The Annals of Statistics
2024, Vol. 52, No. 6, 2913–2930
https://doi.org/10.1214/24-AOS2451

CHANGE-POINT ANALYSIS WITH IRREGULAR SIGNALS

BY TOBIAS KLEY[1,a], YUHAN PHILIP LIU[2,b], HONGYUAN CAO[3,d] AND WEI BIAO WU[2,c]

[1]*Institute for Mathematical Stochastics, Georg-August-Universität Göttingen,* [a]*tobias.kley@uni-goettingen.de*
[2]*Department of Statistics, University of Chicago,* [b]*yuhanphilipliu@uchicago.edu,* [c]*wbwu@uchicago.edu*
[3]*Department of Statistics, Florida State University,* [d]*hcao@fsu.edu*

This paper considers the problem of testing and estimation of change point where signals after the change point can be highly irregular, which departs from the existing literature that assumes signals after the change point to be piecewise constant or vary smoothly. A two-step approach is proposed to effectively estimate the location of the change point. The first step consists of a preliminary estimation of the change point that allows us to obtain unknown parameters for the second step. In the second step, we use a new procedure to determine the position of the change point. We show that, under suitable conditions, the desirable $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point can be obtained. We apply our method to analyze the Baidu search index of COVID-19 related symptoms and find December 8, 2019, to be the starting date of the COVID-19 pandemic.

1. Introduction. Change-point detection and localization are classic and reviving topics in many dynamically evolving systems, where a sequence of measurements are recorded and we are interested in determining whether and at what time or location some aspect of the data, such as mean, variance or distribution changes (Page (1955, 1957)). This problem is of interest in many fields, such as economics, climatology, engineering, genomics, to name just a few. The last few decades witnessed enormous development on this topic from different perspectives including testing the existence of change points and the estimation of their locations. We refer to Csörgő and Horváth (1997), Aue and Horváth (2013), Jandhyala et al. (2013), Niu, Hao and Zhang (2016) for reviews and recent developments on this topic.

An important problem in the detection of structural breaks is the detection of mean changes. The simplest case where there is at most one change point has been studied extensively. The first step is to test whether there is any change point. If we reject the null hypothesis that there is no change point, the next step is to make inference on the location of the change point (Hawkins (1977)). The latter problem is nontrivial even for the normal and homoskedastic model or the one-parameter exponential family (Sen and Srivastava (1975), Hinkley (1970), Worsley (1986), Siegmund (1988)). Recently, the problem of detecting multiple change points has drawn a lot attention (Frick, Munk and Sieling (2014), Fryzlewicz (2014, 2018), Baranowski, Chen and Fryzlewicz (2019)). In particular, functions in the popular R package *changepoint* achieve linear computational cost when the number of change points increases with the number of observations (Killick, Fearnhead and Eckley (2012), Killick and Eckley (2014)). Here, it is assumed that, under the alternative, the mean function is piecewise constant. However, in certain applications, it is more plausible to assume that functions between a finite number of change points vary smoothly and/or the error process is dependent. Relevant statistical methods and theory can be found in Müller (1992), Horváth and Kokoszka (2002), Mallik et al. (2011), Mallik, Banerjee and Sen (2013), Vogt and Dette (2015), Dette, Eckle and Vetter (2020), Bücher, Dette and Heinrichs (2021).

Received December 2023; revised July 2024.
MSC2020 subject classifications. 62M10, 62G20.
Key words and phrases. Change-point analysis, COVID-19, invariance principle, irregular signals, long-run variance estimate, smoothing, weak convergence.

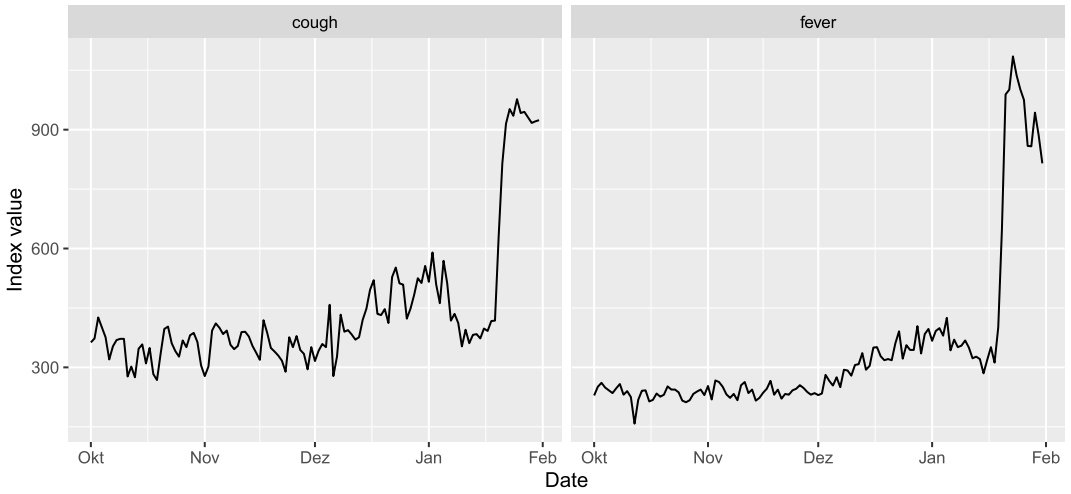

FIG. 1. *Daily Baidu search index values for keywords of COVID-related symptoms. Dates are from October* 1, 2019, *to January* 31, 2020. *Keywords "cough" and "fever" are shown in left and right panel, respectively.*

Unlike existing literature in the previous paragraph, we motivate our research from the fact that signals can be highly irregular after the change point in certain applications. Such irregular signals depart sharply from the constant mean or smoothly varying functions under the alternative—they can vary abruptly. A typical data example with irregular signals is depicted in Figure 1. In this data set, the total number of searches for COVID-19 related symptoms such as "fever" through Baidu (the most used search engine in China) is recorded in Hubei Province from October 1, 2019, to January 31, 2020. An extremely important problem in epidemiology is to identify the starting date of the pandemic. However, the latter problem is very difficult due to lack of access to the data, and researchers have different results on this; see, for example, Worobey (2021), Huang et al. (2020), Huang et al. (2020) and Centre for Disease Control and Prevention (2022) among others. In this paper, we shall address this important and fundamental problem by using the indirect Baidu search data. In particular, we are interested in inferring the start of the COVID-19 pandemic through changes of the Baidu search index by imposing the change-point paradigm (3) and (4). Based on the nature of the problem, it seems plausible to assume a constant mean before the change point. After the change point, it does not make sense to assume constant mean or smooth trend as the data exhibit a high level of variation and irregularity. As far as we know, no results in the current change-point literature can allow such a framework.

In this paper, we consider the testing and estimation of at most one change point where the null assumes a constant mean and signals under the alternative can be quite general. Our method is offline in the sense that we assume data are fully observed, which is different from online change-point detection where information accrue over time and we only have data available that has arrived before the current time. Different from Cao and Wu (2015), where the interest lies in the multiple testing with clustered signals, we propose new test statistics and a two-step method for detection of irregular signals after the change point. We use a CUSUM-type of statistic to test the global null hypothesis that there is no change point. If this global null is rejected, we develop a two-step method to locate the change point. In the first step, we use the minimum of the batched means as a rough estimation of the change-point location. Intuitively speaking, as data before the change point have a constant mean, this minimum falls between the time origin and the true change point. The batched mean effectively smooths out the data and increases the signal-to-noise noise ratio. Equipped with the preliminary estimation from the first step, we are able to estimate the constant mean

before the change point and the minimum distance between signals and the constant mean. This allows us to construct a new test statistic to get a refined estimation of the change point in the second step. We get, under suitable conditions, an $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point to the true change point, which fundamentally improves the results in Cao and Wu (2022), where multiple sequences are needed to estimate the variance due to heteroscedasticity.

The rest of the paper is organized as follows. In Section 2, we introduce our main methodology for the global test and the two-step method to locate the change point. Theoretical results are developed in Section 3, where we show Gaussian approximation to the test statistic and the desirable $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point to the true change point. In Section 4, we investigate the finite sample performance of the new method through simulations. In Section 5, we apply the new change-point estimation method to two data sets: the Baidu search index for COVID-19 related symptoms "fever" and "cough" in 2019–2020. All proofs and technical details are relegated to the Appendix in the Supplementary Material (Kley et al. (2024)).

