Electronic Journal of Statistics
Vol. 10 (2016) 2000–2038
ISSN: 1935-7524
DOI: 10.1214/16-EJS1155

Change-point detection in panel data via double CUSUM statistic

Haeran Cho

School of Mathematics, University of Bristol, UK
e-mail: haeran.cho@bristol.ac.uk

Abstract: In this paper, we consider the problem of (multiple) change-point detection in panel data. We propose the double CUSUM statistic which utilises the cross-sectional change-point structure by examining the cumulative sums of ordered CUSUMs at each point. The efficiency of the proposed change-point test is studied, which is reflected on the rate at which the cross-sectional size of a change is permitted to converge to zero while it is still detectable. Also, the consistency of the proposed change-point detection procedure based on the binary segmentation algorithm, is established in terms of both the total number and locations (in time) of the estimated change-points. Motivated by the representation properties of the Generalised Dynamic Factor Model, we propose a bootstrap procedure for test criterion selection, which accounts for both cross-sectional and within-series correlations in high-dimensional data. The empirical performance of the double CUSUM statistics, equipped with the proposed bootstrap scheme, is investigated in a comparative simulation study with the state-of-the-art. As an application, we analyse the log returns of S&P 100 component stock prices over a period of one year.

Keywords and phrases: Change-point analysis, high-dimensional data analysis, CUSUM statistics, binary segmentation.

Received November 2015.

1. Introduction

Multivariate, possibly high-dimensional observations over time have emerged in many fields, such as economics, finance, natural science, engineering and humanities, thanks to the advances of computing technologies (Fan, Lv and Qi, 2011). Multivariate data observed in practical problems often appear nonstationary in the sense that it is natural to let some quantities or parameters involved in the model to be time-varying. Arguably, the simplest departure from assuming stationarity is to operate under the assumption of piecewise stationarity, which allows more flexibility as well as providing interesting insights into the data with regards to the structural change-points. Besides, in the case of time series analysis, it enables (short-term) prediction of the future process values, by treating the last estimated segment as being stationary. Throughout the paper, the term “multiple change-point detection” is used interchangeably with “segmentation”.

Panel data models are frequently adopted to analyse high-dimensional data involving measurements over time. In this paper, we focus on the problem of detecting (possibly) multiple change-points in the mean of panel data, where n, the dimensionality of the data, may increase with the number of observations T. The panel model is presented as

$$x_{j,t} = f_{j,t} + \varepsilon_{j,t}, \qquad t = 1, \ldots, T;\ j = 1, \ldots, n,$$

where $\{f_{j,t}\}_{t=1}^{T}$, $j = 1, \ldots, n$ are piecewise constant signals which share an unknown number of change-points at unknown locations.

CUSUM statistics have been widely adopted for segmenting both univariate and multivariate data. For univariate data segmentation, CUSUM statistics are computed over time, and this series of CUSUMs is examined to locate a change-point, often as where its maximum in the absolute value is attained. Combined with a binary segmentation (BS) algorithm, the CUSUM statistics can consistently detect multiple change-points in a recursive manner (see e.g., Vostrikova (1981), Venkatraman (1992) and Cho and Fryzlewicz (2012)).

For segmenting n-dimensional panel data, we may apply the above procedure to each univariate component series separately, and then prune down the estimated change-points by identifying those detected for the identical change-point across the panel. However, such pruning may be difficult to accomplish even in moderately large dimensions, due to the estimation bias present in each change-point estimate. Besides, this approach does not take into account, and thus benefit from, the cross-sectional nature of change-points (that they are shared across the panel) which may lead to loss of power in change-point detection. Instead, we propose to segment the n-dimensional data simultaneously by searching for change-points from the *aggregation* of n series of CUSUM statistics, rather than from individual CUSUM series separately.

1.1. Literature review

Let $\mathcal{C}_b$ denote a CUSUM operator which takes $x_{j,t}$ over a generic interval $t \in [s, e]$ with $1 \le s < e \le T$ as an input and returns

$$\begin{aligned} \mathcal{X}^j_{s,b,e} &= \mathcal{C}_b(\{\sigma_j^{-1} x_{j,t}\}_{t=s}^{e}) \\ &= \sqrt{\frac{e-b}{(e-s+1)(b-s+1)}} \sum_{t=s}^{b} \frac{x_{j,t}}{\sigma_j} - \sqrt{\frac{b-s+1}{(e-s+1)(e-b)}} \sum_{t=b+1}^{e} \frac{x_{j,t}}{\sigma_j} \\ &= \frac{1}{\sigma_j} \sqrt{\frac{(b-s+1)(e-b)}{e-s+1}} \left(\frac{1}{b-s+1} \sum_{t=s}^{b} x_{j,t} - \frac{1}{e-b} \sum_{t=b+1}^{e} x_{j,t} \right) \qquad (1) \end{aligned}$$

for $b = s, \ldots, e-1$, with a suitably chosen scaling constant σ_j.

Assuming the presence of at most one change-point, some change-point tests for panel data have been proposed, based on the principle of high-dimensional CUSUM series aggregation. Note that for single change-point detection, $s = 1$ and $e = T$. Zhang et al. (2010) considered a change-point test with the test statistic $\mathcal{T}^{\text{ZSJL}}_{1,T} = \max_{b \in [1,T)} \sum_{j=1}^{n} (\mathcal{X}^j_{1,b,T})^2$ and the test criterion derived under i.i.d. Gaussian setting. A similar change-point statistic was considered by

Horváth and Hušková (2012),

$$\mathcal{T}_{1,T}^{\text{HH}} = \max_{b\in[1,T)} \frac{1}{\sqrt{n}} \frac{b(T-b)}{T^2} \sum_{j=1}^{n} \{(\mathcal{X}_{1,b,T}^j)^2 - 1\},$$

and the limit distribution of the test statistic was derived for independent panel data. Enikeeva and Harchaoui (2014) proposed a change-point test for panel data with i.i.d. Gaussian noise that combines the *linear* statistic, which is constructed similarly as $\mathcal{T}_{1,T}^{\text{HH}}$, and the *scan* statistic, which is aimed at detecting a cross-sectionally sparse change. More specifically, the test statistics are

$$\begin{aligned} \mathcal{T}_{1,T}^{\text{linear}} &= \max_{b\in[1,T)} \frac{1}{H\sqrt{2n}} \sum_{j=1}^{n} \{(\mathcal{X}_{1,b,T}^j)^2 - 1\}, \text{ and} \\ \mathcal{T}_{1,T}^{\text{scan}} &= \max_{b\in[1,T)} \max_{1\le m\le n} \frac{1}{T_m\sqrt{2m}} \sum_{j=1}^{m} \{(\mathcal{X}_{1,b,T}^{(j)})^2 - 1\}, \end{aligned}$$

where $|\mathcal{X}_{1,b,T}^{(1)}| \ge \ldots \ge |\mathcal{X}_{1,b,T}^{(n)}|$ denote the ordered CUSUM statistics at each $b \in [1,T)$. With H and T_m acting as critical values (chosen as approximate quantiles of χ^2-distributions), their proposed change-point test is $\mathcal{T}^{\text{EH}} = (\mathcal{T}_{1,T}^{\text{linear}} > 1) \vee (\mathcal{T}_{1,T}^{\text{scan}} > 1)$ (where $a \vee b = \max(a,b)$).

Allowing for both temporal and cross-sectional dependence, Jirak (2015) proposed a test statistic obtained from taking the pointwise maximum of the multiple CUSUM series:

$$\mathcal{T}_{1,T}^{\text{Jirak}} = \max_{b\in[1,T)} \max_{1\le j\le n} \sqrt{\frac{b(T-b)}{T}} |\mathcal{X}_{1,b,T}^j|, \tag{2}$$

which is compared against a threshold drawn from an extreme value distribution of Gumble type or bootstrap.

Note that with the exception of $\mathcal{T}_{1,T}^{\text{scan}}$, the CUSUM aggregation methods above are not adaptive to the underlying structure of CUSUM statistic values at each b, in the sense that they take either pointwise maximum or sum of (squared) CUSUMs. Empirical studies conducted in Cho and Fryzlewicz (2015) showed that such approaches may lead to inferior performance in detecting and locating change-points in high-dimensional settings. Instead, they proposed the Sparsified Binary Segmentation (SBS) where the change-point test $\mathcal{T}_{1,T}^{\text{SBS}}(\pi_T) > 0$ was based on the following "sparsified" or "thresholded" test statistic

$$\mathcal{T}_{1,T}^{\text{SBS}}(\pi_T) = \max_{b\in[1,T)} \sum_{j=1}^{n} |\mathcal{X}_{1,b,T}^j| \cdot \mathbb{I}(|\mathcal{X}_{1,b,T}^j| > \pi_T)$$

($\mathbb{I}(\mathcal{E}) = 1$ if and only if the event $\mathcal{E}$ is true), with an appropriately bounded threshold π_T chosen to guarantee that $|\mathcal{X}_{s,b,e}^j| < \pi_T$ uniformly over $j \in \{1,\ldots,n\}$ and $\{(s,b,e);\ 1 \le s \le b < e \le T\}$ with probability converging to one under the null hypothesis of no change-point. The intuition behind the construction of

$\mathcal{T}_{1,T}^{\text{SBS}}(\pi_T)$ is that, irrelevant contribution from those components without any change is reflected as small values of $|\mathcal{X}_{1,b,T}^j|$ and hence is disregarded through the thresholding step, while large CUSUMs formed in the vicinity of the true change-points are summed up to strengthen the meaningful contribution from the corresponding components. Combining the BS procedure with the thresholded test statistic, the consistency of the SBS algorithm for multiple change-point detection was established. However, it is empirically observed that the higher autocorrelations in $\varepsilon_{j,t}$ is, the larger π_T is required to grant the consistency in change-point detection, which amounts to selecting n thresholds for segmenting n-dimensional panel data.

1.2. Outline of the paper

In this paper, we propose the *double* CUSUM statistic which accomplishes the high-dimensional CUSUM series aggregation through data-driven partitioning of the panel data at each point, while avoiding the difficulties involved in selecting (possibly n) thresholds that apply directly to individual CUSUMs for establishing change-point detection consistency.

The rest of the paper is organised as follows. We describe the double CUSUM statistic in details in Section 2. In Section 3, we establish the consistency of the change-point test based on the double CUSUM statistics, as well as investigating its efficiency in comparison with the tests discussed in Section 1.1. Also, the double CUSUM Binary Segmentation algorithm is formulated and its consistency in multiple change-point detection is studied. Section 4 discusses the choice of important quantities including the test criterion, for which a bootstrap procedure is introduced. We illustrate its performance on simulated datasets in Section 5 and on log returns of S&P 100 component stock prices in Section 6. Section 7 concludes the paper and the proofs of theoretical results are provided in Section 8. Finally, some auxiliary results and simulation results are reported in the supplementary document (Cho, 2016).

2. Double CUSUM statistic

Recall the panel data model from Introduction

$$x_{j,t} = f_{j,t} + \varepsilon_{j,t}, \quad t = 1, \ldots, T;\ j = 1, \ldots, n. \tag{3}$$

The noise $\{\varepsilon_{j,t}\}_{t=1}^T$ satisfies $\mathbb{E}(\varepsilon_{j,t}) = 0$ for all j and t, and is allowed to be correlated both within-series and cross-sectionally as specified later in Section 3. The piecewise constant signals $\{f_{j,t}\}_{t=1}^T$, $j = 1, \ldots, n$ share N change-points $1 < \eta_1 < \ldots < \eta_N < T$ (possibly with unknown N). That is, at each change-point η_r, there exists an index set $\Pi_r = \{j : \delta_{j,r} = f_{j,\eta_r+1} - f_{j,\eta_r} \neq 0\} \subset \{1, \ldots, n\}$ with $m_r = |\Pi_r| = \sum_{j=1}^n \mathbb{I}(|\delta_{j,r}| > 0) \geq 1$ (where $|\mathcal{S}|$ denotes the cardinality of a set $\mathcal{S}$).

Recall the definition of $\mathcal{X}_{s,b,e}^j$ in (1), which denotes the CUSUM statistic computed on $x_{j,t}$ over a generic interval $t \in [s, e]$ with $1 \leq s < e \leq T$, for

$b = s, \ldots, e-1$. For change-point detection in panel data, we propose to employ the double CUSUM (DC) statistics

$$\begin{aligned} & \mathcal{D}_m^{\varphi}(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n) \\ = \quad & \left\{\frac{m(2n-m)}{2n}\right\}^{\varphi} \left(\frac{1}{m}\sum_{j=1}^m |\mathcal{X}_{s,b,e}^{(j)}| - \frac{1}{2n-m}\sum_{j=m+1}^n |\mathcal{X}_{s,b,e}^{(j)}|\right) \quad (4) \\ = \quad & \left\{\frac{m(2n-m)}{2n}\right\}^{\varphi} \cdot \frac{1}{m}\sum_{j=1}^m \left(|\mathcal{X}_{s,b,e}^{(j)}| - \frac{1}{2n-m}\sum_{j=m+1}^n |\mathcal{X}_{s,b,e}^{(j)}|\right) \quad (5) \end{aligned}$$

for $b \in [s, e)$ and $m \in \{1, \ldots, n\}$, where the DC operator $\mathcal{D}_m^{\varphi}$ takes the *ordered* CUSUM values $|\mathcal{X}_{s,b,e}^{(1)}| \geq |\mathcal{X}_{s,b,e}^{(2)}| \geq \ldots \geq |\mathcal{X}_{s,b,e}^{(n)}|$ at each b, as its input for some $\varphi \in [0, 1]$. Then, the test statistic for detecting the presence of a change-point over a given interval $[s, e]$, is derived as

$$\mathcal{T}_{s,e}^{\varphi} = \max_{b\in[s,e)} \max_{1\leq m\leq n} \mathcal{D}_m^{\varphi}(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n),$$

which is compared against a test criterion $\pi_{n,T}^{\varphi}$. Once $\mathcal{T}_{s,e}^{\varphi} > \pi_{n,T}^{\varphi}$, the location of the change-point is identified as where the pointwise maximum (over m) of the DC statistics is maximised (over b), i.e.,

$$\widehat{\eta} = \arg\max_{b\in[s,e)} \max_{1\leq m\leq n} \mathcal{D}_m^{\varphi}(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n).$$

To understand the properties of the DC statistic, first consider the case where the noise $\varepsilon_{j,t}$ is not present in the panel data. Then the series of DC statistics at each fixed m is always maximised at one of the true change-points within the interval $[s, e)$ and consequently, the maximum over both b and m is guaranteed to be attained at a true change-point. The formal statement and its proof can be found in Appendix B of the supplementary document.

The key feature of the DC statistic is the ordering of the input series to $\mathcal{D}_m^{\varphi}$. In panel data segmentation, in addition to detecting and locating the change-points in time, it is also of interest to locate the change in coordinates as well, which is relatively unexplored in the relevant literature with the exception of Jirak (2015). Not only such information is useful in the interpretation of detected change-points, but also can play an important role in aggregating the high-dimensional CUSUM series efficiently as detailed below.

At a given time point b, one way of partitioning the components into those with changes and those without, is to arrange the modulus of CUSUMs in the decreasing order, and then to label the components which correspond to the first m_b ($\in \{1, \ldots, n\}$) largest values of $|\mathcal{X}_{s,b,e}^j|$ as being likely to have a change-point around b. Note that aggregating the (squared) CUSUMs via pointwise averaging or maximising implicitly takes $m_b = n$ or $m_b = 1$, respectively. In constructing $\mathcal{T}^{\text{SBS}}(\pi_T)$, the choice of m_b is associated with the choice of π_T, i.e., $m_b(\pi_T) = \sum_{j=1}^n \mathbb{I}(|\mathcal{X}^j_{s,b,e}| > \pi_T)$. The DC statistic provides a data-driven alternative for selecting m_b, namely

$$\widehat{m}^{\varphi}_b = \arg\max_{1 \le m \le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}_{j=1}^n). \tag{6}$$

While we do not claim that the thus-obtained partitioning consistently identify Π_r at each detected change-point, we see in simulation studies reported in Section 5 that the DC statistic performs well in change-point detection, by identifying those components that *contribute* to change-point detection according to the above $\widehat{m}^{\varphi}_b$.

The RHS of (5) indicates that the term $(2n-m)^{-1}\sum_{j=m+1}^n |\mathcal{X}^{(j)}_{s,b,e}|$ acts as a threshold on $|\mathcal{X}^{(j)}_{s,b,e}|,\ j = 1, \ldots, m$ at each b, and therefore $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}_{j=1}^n)$ is essentially a scaled average of m largest CUSUMs after soft-thresholding, from which we draw resemblance between $\mathcal{T}^{\varphi}_{s,e}$ and $\mathcal{T}^{\text{SBS}}_{s,e}(\pi_T)$. However, since $\mathcal{T}^{\varphi}_{s,e}$ involves maximisation over both b and m, we avoid explicitly selecting π_T that applies directly to individual CUSUMs while still enjoying the "sparsifying" effect of the thresholding step.

Comparing (1) and (4), we can see the connection between the two operators $\mathcal{D}^{\varphi}_m$ and $\mathcal{C}_b$ especially when $\varphi = 1/2$, as both return the scaled difference between partial averages over two disjoint intervals. However, $\mathcal{D}^{\varphi}_m$ involves $\{m(2n-m)/(2n)\}^{\varphi}$ instead of $\{m(n-m)/n\}^{\varphi}$, although the input series is of length n. This difference comes from the observation that the latter scaling factor may not favour a change-point that is shared by more than $[n/2]$ rows, and even act as a penalty when m_r is close to n, which is against our intuition on the detectability of a change-point. By adopting $\{m(2n-m)/(2n)\}^{\varphi}$, the DC statistic can be regarded as being computed on the panel data of dimension $2n$, where there are additional n "null" components which are known to have no change-point. Since $\{m(2n-m)/(2n)\}^{\varphi}$ is non-decreasing in m, when considering the asymptotic efficiency of the DC statistic-based change-point test, the cross-sectional size of change accounts for both $|\delta_{j,r}|$, the magnitude of jumps at the change-point, and m_r, its "density" (as opposed to the sparsity) across the panel (see Remark 3.1 for further discussion).

For illustration, we computed the DC statistics from panel data generated with a single change-point of different configurations with φ chosen as detailed in Section 4.1. Fixing $n = 250$ and $T = 100$, $\varepsilon_{j,t}$ was simulated as in (N1) of Section 5.1.1 with $\varrho = 0.2$ (which controls the degree of cross-correlations in $\varepsilon_{j,t}$). The piecewise constant signals were generated with a change-point at $\eta_1 = T/2$, where m_1 out of n components contained a shift of magnitude randomly drawn from a uniform distribution $\mathcal{U}(0.75\delta_1, 1.25\delta_1)$ with $(m_1, \delta_1) = ([\log n], 0.24)$ and $([0.5n], 0.05)$. We chose m_1 and δ_1 in order to set $\sum_{j\in\Pi_1} \delta^2_{j,1}$ at approximately the same level.

