Detection of multiple change-points in high-dimensional panel data with cross-sectional and temporal dependence*†

Marie-Christine Düker[‡1], Seok-Oh Jeong[§ 2], Taewook Lee[¶2], and Changryong Baek[‖3]

[1]Cornell University
[2]Hankuk University of Foreign Studies
[3]Sungkyunkwan University

May 20, 2022

Abstract

We consider the detection of multiple change-points in a high-dimensional time series exhibiting both temporal and cross-sectional dependence. Several test statistics based on the celebrated CUSUM statistic are used and discussed. In particular, we propose a novel block wild bootstrap method to address the presence of cross-sectional and temporal dependence. Furthermore, binary segmentation and the moving sum algorithm are considered to detect and locate multiple change-points. We also provide some theoretical justifications for the moving sum method. An extensive numerical study provides insights on the performance of the proposed methods. Finally, our proposed procedures are used to analyze financial stock data in Korea.

*AMS subject classification. Primary: 62M10, 62G10 Secondary: 62H15.

†Keywords and phrases: change point analysis, high-dimensional time series, block wild bootstrap, CUSUM, binary segmentation, moving sum algorithm

‡Department of Statistics and Data Science, Cornell University, Ithaca, NY 14850, U.S.A, md2224@cornell.edu

§Department of Statistics, Hankuk University of Foreign Studies, 81 Oedae-ro, Mohyeon-myeon, Cheoin-gu, Yongin-si, Kyunggi-do, Korea 449-791, seokohj@hufs.ac.kr

¶Dept of Statistics, Hankuk University of Foreign Studies, 81 Oedae-ro, Mohyeon-myeon, Cheoin-gu, Yongin-si, Kyunggi-do, Korea 449-791, twlee@hufs.ac.kr

‖Corresponding author, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, Korea 03063, crbaek@skku.edu

1 Introduction

Let $\{X_t := (X_{1,t}, \ldots, X_{p,t})'\}_{t\in\mathbb{Z}}$ be a p-variate time series. We consider a simple mean plus noise model

$$X_t = \mu + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1.1}$$

where $\mu = (\mu_1, \ldots, \mu_p)'$ is a p-variate (deterministic) mean vector and $\{\varepsilon_t := (\varepsilon_{1,t}, \ldots, \varepsilon_{p,t})'\}_{t\in\mathbb{Z}}$ is a zero-mean process. The spatial (cross-sectional) and temporal dependence structures will be specified later. The problem of interest is to test whether there is a change in mean at an unknown time point τ, while possibly allowing for $p \gg T$, and locate the change-point if it exists. This can be formally formulated as follows. Let $\mu_L(\tau)$ and $\mu_R(\tau)$ denote the mean vectors of X_t for $t \in L(\tau) = \{1, \ldots, \tau\}$ and for $t \in R(\tau) = \{\tau + 1, \ldots, T\}$, respectively. Then, the null hypothesis of no change can be written as

$$H_0 : \mu_{L(\tau)} = \mu_{R(\tau)} \text{ for every } \tau = 1, \ldots, T-1, \tag{1.2}$$

and the alternative hypothesis of a single change-point is given as

$$H_1 : \mu_{L(\tau)} \neq \mu_{R(\tau)} \text{ for some unknown } \tau \in \{1, \ldots, T-1\}.$$

In the univariate setting ($p = 1$), the state-of-the-art method is to use the celebrated CUSUM statistic given by

$$|\chi_1(k)| = \frac{1}{\sqrt{T}} \left| \sum_{t=1}^{k} X_{1,t} - \frac{k}{T} \sum_{t=1}^{T} X_{1,t} \right|. \tag{1.3}$$

The corresponding estimator for the location of the change-point is then given by

$$\widehat{\tau} = \underset{1\le k\le T}{\operatorname{argmax}} |\chi(k)|$$

and the hypothesis testing is based upon the functional central limit theorem

$$\max_{1\le k\le T} \frac{|\chi_1(k)|}{\sigma} \xrightarrow{d} \sup_{0\le u\le 1} |\mathbb{B}^0(u)|, \tag{1.4}$$

where $\mathbb{B}^0(u)$ is a standard Brownian bridge and σ^2 denotes the long-run variance of $\{X_{1,t}\}_{t\in\mathbb{Z}}$. For later use, we will denote $|\chi_i(k)|$ and σ_i as CUSUM and long-run variance of the ith component (dimension) series $\{X_{i,t}\}_{t\in\mathbb{Z}}$.

In the context of high-dimensional time series (HDTS), Bai (2010) proposed a least squares based estimator (LSE) to detect a change-point

$$\widehat{\tau} := \widehat{\tau}(p, T) = \underset{1\le k\le T-1}{\operatorname{argmin}} \sum_{i=1}^{p} \left\{ \sum_{t=1}^{k} \left(X_{i,t} - \overline{X}_{i,L(k)}\right)^2 + \sum_{t=k+1}^{T} \left(X_{i,t} - \overline{X}_{i,R(k)}\right)^2 \right\}, \tag{1.5}$$

where

$$\overline{X}_{i,L(k)} = \frac{1}{k} \sum_{t\in L(k)} X_{i,t}, \quad \overline{X}_{i,R(k)} = \frac{1}{T-k} \sum_{t\in R(k)} X_{i,t}.$$

Bai (2010) showed that the LSE is consistent for the true change-point location τ_0 in the sense that

$$\lim_{p,T\to\infty} \mathrm{P}(\widehat{\tau} = \tau_0) = 1$$

for independent panel data with temporal dependence. Bhattacharjee et al. (2019) extended the consistency result to models with quite general assumptions on the cross-sectional and temporal dependence structure. However, both Bai (2010) and Bhattacharjee et al. (2019) focused on the estimation of a single change-point, without suggesting any formal testing procedure to detect multiple change-points.

Based on the CUSUM statistic $\chi_i(k)$ calculated for each component series as in (1.3), Horváth and Hušková (2012) derived a test from a likelihood argument using the centered sum of CUSUM statistics

$$V(k) = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{|\chi_i(k)|^2}{\sigma_i^2} - \frac{k(T-k)}{T^2}\right) \tag{1.6}$$

for independent panels but with temporal dependence in each panel. Horváth and Hušková (2012) established the convergence of the test statistic (1.6)

$$\max_{1\le k\le T}|V(k)| \xrightarrow{d} \sup_{0\le x\le 1}|\Gamma(x)|, \tag{1.7}$$

where $\Gamma(x)$ is a Gaussian process with zero mean and $\mathbb{E}\Gamma(x)\Gamma(y) = 2x^2(1-y^2)$, $0 \le x \le y \le 1$. However, as discussed in Section 3 of Horváth and Hušková (2012), the convergence is easily affected by even small spatial dependence.