2. Methodology.

2.1. *Model for the data and the problems considered.* Suppose we are given noisy data of the form

$$X_t = \mu_t + Z_t, \quad t = 1, \dots, n, \tag{1}$$

where μ_t are means or signals and $(Z_t)_{t\in\mathbb{Z}}$ is a stationary process with mean 0, autocovariance function $\gamma(k) = \mathrm{Cov}(Z_{t+k}, Z_t)$ and finite long-run variance

$$0 < \sigma_\infty^2 := \sum_{k=-\infty}^{\infty} \gamma(k) < \infty. \tag{2}$$

Consider the following null hypothesis:

$$H_0 : \mu_1 = \cdots = \mu_n, \tag{3}$$

where the signal is constant (i.e., all means μ_j are equal, but not necessarily zero) and the alternative hypothesis

$$H_1 : \exists \tau \in \{2, \dots, n\}, d > 0 : \mu_1 = \cdots = \mu_{\tau-1}, \quad \mu_\tau, \dots, \mu_n \geq \mu_1 + d, \tag{4}$$

where the signal is constant for the first $\tau - 1$ observations and the means from the τth observation onward may then vary arbitrarily as long as they are larger by at least d. We focus on the one-sided case, because such upward shifts to a higher, but nonconstant level are frequently encountered in practice; cf. Figure 1. Yet, to our knowledge, no method that is tailored to this important situation is available to date. Note that this setting includes the case where $\mu_\tau = \cdots = \mu_n \geq \mu_1 + d$, which is to be detected by many traditional methods. The case of multiple changes (as long as $\mu_j \geq \mu_1 + d$, $j \geq \tau$) is also covered. Then τ corresponds to the time of the earliest change. But paradigm (4) goes far beyond these specific cases. In fact, apart from the one-sidedness of the change in means, we do not require any structure and the signal after the change may be arbitrarily wild.

Given the observations $X_1, \dots, X_n$, we aim to develop:

- a hypothesis test to decide whether H_0 holds or H_1 holds (see Section 2.2), and
- a procedure that, under H_1, will estimate τ (see Section 2.3).

Note that Dette and Wu (2019), Heinrichs and Dette (2021), Vogt and Dette (2015), Bücher, Dette and Heinrichs (2021) considered the problem of detecting changes in a sequence of means. Smoothness assumptions are needed for their methods to work. Different from these works, as mentioned before, our methodology does not require such smoothness assumptions.

2.2. *Testing procedure.* Now, we propose a test to distinguish between H_0 and H_1, defined in (3) and (4), respectively, when we have $X_1, \dots, X_n$ that follow (1) available. We use the following quantity:

$$\hat{T} := \min_{j=1,2,\dots,n} \sum_{i=1}^{j} (X_i - \bar{X}_n)/(\sqrt{n}\hat{\sigma}_\infty), \quad \text{where } \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{5}$$

and $\hat{\sigma}_\infty^2$ is a consistent estimator for the long-run variance σ_∞^2, defined in (2). In Section 2.3.2, we propose an estimator for σ_∞^2 that is consistent both under H_0 and H_1. In Section 3.2, we will show that, under H_0, the distribution of $\hat{T}$ is close, asymptotically, to the distribution of a minimum obtained from a standard Brownian bridge, the quantiles of which can be obtained via simulation or approximated asymptotically.

We will further show that, under H_1, we have $\hat{T} \xrightarrow{P} -\infty$, as $n \to \infty$. Therefore, we will use $\{\hat{T} < c\}$ as the rejection area, where c can be chosen as the α-quantile of a minimum of a standard Brownian bridge; cf. Section 3.2.

2.3. *A two-step locating algorithm.*

2.3.1. *Blocking and estimation of the long-run variance.* To reduce the noise from the data and to focus our attention on the signal, we will here and in the following sections split the data set into $m := \lfloor n/k \rfloor$ blocks of size k where $k \to \infty$ and $k/n \to 0$. Then we calculate the blocks' sample means as follows:

$$R_j := \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} X_i, \quad j = 1, 2, \dots, m. \tag{6}$$

Our theoretical results and remarks provide guidance on the choice of k; cf. Section 3.3.

From the R_j, we then obtain

$$\hat{L} := \operatorname*{arg\,min}_{i=1,\dots,m} R_i, \qquad \hat{\ell} := k\hat{L}, \tag{7}$$

where $\hat{L}$ indicates an index of a block likely to have all observations in it prior to the change point and $\hat{\ell}$ points to the last observation in the $\hat{L}$th block. The observations $X_1, \dots, X_{\hat{\ell}}$ are approximately stationary and can be used to obtain an estimate for the long-run variance by computing any consistent estimator for the long-run variance from them. We provide an estimate in Section 2.3.2 and an asymptotic theory in Section 3.3. More generally, than (7), we can let $\hat{L} := \max\{i : R_i \le R_m^{(J)}\}$, for fixed J, where $R_m^{(J)}$ denotes the Jth smallest value among the block averages $R_1, \dots, R_m$. With this generalized definition, we have more data to estimate the long-run variance.

2.3.2. *Estimate for long-run variance.* Having blocked the data as delineated in Section 2.3.1, we derive the index $\hat{\ell}$, which satisfies $\hat{\ell} < \tau$ with high probability. To estimate the long-run variance, we suggest

$$\hat{\sigma}^2 := \frac{k}{\hat{\ell} - k + 1} \sum_{s=k}^{\hat{\ell}} (R_{s/k} - \hat{\mu}_0)^2, \quad \text{where } R_{s/k} := \frac{1}{k} \sum_{i=s-k+1}^{s} X_i, \hat{\mu}_0 := \frac{1}{\hat{\ell}} \sum_{i=1}^{\hat{\ell}} X_i, \tag{8}$$

where the definition of the overlapping block averages $R_{s/k}$ extends the one for the non-overlapping blocks in (6), and $\hat{\mu}_0$ is a preliminary estimate for μ_1. We use overlapping blocks

in this section, as this has been shown to reduce the asymptotic mean squared error; cf. Lahiri (1999). The estimate for the long-run variance is motivated by the fact that

$$\mathbb{E}\big[\big(\sqrt{k}(R_{s/k} - \mathbb{E}R_{s/k})\big)^2\big] \to \sigma_\infty^2, \quad s = 1, \dots, \tau - 1$$

together with $\tau > \hat{\ell} \to \infty$, with high probability (cf. Lemma A.1), and $\hat{\mu}_0$ being a consistent estimator for μ_1 (cf. Lemma A.3. Lemmas A.1 and A.3) such as all other references of the form A.x are in the Supplementary Material. Estimates of a similar nature were previously considered, for example, by Mies and Steland (2023), Zhou (2013) and Peligrad and Shao (1995). The important novelty of the estimate proposed in (8) lies in the data-dependent segment selection via the index $\hat{\ell}$. This ensures consistency under H_0 and H_1, but makes the rigorous analysis more challenging. At a technical level, proving consistency requires a maximal inequality for quadratic forms.

2.3.3. *Locating algorithm*: *Step* 1. The aim of step 1 is to obtain an improved estimate for μ_1 and an estimate toward d, which we describe as follows. We block the data as described in Section 2.3.1 and obtain the blockwise averages R_j, the index $\hat{L}$ that indicates a block with data from before the change, the index $\hat{\ell}$ that points to the last observation of the $\hat{L}$'s block and the preliminary estimate $\hat{\mu}_0$, defined in (8), for μ_1. We then compute "test statistics" $\hat{D}_j$ and "test decisions" $\hat{I}_j$ as

$$\hat{D}_j := \sqrt{k}(R_j - \hat{\mu}_0)/\hat{\sigma}_\infty \quad \text{and} \quad \hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/m}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The long-run variance estimate $\hat{\sigma}_\infty^2$ can be specified by the user. Our estimator $\hat{\sigma}^2$, defined in (8), is a canonical choice, but any consistent estimate can be used. The quantity z_α, $\alpha \in (0, 1)$, denotes the αth quantile of the standard normal distribution, and $m := \lfloor n/k \rfloor$, as before. We then compute

$$\begin{aligned} \hat{\eta} &:= \operatorname*{arg\,min}_{t=1,\dots,m-1} \sum_{j=1}^{m} \big(\hat{I}_j - 1_{[t+1,m]}(j)\big)^2 \\ &= \operatorname*{arg\,min}_{t=1,\dots,m-1} \left[\sum_{j=1}^{t} \hat{I}_j + \sum_{j=t+1}^{m} (1 - \hat{I}_j) \right]. \end{aligned} \tag{10}$$

Finally, we obtain the preliminary estimates for μ_1 and toward d by

$$\hat{\mu}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} X_i, \qquad \hat{d} := \min_{\substack{i=k(\hat{\eta}+1)+1, \\ \dots, n-k+1}} \frac{1}{k} \sum_{j=i}^{i+k-1} (X_j - \hat{\mu}_1). \tag{11}$$

Some comments on the motivation for these estimates are in order. Denote by $\eta := \lfloor \tau/k \rfloor$ the index of the last block for which the signal is still constant; that is, $\eta k + 1 \leq \tau \leq (\eta + 1)k$, where τ is the index of the change; cf. the alternative hypothesis (4) considered. Then we have $\mathbb{E}R_j = \mu_1$ for $j = 1, \dots, \eta$ and $\mathbb{E}R_j > \mu_1$ for $j = \eta + 1, \dots, m$. The intuition behind the estimate $\hat{\mu}_0$ is as follows. Since $R_1, \dots, R_\eta$ fluctuate about their common μ_1, but $R_{\eta+1}, \dots, R_m$ fluctuate about means that are strictly larger than μ_1, we have that $\hat{L}/\eta$ is stochastically bounded away from 0 and 1, namely, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\mathbb{P}(\hat{L}/\eta \in [\delta, 1 - \delta]) \geq 1 - \varepsilon$; cf. Lemma A.1. Thus, we expect to average $k\hat{L} \asymp k\eta$ of the prechange observations to obtain $\hat{\mu}_0$. In particular, if τ diverges at the same rate as n, then $\hat{\mu}_0$ can be expected to be a $\sqrt{n}$-consistent estimate for μ_1.