As shown in Figure 1, the location of the true change-point in time was accurately identified as where the pointwise maximum of DC statistics was maximised in both settings, i.e., $\widehat{\eta}_1 = \arg\max_{b\in[1,T)} \max_{1\le m\le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n)$. Comparing the two heat maps, where $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,\widehat{\eta}_1,T}|\}_{j=1}^n)$ was maximised over

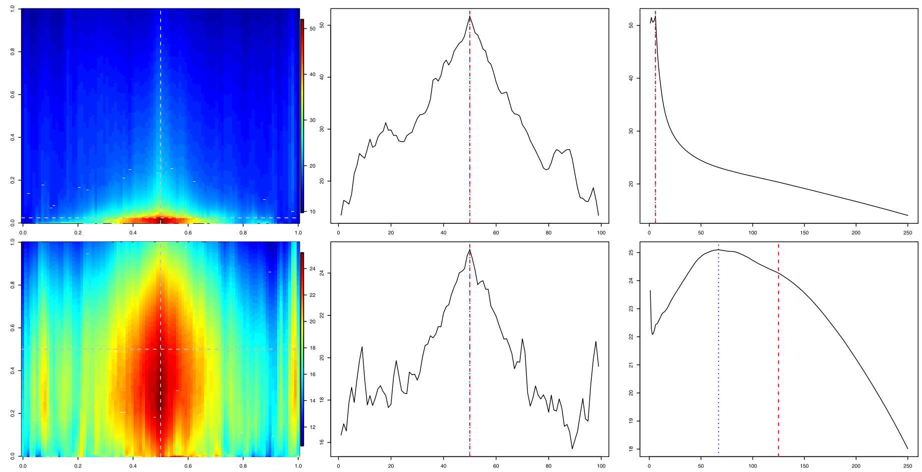

FIG 1. $(m_1, \delta_1) = ([\log n], 0.24)$ *(top) and* $(m_1, \delta_1) = ([0.5n], 0.05)$ *(bottom); the heat map of* $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n)$ *over* b *(x-axis) and* m *(y-axis) (left), the pointwise maximum of* $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n)$ *over* m *for* $b \in [1, T)$*, with broken lines indicating* η_1 *and the dotted ones* $\widehat{\eta}_1$ *(middle), and* $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,\widehat{\boldsymbol{\eta}}_1,T}|\}_{j=1}^n)$, $1 \le m \le n$ *with broken lines indicating* m_1 *and dotted ones* $\widehat{m}_1$.

m, namely $\widehat{m}_1 = \arg\max_{1 \le m \le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,\widehat{\eta}_1,T}|\}_{j=1}^n)$, was closer to m_1 for the larger δ_1. This implies that not all components with the changes contribute to the detection of a change-point in the presence of noise, due to small magnitude of the changes, and using only a subset of Π_1 may serve the purpose better for change-point detection.

Remark 2.1. *For high-dimensional change-point analysis, Aston and Kirch (2014) defined the* high-dimensional efficiency*, a concept closely related to asymptotic relative efficiency. Let* $\boldsymbol{\delta} = (\delta_{1,1}, \ldots, \delta_{n,1})^\top$*. Then, the high-dimensional efficiency is determined by the rate at which the cross-sectional size of change is allowed to converge to zero (*$\|\boldsymbol{\delta}\|_2 \to 0$*) as* T *and, with it,* n *increase, such that the power of the change-point test is strictly between the size and one. They further investigated the problem of single change-point detection using a class of change-point statistics obtained from (i) first projecting the panel data with respect to a projection vector* $\mathbf{p} \in \mathbb{R}^n$*, and (ii) computing the series of CUSUM statistics from the univariate series* $\{\langle \mathbf{x}_t, \mathbf{p} \rangle\}_{t=1}^T$*, with* $\mathbf{x}_t = (x_{1,t}, \ldots, x_{n,t})^\top$*. Assuming that* $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})^\top$ *is independent over time and denoting* $\Sigma_\varepsilon = var(\boldsymbol{\varepsilon}_t)$*, the oracle projection vector with the optimal high-dimensional efficiency was given by maximising* $(\mathbf{p}^\top \Sigma_\varepsilon \mathbf{p})^{-1} |\langle \boldsymbol{\delta}, \mathbf{p} \rangle|^2$ *with respect to* $\mathbf{p}$*, as* $\mathbf{o} := \Sigma_\varepsilon^{-1} \boldsymbol{\delta}$*.*

The DC statistic at fixed b *and* m *coincides with the CUSUM of pointwise projection of the panel data. More specifically,* $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n)$ *is associated* *with the n-vector $\mathbf{p}^{\varphi}_{b,m}$ with its elements*

$$[\mathbf{p}^{\varphi}_{b,m}]_j = \begin{cases} sign(\mathcal{X}^j_{1,b,T}) \cdot \sigma_j^{-1} \left\{\frac{m(2n-m)}{2n}\right\}^{\varphi} \frac{1}{m} & \text{if } |\mathcal{X}^j_{1,b,T}| \geq |\mathcal{X}^{(m)}_{1,b,T}|, \\ -sign(\mathcal{X}^j_{1,b,T}) \cdot \sigma_j^{-1} \left\{\frac{m(2n-m)}{2n}\right\}^{\varphi} \frac{1}{2n-m} & \text{if } |\mathcal{X}^j_{1,b,T}| < |\mathcal{X}^{(m)}_{1,b,T}|, \end{cases}$$

such that $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n) = \mathcal{C}_b(\{\langle \mathbf{x}_t, \mathbf{p}^{\varphi}_{b,m}\rangle\}_{t=1}^T)$. In light of the previous discussion on pointwise ordering and partitioning of the panel data in computing the DC statistics, our approach may be viewed as an attempt at mimicking the performance of the oracle projection without the prior knowledge of either $\boldsymbol{\delta}$ or Σ_ε. We further investigate the high-dimensional efficiency of the DC statistic in comparison with other competitors and the oracle projection in Section 3.1.

Remark 2.2. *The scan statistic (Enikeeva and Harchaoui, 2014) shares with the DC statistic the pointwise maximisation of* cumulative sums of the ordered CUSUMs *over $1 \leq m \leq n$. One key difference is that the former has $(\mathcal{X}^{(j)}_{1,b,T})^2$ in place of $|\mathcal{X}^{(j)}_{1,b,T}|$, with the interpretation of being the marginal log-likelihood ratio at given b and m under the assumption of i.i.d. Gaussian noise. On the other hand, the latter can be seen as a projected change-point statistic as noted in Remark 2.1. Another difference comes from the fact that T_m, acting as thresholds for $\mathcal{T}^{scan}_{s,e}$, are dependent on m, while the test criterion compared against $\mathcal{T}^{\varphi}_{s,e}$ does not depend on the choice of m.*

Enikeeva and Harchaoui (2014) proposed two choices for T_m, a theoretical one from a χ^2_m-distribution, and an empirical one $T_m = 2(2n)^{-1/2}\{m\log(ne/m) + \log(Tn/\alpha)\}$ at a given significance level $\alpha \in (0,1)$. In conducting the simulation studies reported in Section 5, both choices of T_m were observed to be too sensitive to cross-sectional correlations in the panel data. T_m may be chosen numerically from e.g., a bootstrap scheme as indicated by the authors, but this requires selecting n such T_m and moreover, it is unclear how the cross-correlations should be treated.

3. Consistency of the double CUSUM statistic

3.1. Single change-point detection

In this section, we show the consistency of the DC statistic in the single change-point scenario ($N = 1$). More specifically, we consider the null hypothesis

$$\mathcal{H}_0 : f_{j,1} = \cdots = f_{j,T} \quad \text{for all} \quad j = 1, \ldots, n,$$

which indicates structural stability in the mean over time. As an alternative hypothesis, we specify the scenario where the piecewise constant signals $\{f_{j,t}\}_{t=1}^T$, $j = 1, \ldots, n$ contains a single change-point at an unknown location $\eta_1 \in (1, T)$, such that

$$\mathcal{H}_A : f_{j,1} = \cdots = f_{j,\eta_1} \neq f_{j,\eta_1+1} = \cdots = f_{j,T} \text{ for some } j \in \Pi_1 \text{ with } m_1 = |\Pi_1| \geq 1.$$

Throughout the paper, $a \sim b$ is used to denote that a is of the order of b, and $a \wedge b = \min(a, b)$. Then, the consistency of the proposed test is established under the following conditions.

(A1) For each j, $\varepsilon_{j,t}$ denotes a stationary, zero-mean process with the mixing coefficients

$$\alpha_j(k) = \sup_{\substack{G \in \sigma(\varepsilon_{j,t+k}, \varepsilon_{j,t+k+1}, \ldots) \\ H \in \sigma(\varepsilon_{j,t}, \varepsilon_{j,t-1}, \ldots)}} |\mathbb{P}(H \cap G) - \mathbb{P}(H)\mathbb{P}(G)|.$$

Then, there exist fixed $C_\varepsilon, C_\alpha > 0$, $\mu \in (0, 1)$ and $\sigma^* \geq \sigma_* > 0$ with $\bar{\sigma} = \sigma^*/\sigma_* \in [1, \infty)$, such that the followings hold.

(A1.i) $\mathbb{E}(\varepsilon_{j,t}^k) \leq C_\varepsilon^{k-2} k! \mathbb{E}(\varepsilon_{j,t}^2)$ uniformly in j and $k = 3, 4, \ldots$.

(A1.ii) $\max_{1 \leq j \leq n} |\alpha_j(k)| \leq C_\alpha \mu^k$ for all $k = 1, 2, \ldots$.

(A1.iii) σ_j^2, the long-run variance of $\varepsilon_{j,t}$, satisfies $\sigma_*^2 \leq \sigma_j^2 \leq \sigma^{*2}$.

(A2) The dimensionality n satisfies $n \sim T^\omega$ for some fixed $\omega \in [0, \infty)$.

(A3) There exists a fixed constant $\bar{f} > 0$ such that $\max_{1 \leq j \leq n} \max_{1 \leq t \leq T} |f_{j,t}| \leq \bar{f}$.

(A4) There exists a fixed constant $c > 0$ such that $\eta_1 \wedge (T - \eta_1) > cT^\beta$ for some $\beta \in (0, 1]$.

(A5) Let $\widetilde{\delta}_1 = m_1^{-1} \sum_{j \in \Pi_1} |\delta_{j,1}|$, i.e., the average magnitude of non-zero changes at $t = \eta_1$. Then for any $\varphi \in [0, 1]$, we have $(n^\varphi \log T)^{-1} m_1^\varphi \widetilde{\delta}_1 T^{\beta/2} \to \infty$ as $T \to \infty$.

In (A1), each $\varepsilon_{j,t}$ is assumed to be α-mixing (strong-mixing) at a geometric rate, with bounded moments and long-run variance. The condition (A1.i) is met by many distributions such as exponential, gamma and inverse Gaussian, besides the Gaussian distribution. Note that the cross-sectional correlations of the panel data are not explicitly controlled by any of the conditions imposed in (A1). In (A2), the dimensionality can either be fixed or increase with T at a polynomial rate. (A4) imposes a condition on the unbalancedness of the change-point location, permitting $T^{-1}\{\eta_1 \vee (T - \eta_1)\} \to 0$ as $T \to \infty$ when $\beta < 1$. (A5) places a lower bound on the rate of $m_1^\varphi \widetilde{\delta}_1$, which dictates the minimum requirement on the cross-sectional size of the change for the change-point to be detected as well as being located with accuracy. (A3) rules out the trivial case where any $|\delta_{j,1}| \to \infty$ with $T \to \infty$.

Remark 3.1 (High-dimensional efficiency)**.** *Recall the high-dimensional efficiency discussed in Remark 2.1, which was introduced in Aston and Kirch (2014) as a tool that allows us to quantify and compare the power of different change-point tests. Table 1 summarises the high-dimensional efficiency of* $\mathcal{T}_{1,T}^\varphi$ *and other change-point tests discussed in Section 1.1, when the change-point is maximally distanced from the extreme ends of* $[1, T]$ *(i.e.,* $\beta = 1$*). The double vertical line divides the tests into (i) those proposed under the assumption of cross-sectional independence (left) and dependence (right), (ii) those detecting the presence of a change-point only (left) and those identifying its location as well (right), and*

TABLE 1
High-dimensional efficiency of change-point tests when $\beta = 1$.

$\mathcal{T}_{1,T}^{\mathrm{HH}}$	$\frac{(\sum_{j=1}^n \delta_{j,1}^2)^{1/2} T^{1/2}}{n^{1/4}} \to \infty$	$\mathcal{T}_{1,T}^{\mathrm{Jirak}}$	$\frac{\min_{j\in\Pi_1} \lvert\delta_{j,1}\rvert T^{1/2}}{\sqrt{\log T}} \to \infty$
$\mathcal{T}_{1,T}^{\mathrm{linear}}$	$\frac{(\sum_{j=1}^n \delta_{j,1}^2)^{1/2} T^{1/2}}{n^{1/4}} \to \infty$	$\mathcal{T}_{1,T}^{\mathrm{SBS}}(\pi_T)$	$\frac{\min_{j\in\Pi_1} \lvert\delta_{j,1}\rvert T^{1/2}}{\log T} \to \infty$
$\mathcal{T}_{1,T}^{\mathrm{scan}}$	$\frac{(\sum_{j=1}^n \delta_{j,1}^2)^{1/2} T^{1/2}}{\sqrt{m_1 \log(n/m_1)}} > \sqrt{6.6}$	$\mathcal{T}_{1,T}^{\varphi}$	$\frac{m_1^{\varphi} \widetilde{\delta}_1 T^{1/2}}{n^{\varphi} \log T} \to \infty$

(iii) those with the interpretation as projected change-point tests (right) and those without (left).

The oracle projection-based change-point test (see Remark 2.1) achieves high-dimensional efficiency of $T^{1/2}\|\Sigma_{\boldsymbol{\varepsilon}}^{-1/2}\boldsymbol{\delta}\|_2 \to \infty$ and thus attaining better efficiency than $\mathcal{T}_{1,T}^{HH}$ and $\mathcal{T}_{1,T}^{linear}$ by $n^{1/4}$, and than $\mathcal{T}_{1,T}^{scan}$ by $m_1^{1/2}$, for diagonal $\Sigma_{\boldsymbol{\varepsilon}}$ (independent panel). When the change-point is sparse ($m_1 \sim 1$), the high-dimensional efficiency of $\mathcal{T}_{1,T}^{0}$ is comparable to that of the oracle up to a logarithmic factor. When the change-point is dense ($m_1 \sim n$), the high-dimensional efficiency of $\mathcal{T}_{1,T}^{\varphi}$, $\varphi > 0$ is comparable to that of $\mathcal{T}_{1,T}^{0}$. Comparing $\mathcal{T}_{1,T}^{Jirak}$, $\mathcal{T}_{1,T}^{SBS}(\pi_T)$ and $\mathcal{T}_{1,T}^{0}$, the latter attains better high-dimensional efficiency when $m_1^{-1}\sum_{j\in\Pi_1} |\delta_{j,1}| \gg \min_{j\in\Pi_1} |\delta_{j,1}|$. However, the former two achieves partitioning consistency (i.e., consistent estimation Π_1), which is not granted by the latter.

Remark 3.2. *As briefly noted below (1), some studentisation is required for panel data analysis in practice. We estimate σ_j using the flat-top kernel estimator with the automatically chosen bandwidth as discussed in Politis (2011). Let $\widehat{\eta}_{j,1} = \arg\max_{b\in[1,T)} |\mathcal{C}_b(\{x_{j,t}\}_{t=1}^T)|$,*

$$\bar{\varepsilon}_{j,t} = x_{j,t} - \frac{1}{\widehat{\eta}_{j,1}} \sum_{t=1}^{\widehat{\eta}_{j,1}} x_{j,t} \cdot \mathbb{I}(t \le \widehat{\eta}_{j,1}) - \frac{1}{T - \widehat{\eta}_{j,1}} \sum_{t=\widehat{\eta}_{j,1}+1}^{T} x_{j,t} \cdot \mathbb{I}(t > \widehat{\eta}_{j,1}), \quad (7)$$

$$w(t) = \begin{cases} 1 & \textit{for } |t| \le 1/2, \\ 2(1-|t|) & \textit{for } 1/2 < |t| < 1, \\ 0 & \textit{for } |t| \ge 1. \end{cases}$$

and $c_j(k) = \frac{1}{T}\sum_{t=1}^{T-k} \bar{\varepsilon}_{j,t}\bar{\varepsilon}_{j,t+k}$. Note that $\bar{\varepsilon}_{j,t}$ estimates the unobservable $\varepsilon_{j,t}$ by identifying a change-point candidate $\widehat{\eta}_{j,1}$ associated with each $\{x_{j,t}\}_{t=1}^T$. Let τ_j be the smallest positive integer such that $|c_j(\tau_j + k)/c_j(0)| < 1.4\sqrt{T^{-1}\log_{10} T}$ for $k = 1, 2, 3$ (where the constants are chosen as per Hušková and Kirch (2010)). Then, the estimator is given by

$$\widehat{\sigma}_j^2 = \max\left\{ c_j(0) + 2\sum_{k=1}^{2\tau_j} w\left(\frac{k}{2\tau_j}\right) c_j(k),\ \frac{c_j(0)}{2} \right\}, \quad (8)$$

where the second term in the RHS of (8) is present to prevent spuriously large CUSUMs resulting from too small estimates of σ_j (flat-top kernel estimator is *known to produce negative estimators). Hušková and Kirch (2010) showed the consistency of the above estimator in the* single *change-point detection problem, under similar conditions as those imposed on $\varepsilon_{j,t}$ in (A1). Noting that (i) the theoretical results derived in this paper hold without the consistency of the flat-top kernel estimator, provided that $\widehat{\sigma}_j^2$ is uniformly bounded away from zero, and also that (ii) we ultimately consider the problem of* multiple *change-point detection (with adjustment to the estimator of σ_j^2 in order to accommodate the presence of multiple change-points, see Remark 3.3), we assume (A6) below on $\widehat{\sigma}_j^2$ without extending the consistency result of Hušková and Kirch (2010).*

(A6) Define $\mathcal{E}_\sigma = \{\max_{1\le j\le n} |\widehat{\sigma}_j^2 - \sigma_j^2| \le \sigma_*^2/2\}$ where σ_* is the same as that in (A1). Then we assume $\mathbb{P}(\mathcal{E}_\sigma) \to 1$ with $T \to \infty$.

Note that under (A4), short intervals near the extreme points $\{1, T\}$ do not contain the change-point. Hence we search for a change-point within $[1, T] \setminus \mathcal{I}_{1,T}$ only, where $\mathcal{I}_{1,T} = [1, d_T] \cup [T - d_T, T]$ with $d_T := [C \log^2 T]$ for some $C > 0$. That is, $\mathcal{T}_{1,T}^\varphi = \max_{b\in[1,T]\setminus\mathcal{I}_{1,T}} \max_{1\le m\le n} \mathcal{D}_m^\varphi(\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n)$ and $\widehat{\eta}_1 = \arg\max_{b\in[1,T]\setminus\mathcal{I}_{1,T}} \max_{1\le m\le n} \mathcal{D}_m^\varphi(\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n)$.

Under these conditions, we present the following theorems on the consistency of the DC statistic-based test equipped with a test criterion $\pi_{n,T}^\varphi$, which satisfies $C' n^\varphi \log T < \pi_{n,T}^\varphi < C'' m_1^\varphi \widetilde{\delta}_1 T^{\beta/2}$ for some $C', C'' > 0$.

Theorem 3.1. *Assume that (A1)–(A3) and (A6) hold and that there exists no change-point ($N = 0$) in the panel data in (3). Then $\mathbb{P}\left\{\mathcal{T}_{1,T}^\varphi > \pi_{n,T}^\varphi\right\} \to 0$ as $T \to \infty$.*

Theorem 3.1 guarantees that when all signals remain constant, the test does not detect any change-point with probability converging to one.

Theorem 3.2. *Assume that (A1)–(A6) hold. Then there exists $c_0 > 0$ such that*

$$\mathbb{P}\left\{\mathcal{T}_{1,T}^\varphi > \pi_{n,T}^\varphi \text{ and } |\widehat{\eta}_1 - \eta_1| < c_0 \epsilon_T\right\} \to 1$$

as $T \to \infty$, where $\epsilon_T = n^{2\varphi}(m_1^\varphi \widetilde{\delta}_1)^{-2} \log^2 T$.