Another approach takes the maximum over the dimension instead of summation. Jirak (2015) proposed the following test statistic,

$$J = \max_{1\le i\le p}\max_{1\le k\le T}\frac{|\chi_i(k)|}{\sigma_i}$$

and showed that under quite flexible temporal and spatial assumptions on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ in (1.1), J converges to a Gumbel distribution, that is,

$$\lim_{p,T\to\infty}\mathrm{P}\Big(2\sqrt{2\log(2p)}\Big(J - .5\sqrt{2\log(2p)}\Big) < x\Big) = 1 - e^{-x}. \tag{1.8}$$

The asymptotic result in (1.8) relies on statements known from extreme value theory. However, it is widely reported that the convergence to extreme value distributions is relatively slow. Thus, Jirak (2015) proposed to apply a block multiplier bootstrap (BWB) to calculate p-values for better finite sample performance. Furthermore, the convergence result in (1.4) allows a binary segmentation (BS) algorithm to find multiple change-points. The BS algorithm is a recursive way of finding multiple change-points by splitting the data into two sub-samples, namely before and after the break, and test for a possible change-point in each sub-sample. The relation (1.7) can also be used to detect multiple change-points, although it is not explicitly discussed in Hušková and Slabỳ (2001).

A more sophisticated approach is the so-called double CUSUM (DC) method proposed by Cho (2016). The test statistic is given by

$$D = \max_{1\le k<T}\max_{1\le m\le p}\left(\frac{m(2p-m)}{2p}\right)^{\varphi}\left(\frac{1}{m}\sum_{j=1}^{m}|\widetilde{\chi}_{(j)}(k)| - \frac{1}{2p-m}\sum_{j=m+1}^{p}|\widetilde{\chi}_{(j)}(k)|\right), \tag{1.9}$$

where $|\widetilde{\chi}_{(j)}(k)|$ is the j-th largest ordered adjusted CUSUM statistics given by

$$|\widetilde{\chi}_i(k)| = \sqrt{\frac{T^2}{k(T-k)}}|\chi_i(k)|. \tag{1.10}$$

DC uses cross-sectional change-point structure through the cumulative sums of ordered CUSUMs at each point. Cho (2016) also uses a BS algorithm and proposed a Generalized Dynamic Factor Model (GDFM) bootstrap to approximate critical values to identify change-points.

In this paper, we consider the detection of multiple change-points in high-dimensional time series by extending above mentioned previous works. Our contributions are multi-folded. First, we modify the block wild bootstrap to accommodate the cross-sectional and temporal dependence in calculating critical values based on the profound theoretical results in Chernozhukov et al. (2013, 2017). Second, more importantly, we introduce a moving sum (MOSUM) algorithm to detect and locate multiple change-points all at once in a high-dimensional setting. This, in turn, avoids inevitable size distortions of the BS algorithm due to sequential multiple testing while attaining better power owing to localization. Third, we compare various methods through extensive Monte Carlo simulations to evaluate finite sample performances. Finally, we provide some theoretical justifications for the MOSUM method.

The rest of the paper is organized as follows. Our proposed change-points detection methods are detailed in Section 2. We examine finite sample performance of the proposed methods through Monte Carlo simulations in Section 3. We also illustrate our methods to detect change-points in an application to real data in Section 4. We conclude with Section 5. Theoretical results for MOSUM statistics and their proofs are presented in Appendix A.

2 Detection of multiple change-points in HDTS

In this section, we consider three test statistics and two algorithms to detect multiple change-points in HDTS. All three test statistics are based on the adjusted CUSUM statistic (1.10) applied to each individual dimension. We choose to base our test statistics on the adjusted CUSUM statistic because of its well peakedness and superior performance near boundaries. Then, either maximum or summation operation is carried to summarize information across dimensions. The first test statistic is given by

$$\widetilde{T}^2(k) = \max_{1\le i\le p}\left(\frac{|\overline{X}_{i,L(k)} - \overline{X}_{i,R(k)}|}{\widehat{\sigma}_i\sqrt{1/|L(k)| + 1/|R(k)|}}\right)^2 = \max_{1\le i\le p}\frac{|\widetilde{\chi}_i(k)|^2}{\widehat{\sigma}_i^2}, \tag{2.1}$$

which is the maximum of squared t-statistics across dimensions. Hence, this test statistic measures the maximum discrepancy on the equivalence of the two subsample means divided by the time point k. The second test statistic is a variation of the one considered in Jirak (2015). It is obtained by replacing the CUSUM statistic with the adjusted CUSUM statistic (1.10), that is,

$$\widetilde{J}(k) = \max_{1\le i\le p}\frac{|\widetilde{\chi}_i(k)|}{\widehat{\sigma}_i}.$$

This is similar to the $\widetilde{T}^2(k)$-statistic in (2.1), but the L_1 distance is used to measure sample mean differences rather than the squared distance. The last test statistic is a variation of Horváth and

Hušková (2012). It averages the adjusted CUSUM statistic and can be written as

$$\widetilde{V}(k) = \frac{1}{p}\sum_{i=1}^{p} \frac{|\widetilde{\chi}_i(k)|^2}{\widehat{\sigma}_i^2}.$$

All test statistics are scaled by the long-run variance estimator $\widehat{\sigma}_i^2$ to account for temporal dependence. It plays a central role in finite sample performance. Here, we use the so-called Bartlett long-run variance estimator given by

$$\widehat{\sigma}_i^2 = \sum_{j=-q_i}^{q_i} \left(1 - \frac{|j|}{2q_i+1}\right)\widehat{\gamma}_i(j), \tag{2.2}$$

where $q_i = q_i(T)$ denotes the bandwidth parameter and $\widehat{\gamma}_i(h)$ the sample autocovariances calculated from the mean adjusted series

$$Y_{i,t} = X_{i,t} - \overline{X}_{i,L(k)}1_{\{t\le k\}} - \overline{X}_{i,R(k)}1_{\{t>k\}} \tag{2.3}$$

in order to remove possible bias due to changes in mean.

Multiple change-points are found by using two algorithms. The first algorithm is based on the BS method, which essentially applies the single change-point detection algorithm successfully until there are no further change-points. The pseudocode for the BS algorithm can be described as follows:

```
Algorithm BinSeg(data, α, m_T) is
 if T = length(data) < m_T
     STOP searching
 else obtain (τ̂, pval) = SingleBreak(data)
     if pval < α
          add τ̂ to the set of estimated change-points
          split the data into two subsamples, before/after τ̂, then apply
          BinSeg(data[1:τ̂]) and BinSeg(data[τ̂+1:T])
     else STOP searching
return change-points
```

Hence, the key procedure is to obtain a candidate change-point $\widehat{\tau}$ and a p-value by applying a single breakpoint detection algorithm. The `SingleBreak` algorithm is detailed in the below.

Step 1 For a given sample $\{X_t, t = 1, \ldots, T\}$, find change-point candidate $\widehat{\tau}$ from the least-squares method (1.5).

Step 2 For a given change-point candidate $\widehat{\tau}$, calculate the test statistics

$$\widetilde{T}^2(\widehat{\tau}), \quad \widetilde{J}(\widehat{\tau}), \quad \widetilde{V}(\widehat{\tau}). \tag{2.4}$$

Step 3 Calculate the p-value by BWB.