The rationale behind $\hat{\mu}_1$ is that by replacing $\hat{L}$ in $\hat{\mu}_0$ by $\hat{\eta}$ we will obtain an improved estimate for μ_1 as we expect the estimate $\hat{\eta}$ to be closer to η than $\hat{L}$. Note that once $\hat{\eta}$ is

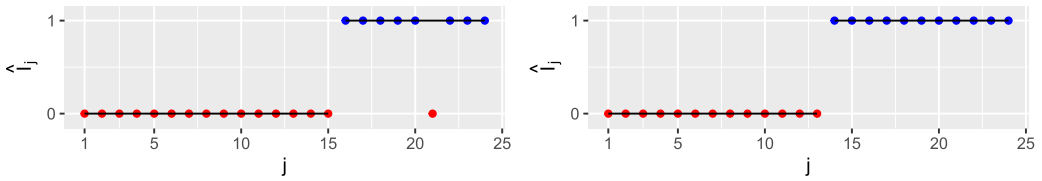

FIG. 2. *Test decisions $\hat{I}_j$ for the cough (left) and fever data (right) analysis described in Section 5. Fitted step function $j \mapsto 1_{[\hat{\eta}+1,m]}(j)$ is indicated by the solid black line.*

available we use $\hat{\eta}$ instead of $\hat{L}$. The intuition behind $\hat{\eta}$ is as follows. The test decisions $\hat{I}_j$ indicate whether a block is from before the change, where we have $\mathbb{E}R_j = \mu_1$, or after the change, where $\mathbb{E}R_j > \mu_1$. Thus, the sequence $\hat{I}_1, \ldots, \hat{I}_m$ of test decisions is an empirical version of the sequence $I_1, \ldots, I_m$ with $I_j := 0$, $j \le \eta$ and $I_j := 1$, $j > \eta$, which is known to be a step function. The estimate $\hat{\eta}$ is obtained by fitting a step function to the sequence $\hat{I}_1, \ldots, \hat{I}_m$ of test decisions that jumps from 0 to 1 at the block that includes the change. Fitting the step function can be seen as a different type of smoothing that we employ to reduce noise in the sequence of test decisions. A graphical illustration of the type of smoothing employed in the estimation of η for the empirical example of Section 5 is shown in Figure 2. The threshold z_α, employed in (9), is chosen with the quantile level $\alpha = 1 - 1/m$ tending to one to avoid too many false rejections among the blocks prior to the change.

2.3.4. *Locating algorithm*: *Step* 2. In this section, we define the novel estimate for the time τ where the change occurs; cf. (4). Consider

$$\hat{\tau} := \arg \min_{j=2,\ldots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho\hat{d}), \tag{12}$$

where $\rho \in (0, 1)$ is a tuning parameter. In Section 3.3, we provide rigorous theory for $\hat{\tau}$, which sheds light on the choice of the tuning parameter. As a rule of thumb, we use $\rho = 1/2$. It should be noted that, in a similar vein, Chen, Wang and Samworth (2022) also uses $1/2$ as a factor for an online change-point procedure based on likelihood ratio test statistics.

In Section 5, we study the fundamentally important and much debated problem of inferring the beginning of the COVID-19 pandemic. To this end, we employ our estimate $\hat{\tau}$ to search indices obtained from Baidu. We will see that our method reveals a plausible date, where traditional methods fail completely.

3. Theory.

3.1. *Assumptions on the noise process*. To derive meaningful results regarding the statistical properties of our proposed methods, some assumptions regarding the noise process $(Z_t)_{t\in\mathbb{Z}}$ that appears in model (1) are in order.

We employ the framework of functional dependence measure introduced in Wu (2005). In this framework, we view the causal stationary process $(Z_t)_{t\in\mathbb{Z}}$ as outputs from a physical system as follows:

$$Z_t = G(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{13}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$, i. i. d., is the input information of this system and G is an $\mathbb{R}$-valued measurable function that can be thought of as a filter or, intuitively, "mechanism" of this system. Many widely used, linear and nonlinear time series, including ARCH, threshold autoregressive, random coefficient autoregressive and bilinear autoregressive processes follow the framework

of (13); see, for example, Tong (1990), Priestley (1988), Wu (2011) among others. Then with this system, we measure the dependence from how much the outputs of this system will change if we replace the input information at time $t = 0$ with an i.i.d. copy ε_0'. Assume $\mathbb{E}|Z_i|^\theta < \infty$, $\theta \geq 1$. For a single observation at time i, we define the functional dependence measure as follows:

$$\delta_{i,\theta} = \big(\mathbb{E}|Z_i - Z_{i,\{0\}}|^\theta\big)^{1/\theta}, \quad \text{where } Z_{i,\{0\}} = G\big(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i\big). \tag{14}$$

To measure the temporal dependence for the whole time series, we define the cumulative dependence measure of $(Z_i)_{i\geq n}$ on ε_0:

$$\Theta_{n,\theta} = \sum_{i\geq n} \delta_{i,\theta}, \quad n \geq 0. \tag{15}$$

3.2. *Testing procedure.* We now state results on the test described in Section 2.2. The first result provides the asymptotic distribution under the null hypothesis, and the second one asserts asymptotic consistency under H_1, where we will allow $d = d_n$ and $\tau = \tau_n$ to depend on n without making this explicit in the notation.

THEOREM 3.1. *Assume that the short-range dependence condition holds*:

$$\Theta_{0,2} = \sum_{i\geq 0} \delta_{i,2} < \infty. \tag{16}$$

(i) *Under H_0, we have, as $n \to \infty$, that*

$$\sup_{x\leq 0} \big|\mathbb{P}(\hat{T} \leq x) - e^{-2x^2}\big| \to 0. \tag{17}$$

(ii) *Under H_1, assume $(\tau/n)(1-\tau/n)d\sqrt{n} \to \infty$. Then we have, as $n \to \infty$, that*

$$\hat{T} \to -\infty \quad \textit{in probability}.$$

The proof is deferred to Section A.1. Theorem 3.1(i) suggests that, given level $\alpha \in (0, 1)$, we can use the α-quantile of the limit, $-(-0.5\log\alpha)^{1/2}$, as the cutoff value to test H_0. Let $\mathbb{B}$ denote a standard Brownian motion and $\mathbb{B}_1(u) = \mathbb{B}(u) - u\mathbb{B}(1)$ be the Brownian bridge. For all $x \leq 0$, we have

$$\mathbb{P}\Big(\inf_{u\in[0,1]} \mathbb{B}_1(u) \leq x\Big) = e^{-2x^2}; \tag{18}$$

cf. equation (9.41) in Billingsley (1999). When n is relatively small, a refined approximation of $\mathbb{P}(\hat{T} \leq x)$ is $\mathbb{P}(T^\circ \leq x)$, where the discretized version $T^\circ = \min_{j\in\{1,2,\ldots,n\}} \mathbb{B}_1(j/n)$ and its distribution can be obtained by extensive simulation. The test based on the latter can have a more accurate performance. Theorem 3.1(ii) implies that for any $q \in \mathbb{R}$, $\mathbb{P}(\hat{T} \leq q) \to 1$.

3.3. *Locating algorithm.* To establish a convergence theory for the estimated change points, we will require the following assumption on temporal dependence.

CONDITION 3.1. $(Z_i)_{i\in\mathbb{Z}}$ satisfies that the θth moment $H_\theta := (\mathbb{E}|Z_i|^\theta)^{1/\theta} < \infty$, where $\theta > 2$. Assume that any one of the following holds:

- $\theta > 4$ and $\Theta_{n,\theta} = \mathcal{O}(n^{-\gamma_\theta}(\log n)^{-A})$, as $n \to \infty$, for $A > 2(1/\theta + 1 + \gamma_\theta)/3$, where
$$\gamma_\theta = \big(\theta^2 - 4 + (\theta-2)\sqrt{\theta^2 + 20\theta + 4}\big)/(8\theta);$$
- $2 < \theta \leq 4$ and $\Theta_{n,\theta} = \mathcal{O}(n^{-1}(\log n)^{-A})$, as $n \to \infty$, with $A > 3/2$.

