Received: 11 January 2022 | Accepted: 19 August 2022 | Published on: 30 October 2022

DOI: 10.1111/rssb.12552

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY Series B Statistical Methodology

Segmenting time series via self-normalisation

Zifeng Zhao[1] | **Feiyu Jiang**[2] | **Xiaofeng Shao**[3]

[1]Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana, USA

[2]Department of Statistics and Data Science, Fudan University, Shanghai, China

[3]Department of Statistics, University of Illinois at Urbana Champaign, Champaign, Illinois, USA

Correspondence
Feiyu Jiang, Department of Statistics and Data Science, Fudan University, Shanghai, China.
Email: jiangfy@fudan.edu.cn

Funding information
National Science Foundation, Grant/Award Numbers: DMS-1807032, DMS-2014018, DMS-2014053; Shanghai Sailing Program, Grant/Award Number: 22YF1402400; National Natural Science Foundation of China, Grant/Award Numbers: NSFC-71973077, NSFC-11771239

Abstract

We propose a novel and unified framework for change-point estimation in multivariate time series. The proposed method is fully non-parametric, robust to temporal dependence and avoids the demanding consistent estimation of long-run variance. One salient and distinct feature of the proposed method is its versatility, where it allows change-point detection for a broad class of parameters (such as mean, variance, correlation and quantile) in a unified fashion. At the core of our method, we couple the self-normalisation- (SN) based tests with a novel nested local-window segmentation algorithm, which seems new in the growing literature of change-point analysis. Due to the presence of an inconsistent long-run variance estimator in the SN test, non-standard theoretical arguments are further developed to derive the consistency and convergence rate of the proposed SN-based change-point detection method. Extensive numerical experiments and relevant real data analysis are conducted to illustrate the effectiveness and broad applicability of our proposed method in comparison with state-of-the-art approaches in the literature.

KEYWORDS

binary segmentation, change-point detection, long-run variance, scanning, studentisation, temporal dependence

1 | INTRODUCTION

Change-point detection has been identified as one of the major challenges for modern data applications (National Research Council, 2013). There is a vast literature on change-point estimation and testing in statistics, in part due to its broad applications in bioinformatics, climate science, economics, finance, genetics, medical science and signal processing among many other areas. See Csörgő and Horváth (1997), Brodsky and Darkhovsky (2013) and Tartakovsky et al. (2014) for book-length treatments of the subject. We also refer to Aue and Horváth (2013), Casini and Perron (2019) and Truong et al. (2020) for excellent reviews.

In this paper, we study the problem of time series segmentation, also known as (offline) change-point estimation, where the task is to partition a sequence of potentially non-homogeneous ordered observations into piecewise homogeneous segments. Many change-point problems arise within a time series context (e.g. climate, epidemiology, economics and financial data), where there is a natural temporal ordering in the observations. Although temporal dependence is the norm rather than the exception for time series, most literature in change-point analysis assume and require independence of observations $\{Y_t\}_{t=1}^n$ over time for methodological and theoretical validity; see for example Olshen et al. (2004), Killick et al. (2012), Matteson and James (2014), Fryzlewicz (2014) and Baranowski et al. (2019) among others. One stream of literature addresses temporal dependence via the assumption of parametric models, see Davis et al. (2006) and Yau and Zhao (2016) for change-point detection in AR process and Fryzlewicz and Subba-Rao (2014) in ARCH process. However, parametric approaches generally require stronger conditions and potential violation of parametric assumptions can inevitably cast doubts on the estimation result.

Existing non-parametric approaches for change-point estimation in temporally dependent observations primarily focus on first or second-order moments, see Bai and Perron (1998), Eichinger and Kirch (2018) for change-point estimation in mean, Aue et al. (2009), Preuss et al. (2015) in (auto)-covariance, and Cho and Fryzlewicz (2012), Casini and Perron (2021a) in spectral density function (thus second-order properties). However, for many applications, the key interest can go beyond mean or covariance. For example, detecting potential changes in extreme quantiles is critical for monitoring systemic risk (i.e. Value-at-Risk) in finance and for studying evolving behaviour of severe weather systems such as hurricanes in climate science. Moreover, existing non-parametric methods are mostly designed for detecting only one specific type of change (e.g. mean or variance) and cannot be universally used for examining changes in different aspects of the data, which may limit its applications and cause inconvenience of implementation for practitioners. Additionally, existing nonparametric procedures typically involve certain tuning or smoothing parameters, such as the bandwidth parameter involved in the consistent estimation of the long-run variance (LRV), and how to choose these tuning parameters is important yet highly challenging in practice.

To fill in the gap in the literature, we propose a new multiple change-point estimation framework that is fully non-parametric, robust to temporal dependence, enjoys effortless tuning, and works universally for various parameters of interest for a multivariate time series $\{Y_t\}_{t=1}^n$ where $Y_t \in \mathbb{R}^p$ with a fixed dimension $p \geq 1$. Specifically, denote F_t as the cumulative distribution function (CDF) of Y_t, the proposed procedure allows change-point detection for any θ such that $\theta = \theta(F_t)$, where $\theta(\cdot)$ is a functional that takes value in $\mathbb{R}^d$ with $d \geq 1$. This is a broad framework that covers important quantities such as mean, variance, quantile, (auto)-correlation and (auto)-covariance among others, see Künsch (1989) and Shao (2010).

As in the standard change-point literature, we assume the change happens in a piecewise constant fashion. Specifically, we assume $\{Y_t\}_{t=1}^n$ is a piecewise stationary time series and there exist $m_o \geq 0$ unknown number of change-points $0 < k_1 < \cdots < k_{m_o} < n$ that partition $\{Y_t\}_{t=1}^n$ into $m_o + 1$ stationary segments. Define $k_0 = 0$ and $k_{m_o+1} = n$, the ith segment contains stationary observations $\{Y_t\}_{t=k_{i-1}+1}^{k_i}$ that share common behaviour characterised by θ_i (e.g. mean, variance, correlation, quantile), where we require $\theta_i \neq \theta_{i+1}$ for $i = 1, \ldots, m_o$ due to the structural break. Our primary interest is to recover the unknown number and locations of the change-points.

To achieve broad applicability and robustness against temporal dependence, our proposed multiple change-point estimation method is built upon self-normalisation (SN, hereafter), a nascent inference technique for time series (Shao, 2010, 2015). We note that since its first proposal in Shao (2010), SN has been extended to retrospective change-point testing by Shao and Zhang (2010), Hoga (2018), Betken and Wendler (2018), Zhang and Lavitas (2018) and Dette et al. (2020), and to sequential change-point monitoring by Dette and Gösmann (2020) and Chan et al. (2021). However, the primary focus of these papers is to construct SN-based change-point testing procedures (either retrospective or sequential) but not change-point estimation. Compared to change-point testing, change-point estimation is a much more challenging task both methodologically and theoretically: it further requires the estimation of the unknown number and locations of change-points, which involves substantially different techniques and analysis.

Indeed, the use of SN for time series segmentation (i.e. multiple change-point estimation) seems largely unexplored, with the exception of (Jiang et al., 2020, 2022) for piecewise linear and quantile trend models designed for COVID-19 time series. One notable reason for the scarcity of SN-based time series segmentation algorithms is that, unlike the classical CUSUM-based change-point test, the SN-based change-point testing cannot be easily extended to multiple change-point estimation by combining with the standard binary segmentation algorithm (Vostrikova, 1981). Such a combination simply fails due to the potential inflation of the self-normaliser under the presence of multiple change-points. We discuss this point in more details later in Section 3 and provide further illustration via both theory and numerical experiments in Section S.1 of Appendix S1.

To bypass this difficulty, we propose a novel nested local-window segmentation algorithm, which is then combined with an SN test to achieve multiple change-point estimation. We name the procedure SNCP. Through a series of carefully designed nested local-windows, the proposed procedure can isolate each true change-point adaptively and thus achieves respectable detection power and estimation accuracy. The statistical and computational efficiency of the nested local-window segmentation algorithm is further illustrated via extensive numerical comparison with popular segmentation algorithms such as SaRa in Niu and Zhang (2012), WBS in Fryzlewicz (2014) and SBS in Kovacs et al. (2020).

In addition to methodological advances, new theoretical arguments based on the partial influence functions (Pires & Branco, 2002) are further developed to establish the consistency and convergence rate of the proposed change-point estimation procedure, which seems to be the first in the SN literature. The proof is non-standard and built on a subtle analysis of the behaviour of SN-based test statistic around change-points. It differs from existing techniques in the change-point literature due to the presence of the self-normaliser (an inconsistent LRV estimator) and is of independent interest.

To our best knowledge, the proposed method (SNCP) is the first to address multiple change point estimation for a general parameter in the time series setting. One salient and distinct feature of SNCP is its versatility: it allows the user to examine potential change in virtually any parameter

of interest in an effortless fashion. This is valuable as in practice, the ground truth is unknown and it is important to examine the behaviour change of the data via different angles. In addition, due to its versatility and robustness to temporal dependence, SNCP can serve as a numerically credible and theoretically valid benchmark for almost all algorithms designed for multiple change-point estimation in a fixed-dimensional time series, which is of interest to both practical applications and academic research.

The rest of the paper is organised as follows. We first provide background of SN and introduce the SN-based detection method for single change-point estimation in Section 2. Building upon a novel nested local-window segmentation algorithm, Section 3 proposes a unified SN-based framework (SNCP) for multiple change-point estimation and further studies its theoretical properties. Extensive numerical experiments are conducted in Section 4 to demonstrate the promising performance of SNCP when compared with state-of-the-art methods for change-point estimation in mean, variance, quantile of univariate time series and correlation and covariance matrix of multivariate time series. Section 5 concludes. Technical proofs and additional simulation and real data application results can be found in Appendix S1.

Some notations used throughout the paper are defined as follows. Let $D[0,1]$ denote the space of functions on $[0,1]$ which are right continuous with left limits, endowed with the Skorokhod topology (Billingsley, 1968). We use $\Rightarrow$ to denote weak convergence in $D[0,1]$ or more generally in $\mathbb{R}^m$-valued function space $D^m[0,1]$, where $m \in \mathbb{N}$. We use $\xrightarrow{D}$ to denote convergence in distribution. We use $\|\cdot\|_2$ to denote the l_2 norm of a vector and use $\|\cdot\|$ to denote the spectral norm of a matrix.

2 | SINGLE CHANGE-POINT ESTIMATION

In this section, we provide some background on the SN test and propose an SN test-based method for single change-point estimation, which serves as a building block for the proposed multiple change-point estimation procedure in Section 3. Model assumptions and consistency results are discussed in details to provide intuition and foundation for more involved results in Section 3. For ease of presentation, in the following we assume $d = 1$, in other words, the parameter of interest θ is univariate, and postpone the results for the multivariate case of $d > 1$ to Section 3.3.

2.1 | An SN-based estimation procedure

We start with single change-point estimation in a general parameter $\theta = \theta(F_t)$ for a univariate time series $\{Y_t\}_{t=1}^n$, where F_t denotes the CDF of Y_t and $\theta(\cdot)$ is a general functional. Under the no change-point scenario, $\{Y_t\}_{t=1}^n$ is a stationary time series. Under the single change-point alternative, we follow the framework of Dette and Gösmann (2020) and assume $\{Y_t\}_{t=1}^n$ is generated by

$$Y_t = \begin{cases} Y_t^{(1)}, & 1 \le t \le k_1 \\ Y_t^{(2)}, & k_1 + 1 \le t \le n, \end{cases} \tag{1}$$

where $\{Y_t^{(i)}\}_{t\in\mathbb{Z}}$ is a stationary time series with $Y_t^{(i)} \sim F^{(i)}$ for $i = 1, 2$. Thus we have $F_t = F^{(1)}\mathbf{1}(t \le k_1) + F^{(2)}\mathbf{1}(t > k_1)$. Denote $\theta_1 = \theta(F^{(1)})$ and $\theta_2 = \theta(F^{(2)})$, we have $\delta = \theta_2 - \theta_1 \neq 0$ and the change-point $k_1 = \lfloor n\tau_1 \rfloor$ with $\tau_1 \in (0,1)$. Note that the dependence between $\{Y_t^{(1)}\}$ and

$\{Y_t^{(2)}\}$ is deliberately left unspecified, as the validity of the proposed method does not rely on the specification of the dependence (see Assumption 1(i) for more details).