Step 3.1 Generate bootstrap sample first by demeaning X_t to get the residual vector time series

$$e_{i,t} = X_{i,t} - \frac{1}{T}\sum_{t=1}^{T} X_{i,t}, \quad t = 1, \ldots, T,$$

and multiply i.i.d. zero mean unit variance random variables $w_{h,\ell}$ independent of $\{e_t := (e_{1,t}, \ldots, e_{p,t})'\}$

$$X^*_{i,t} = w_{h,\ell} e_{i,t}$$

if the t-th observations fall into the ℓ-th block of length b_T, and i-th dimension falls into the h-th block of length c_T, that is, $\ell = \lceil t/b_T \rceil$ and $h = \lceil i/c_T \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function.

Step 3.2 Calculate test statistics (2.4) from the bootstrap sample $\{X^*_t\}$ denoted as $\widetilde{T}^2_{*,b}$, $\widetilde{J}_{*,b}$ and $\widetilde{V}_{*,b}$.

Step 3.3 Repeat Step 3.1-3.2 over B bootstrap replications, calculate the empirical p-value by counting the number of observations greater than the test statistic from the sample. For instance, the p-value for the T^2 statistic is computed as

$$\frac{1}{B} \sum_{b=1}^{B} 1_{\{\widetilde{T}^2(\widehat{\tau}) < T^2_{*,b}\}}.$$

Step 4 If p-value is less than a significance level α, then reject the null hypothesis of no change-point and locate $\widehat{\tau}$ as the change-point.

Remark 1. The basic idea of BWB used here is to apply the same multiplier $w_{h,\ell}$ to the block of observations of dimension $c_T \times b_T$. Because the multiplier has zero mean and unit variance, the bootstrap sample has zero mean and the same autocovariance matrix as the original time series if it exists. The popular choices of multipliers are i.i.d. $\mathcal{N}(0,1)$ random variables due to Chernozhukov et al. (2013, 2017). Also, the Rademacher distribution $w_j = \pm 1$ with probability $1/2$ is widely used in the literature, e.g. Lee and Baek (2020).

Remark 2. We use the change-point estimator (1.5) due to its good theoretical properties provided in Bai (2010) and Bhattacharjee et al. (2019). A more natural change-point estimator would be based on the test statistic itself, for example,

$$\widetilde{\tau} := \underset{1 \le k \le T}{\operatorname{argmax}} \widetilde{T}^2(k) = \underset{1 \le k \le T}{\operatorname{argmax}} \max_{1 \le i \le p} \frac{|\widetilde{\chi}_i(k)|^2}{\widehat{\sigma}_i^2}. \tag{2.5}$$

However, this change-point estimator requires a consistent estimation of the long-run variance σ_i^2 under possible multiple changes-points. This is not feasible in practice without estimating change-points *a priori*. In this case, a two-stage approach should be used. First, we estimate the long-run variance from the demeaned series obtained by applying the BS algorithm based on the LSE change-point estimator. Then, we reapply the BS algorithm and obtain the change-point estimator $\widehat{\tau}$ as in (2.5). Cho (2016) uses such an algorithm which is available in the package `hdbinseg` (Cho and Fryzlewicz, 2018).

The BS method is straightforward to implement and is computationally fast, even for HDTS. This is because the BS algorithm finds change-points sequentially and depends on the previous partition. In other words, due to its tree structure, the BS algorithm never revisits a previous stage. This implies that once a change-point is selected, it remains to be selected for the rest of the procedure. It increases the false detection rate. More critically, BS algorithm may offset multiple change-points with each other when jumps have the same magnitude but point in opposite

directions. Fryzlewicz (2014) proposed the so-called wild binary segmentation (WBS) to address these shortcomings by localization. The other algorithm is called the moving sum (MOSUM) method, which scans change-points by rolling a fixed window through the series. MOSUM was first proposed by Hušková and Slabỳ (2001) for i.i.d. univariate data and studied further in Eichinger and Kirch (2018) for dependent errors. Here, we further extend this algorithm to the HDTS context.

Let G be the window size. Note that the adjusted CUSUM statistic evaluated at the midpoint of the $2G$-windowed sample $\{X_{k-G+1}, \ldots, X_k, X_{k+1}, \ldots, X_{k+G}\}$ is given by

$$\widetilde{\chi}_i(k, G) = \frac{1}{\sqrt{2G}}\Big(\sum_{t=k+1}^{k+G} X_{i,t} - \sum_{t=k-G+1}^{k} X_{i,t} \Big). \tag{2.6}$$

Then, for $G \leq k \leq T - G$, we calculate test statistics as follows

$$\widetilde{T}^2(k, G) = \max_{1 \leq i \leq p} \frac{|\widetilde{\chi}_i(k, G)|^2}{\widehat{\sigma}_i^2}, \;\; \widetilde{J}(k, G) = \max_{1 \leq i \leq p} \frac{|\widetilde{\chi}_i(k, G)|}{\widehat{\sigma}_i}, \;\; \widetilde{V}(k, G) = \frac{1}{p} \sum_{i=1}^{p} \frac{|\widetilde{\chi}_i(k, G)|^2}{\widehat{\sigma}_i^2}. \tag{2.7}$$

That is, we evaluate the test statistic at the midpoint of the subsample of length $2G$. The MOSUM-based testing procedures are described below.

Step 1 From $G \leq k \leq T - G$, calculate test statistics (either $\widetilde{T}^2(k, G)$ or $\widetilde{J}(k, G)$ or $\widetilde{V}(k, G)$) from the $2G$-windowed sample $\{X_{k-G+1}, \ldots, X_k, X_{k+1}, \ldots, X_{k+G}\}$.

Step 2 Apply BWB for each $2G$-windowed sample to calculate a threshold ζ_T.

Step 2.1 For each $2G$-windowed sample $\{X_{k-G+1}, \ldots, X_k, X_{k+1}, \ldots, X_{k+G}\}$ in Step 1, get the residual vector time series

$$e_{i,t} = X_{i,t} - \frac{1}{T} \sum_{t=1}^{T} X_{i,t}, \quad t = k - G + 1, \ldots, k + G,$$

and multiply i.i.d. zero mean unit variance random variables $w_{h,\ell}$ independent of $\{e_t, t = k - G + 1, \ldots, k + G\}$

$$X^*_{i,t} = w_{h,\ell} e_{i,t}, \quad t = k - G + 1, \ldots, k + G,$$

if the t-th observation falls into the ℓ-th block of length b_T, and the i-th dimension belongs to the h-th block of length c_T, that is $k = \lceil (t-k+G-1)/b_T \rceil$ and $h = \lceil i/c_T \rceil$.

Step 2.2 Calculate test statistics from the bootstrap sample $\{X^*_{k-G+1}, \ldots, X^*_k, X^*_{k+1}, \ldots, X^*_{k+G}\}$ denoted as $\widetilde{T}^2_*(k, G)$, $\widetilde{J}_*(k, G)$ and $\widetilde{V}_*(k, G)$.

Step 2.3 Repeat Step 2.1-2.2 b times for each k. The empirical threshold ζ_T is the $100(1-\alpha)$-percentile of $b(T - 2G + 1)$ bootstrap test statistics.

Step 3 Find all pairs of indices (s_j, e_j) such that i) test statistic is above the threshold ζ_T within interval (s_j, e_j), but ii) below threshold for $s_j - 1$ and $e_j + 1$ while keeping the length of the interval to be ηG.