Condition 3.1 holds, for example, under the geometric moment contraction $\delta_{n,\theta} = \mathcal{O}(\rho^n)$ for some $\rho \in (0,1)$, which is satisfied for many nonlinear time-series models; see, for example, Shao and Wu (2007) or Wu (2011). In general, it can be weaker since it allows polynomially decaying functional dependence measures. By Corollary 2.1 in Berkes, Liu and Wu (2014), Condition 3.1 implies the following optimal Komlós–Major–Tusnády result: on a possibly richer probability space $(\Omega_c, \mathcal{A}_c, \mathrm{P}_c)$, there exists $(Z_i^c)_{i\in\mathbb{Z}} \overset{\mathcal{D}}{=} (Z_i)_{i\in\mathbb{Z}}$, and a standard Brownian motion $\mathbb{B}_c(\cdot)$ such that

$$\sum_{i=1}^{n} Z_i^c = \sigma_\infty \mathbb{B}_c(n) + o_{a.s.}\big(n^{1/\theta}\big). \tag{19}$$

The next result asserts that $\hat{\sigma}^2$, defined in (8), is a consistent estimator for σ_∞^2. In the statement of the proof, we write $a_n \ll b_n$ or $b_n \gg a_n$ to mean that $a_n = o(b_n)$, as $n \to \infty$. The quantities $d = d_n$ and $\tau = \tau_n$ are the ones from (4), which we allow to depend on n without making this explicit in the notation, and k is the user-chosen block size; cf. Section 2.3.1. Further, $m := \lfloor n/k \rfloor$ and $\eta := \lfloor \tau/k \rfloor$, as before.

THEOREM 3.2. *Assume that Condition 3.1 holds, $d \gg n^{-1/\theta}$ and $n^{2/\theta}\log(n) \ll k \ll \tau$. Then*

$$\hat{\sigma}^2 = \sigma_\infty^2 + \mathcal{O}(1/k) + \mathcal{O}_{\mathbb{P}}\big(\eta^{2/\min(4,\theta)-1}\big), \quad \text{as } n \to \infty.$$

The proof is deferred to Section A.2.

REMARK 1. (i) When $\theta \geq 4$, the bound in Theorem 3.2 becomes $\mathcal{O}(1/k) + \mathcal{O}_{\mathbb{P}}(\eta^{-1/2})$. Intuitively, the $\mathcal{O}(1/k)$ and $\mathcal{O}_{\mathbb{P}}(\eta^{-1/2})$ can be seen as the rate for the bias and a centered version of the estimate $\hat{\sigma}^2$, respectively. The rate of the bias follows from $\sum_{u=-\infty}^{\infty} |u\gamma(u)| < \infty$, which is satisfied under Condition 3.1; cf. Lemma A.2.

(ii) The choice of the block size k, in Theorem 3.2, is limited by the rates of $n^{2/\theta}\log(n)$ from below and by τ from above. The lower bound assures a sufficient noise reduction and is smaller if tails of the noise are lighter. The upper bound implies $\eta \to \infty$. Generally, in practice, the moment of the noise θ is unknown. The nonadaptive block length $k = \lceil n^{1/3} \rceil$ is a simple, yet effective choice that satisfies the condition of Theorems 3.2 and 3.3 if $\theta > 6$. Bühlmann and Künsch (1999) found that the $n^{1/3}$-choice performs quite well in most cases.

(iii) The gap d is allowed to vanish asymptotically for our result, as long as the rate of decay is slower than $n^{-1/\theta}$. The conditions on d and k have to be satisfied for the same θ. This means that if d decays slowly and Condition 3.1 is satisfied for a large θ, then k can be chosen smaller and the result still holds.

Our main result of this section is regarding a bound for the error of estimating τ by $\hat{\tau}$. We provide this bound in terms of the minimum gap to the signal averaged over sliding blocks. More precisely, defining

$$d_* := \min_{\substack{i=k(\eta+1)+1,\\ \ldots,n-k+1}} \frac{1}{k} \sum_{j=i}^{i+k-1} (\mu_j - \mu_1), \tag{20}$$

then we have the following.

THEOREM 3.3. *Assume Condition 3.1, $d \gg n^{-1/\theta}$, $n^{2/\theta}\log(n) \ll k \ll \tau$, $n - \tau \geq 2k$, and that there exists a constant $K > \rho$ with $d > Kd_*$, where ρ is the tuning parameter from*

the definition of $\hat{\tau}$. *Let the estimator* $\hat{\sigma}_\infty^2$, *used in* (9), *be consistent for the long-run variance* σ_∞^2; *that is*, $\hat{\sigma}_\infty^2 = \sigma_\infty^2 + o_{\mathbb{P}}(1)$. *Then*

$$\hat{\tau} = \tau + \mathcal{O}_{\mathbb{P}}(d_*^{-\theta/(\theta-1)}),$$

as $n \to \infty$.

The proof is deferred to Section A.3.

COROLLARY 3.1. *Under the conditions of Theorem* 3.3, *we have*:

(i) *If* d *is bounded away from zero* (*i.e.*, *if there exists a constant* M *with* $0 < M < d$), *then* $\hat{\tau}_n = \tau + \mathcal{O}_{\mathbb{P}}(1)$, *as* $n \to \infty$.

(ii) *If* d *is unbounded* (*i.e.*, $d \to \infty$, *as* $n \to \infty$), *then* $\mathbb{P}(\hat{\tau}_n = \tau) \to 1$, *as* $n \to \infty$.

REMARK 2. (i) Under the general alternative (4), we do not impose any regularity conditions, apart from the one sidedness of the change. Hence, it is possible that the gap $d \le \min_{t=\tau,\dots,n}(\mu_t - \mu_1)$ is determined by an individual (noisy) observation, which is not enough to consistently estimate d itself. We show (Lemma A.5) that our proposed $\hat{d}$ consistently estimates d_*. By definition, $d_* \ge d$; that is, d_* provides an upper bound for the gap d. The regularity condition $d > Kd_*$, for a constant $K > \rho$, requires that d_*, which is tractable, also facilitates a lower bound with respect to d. The regularity condition $d > Kd_*$, for a constant $K > \rho$, is sufficient to estimate τ by $\hat{\tau}$.

(ii) The following example illustrates a situation where $d > Kd_*$ is satisfied, for a constant $K > \rho$: say $\mu_j - \mu_1 = m((j-\tau)/(n-\tau))$ for a function $m : [0,1] \to (0,\infty)$ of bounded total variation $\|m\|_{\mathrm{TV}} < \infty$ and let $K := 1/(1 + \|m\|_{\mathrm{TV}}/(k \inf_{x\in[0,1]} m(x)))$. Note that $K > \rho$ if $k > \rho\|m\|_{\mathrm{TV}}/((1-\rho)\inf_{x\in[0,1]} m(x))$ and

$$\begin{aligned} d_* &\le \inf_{x\in[0,1]} m(x) + \frac{1}{k}\|m\|_{\mathrm{TV}} \\ &= K^{-1} \inf_{x\in[0,1]} m(x) \\ &\le K^{-1} \min_{t\in\tau,\dots,n} m\left(\frac{j-\tau}{n-\tau}\right) \\ &=: K^{-1} d. \end{aligned}$$

As an example, take a continuously differentiable function f and add a finite number of jump discontinuities at distinct $x_1, \dots, x_b$: that is, $m(x) := f(x) + \sum_{i=1}^b \delta_i I\{x \le x_i\}$. Then m is of bounded variation: $\|m\|_{\mathrm{TV}} = \int_0^1 |f'(t)|\mathrm{d}t + \sum_{i=1}^b |\delta_i|$.

(iii) If the true d were known, we could use the following estimate for τ:

$$\tilde{\tau} := \arg\min_{j=2,\dots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho d). \tag{21}$$

Following the lines of our proof for Theorem 3.3, it can be shown that $\tilde{\tau} = \tau + \mathcal{O}_{\mathbb{P}}(d^{-\theta/(\theta-1)})$, for all $\rho \in (0,1)$, without an assumption regarding d_*. Note that $\tilde{\tau}$ is only available if d is known.

(iv) The conditions regarding τ allow for "early" and "late" changes. In particular, we do not require that $\tau \asymp n$. The requirement $k \ll \tau$ ensures that there is an increasing number of blocks before the change. The requirement $n - \tau \ge 2k$ is slightly weaker and ensures that there is at least one complete block after the change. The requirement $n - \tau \ge 2k$ is needed to estimate d_* (see Lemma A.5).