Theorem 3.2 states that in the presence of a single change-point, the proposed test detects its presence as well as accurately identifying its location. From (A5), it is easily seen that $\epsilon_T / T^\beta \to 0$ as $T \to \infty$, i.e., in the rescaled time, the estimated change-point location is consistent since $T^{-1}|\widehat{\eta}_1 - \eta_1| \le T^{-\beta}|\widehat{\eta}_1 - \eta_1| \to 0$. We may define the optimality in change-point detection as when the true change-point and its estimate are within the distance of $O_p(1)$, see Korostelev (1987). Then, near-optimality in change-point estimation is achieved up to a logarithmic factor ($\log^2 T$) with the choice (i) $\varphi = 0$ when $\widetilde{\delta}_1 \sim 1$, or (ii) $\varphi > 0$ when $m_1 \sim n$ and $\widetilde{\delta}_1 \sim 1$.

3.2. Binary segmentation for multiple change-point detection

In this section, we show the consistency of the DC statistic for multiple change-point detection when applied jointly with a BS algorithm. We first formulate the double CUSUM Binary Segmentation (DCBS) algorithm for panel data segmentation, which is equipped with a test criterion $\pi^{\varphi}_{n,T}$. As in Section 3.1, let $\mathcal{I}_{s,e} = [s, s + d_T] \cup [e - d_T, e]$ denote a fraction of the interval $[s, e]$ on which we do not search for change-points, in order to account for possible bias in the previously detected change-points. This does not affect the asymptotic consistency of the estimated change-point locations under the assumption (B1) below, on the dispersion of change-points. The index p is used to denote the level (indicating the progression of the segmentation procedure) and q is used to denote the location of the node at each level.

The double CUSUM Binary Segmentation algorithm

Step 0 Set $(p, q) = (1, 1)$, $s_{p,q} = 1$ and $e_{p,q} = T$.

Step 1 At the current level p, repeat the following for all q.

Step 1.1 Letting $s = s_{p,q}$ and $e = e_{p,q}$, obtain the series of CUSUMs $\{\mathcal{X}^j_{s,b,e}\}$ for $b \in [s, e)$ and $j = 1, \ldots, n$, on which $\mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}^n_{j=1})$ is computed over all b and m.

Step 1.2 Obtain the test statistic

$$\mathcal{T}^{\varphi}_{s,e} = \max_{b\in[s,e]\setminus\mathcal{I}_{s,e}} \max_{1\le m\le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}^n_{j=1}).$$

Step 1.3 If $\mathcal{T}^{\varphi}_{s,e} \le \pi^{\varphi}_{n,T}$, quit searching for change-points on the interval $[s, e]$. On the other hand, if $\mathcal{T}^{\varphi}_{s,e} > \pi^{\varphi}_{n,T}$, locate

$$\widehat{\eta} = \arg \max_{b\in[s,e]\setminus\mathcal{I}_{s,e}} \max_{1\le m\le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}^n_{j=1})$$

and proceed to Step 1.4.

Step 1.4 Add $\widehat{\eta}$ to the set of estimated change-points and divide the interval $[s_{p,q}, e_{p,q}]$ into two sub-intervals $[s_{p+1,2q-1}, e_{p+1,2q-1}]$ and $[s_{p+1,2q}, e_{p+1,2q}]$, where $s_{p+1,2q-1} = s_{p,q}$, $e_{p+1,2q-1} = \widehat{\eta}$, $s_{p+1,2q} = \widehat{\eta} + 1$ and $e_{p+1,2q} = e_{p,q}$.

Step 2 Once $[s_{p,q}, e_{p,q}]$ for all q are examined at level p, set $p \leftarrow p + 1$ and go to Step 1.

Step 1.3 furnishes a stopping rule to the DCBS algorithm: the search for further change-points is terminated once $\mathcal{T}^{\varphi}_{s,e} \le \pi^{\varphi}_{n,T}$ on every $[s, e]$ defined by two adjacent estimated change-points.

Remark 3.3. *An adjustment is required to extend the scaling estimation procedure in Remark 3.2, originally designed for single change-point detection, to be applicable to the multiple change-point scenario. More specifically, $\bar{\varepsilon}_{j,t}$ in (7) can no longer be regarded as well-estimating $\varepsilon_{j,t}$ when $[1, \widehat{\eta}_{j,1}]$ or $(\widehat{\eta}_{j,1}, T]$ may* *contain further change-points. Seeing that $\widehat{\eta}_{j,1}$ is the top node of a binary tree that results from applying a BS algorithm to $\{x_{j,t}\}_{t=1}^T$, we remedy this by growing a binary tree of* depth L_T *($= O(\log T)$) from each $\{x_{j,t}\}_{t=1}^T$, $j = 1, \ldots, n$. Such a tree represents a maximal segmentation of $x_{j,t}$ and therefore, regarding each segment as being stationary, $\bar{\varepsilon}_{j,t}$ is derived by subtracting the sample means within those intervals. Then, the scaling estimation procedure is applied to the thus-obtained $\bar{\varepsilon}_{j,t}$.*

Note that this approach requires an arbitrary choice of L_T when there is no prior information on the upper bound on the total number of change-points, N. For multiple change-point detection, such a requirement (or its equivalent) is commonly found; see e.g., Soh and Chandrasekaran (2015) and Kirch and Muhsal (2015) where the minimum distance between two consecutive change-points plays an essential role in guaranteeing the consistency of the proposed procedures.

The consistency of the DCBS algorithm is established under (A1)–(A3), (A6) and the following conditions extending (A4)–(A5) in order to allow for the presence of multiple change-points.

(B1) There exists a fixed constant $c > 0$ such that $\min_{0\le r\le N}(\eta_{r+1} - \eta_r) \ge cT^\beta$ for some $\beta \in (6/7, 1]$, using the notational convention that $\eta_0 = 0$ and $\eta_{N+1} = T$.

(B2) At each η_r, let $\widetilde{\delta}_r = m_r^{-1}\sum_{j\in\Pi_r}|\delta_{j,r}|$. Then for any $\varphi \in [0, 1]$, we have $\underline{\Delta}_\varphi = \min_{1\le r\le N} m_r^\varphi \widetilde{\delta}_r$ satisfy $(n^\varphi \log T)^{-1}\underline{\Delta}_\varphi T^{7\beta/4-3/2} \to \infty$ as $T \to \infty$.

Since the unbalancedness of a change-point location is closely related to the distance between two adjacent change-points, (B1) is formulated in terms of the latter. Note that we do not impose any upper bound on the number of total change-points N except for the implication that can be derived from (B1), namely that N may grow with T provided that any two adjacent change-points is sufficiently distanced. Comparing (B2) to (A5), the high-dimensional efficiency is worsen by $T^{3/2-5\beta/4}$ when detecting multiple change-points, which is also the case when applying the BS algorithm to univariate data. This rate can be improved when the DC statistic is combined with e.g., the Wild Binary Segmentation (WBS) (Fryzlewicz, 2014) which applies the CUSUM principle to randomly drawn intervals; we leave the exploration in this direction for the future research.

Theorem 3.3. *Let $\widehat{\eta}_r$, $r = 1, \ldots, \widehat{N}$ ($1 < \widehat{\eta}_1 < \ldots < \widehat{\eta}_{\widehat{N}} < T$) denote the change-points detected by the DCBS algorithm with a test criterion $\pi_{n,T}^\varphi$ satisfying $C'n^{2\varphi}\underline{\Delta}_\varphi^{-1}T^{5(1-\beta)/2}\log T < \pi_{n,T}^\varphi < C''\underline{\Delta}_\varphi T^{\beta-1/2}$. Assuming (A1)–(A3), (A6) and (B1)–(B2), there exists $c_0 > 0$ such that*

$$\mathbb{P}\left\{\widehat{N} = N;\ |\widehat{\eta}_r - \eta_r| < c_0\epsilon_T \text{ for } r = 1, \ldots, N\right\} \to 1$$

as $T \to \infty$, where $\epsilon_T = n^{2\varphi}\underline{\Delta}_\varphi^{-2}T^{5(1-\beta)}\log^2 T$.

Unlike in the single change-point scenario, ϵ_T now depends on the dispersion of the change-points through β. From (B1)–(B2), it is easily seen that $\epsilon_T/T^\beta \to 0$ as $T \to \infty$, and hence the multiple change-points are consistently located for all $r = 1, \ldots, N$.

Remark 3.4 (Post-processing of estimated change-points)**.** *In this paper, we focus on establishing the good performance of the DC statistic in high-dimensional CUSUM series aggregation when combined with a simple segmentation method, rather than considering more sophisticated algorithms such as the WBS (Fryzlewicz, 2014) or MOSUM (Kirch and Muhsal, 2015) procedures, where the DC principle is straightforwardly applicable. The BS algorithm is known to perform sub-optimally in certain unfavourable settings since, at each iteration, it fits a step function with a single break to the data over a segment that contains possibly multiple change-points. Hence, we equip the DCBS algorithm with an extra step aiming at removing spuriously detected change-points which is in line with the post-processing steps proposed in Cho and Fryzlewicz (2012) and Cho and Fryzlewicz (2015). It checks whether* $\mathcal{T}^{\varphi}_{\widehat{\eta}_r-d_r,\widehat{\eta}_r+d_r} > \pi^{\varphi}_{n,T}$ *for* $r = 1, \ldots, \widehat{N}$ *with* $d_r = \min(\widehat{\eta}_r - \widehat{\eta}_{r-1}, \widehat{\eta}_{r+1} - \widehat{\eta}_r)/2$*. In other words, we compute the DC statistics within each segment containing a single estimated change-point, and retain only those* $\widehat{\eta}_r$ *that survive the thresholding. More details on the post-processing step can be found in Section 3.2.1 of Cho and Fryzlewicz (2012).*

4. Choice of φ and test criterion

4.1. Choice of φ

Remark 3.1 indicates that $\varphi = 0$ is preferable in terms of the high-dimensional efficiency for detecting a change-point that is known to be sparse across the panel, while $\mathcal{T}^{\varphi}_{1,T}$ with $\varphi > 0$ achieves the same high-dimensional efficiency as $\mathcal{T}^{0}_{1,T}$ when the change is dense across the panel. In practice, such information is often unavailable a priori and therefore it is of interest to find a way of combining the information embedded in an *array* of DC statistics indexed by $\varphi \in [0, 1]$, which works universally well over different ranges of change-point configurations determined by η_r, m_r and $\widetilde{\delta}_r$.

Recalling the data-driven partitioning of the panel achieved by the DC statistic, the pointwise partitioning of the ordered CUSUM values may be regarded as analogous to fitting a step function with a single break to $|\mathcal{X}^{(j)}_{s,b,e}|$, $j = 1, \ldots, n$ at each b. Suppose that $\{|\mathcal{X}^{(j)}_{s,b,e}|\}_{j=1}^n$ arises from an additive model with a piecewise constant signal that contains a single break at $j = m_r$ for some $r \in \{1, \ldots, N\}$ and otherwise constant. This setting is unlikely to be satisfied by the ordered CUSUM values in general, but provides a framework for linking the optimality in panel partitioning to that in locating a single change-point. Then, Brodsky and Darkhovsky (1993) showed that the choice of $\varphi = 1/2$ leads to the best rate of convergence in the bias $|\widehat{m}^{\varphi}_b - m_r|$ (recall (6)) among $\varphi \in [0, 1]$, i.e., asymptotically, the optimal partitioning is attained by the choice of $\varphi = 1/2$.

Taking into account these observations, we propose a new DC statistic

$$\widetilde{\mathcal{D}}_m(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n) = \gamma_n \mathcal{D}_m^0(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n) + \mathcal{D}_m^{1/2}(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n),$$

where γ_n acts as a scaling factor that enables treating $\mathcal{D}_m^0$ and $\mathcal{D}_m^{1/2}$ on an equal footing. Heuristically, the discrepancy between $\mathcal{D}_m^0$ and $\mathcal{D}_m^{1/2}$ increases at the rate $m^{1/2}$, but is not so pronounced for small values of m. Therefore, $\widetilde{\mathcal{D}}_m$ can be viewed as an attempt at combining the different ranges of change-point configurations over which either $\mathcal{D}_m^0$ or $\mathcal{D}_m^{1/2}$ achieves consistency, and the following conditions modifying (A5) and (B2) reflect this point.

(A5') $(n^{1/2}\log T)^{-1}(\gamma_n \vee m_1^{1/2})\widetilde{\delta}_1 T^{\beta/2} \to \infty$ as $T \to \infty$.

(B2') We have $\underline{\Delta} = \min_{1\le r\le N}(\gamma_n \vee m_r^{1/2})\widetilde{\delta}_r$ satisfy $(n^{1/2}\log T)^{-1}\underline{\Delta}T^{7\beta/4-3/2} \to \infty$ as $T \to \infty$.

Then, the consistency of DC statistics carry over to that of the newly defined DC statistic $\widetilde{\mathcal{D}}_m$ as below.

Theorem 4.1. *Assume $\omega > 0$ such that $n \to \infty$ as $T \to \infty$, and $n^{-1/2}\gamma_n \to 0$ as $n \to \infty$. Let $\widetilde{\mathcal{D}}_m$ replace $\mathcal{D}_m^{\varphi}$ in deriving the test statistics where applicable.*

- *(a) Assume that (A1)–(A4), (A5') and (A6) hold. Equipped with the test criterion $\widetilde{\pi}_{n,T}$ satisfying $C' n^{1/2}\log T < \widetilde{\pi}_{n,T} < C''(\gamma_n \vee m_1^{1/2})\widetilde{\delta}_1 T^{\beta/2}$ for some $C', C'' > 0$, Theorems 3.1–3.2 hold with $\epsilon_T = n\{(\gamma_n \vee m_1^{1/2})\widetilde{\delta}_1\}^{-2}\log^2 T$ in the latter.*
- *(b) Assume that (A1)–(A3), (A6), (B1) and (B2') hold. Equipped with the test criterion satisfying $C' n\underline{\Delta}^{-1}T^{5(1-\beta)/2}\log T < \widetilde{\pi}_{n,T} < C''\underline{\Delta}T^{\beta-1/2}$ for some $C', C'' > 0$, Theorem 3.3 holds with $\epsilon_T = n\underline{\Delta}^{-2}T^{5(1-\beta)}\log^2 T$.*

For the purpose of testing only, it is reasonable to combine the two tests based on $\mathcal{D}_m^0$ and $\mathcal{D}_m^{1/2}$ as in Enikeeva and Harchaoui (2014), namely, $(\mathcal{T}_{1,T}^0 > \pi_{n,T}^0) \vee (\mathcal{T}_{1,T}^{1/2} > \pi_{n,T}^{1/2})$. Adopting such an approach, however, it is not clear how to identify the location of a change-point once its presence is detected, since we cannot exclude the possibility that it is detected by both $\mathcal{T}_{1,T}^0$ and $\mathcal{T}_{1,T}^{1/2}$, each of which estimates its location with bias. This becomes further complicated in the presence of multiple change-points of different change-point configurations. On the other hand, while γ_n needs to be selected additionally, we can readily identify the change-point locations using $\widetilde{\mathcal{D}}_m$.

Theorem 4.1 assumes that $n^{-1/2}\gamma_n \to 0$ only. However, as the formulation of (A5') and (B2') suggests, guaranteed improvement of the high-dimensional efficiency of $\widetilde{\mathcal{D}}_m$ over that of both $\mathcal{D}_m^0$ and $\mathcal{D}_m^{1/2}$, still requires the knowledge of unknown sparsity or density of the change-point for the choice of γ_n. In Section 5, $\gamma_n = \log n$ is considered in investigating the finite sample performance of $\widetilde{\mathcal{D}}_m$ along with that of $\mathcal{D}_m^0$ and $\mathcal{D}_m^{1/2}$.

4.2. Bootstrap for test criterion selection

Theorems 3.1–4.1 provide ranges for the rate of $\pi^{\varphi}_{n,T}$ which grant size and power consistency for both single and multiple change-point detection problems. However, the theoretical rates involve typically unattainable knowledge on the quantities such as β and $\underline{\Delta}_{\varphi}$ and, even with such knowledge available, finite sample performance may be heavily influenced by the choice of the multiplicative constant applied to the given rate. Therefore, we propose a resampling procedure that enables us to approximate the quantiles of $\mathcal{T}^{\varphi}_{s,e}$ under the null hypothesis of no change-points.

Originally proposed as a data-based simulation method for statistical inference with i.i.d. random samples, the bootstrap principle has been extended to a variety of non-i.i.d. situations; see Härdle, Horowitz and Kreiss (2003) for a comprehensive survey. Although some heuristic attempts have been made (Fiecas and von Sachs, 2014), applying bootstrap methods developed for time series of fixed dimensions to high-dimensional settings is challenging. Some recent efforts in this direction include Jentsch and Politis (2015), who proposed the Linear Process Bootstrap for multivariate time series and established its asymptotic validity for the sample mean when n is allowed to increase with T. The procedure involves the estimation of $(nT) \times (nT)$-dimensional covariance matrix of the vectorised version of $[\boldsymbol{\varepsilon}_t,\, t = 1, \ldots, T]$ and the computation of its Cholesky decomposition, where the latter task alone is of computational complexity $O(n^3T^3)$, which makes it difficult to apply the method even to panel data of moderately large dimensions.

Instead, we propose a bootstrap procedure motivated by the representation theory developed for the Generalised Dynamic Factor Model (GDFM). Factor analysis is a popular dimension reduction technique used in many disciplines, such as econometrics, statistics, signal processing, psychometrics, and chemometrics. The key idea is that pervasive cross-correlations in $\boldsymbol{\varepsilon}_t$ are modelled by the *common* component $\boldsymbol{\chi}_t$, and $\boldsymbol{\xi}_t = \mathbf{x}_t - \boldsymbol{\chi}_t$ with moderate degree of cross-correlations denotes the *idiosyncratic* component. The GDFM introduced in Forni et al. (2000) steps further and admits the following representation theorem (Forni and Lippi, 2001): any $\boldsymbol{\varepsilon}_t$ with a finite number ($q < \infty$) of diverging dynamic eigenvalues is decomposed into $\boldsymbol{\chi}_t$ driven by a q-tuple of white noises $\mathbf{u}_t$ (*common shocks*) as $\boldsymbol{\chi}_t = \mathbf{b}(L)\mathbf{u}_t$ (L denotes the lag operator and $\mathbf{b}(L)$ is an $n \times q$-matrix of one-sided and square-summable filters $b_{ik}(L)$), and $\boldsymbol{\xi}_t$ with bounded dynamic eigenvalues. Referred to as the GDFM-boot, the proposed bootstrap method utilises this representation property of the GDFM in order to effectively handle the cross-correlations as well as within-series correlations present in $\boldsymbol{\varepsilon}_t$ of large dimensions.

The GDFM-boot algorithm

Step 1 Obtain $\mathbf{E} = [\widehat{\varepsilon}_t,\, t = 1, \ldots, T]$ where $\widehat{\varepsilon}_{j,t} = \widehat{\sigma}_j^{-1}\bar{\varepsilon}_{j,t}$ (refer to (7) and Remark 3.3 for the definition of $\bar{\varepsilon}_{j,t}$).