To detect the existence and further estimate the location of the (potential) single change-point $k_1 = \lfloor n\tau_1 \rfloor$, we propose an SN-based testing approach. Specifically, we define

$$\mathrm{SN}_n = \max_{k=1,\dots,n-1} T_n(k), \quad T_n(k) = D_n(k)^2 / V_n(k), \tag{2}$$

where

$$\begin{aligned} D_n(k) &= \frac{k(n-k)}{n^{3/2}}\big(\widehat{\theta}_{1,k} - \widehat{\theta}_{k+1,n}\big), \\ V_n(k) &= \sum_{i=1}^{k} \frac{i^2(k-i)^2}{n^2k^2}(\widehat{\theta}_{1,i} - \widehat{\theta}_{i+1,k})^2 + \sum_{i=k+1}^{n} \frac{(n-i+1)^2(i-k-1)^2}{n^2(n-k)^2}\big(\widehat{\theta}_{i,n} - \widehat{\theta}_{k+1,i-1}\big)^2, \end{aligned} \tag{3}$$

and for any $1 \le a < b \le n$, $\widehat{\theta}_{a,b} = \theta(\widehat{F}_{a,b})$ where $\widehat{F}_{a,b}$ is the empirical distribution of $\{Y_t\}_{t=a}^{b}$. In other words, $\widehat{\theta}_{a,b}$ denotes the non-parametric estimator of θ based on the sub-sample $\{Y_t\}_{t=a}^{b}$.

When $\theta(\cdot)$ is the mean functional, that is, $\theta(F_t) = \int xF_t(dx)$, the newly defined contrast-based test SN_n in (2) reduces to the CUSUM-based SN test statistic in Shao and Zhang (2010) (cf. equation 4 therein). However, for a non-linear functional $\theta(\cdot)$, such as variance, correlation and quantile, SN_n is not equivalent to the CUSUM-based counterpart and is preferred due to its contrast nature. We refer to Zhang and Lavitas (2018) for more discussion.

Built upon the test statistic defined in (2), the SN-based change-point detection procedure proceeds as follows. For a pre-specified threshold K_n, we declare no change-point if $\mathrm{SN}_n \le K_n$. Given that SN_n exceeds the threshold, we estimate the single change-point location via

$$\widehat{k} = \arg\max_{k=1,\dots,n-1} T_n(k).$$

This SN-based procedure provides a general and unified change-point estimation framework, as it can be implemented for any functional $\theta(\cdot)$ with a non-parametric estimator based on the empirical distribution.

2.2 | Assumptions and theoretical results

To establish the consistency of the SN-based estimation procedure under the general functional setting (1), the key is to track the asymptotic behaviour of $\widehat{\theta}_{a,b}$ for $1 \le a < b \le n$. To achieve this, we operate under the framework of approximately linear functional, which covers important quantities such as mean, variance, covariance, correlation and quantile (Künsch, 1989; Shao, 2010).

Specifically, we assume the sub-sample estimator $\widehat{\theta}_{a,b}$ admits the following expansion on the stationary time series $\{Y_t^{(i)}\}, i = 1, 2$, where

$$\begin{aligned} \widehat{\theta}_{a,b} &= \theta_1 + \frac{1}{b-a+1}\sum_{t=a}^{b} \xi_1\left(Y_t^{(1)}\right) + r_{a,b}^{(1)}, \quad \text{for } b \le k_1, \\ \widehat{\theta}_{a,b} &= \theta_2 + \frac{1}{b-a+1}\sum_{t=a}^{b} \xi_2\left(Y_t^{(2)}\right) + r_{a,b}^{(2)}, \quad \text{for } a > k_1. \end{aligned} \tag{4}$$

In other words, $\widehat{\theta}_{a,b}$ is approximately linear when the sub-sample $\{Y_t\}_{t=a}^{b}$ is stationary. Note that $\xi_1\left(Y_t^{(1)}\right)$ and $\xi_2\left(Y_t^{(2)}\right)$ are indeed the influence functions of the functional $\theta(\cdot)$ (Hampel et al., 1986), which is the leading term for asymptotic behaviour of $\widehat{\theta}_{a,b}$, and $r_{a,b}^{(1)}, r_{a,b}^{(2)}$ are the remainder terms.

To further regulate the behaviour of $\widehat{\theta}_{a,b}$ when the sub-sample $\{Y_t\}_{t=a}^{b}$ is a mixture of two stationary segments, we utilise the concept of *partial influence functions* originated from the robust statistics literature (Pires & Branco, 2002). Specifically, for $a \le k_1 < b$, we assume

$$\widehat{\theta}_{a,b} = \theta(\omega_{a,b}) + \frac{1}{b-a+1}\left[\sum_{t=a}^{k_1}\xi_1\left(Y_t^{(1)}, \omega_{a,b}\right) + \sum_{t=k_1+1}^{b}\xi_2\left(Y_t^{(2)}, \omega_{a,b}\right)\right] + r_{a,b}(\omega_{a,b}), \tag{5}$$

where $\omega_{a,b} = \left(\omega_{a,b}^{(1)}, \omega_{a,b}^{(2)}\right)^\top = \left(\frac{k_1-a+1}{b-a+1}, \frac{b-k_1}{b-a+1}\right)^\top$ denotes the proportion of each stationary segment in $\{Y_t\}_{t=a}^{b}$, $\theta(\omega_{a,b})$ denotes $\theta(\cdot)$ evaluated at the mixture distribution $F^{\omega_{a,b}} = \omega_{a,b}^{(1)}F^{(1)} + \omega_{a,b}^{(2)}F^{(2)}$ and $r_{a,b}(\omega_{a,b})$ is the remainder term. The terms $\xi_1\left(Y_t^{(1)}, \omega_{a,b}\right)$ and $\xi_2\left(Y_t^{(2)}, \omega_{a,b}\right)$ are related to the partial influence functions of the functional $\theta(\cdot)$ evaluated at the mixture distribution $F^{\omega_{a,b}}$. See detailed discussion later.

Note that the expansion (5) generalises (4) under the single change-point scenario. Specifically, define $\omega_{a,b} = (1,0)^\top$ and $(0,1)^\top$ for $b \le k_1$ and $a > k_1$, respectively, (4) can be viewed as a special case of (5) where the mixture distribution is *pure* such that $\xi_1\left(Y_t^{(1)}\right) = \xi_1\left(Y_t^{(1)}, (1,0)^\top\right)$, $r_{a,b}^{(1)} = r_{a,b}\left((1,0)^\top\right)$ and $\xi_2\left(Y_t^{(2)}\right) = \xi_2\left(Y_t^{(2)}, (0,1)^\top\right)$, $r_{a,b}^{(2)} = r_{a,b}\left((0,1)^\top\right)$, respectively.

We now work out the explicit formulation of the expansion (5) under the framework of partial influence function (Pires & Branco, 2002). Denote the mixture weight $\omega = \left(\omega^{(1)}, \omega^{(2)}\right)^\top$ such that $\omega^{(i)} \in [0,1]$, $i = 1, 2$ and $\omega^{(1)} + \omega^{(2)} = 1$. Denote $\theta(\omega, F^{(1)}, F^{(2)}) := \theta\left(\omega^{(1)}F^{(1)} + \omega^{(2)}F^{(2)}\right)$ as the functional $\theta(\cdot)$ evaluated at the mixture $F^{\omega} := \omega^{(1)}F^{(1)} + \omega^{(2)}F^{(2)}$. Definition 1 defines the partial influence function as in Pires and Branco (2002).

Definition 1. The partial influence functions of the functional $\theta(F^{\omega}) = \theta\left(\omega, F^{(1)}, F^{(2)}\right)$ with relation to $F^{(1)}$ and $F^{(2)}$, respectively, are given by

$$IF_1\left(y, \theta(\omega, F^{(1)}, F^{(2)})\right) = \lim_{\epsilon\to 0}\epsilon^{-1}\left[\theta\left(\omega, (1-\epsilon)F^{(1)} + \epsilon\delta_y, F^{(2)}\right) - \theta\left(\omega, F^{(1)}, F^{(2)}\right)\right],$$
$$IF_2\left(y, \theta(\omega, F^{(1)}, F^{(2)})\right) = \lim_{\epsilon\to 0}\epsilon^{-1}\left[\theta\left(\omega, F^{(1)}, (1-\epsilon)F^{(2)} + \epsilon\delta_y\right) - \theta\left(\omega, F^{(1)}, F^{(2)}\right)\right],$$

provided the limits exist, where δ_y is the Dirac mass at y.

To understand the partial influence functions, define $\zeta = \omega^{(1)}\epsilon$, by Definition 1, we have

$$IF_1\left(y, \theta\left(\omega, F^{(1)}, F^{(2)}\right)\right) = \omega^{(1)}\lim_{\zeta\to 0}\zeta^{-1}\left[\theta\left(\left(\delta_y - F^{(1)}\right)\zeta + F^{\omega}\right) - \theta\left(F^{\omega}\right)\right] = \omega^{(1)}\xi_1(y, \omega),$$

where $\xi_1(y, \omega)$ is the Gâteaux derivative of $\theta\left(F^{\omega}\right)$ in the direction $\delta_y - F^{(1)}$. Similarly, $IF_2\left(y, \theta(\omega, F^{(1)}, F^{(2)})\right) = \omega^{(2)}\xi_2(y, \omega)$, where $\xi_2(y, \omega)$ is the Gâteaux derivative of $\theta\left(F^{\omega}\right)$ in the direction $\delta_y - F^{(2)}$.

To establish the expansion (5), note that $\widehat{\theta}_{a,b} = \theta(\widehat{F}_{a,b})$, where $\widehat{F}_{a,b}$ denotes the empirical CDF based on the sub-sample $\{Y_t\}_{t=a}^{b}$. The key observation is that $\widehat{F}_{a,b} = \omega_{a,b}^{(1)}\widehat{F}_{a,k_1} + \omega_{a,b}^{(2)}\widehat{F}_{k_1+1,b}$ with $\omega_{a,b} = \left(\frac{k_1-a+1}{b-a+1}, \frac{b-k_1}{b-a+1}\right)^\top$. In other words, $\widehat{F}_{a,b}$ can be viewed as a mixture of two empirical CDFs

$\widehat{F}_{a,k_1}$ and $\widehat{F}_{k_1+1,b}$ based on stationary segments with CDF $F^{(1)}$ and $F^{(2)}$ respectively. Thus, by the results in Pires and Branco (2002), we have

$$\theta\left(\widehat{F}_{a,b}\right) = \theta\left(F^{\omega_{a,b}}\right) + \frac{1}{k_1 - a + 1}\sum_{t=a}^{k_1} IF_1\left(Y_t^{(1)}, \theta\left(\omega_{a,b}, F^{(1)}, F^{(2)}\right)\right) + \frac{1}{b - k_1}\sum_{t=k_1+1}^{b} IF_2\left(Y_t^{(2)}, \theta\left(\omega_{a,b}, F^{(1)}, F^{(2)}\right)\right) + R\left(\widehat{F}_{a,b} - F^{\omega_{a,b}}\right),$$

where $R\left(\widehat{F}_{a,b} - F^{\omega_{a,b}}\right)$ denotes the remainder term. The expansion (5) follows immediately by substituting the partial influence functions with the Gâteaux derivatives $\xi_1(y, \omega_{a,b})$ and $\xi_2(y, \omega_{a,b})$.

We proceed by imposing the following Assumptions 1–3 on the approximately linear functional $\theta(\cdot)$, which are further verified in Section S.4 of Appendix S1 for the smooth function model (including mean, variance, (auto)-covariance, (auto)-correlation) and in Section S.5 of Appendix S1 for quantile. We refer to Remark 1 in Section 3.2 for more detailed discussion on the verification of assumptions.

Assumption 1. (i) For some $\sigma_1 > 0$ and $\sigma_2 > 0$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\left(\xi_1\left(Y_t^{(1)}\right), \xi_2\left(Y_t^{(2)}\right)\right) \Rightarrow \left(\sigma_1 B^{(1)}(r), \sigma_2 B^{(2)}(r)\right),$$

where $B^{(1)}(\cdot)$ and $B^{(2)}(\cdot)$ are standard Brownian motions.

(ii) $\sup_{k<k_1}\left|\sum_{t=k+1}^{k_1}\xi_1\left(Y_t^{(1)}, \omega_{k+1,n}\right) + \sum_{t=k_1+1}^{n}\xi_2\left(Y_t^{(2)}, \omega_{k+1,n}\right)\right| = O_p(n^{1/2})$,

$\sup_{k>k_1}\left|\sum_{t=1}^{k_1}\xi_1\left(Y_t^{(1)}, \omega_{1,k}\right) + \sum_{t=k_1+1}^{k}\xi_2\left(Y_t^{(2)}, \omega_{1,k}\right)\right| = O_p(n^{1/2})$.