TABLE 1
Simulated expectation $\mathbb{E}Z_i'$ and long-run variance σ_∞^2 of (Z_i'), defined in (22), for the case when $\varepsilon_i \sim \mathcal{N}(0,1)$

θ	$\mathbb{E}Z_i'$	σ_∞^2
0.2	0.343	1.332
0.3	0.577	2.104
0.4	0.988	5.782

4. Monte Carlo studies.

4.1. *Models considered.* We assess the finite sample performance of both the testing procedure (Section 2.2) and the two-stage locating algorithm (Section 2.3). Our experiments employ data crafted via the signal plus noise model delineated in (1).

For the noise component, we utilize a threshold AR model (Tong (1990)) as follows:

$$Z_i' = \theta\big(|Z_{i-1}'| + |Z_{i-2}'|\big) + \varepsilon_i, \tag{22}$$

where θ is the parameter governing temporal dependence, and the i.i.d. innovations ε_i follow the normal distribution $\mathcal{N}(0, 0.5^2)$. Within this model, a higher absolute value of θ indicates stronger temporal dependence. The process remains stationary provided $|\theta| < 0.5$. The noise process (Z_i) is obtained by centering Z_i as $Z_i := Z_i' - \mathbb{E}(Z_i')$. Three digits behind the comma approximations to the values we used are in Table 1. For $\theta < 0$, we use that the expectation of the process for θ and $-\theta$ have the same long-run variance and the expectation differs only in sign. Further, for $\varepsilon_i \sim \mathcal{N}(0, \xi^2)$, $\xi > 0$, we obtain expectation and long-run variances by multiplying the ones from Table 1 with ξ and ξ^2, respectively. For example, for $\theta = -0.2$ and $\varepsilon_i \sim \mathcal{N}(0, 0.5^2)$, we use $\mathbb{E}Z_i' = -0.343 \cdot 0.5$ and $\sigma_\infty^2 = 1.332 \cdot 0.5^2$.

Regarding the signal μ_t, we examine two scenarios: (i) under the null hypothesis H_0, as defined in (3), the signal remains constant at $\mu_1 = 0$. (ii) Under the alternative hypothesis H_1, as defined in (4), the signal generation model is as follows:

$$\mu_t = \begin{cases} \mu_1 := 0 & \text{for } t = 1, \dots, \tau - 1, \\ \mu_1 + s\left(\dfrac{2t - 3\tau + \tau'}{\tau' - \tau}\right) & \text{for } t = \tau, \dots, \tau', \\ \mu_1 + s\left(2 + \exp\left(\dfrac{2(t - \tau')}{\tau'' - \tau'}\right)\right) & \text{for } t = \tau' + 1, \dots, \tau'', \\ \mu_1 + s\left(2 + \exp(2) \cdot \dfrac{2n - \tau'' - t}{2n - 2\tau''}\right) & \text{for } t = \tau'' + 1, \dots, n, \end{cases} \tag{23}$$

In this model, the parameter s defines the magnitude of deviation from the baseline mean state ($\mu_1 = 0$) for $t < \tau$ to the varied mean state for $t \geq \tau$. This model is designed to reflect trends similar to those observed in the search engine index data depicted in Figure 1.

Figure 3 showcases an example of the signal (μ_i). It highlights the increase in signal strength after the initial change point at $\tau = 320$, where it rises by at least $s = 0.5$ above the stable level of $\mu_1 = 0$. Beyond the first change-point τ, a second significant change occurs at $\tau'' = 640$, where the signal further elevates, reaching at least $8s = 4$ above the initial $\mu_1 = 0$ level. This pattern echoes our observations in real-world data.

4.2. *Synthetic data under the null hypothesis.* In this section, we illustrate that the testing procedure described in Section 2.2 has the correct size, asymptotically. We employ data structured as detailed in Section 4.1, operating under a constant signal (i.e., H_0).

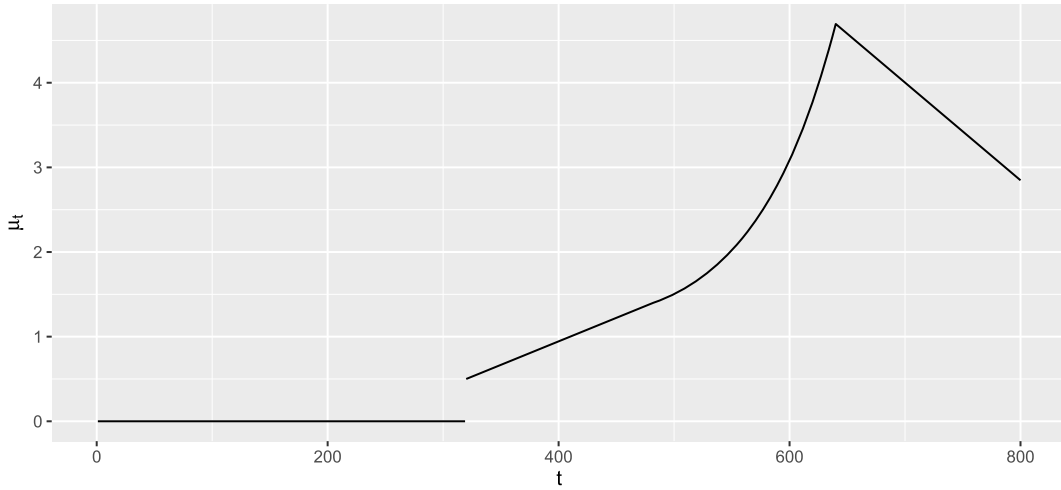

FIG. 3. *Illustration of* (μ_t) *with* $n = 800$, $\tau = 320$, $\tau' = 500$, $\tau'' = 640$ *and* $s = 0.5$.

We modulate the sample size, selecting n from 50, 100, 300, 500, 2000, and adjust the dependence parameter θ from -0.4, -0.2, 0 (independence), 0.2, 0.4. The significance level remains fixed at $\alpha = 0.05$. We use the true long-run variance σ_∞^2 instead of $\hat{\sigma}_\infty^2$ in (5); cf. Table 1. The empirical sizes, derived from 100,000 replications, are summarized in Table 2.

Analyzing Table 2, it is evident that the rejection ratios—serving as proxies for type-I error—gravitate closer to the target significance level of $\alpha = 0.05$ as the sample size n expands and temporal dependence weakens (absolute value of θ shrinks). This observation aligns seamlessly with our theoretical framework presented in Section 3.2. On juxtaposing the two methodologies, the finite-sample Gaussian approximation-based testing procedure emerges superior in smaller data sets ($n = 50, 100, 300, 500$), compared to its asymptotic counterpart. However, the latter's performance converges with the finite-sample approach as the sample size surges to $n = 2000$. This implies that, for shorter data sets, the finite-sample Gaussian approximation can be advantageous. Conversely, for longer data sets, the more computationally economical asymptotic approach becomes viable.

Applying the test with the estimated long-run variance leads to higher error rates, in particular for small sample sizes, when it its difficult to estimate the long-run variance. Results are provided in Appendix B (in the Supplementary Material).

TABLE 2
Rejection ratios for the change-point testing procedure under the null hypothesis; *cf.* (3)

		θ				
Approximation Method	n	-0.4	-0.2	0	0.2	0.4
Asymptotic	50	1.41%	2.70%	3.28%	3.18%	1.63%
	100	2.26%	3.40%	3.74%	3.69%	2.41%
	300	3.23%	3.92%	4.21%	4.08%	3.39%
	500	3.50%	4.17%	4.38%	4.30%	3.56%
	2000	4.18%	4.54%	4.54%	4.65%	4.23%
Finite sample	50	2.10%	4.08%	5.00%	4.79%	2.44%
	100	2.96%	4.54%	5.00%	4.90%	3.14%
	300	3.80%	4.71%	4.97%	4.86%	3.92%
	500	3.98%	4.69%	5.01%	4.92%	4.04%
	2000	4.41%	4.85%	4.82%	4.95%	4.47%

4.3. *Synthetic data under alternative hypotheses*. This section provides an in-depth analysis of our testing procedure's power and evaluates the efficacy of the algorithm used for locating the first change point, employing synthetic data. We adopt the data structure described in Section 4.1, with the signal defined as per (23) (i.e., H_1).

