0), r, s, t)| = o_p(1), \quad (S31)
$$

where $D(L_N(\cdot; \theta^*_0), r, s, t)$ is similarly defined as $D(L_0, r, s, t)$. As a result, the statement in Theorem 3.1 (i) follows by a similar argument for Theorem 2.1.

Moreover, under $H_1$, we focus on the unique triple $(t_1, \lambda^*, t_3) \in \Omega(\epsilon)$. By Assumption 3.1, we have

$$
D([Nt_1] \lor 1, [N\lambda^*], [Nt_3]) = -\frac{([N\lambda^*] - [Nt_1] \lor 1 + 1)([Nt_3] - [N\lambda^*])}{([Nt_3] - [Nt_1] \lor 1 + 1)^{3/2}}(\theta^*_1 - \theta^*_0) + O_p(1).
$$

In the meantime, given Assumption 3.1 and the invertibility of $V(L_i, t_1, \lambda^*, t_3)$, one can easily show that $V_N^{-1}([Nt_1] \lor 1, [N\lambda^*], [Nt_3])$ must be invertible and of the order $O_p(1)$ using the continuous mapping theorem. Then, $\sqrt{N}|\theta^*_1 - \theta^*_0| \to \infty$ implies the statement (ii) in Theorem 3.1.

On the other hand, when $\delta$ is set to be 0, by the additional assumption $\sup_t |L_i(t, t)| = 0$ almost surely and the tightness of $L_i(s, t)$ on $l^\infty(\mathbb{D}_0)$, it follows that for any $\delta_1, \delta_2 > 0$, there exists one $\eta > 0$, such that

$$
\limsup_{n \to \infty} P\left(\sup_{|s - t| < \eta} |L_i(s, t)| > \delta_1 \right) < \delta_2.
$$

Then, following the similar proof procedure for $(S10)$ and $(S11)$, the conclusion can be easily proved. This completes the whole proof.

2 Proof of Theorem A.1

We consider to show the two statement in Theorem A.1 separately.
Proof of Theorem A.1 (i). Under $H_1$, if the condition (2.1) holds, we have

$$\sup_{k \in 1, \ldots, n-1} |Z_n(k) - M_n(k)| = O_p(1),$$

where $M_n(k) = n^{-1/2} \sum_{t=1}^{k} (EX_t - n^{-1} \sum_{t=1}^{n} EX_t)$. We further get that

(a) when $k > k^*$,

$$Z^2_n(k) = \frac{(k^*)^2}{n} \left(1 - \frac{k}{n}\right)^2 (\mu_1 - \mu_2)^2 + O_p(n^{1/2}); \quad (S32)$$

(b) when $k \leq k^*$,

$$Z^2_n(k) = \frac{k^2}{n} \left(1 - \frac{k}{n}\right)^2 (\mu_1 - \mu_2)^2 + O_p(n^{1/2}). \quad (S33)$$

Thus, it follows from $\max_{1 \leq k \leq n-1} Z^2_n(k) \geq Z^2_n(k^*)$ and (S32)-(S33) that

$$\max_{1 \leq k \leq n-1} Z^2_n(k)/n = \left(\frac{k^*}{n}\right)^2 \left(1 - \frac{k^*}{n}\right)^2 (\mu_1 - \mu_2)^2 + O_p(n^{-1/2}). \quad (S34)$$

Furthermore, according to (S32)-(S34) and $\mu_1 \neq \mu_2$, for $NG_n$, one can show that

$$NG_n = \frac{k^* n^{-2}(1 - k^* n^{-1})^2 + o_p(1)}{\sum_{k=1}^{k^*} k^2 n^{-3}(1 - k^* n^{-1})^2 + \sum_{k=k^*+1}^{n} k^* n^{-3}(1 - k n^{-1})^2 + o_p(1)}.$$

Note that $k^* = \lceil n \lambda^* \rceil$ and $\sum_{k=1}^{m} k^2 = m(m+1)(2m+1)/6$. Then, we have

$$NG_n = 3 + o_p(1) \xrightarrow{p} 3,$$

as $n \to \infty$. This completes the proof. \qed

Proof of Theorem A.1 (ii). Under $H_2$, denote $d_1 = \mu_1 - \mu_2$ and $d_2 = \mu_3 - \mu_2$. By the condition (2.1), it is straightforward to show that