Assumption 2. $\sup_{1\le k\le n} k|r_{1,k}(\omega_{1,k})| + \sup_{1\le k\le n}(n-k+1)|r_{k,n}(\omega_{k,n})| = o_p(n^{1/2})$.

Assumption 1 regulates the behaviour of the (partial) influence function $\xi_1(\cdot)$ and $\xi_2(\cdot)$. Specifically, Assumption 1(i) requires the invariance principle to hold for each stationary segment. Note that the dependence of the two Brownian motions $B^{(1)}(\cdot)$ and $B^{(2)}(\cdot)$ are left unspecified as we do not require a specific dependence structure on $\left\{Y_t^{(1)}\right\}$ and $\left\{Y_t^{(2)}\right\}$. Assumption 1(ii) are tailored to regulate $\widehat{\theta}_{a,b}$ estimated on a mixture of two stationary segments. Assumption 2 requires that the remainder term is asymptotically negligible and is a commonly used assumption in the SN literature (Shao, 2010, 2015).

Assumption 3. Denote $\theta(\omega) = \theta\left(\omega^{(1)}F^{(1)} + \omega^{(2)}F^{(2)}\right)$, where $\omega = \left(\omega^{(1)}, \omega^{(2)}\right)^\top$ is the mixture weight with $\omega^{(i)} \in [0,1], i = 1,2$ and $\omega^{(1)} + \omega^{(2)} = 1$. There exist some constants $0 < C_1 < C_2 < \infty$ such that for any mixture weight ω, we have

$$C_1\omega^{(2)}|\theta_1 - \theta_2| \le |\theta_1 - \theta(\omega)| \le C_2\omega^{(2)}|\theta_1 - \theta_2| \text{ and } C_1\omega^{(1)}|\theta_1 - \theta_2| \le |\theta_2 - \theta(\omega)| \le C_2\omega^{(1)}|\theta_1 - \theta_2|.$$

Assumption 3 regulates the smoothness of $\theta(\omega)$. Intuitively, it means that the functional $\theta(\cdot)$ can distinguish the mixture distribution $\omega^{(1)}F^{(1)} + \omega^{(2)}F^{(2)}$ from $F^{(1)}$ and $F^{(2)}$. For mean functional, we have $\theta(\omega) = \omega^{(1)}\theta_1 + \omega^{(2)}\theta_2$, thus we can set $C_1 = C_2 = 1$ as $\theta(\omega)$ is linear in ω.

Assumption 4. $n\delta^2 \to \infty$ as $n \to \infty$, and K_n satisfies $K_n = (n\delta^2)^\kappa$ for some $\kappa \in \left(\frac{1}{2}, 1\right)$.

Assumption 4 quantifies the asymptotic order of the change size δ and the threshold K_n. Under Assumptions 1–4, Theorem 1 gives the consistency results of the SN-based change-point estimation method for approximately linear functionals.

Theorem 1. (i) *Under the no change-point scenario, suppose Assumptions 1(i) and 2 hold, we have* $SN_n \xrightarrow{\mathcal{D}} G = \sup_{r\in[0,1]} \{B(r) - rB(1)\}^2/V(r)$, *where* $B(\cdot)$ *denotes a standard Brownian motion and* $V(r) = \int_0^r \left[B(s) - (s/r)B(r)\right]^2 ds + \int_r^1 \left[B(1) - B(s) - (1-s)/(1-r)\{B(1) - B(r)\}\right]^2 ds$.

(ii) *Under the one change-point scenario, suppose Assumptions 1–4 hold, we have*

$$\lim_{n\to\infty} P\Big(T_n(\widehat{k}) > K_n \quad \text{and} \quad |\widehat{k} - k_1| \le \iota_n \Big) = 1,$$

for any sequence ι_n *such that* $\iota_n/n \to 0$ *and* $\iota_n^{-2}\delta^{-2}n \to 0$ as $n \to \infty$.

Theorem 1(i) indicates that the asymptotic distribution of SN_n for a general functional $\theta(\cdot)$ coincides with the asymptotic distribution of the CUSUM-based SN test for mean (see theorem 3.1 in Shao & Zhang, 2010). This implies that the same threshold K_n can be used to control false positives (i.e. Type-I error) for change-point detection in various parameters and thus greatly simplifies the implementation of the proposed method. In practice, we recommend to set K_n as the 90% or 95% quantile of G, which can be obtained via simulation as G is pivotal. See Shao and Zhang (2010) for tabulated critical values of G.

Theorem 1(ii) gives the convergence rate of the estimated change-point $\widehat{k}$, providing a unified theoretical justification of the SN-based method for a broad class of functionals. Due to the presence of the self-normaliser $V_n(k)$, which is complex and further varies by k, non-standard technical arguments different from existing techniques in the change-point literature are developed to establish the consistency result. It involves a simultaneous analysis of the contrast statistic $D_n(k)$ and the self-normaliser $V_n(k)$. In general, the localization error rate of SNCP is not optimal (at least for change in mean). However, a simple local refinement procedure can be performed to help achieve the optimal rate. We refer to the discussion following Theorem 2 in Section 3.2 for more details on this matter.

The traditional CUSUM-based estimation procedure in the change-point literature typically admits the form $\max_{k=1,\dots,n-1} |D_n(k)|/\widehat{\sigma}_n$, where theoretical results are derived under the assumption that $\widehat{\sigma}_n$ is a consistent estimator of the LRV, leading to less involved technical analysis than the proposed SN-based estimation. However, in practice, the construction of a consistent $\widehat{\sigma}_n$ involves a bandwidth tuning parameter that is difficult to select, especially under the presence of change-points. For example, in the mean case, using a data-driven bandwidth with the estimation-optimal bandwidth formula in Andrews (1991) could lead to non-monotonic power under the change-point alternative and large size distortion under the null, see Crainiceanu and Vogelsang (2007) and Shao and Zhang (2010). Casini et al. (2021) and Casini and Perron (2021b) further provide a comprehensive theoretical analysis of such phenomenon based on Edgeworth expansion. Additionally, different construction of $\widehat{\sigma}_n$ is required for different functional $\theta(\cdot)$, which can be highly involved and non-trivial for parameters such as correlation and quantile, making the practical implementation challenging.

In contrast, thanks to the self-normaliser $V_n(k)$, the proposed SN-based procedure avoids the challenging estimation of LRV and provides a robust framework that works universally for a broad class of functionals under temporal dependence.

3 | MULTIPLE CHANGE-POINT ESTIMATION

In this section, we further extend the proposed SN-based test to multiple change-point estimation. As in standard change-point literature, we assume $\{Y_t\}_{t=1}^n$ is a piecewise stationary time series and there exist $m_o \geq 0$ unknown number of change-points $0 < k_1 < \cdots < k_{m_o} < n$ that partition $\{Y_t\}_{t=1}^n$ into $m_o + 1$ stationary segments. Define $k_0 = 0$ and $k_{m_o+1} = n$, the ith segment contains stationary observations $\{Y_t\}_{t=k_{i-1}+1}^{k_i}$ that share common behaviour characterised by θ_i, for $i = 1, \ldots, m_o + 1$.

More specifically, we operate under the following data generating process for $\{Y_t\}_{t=1}^n$ such that

$$Y_t = Y_t^{(i)}, \ k_{i-1} + 1 \leq t \leq k_i, \ \text{for} \ \ i = 1, \ldots, m_o + 1, \tag{6}$$

where $\{Y_t^{(i)}\}_{t\in\mathbb{Z}}$ is a stationary time series with CDF $F^{(i)}$ and we require $\theta_i = \theta(F^{(i)}) \neq \theta_{i+1} = \theta(F^{(i+1)})$ for $i = 1, \ldots, m_o$ due to the structural break. Our primary interest is to recover the unknown number and locations of the change-points.

To proceed, we first introduce some notations. For $1 \leq t_1 < k < t_2 \leq n$, we define

$$T_n(t_1, k, t_2) = D_n(t_1, k, t_2)^2 / V_n(t_1, k, t_2), \tag{7}$$

where $D_n(t_1, k, t_2) = \frac{(k-t_1+1)(t_2-k)}{(t_2-t_1+1)^{3/2}} \left(\widehat{\theta}_{t_1,k} - \widehat{\theta}_{k+1,t_2}\right)$, $V_n(t_1, k, t_2) = L_n(t_1, k, t_2) + R_n(t_1, k, t_2)$, and

$$L_n(t_1, k, t_2) = \sum_{i=t_1}^{k} \frac{(i - t_1 + 1)^2 (k - i)^2}{(t_2 - t_1 + 1)^2 (k - t_1 + 1)^2} \left(\widehat{\theta}_{t_1,i} - \widehat{\theta}_{i+1,k}\right)^2,$$

$$R_n(t_1, k, t_2) = \sum_{i=k+1}^{t_2} \frac{(t_2 - i + 1)^2 (i - 1 - k)^2}{(t_2 - t_1 + 1)^2 (t_2 - k)^2} \left(\widehat{\theta}_{i,t_2} - \widehat{\theta}_{k+1,i-1}\right)^2.$$

Note that $T_n(t_1, k, t_2)$ is essentially the proposed SN test defined on sub-sample $\{Y_t\}_{t=t_1}^{t_2}$. Set $t_1 = 1$ and $t_2 = n$, $T_n(t_1, k, t_2) = T_n(1, k, n)$ reduces to the *global* SN test defined in (2) of Section 2.1.

The key observation is that, due to the presence of the self-normaliser V_n, the *global* test statistic $T_n(1, k, n)$ may experience severe power loss under multiple change-point scenarios. The intuition is as follows. Suppose k is a true change-point and $\{Y_t\}_{t=1}^n$ has other change-points besides k. Intuitively, $V_n(1, k, n)$ may observe significant inflation as $L_n(1, k, n)$ and $R_n(1, k, n)$ are based on contrast statistics and their values could significantly inflate due to the existence of other change-points besides k. This can in turn cause $T_n(1, k, n)$ to suffer severe deflation and thus a loss of power. Consequently, a naive combination of the standard binary segmentation (Vostrikova, 1981) and the SN test cannot serve as a viable option for multiple change-point estimation (see both theoretical evidence and numerical illustration of this phenomenon in Section S.1 of Appendix S1).

3.1 | The nested local-window segmentation algorithm

To bypass this issue, we combine the SN test with a novel nested local-window segmentation algorithm, where for each k, instead of one global SN test $T_n(1, k, n)$, we compute a maximal SN test based on a collection of nested windows covering k. Specifically, fix a small $\epsilon \in (0, 1/2)$ such

as $\epsilon = 0.05, 0.1$, define the window size $h = \lfloor n\epsilon \rfloor$. For each $k = h, \ldots, n-h$, we define its nested window set $H_{1:n}(k)$ where

$$H_{1:n}(k) = \left\{ (t_1, t_2) \middle| t_1 = k - j_1 h + 1,\ j_1 = 1, \ldots, \lfloor k/h \rfloor; t_2 = k + j_2 h,\ j_2 = 1, \ldots, \lfloor (n-k)/h \rfloor \right\}.$$

Note that for $k < h$ and $k > n - h$, by definition, we have $H_{1:n}(k) = \emptyset$.

For each $k = 1, \ldots, n$, based on its nested window set $H_{1:n}(k)$, we define a maximal SN test statistic $T_{1,n}(k)$ such that

$$T_{1,n}(k) = \max_{(t_1,t_2)\in H_{1:n}(k)} T_n(t_1, k, t_2),$$

where we set $\max_{(t_1,t_2)\in\emptyset} T_n(t_1, k, t_2) := 0$. Note that unlike the standard binary segmentation, the test statistic $T_{1,n}(k)$ is calculated based on a set of nested *local*-window observations $\{Y_t\}_{t=t_1}^{t_2}$ surrounding the time point k instead of directly based on the *global* observations $\{Y_t\}_{t=1}^{n}$.

This mechanism is precisely designed to alleviate the inflation of the self-normaliser V_n for the SN test under multiple change-point scenarios. With a sufficiently small window size ϵ, for any change-point k, there exists some local-window $(\tilde{t}_1, \tilde{t}_2)$ which contains k as the only change-point, thus the maximal statistic $T_{1,n}(k)$ remains effective thanks to $T_n(\tilde{t}_1, k, \tilde{t}_2)$. In the literature, there exists *pure* local-window-based segmentation algorithms, for example, SaRa in Niu and Zhang (2012) for change in mean, LRSM in Yau and Zhao (2016) for change in AR models. The pure local-window approach only considers the smallest local-window $(k-h+1, k+h)$ when constructing change-point tests for k given a window size h. Such an approach is also employed in the literature of 'piecewise smooth' change, see Wu and Zhao (2007), Bibinger et al. (2017) and Casini and Perron (2021a).