Step 4 The change-point is given by the point that maximizes the test statistic within the interval (s_j, e_j), for example,

$$\operatorname*{argmax}_{k \in (s_j, e_j)} \widetilde{T}^2(k, G) = \operatorname*{argmax}_{k \in (s_j, e_j)} \max_{1 \leq i \leq p} \frac{|\widetilde{\chi}_i(k, G)|^2}{\widehat{\sigma}_i^2}. \tag{2.8}$$

Remark 3. The other disadvantage of BS algorithm is that it may suffer from size distortions due to multiple testing. A naive approach is to apply Bonferroni's correction to correct the size. Alternatively, post model selection as in Lavielle (2005) or false discovery rate approach as in Li et al. (2016) could be considered.

Remark 4. The proposed change-point estimator (2.8) suffers less from a consistent estimation of the long-run variance since it is localized. Once the window size G is suitably chosen so that at most one change-point exits within the window, the proposed long-run variance estimator (2.2) provides a good approximation of the true long-run variance.

3 Finite sample performances

In this section we evaluate the finite sample performance of the proposed methods. Three dependent errors are considered.

(E1) AR(1) model with only temporal correlation:

$$e_{i,t} = \phi e_{i,t-1} + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}(0, .2^2(1-\phi^2)))$$

(E2) ARMA(2, 1) model with cross-sectional correlation:

$$u_{i,t} = \sum_{0 \le j < 100} \frac{\rho}{i+1} v_{i-j,t}, \quad v_{j,t} \sim \mathcal{N}(0, \sigma_v^2), \tag{3.1}$$

$$e_{i,t} = .2e_{i,t-1} - .3e_{i,t-2} + u_{i,t} + .2u_{i,t-1}$$

with $\sigma_v = .2$.

(E3) Factor model:

$$e_{i,t} = \rho_h h_t + .2e_{i,t-1} - .3e_{i,t-2} + u_{i,t} + .2u_{i,t-1}, \quad h_t \sim \mathcal{N}(0, .1^2),$$

where $u_{i,t}$ is generated by (3.1) with $\sigma_v^2 = .2^2(1-\rho_h^2)$.

(E1) is considered to assess the performance of the proposed method under no spatial dependence. Models (E2) and (E3) are adapted from Cho (2016). We use E1(ϕ), E2(ρ) and E3(ρ_h) to indicate different choices of model parameters if necessary. To simulate changes in mean at locations τ_r, we consider a piecewise constant function

$$\mu_{i,t} = \Delta_{i,r} 1_{\{\tau_r \le t \le T\}}.$$

For a given jump size δ_r at the r-th location, the jump size in the i-th dimenison is randomly selected from

$$|\Delta_{i,r}| \sim i.i.d.\ \text{Unif}(.75\delta_r, 1.25\delta_r).$$

The number of dimensions subject to change is denoted by m_r. Once m_r is determined, the corresponding dimensions are uniformly selected from $\{1, \ldots, p\}$.

Tuning parameters are selected as follows. The bandwidth parameter q_i for estimating the long-run variance in (2.2) is selected based on the method suggested in Andrews (1991). More precisely, it is given by

$$\hat{q}_i = [1.147\big(4T\widehat{\rho}_i^2/(1-\widehat{\rho}_i^2)^2\big)^{1/3}] \tag{3.2}$$

with $\widehat{\rho}_i$ being the AR(1) coefficient calculated from the ordinary least squares (OLS) estimator by regressing $Y_{i,t}$ on $Y_{i,t-1}$, $t = 2, \ldots, T$. The block size for BWB in `SingleBreak` is also selected by using q_i as

$$\widehat{b}_T = [\underset{1\le i\le p}{\text{median}}(\widehat{q}_i)].$$

Similarly, we also use a data-adaptive method to determine the block size that accommodates spatial dependence

$$\widehat{c}_T = [\underset{1\le t\le T}{\text{median}}(\widehat{r}_t)], \tag{3.3}$$

where $r_t = [1.147\big(4T\widehat{\phi}_t^2/(1-\widehat{\phi}_t^2)^2\big)^{1/3}]$ with $\widehat{\phi}_t$ being the OLS estimator by regressing $Y_{j,t}$ on $Y_{j-1,t}$, $j = 2, \ldots, p$.

The window size parameter G in MOSUM also plays a key role in finite sample performance. It should be long enough to detect changes with the CUSUM statistic while short enough to catch immediate changes. We also used data-adaptive selection by modifying Andrews (1991)

$$\widehat{G} = \min([2\max_{1\le i\le p}(\widehat{q}_i^*)], [.2T]),$$

where $\widehat{q}_i^*$ is obtained by replacing $\widehat{\rho}_i$ by $\widehat{\rho}_i^*$ in (3.2), where $\widehat{\rho}_i^*$ is the OLS estimator of $X_{i,t}$ regressed on $X_{i,t-1}$, $t = 2, \ldots, T$. The block length of the MOSUM statistic is chosen as $\widehat{b}_T = [\sqrt{2\widehat{G}}]$ and the block length for spatial dependence c_T is estimated as in (3.3). The number of bootstrap replications b in MOSUM is determined by setting the total number of bootstrap samples $b(T-2G+1) = 500$. Finally, we set $\eta G = [1.5\log(T)]$.

We considered 7 tests. The first three tests are denoted as MS-max2, MS-max and MS-sum referring to the MOSUM method with $\widetilde{T}^2(k,G)$, $\widetilde{J}(k,G)$ and $\widetilde{V}(k,G)$, respectively. Similarly, BS-max2, BS-max and BS-sum represent the tests based on BS method. We also considered the double CUSUM statistics with the proposed wild block bootstrap and BS, denoted as BS-DC. DC-GDFM is the proposed method of Cho (2016) from the package `hdbinseg` with defaults suggested in the package, for example, $\varphi = 1/2$. Table 1 summarizes the list of tests considered in the simulation. All results are based on 500 replications.

Table 2 shows the empirical sizes and powers when the true change-point occurs at $\tau = 40$ with $T = 100$, dimension $p = 100$ and errors follow E2(.8). The parameter m_1 is the number of dimensions subject to the change. A smaller m_1 mimics sparse signals, whereas larger values of m_1 represent dense signals. Since δ_r is the magnitude of the jump, the sizes of tests correspond to $\delta_1 = 0$, and a positive δ_1 represents powers of the tests. No serious size distortions are observed in any of the tests considered. Indeed, the test seems to be conservative, except MS-sum (and slightly BS-sum). The power of the tests increases as the jump size δ_1 increases. Observe that the power is quite low for sparse signal $m_1 = 4$ when the jump size is small as $\delta_1 = .2$. Also, the blessing of dimensionality is observed because the power increases as the number of dimensions subject to change increases. It is rather difficult to observe noticeable differences between BS and MOSUM

Test	Description
MS-max2	MOSUM based on test statistic $\widetilde{T}^2(k, G)$ in (2.7).
MS-max	MOSUM based on test statistic $\widetilde{J}(k, G)$ in (2.7).
MS-sum	MOSUM based on test statistic $\widetilde{V}(k, G)$ in (2.7).
BS-max2	BS based on test statistic $\widetilde{T}^2$ in (2.4).
BS-max	BS based on test statistic $\widetilde{J}$ in (2.4).
BS-sum	BS based on test statistic $\widetilde{V}$ in (2.4).
BS-DC	BS based on a double CUSUM statistic (1.9).
DC-GDFM	Double CUSUM method of Cho (2016) with GDFM bootstrap.