Our experimental setup is as follows: we vary the sample size, choosing n from the values 50, 100, 300, 500, 2000. We select the dependence parameter θ from the values -0.4, -0.2, 0 (representing independence), 0.2 and 0.4. The gap parameter s ranges from 0 to 0.045, increasing in steps of 0.0006. It is important to note that the standard deviation of the innovation in the dependent process is fixed at 0.5, and we keep $\mu_1 = 0$. We standardize the ratios $\tau/n = 0.4$, $\tau'/n = 0.6$, $\tau''/n = 0.8$, maintaining $\mu_1 = 0$. For our testing methods, we consistently set the significance level at $\alpha = 0.05$. Once the parameters for an experiment are established, we generate the trend (μ_i) using the aforementioned methodology. Subsequently, the additive noise process is simulated repeatedly, and this data is input into our testing and locating algorithms. For testing, we use the quantile obtained from the asymptotic approximation and the true long-run variance σ_∞^2 instead of $\hat{\sigma}_\infty^2$ in (5); cf. Table 1. For our locating algorithm, we first apply the test and continue only if it rejects. We use the long-run variance estimator defined in Section 2.3.2; that is, $\hat{\sigma}_\infty^2 := \hat{\sigma}^2$; cf. (8) and (9). The results are derived from 100,000 independent simulations.

It is crucial to observe that testing for a change point remains challenging, even in scenarios with the largest gap parameter $s = 0.045$. This difficulty arises because, as indicated in Table 1, the gap $s = 0.045$ is considerably smaller than the noise levels, complicating the detection of the change-point's presence significantly.

Initial observations indicate variations in the rejection ratio, an estimate of the true power, in relation to the gap parameter s. These variations are evident across different combinations of the dependence parameter θ and sample size n, as depicted in Figure 4. In each experiment, the rejection ratio progresses from the nominal level ($\alpha = 0.05$) to nearly 1 as s increases from 0 to 0.045. This trend suggests that as the task of detecting change points becomes less challenging, the power of our test approaches unity.

The graphic shows an increase in the rejection ratio with sample sizes expanding from 50 to 2000. This trend is in alignment with the theoretical insights presented in Theorem 3.1(ii). Additionally, it is noteworthy that despite a diminished test power under conditions of strong temporal dependence (with $|\theta| = 0.4$), the power can still approach unity given a sufficient sample size. This observation implies the efficacy of our testing procedure even under the influence of temporal dependent noise in the data. Results for the case when the long-run variance is estimated are delegated to Appendix B (in the Supplementary Material).

Next, we showcase the absolute errors normalized by sample size $\mathbb{E}|\hat{\tau} - \tau|/n$ of our two-step locating algorithm across experiments with diverse parameters in Figure 5.

Without temporal dependence, error rates are smaller. As the disparity between the signal and nonsignal segments grows from 0.4 to 0.8, the error rate decreases. As the sample size n expands from 50 to 2000, the MAE/n progressively diminishes. These findings resonate with Theorem 3.3, discussed in Section 3.3. For scenarios characterized by heightened dependence and minimal gap, error rates can be larger. Yet, in more favorable conditions, the error remains relatively stable or increases only marginally. This fact underscores the robustness of our methodology.

We expanded our analysis to compare the performance of our locating algorithm with four established change-point estimation techniques: a CUSUM-type method, a likelihood-based method (AMOC), the earliest change point from the standard binary segmentation (1SBS) method and a modified 1SBS method where the marginal variance in the threshold is replaced by the estimated long-run variance. More precisely, we obtain $\arg\min_{j=2,3,\ldots,n+1} \sum_{i=1}^{j-1} (X_i - \bar{X}_n)$ and refer to it as CUSUM. This is related to our test statistic $\hat{T}$, defined in (5). Second, we

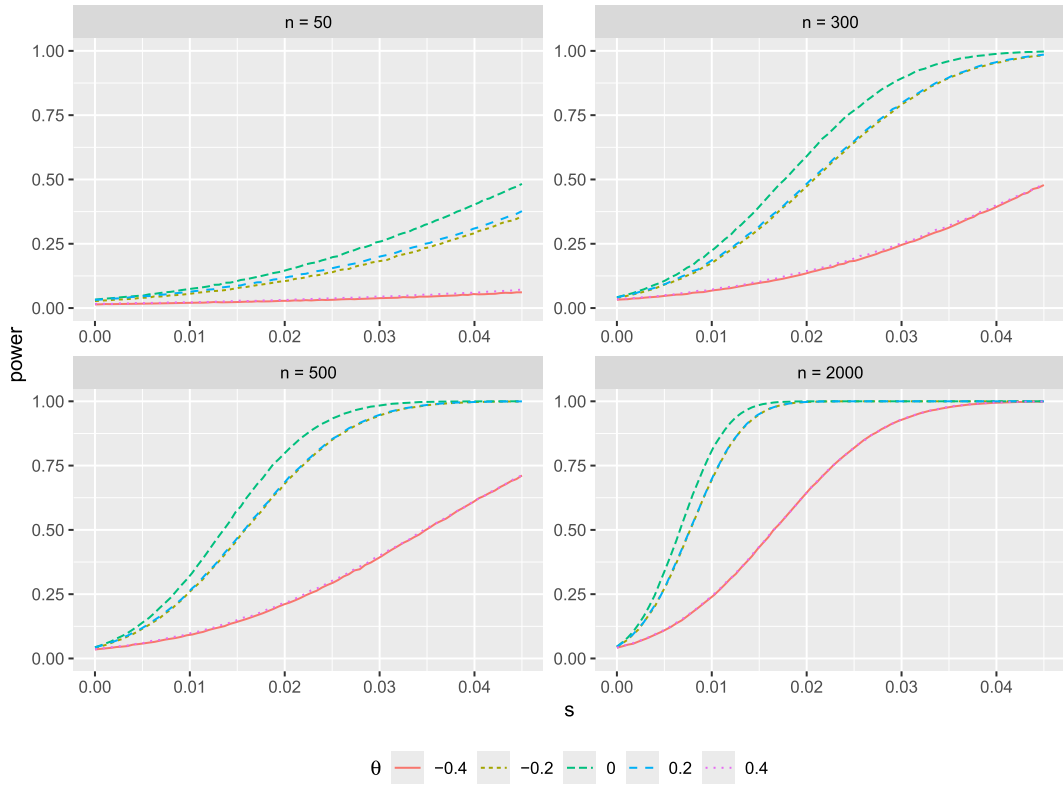

FIG. 4. *Rejection ratios for the testing procedure under the alternative hypothesis*: $\mu_i = \mu_1$ *for* $i = 1, 2, \ldots, \tau - 1$; $\mu_i > \mu_1 + s$ *for* $i = \tau, \tau + 1, \tau + 2, \ldots, n$. *The noise process is shaped by the dependence parameter* θ. *We adjust the gap parameter* s *over the set* $\{0, 0.0006, 0.0012, \ldots, 0.045\}$, n *over* $\{50, 300, 500, 2000\}$ *and* θ *over* $\{-0.4, -0.2, 0, 0.2, 0.4\}$. *Each data point represents* 100,000 *replications*.

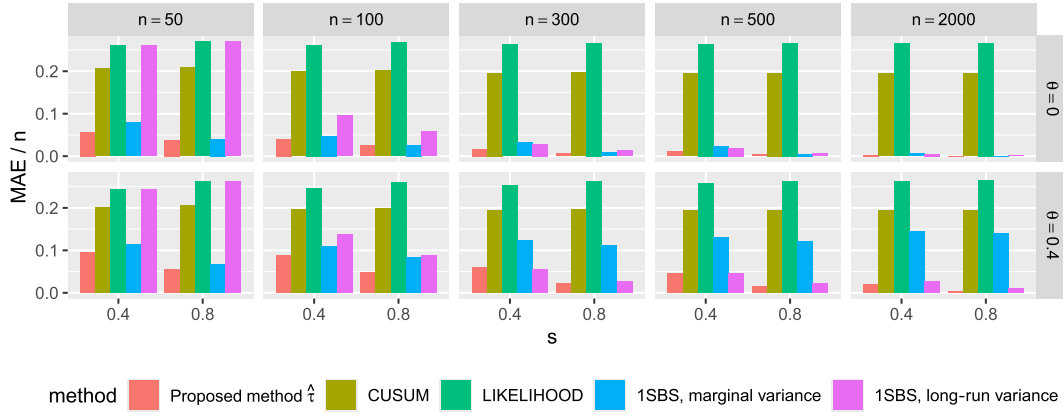

FIG. 5. *Expected absolute errors normalized by sample size* (*MAE/n*) *across four change-point detection methods. The red, olive, green, blue and purple bars show* $\mathbb{E}|\hat{\tau} - \tau|/n$ *of our proposed method* $\hat{\tau}$, *CUSUM, likelihood-based method* (*AMOC*), *the earliest change point from the standard binary segmentation* (1*SBS*) *method and the earliest change point from the* 1*SBS method with marginal variance in the threshold replaced by long-run variance, respectively. The parameters varied in this study include the gap parameter* s, *the sample size* n *and the parameter* θ *of the threshold autoregression noise process. Each bar in the graph represents the average result from* 100,000 *replications*.

apply the functions `cpt.mean`, with `method = "AMOC"` (at most one change), and `cpts` from the R package *changepoint* (Killick and Eckley (2014), Killick, Haynes and Eckley (2022)) and refer to the obtained value as AMOC.