Compared to the *pure* local-window approach, the constructed nested window set $H_{1:n}(k)$ makes our algorithm more adaptive as it helps $T_{1,n}(k)$ retain more power when k is far away from other change-points by utilising larger windows that cover k. We refer to Section S.1.2 of Appendix S1 for more detailed discussion of this point and numerical evidence of the substantial advantage in detection power and estimation accuracy of the proposed nested local-window approach over the pure local-window approach. In addition, since the nested local-window algorithm examines a set of expanding windows instead of a single window, its performance is more robust to the choice of the bandwidth h. This is confirmed by numerical experiments in Section S.2.1 of Appendix S1, where we conduct sensitivity analysis of h and it is seen that performance of the nested local-window is robust and stable w.r.t. the choice of h.

Note that the nested window-based SN statistic $T_{1,n}(k)$ can be viewed as a discretised version of the SN test statistic $\tilde{T}_{1,n}(k) = \max_{1\le t_1<k<t_2\le n} T_n(t_1, k, t_2)$, which is related to the scan statistics (Chan & Walther, 2013) and multi-scale statistics (Frick et al., 2014). However, $\tilde{T}_{1,n}(k)$ is computationally impractical, thus we instead approximate $\tilde{T}_{1,n}(k)$ by $T_{1,n}(k)$ computed on the nested window set $H_{1:n}(k)$.

Based on the maximal test statistic $T_{1,n}(k)$ and a pre-specified threshold K_n, the SN-based multiple change-point estimation (SNCP) proceeds as follows. Starting with the full sample $\{Y_t\}_{t=1}^{n}$, we calculate $T_{1,n}(k), k = 1, \ldots, n$. Given that $\max_{k=1,\ldots,n} T_n(k) \le K_n$, SNCP declares no change-point. Otherwise, SNCP sets $\widehat{k} = \arg\max_{k=1,\ldots,n} T_{1,n}(k)$ and we recursively perform SNCP on the sub-sample $\{Y_t\}_{t=1}^{\widehat{k}}$ and $\{Y_t\}_{t=\widehat{k}+1}^{n}$ until no change-point is declared.

Denote $W_{s,e} = \left\{(t_1, t_2) \middle| s \le t_1 < t_2 \le e\right\}$ and $H_{s:e}(k) = H_{1:n}(k) \bigcap W_{s,e}$, which is the nested window set of k on the sub-sample $\{Y_t\}_{t=s}^{e}$. Define the sub-sample maximal SN test statistic as $T_{s,e}(k) = \max_{(t_1,t_2)\in H_{s:e}(k)} T_n(t_1, k, t_2)$. Algorithm 1 gives the formal description of SNCP.

Algorithm 1. SNCP for multiple change-point estimation

Input: Time series $\{Y_t\}_{t=1}^{n}$, threshold K_n, window size $h = \lfloor n\epsilon \rfloor$.
Output: Estimated change-points set $\widehat{\mathbf{k}} = \left(\widehat{k}_1, \cdots, \widehat{k}_{\widehat{m}}\right)$
Initialization: SNCP$(1, n, K_n, h)$, $\widehat{\mathbf{k}} = \varnothing$
Procedure: SNCP(s, e, K_n, h)

```
if e − s + 1 < 2h then
    Stop
else
    k̂* = arg max_{k=s,···,e} T_{s,e}(k);
    if T_{s,e}(k̂*) ≤ K_n then
        Stop
    else
        k̂ = k̂ ∪ k̂*;
        SNCP(s, k̂*, K_n, h);
        SNCP(k̂* + 1, e, K_n, h);
    end
end
```

3.1.1 | Comparison with popular segmentation algorithms in the literature

We remark that it is possible to combine the proposed SN test statistic with other segmentation algorithms designed for multiple change-point estimation, such as wild binary segmentation (WBS) (Fryzlewicz, 2014) or its variants including narrowest-over-threshold (NOT) (Baranowski et al., 2019) and seeded binary segmentation (SBS) (Kovacs et al., 2020). WBS and NOT use randomly generated intervals for searching multiple change-points, whereas SBS employs deterministic intervals. However, theoretical guarantees for such procedures can be challenging to establish as the above-mentioned segmentation algorithms are mainly used for change-point estimation in a sequence of independent data. Nevertheless, in Section S.2.2 of Appendix S1, we provide an extensive numerical comparison between the proposed nested local-window segmentation algorithm (SNCP) and the combinations of the SN test with WBS, NOT and SBS, where the performance of SNCP is seen to be very competitive in terms of both statistical accuracy and computational efficiency.

3.2 | Assumptions and theoretical results

In this section, we study the theoretical properties of the proposed SNCP for multiple change-point estimation. We operate under the classical infill framework where we assume $k_i/n \to \tau_i \in (0, 1)$ for $i = 1, \ldots, m_o$ as $n \to \infty$. Define $\tau_0 = 0$ and $\tau_{m_o+1} = 1$, we further assume

that $\min_{1\le i\le m_o+1}(\tau_i-\tau_{i-1})=\epsilon_o>\epsilon$, where ϵ is the window size parameter used in SNCP, which imposes an implicit upper bound for m_o such that $m_o\le 1/\epsilon$. This is a common assumption in the literature for change-point testing and estimation under temporal dependence, see Andrews (1993), Bai and Perron (2003), Davis et al. (2006) and Yau and Zhao (2016). In practice, we set ϵ to be a small constant such as $\epsilon=0.05, 0.10, 0.15$, which can be based on prior information about the minimum spacing between consecutive change-points.

In Section S.2.1 of Appendix S1, we conduct an extensive sensitivity analysis of SNCP w.r.t. the window size ϵ and the threshold K_n, and the result indicates SNCP is rather robust to the choices of (ϵ, K_n) as long as $\epsilon_o>\epsilon$, the violation of which could lead to unsatisfactory segmentation results. This suggests that the assumption $\epsilon_o>\epsilon$ is necessary both theoretically and empirically, and hence the proposed SNCP may not be suitable for time series with frequent change-points where ϵ_o is vanishing with $\epsilon_o=o(1)$; see Fryzlewicz (2020) for a recent contribution to detecting frequent change-points.

Denote the true parameter for the ith segment by θ_i and denote the change size by $\delta_i=\theta_{i+1}-\theta_i$ for $i=1,\ \dots\ ,m_o$. For ease of presentation, we assume that $\delta_i=c_i\delta$ for $i=1,\ \dots\ ,m_o$, where $c_i\neq 0$ is a fixed constant. Thus, the overall change size is controlled by δ.

We assume the following expansions for the empirical functional $\widehat{\theta}_{a,b}=\theta(\widehat{F}_{a,b})$, which is a natural extension of the expansions (4) and (5) from the single change-point setting in Section 2.2 to the multiple change-point setting. Specifically, for $\widehat{\theta}_{a,b}$ computed exclusively on the ith stationary segments with $i=1,\ \dots\ ,m_o+1$, we assume

$$\widehat{\theta}_{a,b}=\theta_i+\frac{1}{b-a+1}\sum_{t=a}^{b}\xi_i\big(Y_t^{(i)}\big)+r_{a,b}^{(i)},\quad \text{for } k_{i-1}+1\le a<b\le k_i, \tag{8}$$

where $\xi_i\big(Y_t^{(i)}\big)$ is the influence function of the functional $\theta(\cdot)$ for the ith segment and $r_{a,b}^{(i)}$ denotes the remainder term. For $\widehat{\theta}_{a,b}$ computed based on a mixture of stationary segments, we further assume

$$\begin{aligned}\widehat{\theta}_{a,b}&=\theta(\omega_{a,b})+\frac{1}{b-a+1}\left[\sum_{t=a}^{k_i}\xi_i\big(Y_t^{(i)},\omega_{a,b}\big)+\sum_{l=1}^{j-i}\sum_{t=k_{l+i-1}+1}^{k_{l+i}}\xi_{i+l}\big(Y_t^{(l+i)},\omega_{a,b}\big)+\sum_{t=k_j+1}^{b}\xi_{j+1}\big(Y_t^{(j+1)},\omega_{a,b}\big)\right]\\&\quad+r_{a,b}(\omega_{a,b}):=\theta_{a,b}+\bar{\xi}_{a,b}(\omega_{a,b})+r_{a,b}(\omega_{a,b}),\end{aligned} \tag{9}$$

where $(k_i,k_{i+1},\ \dots\ ,k_j)$ with $i\le j$ denotes the $j-i+1$ true change-points between a and b such that $k_{i-1}+1\le a\le k_i$ and $k_j+1\le b\le k_{j+1}$, and

$$\begin{aligned}\omega_{a,b}&=\Big(\omega_{a,b}^{(1)},\ \dots\ ,\omega_{a,b}^{(m_o+1)}\Big)^{\top}\\&=\left(\overbrace{0,\ \dots\ ,0}^{\text{of } i-1},\frac{k_i-a+1}{b-a+1},\frac{k_{i+1}-k_i}{b-a+1},\ \dots\ ,\frac{k_j-k_{j-1}}{b-a+1},\frac{b-k_j}{b-a+1},\overbrace{0,\ \dots\ ,0}^{\text{of } m_o-j}\right)^{\top},\end{aligned}$$

denotes the proportion of each stationary segment in $\{Y_t\}_{t=a}^{b}$, $\theta(\omega_{a,b})$ denotes $\theta(\cdot)$ evaluated at the mixture distribution $F^{\omega_{a,b}}=\sum_{i=1}^{m_o+1}\omega_{a,b}^{(i)}F^{(i)}$ and $r_{a,b}(\omega_{a,b})$ denotes the remainder term.

Similar to the single change-point scenario, the expansion (8) of $\widehat{\theta}_{a,b}$ with $k_{i-1}+1 \le a < b \le k_i$ can be viewed as a special case of (9) where the mixture distribution is pure and $\omega_{a,b}$ is defined as $\omega_{a,b}^{(i)} = 1$ and $\omega_{a,b}^{(i')} = 0, i' \neq i$. We proceed by making the following assumptions.

Assumption 5. (i) For some $\sigma_i > 0, i = 1, \ldots, m_o + 1$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \left(\xi_1\left(Y_t^{(1)}\right), \ldots, \xi_{m_o+1}\left(Y_t^{(m_o+1)}\right) \right) \Rightarrow \left(\sigma_1 B^{(1)}(r), \ldots, \sigma_{m_o+1} B^{(m_o+1)}(r) \right),$$

where $B^{(i)}(\cdot), i = 1, \ldots, m_o + 1$ are standard Brownian motions.

(ii) $\sup_{1 \le a < b \le n} |(b-a+1)\bar{\xi}_{a,b}(\omega_{a,b})| = O_p(n^{1/2})$.

Assumption 6. $\sup_{1 \le a < b \le n} |(b-a+1) r_{a,b}(\omega_{a,b})| = o_p(n^{1/2})$.

Assumptions 5 and 6 are natural extensions of Assumptions 1 and 2 to the multiple change-point setting and can also be verified for smooth function models and quantile under mild conditions. We refer to Sections S.4 and S.5 of Appendix S1 for more details.

Assumption 7. For $1 \le a < b \le n$, $\theta_{a,b} = \theta(\omega_{a,b})$ can be expressed almost linearly such that

$$\sup_{1 \le a < b \le n} \left| \theta_{a,b} - (\theta_1, \ldots, \theta_{m_o+1}) \omega_{a,b} \right| = \sup_{1 \le a < b \le n} \left| \theta_{a,b} - \sum_{i=1}^{m_o+1} \omega_{a,b}^{(i)} \theta_i \right| = o(n^{-1/2}).$$

Assumption 7 imposes a relatively strong technical condition on the functional $\theta(\cdot)$ such that $\theta_{a,b} \approx \sum_{i=1}^{m_o+1} \omega_{a,b}^{(i)} \theta_i$. Assumption 7 holds trivially for mean change and is typically satisfied when $\theta(\cdot)$ is the only quantity that changes, which is a common assumption in testing-based change-point estimation literature. For example, Assumption 7 holds for variance, (auto)-covariance change with constant mean (Aue et al., 2009; Cho & Fryzlewicz, 2012) and (auto)-correlation change with constant mean and variance (Wied et al., 2012). Numerical experiments conducted in Section 4.6 and Sections S.2.6–S.2.9 of Appendix S1 indicate that SNCP is robust and continues to perform well when Assumption 7 can not be easily verified.

An alternative Assumption 3.3* is provided in Section S.4.2.3 of Appendix S1, which is a natural extension of Assumption 3 to the multiple change-point setting and further includes Assumption 7 as a special case. We defer Assumption 3.3* to the supplement as it is a more involved technical assumption.