Table 1: List of tests considered in Section 3

Method	m_1	$\delta_1 = 0$			$\delta_1 = .2$				$\delta_1 = .4$			
		$\widehat{R}$			$\widehat{R}$				$\widehat{R}$			
	$(p = 100)$	0	1	≥ 2	0	1	2	≥ 3	0	1	2	≥ 3
MS-max2	4	497	3		489	11			25	472	3	
MS-max		497	3		489	11			25	472	3	
MS-sum		479	20	1	441	51	8		87	369	44	
BS-max2		500			459	41			29	448	23	
BS-max		500			456	44			27	452	21	
BS-sum		498	2		463	33	1	3	92	350	37	21
BS-DC		499	1		491	9			32	449	18	1
DC-GDFM		499	1		482	16	2		7	488	5	
MS-max2	25	498	2		268	232				499	1	
MS-max		498	2		268	232				499	1	
MS-sum		466	32	2	53	412	31	4		500		
BS-max2		500			251	249				500		
BS-max		500			257	243				500		
BS-sum		491	9			471	25	4		484	13	3
BS-DC		499	1			489	10	1		499	1	
DC-GDFM		495	5			495	5			496	4	
MS-max2	75	499	1		56	443	1			500		
MS-max		499	1		56	443	1			500		
MS-sum		474	24	1		499	1			500		
BS-max2		500			142	358				500		
BS-max		500			150	350				500		
BS-sum		491	7	2		493	7			500		
BS-DC		497	3			496	4			500		
DC-GDFM		498	1	1		497	3			496	4	

Table 2: Empirical sizes/powers of the proposed change-point detection methods. The true change-point $\tau = 40$ with $T = 100$, $p = 100$ and E2(.8) errors are used.

algorithms. However, the max-type tests showed less power than the sum-type tests when m_1 is 25 or 75. Conversely, for sparse signal with a moderate jump size $\delta_1 = .4$, max-type tests showed higher power. This is consistent with previous studies reporting that max-type tests perform better than sum-type tests when the model is sparse and vice versa. See for example, Chang et al. (2017) and Baek et al. (2021). Table 2 shows that all our proposed methods successfully detect a single change-point while keeping the nominal size.

More interesting simulations are presented in Tables 3–4 where the two change-points are located at $\tau_1 = 50$ and $\tau_2 = 100$. The sample size $T = 150$ is fixed, while the dimension is smaller than the sample size $p = 20 < T$, comparable to $p = 150 = T$, and high-dimensional low sample size case $p = 200 > T$. The mean squared error (MSE) of the estimated mean function to the true mean function and adjusted Rand Index (adjRand) are reported to measure the similarity between estimated and true mean functions.

Table 3 shows the empirical frequency table on the number of change-points detected for dense models. Dense models are generated by setting $\delta_1 = .3$ and $\delta_2 = .2$ and the cardinality of dimension change is given by $m_1 = [.75p]$, $m_2 = [.25p]$. First, observe that MOSUM based tests find the true number of change-points quite well in all cases considered. The margin is small but sum-type tests such as MS-sum has better detection rate than max-type tests. The performance of BS methods, on the other hand, seems to depends on models. For example, BS-max2 or BS-max often fail to find change-points with E1(.5) errors although it improves as dimensions grow. BS-sum generally performs well but it slightly overestimates change-points in smaller dimension. Regarding DC-based methods, they perform quite comparable to MOSUM methods. BS-DC performs slightly better than DC-GDFM in our setting. The only minor imperfection of DC-GDFM is that it overestimates change-points when the dimension is as small as $p = 20$. Nevertheless, DC-GDFM may improve performance by selecting tuning parameters more carefully. For dense models, MOSUM-based methods performed best in detecting correct number of change-points.

By contrast, in terms of MSE and adjRand, the results are reversed in some cases. For example, when E3(.8) is used with dimension $p = 200$, both MS-sum and BS-sum find the correct number of change-points. However, the MSE for MS-sum is .211 and BS-sum is .059, so BS-sum has smaller MSE than MS-sum, and observed other cases as well. This is because the change-point estimators of MOSUM-based methods are slightly biased. For instance, the average location of the first change-point is 50.08 for MS-sum while it is 50.00 for BS-sum. This could be a possible drawback of MOSUM methods.

Table 4 reports results on sparse models. These models are generated by setting $\delta_1 = .3$ and $\delta_2 = .4$ with $m_1 = [2 \log p]$ and $m_2 = [\log p]$. Observe that MOSUM methods perform generally well in all cases considered. In particular, they have the highest detection rate compared to other methods. Specifically, MS-max or MS-max2 performed slightly better as dimension increases. DC-GDFM performed comparable to MOSUM based method for E2(.8) and E3(.8), but it tends to underestimate change-points for independent panels. It may due to the tuning parameter selection in DC-GDFM method. BS methods, on the other hand, showed clear differences between methods. First, BS-sum generally tends to overestimate change-points while BS-max2 or BS-max are more conservative in finding change-points. This may be related to the bias in the long-run variance estimation when multiple change-points exists. In other words, MOSUM-based methods reduce the bias in long-run variance estimation by localization.

All in all, our proposed block wild bootstrap methods that consider both cross-sectional and temporal dependence work well in practice. MOSUM-based tests performed generally well and were particularly useful in identifying the correct number of change-points. However, the location of change-points is determined by the local maximum, so it has a small bias compared to the estimator of Bai (2010). In addition, max-type tests performed better for sparse models, whereas sum-type tests outperformed for dense models as reported in the previous studies.