Step 2 Let $\mathbf{e}_l$ denote the eigenvector of the covariance matrix $T^{-1}\mathbf{E}\mathbf{E}^\top$, which corresponds to its l-th largest eigenvalue. Then, estimate q, the number of common shocks driving the cross-correlations of $\mathbf{E}$ using the information criterion proposed by Bai and Ng (2002): $IC(k) = \log(T^{-1}\sum_{t=1}^T \|\widehat{\boldsymbol{\varepsilon}}_t - \sum_{l=1}^k \mathbf{e}_l\mathbf{e}_l^\top \widehat{\boldsymbol{\varepsilon}}_t\|_2^2) + kC_{n,T}^{-1}\log(C_{n,T})$, as $q = \arg\min_{0\le k\le Q} IC(k)$ with $Q = \lfloor C_{n,T}/\log(C_{n,T})\rfloor$ and $C_{n,T} = n \wedge T$.

Step 3 Estimate the q-dimensional common shocks ($\widehat{\mathbf{u}}_t$) and the associated filter ($\widehat{\mathbf{b}}(L)$), such that $\widehat{\boldsymbol{\varepsilon}}_t$ is decomposed into $\widehat{\boldsymbol{\chi}}_t = \widehat{\mathbf{b}}(L)\widehat{\mathbf{u}}_t$ and $\widehat{\boldsymbol{\xi}}_t = \widehat{\boldsymbol{\varepsilon}}_t - \widehat{\boldsymbol{\chi}}_t$.

Step 4 For $l = 1, \ldots, B$, repeat the following steps.

Step 4.1 For each $k = 1, \ldots, q$, draw $\{u_{k,t}^l\}_{t=1}^T$ independently from the empirical distribution of $\{\widehat{u}_{k,t}\}_{t=1}^T$, from which $\boldsymbol{\chi}_t^l = \widehat{\mathbf{b}}(L)\mathbf{u}_t^l$ is obtained.

Step 4.2 Generate a bootstrap sequence of the Fourier coefficients of $\widehat{\boldsymbol{\xi}}_t$ using the Time Frequency Toggle (TFT)-Bootstrap (Kirch and Politis, 2011), to which the inverse fast Fourier transform algorithm is applied to produce $\boldsymbol{\xi}_t^l$.

Step 4.3 On the bootstrap series $\boldsymbol{\varepsilon}_t^l = \boldsymbol{\chi}_t^l + \boldsymbol{\xi}_t^l$, compute $\mathcal{E}_{1,b,T}^{j,l} = \mathcal{C}_b(\{\varepsilon_{j,t}^l\}_{t=1}^T)$, from which $\mathcal{T}_{1,T}^{\varphi,l} = \max_{b\in[1,T),\,1\le m\le n} \mathcal{D}_m^\varphi(\{|\mathcal{E}_{1,b,T}^{(j),l}|\}_{j=1}^n)$ is obtained where $|\mathcal{E}_{1,b,T}^{(1),l}| \ge \ldots \ge |\mathcal{E}_{1,b,T}^{(n),l}|$.

Step 5 Select the $(1-\alpha)$-quantile of $\mathcal{T}_{1,T}^{\varphi,l}$, $l = 1, \ldots, B$ as $\pi_{n,T}^\varphi$.

In Step 1, we adopt the same approach taken in estimating σ_j, namely first to estimate the locations of the change-points coordinate-wise and then to take away the estimated piecewise constant signal from each $x_{j,t}$, in order to estimate the unobservable $\varepsilon_{j,t}$. Note that the practice of input data studentisation is often adopted prior to factor modelling. Step 2 is justified by the observations that (i) the number of factors in the static representation of factor models is typically greater than the number of common shocks in the GDFM, and (ii) over-estimated q still returns mean-square consistent common components (Forni et al., 2000, Corollary 2). Section 4 of Forni et al. (2000) provides an estimator of $\mathbf{b}(L)$ (and hence $\mathbf{u}_t$) based on the windowed estimator of the spectral density matrix of $\mathbf{E}$, which is adopted for Step 3.

Step 4.1 produces a bootstrap sample of $\widehat{\boldsymbol{\chi}}_t$ by treating the white noise estimates $\widehat{u}_{k,t}$ as being i.i.d. over time. In Step 4.2, the Local Bootstrap, originally proposed in Kirch and Politis (2011) as a part of TFT-Bootstrap for resampling univariate time series, is applied to the n-dimensional $\widehat{\boldsymbol{\xi}}_t$ of bounded cross-sectional correlations. It does not require an initial estimate of the spectral density matrix of $\widehat{\boldsymbol{\xi}}_t$. Instead, the Fourier coefficients of $\widehat{\boldsymbol{\xi}}_t$ are resampled according to a kernel-based probability vector, under the observation that in a neighbourhood of each frequency, the distribution of different coefficients is almost identical (if the spectral density is smooth). A detailed description of the Local Bootstrap is provided in the Appendix A of the supplement document. In simulation studies, we used the Daniell kernel with the window size $[0.05T]$. Since the common and idiosyncratic components are handled independently, generating a smaller size bootstrap sample (e.g., $\lceil \sqrt{B} \rceil$) for each component may be sufficient to generate the bootstrap sample of size B for $\boldsymbol{\varepsilon}_t$.

Once a bootstrap sample is generated at level $p = 1$ of the DCBS algorithm, the same sample can be used repeatedly for critical value selection at levels $p \geq 2$. That is, in Step 4.3 above, the test statistics are computed as $\mathcal{T}^{\varphi,l}_{s',e'} = \max_{b \in [s',e'),\, 1 \leq m \leq n} \mathcal{D}^{\varphi}_m(\{|\mathcal{E}^{(j),l}_{s',b,e'}|\}_{j=1}^n)$ over a moving window of size $e - s + 1$ for $1 \leq s' \leq T - e + s$, $e' = s' - s + e$ and $l = 1, \ldots, B$, from which the test criterion is drawn.

Establishing the validity of the GDFM-boot algorithm for change-point analysis is beyond the scope of the current paper and hence is left for the future investigation. However, simulation studies in Section 5 confirm its good performance in various settings together with the proposed change-point statistic. Computationally, the GDFM-boot algorithm benefits from the dimension reduction via factor modelling. When applied to generate a bootstrap sample of size $B = 100$ with $n = 25$ and $T = 100$ (executed on a 3.10GHz quad-core with 8GB of RAM running Windows 7), the `R` code implementing the algorithm took 0.38 seconds, compared to 62 seconds taken by the Linear Process Bootstrap. Further, applying the latter was not computationally feasible for any data of the size and dimensionality considered in our simulation study.

5. Simulation study

5.1. Single change-point detection

In this section, we evaluate the empirical performance of the DC statistic on simulated panel data with (at most) a single change-point. For comparison, change-point tests with two different choices of $\varphi = 0, 1/2$ (referred to as $\mathcal{T}^0$ and $\mathcal{T}^{1/2}$) as well as the combined DC statistic ($\widetilde{\mathcal{T}}$) from Section 4.1 are considered, with $\pi^{\varphi}_{n,T}$ computed from the GDFM-boot algorithm. Also included are

- $\mathcal{T}^{\text{EH}}_{1,T} = (\mathcal{T}^{\text{linear}}_{1,T} > 1) \vee (\mathcal{T}^{\text{scan}}_{1,T} > 1)$ equipped with the thresholds H chosen as $(1 - \alpha/2)$-quantile of the χ^2_n-distribution and $T_m = \frac{2}{\sqrt{2m}}\{m \log(ne/m) + \log(nT/\alpha)\}$, as per the recommendation made in Section 5 of Enikeeva and Harchaoui (2014) (referred to as $\mathcal{T}^{\text{EH}}$);
- $\mathcal{T}^{\text{Jirak}}_{1,T}$ with the test criterion chosen according to the bootstrap algorithm (Algorithm 4.6) proposed in Jirak (2015) with the block size $K = 4$ (referred to as $\mathcal{T}^{\text{Jirak}}$).

Since the scaling of component series is not discussed in Enikeeva and Harchaoui (2014), we choose to apply $\widehat{\sigma}_j$ estimated as in Remark 3.2 in deriving $\mathcal{T}^{\text{EH}}_{1,T}$. Jirak (2015) showed the consistency of the proposed test when applied with a set of long-run variance estimators for σ_j. As discussed in Section 5.1.2, $\mathcal{T}^{\text{Jirak}}$ is highly sensitive to the scaling terms estimated spuriously small, and therefore we use the most conservative estimator among those included in Proposition 3.5 of Jirak (2015). $\mathcal{T}^{\text{HH}}_{1,T}$ is similarly constructed as $\mathcal{T}^{\text{linear}}_{1,T}$ with the same high-dimensional efficiency, and $\mathcal{T}^{\text{EH}}$ is expected to perform better than either of the two component tests. Hence $\mathcal{T}_{1,T}^{\text{HH}}$ is omitted from our study.

As described in Section 5.1.1, each coordinate of $\boldsymbol{\varepsilon}_t$ is generated from the same model, which enables us to select a single threshold π_T applicable to all n component series in deriving $\mathcal{T}_{1,T}^{\text{SBS}}(\pi_T)$. Hence we include $\mathcal{T}_{1,T}^{\text{SBS}}(\pi_T)$ equipped with the "oracle" threshold in the comparative study (referred to as $\mathcal{T}^{\text{SBS}}$), where π_T is chosen from the GDFM-boot algorithm with (unobservable) $\boldsymbol{\varepsilon}_t$ replacing the estimated $\bar{\boldsymbol{\varepsilon}}_t$. Then, π_T is the $(1-\alpha)$-quantile of $\max_{b\in[1,T),\,1\le j\le n}|\mathcal{E}_{1,b,T}^{j,l}|$, $l = 1,\ldots,B$.

When a bootstrap procedure is involved for test criterion selection, the bootstrap sample size is fixed at $B = 100$. We report the Type I error and the size-corrected power at the significance level $\alpha = 0.05$, as well as the location accuracy ($|\widehat{\eta}_1 - \eta_1| < \log T$, in %) for $\mathcal{T}^0$, $\mathcal{T}^{1/2}$, $\widetilde{\mathcal{T}}$, $\mathcal{T}^{\text{Jirak}}$ and $\mathcal{T}^{\text{SBS}}$ over 100 simulated sample paths for each setting; for $\mathcal{T}^{\text{EH}}$, it is not suggested how the change-point location is to be estimated. For all tests, $d_T = 5$ is used to trim off the extreme ends of the interval $[1, T]$. Also, we investigate the partitioning accuracy of $\mathcal{T}^0$, $\mathcal{T}^{1/2}$, $\widetilde{\mathcal{T}}$, $\mathcal{T}^{\text{Jirak}}$ and $\mathcal{T}^{\text{SBS}}$ by reporting the Rand index, the sum of true positives and true negatives divided by the total (n), where Rand index close to 1 indicates more accurate partitioning (binary classification).

5.1.1. *Data generating models*

We consider piecewise constant signals $\{f_{j,t}\}_{t=1}^T$ with varying sparsity/density (m_1), size of jumps ($|\delta_{j,1}| \sim \mathcal{U}(0.75\delta_1, 1.25\delta_1)$ for $j \in \Pi_1$ with randomly assigned signs) and locations $t = \eta_1$.

Motivated by Jirak (2015), $\boldsymbol{\varepsilon}_t$ is generated from the following two models:

(N1) ARMA(2,2) model. With $\varrho_i = \varrho(i+1)^{-1}$ and $\sigma_v = 0.1\varrho^{-1}$,

$$\begin{aligned} u_{j,t} &= \sum_{i=0}^{99} \varrho_i v_{j-i,t}, \qquad v_{j,t} \sim_{\text{i.i.d}} \mathcal{N}(0,\sigma_v^2), \\ \varepsilon_{j,t} &= 0.2\varepsilon_{j,t-1} - 0.3\varepsilon_{j,t-2} + u_{j,t} + 0.2u_{j,t-1}. \end{aligned} \tag{9}$$

(N2) Factor model. $u_{j,t}$ is generated as in (9) with $\varrho = 0.2$ and $\sigma_v = 0.5\sqrt{1-\varrho_h^2}$, and

$$\varepsilon_{j,t} = \varrho_h h_t + 0.2\varepsilon_{j,t-1} - 0.3\varepsilon_{j,t-2} + u_{j,t} + 0.2u_{j,t-1},\ h_t \sim_{\text{i.i.d}} \mathcal{N}(0, 0.1^2).$$

It is easily seen that the degree of cross-sectional correlations is controlled by the choice of $\varrho \in \{0.2, 0.5\}$ in (N1) and $\varrho_h \in \{0.5, 0.9\}$ in (N2).

5.1.2. *Results*

Table 2 compares the Type I errors of different tests in consideration. Combined with the DGFM-boot procedure, the DC statistic-based tests generally manages to control the Type I errors below the nominal level ($\alpha = 0.05$) or slightly above, with the exception of the case when $n = T = 250$ and the ARMA model (N1) is used to generate the noise with $\varrho = 0.5$. The oracle threshold for $\mathcal{T}^{\text{SBS}}$ leads to a very conservative test, which is also evident in the power performance of $\mathcal{T}^{\text{SBS}}$.

$\mathcal{T}^{\text{Jirak}}$ appears to be highly sensitive and therefore vulnerable to the choice of $\widehat{\sigma}_j$ particularly when the critical value is selected by the parametric bootstrap (not reported here), since the test statistic directly depends on the largest CUSUM value attained by a single component series. When the noise is generated as in (N2) with $\varrho_h = 0.9$, the size of $\mathcal{T}^{\text{Jirak}}$ is the closest to the nominal level, which is attributed to the presence of a strong factor as it leads to the scaling terms being estimated homogeneously across the panel. This presence of a strong factor has an adverse effect on the size of $\mathcal{T}^{\text{EH}}$, which is originally proposed for independent panel data (see also Remark 2.2).

To observe the power behaviour, we present the results when $n = 250$ and $T = 100$ as a representative example in Tables 3–4 and report the rest in Appendix C. Increasing sample size generally leads to improved power and accuracy in estimated change-point location. As for the dimensionality, its increase has different effects depending on the error generating model: while there is no strong discernible trend with regards to increasing n in the case of (N1), it brings in dramatic improvement in the performance of $\mathcal{T}^0$ for the noise generated from (N2), especially with increasing influence of the factor (when $\varrho_h = 0.9$).

The tests tend to lose power when the change-point is sparse, the jumps are of smaller magnitude and its location is distanced away from the centre, and the similar arguments apply to the location and partitioning accuracy. Note that the Rand index occasionally decreases with increasing m_1, as it measures both true positives and true negatives where the former may decrease with growing density of the change-point. This indicates that over the range of δ_1 considered here, some $x_{j,t}$, $j \in \Pi_1$ may not contribute to change-point detection due to small jump size $|\delta_{j,1}|$.

When $\boldsymbol{\varepsilon}_t$ is generated from (N1) (where the cross-sectional correlations are relatively small), $\mathcal{T}^{\text{EH}}$ and $\widetilde{\mathcal{T}}$ consistently achieve high (size-corrected) power over the entire range of change-point configurations, and the latter also achieves high location accuracy and Rand index. In the case of (N2), $\mathcal{T}^0$ outperforms all the other tests considered in our study in all change-point configurations and criteria for evaluation. Although not reported here, the performance of $\widetilde{\mathcal{T}}$ in this particular setting can be improved by using a greater value for γ_n, which prompts the development of a data-driven way of exploiting the information embedded in an array of DC statistics over $\varphi \in [0, 1]$. When the size of the data increases ($T = 250$), the performance of $\widetilde{\mathcal{T}}$ catches up with that of $\mathcal{T}^0$ in this setting. Comparing $\mathcal{T}^0$ and $\mathcal{T}^{1/2}$, the former performs superior to the latter except for when the change-point is highly dense and the noise is generated from (N1).

$\mathcal{T}^{\text{Jirak}}$ performs as well as $\mathcal{T}^0$ and $\widetilde{\mathcal{T}}$ or, occasionally, even better when the change-point is centrally located, but its performance deteriorates greatly when $\eta_1 = 0.1T$. This can be explained by the presence of a multiplicative factor

TABLE 2
Type I error when $\alpha = 0.05$; $n = 100$ (top) and $n = 250$ (bottom).

	ϱ/ϱ_h	$T = 100$						$T = 250$					
		$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{EH}}$	$\mathcal{T}^{\text{SBS}}$	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{EH}}$	$\mathcal{T}^{\text{SBS}}$
(N1)	0.2	0.06	0.05	0.06	0.15	0.08	0	0.01	0.07	0.01	0.11	0.1	0
	0.5	0.04	0.03	0.04	0.19	0.08	0	0.02	0.06	0.02	0.14	0.11	0
(N2)	0.5	0.08	0.04	0.07	0.13	0.16	0	0.01	0.08	0.01	0.18	0.23	0
	0.9	0.04	0.04	0.04	0.1	0.57	0	0.04	0.05	0.05	0.05	0.64	0
(N1)	0.2	0.06	0	0.06	0.15	0.07	0	0.07	0.03	0.07	0.18	0.09	0
	0.5	0.06	0.01	0.04	0.21	0.09	0	0.12	0.01	0.1	0.22	0.05	0
(N2)	0.5	0.07	0.04	0.05	0.04	0.35	0	0.02	0.05	0.05	0.08	0.34	0
	0.9	0.04	0.07	0.07	0.1	0.61	0	0.05	0.05	0.05	0.06	0.75	0

involving b in (2). $\sqrt{b(T-b)/T}|\mathcal{X}^j_{1,b,T}|$ amounts to a CUSUM-based change-point test that attains the slowest rate at which its Type II error converges to zero among the set of CUSUM statistics studied in Section 3 of Brodsky and Darkhovsky (1993), with the rate depending on $T^{\beta-1}$ instead of $T^{\beta/2-1/2}$ as is the case with $|\mathcal{X}^j_{1,b,T}|$; see Theorem 3.5.2 of Brodsky and Darkhovsky (1993) for further details. Due to the conservative behaviour of the oracle threshold as observed in Table 2, $\mathcal{T}^{\text{SBS}}$ does not outperform the other tests.

Interestingly, when $\mathcal{T}^{\text{Jirak}}$, $\mathcal{T}^0$, $\mathcal{T}^{1/2}$ and $\widetilde{\mathcal{T}}$ attain the similar level of power, the latter two achieve higher accuracy in locating the change-point. It is attributed to the fact that the latter usually select greater $\widehat{m}^{\varphi}_b$ in (6), which is evidenced by the higher Rand index values.

5.2. Multiple change-point detection

In this section, we evaluate the empirical behaviour of the DCBS algorithm with $\mathcal{T}^0$ and $\widetilde{\mathcal{T}}$, based on their good performance observed in Section 5.1.2. For comparison, we also investigate the performance of the SBS algorithm furnished with the oracle threshold.

Fixing the dimensionality and the size of data at $n = 250$ and $T = 250$, we consider piecewise constant signal $\{f_{j,t}\}_{t=1}^T$ containing three change-points as follows: at $t = \eta_r$, an index set Π_r of cardinality m_r is randomly drawn from $\{1, \ldots, n\}$, where $|\delta_{j,r}| \sim_{\text{i.i.d}} \mathcal{U}(0.75\delta_r, 1.25\delta_r)$ for $j \in \Pi_r$. We set $(\eta_1, m_1, \delta_1) = ([0.3T], [0.75n], 0.050)$, $(\eta_2, m_2, \delta_2) = ([0.6T], [0.25n], 0.087)$, and $(\eta_3, m_3, \delta_3) = ([0.8T], [0.1n], 0.140)$ such that $m_r\delta_r^2$ remains identical over all $r = 1, 2, 3$. The noise $\boldsymbol{\varepsilon}_t$ is generated as in (N1) and (N2) of Section 5.1.1 with varying ϱ and ϱ_h. We set $B = 100$, $d_T = 5$ and $L_T = [\log_2(\log T + 1)]$ where the latter is chosen to permit a growing number ($\log T$) of change-points in the data. To account for multiple testing, the Bonferroni's correction is adopted by setting $\alpha = \alpha^*/(2^{L_T} - 1)$ with $\alpha^* = 0.05$. We report the total number of estimated change-points ($\widehat{N}$, in %) and their location accuracy ($|\widehat{\eta}_r - \eta_r| < \log T$, in %), over 100 simulated sample paths for each setting in Table 5 and Figures 2–3.