Remark 1. (Verification of assumptions): Assumptions 5–7 are high-level assumptions made on a general functional $\theta(\cdot)$ to facilitate presentation. In Sections S.4 and S.5 of Appendix S1, under mild conditions, we provide verification of Assumptions 5–7 for commonly used functionals including the smooth function model and quantile. In general, the assumptions can be verified for mean change, variance and (auto)-covariance change with constant mean or with concurrent small-scale mean change, (auto)-correlation change with constant mean and variance or with concurrent small-scale mean and variance change, and quantile change with density functions that are smooth and bounded. In particular, the verification of Assumption 6 for quantile is highly nontrivial and of independent interest. It essentially provides a uniform Bahadur representation for quantiles in sub-samples. Our result allows for change-points and temporal dependence, and thus generalises the ones in Wu (2005) and Dette and Gösmann (2020).

For $u \in (\epsilon, 1-\epsilon)$, define the scaled limit of $H_{1:n}(k)$ by $H_\epsilon(u) = \left\{ (u_1, u_2) \middle| u_1 = u - j_1\epsilon, j_1 = 1, \ldots, \lfloor u/\epsilon \rfloor; u_2 = u + j_2\epsilon, j_2 = 1, \ldots, \lfloor (1-u)/\epsilon \rfloor \right\}$ and define $\Delta(u_1, u, u_2) =$

$B(u) - B(u_1) - \frac{u-u_1}{u_2-u_1}\{B(u_2) - B(u_1)\}$, where $B(\cdot)$ is a standard Brownian motion. Theorem 2 gives the consistency result of SNCP for multiple change-point estimation.

Theorem 2. *(i) Under the no change-point scenario, and Assumptions* 5(i) *and* 6, *we have*

$$\max_{k=1,\dots,n} T_{1,n}(k) \xrightarrow{\mathcal{D}} G_\epsilon = \sup_{u\in(\epsilon,1-\epsilon)} \max_{(u_1,u_2)\in H_\epsilon(u)} D(u_1,u,u_2)^2/V(u_1,u,u_2), \tag{10}$$

where $D(u_1,u,u_2) = \frac{1}{\sqrt{u_2-u_1}}\Delta(u_1,u,u_2)$ *and* $V(u_1,u,u_2) = \frac{1}{(u_2-u_1)^2}\left(\int_{u_1}^{u}\Delta(u_1,s,u)^2ds + \int_{u}^{u_2}\Delta(u,s,u_2)^2ds\right)$.

(ii) Under the multiple change-point scenario, suppose Assumption 4, *Assumptions* 5, 6 *and* 7 (or 3.3*) *hold and suppose* $\epsilon < \epsilon_o$, *we have*

$$\lim_{n\to\infty} P\Big(\widehat{m} = m_o \quad \text{and} \quad \max_{1\le i\le m_o} |\widehat{k}_i - k_i| \le \iota_n\Big) = 1,$$

for any sequence ι_n *such that* $\iota_n/n \to 0$ *and* $\iota_n^{-2}\delta^{-2}n \to 0$ as $n \to \infty$.

Theorem 2(i) characterises the asymptotic behaviour of SNCP under no change-point and thus provides a natural choice of threshold K_n. In practice, we set K_n as a high quantile, for example, 90% or 95% quantile of G_ϵ to control the Type-I error of SNCP. For a given window size ϵ, G_ϵ is a pivotal distribution and its critical values can be obtained via simulation. Theorem 2(ii) indicates that SNCP can correctly identify the number of change-points m_o with an increasing threshold K_n of a proper order. Note that the localization error rate of SNCP is the same as the single change-point scenario in Theorem 1.

Theorem 2(ii) assumes all changes have the same order δ and requires $\iota_n^{-2}\delta^{-2}n \to 0$ to achieve consistency. In fact, this can be relaxed to allow multi-scale changes and we then require $\iota_n^{-2}\delta_{\max}^2\delta_{\min}^{-4}n \to 0$, where $\delta_{\max}$ and $\delta_{\min}$ denote the maximum and minimum change size. This multi-scale condition matches the one required by Lavielle and Moulines (2000) for multiple change-point estimation in mean under temporal dependence (cf. theorem 3 therein).

Remark 2. (Localization error rate and local refinement): Set the change size $\delta = D_0 n^{-c}$ with $c \in [0, 1/2)$ and $D_0 \neq 0$, Theorem 2(ii) implies that $n^{1/2+c} = o(\iota_n)$. Under the fixed change size ($c = 0$), it implies that the convergence rate ι_n/n of SNCP is at best $1/\sqrt{n}$, which is slower than the optimal rate $1/n$ for change-point estimation in mean under temporal dependence, see Bai (1994) and Lavielle and Moulines (2000).[1] We note that the derived rate is technically difficult to be further improved due to the complex nature of the self-normaliser $V_n(k)$. On the other hand, the derived rate applies to a general functional, which seems not well studied in the literature. Nevertheless, in Section S.8 of Appendix S1, we further propose a simple and intuitive local refinement procedure, which provably improves the localization error rate of SNCP to $1/n$ for the mean functional. The key observation is that by Theorem 2, SNCP can asymptotically isolate each single change-point and thus a simple CUSUM statistic can be used within a well-designed local interval around each estimated change-point $\widehat{k}_i$ by SNCP to achieve further refinement. We refer

[1]For multiple change-point estimation of univariate mean in a sequence of independent sub-Gaussian observations, this is further shown as the minimax optimal localization rate, see Wang et al. (2020), Verzelen et al. (2020) and references therein.

to Sections S.8.1 and S.8.2 for Appendix S1 for more detailed theoretical and numerical results of the procedure.

3.3 | Extension to vector-valued functionals

In this section, we discuss the extension of SNCP to a vector-valued functional, where $\boldsymbol{\theta}(\cdot) \in \mathbb{R}^d$ with $d > 1$. A natural example is change-point detection in mean or covariance matrix of multivariate time series, see for example Aue et al. (2009). Additionally, for a univariate time series, we may be interested in detecting any structural break among multiple parameters of interest, such as examining mean and variance together or examining multiple quantile levels simultaneously.

Note that the dimension of the underlying time series $\{Y_t\}_{t=1}^n$ may or may not equal to that of $\boldsymbol{\theta}$ (i.e. d). For change-point estimation in mean of multivariate time series, we have $\boldsymbol{\theta} = E(Y_t)$ and the dimension of Y_t is d. However, for change-point estimation in covariance matrix ($\boldsymbol{\theta} = \mathrm{Cov}(Y_t)$) or multiple parameters (e.g. $Y_t \in \mathbb{R}$ and $\boldsymbol{\theta} = (E(Y_t), \mathrm{Var}(Y_t))^\top$), the dimension of Y_t can be smaller than d. We examine the performance of SNCP for all three cases via numerical experiments in Section 4.

To accommodate the vector-valued functional, we modify the SN test statistic in (7) such that

$$T_n^*(t_1, k, t_2) = D_n^*(t_1, k, t_2)^\top V_n^*(t_1, k, t_2)^{-1} D_n^*(t_1, k, t_2), \tag{11}$$

where $\widehat{\boldsymbol{\theta}}_{a,b} = \boldsymbol{\theta}(\widehat{F}_{a,b})$ with $\widehat{F}_{a,b}$ being the empirical distribution of $\{Y_t\}_{t=a}^b$ and

$$D_n^*(t_1, k, t_2) = \frac{(k - t_1 + 1)(t_2 - k)}{(t_2 - t_1 + 1)^{3/2}} \left(\widehat{\boldsymbol{\theta}}_{t_1,k} - \widehat{\boldsymbol{\theta}}_{k+1,t_2}\right), \quad V_n^*(t_1, k, t_2) = L_n^*(t_1, k, t_2) + R_n^*(t_1, k, t_2),$$

$$L_n^*(t_1, k, t_2) = \sum_{i=t_1}^{k} \frac{(i - t_1 + 1)^2 (k - i)^2}{(t_2 - t_1 + 1)^2 (k - t_1 + 1)^2} \left(\widehat{\boldsymbol{\theta}}_{t_1,i} - \widehat{\boldsymbol{\theta}}_{i+1,k}\right) \left(\widehat{\boldsymbol{\theta}}_{t_1,i} - \widehat{\boldsymbol{\theta}}_{i+1,k}\right)^\top,$$

$$R_n^*(t_1, k, t_2) = \sum_{i=k+1}^{t_2} \frac{(t_2 - i + 1)^2 (i - 1 - k)^2}{(t_2 - t_1 + 1)^2 (t_2 - k)^2} \left(\widehat{\boldsymbol{\theta}}_{i,t_2} - \widehat{\boldsymbol{\theta}}_{k+1,i-1}\right) \left(\widehat{\boldsymbol{\theta}}_{i,t_2} - \widehat{\boldsymbol{\theta}}_{k+1,i-1}\right)^\top.$$

With a pre-specified threshold K_n, SNCP proceeds as in Algorithm 1 where the only difference is that we replace $T_{s,e}(k)$ with $T_{s,e}^*(k) = \max_{(t_1,t_2) \in H_{s:e}(k)} T_n^*(t_1, k, t_2)$ as defined in (11).

3.3.1 | Limiting distribution under no change-point scenario

We first derive the limiting null distribution of $\max_{k=1,\dots,n} T_{1,n}^*(k)$, which is pivotal and thus provides natural choices of the threshold K_n. We assume the sub-sample estimator $\widehat{\boldsymbol{\theta}}_{a,b}$ for the parameter of interest $\boldsymbol{\theta} \in \mathbb{R}^d$ admits the following expansion

$$\widehat{\boldsymbol{\theta}}_{a,b} = \boldsymbol{\theta}_0 + \frac{1}{b - a + 1} \sum_{t=a}^{b} \xi(Y_t) + r_{a,b},$$

where $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$, $\xi(Y_t) \in \mathbb{R}^d$ denotes the influence function of $\boldsymbol{\theta}$ and $r_{a,b} \in \mathbb{R}^d$ is the remainder term. We further impose the following mild assumptions.

TABLE 1 Critical values of the limiting null distribution $G^*_{\epsilon,d}$ with $\epsilon = 0.05$

$1-\alpha/d$	1	2	3	4	5	6	7	8	9	10
90%	141.9	208.2	275.0	344.4	415.9	492.5	568.4	651.4	740.3	823.5
95%	165.5	237.5	309.1	387.5	464.5	541.7	624.1	713.3	808.6	898.9

Assumption 8. For some positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\xi(Y_t) \Rightarrow \Sigma^{1/2}\mathcal{B}_d(r),$$

where $\mathcal{B}_d(\cdot)$ is a d-dimensional Brownian motion with independent entries.

Assumption 8 is a standard functional central limit theorem (FCLT) result commonly assumed in the SN literature under the no change-point scenario, and can be verified under mild moment and weak dependence conditions, see for example, Shao (2010, assumption 2.1, 2010, assumption 3.1) and Dette and Gösmann (2020, assumption 3.1).

Assumption 9. The remainder term $r_{a,b}$ is asymptotically negligible such that

$$\sup_{1\le a<b\le n}(b-a+1)\|r_{a,b}\|_2 = o_p(n^{1/2}).$$

Proposition 1. *Under the no change-point scenario, given Assumptions 8 and 9, we have*

$$\max_{k=1,\dots,n} T^*_{1,n}(k) \xrightarrow{D} G^*_{\epsilon,d} = \sup_{u\in(\epsilon,1-\epsilon)}\max_{(u_1,u_2)\in H_\epsilon(u)} D^*(u_1,u,u_2)^\top V^*(u_1,u,u_2)^{-1}D^*(u_1,u,u_2),$$

where $D^*(u_1,u,u_2) = \frac{1}{\sqrt{u_2-u_1}}\mathbf{\Delta}(u_1,u,u_2)$ *and* $V^*(u_1,u,u_2) = \frac{1}{(u_2-u_1)^2}\left(\int_{u_1}^{u}\mathbf{\Delta}(u_1,s,u)\,\mathbf{\Delta}(u_1,s,u)^\top ds + \int_{u}^{u_2}\mathbf{\Delta}(u,s,u_2)\mathbf{\Delta}(u_1,s,u)^\top ds\right)$ *with* $\mathbf{\Delta}(u_1,u,u_2) = \mathcal{B}_d(u) - \mathcal{B}_d(u_1) - \frac{u-u_1}{u_2-u_1}[\mathcal{B}_d(u_2) - \mathcal{B}_d(u_1)]$.