Method	Errors	$p = 20$							$p = 150$							$p = 200$						
		$\hat{R}$							$\hat{R}$							$\hat{R}$						
		0	1	2	3	≥ 4	MSE	adjRand	0	1	2	3	≥ 4	MSE	adjRand	0	1	2	3	≥ 4	MSE	adjRand
MS-max2	E1(.5)		7	471	22		0.558	0.068			487	13		0.495	0.009			497	3		0.485	0.007
MS-max			7	471	22		0.558	0.068			487	13		0.495	0.009			497	3		0.485	0.007
MS-sum				500			0.431	0.070			500			0.426	0.010			500			0.431	0.007
BS-max2		189	3	308			1.199	0.064	116		384			0.846	0.009	95		405			0.747	0.007
BS-max		182	2	315	1		1.165	0.064	105		395			0.792	0.009	95		405			0.747	0.007
BS-sum				497	3		0.287	0.073			500			0.282	0.010			500			0.286	0.008
BS-DC				489	10	1	0.306	0.073			500			0.282	0.010			500			0.286	0.008
DC-GDFM				449	48	3	0.485	0.061			500			0.282	0.009			500			0.273	0.007
MS-max2	E2(.8)	1	3	493	3		0.330	0.069			500			0.294	0.010			499	1		0.291	0.007
MS-max		1	3	493	3		0.330	0.069			500			0.294	0.010			499	1		0.291	0.007
MS-sum				499	1		0.271	0.070			500			0.266	0.010			500			0.268	0.007
BS-max2		49		451			0.369	0.071	17		483			0.208	0.010	23		477			0.241	0.007
BS-max		48		452			0.366	0.071	28		472			0.264	0.010	23		477			0.240	0.007
BS-sum				458	38	4	0.144	0.072			497	3		0.122	0.010			499	1		0.124	0.008
BS-DC		1		483	16		0.145	0.073			498	2		0.122	0.010			499	1		0.124	0.008
DC-GDFM				466	33	1	0.511	0.061		2	497	1		0.319	0.009	4	25	471			0.175	0.007
MS-max2	E3(.8)		1	497	2		0.219	0.070			500			0.212	0.010			500			0.211	0.007
MS-max			1	497	2		0.219	0.070			500			0.212	0.010			500			0.211	0.007
MS-sum				500			0.203	0.070			500			0.202	0.010			500			0.204	0.007
BS-max2		2		498			0.071	0.073			500			0.058	0.010	1		499			0.064	0.008
BS-max		2		498			0.070	0.073			500			0.058	0.010			500			0.059	0.008
BS-sum				474	23	3	0.066	0.073			467	32	1	0.062	0.010			460	36	4	0.063	0.007
BS-DC				488	12		0.065	0.073			481	19		0.061	0.010			476	22	2	0.062	0.007
DC-GDFM				450	47	3	0.428	0.054			473	25	2	0.397	0.005		1	487	12		0.061	.007

Table 3: Empirical frequency for the number of changes. Dense model.

Method	Errors	$p = 20$							$p = 150$							$p = 200$						
		$\hat{R}$					MSE	adjRand	$\hat{R}$					MSE	adjRand	$\hat{R}$					MSE	adjRand
		0	1	2	3	≥ 4			0	1	2	3	≥ 4			0	1	2	3	≥ 4		
MS-max2	E1(.5)		26	438	35	1	0.266	0.128		8	466	26		0.065	0.052	1	14	453	32		0.050	0.052
MS-max			26	438	35	1	0.266	0.128		8	466	26		0.065	0.052	1	14	453	32		0.050	0.052
MS-sum		1	9	480	10		0.205	0.133	2	17	472	9		0.062	0.053	1	13	478	8		0.052	0.047
BS-max2		18	74	406	2		0.290	0.128	29	75	392	3	1	0.099	0.051	48	105	344	3		0.047	0.099
BS-max		16	72	410	2		0.289	0.128	28	75	395	1	1	0.098	0.052	46	103	344	7		0.047	0.098
BS-sum		1		458	37	4	0.181	0.134			48	94	358	0.087	0.035			15	30	455	0.028	0.077
BS-DC		3		464	33		0.174	0.136			60	104	336	0.077	0.039			25	46	429	0.033	0.068
DC-GDFM		37	85	323	42	13	0.389	0.113	311	115	73	1		0.512	0.022	407	58	34	1		0.282	0.034
MS-max2	E2(.8)	2	56	440	2		0.190	0.125		19	481			0.042	0.053	1	24	474	1		0.033	0.052
MS-max		2	56	440	2		0.190	0.125		19	481			0.042	0.053	1	24	474	1		0.033	0.052
MS-sum		2	158	468	15		0.138	0.128		16	459	25		0.039	0.053		24	442	33	1	0.033	0.052
BS-max2		7	13	475	5		0.150	0.127	18	74	405	3		0.080	0.050	23	100	375	2		0.075	0.047
BS-max		7	15	471	7		0.152	0.127	16	70	413	1		0.078	0.050	23	113	362	2		0.076	0.047
BS-sum		3	1	407	78	11	0.134	0.126	7	19	423	40	11	0.060	0.050	26	45	388	31	10	0.065	0.047
BS-DC		4	1	474	21		0.078	0.133	2	2	481	15		0.019	0.055	13	19	459	9		0.025	0.053
DC-GDFM				466	33	1	0.073	0.136		2	497	1		0.016	0.056	4	25	471			0.018	0.055
MS-max2	E3(.8)		1	499			0.094	0.133			500			0.026	0.054			500			0.018	0.053
MS-max			1	499			0.094	0.133			500			0.026	0.054			500			0.018	0.053
MS-sum			19	473	8		0.089	0.133		31	449	20		0.027	0.053		23	448	27	2	0.022	0.052
BS-max2				487	13		0.074	0.134	2	21	416	58	3	0.060	0.049	5	34	376	78	7	0.058	0.046
BS-max			1	489	10		0.074	0.134	1	21	417	58	3	0.059	0.049	7	34	373	81	5	0.059	0.046
BS-sum		1		322	129	48	0.082	0.128	11	16	152	154	167	0.067	0.044	50	22	104	121	203	0.070	0.040
BS-DC				451	47	2	0.037	0.137	2	2	302	136	58	0.011	0.053	2	7	283	146	62	0.015	0.051
DC-GDFM				450	47	3	0.047	0.138			473	25	2	0.036	0.056		1	487	12		0.018	0.054

Table 4: Empirical frequency for the number of changes. Sparse model.

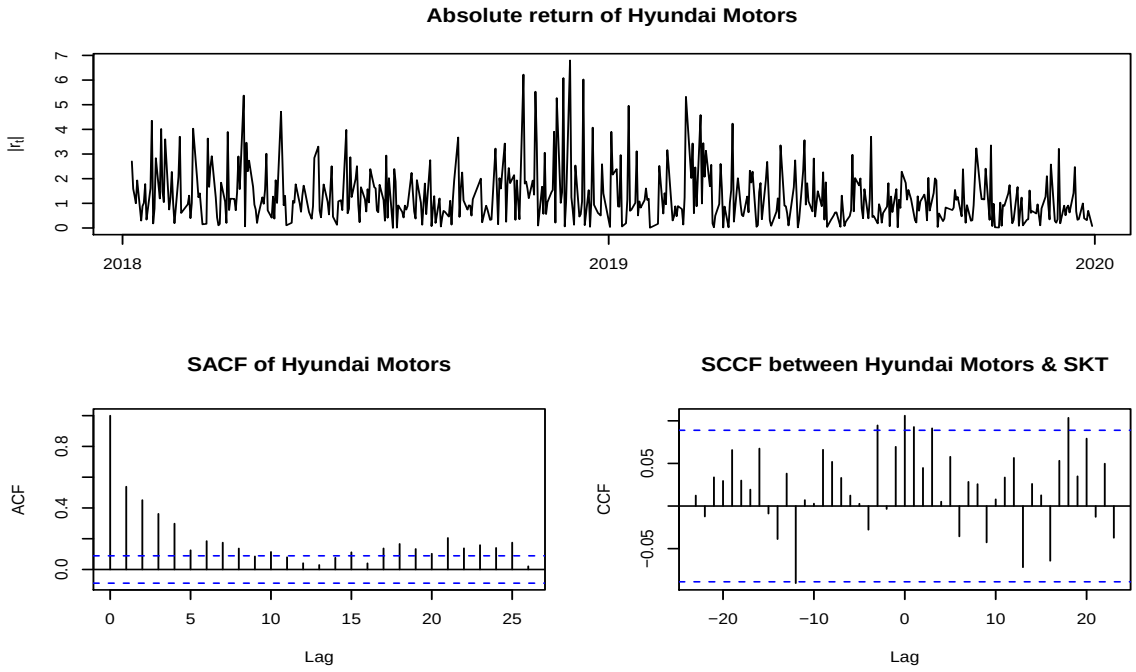

Figure 1: Time plots of absolute return of Hyundai Motors (top), sample autocorrelations (bottom left) and sample cross-correlations with SK telecom after prewhitening by AR(1).