Overall, the BS algorithm applied with $\widetilde{\mathcal{T}}$ performs the best in detecting all three change-points as well as identifying their locations over 80% of the simulated data. In comparison, $\mathcal{T}^0$ or $\mathcal{T}^{\text{SBS}}$ tend to miss η_1 which is associated with the smallest jump size, while all three methods tend to detect η_3 the

Table 3. $n = 250$, $T = 100$ *and* $\alpha = 0.05$*: (N1) with* $\varrho = 0.2$ *(top) and* $\varrho = 0.5$ *(bottom).*

			size-corrected power						location accuracy (%)					Rand index				
δ_1	m_1	η_1	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{EH}}$	$\mathcal{T}^{\text{SBS}}$	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{SBS}}$	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{SBS}}$
0.05	$\log\, n$	$0.1T$	0.08	0.02	0.08	0.03	0.06	0.08	2	0	2	0	2	0.08	0.02	0.08	0.03	0.08
		$0.5T$	0.1	0.03	0.09	0.11	0.06	0.08	0	0	0	6	0	0.10	0.03	0.09	0.11	0.08
	$\sqrt{n}$	$0.1T$	0.11	0.03	0.09	0.07	0.02	0.07	1	0	1	0	1	0.10	0.03	0.08	0.06	0.06
		$0.5T$	0.24	0.1	0.25	0.3	0.19	0.21	10	2	10	18	9	0.21	0.10	0.22	0.27	0.19
	$0.4n$	$0.1T$	0.14	0.17	0.14	0.06	0.15	0.11	2	13	2	0	1	0.03	0.11	0.03	0.01	0.02
		$0.5T$	0.56	1	0.65	0.77	0.97	0.6	23	96	31	53	26	0.12	0.77	0.16	0.18	0.13
0.075	$\log\, n$	$0.1T$	0.09	0.01	0.09	0.05	0.04	0.07	0	0	0	0	1	0.09	0.01	0.09	0.05	0.07
		$0.5T$	0.31	0.05	0.31	0.41	0.16	0.25	16	0	17	31	14	0.30	0.05	0.30	0.40	0.24
	$\sqrt{n}$	$0.1T$	0.1	0.03	0.1	0.06	0.03	0.1	1	0	1	0	3	0.09	0.03	0.09	0.05	0.09
		$0.5T$	0.51	0.15	0.53	0.69	0.53	0.48	22	5	23	47	22	0.45	0.15	0.47	0.62	0.43
	$0.4n$	$0.1T$	0.23	0.87	0.24	0.11	0.75	0.24	12	86	12	0	16	0.05	0.66	0.05	0.02	0.05
		$0.5T$	0.89	1	1	1	0.97	0.98	55	100	90	75	75	0.19	0.87	0.53	0.30	0.23
0.1	$\log\, n$	$0.1T$	0.12	0.02	0.14	0.04	0.04	0.11	5	0	5	0	5	0.12	0.02	0.14	0.04	0.11
		$0.5T$	0.55	0.08	0.55	0.74	0.34	0.49	31	0	33	58	31	0.53	0.08	0.53	0.72	0.48
	$\sqrt{n}$	$0.1T$	0.15	0.05	0.15	0.09	0.09	0.1	7	1	7	0	4	0.13	0.05	0.13	0.08	0.09
		$0.5T$	0.76	0.47	0.82	0.94	0.96	0.74	52	33	58	74	57	0.68	0.47	0.74	0.86	0.66
	$0.4n$	$0.1T$	0.42	1	0.68	0.11	0.97	0.51	28	100	55	0	34	0.09	0.84	0.26	0.02	0.11
		$0.5T$	1	1	1	1	0.97	1	76	100	99	85	97	0.21	0.93	0.69	0.47	0.29
0.05	$\log\, n$	$0.1T$	0.08	0.02	0.08	0.03	0.06	0.08	2	0	2	0	2	0.08	0.02	0.08	0.03	0.08
		$0.5T$	0.1	0.03	0.09	0.11	0.06	0.08	0	0	0	6	0	0.10	0.03	0.09	0.11	0.08
	$\sqrt{n}$	$0.1T$	0.11	0.03	0.09	0.07	0.02	0.07	1	0	1	0	1	0.10	0.03	0.08	0.06	0.06
		$0.5T$	0.24	0.1	0.25	0.3	0.19	0.21	10	2	10	18	9	0.21	0.10	0.22	0.27	0.19
	$0.4n$	$0.1T$	0.14	0.17	0.14	0.06	0.15	0.11	2	13	2	0	1	0.03	0.11	0.03	0.01	0.02
		$0.5T$	0.56	1	0.65	0.77	0.97	0.6	23	96	31	53	26	0.12	0.77	0.16	0.18	0.13
0.075	$\log\, n$	$0.1T$	0.09	0.01	0.09	0.05	0.04	0.07	0	0	0	0	1	0.09	0.01	0.09	0.05	0.07
		$0.5T$	0.31	0.05	0.31	0.41	0.16	0.25	16	0	17	31	14	0.30	0.05	0.30	0.40	0.24
	$\sqrt{n}$	$0.1T$	0.1	0.03	0.1	0.06	0.03	0.1	1	0	1	0	3	0.09	0.03	0.09	0.05	0.09
		$0.5T$	0.51	0.15	0.53	0.69	0.53	0.48	22	5	23	47	22	0.45	0.15	0.47	0.62	0.43
	$0.4n$	$0.1T$	0.23	0.87	0.24	0.11	0.75	0.24	12	86	12	0	16	0.05	0.66	0.05	0.02	0.05
		$0.5T$	0.89	1	1	1	0.97	0.98	55	100	90	75	75	0.19	0.87	0.53	0.30	0.23
0.1	$\log\, n$	$0.1T$	0.12	0.02	0.14	0.04	0.04	0.11	5	0	5	0	5	0.12	0.02	0.14	0.04	0.11
		$0.5T$	0.55	0.08	0.55	0.74	0.34	0.49	31	0	33	58	31	0.53	0.08	0.53	0.72	0.48
	$\sqrt{n}$	$0.1T$	0.15	0.05	0.15	0.09	0.09	0.1	7	1	7	0	4	0.13	0.05	0.13	0.08	0.09
		$0.5T$	0.76	0.47	0.82	0.94	0.96	0.74	52	33	58	74	57	0.68	0.47	0.74	0.86	0.66
	$0.4n$	$0.1T$	0.42	1	0.68	0.11	0.97	0.51	28	100	55	0	34	0.09	0.84	0.26	0.02	0.11
		$0.5T$	1	1	1	1	0.97	1	76	100	99	85	97	0.21	0.93	0.69	0.47	0.29
		average	0.35	0.34	0.38	0.37	0.41	0.35	19.06	29.78	24.78	24.83	22.11	0.20	0.29	0.26	0.24	0.19

Table 4. $n = 250$, $T = 100$ *and* $\alpha = 0.05$*: (N2) with* $\varrho_h = 0.5$ *(top) and* $\varrho_h = 0.9$ *(bottom).*

			size-corrected power						location accuracy (%)					Rand index				
δ_1	m_1	η_1	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{EH}}$	$\mathcal{T}^{\text{SBS}}$	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{SBS}}$	$\mathcal{T}^0$	$\mathcal{T}^{1/2}$	$\widetilde{\mathcal{T}}$	$\mathcal{T}^{\text{Jirak}}$	$\mathcal{T}^{\text{SBS}}$
0.05	$\log n$	$0.1T$	0.22	0.06	0.14	0.04	0	0.02	4	1	1	0	0	0.21	0.06	0.14	0.04	0.02
		$0.5T$	0.3	0.07	0.19	0.08	0	0.05	12	0	6	6	1	0.29	0.07	0.18	0.08	0.05
	$\sqrt{n}$	$0.1T$	0.2	0.1	0.16	0.02	0	0.02	5	1	4	0	0	0.18	0.10	0.15	0.02	0.02
		$0.5T$	0.45	0.1	0.21	0.09	0	0.03	18	1	6	4	0	0.40	0.10	0.19	0.08	0.03
	$0.4n$	$0.1T$	0.4	0.08	0.25	0.07	0	0.02	20	2	11	0	0	0.08	0.05	0.10	0.01	0.00
		$0.5T$	0.84	0.49	0.88	0.55	0.62	0.1	40	39	67	39	6	0.18	0.38	0.45	0.14	0.02
0.075	$\log n$	$0.1T$	0.24	0.04	0.16	0.02	0	0.03	7	0	6	0	0	0.23	0.04	0.16	0.02	0.03
		$0.5T$	0.68	0.08	0.48	0.26	0	0.11	40	1	27	16	8	0.66	0.08	0.47	0.25	0.11
	$\sqrt{n}$	$0.1T$	0.32	0.09	0.21	0.07	0	0.03	20	1	8	0	1	0.28	0.09	0.19	0.06	0.03
		$0.5T$	0.91	0.13	0.76	0.55	0.05	0.17	53	2	54	41	10	0.81	0.13	0.69	0.50	0.15
	$0.4n$	$0.1T$	0.7	0.26	0.61	0.07	0.33	0.03	44	19	48	0	2	0.15	0.19	0.28	0.02	0.01
		$0.5T$	1	1	1	0.9	0.7	0.7	54	100	100	65	50	0.21	0.90	0.72	0.31	0.15
0.1	$\log n$	$0.1T$	0.37	0.05	0.21	0.06	0	0.03	18	1	11	0	0	0.36	0.05	0.20	0.06	0.03
		$0.5T$	0.95	0.08	0.85	0.68	0	0.49	71	0	64	55	41	0.92	0.08	0.83	0.67	0.48
	$\sqrt{n}$	$0.1T$	0.66	0.07	0.35	0.03	0	0.04	46	2	23	0	2	0.59	0.07	0.32	0.03	0.04
		$0.5T$	0.99	0.15	0.98	0.88	0.69	0.7	80	3	87	74	60	0.88	0.15	0.90	0.82	0.63
	$0.4n$	$0.1T$	0.92	0.86	0.99	0.04	0.7	0.2	77	84	96	1	17	0.19	0.73	0.63	0.01	0.04
		$0.5T$	1	1	1	0.97	0.7	0.99	77	100	100	86	96	0.21	0.94	0.84	0.52	0.27
0.05	$\log n$	$0.1T$	0.18	0.05	0.05	0.05	0	0.02	12	2	2	0	0	0.17	0.05	0.05	0.05	0.02
		$0.5T$	0.75	0.05	0.05	0.24	0	0.03	43	0	0	12	2	0.73	0.05	0.05	0.23	0.03
	$\sqrt{n}$	$0.1T$	0.33	0.06	0.06	0.05	0	0.03	21	1	1	0	1	0.29	0.06	0.06	0.04	0.03
		$0.5T$	0.96	0.07	0.07	0.39	0	0.1	69	0	0	28	7	0.85	0.07	0.07	0.35	0.09
	$0.4n$	$0.1T$	0.69	0.07	0.07	0.05	0	0.03	47	2	2	0	1	0.14	0.06	0.06	0.01	0.01
		$0.5T$	1	0.12	0.17	0.74	0.39	0.21	60	2	7	47	13	0.21	0.08	0.11	0.24	0.05
0.075	$\log n$	$0.1T$	0.6	0.05	0.05	0.04	0	0.04	43	2	2	0	2	0.58	0.05	0.05	0.04	0.04
		$0.5T$	1	0.07	0.09	0.76	0	0.27	83	0	2	64	22	0.97	0.07	0.09	0.74	0.26
	$\sqrt{n}$	$0.1T$	0.89	0.09	0.09	0.06	0	0.05	76	2	2	0	2	0.79	0.09	0.09	0.05	0.04
		$0.5T$	1	0.09	0.11	0.85	0.28	0.43	88	0	5	70	37	0.89	0.09	0.10	0.79	0.39
	$0.4n$	$0.1T$	0.94	0.09	0.09	0.12	0.25	0.12	70	2	2	1	6	0.20	0.07	0.07	0.03	0.03
		$0.5T$	1	0.73	0.92	0.99	0.44	0.77	81	64	87	79	59	0.21	0.65	0.75	0.51	0.22
0.1	$\log n$	$0.1T$	0.9	0.08	0.08	0.05	0	0.06	86	2	2	1	4	0.87	0.08	0.08	0.05	0.06
		$0.5T$	1	0.06	0.25	0.97	0.14	0.8	92	0	21	85	73	0.97	0.06	0.24	0.96	0.78
	$\sqrt{n}$	$0.1T$	1	0.06	0.06	0.06	0	0.14	95	2	2	2	9	0.89	0.06	0.06	0.05	0.13
		$0.5T$	1	0.07	0.59	1	0.44	0.96	90	0	57	90	89	0.89	0.07	0.55	0.97	0.88
	$0.4n$	$0.1T$	0.99	0.1	0.29	0.16	0.44	0.23	85	6	25	3	17	0.21	0.08	0.21	0.04	0.06
		$0.5T$	1	1	1	1	0.44	0.99	86	99	99	93	92	0.21	0.96	0.90	0.80	0.45
		average	0.73	0.21	0.38	0.36	0.18	0.25	53.14	15.08	28.81	26.72	20.28	0.47	0.19	0.31	0.27	0.16

TABLE 5
Summary of the total number of estimated change-points and their location accuracy: $n = 250$, $T = 250$ and $\alpha^* = 0.05$.

	ϱ/ϱ_h		$\widehat{N}$ (%) 0	1	2	**3**	4	≥ 5	accuracy (%) η_1	η_2	η_3
(N1)	0.2	$\mathcal{T}^0$	0	9	37	53	0	1	35	71	89
		$\widetilde{\mathcal{T}}$	0	1	12	85	1	1	90	87	93
		$\mathcal{T}^{\text{SBS}}$	8	66	26	0	0	0	0	37	77
	0.5	$\mathcal{T}^0$	0	4	37	56	3	0	46	71	92
		$\widetilde{\mathcal{T}}$	0	1	8	88	3	0	95	89	96
		$\mathcal{T}^{\text{SBS}}$	4	67	29	0	0	0	0	46	75
(N2)	0.5	$\mathcal{T}^0$	1	3	31	65	0	0	34	81	97
		$\widetilde{\mathcal{T}}$	0	2	7	83	8	0	96	95	97
		$\mathcal{T}^{\text{SBS}}$	2	21	76	1	0	0	2	83	90
	0.9	$\mathcal{T}^0$	0	2	10	87	1	0	74	97	98
		$\widetilde{\mathcal{T}}$	3	35	37	24	1	0	44	38	92
		$\mathcal{T}^{\text{SBS}}$	0	22	69	9	0	0	9	76	100

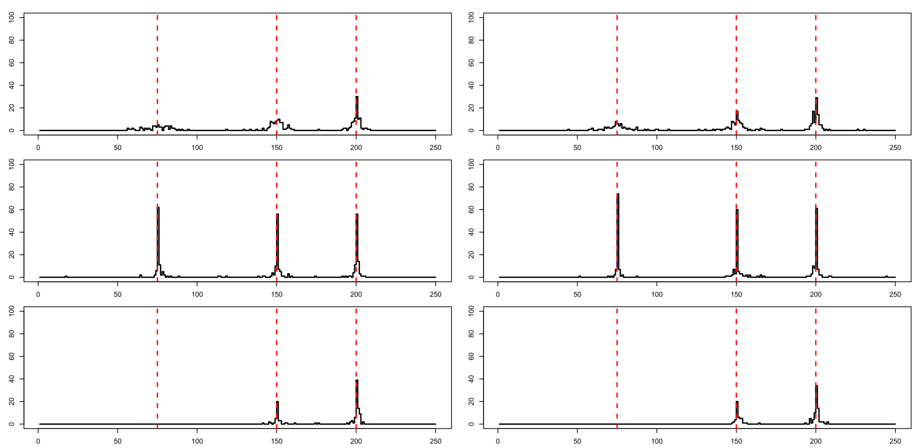

FIG 2. *(N1) with $\varrho = 0.2$ (left) and $\varrho = 0.5$ (right): the locations of the change-points detected by the BS algorithm in combination with $\mathcal{T}^0$ (top), $\widetilde{\mathcal{T}}$ (middle) and $\mathcal{T}^{SBS}$ (bottom); vertical lines indicate the locations of true η_r, $r = 1, 2, 3$.*

best, which is associated with the largest jump size. The behaviour of $\mathcal{T}^0$ and $\widetilde{\mathcal{T}}$ change dramatically when $\boldsymbol{\varepsilon}_t$ is generated from (N2) with a strong factor ($\varrho_h = 0.9$), where the performance of $\mathcal{T}^0$ improves (identifying all three change-points over 70% of the simulated data) whereas that of the latter deteriorates greatly. As observed in Section 5.1.2, using a larger scaling term γ_n may improve the performance of $\widetilde{\mathcal{T}}$ in this setting. While the performance of $\mathcal{T}^{\text{SBS}}$ is better when $\boldsymbol{\varepsilon}_t$ is generated from (N2) rather than (N1), the choice of threshold for $\mathcal{T}^{\text{SBS}}$ turns out to be too conservative overall. Although each $\varepsilon_{j,t}$ is generated from the identical model, employing n thresholds for the n-dimensional panel data may lead to better small sample performance of $\mathcal{T}^{\text{SBS}}$.

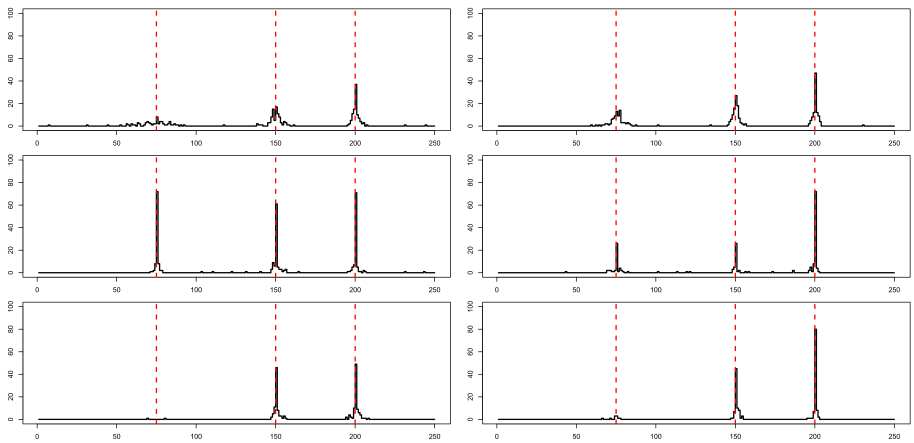

FIG 3. *(N2) with* $\varrho_h = 0.5$ *(left) and* $\varrho_h = 0.9$ *(right).*

6. Application to financial time series data

We analysed the log returns of the closing prices of all S&P 100 component stocks between February 24, 2015 and February 23, 2016, which are denoted by $y_{i,t}$, $i = 1, \ldots, \widetilde{n}$; $t = 1, \ldots, T$ with $\widetilde{n} = 88$ (only those components which remained in the index for a longer period were included) and $T = 252$. Jirak (2015) analysed a similar financial dataset for a single change in its mean and variance, respectively. However, considering that (i) log returns are often modelled to have zero-mean and time-varying conditional variance using conditionally heteroscedastic models, and (ii) it is difficult to rule out the possible existence of multiple change-points, we chose to perform the change-point analysis in the (unconditional) second-order structure of $y_{i,t}$ using the DCBS algorithm. For the purpose, wavelet-based periodogram and cross-periodogram sequences of $y_{i,t}$ were computed, which were also adopted to comprise the input panel data to the SBS-MVTS algorithm in Cho and Fryzlewicz (2015). Any change-point in the autocovariance and cross-covariance structure of $y_{i,t}$ is detectable from examining the wavelet (cross-)periodogram sequences; for further details, see Section 3.1 of Cho and Fryzlewicz (2015).