(a) when $k \leq k^*_1$,

$$Z^2_n(k) = \frac{k^2}{n} \left[\left(1 - \frac{k^*_1}{n}\right) d_1 - \left(1 - \frac{k^*_2}{n}\right) d_2\right]^2 + O_p(n^{1/2}); \quad (S35)$$

$$NG_n = 3 + o_p(1) \xrightarrow{p} 3,$$
(b) when \(k^*_1 < k \leq k^*_2\),
\[
Z_n^2(k) = \left[ \frac{k^*_1}{\sqrt{n}} (1 - \frac{k}{n}) d_1 - \frac{k}{\sqrt{n}} (1 - \frac{k^*_2}{n}) d_2 \right]^2 + O_p(n^{1/2}); \quad (S36)
\]

(c) when \(k > k^*_2\),
\[
Z_n^2(k) = \left(1 - \frac{k}{n}\right)^2 \left[ \frac{k^*}{\sqrt{n}} d_1 - \frac{k^*_2}{\sqrt{n}} d_2 \right]^2 + O_p(n^{1/2}). \quad (S37)
\]

On the other hand, for the denominator \(V_n(k)\), we need do further analysis. Let \(S_{i,j}\) be the sum of subsamples \(X_i, \cdots, X_j\). Then, it follows that

**Case I:** when \(k = \lfloor nr \rfloor\) with \(r \leq \lambda^*_1\), we have that for \(t \leq k\),
\[
\frac{1}{n} (S_{t,n} - \frac{t}{k} S_{1,k}) = O_p(n^{-1/2});
\]
for \(k + 1 \leq t \leq k^*_1 + 1\),
\[
\frac{1}{n} (S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n}) = \frac{t-1-k}{n-k} \left[ \left(1 - \frac{k^*_1}{n}\right) d_2 - \left(1 - \frac{k^*_1}{n}\right) d_1 \right] + O_p(n^{-1/2});
\]
for \(k^*_1 + 1 < t \leq k^*_2 + 1\),
\[
\frac{1}{n} (S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n}) = \frac{(k^*_1 - k)(t-1-n)}{n(n-k)} d_1 + \frac{(t-1-k)(n-k^*_2)}{n(n-k)} d_2 + O_p(n^{-1/2});
\]
for \(k^*_2 + 1 < t \leq n\),
\[
\frac{1}{n} (S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n}) = \frac{t-1-n}{n(n-k)} \left[(k^*_1 - k)d_1 - (k^*_2 - k)d_2 \right] + O_p(n^{-1/2}).
\]

Thus, by the above four equalities and (S35), one can show that
\[
\sup_{r \in [0, \lambda^*_1]} |Z_n^2(\lfloor nr \rfloor) V_n^{-1}(\lfloor nr \rfloor) - C(r)| \xrightarrow{p} 0, \quad (S38)
\]
as \(n \to \infty\), where \(C(r)\) is equal to
\[
\frac{3r^2(1-r)[(1-\lambda^*_1) - (1-\lambda^*_2)d]^2}{[(\lambda^*_1 - r)(1-\lambda^*_1) - (1-\lambda^*_2)(\lambda^*_1 - r)d]^2 + (\lambda^*_1 - r)(1-\lambda^*_2)(\lambda^*_1 - \lambda^*_2)^2d^2}
\]
Case II: when \( k = \lfloor nr \rfloor \) with \( r > \lambda_2^* \), the proof is very similar. So we omit the details and directly write the result

\[
\sup_{r \in (\lambda_2^*, 1]} |Z_n^2([nr])V_n^{-1}([nr]) - C(r)| \xrightarrow{p} 0, \text{ as } n \to \infty, \tag{S39}
\]

where \( C(r) \) is equal to

\[
\frac{3r(1-r)[\lambda_1^* - \lambda_2^* d]^2}{[\lambda_1^*(r - \lambda_1^*) - \lambda_2^*(r - \lambda_2^*)d]^2 + \lambda_1^*(r - \lambda_2^*)(\lambda_1^* - \lambda_2^*)^2 d^2}.
\]