The proof of Proposition 1 is straightforward and follows the same argument as the proof of theorem 2.1 in Shao (2010) and the continuous mapping theorem, hence omitted. For a given dimension d and window size ϵ, the limiting distribution $G^*_{\epsilon,d}$ is pivotal and its critical values can be obtained via simulation. Table 1 tabulates the critical values of $G^*_{\epsilon,d}$ for $\epsilon = 0.05$ and $d = 1, \dots, 10$. Note that for $d = 1$, $G^*_{\epsilon,d}$ coincides with the univariate limiting distribution G_ϵ derived in Theorem 2(i).

3.3.2 | Consistency of SNCP

To ease presentation and facilitate understanding, we first establish the consistency of SNCP for change-point estimation in mean of multivariate time series. We then provide further discussions on how to extend the consistency result to a general vector-valued functionals.

Specifically, we operate under the following data generating process for $\{Y_t \in \mathbb{R}^d\}_{t=1}^n$ such that

$$Y_t = X_t + \theta_i, \quad k_{i-1}+1 \le t \le k_i, \quad \text{for } i = 1, \dots, m_o+1,$$

where $\{X_t\}_{t=1}^n$ is a d-dimensional stationary time series with $E(X_t) = 0$, $k_0 := 0 < k_1 < \cdots < k_{m_o} < k_{m_o+1} := n$ denote the (potential) change-points, and $\theta_i \in \mathbb{R}^d$ denotes the mean of the ith segment. We assume that, for $i = 1, \ldots, m_o$, $\theta_{i+1} - \theta_i = \eta_i \delta$ where $\eta_i \in \mathbb{R}^d / \{\mathbf{0}\}$ is a non-zero vector. Thus, the overall change size is controlled by δ.

Same as in Section 3.2, we use the infill framework where we assume $k_i/n \to \tau_i \in (0, 1)$ for $i = 1, \ldots, m_o$ as $n \to \infty$. Define $\tau_0 = 0$ and $\tau_{m_o+1} = 1$, we again require that $\min_{1 \le i \le m_o+1}(\tau_i - \tau_{i-1}) = \epsilon_o > \epsilon$, where ϵ is the window size parameter used in SNCP.

Theorem 3. *Suppose $\{X_t\}_{t=1}^n$ satisfies the invariance principle such that $n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} X_t \Rightarrow \Sigma_X^{1/2} \mathcal{B}_d(r)$, where Σ_X is a positive definite matrix.*

(i) *Under the no change-point scenario, we have* $\max_{k=1,\ldots,n} T^*_{1,n}(k) \xrightarrow{\mathcal{D}} G^*_{\epsilon,d}$.

(ii) *Under the multiple change-point scenario, suppose Assumption 4 hold and suppose $\epsilon < \epsilon_o$, we have*

$$\lim_{n\to\infty} P\Big(\widehat{m} = m_o \quad \text{and} \quad \max_{1\le i\le m_o} |\widehat{k}_i - k_i| \le \iota_n\Big) = 1,$$

for any sequence ι_n such that $\iota_n/n \to 0$ and $\iota_n^{-2}\delta^{-2}n \to 0$ as $n \to \infty$.

Compared to the univariate result in Theorem 2(ii), it can be seen that the same localization rate is obtained in Theorem 3(ii) for the multivariate mean case. However, compared to the univariate proof, the technical argument needed for Theorem 3 is substantially different, which is indeed much more challenging as it requires the analysis of a random matrix and its inverse, since the self-normaliser $V_n^*(t_1, k, t_2)$ is a random matrix in $\mathbb{R}^{d\times d}$ due to the vector nature of the functional $\boldsymbol{\theta}(\cdot)$.

It is easy to see that the result of Theorem 3 can be directly used to establish consistency of SNCP for change-point estimation in covariance matrix of $\{Y_t \in \mathbb{R}^d\}_{t=1}^n$ (assuming constant mean $E(Y_t)$), as the problem can be transformed into multivariate mean change-point estimation for the $(d + d^2)/2$-dimensional time series $\{(Y_{ti} \cdot Y_{tj})_{i\le j}\}_{t=1}^n$, see for example Aue et al. (2009).

Remark 3. (Extension to general vector-valued functionals): To further extend the consistency result in Theorem 3 to a general vector-valued functional $\boldsymbol{\theta}(\cdot)$, we need an additional assumption on the (approximate) linearity of $\boldsymbol{\theta}$, similar to Assumption 7 of the univariate case. Combined with Assumption 8 (FCLT) and 9 (asymptotic negligibility of reminder terms), the same argument used for the multivariate mean in Theorem 3 can then be applied to establish consistency of SNCP for the general functional $\boldsymbol{\theta}$. We omit the details to conserve space.

4 | SIMULATION STUDIES

In this section, we conduct extensive numerical experiments to demonstrate the promising performance of SNCP for a wide range of change-point detection problems under temporal dependence. Under the unified framework of SNCP, we consider change-point estimation for four different settings: mean, covariance matrix, multi-parameter and correlation. In Appendix S1, we further consider change-point estimation for variance, autocorrelation and quantile.

For comparison, we further implement several state-of-the-art non-parametric change-point detection methods in the literature that are explicitly designed to accommodate temporal dependence. Specifically, (A) For mean change, we compare with the classical CUSUM with binary

segmentation (Csörgő & Horváth, 1997) (hereafter CUSUM) and Bai and Perron (2003) (hereafter BP), which are designed for detecting mean change in time series and uses a model selection approach to simultaneously detect all change-points. (B) For covariance matrix change, we compare with the CUSUM method in Aue et al. (2009) (hereafter AHHR). (C) For correlation change, we compare with Galeano and Wied (2017) (hereafter GW), which is essentially a combination of binary segmentation and the correlation change test proposed in Wied et al. (2012). (D) For variance change and autocorrelation change, we compare with Cho and Fryzlewicz (2012) (hereafter MSML) and Korkas and Fryzlewicz (2017) (hereafter KF). Both methods are designed for detecting second-order structural change in time series based on wavelet representation. (E) For multi-parameter change and quantile change, to our best knowledge, there is no existing non-parametric method that works under temporal dependence. For illustration, we compare with the energy statistics-based segmentation in Matteson and James (2014) (hereafter ECP) for multi-parameter change and with the multi-scale quantile segmentation in Vanegas et al. (2021) (hereafter MQS) for quantile change. Both ECP and MQS require temporal independence. All methods are implemented using the recommended setting in the corresponding `R` packages or papers. We refer to Section S.2.12 of Appendix S1 for implementation details of these methods.

4.1 | Implementation details of SNCP

Throughout Section 4, we set the window size ϵ of SNCP to be $\epsilon = 0.05$. We denote SNCP for mean as SNM, for covariance matrix as SNCM, for multi-parameter as SNMP, for correlation as SNC, for variance as SNV, for autocorrelation as SNA, and for quantile as SNQ. In addition, SNM90 denotes SNM using 90% quantile (i.e. critical value at $\alpha = 0.1$) of the limiting null distribution $G^*_{\epsilon,d}$ as the threshold K_n, and similarly for other types of change and levels of critical value. For the power analysis in Sections 4 and real data applications in Section S.3, the threshold K_n for SNCP is set at 90% quantile of $G^*_{\epsilon,d}$ (i.e. $\alpha = 0.1$), which can be found in Table 1 for $d = 1, 2, \ldots, 10$.

We remark that the performance of SNCP is robust w.r.t. the window size ϵ and the quantile level α as the limiting distribution $G^*_{\epsilon,d}$, and thus the threshold K_n, adapt to the effect of ϵ and α. We refer to Section S.2.1 of Appendix S1 for a detailed sensitivity analysis.

4.2 | Error measures of change-point estimation

To assess the accuracy of change-point estimation, we use the Hausdorff distance and adjusted Rand index (ARI). The Hausdorff distance is defined as follows. Denote the set of true (relative) change-points as $\boldsymbol{\tau}_o$ and the set of estimated (relative) change-points as $\hat{\boldsymbol{\tau}}$, we define $d_1(\boldsymbol{\tau}_o, \hat{\boldsymbol{\tau}}) = \max_{\tau_1 \in \hat{\boldsymbol{\tau}}} \min_{\tau_2 \in \boldsymbol{\tau}_o} |\tau_1 - \tau_2|$ and $d_2(\boldsymbol{\tau}_o, \hat{\boldsymbol{\tau}}) = \max_{\tau_1 \in \boldsymbol{\tau}_o} \min_{\tau_2 \in \hat{\boldsymbol{\tau}}} |\tau_1 - \tau_2|$, where d_1 measures the over-segmentation error of $\hat{\boldsymbol{\tau}}$ and d_2 measures the under-segmentation error of $\hat{\boldsymbol{\tau}}$. The Hausdorff distance is $d_H(\boldsymbol{\tau}_o, \hat{\boldsymbol{\tau}}) = \max(d_1(\boldsymbol{\tau}_o, \hat{\boldsymbol{\tau}}), d_2(\boldsymbol{\tau}_o, \hat{\boldsymbol{\tau}}))$. The ARI is originally proposed in Morey and Agresti (1984) as a measure of similarity between two different partitions of the same observations for evaluating the accuracy of clustering. Under the change-point setting, we calculate the ARI between partitions of the time series given by $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_o$. Ranging from 0 to 1, a higher ARI indicates more coherence between the two partitions by $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_o$ and thus more accurate change-point estimation.

TABLE 2 Performance under no change-point scenario with $m_o = 0$

$n = 1024$	$\rho = -0.8$			$\rho = -0.5$			$\rho = 0$			$\rho = 0.5$			$\rho = 0.8$		
$\hat{m}$	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**
SNM90	0.99	0.01	0.00	0.96	0.04	0.00	0.93	0.06	0.00	0.87	0.12	0.01	0.60	0.30	0.10
BP	1.00	0.00	0.00	1.00	0.00	0.00	0.99	0.01	0.00	0.35	0.12	0.53	0.00	0.00	1.00
SNV90	0.80	0.18	0.02	0.90	0.09	0.01	0.90	0.09	0.01	0.86	0.12	0.01	0.73	0.22	0.05
KF	0.18	0.20	0.63	0.76	0.14	0.10	0.96	0.03	0.01	0.95	0.04	0.01	0.94	0.04	0.02
MSML	0.48	0.33	0.19	0.84	0.15	0.01	0.92	0.08	0.00	0.92	0.08	0.00	0.90	0.09	0.00
$n = 4096$	$\rho = -0.8$			$\rho = -0.5$			$\rho = 0$			$\rho = 0.5$			$\rho = 0.8$		
$\hat{m}$	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**	**0**	**1**	**≥ 2**
SNM90	0.94	0.06	0.00	0.89	0.10	0.00	0.89	0.10	0.01	0.88	0.11	0.01	0.84	0.14	0.02
BP	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.49	0.13	0.38	0.00	0.00	1.00
SNV90	0.88	0.12	0.00	0.90	0.10	0.01	0.91	0.08	0.00	0.90	0.09	0.01	0.85	0.13	0.02
KF	0.02	0.01	0.97	0.54	0.17	0.29	0.90	0.06	0.04	0.92	0.05	0.04	0.88	0.06	0.06
MSML	0.38	0.27	0.36	0.80	0.18	0.02	0.92	0.08	0.00	0.92	0.08	0.00	0.90	0.10	0.00

4.3 | No change

We first investigate the performance of SNCP under the null, where the time series is stationary with no change-point. We report the performance of SNM and SNV observed in extensive numerical experiments. The performance of SNCP for other functionals is similar and thus omitted.

We simulate a stationary univariate time series $\{Y_t\}_{t=1}^n$ from an AR(1) process $Y_t = \rho Y_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is *i.i.d.* standard normal $N(0,1)$. We set $n = 1024, 4096$[2] and vary $\rho \in \{-0.8, -0.5, 0, 0.5, 0.8\}$ to examine robustness of SNCP against false positives (i.e. Type-I error) under different direction and strength of temporal dependence. Section S.2.3 of Appendix S1 further provides the simulation results for $n = 512$. For each combination of (n, ρ), we repeat the simulation 1000 times.

The numerical result is summarised in Table 2, where we report the proportion of $\hat{m} = 0$, $\hat{m} = 1$ and $\hat{m} \geq 2$ among 1000 experiments. In general, the observation is as follows. SNCP gives satisfactory performance under moderate temporal dependence with $|\rho| \leq 0.5$ for all sample sizes and its performance further improves as the sample size n increases.

BP performs well under $\rho = -0.8, -0.5, 0$ but exhibits severe over-rejection under positive temporal dependence for $\rho = 0.5, 0.8$ and the performance does not improve as n increases. KF and MSML perform well under $\rho = 0, 0.5, 0.8$ but produce high proportion of false positives under negative temporal dependence for $\rho = -0.5, -0.8$ and the performance does not improve as n increases. Overall, SNCP provides reasonably accurate size under different direction and strength of temporal dependence and achieves the target size as the sample size n increases.