Third, we apply the functions `sbs` and `changepoints` from the R package *wbs* (Baranowski and Fryzlewicz (2019)) and refer to the minimum of the obtained values as 1SBS (first time of change obtained with standard binary segmentation). Fourth, we replace the marginal variance used in the definition of the threshold of the standard binary segmentation method by the long-run variance, which we estimate by $\hat{\sigma}^2$, from Section 2.3.2, with $k = \lceil n^{1/3} \rceil$ and $J = 3$.

We add $+1$ to AMOC, 1SBS and 1SBS based on long-run variance, to account for the fact that in our notation the change occurs from $\tau - 1$ to τ while there it occurs from τ to $\tau + 1$. Note that, while the first two methods estimate a single change point, the binary segmentation methods estimate multiple change points of which we select the earliest.

The outcome of this comparative study is detailed in Figure 5. The results reveal that the errors associated with the four alternative methods are somewhat unstable, and depending on the scenario, can perform poorly. The 1SBS method with the standard threshold performs roughly equally well under independence when the gap size is $s = 0.8$. When the gap size takes the smaller value $s = 0.4$, where the locating problem is harder, the proposed method has slightly better performance. The 1SBS method with the threshold adjusted for long-run variance also performs well when there is serial dependence, but only for larger sample sizes. We also note that keeping the gap parameter (s) and the dependence parameter (θ) constant while increasing the sample size (n) from 50 to 2000 results in the mean absolute error (MAE/n) normalized by n for our proposed method approaching zero. This observation confirms our theory and previous numerical analysis that our method's error remains relatively constant with larger sample sizes. In contrast, for the other four methods, we observe less stable behavior of the MAE/n, which either remains relatively unchanged as n increases (CUSUM and AMOC), indicating that their errors grow with the sample size, or behave reasonably under independence but struggle in the presence of serial dependence (1SBS), or do not perform well for small sample sizes (1SBS with LRV).

Furthermore, when θ and n are fixed and s is varied, our method demonstrates a steady decline in error as s increases. Such a consistent pattern of reducing error is also not observed in all of the other methods.

Overall, these results highlight the shortcomings of traditional methods in handling nonstandard or complex data configurations, emphasizing the versatility of our proposed method. The bar plots in Figure 5 provide visual evidence of the consistently satisfactory performance of our method across various parameter settings, while the alternative methods exhibit unstable and sometimes poor performance, particularly in the presence of serial dependence.

5. Baidu search index for COVID-19 related symptoms. Numerous studies have endeavored to pinpoint the initial emergence of the SARS-CoV-2 virus among humans. The initial cases were likely linked to the Huanan Seafood Wholesale Market in late December 2019. However, this cluster is not believed to signify the pandemic's inception. To deduce the possible duration SARS-CoV-2 circulated in China before detection, we analyzed Baidu's search index (China's leading search engine) for COVID-19 symptom-related keywords between October 1, 2019, and January 31, 2020, in the Hubei Province, China. We focused on the terms "fever" and "cough," aggregating searches from both desktop and mobile platforms. As depicted in Figure 6, the counts exhibit regular fluctuations until the series' end. Given the rapid transmission capability of COVID-19, the constant mean assumption post-change point in conventional methods is inapplicable. Applying the test proposed in Section 2.2 for the null hypothesis H_0 of constant mean, defined in (3), against the alternative hypothesis H_1

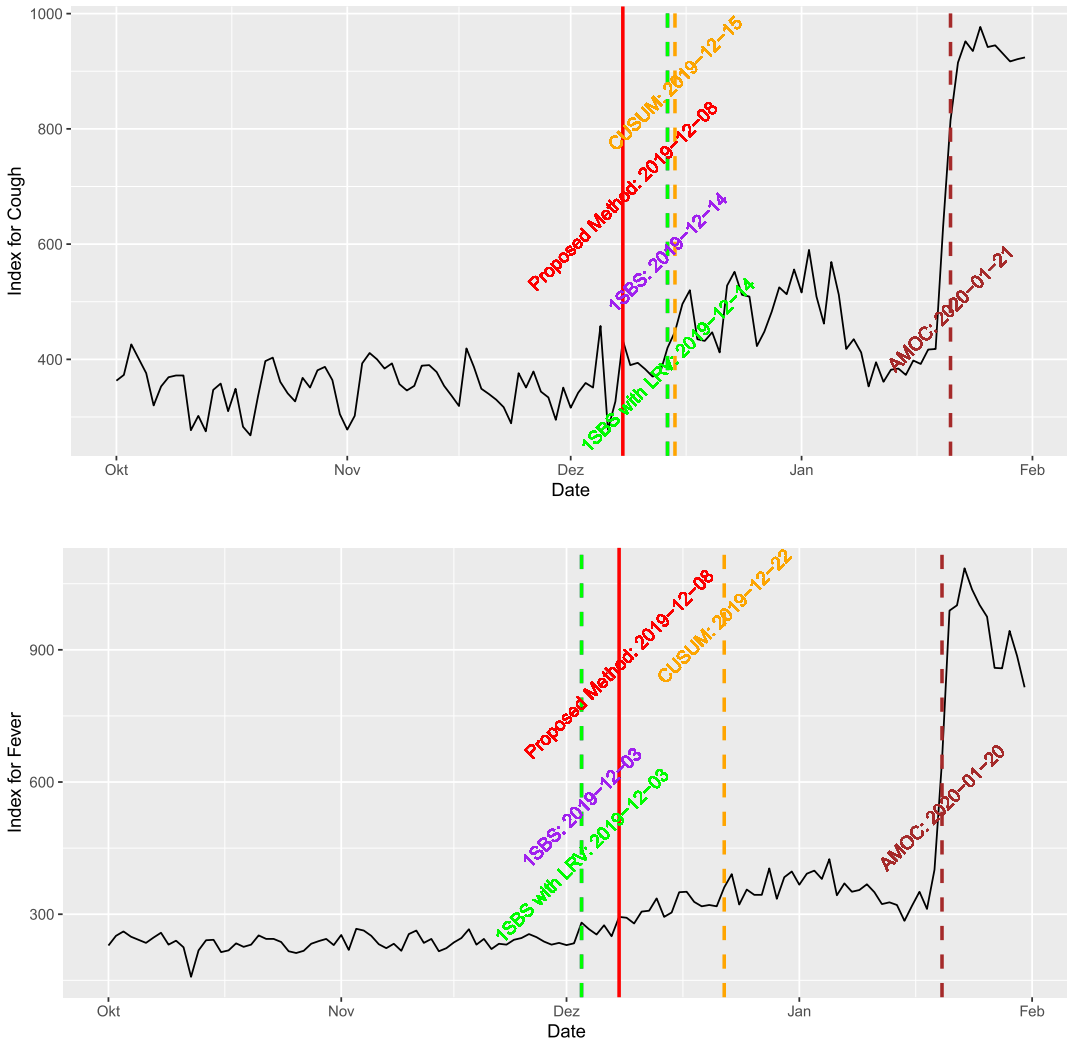

FIG. 6. *BAIDU search index for "fever" and "cough" from October* 1, 2019, *to January* 31, 2020. *The red solid vertical line indicates the change point detected by our proposed method* $\hat{\tau}$; *the dashed lines in orange, brown, purple and green represent the change points detected by CUSUM, likelihood-based* (*AMOC*), *earliest change detected by standard binary segmentation* (1*SBS*) *and binary segmentation with long-run variance modification* (1*SBS with LRV*), *respectively.*

of a one-sided upwards change, defined in (4), yields test statistics $\hat{T} < -9$ and $\hat{T} < -22$, for Baidu search indices "cough" and "fever," respectively. The p-values implied by Theorem 3.1(i) are essentially zero such that we reject the null hypothesis in both cases.