4 Real data application

In this section, we illustrate our proposed methods using a real-life application. We consider the absolute log-returns of stocks given by

$$|r_{i,t}| = |\log P_{i,t} - \log P_{i,t-1}|\,,$$

where $P_{i,t}$ is the daily closing price of the i-th stock at time t. We considered stocks consisting of the KOSPI200 index. The KOSPI200 is a capitalization-weighted index based on 200 major stocks in the Korea Stock Exchange market. We calculated the absolute return from Jan 8, 2018 to Dec 30, 2019. KOSPI200 is renewed annually, so we only considered stocks that remained in KOSPI200 for two years. As a result, our data consist of 197 stocks over 486 time points.

Figure 1 shows the time plot (top) and sample autocorrelations (bottom left) for the absolute return of Hyundai Motors. It clearly shows non-ignorable temporal correlations. Furthermore, the bottom right panel shows the sample cross-correlations with SK Telecom after prewhitening the series by an AR(1) model. It shows moderate cross-correlations on small lags. It indicates that the absolute return of major Korean stocks have both temporal and cross-sectional correlations.

All seven methods used in Section 3 are applied to find change-points with the same tuning parameters selection as used in our simulation study in Section 3. Figure 2 shows how MOSUM-based methods find change-points. The top panel shows the MS-max statistics calculated over windows of length $2G = 48$. The critical value 5.919 is depicted by the red line. We find a local maximum over the critical value where the test statistics are greater than the critical value by at least $\eta G = 10$ times consecutively. For example, MS-max (top-panel) finds three change-points on

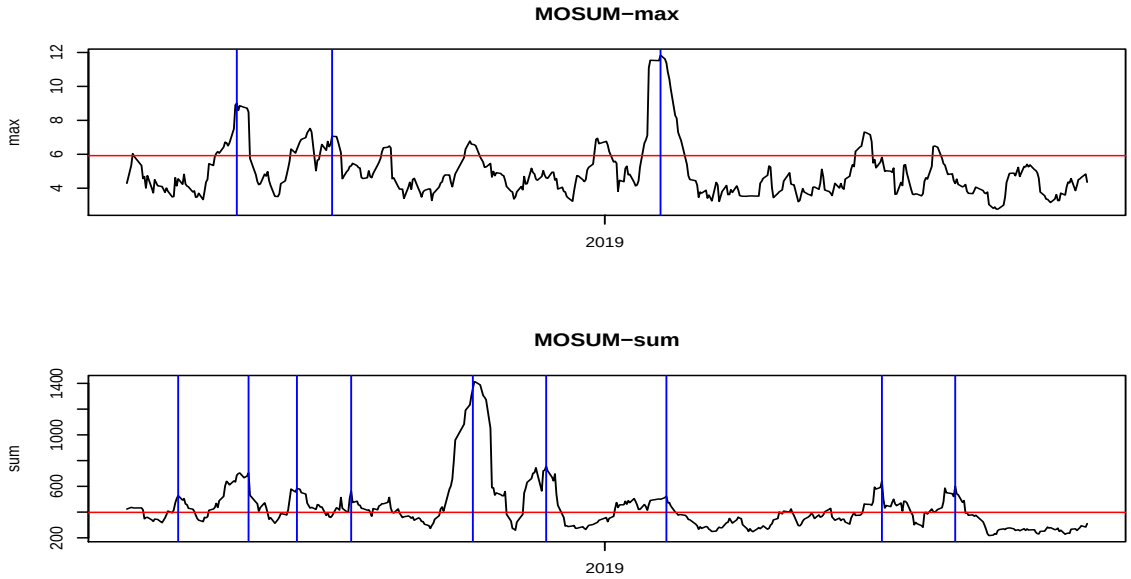

Figure 2: The moving sum test statistics for max-type (top) and sum (bottom). The vertical lines indicate change-points.

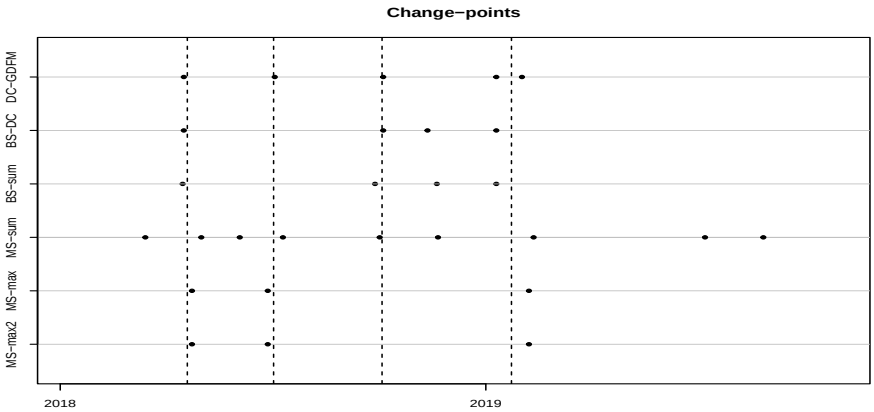

Figure 3: The estimated change-points from all methods considered. Vertical lines are final change-points selected by the majority vote from various methods considered.

April 24, 2018, June 28, 2018 and Feb 7, 2019 while MS-sum (bottom-panel) finds more change-points on March 15, 2018, May 2, 2018, June 4, 2018, July 11, 2018, Oct 2, 2018, Nov 21, 2018, Feb 11, 2019, July 8, 2019 and Aug 27, 2019. On the other hand, BS-max and BS-max2 find no change-point.

All change-points found are summarized in Figure 3 with circles indicating the estimated change-points. Final change-points are detected in the spirit of model ensemble. In other words, we estimated the final change-points by taking the average of nearby change-points selected by more than four different methods. The final change-points are represented as the vertical dotted line in Figure 3. They are April 20, 2018, July 3, 2018, Oct 4, 2018 and Jan 23, 2019. Figure 4 shows the KOSPI200 index and absolute return together with the final estimated change-points.