[2] n is deliberately set as power of 2 as MSML in Cho and Fryzlewicz (2012) can only handle such sample size.

4.4 | Change in mean

For mean change, we first simulate a stationary d-dimensional time series $\{X_t = (X_{t1}, \ldots, X_{td})\}_{t=1}^n$ from a VAR(1) process with $X_t = \rho \mathbf{I}_d X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is *i.i.d.* standard d-variate normal $N(0, \mathbf{I}_d)$, and $\mathbf{I}_d$ denotes the d-dimensional identity matrix. We then generate time series $\{Y_t\}_{t=1}^n$ with piecewise constant mean based on $\{X_t\}_{t=1}^n$.

$$\text{(M1)}: n = 600, \quad \rho = 0.2, \quad Y_t = \begin{cases} 0 + X_t, & t \in [1, 100], [201, 300], [401, 500], \\ 2/\sqrt{d} + X_t, & t \in [101, 200], [301, 400], [501, 600]. \end{cases}$$

$$\text{(M2)}: n = 1000, \quad \rho = 0.5, \quad Y_t = \begin{cases} -3/\sqrt{d} + X_t, & t \in [1, 75], [526, 575], \\ 0 + X_t, & t \in [76, 375], [426, 525], [576, 1000], \\ 3/\sqrt{d} + X_t, & t \in [376, 425]. \end{cases}$$

$$\text{(M3)}: n = 2000, \quad \rho = -0.7, \quad Y_t = \begin{cases} 0.4/\sqrt{d} + X_t, & t \in [1, 1000], [1501, 2000], \\ 0 + X_t, & t \in [1001, 1500]. \end{cases}$$

(M1) has evenly spaced change-points with moderate temporal dependence, (M2) features abrupt changes where shortest segments have only 50 or 75 time points with change-points mainly located at the first half of the time series, and (M3) has longer segments with small-scale changes. Typical realisations of (M1)–(M3) for $d = 1$ can be found in Figure S2 of Appendix S1.

Note that the change size in (M1)–(M3) is normalised by $\sqrt{d}$ to keep the signal-to-noise ratio the same across time series of different dimensions. This enables us to isolate and examine the effect of dimension d on estimation. Intuitively, a larger d makes the estimation more difficult as the quality of finite sample approximation by FCLT worsens for higher dimension.

We set the dimension $d = 1, 5, 10$. Note that BP only works for $d = 1$ (i.e. univariate time series) and thus is not included in the comparison for $d = 5, 10$. The estimation results for $d = 1$ and $d = 5$ are summarised in Table 3, where we report the distribution of $\hat{m} - m_o$, average ARI, over- and under-segmentation errors d_1, d_2 and Hausdorff distance d_H among 1000 experiments. The estimation result for $d = 10$ can be found in Table S9 of Appendix S1.

4.4.1 | Univariate time series

$d = 1$: For (M1), all methods perform well overall, though CUSUM tends to greatly over-estimate the number of change-points m_o, as reflected by the distribution of $\hat{m} - m_o$. For (M2), SNM tends to slightly under-estimate m_o (missing a short segment) while BP and CUSUM severely over-estimate m_o and provide much less accurate estimation with noticeably larger Hausdorff distance d_H and smaller ARI. For (M3), which corresponds to strong negative dependence, BP experiences severe power loss and have large under-segmentation error d_2. In summary, BP and CUSUM are prone to produce false positives under positive dependence, and BP may lose power under strong negative dependence. SNM is robust but may experience power loss when detecting short segment changes.

TABLE 3 Performance of SNM, BP, CUSUM under change in mean for $d = 1$ and 5

		$\hat{m} - m_o$											
Method	**Model**	**≤ −3**	**−2**	**−1**	**0**	**1**	**2**	**≥ 3**	**ARI**	$d_1 \times 10^2$	$d_2 \times 10^2$	$d_H \times 10^2$	**time**
SNM	(*M*1)	0	0	9	974	17	0	0	0.960	0.87	0.90	1.01	1.75
BP		0	0	0	847	142	11	0	0.974	1.48	0.50	1.48	9.10
CUSUM		0	0	0	438	414	119	29	0.944	4.43	0.53	4.43	0.05
SNM	(*M*2)	0	11	196	749	43	1	0	0.970	1.33	1.77	2.67	3.55
BP		0	0	0	425	226	203	146	0.863	11.68	0.19	11.68	34.04
CUSUM		2	0	15	365	341	190	87	0.821	10.63	2.86	10.76	0.06
SNM	(*M*3)	0	0	1	986	13	0	0	0.969	1.11	0.80	1.14	10.59
BP		0	371	6	623	0	0	0	0.616	0.33	19.03	19.03	179.75
CUSUM		0	0	0	947	53	0	0	0.965	1.32	0.88	1.32	0.09
		$\hat{m} - m_o$											
Method	**Model**	**≤ −3**	**−2**	**−1**	**0**	**1**	**2**	**≥ 3**	**ARI**	$d_1 \times 10^2$	$d_2 \times 10^2$	$d_H \times 10^2$	**time**
SNM	(*M*1)	0	0	13	946	41	0	0	0.953	1.16	1.12	1.37	12.48
CUSUM	$d = 5$	167	0	0	230	336	189	78	0.783	5.18	11.41	15.71	0.04
SNM	(*M*2)	0	11	175	628	166	18	2	0.937	4.59	1.93	5.68	22.88
CUSUM	$d = 5$	63	5	5	98	161	213	455	0.626	18.02	5.77	20.88	0.07
SNM	(*M*3)	0	0	4	993	3	0	0	0.968	0.93	0.96	1.03	60.00
CUSUM	$d = 5$	0	70	0	928	2	0	0	0.896	1.02	4.50	4.52	0.07

4.4.2 | Multivariate time series

$d = 5, 10$: For (M1) and (M3), the estimation accuracy of SNM is remarkably robust to the increasing dimension, where the ARI and d_H achieved by SNM only worsen slightly from $d = 1$ to $d = 5$. This also holds true for $d = 10$ (see Table S9 of Appendix S1). For (M2), with abrupt changes and strong positive temporal dependence, SNM is less robust to the increasing dimension and gives more false positives for $d = 5, 10$, however, its performance is still decent as measured by ARI and d_H. On the contrary, for all three models (M1)–(M3), the performance of CUSUM worsens significantly from $d = 1$ to $d = 5$ (and even more so for $d = 10$).

4.5 | Change in covariance matrix

For covariance matrix change, we adopt the simulation settings in Aue et al. (2009) and detect change in covariance matrices of a four-dimensional time series $\{Y_t = (Y_{t1}, \ldots, Y_{t4})\}_{t=1}^{n}$ with $n = 1000$. Thus, the number of parameters in the covariance matrix is $d = (4 \times 5)/2 = 10$. Denote Σ_ρ as an exchangeable covariance matrix with unit variance and equal covariance ρ, we consider

TABLE 4 Performance of SNCM and AHHR under change in covariance matrix

		$\hat{m}-m_o$											
Method	**Model**	≤ -3	-2	-1	**0**	**1**	**2**	≥ 3	**ARI**	$d_1 \times 10^2$	$d_2 \times 10^2$	$d_H \times 10^2$	**time**
SNCM	($C1$)	0	1	19	951	29	0	0	0.923	2.13	2.46	2.78	56.44
AHHR		0	221	0	687	82	10	0	0.721	2.45	12.37	13.50	0.44
SNCM	($C2$)	0	0	59	902	39	0	0	0.898	2.53	3.95	4.37	55.17
AHHR		0	0	1	792	168	32	7	0.896	4.97	2.34	5.00	0.56
Method	**Model**	$\hat{m}=0$	$\hat{m}=1$	$\hat{m}\geq 2$									
SNCM	($C0$)	916	80	4									
AHHR		932	59	9									

$$\text{(C0)}: Y_t = 0.3\mathbf{I}_4 Y_{t-1} + \mathbf{e}_t,\ \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \Sigma_{0.5}),\ t \in [1, 1000].$$

$$\text{(C1)}: Y_t = \begin{cases} L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [1, 333], \\ \sqrt{3} L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [334, 667], \\ L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [668, 1000]. \end{cases}$$

$$\text{(C2)}: Y_t = \begin{cases} L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [1, 333], \\ \sqrt{3} L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [334, 667], \\ 3 L_0 F_t + \mathbf{e}_t, & \mathbf{e}_t \overset{i.i.d.}{\sim} N(0, \mathbf{I}_4), \quad t \in [668, 1000]. \end{cases}$$

Here, $\{F_t\}_{t=1}^n$ is a two-dimensional stationary VAR(1) process with the transition matrix $0.3\mathbf{I}_2$ and $L_0 = [1, 1, 0, 0; 0, 0, 1, 1]$ denotes the factor loading matrix. (C1) and (C2) generate covariance changes in the dynamic factor model, which is widely used in the time series literature. We refer to Section S.2.10 of Appendix S1 for additional simulation settings with covariance changes in VAR models. The estimation result is reported in Table 4. For monotonic changes (C2), both methods perform well though AHHR tends to over-estimate the number of change-points, while for non-monotonic changes (C1), AHHR seems to over-estimate and experience power loss at the same time and is outperformed by SNCM. For (C0), both methods give decent performance under moderate temporal dependence with SNCM achieving the target size more accurately.

4.6 | Change in multi-parameter

As discussed before, one notable advantage of SNCP is its universal applicability, where it treats change-point detection for a broad class of parameters in a unified fashion. To conserve space, we refer to Sections S.2.5, S.2.6, S.2.7 and S.2.8 of the Appendix S1 for extensive numerical evidence of the favourable performance of SNCP for change-point detection in variance, auto-correlation, correlation and quantile.

In this section, we further consider change-point estimation for multi-parameter of a univariate time series, where we aim to detect any structural break among multiple parameters of

interest. This can be useful for practical scenarios where one does not know the exact nature of the change but wishes to detect any change among a group of parameters of interest. For example, if one is interested in central tendency of the time series, SNMP can be used to simultaneously detect change in mean and median, while if the user suspects there is change in the dispersion/volatility of the data, SNMP can be used to detect change jointly in variance and high quantiles.

In some sense, this is related to change-point detection in distribution (e.g. ECP, Matteson & James, 2014), where the focus is to detect any change in the marginal distribution of a univariate time series. In theory, algorithms that target distributional change can capture all potential changes in the data. However, it only informs users the existence of a change but is unable to narrow down the specific type of change (e.g. is the detected change in central tendency or in volatility?). This can be less informative in real data analysis when the practitioner is particularly concerned about one certain behaviour change of the data and may also lead to potential power loss compared to methods that target a specific type of change. In addition, existing methods on distributional change typically require the temporal independence assumption, such as ECP in Matteson and James (2014).

We consider two simulation settings with $n = 1000$, and compare the performance of SNMP and ECP.

$$\text{(MP1)}: Y_t = \begin{cases} X_t, & t \in [1, 333], \\ F^{-1}(\Phi(X_t)), & t \in [334, 667], \\ X_t, & t \in [668, 1000]. \end{cases} \qquad \text{(MP2)}: Y_t = \begin{cases} \epsilon_t, & t \in [1, 333], \\ 1.6\epsilon_t, & t \in [334, 667], \\ \epsilon_t, & t \in [668, 1000]. \end{cases}$$

For (MP1), $\{X_t\}_{t=1}^n$ follows an AR(1) process with $X_t = \rho X_{t-1} + \sqrt{1-\rho^2}\epsilon_t$ where $\rho = 0.2$ and $\{\epsilon_t\}$ is *i.i.d.* $N(0, 1)$, $\Phi(\cdot)$ denotes the CDF of $N(0, 1)$, and $F(\cdot)$ denotes a mixture of a truncated normal and a generalised Pareto distribution such that $F^{-1}(q) = \Phi^{-1}(q)$ for $q \leq 0.5$ and $F^{-1}(q) \neq \Phi^{-1}(q)$ for $q > 0.5$. Thus, for (MP1), the change originates from upper quantiles. We refer to Section S.2.8 of Appendix S1 for the detailed definition of $F(\cdot)$ and its motivation from financial applications. For (MP2), $\{\epsilon_t\}_{t=1}^n$ is *i.i.d.* $N(0, 1)$, thus we have temporal independence and the change is solely driven by variance.