We used Haar wavelets at the two finest scales to produce the periodogram sequences, which are denoted by $x_{j,t}$, $j = 1, \ldots, n = \widetilde{n}(\widetilde{n} + 1)$; $t = 1, \ldots, T$. The DCBS algorithm with $\widetilde{\mathcal{D}}_m$ detects a single change-point at $t = 220$, which corresponds to January 6, 2016. It has been noted that the first week of trading in 2016 marked the worst five-day start to a year ever, according to S&P Dow Jones Indices (Financial Times, http://www.ft.com/fastft/2016/01/07/sp-500-logs-worst-annual-kick-off-on-record/). For example, the S&P 500 index dropped by 4.9% during the period, and the Dow Jones Industrial Average by 6.19%. Figure 4 shows $y_{i,t}$ (left) and the pointwise maximum of the DC statistics at the first iteration of the DCBS algorithm (right), where

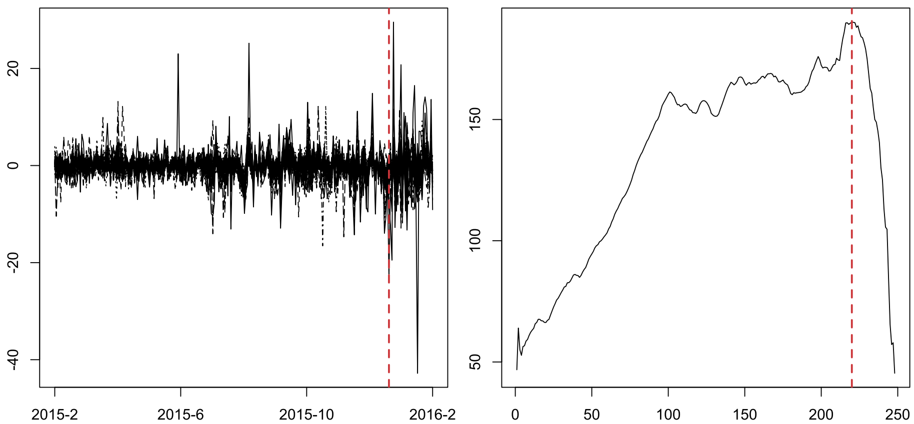

FIG 4. *Log returns of S&P* 100 *component stock prices* ($y_{i,t}$) *between February* 24, 2015 *and February* 23, 2016 *(left); pointwise maximum of* $\widetilde{\mathcal{D}}_m(\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n)$ *over* $b = 1, \ldots, T-1$ *(right); the vertical broken line denotes the location of the estimated change-point.*

such behaviour of the financial market at the beginning of 2016 is reflected as a large peak in the latter.

7. Conclusions

In this paper, we have proposed the DC statistic, a novel way of aggregating high-dimensional CUSUM series across the panel for change-point analysis, and showed its consistency in single and multiple change-point detection both theoretically and empirically. We conclude by listing possible future research projects.

- The DC statistic can be applied to detect changes in high-dimensional time series besides those in the mean of panel data. For example, the DCBS algorithm is easily extended to detect change-points in both autocovariance and cross-covariance structure of n-dimensional time series, by taking panel data consisting of local periodogram and cross-periodogram sequences as an input, see the real data analysis in Section 6.
  Moreover, the DC operator may be regarded as a generic tool that can be adopted to aggregate multiple series of statistics cross-sectionally, the result of which can be utilised for panel data analysis beyond change-point detection.
- Empirically, it was shown that $\widetilde{\mathcal{T}}$ generally outperforms $\mathcal{T}^\varphi$ for any $\varphi \in \{0, 1/2\}$, while in the presence of strong cross-correlations, $\mathcal{T}^0$ performs better than the rest. This opens up several possibilities for future research, including the investigation of the "optimal" way of exploiting the information contained in the infinite array of DC statistics over $\varphi \in [0, 1]$.

- Proposed here for test criterion selection, the GDFM-boot algorithm showed good empirical performance, but it remains to investigate the validity of its application to change-point analysis by establishing that the bootstrap scheme mimics the correct second-order structure for a large class of time series processes. The GDFM-boot will have applicability to a wide range of inference and estimation problems involving high-dimensional time series beyond the context of change-point analysis. For example, a key task in factor analysis is to estimate the number of common factors that drive the pervasive cross-sectional correlations, and several information criterion-type estimators have been proposed. However, there is lack of any attempt at statistical inference on the factor number, e.g., by constructing its confidence interval, and the GDFM-boot can be adopted for such tasks.

8. Proofs

8.1. Preliminary results

We first prove a set of lemmas that are stepping stones for the proofs of Theorems 3.1–4.1. We assume that (A1)–(A3) and (A6) hold in all the lemmas where applicable, and the notations $C_i,\ c_i,\ i = 0, 1, \ldots$ are adopted to denote fixed positive constants throughout the proofs. Further, $\widetilde{\varepsilon}_{j,t}$ denotes the scaled $\varepsilon_{j,t}$ with respect to an estimator $\widehat{\sigma}_j$ satisfying (A6), and $\widetilde{x}_{j,t}$ and $\widetilde{f}_{j,t}$ are defined similarly. Throughout, we operate within $\mathcal{E}_\sigma$ defined in (A6).

Lemma 8.1. *Defining the set* $\mathcal{I}_1 = \{(s, e) : 1 \le s < e \le T,\ e - s + 1 > d_T = [C \log^2 T]\}$ *and the event* $\mathcal{E}_1 = \{\max_{1\le j\le n} \max_{(s,e)\in\mathcal{I}_1} (e - s + 1)^{-1/2} |\sum_{t=s}^{e} \widetilde{\varepsilon}_{j,t}| \le \log T\}$ *with some fixed constant* $C > 0$*, we have* $\mathbb{P}(\mathcal{E}_1) \to 1$ *as* $T \to \infty$.

Proof. We first study the following probability

$$\mathbb{P}\left(\frac{1}{\sqrt{e-s+1}}\left|\sum_{t=s}^{e}\widetilde{\varepsilon}_{j,t}\right| > \log T\right). \tag{10}$$

Let $d = e-s+1$. Theorem 1.4 of Bosq (1998) showed that under the conditions imposed in (A1), the probability in (10) is bounded from the above by

$$\left\{\frac{2d}{q} + 2\left(1 + \frac{\log^2 T}{25d\mathbb{E}(\widetilde{\varepsilon}_{j,t}^2) + 5C_\varepsilon\sqrt{d}\log T}\right)\right\}\exp\left(-\frac{q\log^2 T}{25d\mathbb{E}(\widetilde{\varepsilon}_{j,t}^2) + 5C_\varepsilon\sqrt{d}\log T}\right)$$
$$+ 11C_\alpha d\left[1 + \frac{5\{\mathbb{E}(\widetilde{\varepsilon}_{j,t}^2)\}^{2/5}\sqrt{d}}{\log T}\right](\mu^{[d/(q+1)]})^{4/5} \tag{11}$$

where $q \in \{1, \ldots, [d/2]\}$. With the choice $q = [c_q d/\log T]$ for some $c_q > 0$, (11) is bounded by

$$2\left(\frac{\log T}{c_q}+2\right)\exp\left\{-\frac{c_q\log T}{25\mathbb{E}(\widetilde{\varepsilon}_{j,t}^2)+5C_\varepsilon C^{-1/2}}\right\}+11C_\alpha d\left[1+\frac{5\{\mathbb{E}(\widetilde{\varepsilon}_{j,t}^2)\}^{2/5}\sqrt{d}}{\log T}\right]$$
$$\times\exp\left\{-\frac{4\log T}{5c_q\log(1/\mu)}\right\}\le C_0T^{3/2}\log^{-1}T\cdot T^{-C_1}$$

for some $C_0, C_1 > 0$, where the latter depends on μ, c_q, C_ε, C and σ^{*2}/σ_*^2. More specifically, we can impose appropriate conditions on the above parameters in order that $C_1 > 7/2+\omega$. Therefore, $\mathbb{P}(\mathcal{E}_1^c)$ is bounded from the above by $nT^2\cdot C_0T^{3/2}\log^{-1}T\cdot T^{-C_1}\to 0$ as $T\to\infty$. □

Lemma 8.2. *Define the set* $\mathcal{I}_2 = \{(s,b,e) : 1\le s<b<e\le T,\ (b-s+1)\wedge(e-b) > d_T\}$ *with* d_T *given in Lemma 8.1, and the event* $\mathcal{E}_2 = \{\max_{1\le j\le n}\max_{(s,b,e)\in\mathcal{I}_2}|\mathcal{C}_b(\{\widetilde{\varepsilon}_{j,t}\}_{t=s}^e)|\le\sqrt{2}\log T\}$*. Then as* $T\to\infty$*, we have* $\mathbb{P}(\mathcal{E}_2)\to 1$*.*

Proof. Note that $\mathcal{I}_2\subset\mathcal{I}_1$, and

$$|\mathcal{C}_b(\{\widetilde{\varepsilon}_{j,t}\}_{t=s}^e)|\le\sqrt{\frac{e-b}{(e-s+1)(b-s+1)}}\left|\sum_{t=s}^b\widetilde{\varepsilon}_{j,t}\right|+$$
$$\sqrt{\frac{b-s+1}{(e-s+1)(e-b)}}\left|\sum_{t=b+1}^e\widetilde{\varepsilon}_{j,t}\right|=I+II.$$

On the event $\mathcal{E}_1$ defined in Lemma 8.1,

$$I=\sqrt{\frac{e-b}{e-s+1}}\cdot\frac{1}{\sqrt{b-s+1}}\left|\sum_{t=s}^b\widetilde{\varepsilon}_{j,t}\right|\le\sqrt{\frac{e-b}{e-s+1}}\log T,$$

and similarly $II\le\sqrt{(b-s+1)/(e-s+1)}\log T$, uniformly over all j and $(s,b,e)\in\mathcal{I}_2$. Hence

$$\begin{aligned}\max_{1\le j\le n}\max_{(s,b,e)\in\mathcal{I}_2}|\mathcal{C}_b(\{\widetilde{\varepsilon}_{j,t}\}_{t=s}^e)| &\le \max_{(s,b,e)\in\mathcal{I}_2}\left\{\sqrt{\frac{e-b}{e-s+1}}+\sqrt{\frac{b-s+1}{e-s+1}}\right\}\log T\\ &\le \sqrt{2}\log T,\end{aligned}$$

with probability tending to one as $T\to\infty$, and therefore $\mathbb{P}(\mathcal{E}_2^c)\le\mathbb{P}(\mathcal{E}_2^c|\mathcal{E}_1)\mathbb{P}(\mathcal{E}_1)+\mathbb{P}(\mathcal{E}_1^c)\to 0$. □

Next, we introduce N additive models $y_{r,t}=g_{r,t}+\xi_{r,t}$, $r=1,\ldots,N$, which play a vital role in the following proofs. For each r, let $\{k_1^r,\ldots,k_n^r\}$ denote a permutation of $\{1,\ldots,n\}$, and $\{i_1^r,\ldots,i_n^r\}$ a set of signs (taking values from {-1, 1} with repetitions), for which the followings hold:

$$i_1^r\cdot\delta_{k_1^r,r}\ge i_2^r\cdot\delta_{k_2^r,r}\ge\ldots\ge i_n^r\cdot\delta_{k_n^r,r}\ge 0$$

(since $\delta_{k_j^r,r} = f_{k_j^r,\eta_r+1} - f_{k_j^r,\eta_r} = 0$ for all $j \geq m_r + 1$, the ordering and the set of signs are not unique). Then $y_{r,t}$, $g_{r,t}$ and $\xi_{r,t}$ are defined as

$$y_{r,t} = \left\{\frac{m_r(2n-m_r)}{2n}\right\}^{\varphi} \left\{\frac{1}{m_r}\sum_{j=1}^{m_r} i_j^r \cdot \widetilde{x}_{k_j^r,t} - \frac{1}{2n-m_r}\sum_{j=m_r+1}^{n} i_j^r \cdot \widetilde{x}_{k_j^r,t}\right\}, \tag{12}$$

$$g_{r,t} = \left\{\frac{m_r(2n-m_r)}{2n}\right\}^{\varphi} \left\{\frac{1}{m_r}\sum_{j=1}^{m_r} i_j^r \cdot \widetilde{f}_{k_j^r,t} - \frac{1}{2n-m_r}\sum_{j=m_r+1}^{n} i_j^r \cdot \widetilde{f}_{k_j^r,t}\right\}, \tag{13}$$

$$\xi_{r,t} = \left\{\frac{m_r(2n-m_r)}{2n}\right\}^{\varphi} \left\{\frac{1}{m_r}\sum_{j=1}^{m_r} i_j^r \cdot \widetilde{\varepsilon}_{k_j^r,t} - \frac{1}{2n-m_r}\sum_{j=m_r+1}^{n} i_j^r \cdot \widetilde{\varepsilon}_{k_j^r,t}\right\}. \tag{14}$$

By its definition, $\{g_{r,t}\}_{t=1}^T$ is a piecewise constant signal with change-points at $t = \eta_1, \ldots, \eta_N$, and its jump at $t = \eta_r$ is of size satisfying the following:

$$g_{r,\eta_r+1} - g_{r,\eta_r} = \left\{\frac{m_r(2n-m_r)}{2n}\right\}^{\varphi} \frac{1}{m_r}\sum_{j\in\Pi_r} \frac{|\delta_{j,r}|}{\widetilde{\sigma}_j} \geq \left\{\frac{m_r n}{2n}\right\}^{\varphi} \frac{2\widetilde{\delta}_r}{3\sigma^*} \geq c_1 m_r^{\varphi} \widetilde{\delta}_r. \tag{15}$$

Then Lemma 8.1 implies that for all $r = 1, \ldots, N$,

$$\max_{(s,e)\in\mathcal{I}_1} \frac{1}{\sqrt{e-s+1}}\left|\sum_{t=s}^{e} \xi_{r,t}\right| \leq 2m_r^{\varphi}\log T \leq 2n^{\varphi}\log T \tag{16}$$

with probability converging to one, since within the event $\mathcal{E}_1$ of Lemma 8.1,

$$\frac{1}{\sqrt{e-s+1}}\left|\sum_{t=s}^{e}\xi_{r,t}\right| \leq \left\{\frac{m_r(2n-m_r)}{2n}\right\}^{\varphi}\left\{\frac{1}{m_r}\sum_{j=1}^{m_r}\left|\frac{1}{\sqrt{e-s+1}}\sum_{t=s}^{e}\varepsilon_{k_j^r,t}\right|\right.$$
$$\left. + \frac{1}{2n-m_r}\sum_{j=m_r+1}^{n}\left|\frac{1}{\sqrt{e-s+1}}\sum_{t=s}^{e}\varepsilon_{k_j^r,t}\right|\right\} \leq 2m_r^{\varphi}\log T.$$

Similarly, Lemma 8.2 implies that for all r,

$$\max_{(s,b,e)\in\mathcal{I}_2} |\mathcal{C}_b(\{\xi_{r,t}\}_{t=s}^{e})| \leq 2\sqrt{2}m_r^{\varphi}\log T \leq 2\sqrt{2}n^{\varphi}\log T \tag{17}$$

with probability tending to one.

Now we consider a generic additive model

$$y_t = g_t + \xi_t, \tag{18}$$

where y_t, g_t and ξ_t are obtained from the panel data $\{x_{j,t}\}_{t=1}^T$, $j = 1, \ldots, n$ in the same manner as $y_{r,t}$, $g_{r,t}$ and $\xi_{r,t}$ in (12)–(14), with respect to some $m \in \{1, \ldots, n\}$, a permutation of index $\{k_1, \ldots, k_n\}$ and a sign sequence $\{i_1, \ldots, i_n\}$. Then g_t is a piecewise constant signal with change-points at $t = \eta_1, \ldots, \eta_N$ satisfying (B1). Also, it is easily seen that the inequalities (16)–(17) hold with the zero-mean noise series ξ_t in place of $\xi_{r,t}$ with probability converging to one.

Recall that s and e denote the start and the end of an interval which is examined at some stage of our search for the change-points. Let s and e satisfy

$$\eta_{q_1} \le s < \eta_{q_1+1} < \ldots < \eta_{q_2} < e \le \eta_{q_2+1} \tag{19}$$

for $0 \le q_1 < q_2 \le N$. In some of the following lemmas, we impose at least one of following conditions:

$$s < \eta_q - c_2 T^\beta < \eta_q + c_2 T^\beta < e \text{ for some } q \in \{q_1 + 1, \ldots, q_2\}, \tag{20}$$

$$\{(\eta_{q_1+1} - s) \wedge (s - \eta_{q_1})\} \vee \{(\eta_{q_2+1} - e) \wedge (e - \eta_{q_2})\} \le c_3 \bar{\epsilon}_T, \tag{21}$$

with $\bar{\epsilon}_T$ to be defined later. The condition (20) implies that there is at least one change-point to be detected which is sufficiently distanced away from the previously detected change-points s and e, and (21) indicates that each of s and e is detected for one of the true change-points.

Since the CUSUM statistics are not affected by the shift in the overall level of g_t, we assume that $\sum_{t=s}^e g_t = 0$ without loss of generality. Then the CUSUM statistics computed on $\{g_t\}_{t=s}^e$ can be re-written as $\mathcal{C}_b(\{g_t\}_{t=s}^e) = \sqrt{\frac{e-s+1}{(b-s+1)(e-b)}} \sum_{t=s}^b g_t$, $b = s, \ldots, e-1$.

Lemma 8.3. *For s and e satisfying (19), there exists $q' \in \{q_1+1, \ldots, q_2\}$ which satisfies $\eta_{q'} = \arg\max_{b \in [s,e)} |\mathcal{C}_b(\{g_t\}_{t=s}^e)|$.*

Proof. The proof follows directly from Lemmas 2.2–2.3 of Venkatraman (1992). ☐

Lemma 8.4. *Let (20) and (21) hold. For $\widehat{\eta} = \arg\max_{b \in [s,e)} |\mathcal{C}_b(\{y_t\}_{t=s}^e)|$, there exists a true change-point $\eta_q \equiv \eta \in (s, e)$ satisfying $|\widehat{\eta} - \eta| < c_0 \bar{\epsilon}_T$ with probability converging to one, provided that*

$$|\mathcal{C}_\eta(\{g_t\}_{t=s}^e)||\mathcal{C}_\eta(\{g_t\}_{t=s}^e) - \mathcal{C}_{\widehat{\eta}}(\{g_t\}_{t=s}^e)| > C_1 n^\varphi \log T \times$$
$$\max\left\{\begin{array}{l} n^\varphi \log T, \\ \sqrt{\eta - s + 1}\left|\frac{1}{\widehat{\eta}-s+1}\sum_{t=s}^{\widehat{\eta}} g_t - \frac{1}{\eta-s+1}\sum_{t=s}^{\eta} g_t\right|, \\ \sqrt{\bar{\epsilon}_T}\left|\frac{1}{\widehat{\eta}-s+1}\sum_{t=s}^{\widehat{\eta}} g_t - \frac{1}{e-\eta}\sum_{t=\eta+1}^{e} g_t\right| \end{array}\right\} \quad \textit{for some } C_1 > 0.$$

Proof. The following proof is an adaptation of the proof of Theorem 3.3.1 in Fryzlewicz (2014) to a non-i.i.d. case.