We continue the analysis by employing the two-stage locating method (Section 2.3). For the keyword "cough" (comprising $n = 123$ data points), the initial stage estimates the equilibrium data state's mean, μ_1, and the state gap parameter, d, guiding the subsequent stage. We defined $k = \lceil n^{1/3} \rceil = 5$ for the batched mean length and computed

$$R_j = \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} X_j, \quad j = 1, \ldots, m,$$

as defined in (6). We obtain $\hat{L} := \max\{i : R_i \le R_m^{(3)}\} = 11$ and $\hat{\ell} := k\hat{L} = 55$, as defined in (7). We find a prechange sample mean of $\hat{\mu}_0 = 352.84$ obtained from the initial $\hat{\ell}$ data. The test statistics $\hat{D}_j = \sqrt{k}(R_j - \hat{\mu}_0)/\hat{\sigma}_\infty$ using $\hat{\sigma}_\infty \approx 48.68$, which is the square root of the

estimated long-term variance from the initial $\hat{\ell}$ observations; cf. (8). Then we obtain the test decisions $\hat{I}_j$, defined in (9), as

$$\hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/m}, \\ 0 & \text{otherwise,} \end{cases}$$

where $z_{1-1/m}$ is the $1 - 1/m$ quantile of the standard normal distribution. We obtain

$$\hat{\eta} := \arg\min_t \sum_{j=1}^{m} \{I_j - 1_{[t+1,m]}(j)\}^2 = 15,$$

as defined in (10). A graphical representation of the test decisions and smoothing can be seen in Figure 2. The first-stage estimates thus are

$$\hat{\mu}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} X_i \approx 355.43, \quad \hat{d} := \min_{\substack{i=k(\hat{\eta}+1)+1, \\ \ldots, n-k+1}} \frac{1}{k} \sum_{j=i}^{i+k-1} (X_j - \hat{\mu}_1) \approx 19.24.$$

Setting $\rho = 0.5$, our refined change-point estimate in the second phase is

$$\hat{\tau} := \arg\min_{j=2,\ldots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho\hat{d}) = 69,$$

as defined in (12), which translates to December 8, 2019.

For comparative purposes, we also compute the CUSUM, AMOC and 1SBS estimates considered in Section 4.3. The CUSUM approach identified December 15, 2019, as the change point. The AMOC approach, pinpointed January 21, 2020, as the change point. In contrast, the binary segmentation method and its modified version, where the marginal variance in the threshold is replaced by the long-run variance, focusing on the first change point, suggests December 14, 2019, as the initial outbreak date, which precedes the dates indicated by the other two methods but is still later than our findings.

Interestingly and surprisingly, but reasonably, our analysis of the Baidu "fever" search index corroborates our findings by *also* indicating December 8, 2019, as the change point, consistent with the "cough" data set results. Conversely, the CUSUM, AMOC and 1SBS methods suggest change points on December 22, 2019, January 20, 2020, and December 3, 2019, respectively. A graphical representation of the results can be seen in Figure 6.

Reports such as Worobey (2021) mention an early COVID-19 case, a 41-year-old male, showing symptoms on December 16, 2019, suggesting community transmission. Another case, a female seafood vendor, exhibited symptoms on December 10, 2019, and was aware of potential COVID-19 cases near Huanan Market from December 11, 2019. Other studies and organizations like the CDC mention early December 2019 as significant. Given these findings, our change-point detection appears plausible.

It is noteworthy that the Chinese government officially announced the outbreak on January 20, 2020, a discernible tipping point. This is not our primary focus, as our aim is to identify the initial outbreak, which undoubtedly predates January 1, 2020. Classical methods seem ill-equipped to discern this early change point, potentially overshadowed by subsequent tipping points. This is understandable, as such methods rely heavily on the sample mean, which can be skewed by later data points, leading to inaccurate estimations.

Acknowledgments. The authors would like to thank the Associate Editor and the three reviewers for their constructive comments that helped to improve the paper.

Funding. This research is partially supported by NSF Grants DMS-2311249 and NSF DMS-2027723.

SUPPLEMENTARY MATERIAL

Appendix (DOI: 10.1214/24-AOS2451SUPPA; .pdf). Contains the proofs and additional simulation results.

Replication package (DOI: 10.1214/24-AOS2451SUPPB; .zip). R code implementing the proposed method and scripts to replicate the simulation and empirical results in the paper are available on https://github.com/tobiaskley/cp_analysis_w_irreg_signals_replication_package.

REFERENCES

AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 649–672. MR3961502 https://doi.org/10.1111/rssb.12322

BARANOWSKI, R. and FRYZLEWICZ, P. (2019). wbs: Wild Binary Segmentation for Multiple Change-Point Detection. R package version 1.4.

BERKES, I., LIU, W. and WU, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

BÜCHER, A., DETTE, H. and HEINRICHS, F. (2021). Are deviations in a gradually varying mean relevant? A testing approach based on sup-norm estimators. *Ann. Statist.* **49** 3583–3617. MR4352542 https://doi.org/10.1214/21-aos2098

BÜHLMANN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31** 295–310.

CAO, H. and WU, W. B. (2015). Changepoint estimation: Another look at multiple testing problems. *Biometrika* **102** 974–980. MR3431567 https://doi.org/10.1093/biomet/asv031

CAO, H. and WU, W. B. (2022). Testing and estimation for clustered signals. *Bernoulli* **28** 525–547. MR4337715 https://doi.org/10.3150/21-bej1355

CHEN, Y., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

DETTE, H., ECKLE, T. and VETTER, M. (2020). Multiscale change point detection for dependent data. *Scand. J. Stat.* **47** 1243–1274. MR4178193 https://doi.org/10.1111/sjos.12465

DETTE, H. and WU, W. (2019). Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *Ann. Statist.* **47** 3578–3608. MR4025752 https://doi.org/10.1214/19-AOS1811

CENTRE FOR DISEASE CONTROL AND PREVENTION (2022). CDC museum COVID-19 timeline. *website*. Accessed: 2022-05-18.

FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 https://doi.org/10.1111/rssb.12047

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

FRYZLEWICZ, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *Ann. Statist.* **46** 3390–3421. MR3852656 https://doi.org/10.1214/17-AOS1662

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **72** 180–186. MR0451496

HEINRICHS, F. and DETTE, H. (2021). A distribution free test for changes in the trend function of locally stationary processes. *Electron. J. Stat.* **15** 3762–3797. MR4298981 https://doi.org/10.1214/21-ejs1871

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727 https://doi.org/10.1093/biomet/57.1.1

HORVÁTH, L. and KOKOSZKA, P. (2002). Change-point detection with non-parametric regression. *Statistics* **36** 9–31. MR1906372 https://doi.org/10.1080/02331880210930

HUANG, C., WANG, Y., LI, X., REN, L., ZHAO, J., HU, Y., ZHANG, L., FAN, G., XU, J. et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395** 497–506.

JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866 https://doi.org/10.1111/jtsa.12035

KILLICK, R. and ECKLEY, I. A. (2014). Changepoint: An R package for changepoint analysis. *J. Stat. Softw.* **58** 1–19.

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KILLICK, R., HAYNES, K. and ECKLEY, I. A. (2022). changepoint: an R package for changepoint analysis. R package version 2.2.4.

KLEY, T., LIU, Y. P, CAO, H. and WU, W. B (2024). Supplement to "Change-point analysis with irregular signals." https://doi.org/10.1214/24-AOS2451SUPPA, https://doi.org/10.1214/24-AOS2451SUPPB

LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117 https://doi.org/10.1214/aos/1018031117

MALLIK, A., BANERJEE, M. and SEN, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electron. J. Stat.* **7** 2477–2515. MR3117104 https://doi.org/10.1214/13-EJS845

MALLIK, A., SEN, B., BANERJEE, M. and MICHAILIDIS, G. (2011). Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika* **98** 887–900. MR2860331 https://doi.org/10.1093/biomet/asr051

MIES, F. and STELAND, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli* **29** 3114–3140. MR4632133 https://doi.org/10.3150/22-bej1577

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761. MR1165590 https://doi.org/10.1214/aos/1176348654

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44** 248–252.

PELIGRAD, M. and SHAO, Q. M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *J. Multivariate Anal.* **52** 140–157. MR1325375 https://doi.org/10.1006/jmva.1995.1008

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, London. MR0991969

SEN, A. and SRIVASTAVA, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3** 98–108. MR0362649

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105 https://doi.org/10.1214/009053606000001479

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Int. Stat. Rev.* **56** 31–48. MR0963139 https://doi.org/10.2307/1403360

TONG, H. (1990). *Nonlinear Time Series*. *Oxford Statistical Science Series* **6**. Clarendon Press, Oxford University Press, New York. MR1079320

VOGT, M. and DETTE, H. (2015). Detecting gradual changes in locally stationary processes. *Ann. Statist.* **43** 713–740. MR3319141 https://doi.org/10.1214/14-AOS1297

WOROBEY, M. (2021). Dissecting the early COVID-19 cases in Wuhan. *Science* **374** 1202–1204.

WORSLEY, K. J. (1986). Confidence regions and test for a change-point in a sequence of exponential family random variables. *Biometrika* **73** 91–104. MR0836437 https://doi.org/10.1093/biomet/73.1.91

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. MR2812816 https://doi.org/10.4310/SII.2011.v4.n2.a15

ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184