The estimation result is summarised in Table 5. We compare the performance of SNCP based on individual parameters and their multi-parameter combination. For clarity, we specify the multi-parameter set that SNMP targets. For example, SNQ_{90}V denotes the SNMP that targets 90% quantile and variance simultaneously. For (MP1), SNQ_{90} and SNQ_{95} perform well as the change originates from upper quantiles, and further improvement can be achieved by combining them into multi-parameter $\text{SNQ}_{90,95}$. Similarly, including variance in the multi-parameter set further improves the estimation accuracy. ECP provides decent performance but tends to over-estimate due to the temporal dependence of the time series. For (MP2), since the change is solely driven by variance, SNV gives the best performance, while quantile-based detection, such as SNQ_{90} experiences power loss. However, the multi-parameter detection based on $\text{SNQ}_{10,90}$ and $\text{SNQ}_{10,20,80,90}$ provide much improved performance over SNQ_{90}, though similar to ECP, they do experience certain power loss compared to SNV. Moreover, SNMP performs competently compared to SNV once variance is included in the multi-parameter set.

This numerical study clearly demonstrates the versatility of SNCP, where it can be effortlessly tailored to target various types of parameter change and their multi-parameter combination. Moreover, compared to detection based on an individual parameter, multi-parameter detection tends to enhance power and improve estimation accuracy when the underlying change affects

TABLE 5 Performance of SNMP and ECP under change in multi-parameter

		$\hat{m}-m_o$											
Method	**Model**	≤ -3	-2	-1	**0**	**1**	**2**	≥ 3	**ARI**	$d_1 \times 10^2$	$d_2 \times 10^2$	$d_H \times 10^2$	**time**
SNQ_{90}	(*MP*1)	0	10	132	805	50	3	0	0.839	3.25	7.26	7.85	17.74
SNQ_{95}		0	5	100	820	73	2	0	0.868	3.16	5.70	6.62	17.20
SNV		0	2	110	832	54	2	0	0.869	2.45	5.47	6.06	12.20
$SNQ_{90,95}$		0	3	82	850	62	3	0	0.878	3.01	4.88	5.67	39.56
SNQ_{90}V		0	0	56	869	70	5	0	0.891	3.04	3.95	4.77	30.96
SNQ_{95}V		0	2	64	861	68	5	0	0.889	2.92	4.30	5.14	30.81
$SNQ_{90,95}$V		0	2	48	882	66	2	0	0.894	2.95	3.79	4.58	49.72
ECP		0	0	0	730	144	92	34	0.850	6.33	3.68	6.41	10.58
SNV	(*MP*2)	0	0	14	956	28	2	0	0.928	2.15	2.13	2.60	12.28
SNQ_{90}		0	71	282	596	48	3	0	0.705	4.10	15.72	16.33	17.50
$SNQ_{10,90}$		0	13	165	788	32	2	0	0.826	3.00	8.36	8.84	39.62
SNQ_{90}V		0	0	32	929	39	0	0	0.913	2.42	2.95	3.45	30.92
$SNQ_{10,90}$V		0	1	50	917	32	0	0	0.903	2.37	3.67	4.06	49.74
$SNQ_{10,20,80,90}$		0	5	118	816	60	1	0	0.849	3.41	6.51	7.27	68.96
ECP		0	49	46	807	79	15	4	0.833	3.43	6.78	7.58	9.96

several parameters in the considered multi-parameter set. We further illustrate this point in more details via real data analysis in Section S.3.2.

For each estimated change-point by SNMP, one may want to identify which features actually changed. One informal strategy is to further conduct a subsequent SN-test. Specifically, for each estimated change-point, based on a well-designed local interval, we can further conduct a single change-point SN test via (2) for each feature and determine if it is changed at this very change-point. Though this procedure is obviously subject to multiple testing issues, it can shed some light on which feature actually changed. We refer to Section S.8.1 for more details of this informal procedure.

5 | CONCLUSION

In this paper, we present a novel and unified framework for time series segmentation in multivariate time series with rigorous theoretical guarantees. Our proposed method is motivated by the recent success of the SN method (Shao, 2015) and advances the methodological and theoretical frontier of statistics literature on change-point estimation by adapting the general framework of approximately linear functional in Künsch (1989). Our method is broadly applicable to the estimation of piecewise stationary models defined in a general functional. In terms of statistical theory, the consistency and convergence rate of change-point estimation are established under the multiple change-points setting for the first time in the literature of SN-based change-point analysis.

For future research, it may be desirable to relax the piecewise constant assumption and allow the parameter to vary smoothly within each segment; see Wu and Zhou (2019) for such a formulation in non-parametric trend models and Casini and Perron (2021a) in locally stationary time series.

ACKNOWLEDGEMENTS

We are grateful to Editor Professor Qiwei Yao, Associate Editor and two anonymous referees for their insightful comments and suggestions, which led to substantial improvements. Zifeng Zhao is supported in part by National Science Foundation grants DMS-2014053; Feiyu Jiang is supported in part by Shanghai Sailing Program No.22YF1402400, NSFC 71973077 and 11771239; Xiaofeng Shao is supported in part by National Science Foundation grants DMS-1807032, DMS-2014018.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Zifeng Zhao https://orcid.org/0000-0001-8466-1321
Feiyu Jiang https://orcid.org/0000-0002-6243-6200
Xiaofeng Shao https://orcid.org/0000-0001-5822-8209

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817–858.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4), 821–856.

Aue, A., Hörmann, S., Horváth, L. & Reimherr, M. (2009) Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B), 4046–4087.

Aue, A. & Horváth, L. (2013) Structural breaks in time series. *Journal of Time Series Analysis*, 34(1), 1–16.

Bai, J. (1994) Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15(5), 453–472.

Bai, J. & Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47–78.

Bai, J. & Perron, P. (2003) Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1–22.

Baranowski, R., Chen, Y. & Fryzlewicz, P. (2019) Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B*, 81(3), 649–672.

Betken, A. & Wendler, M. (2018) Subsampling for general statistics under long range dependence with application to change point analysis. *Statistica Sinica*, 28(3), 1199–1224.

Bibinger, M., Jirak, M. & Vetter, M. (2017) Nonparametric change-point analysis of volatility. *The Annals of Statistics*, 45(4), 1542–1578.

Billingsley, P. (1968) *Convergence of probability measures*. Hoboken: John Wiley & Sons.

Brodsky, E. & Darkhovsky, B.S. (2013) *Nonparametric methods in change point problems*. Netherlands: Springer.

Casini, A., Deng, T. & Perron, P. (2021) *Theory of low frequency contamination from nonstationarity and misspecification: Consequences for HAR inference*. arXiv preprint arXiv:2103.01604.

Casini, A. & Perron, P. (2019) Structural breaks in time series. In: *Oxford Research Encyclopedia of Economics and Finance*. Oxford: Oxford University Press.

Casini, A. & Perron, P. (2021a) *Change-point analysis of time series with evolutionary spectra*. arXiv preprint arXiv:2106.02031.

Casini, A. & Perron, P. (2021b) *Minimax MSE bounds and nonlinear VAR prewhitening for long-run variance estimation under nonstationarity*. arXiv preprint arXiv:2103.02235.

Chan, H.P. & Walther, G. (2013) Detection with the scan and the average likelihood ratio. *Statistica Sinica*, 23(1), 409–428.

Chan, N.H., Ng, W.L. & Yau, C.Y. (2021) A self-normalized approach to sequential change-point detection for time series. *Statistica Sinica*, 31(1), 491–517.

Cho, H. & Fryzlewicz, P. (2012) Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22(1), 207–229.

Crainiceanu, C. & Vogelsang, T. (2007) Nonmonotonic power for tests of mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77(6), 457–476.

Csörgő, M. & Horváth, L. (1997) *Limit theorems in change-point analysis. Wiley Series in Probability and Statistics*. New York: Wiley.

Davis, R., Lee, T.C.M. & Rodriguez-Yam, G. (2006) Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473), 223–239.

Dette, H. & Gösmann, J. (2020) A likelihood ratio approach to sequential change point detection. *Journal of the American Statistical Association*, 115(531), 1361–1377.

Dette, H., Kokot, K. & Volgushev, S. (2020) Testing relevant hypotheses in functional time series via self-normalization. *Journal of Royal Statistical Society: Series B*, 82(3), 629–660.

Eichinger, B. & Kirch, C. (2018) A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1), 526–564.

Frick, K., Munk, A. & Sieling, H. (2014) Multiscale change point inference. *Journal of the Royal Statistical Society: Series B*, 76(3), 495–580.

Fryzlewicz, P. (2014) Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42(6), 2243–2281.

Fryzlewicz, P. (2020) Detecting possibly frequent change-points: wild binary segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society (with discussion)*, 49(4), 1027–1070.

Fryzlewicz, P. & Subba-Rao, S. (2014) Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *Journal of the Royal Statistical Society: Series B*, 76(5), 903–924.

Galeano, P. & Wied, D. (2017) Dating multiple change points in the correlation matrix. *Test*, 26(2), 331–352.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A. (1986) *Robust statistics: the approach based on influence functions*. New York: John Wiley.

Hoga, Y. (2018) A structural break test for extremal dependence in β-mixing random vectors. *Biometrika*, 105(3), 627–643.

Jiang, F., Zhao, Z. & Shao, X. (2020) Time series analysis of COVID-19 infection curve: a change-point perspective. *Journal of Econometrics to appear*.

Jiang, F., Zhao, Z. & Shao, X. (2022) Modelling the COVID-19 infection trajectory: a piecewise linear quantile trend model. *Journal of the Royal Statistical Society: Series B to appear*.

Killick, R., Fearnhead, P. & Eckley, I. (2012) Optimal detection of change-points with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590–1598.

Korkas, K.K. & Fryzlewicz, P. (2017) Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, 27(1), 287–311.

Kovacs, S., Li, H., Bühlmann, P. & Munk, A. (2020) *A seeded binary segmentation: a general methodology for fast and optimal change point detection*. arxiv. Available from: https://arxiv.org/abs/2002.06633.

Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.

Lavielle, M. & Moulines, E. (2000) Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1), 33–59.

Matteson, D. & James, N. (2014) A nonparametric approach for multiple change-point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505), 334–345.

Morey, L.C. & Agresti, A. (1984) The measurement of classification agreement: an adjustment to the rand statistic for chance agreement. *Educational and Psychological Measurement*, 44(1), 33–37.

National Research Council. (2013) *Frontiers in massive data analysis*. Washington, DC: The National Academies Press.

Niu, Y.S. & Zhang, H. (2012) The screening and ranking algorithm to detect DNA copy number variations. *Annals of Applied Statistics*, 6(3), 1306–1326.

Olshen, A.B., Venkatraman, S., Lucito, R. & Wigler, M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, 5(4), 557–572.
Pires, A.M. & Branco, J.A. (2002) Partial influence functions. *Journal of Multivariate Analysis*, 83(2), 451–468.
Preuss, P., Puchstein, R. & Dette, H. (2015) Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, 110(510), 654–668.
Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B*, 72(3), 343–366.
Shao, X. (2015) Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Association*, 110(512), 1797–1817.
Shao, X. & Zhang, X. (2010) Testing for change points in time series. *Journal of the American Statistical Association*, 105(491), 1228–1240.
Tartakovsky, A., Nikiforov, I. & Basseville, M. (2014) *Sequential analysis: hypothesis testing and change-point detection*. Boca Raton: CRC Press.
Truong, C., Oudre, L. & Vayatis, N. (2020) Selective review of offline change point detection methods. *Signal Processing*, 167, 107299.
Vanegas, L.J., Behr, M. & Munk, A. (2021) Multiscale quantile segmentation. *Journal of the American Statistical Association to appear*.
Verzelen, N., Fromont, M., Lerasle, M. & Reynaud-Bouret, P. (2020) *Optimal change-point detection and localization*. arXiv preprint arXiv:2010.11470.
Vostrikova, L.Y. (1981) Detecting disorder in multidimensional random processes. *Doklady Akademii Nauk*, 259, 270–274.
Wang, D., Yu, Y. & Rinaldo, A. (2020) Univariate mean change point detection: penalization, CUSUM and optimality. *Electronic Journal of Statistics*, 14(1), 1917–1961.
Wied, D., Krämer, W. & Dehling, H. (2012) Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3), 570–589.
Wu, W. & Zhou, Z. (2019) *Multiscale jump testing and estimation under complex temporal dynamics*. arXiv preprint arXiv:1909.06307.
Wu, W.B. (2005) On the bahadur representation of sample quantiles for dependent sequences. *Annals of Statistics*, 33, 1934–1963.
Wu, W.B. & Zhao, Z. (2007) Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, 69(3), 391–410.
Yau, C.Y. & Zhao, Z. (2016) Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B*, 78(4), 895–916.
Zhang, T. & Lavitas, L. (2018) Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association*, 113(522), 637–648.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhao, Z., Jiang, F. & Shao, X. (2022) Segmenting time series via self-normalisation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(5), 1699–1725. Available from: https://doi.org/10.1111/rssb.12552