On the segment $[s, e]$, detecting a change-point is equivalent to fitting the best step function $\widehat{g}_t$ (a piecewise constant function with one change-point) which minimises $\sum_{t=s}^e (y_t - h_t)^2$ among all step functions h_t defined on $[s, e]$. Let g_t^* be the best step function approximation to g_t on $[s, e]$, which may not be unique. From Lemma 8.3, g_t^* needs to have its change-point η coincide with one of the true change-points η_q, $q \in \{q_1+1, \ldots, q_2\}$.

Let us assume that $|\widehat{\eta} - \eta| = c_0\bar{\epsilon}_T$. Due to the fact that $|\mathcal{C}_b(\{g_t\}_{t=s}^e)|$ is monotonic, or decreasing and then increasing in b between two adjacent change-points of g_t (Lemma 2.7 of Venkatraman (1992)), it is enough to consider the case when $\widehat{\eta}$ satisfies $|\widehat{\eta} - \eta| = c_0\bar{\epsilon}_T$. Then, if it is shown that

$$\sum_{t=s}^{e}(y_t - g_t^*)^2 - \sum_{t=s}^{e}(y_t - \widehat{g}_t)^2 < 0, \tag{22}$$

the claim would be proved to be contradiction. Expanding the LHS of (22), we obtain

$$\sum_{t=s}^{e}(\xi_t + g_t - g_t^*)^2 - \sum_{t=s}^{e}(\xi_t + g_t - \widehat{g}_t)^2 = 2\sum_{t=s}^{e}\xi_t(\widehat{g}_t - g_t^*) + \sum_{t=s}^{e}\{(g_t - g_t^*)^2 - (g_t - \widehat{g}_t)^2\} = I + II.$$

Clearly, $II < 0$ from the definition of g_t^*.

Let $\mathbf{\Psi}$ be the set of vectors of length $(e-s+1)$ whose elements are initially positive and constant, then after a break, are negative and constant; moreover, the elements sum to zero and when squared, sum to one. Since we assume that $\sum_{t=s}^{e} g_t = 0$, we can find a vector $\boldsymbol{\psi}^* \in \mathbf{\Psi}$ satisfying $\mathbf{g}^* = \langle \mathbf{g}, \boldsymbol{\psi}^*\rangle \boldsymbol{\psi}^*$ where $\mathbf{g} = (g_s, \ldots, g_e)^\top$ and $\mathbf{g}^* = (g_s^*, \ldots, g_e^*)^\top$. Then we have

$$\sum_{t=s}^{e}(g_t - g_t^*)^2 = \sum_{t=s}^{e} g_t^2 - \langle \mathbf{g}, \boldsymbol{\psi}^*\rangle^2.$$

Let a step function $\widetilde{g}_t$ be chosen to minimise $\sum_{t=s}^{e}(g_t - h_t)^2$ among all the step functions h_t defined on $[s, e]$, under the constraint that h_t has its change-point at $t = \widehat{\eta}$. For such $\widetilde{g}_t$, we have $\sum_{t=s}^{e}(g_t - \widetilde{g}_t)^2 \le \sum_{t=s}^{e}(g_t - \widehat{g}_t)^2$. Again, there exists a vector $\widetilde{\boldsymbol{\psi}} \in \mathbf{\Psi}$ satisfying $\widetilde{\mathbf{g}} = \langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle \widetilde{\boldsymbol{\psi}}$ with $\widetilde{\mathbf{g}} = (\widetilde{g}_s, \ldots, \widetilde{g}_e)^\top$. Then

$$\begin{aligned} |II| &\ge \sum_{t=s}^{e}\{(g_t - \widetilde{g}_t)^2 - (g_t - g_t^*)^2\} = \langle \mathbf{g}, \boldsymbol{\psi}^*\rangle^2 - \langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle^2 \\ &= (\langle \mathbf{g}, \boldsymbol{\psi}^*\rangle + \langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle)(\langle \mathbf{g}, \boldsymbol{\psi}^*\rangle - \langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle) \ge |\langle \mathbf{g}, \boldsymbol{\psi}^*\rangle||\langle \mathbf{g}, \boldsymbol{\psi}^*\rangle - \langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle|, \end{aligned} \tag{23}$$

since $|\mathcal{C}_\eta(\{g_t\}_{t=s}^e)| = |\langle \mathbf{g}, \boldsymbol{\psi}^*\rangle| \ge |\langle \mathbf{g}, \widetilde{\boldsymbol{\psi}}\rangle| = |\mathcal{C}_{\widehat{\eta}}(\{g_t\}_{t=s}^e)|$.

Turning to I, it is decomposed as

$$2\sum_{t=s}^{e}\xi_t(\widehat{g}_t - g_t^*) = 2\sum_{t=s}^{e}\xi_t(\widehat{g}_t - \widetilde{g}_t) + 2\sum_{t=s}^{e}\xi_t(\widetilde{g}_t - g_t^*),$$

and each of the two terms are split into sub-sums computed over the intervals where $\widehat{g}_t - \widetilde{g}_t$ and $\widetilde{g}_t - g_t^*$ are constant, respectively. Letting $\widehat{\eta} > \eta$ without loss of generality, we have

$$\sum_{t=s}^{e} \xi_t(\widetilde{g}_t - g_t^*) = \left(\sum_{t=s}^{\eta} + \sum_{t=\eta+1}^{\widehat{\eta}} + \sum_{t=\widehat{\eta}+1}^{e}\right) \xi_t(\widetilde{g}_t - g_t^*) = III + IV + V.$$

Then $|III| \le 2n^{\varphi} \log T\sqrt{\eta - s + 1}|(\widehat{\eta} - s + 1)^{-1} \sum_{t=s}^{\widehat{\eta}} g_t - (\eta - s + 1)^{-1} \sum_{t=s}^{\eta} g_t|$ with probability tending to one, following (16). $|V|$ is of the same order as $|III|$, and similar arguments lead to $|IV| \le 2n^{\varphi} \log T\sqrt{\widehat{\eta} - \eta + 1}|(\widehat{\eta} - s + 1)^{-1} \sum_{t=s}^{\widehat{\eta}} g_t - (e - \eta)^{-1} \sum_{t=\eta+1}^{e} g_t|$.

As for $\sum_{t=s}^{e} \xi_t(\widehat{g}_t - \widetilde{g}_t)$, we have

$$\sum_{t=s}^{e} \xi_t(\widehat{g}_t - \widetilde{g}_t) = \left(\sum_{t=s}^{\widehat{\eta}} + \sum_{t=\widehat{\eta}+1}^{e}\right) \xi_t(\widehat{g}_t - \widetilde{g}_t) = VI + VII.$$

$|VI|$ and $|VII|$ are of the same order, and with probability converging to one,

$$|VI| = |\sum_{t=s}^{\widehat{\eta}} \xi_t| \cdot \frac{1}{\widehat{\eta} - s + 1} \left|\sum_{t=s}^{\widehat{\eta}} (y_t - g_t)\right| = \frac{1}{\widehat{\eta} - s + 1} \left(\sum_{t=s}^{\widehat{\eta}} \xi_t\right)^2 \le 4n^{2\varphi} \log^2 T.$$

Putting together (23) and the upper bound on $|III|$–$|VII|$, we have the dominance of II over I under the conditions given in the lemma. ☐

8.2. Proofs of Theorems 3.1–3.2

Throughout the section, we assume (A1)–(A3) and (A6), and (A4)–(A5) where applicable. In the problem of detecting (at most) a single change-point, (20)–(21) are met with $s = 1$ and $e = T$ under (A4) with $\bar{\epsilon}_T = (n^{\varphi}/m_1)^2 \widetilde{\delta}_1^2 \log^2 T$. For $(\widehat{\eta}_1, \widehat{m}_1) = \arg\max_{b \in [1,T] \setminus \mathcal{I}_{1,T},\, 1 \le m \le n} \mathcal{D}_m^{\varphi}(\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n)$, let $\{k_1^0, \ldots, k_n^0\}$ denote a permutation of $\{1, 2, \ldots, n\}$ satisfying $|\mathcal{X}_{1,\widehat{\eta}_1,T}^{k_1^0}| \ge |\mathcal{X}_{1,\widehat{\eta}_1,T}^{k_2^0}| \ge \ldots \ge |\mathcal{X}_{1,\widehat{\eta}_1,T}^{k_n^0}|$, and $i_j^0 \in \{-1, 1\}$ satisfy $|\mathcal{X}_{1,\widehat{\eta}_1,T}^{k_j^0}| = i_j^0 \cdot \mathcal{X}_{1,\widehat{\eta}_1,T}^{k_j^0}$ for all $j = 1, \ldots, n$. With $\widehat{m}_1$, $\{k_j^0\}$ and $\{i_j^0\}$ replacing m_r, $\{k_j^r\}$ and $\{i_j^r\}$, respectively, we have an additive model $y_{0,t} = g_{0,t} + \xi_{0,t}$ with its components defined in the same manner as $y_{r,t}$, $g_{r,t}$ and $\xi_{r,t}$ in (12)–(14). Then $g_{0,t}$ is piecewise constant with (at most) one change-point at $t = \eta_1$, and $\{\xi_{0,t}\}$ satisfies (16)–(17) in place of $\{\xi_{r,t}\}$. Note that the DC statistic at $m = \widehat{m}_1$ and $b = \widehat{\eta}_1$ can equivalently be represented using $\{y_{0,t}\}$ as

$$\mathcal{D}_{\widehat{m}_1}^{\varphi}(\{|\mathcal{X}_{1,\widehat{\eta}_1,T}^{(j)}|\}_{j=1}^n) = \mathcal{D}_{\widehat{m}_1}^{\varphi}(\{i_j^0 \cdot \mathcal{X}_{1,\widehat{\eta}_1,T}^{k_j^0}\}_{j=1}^n) = \mathcal{C}_{\widehat{\eta}_1}(\{y_{0,t}\}_{t=1}^T).$$

In the case when there exists no change-point, $g_{0,t}$ is constant and therefore $\mathcal{C}_b(\{g_{0,t}\}_{t=1}^T) = 0$ for all b. As $(1, \widehat{\eta}_1, T) \in \mathcal{I}_2$, we have $\mathcal{D}_{\widehat{m}_1}^{\varphi}(\{|\mathcal{X}_{1,\widehat{\eta}_1,T}^{(j)}|\}_{j=1}^n) = \mathcal{C}_{\widehat{\eta}_1}(\{\xi_{0,t}\}_{t=1}^T) < C' n^{\varphi} \log T$ for some $C' > 2\sqrt{2}$, which proves Theorem 3.1.

To prove Theorem 3.2, we need additional lemmas stated with the generic additive model in (18).

Lemma 8.5. *Assume that there exists a single change-point η_1 in g_t which satisfies (A4) and $|g_{\eta_1+1} - g_{\eta_1}| \geq \delta$. Then for some $C_2 > 0$,*

$$|\mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T)| = \max_{b \in [1,T)} |\mathcal{C}_b(\{g_t\}_{t=1}^T)| = \sqrt{\frac{\eta_1(T-\eta_1)}{T}}\delta \geq C_2 \delta T^\beta. \tag{24}$$

Proof. The first equality (24) is a direct result of Lemma 8.3. The second equality follows from the definition of $|\mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T)|$. □

Lemma 8.6. *Assume that the conditions imposed in Lemma 8.5 are met. Then for some ϵ_T, we have*

$$|\mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T) - \mathcal{C}_{b'}(\{g_t\}_{t=1}^T)| \geq C_3 \epsilon_T |\mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T)| \frac{T}{\eta_1(T-\eta_1)}$$

with any b' satisfying $|b' - \eta_1| \geq c_0 \epsilon_T$.

Proof. Without loss of generality, let $g_{\eta_1} = g_1^* > g_2^* = g_{\eta_1+1}$ and that $b' > \eta_1$. Then

$$\begin{aligned}
&\mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T) - \mathcal{C}_{b'}(\{g_t\}_{t=1}^T) \\
&\geq \sqrt{\frac{\eta_1(T-\eta_1)}{T}}\delta - \sqrt{\frac{(\eta_1 + c_0\epsilon_T)(T - \eta_1 - c_0\epsilon_T)}{T}} \left(\frac{g_1^*\eta_1 + g_2^* c_0 \epsilon_T}{\eta_1 + c_0\epsilon_T} - g_2^* \right) \\
&= \sqrt{\frac{\eta_1(T-\eta_1)}{T}}\delta \left(1 - \sqrt{\frac{\eta_1(T-\eta_1-c_0\epsilon_T)}{(\eta_1+c_0\epsilon_T)(T-\eta_1)}} \right) \\
&= \mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T) \cdot \frac{\sqrt{1 + \frac{c_0\epsilon_T}{\eta_1}} - \sqrt{1 - \frac{c_0\epsilon_T}{T-\eta_1}}}{\sqrt{1+\frac{c_0\epsilon_T}{\eta_1}}} \\
&\geq \mathcal{C}_{\eta_1}(\{g_t\}_{t=1}^T) \cdot \frac{c_0\epsilon_T}{2\sqrt{2}} \frac{T}{\eta_1(T-\eta_1)},
\end{aligned}$$

where the inequality follows from the Taylor expansion. □

Lemma 8.3 implies that $|\mathcal{C}_{\eta_1}(\{g_{0,t}\}_{t=1}^T)| \geq |\mathcal{C}_{\widehat{\eta}_1}(\{g_{0,t}\}_{t=1}^T)|$ while $|\mathcal{C}_{\eta_1}(\{y_{0,t}\}_{t=1}^T)| \leq |\mathcal{C}_{\widehat{\eta}_1}(\{y_{0,t}\}_{t=1}^T)|$ by assumption. From (17), we have $|\mathcal{C}_b(\{g_{r,t}\}_{t=1}^T)| \geq |\mathcal{C}_b(\{y_{r,t}\}_{t=1}^T)| - c_4 n^\varphi \log T$ at any $b \in \mathcal{I}_{1,T}$ and $r = 0, 1$ for some $c_4 > 0$. Note that for all $\varphi \in [0,1]$ and $1 \leq m \leq n$, on a given interval $[s,e]$, the ordering and the set of signs that are applied to $\{\mathcal{X}_{1,b,T}^j\}_{j=1}^n$ in order to produce $\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n$ (identical to $\{k_j^0\}$ and $\{i_j^0\}$ when $b = \widehat{\eta}_1$), leads to the maximum value of $\mathcal{D}_m^\varphi$ at any b, among all possible index permutations and the sets of signs. Based on the above observations, the following holds with probability tending to one:

$$|\mathcal{C}_{\widehat{\eta}_1}(\{y_{0,t}\}_{t=1}^T)| \geq \max_{1 \leq m \leq n} \mathcal{D}_m^\varphi(\{|\mathcal{X}_{1,\eta_1,T}^{(j)}|\}_{j=1}^n) \geq |\mathcal{C}_{\eta_1}(\{y_{1,t}\}_{t=1}^T)|, \text{ and}$$

$$|\mathcal{C}_{\widehat{\eta}_1}(\{g_{0,t}\}_{t=1}^T)| \geq |\mathcal{C}_{\eta_1}(\{g_{1,t}\}_{t=1}^T)| - 2c_4 n^\varphi \log T \geq \frac{c_1 m_1^\varphi \widetilde{\delta}_1}{2} \sqrt{\frac{\eta_1(T-\eta_1)}{T}}, \tag{25}$$

where the last inequality of (25) follows from (15) and Lemma 8.5. Recalling the rates given above Theorem 3.1 for $\pi^{\varphi}_{n,T}$, we have $\mathcal{D}^{\varphi}_{\widehat{m}_1}(\{|\mathcal{X}^{(j)}_{1,\widehat{\eta}_1,T}|\}_{j=1}^n) > \pi^{\varphi}_{n,T}$.

To prove the consistency in the location of the estimated change-point, let $|\widehat{\eta}_1 - \eta_1| = c_0\epsilon_T$ and assume that $\widehat{\eta}_1 > \eta_1$ without loss of generality. We can view the problem of deriving the upper bound on the bias $|\widehat{\eta}_1 - \eta_1|$ for $\widehat{\eta}_1 = \arg\max_{b\in[1,T]\setminus\mathcal{I}_{1,T}} \max_{1\le m\le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{1,b,T}|\}_{j=1}^n)$, as that for

$$\widehat{\eta}_1 = \arg\max_{b\in[1,T]\setminus\mathcal{I}_{1,T}} \mathcal{C}_b(\{y_{0,t}\}_{t=1}^T).$$

Adopting the notations in Lemma 8.6 with $g_{0,t}$ replacing g_t,

$$\begin{aligned}\left|\frac{1}{\widehat{\eta}_1}\sum_{t=1}^{\widehat{\eta}_1} g_{0,t} - \frac{1}{\eta_1}\sum_{t=s}^{\eta_1} g_{0,t}\right| &= \left|\frac{g_1^*\eta_1 + g_2^* c_0\epsilon_T}{\widehat{\eta}_1} - g_1^*\right| \le |g_1^* - g_2^*|\frac{c_0\epsilon_T}{\eta_1},\\ \left|\frac{1}{\widehat{\eta}_1}\sum_{t=1}^{\widehat{\eta}_1} g_{0,t} - \frac{1}{T-\eta_1}\sum_{t=\eta_1+1}^{e} g_{0,t}\right| &= \left|\frac{g_1^*\eta_1 + g_2^* c_0\epsilon_T}{\widehat{\eta}_1} - g_2^*\right| \le |g_1^* - g_2^*|,\end{aligned}$$

and (25) indicates that $|g_1^* - g_2^*| \ge c_5 m_1^{\varphi}\widetilde{\delta}_1$ for some $c_5 > 0$. Combining the above bounds with Lemmas 8.5–8.6, the conditions in Lemma 8.4 are met for $\eta = \eta_1$, $\widehat{\eta} = \widehat{\eta}_1$, $s = 1$, $e = T$ and $\epsilon_T = (n^{\varphi}/m_1)^2\widetilde{\delta}_1^2 \log^2 T$, and thus Theorem 3.2 is proved. □

8.3. Proof of Theorem 3.3

Throughout the section, we assume (A1)–(A3), (A6) and (B1)–(B2). Let s, e satisfy (20)–(21) with $\bar{\epsilon}_T = n^{2\varphi}\underline{\Delta}_{\varphi}^{-2}T^{5(1-\beta)}\log^2 T$. For

$$(\widehat{\eta}, \widehat{m}) = \arg\max_{b\in[s,e]\setminus\mathcal{I}_{s,e},\, 1\le m\le n} \mathcal{D}^{\varphi}_m(\{|\mathcal{X}^{(j)}_{s,b,e}|\}_{j=1}^n),$$

let $\{k_1^0, \ldots, k_n^0\}$ denote a permutation of $\{1, 2, \ldots, n\}$ satisfying $|\mathcal{X}^{k_1^0}_{s,\widehat{\eta},e}| \ge |\mathcal{X}^{k_2^0}_{s,\widehat{\eta},e}| \ge \ldots \ge |\mathcal{X}^{k_n^0}_{s,\widehat{\eta},e}|$, and $i_j^0 \in \{-1, 1\}$ satisfy $|\mathcal{X}^{k_j^0}_{s,\widehat{\eta},e}| = i_j^0 \cdot \mathcal{X}^{k_j^0}_{s,\widehat{\eta},e}$ for all $j = 1, \ldots, n$. As before, with $\widehat{m}$, $\{k_j^0\}$ and $\{i_j^0\}$ replacing m_r, $\{k_j^r\}$ and $\{i_j^r\}$, respectively, we define an additive model $y_{0,t} = g_{0,t} + \xi_{0,t}$ with its components obtained in the same manner as those in (12)–(14). Then $g_{0,t}$ is a piecewise constant signal with change-points at $t = \eta_r$, $r = 1, \ldots, N$, and $\{\xi_{0,t}\}$ satisfies (16)–(17) in place of $\{\xi_{r,t}\}$. Also, the DC statistic at $m = \widehat{m}$ and $b = \widehat{\eta}$ can equivalently be represented by $\mathcal{D}^{\varphi}_{\widehat{m}}(\{|\mathcal{X}^{(j)}_{s,\widehat{\eta},e}|\}_{j=1}^n) = \mathcal{D}^{\varphi}_{\widehat{m}}(\{i_j^0 \cdot \mathcal{X}^{k_j^0}_{s,\widehat{\eta},e}\}_{j=1}^n) = \mathcal{C}_{\widehat{\eta}}(\{y_{0,t}\}_{t=s}^e)$.