Case III: when \( k = \lfloor nr \rfloor \) with \( r \in (\lambda_1^*, \lambda_2^*) \), one has that for \( t \leq k_1^* \),

\[
\frac{1}{n} \left( S_{1,t} - \frac{t}{k} S_{1,k} \right) = \frac{t(k - k_1^*)}{nk} d_1 + O_p(n^{-1/2});
\]

for \( k_1^* < t \leq k \), we have

\[
\frac{1}{n} \left( S_{1,t} - \frac{t}{k} S_{1,k} \right) = \frac{k_1^*(k - t)}{nk} d_1 + O_p(n^{-1/2});
\]

for \( k < t \leq k_2^* + 1 \), we can get that

\[
\frac{1}{n} \left( S_{t,n} - \frac{n - t + 1}{n - k} S_{k+1,n} \right) = \frac{(t - 1 - k)(n - k_2^*)}{n(n - k)} d_2 + O_p(n^{-1/2});
\]

for \( t > k_2^* + 1 \),

\[
\frac{1}{n} \left( S_{t,n} - \frac{n - t + 1}{n - k} S_{k+1,n} \right) = \frac{(n + 1 - t)(k_2^* - k)}{n(n - k)} d_2 + O_p(n^{-1/2}).
\]

Then, by (S36), it is not hard to get that

\[
\sup_{r \in (\lambda_1^*, \lambda_2^*)} |Z_n^2([nr])V_n^{-1}([nr]) - C(r)| \xrightarrow{p} 0, \text{ as } n \to \infty, \tag{S40}
\]

where \( C(r) \) is equal to

\[
\frac{3r(1-r)[\lambda_1^*(1-r) - r(1-\lambda_2^*)d]^2}{(1-r)\lambda_1^*(\lambda_1^* - r)^2 + r(1-\lambda_2^*)^2(\lambda_2^* - r)^2 d^2}.
\]

Thus, by (S38)-(S40) and the continuous mapping theorem, the conclusion holds.
3 Additional Simulation for LASN$_0$

In this section, we further give the size performance of the test $G_{n}^{LA}$ when the trimming parameter is removed (i.e., $\epsilon = 0$). When the limiting distribution $G^{LA}$ is well-defined with $\epsilon = 0$, then its critical value at the nominal level 90% and 95% can be simulated as 90.463 and 117.465. Table 1 reports the empirical size of $G_{n}^{LA}$ when $\epsilon = 0$. We can observe that the size-distortion is very serious and its improvement is very limited as the sample size increases. In particular, when $\rho = 0.6$, the size performance becomes worse. This phenomenon provides evidence for the conjecture that the limiting distribution $G^{LA}$ under $\epsilon = 0$ may be divergent.

Table 1: Empirical size of the test $G_{n}^{LA}$ for a change in the mean with $\epsilon = 0$

| $n$  | $\alpha$ | -0.6 | -0.3 | 0   | 0.3 | 0.6 |
|------|----------|------|------|-----|-----|-----|
| 500  | 5%       | 0.011| 0.019| 0.032| 0.067| 0.116|
|      | 10%      | 0.015| 0.037| 0.061| 0.124| 0.177|
| 1000 | 5%       | 0.018| 0.021| 0.034| 0.063| 0.112|
|      | 10%      | 0.034| 0.044| 0.061| 0.103| 0.194|
| 2000 | 5%       | 0.023| 0.023| 0.046| 0.063| 0.144|
|      | 10%      | 0.046| 0.060| 0.079| 0.115| 0.227|

References

Dette, H., Kokot, K. and Volgushev, S. (2020), Testing relevant hypotheses in functional time series via self-normalization, *J. R. Stat. Soc. Ser. B*, 82, 629–660.

Jiang, F., Zhao, Z., and Shao, X. (2020), Time series analysis of COVID-19 infection curve: a change-point perspective, *J. Econom.*, in press, 2020.