Below we introduce additional lemmas stated with the generic additive model in (18).

Lemma 8.7. *Let s and e satisfy (20) and assume that there exists a change-point $\eta_q \in (s, e)$ at which $(\eta_q - s + 1) \wedge (e - \eta_q) > c_2 T^{\beta}$ and $|g_{\eta_q+1} - g_{\eta_q}| \ge \delta$. Then there exists $q' \in \{q_1 + 1, \ldots, q_2\}$ and $C_4 > 0$ such that $|\mathcal{C}_{\eta_{q'}}(\{g_t\}_{t=s}^e)| = \max_{b\in[s,e)} |\mathcal{C}_b(\{g_t\}_{t=s}^e)| \ge C_4\delta T^{\beta-1/2}$.*

Proof. The equality part is a direct result of Lemma 8.3. Since $|g_{\eta_q+1}| \vee |g_{\eta_q}| \geq \delta/2$, we have $|\sum_{t=\eta_q-c_2T^\beta+1}^{\eta_q} g_t| \vee |\sum_{t=\eta_q+1}^{\eta_q+c_2T^\beta} g_t| \geq c_2\delta T^\beta/2$. Hence, $\max_{b\in[s,e)} |\sum_{t=s}^{b} g_t| \geq c_2\delta T^\beta/4$, from which it is derived that

$$\max_{b\in[s,e)} |\mathcal{C}_b(\{g_t\}_{t=s}^e)| \geq \min_{b\in[s,e)} \sqrt{\frac{e-s+1}{(b-s+1)(e-b)}} \cdot \max_{b\in[s,e)} |\sum_{t=s}^{b} g_t| \geq C_4\delta T^{\beta-1/2}.$$

□

Lemma 8.8. *Assume that s and e meet the conditions (20)–(21) and let $\eta \in (s,e)$ denote a change-point that satisfies*

$$|\mathcal{C}_\eta(\{g_t\}_{t=s}^e)| > \max_{b\in[s,e)} |\mathcal{C}_b(\{g_t\}_{t=s}^e)| - C_5 n^\varphi \log T \tag{26}$$

for some positive constant C_5. Then for some ϵ_T and $C_6 > 0$, we have

$$|\mathcal{C}_\eta(\{g_t\}_{t=s}^e) - \mathcal{C}_{b'}(\{g_t\}_{t=s}^e)| > C_6 T^{\beta-2}\epsilon_T |\mathcal{C}_\eta(\{g_t\}_{t=s}^e)| \tag{27}$$

with any b' satisfying $|b'-\eta| \geq \epsilon_T$.

Proof. The result is a modification of Lemma 2.6 in Venkatraman (1992) and the arguments therein are directly applicable to show (27). □

For any interval $[s,e]$, we define an index set $\mathcal{R}_{s,e} \subset \{1,\ldots,N\}$ as $\mathcal{R}_{s,e} = \{1 \leq r \leq N : \eta_r \in [s,e] \setminus \mathcal{I}_{s,e}\}$. Adopting the same arguments as in Section 8.2,

$$\begin{aligned}
|\mathcal{C}_{\widehat{\eta}}(\{y_{0,t}\}_{t=s}^e)| &\geq \max_{q\in\mathcal{R}_{s,e}} \max_{1\leq m\leq n} \mathcal{D}_m^\varphi(\{|\mathcal{X}_{s,\eta_q,e}^{(j)}|\}_{j=1}^n) \geq \max_{q\in\mathcal{R}_{s,e}} \max_{1\leq r\leq N} |\mathcal{C}_{\eta_q}(\{y_{r,t}\}_{t=s}^e)|,\\
|\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^e)| &\geq \max_{q\in\mathcal{R}_{s,e}} \max_{1\leq r\leq N} |\mathcal{C}_{\eta_q}(\{g_{r,t}\}_{t=s}^e)| - 2c_4 n^\varphi \log T \geq \frac{C_4}{2} c_1 \underline{\Delta}_\varphi T^{\beta-1/2},
\end{aligned} \tag{28}$$

where the last inequality of (28) follows from (15), (20) and Lemma 8.7.

Lemma 8.9. *Let (20) and (21) hold. For $\widehat{\eta} = \arg\max_{b\in[s,e]\setminus\mathcal{I}_{s,e}} |\mathcal{C}_b(\{y_{0,t}\}_{t=s}^e)|$, there exists a true change-point $\eta_q \equiv \eta \in (s,e)$ satisfying $|\widehat{\eta}-\eta| < c_0\epsilon_T$ with probability converging to one, where $\epsilon_T = n^{2\varphi}\underline{\Delta}_\varphi^{-2} T^{5(1-\beta)} \log^2 T$.*

Proof. We adopt the notations from the proof of Lemma 8.4 with $g_{0,t}$ in place of g_t.

Recall that from Lemma 8.3, g_t^* needs to have its change-point η coincide with one of the true change-points $\eta_{q_1+1},\ldots,\eta_{q_2}$. Trivially, such η satisfies (26) since $|\mathcal{C}_\eta(\{g_{0,t}\}_{t=s}^e)| = \max_{b\in[s,e)} |\mathcal{C}_b(\{g_{0,t}\}_{t=s}^e)|$. Under (20)–(21), we have either $(\eta-s+1)\wedge(e-\eta) < c_3\epsilon_T$ or $(\eta-s+1)\wedge(e-\eta) > c_2T^\beta$. If the former is the case, since $|g_{0,t}| \leq 2\bar{f}n^\varphi$ uniformly in t under (A3), $|\mathcal{C}_\eta(\{g_{0,t}\}_{t=s}^e)| \leq 4\bar{f}n^\varphi(c_3\epsilon_T)^{1/2} < C_4\underline{\Delta}_\varphi T^{\beta-1/2}$, which leads to contradict (28) and thus $(\eta-s+1)\wedge(e-\eta) > c_2T^\beta$.

Now we turn our attention to bound the terms $|III|$ and $|IV|$ in the presence of multiple change-points. Firstly,
$|(\widehat{\eta}-s+1)^{-1}\sum_{t=s}^{\widehat{\eta}} g_{0,t} - (\eta-s+1)^{-1}\sum_{t=s}^{\eta} g_{0,t}|$

$$
\begin{aligned}
&= \left| \tfrac{1}{\widehat{\eta}-s+1}\sqrt{\tfrac{(\widehat{\eta}-s+1)(e-\widehat{\eta})}{e-s+1}}\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) - \tfrac{1}{\eta-s+1}\sqrt{\tfrac{(\eta-s+1)(e-\eta)}{e-s+1}}\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e}) \right| \\
&= \tfrac{1}{\sqrt{e-s+1}} \left| \sqrt{\tfrac{e-\widehat{\eta}}{\widehat{\eta}-s+1}}\{\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) - \mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})\} \right. \\
&\qquad \left. - \left(\sqrt{\tfrac{e-\eta}{\eta-s+1}} - \sqrt{\tfrac{e-\widehat{\eta}}{\widehat{\eta}-s+1}}\right)\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e}) \right| \\
&\le \sqrt{\tfrac{e-\eta}{(e-s+1)(\eta-s+1)}} \left\{ \left| 1 - \frac{\sqrt{1-\frac{\widehat{\eta}-\eta}{e-\eta}}}{\sqrt{1+\frac{\widehat{\eta}-\eta}{\eta-s+1}}} \right| |\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})| + \right. \\
&\qquad \left. |\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) - \mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})| \right\} \\
&\le \sqrt{\tfrac{e-\eta}{(e-s+1)(\eta-s+1)}} \left\{ C_7 T^{-\beta}\epsilon_T |\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})| + |\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) - \mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})| \right\}
\end{aligned}
$$

for some fixed $C_7 > 0$, where the last inequality follows from the Taylor expansion. Since $T^{\beta-2} \le T^{-\beta}$, Lemma 8.8 leads to

$$|III| \le 2n^{\varphi}\epsilon_T T^{-\beta} \log T |\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})|$$

(recall the notation from Lemma 8.4) with probability tending to one. Similarly,
$|(\widehat{\eta}-s+1)^{-1}\sum_{t=s}^{\widehat{\eta}} g_{0,t} - (e-\eta)^{-1}\sum_{t=\eta+1}^{e} g_{0,t}|$

$$
\begin{aligned}
&= \left| \tfrac{1}{\widehat{\eta}-s+1}\sqrt{\tfrac{(\widehat{\eta}-s+1)(e-\widehat{\eta})}{e-s+1}}\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) + \tfrac{1}{e-\eta}\sqrt{\tfrac{(\eta-s+1)(e-\eta)}{e-s+1}}\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e}) \right| \\
&= \tfrac{1}{\sqrt{e-s+1}} \left| \left(\sqrt{\tfrac{e-\widehat{\eta}}{\widehat{\eta}-s+1}} + \sqrt{\tfrac{\eta-s+1}{e-\eta}}\right)\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e}) \right. \\
&\qquad \left. + \sqrt{\tfrac{e-\widehat{\eta}}{\widehat{\eta}-s+1}}\{\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e}) - \mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})\} \right| \\
&\le 2\sqrt{\tfrac{e-s+1}{(\eta-s+1)(e-\eta)}} |\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})|,
\end{aligned}
$$

and thus $|IV| \le 4n^{\varphi}\sqrt{c_0\epsilon_T}T^{-\beta/2}\log T|\mathcal{C}_{\eta}(\{g_t\}_{t=s}^{e})|$. Plugging in the above bounds to the condition given in Lemma 8.4, we obtain

$$
\begin{aligned}
&|\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})||\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e}) - \mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^{e})| > \\
&C_8 n^{\varphi}\log T \left\{ (\epsilon_T T^{-\beta}|\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})|) \vee (\sqrt{\epsilon_T}T^{-\beta/2}|\mathcal{C}_{\eta}(\{g_{0,t}\}_{t=s}^{e})|) \vee (n^{\varphi}\log T) \right\}
\end{aligned}
$$

for some $C_8 > 0$. Due to (28), the above is met with $\epsilon_T = n^{2\varphi}\underline{\Delta}_{\varphi}^{-2}T^{5(1-\beta)}\log^2 T$. □

With the above lemmas, we are now ready to prove the theorem. At the beginning of the DCBS algorithm, we have $s = 1$ and $e = T$ for which both (20) and (21) hold, and thus η_{r_1} is estimated by $\widehat{\eta}_{r_1}$ within the distance of ϵ_T from a true change-point (Lemma 8.9). Then both of the two segments defined to the left and to the right of $\widehat{\eta}_{r_1}$ satisfy (20)–(21) and so do all the subsequently defined segments, and therefore the same arguments apply to show the consistency of the estimates $\widehat{\eta}_{r_2}, \widehat{\eta}_{r_3}, \ldots, \widehat{\eta}_{r_N}$.

Once all the N change-points are detected, any segment $[s, e]$ determined by two adjacent $\widehat{\eta}_1, \ldots, \widehat{\eta}_N$ (including $\eta_0 = 1$ and $\eta_{N+1} = T$) satisfy either

(i) $\exists\, 1 \le r \le N$ such that $r = q_1 + 1 = q_2$ in (19) and $(\eta_r - s + 1) \wedge (e - \eta_r) \le c_6 \epsilon_T$, or
(ii) $\exists\, 1 \le r \le N - 1$ such that $r = q_1 + 1$ and $r + 1 = q_2$, and $(\eta_r - s + 1) \vee (e - \eta_{r+1}) \le c_6 \epsilon_T$

for some positive constant c_6. Under (i), let

$$(\widehat{\eta}, \widehat{m}) = \arg \max_{b \in [s,e] \setminus \mathcal{I}_{s,e}} \max_{1 \le m \le n} \mathcal{D}_m^{\varphi}(\{|\mathcal{X}_{s,b,e}^{(j)}|\}_{j=1}^n),$$

and we adopt the notations $y_{0,t}$, $g_{0,y}$ and $\xi_{0,t}$ introduced at the beginning of this section with respect to $\widehat{\eta}$ and $\widehat{m}$. Then,

$$\begin{aligned} |\mathcal{C}_{\widehat{\eta}}(\{y_{0,t}\}_{t=s}^e)| &\le |\mathcal{C}_{\widehat{\eta}}(\{g_{0,t}\}_{t=s}^e)| + c_4 n^{\varphi} \log T \le |\mathcal{C}_{\eta_r}(\{g_{0,t}\}_{t=s}^e)| + c_4 n^{\varphi} \log T \\ &\le 4\bar{f} n^{\varphi} (c_6 \epsilon_T)^{1/2} + c_4 n^{\varphi} \log T < \pi_{n,T}^{\varphi}, \end{aligned}$$

from the fact that $|g_{0,t}| \le 2\bar{f} n^{\varphi}$. A similarly conclusion can be drawn in the case of (ii) as well, and thus the proof is completed. $\square$

8.4. Proof of Theorem 4.1

We first prove (a) of Theorem 4.1, which states that $\widetilde{\mathcal{D}}_m$ achieves consistency in identifying and locating the change-point under the single change-point scenario. As in Section 8.2, let $(\widehat{\eta}_1, \widehat{m}_1) = \arg\max_{b \in [1,T] \setminus \mathcal{I}_{1,T}, 1 \le m \le n} \widetilde{\mathcal{D}}_m(\{|\mathcal{X}_{1,b,T}^{(j)}|\}_{j=1}^n)$, and adopt the notations $\{k_1^0, \ldots, k_n^0\}$ and $i_j^0 \in \{-1, 1\}$ therein. We define an additive model $y_{0,t} = g_{0,t} + \xi_{0,t}$, where

$$y_{0,t} = \left\{ \gamma_n + \sqrt{\frac{\widehat{m}_1(2n - \widehat{m}_1)}{2n}} \right\} \left\{ \frac{1}{\widehat{m}_1} \sum_{j=1}^{\widehat{m}_1} i_j^0 \cdot \widetilde{x}_{k_j^0,t} - \frac{1}{2n - \widehat{m}_1} \sum_{j=\widehat{m}_1+1}^{n} i_j^0 \cdot \widetilde{x}_{k_j^0,t} \right\},$$

and $g_{0,t}$ and $\xi_{0,t}$ are constructed analogously with $\widetilde{f}_{k_j^0,t}$ and $\widetilde{\varepsilon}_{k_j^0,t}$ in place of $\widetilde{x}_{k_j^0,t}$, respectively.

Provided that $n^{-1/2}\gamma_n \to 0$ with $n \to \infty$, we have (16)–(17) satisfied with $\{\xi_{0,t}\}$ replacing $\{\xi_{r,t}\}$ with the bound of the order $n^{1/2} \log T$. Also as in (25), it is shown that $|\mathcal{C}_{\widehat{\eta}_1}(\{g_{0,t}\}_{t=1}^T)| \ge C_9(\gamma_n \vee m_1^{1/2})\widetilde{\delta}_1 \sqrt{\frac{\eta_1(T-\eta_1)}{T}}$ for some constant $C_9 > 0$, and thus $\widetilde{\mathcal{D}}_{\widehat{m}_1}(\{|\mathcal{X}_{1,\widehat{\eta}_1,T}^{(j)}|\}_{j=1}^n) > \widetilde{\pi}_{n,T}$ with probability tending to one, which indicates the consistency of the test $\widetilde{\mathcal{T}}_{1,T} > \widetilde{\pi}_{n,T}$. Further, the same arguments adopted in bounding $|\widehat{\eta}_1 - \eta_1|$ for Theorem 3.2 are applicable to show that $\epsilon_T = n\{(\gamma_n \vee m_1^{1/2})\widetilde{\delta}_1\}^{-2} \log^2 T$. Similarly, (28) can be modified for $\widetilde{\mathcal{D}}_m$ and thus the proof of (b) follows verbatim. $\square$

Supplementary Material

Supplement to "Change-point detection in panel data via double CUSUM statistic"
(doi: 10.1214/16-EJS1155SUPP; .pdf). We provide the detailed description of the Local Bootstrap and proof of some auxillary results. In addition, the tables and plots summarising the outcome of the simulation studies conducted in Section 5 are presented.

References

Aston, J. A. and Kirch, C. (2014). Change points in high dimensional settings. *arXiv preprint, arXiv:1409.1771*.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Process: Estimation and Prediction*. Springer. MR1640691

Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-point Problems*. Springer. MR1228205

Cho, H. (2016). Supplement to "Change-point detection in panel data via double CUSUM statistic". *Electronic Journal of Statistics*. DOI: 10.1214/16-EJS1155SUPP

Cho, H. and Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica* **22** 207–229. MR2933173

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B* **77** 475–507. MR3310536

Enikeeva, F. and Harchaoui, Z. (2014). High-dimensional change-point detection with sparse alternatives. *arXiv preprint, arXiv:1312.1900*.

Fan, J., Lv, J. and Qi, L. (2011). Sparse high dimensional models in economics. *Annual Review of Economics* **3** 291–317.

Fiecas, M. and von Sachs, R. (2014). Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electronic Journal of Statistics* **8** 2975–3003. MR3301298

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric Theory* **17** 1113–1141. MR1867540

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The Generalized Dynamic Factor Model: identification and estimation. *The Review of Economics and Statistics* **82** 540–554.

Fryzlewicz, P. (2014). Wild Binary Segmentation for multiple change-point detection. *Annals of Statistics* **42** 2243–2281. MR3269979

Härdle, W., Horowitz, J. and Kreiss, J.-P. (2003). Bootstrap methods for time series. *International Statistical Review* **71** 435–459.

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis* **33** 631–648. MR2944843

Hušková, M. and Kirch, C. (2010). A note on studentized confidence intervals for the change-point. *Computational Statistics* **25** 269–289. MR2639810

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Annals of Statistics* **43** 1117–1140. MR3346699

Jirak, M. (2015). Uniform change point tests in high dimension. *Annals of Statistics* **43** 2451–2483. MR3405600

Kirch, C. and Muhsal, B. (2015). A MOSUM procedure for the estimation of multiple random change points. *Preprint*.

Kirch, C. and Politis, D. N. (2011). TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Annals of Statistics* **39** 1427–1470. MR2850208

Korostelev, A. (1987). On minimax estimation of a discontinuous signal. *Theory of Probability and its Applications* **32** 727–730. MR0927265

Politis, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744. MR2822363

Soh, Y. S. and Chandrasekaran, V. (2015). High-dimensional change-point estimation: combining filtering with convex optimization. *Applied and Computational Harmonic Analysis (to appear)*.

Venkatraman, E. S. (1992). Consistency results in multiple change-point problems. *Technical Report No. 24, Department of Statistics, Stanford University*. MR2687536

Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random processes. *Soviet Doklady Mathematics* **24** 55–59.

Zhang, N. R., Siegmund, D. O., Ji, H. and Li, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488