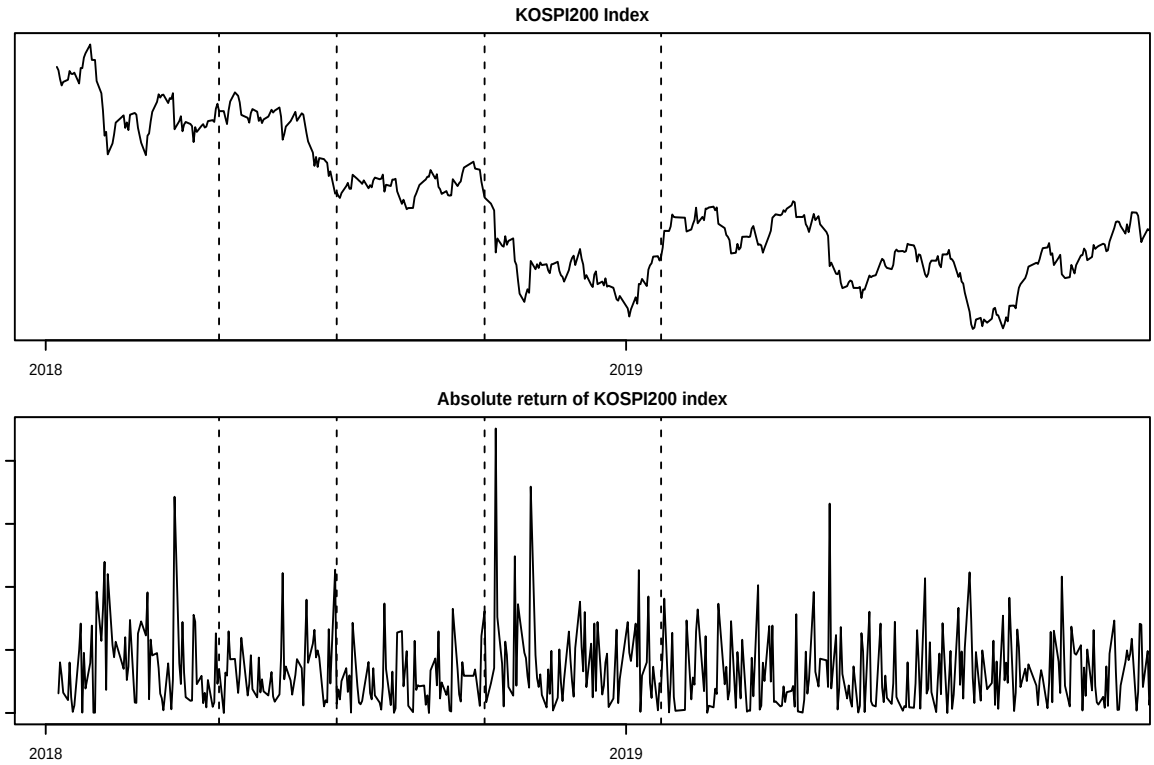

Figure 4: The final change-points on Apr 20, 2018, July 3, 2018, Oct 4, 2018 and Jan 23, 2019 with KOSPI200 index (top), and the absolute return of KOSPI200 index (bottom).

5 Conclusions

This study considered several methods to detect and locate multiple change-points in HDTS with temporal and spatial dependence. The test statistics take either sum or maximum of the adjusted CUSUM over the dimensions. Two algorithms, BS and MOSUM are used to find multiple change-points. However, a novelty comes from a new block wild bootstrap procedure that accommodates both temporal and cross-sectional dependence. Moreover, we examined how MOSUM algorithm can be used in the HDTS context with some theoretical justifications. Our simulations study indicates that MOSUM-based tests have sufficient power while keeping the nominal sizes. However, effective way of reducing bias in estimating the location of change-points in MOSUM methods remains an interesting open question.

Acknowledgements: The first author gratefully acknowledges financial support from the National Science Foundation 1934985, 1940124, 1940276, 2114143.

A Theoretical results

In this section we discuss some asymptotic results for the considered test statistics. We gather some preliminaries and assumptions in Section A.1, subsequently we present our results and their proofs in Section A.2. Section A.3 provides a rather technical but crucial results necessary to prove our main results.

A.1 Preliminaries and assumptions

Note that we will first consider the test statistics with the true long-run variance. Subsequently, Lemma 1 states that the true long-run variance can be replaced consistently by the Bartlett long-run variance estimator (2.2).

Instead of the notation $\widetilde{T}^2(k,G), \widetilde{J}(k,G), \widetilde{V}(k,G)$ in (2.2) for the statistics based on the estimated true long-run variance, we use slight alter the notation to $\widetilde{\mathtt{T}}_k^2(G), \widetilde{\mathtt{J}}_k(G), \widetilde{\mathtt{V}}_k(G)$ to indicate the dependence on the true long-run variance. Then, our test statistics with true long-run variance read as follows

$$\widetilde{\mathtt{T}}_k^2(G) = \max_{1\le i\le p} \frac{|\widetilde{\chi}_i(k,G)|^2}{\sigma_i^2}, \ \ \widetilde{\mathtt{J}}_k(G) = \max_{1\le i\le p} \frac{|\widetilde{\chi}_i(k,G)|}{\sigma_i}, \ \ \widetilde{\mathtt{V}}_k(G) = \frac{1}{p}\sum_{i=1}^{p} \frac{|\widetilde{\chi}_i(k,G)|^2}{\sigma_i^2}. \tag{A.1}$$

We are interested in deriving the asymptotic behavior of the following three test statistics

$$\max_{G\le k\le T-G} \widetilde{\mathtt{J}}_k(G), \quad \max_{G\le k\le T-G} \widetilde{\mathtt{T}}_k^2(G), \quad \max_{G\le k\le T-G} \widetilde{\mathtt{V}}_k(G) \tag{A.2}$$

under the null hypothesis of no change-point.

Before we proceed to our main results, we state some assumptions. The following Assumption 1 specifies the spatial dependence.

Assumption 1. *Assume that* $\mathbb{E}(\varepsilon_{i,1}\varepsilon_{j,1}) = \sigma_i\sigma_j\gamma_{i,j}$ *where* $\Gamma = (\gamma_{i,j})_{i,j=1,\dots,p}$ *is the correlation matrix with* $\gamma_{ii} = 1$ *for* $i = 1,\dots,p$ *and* $\max_{1\le i\ne j\le p}|\gamma_{ij}| < 1$.

Assumption 2. *There exist a p-dimensional Brownian motion* $\mathbb{B}(t) = (\mathbb{B}_1(t),\dots,\mathbb{B}_p(t))'$ *and a* $\nu > 0$ *such that*

$$\|\Sigma^{-\frac{1}{2}}\sum_{k=1}^{t}\varepsilon_k - \mathbb{B}(t)\| = O(t^{\frac{1}{2+\nu}}) \ \ a.s. \tag{A.3}$$

with $\Sigma = \mathrm{diag}(\sigma_1^2,\dots,\sigma_p^2)$ *and correlation matrix* $t\Gamma = t(\gamma_{i,j})_{i,j=1,\dots,p} = \mathbb{E}(\mathbb{B}(t)\mathbb{B}(t)')$.

Assumption 2 is satisfied for time series under general assumptions on the dependence structure. See for instance Theorem 2 in Kuelbs and Philipp (1980) for multivariate mixing time series.

The following assumption specifies how fast the window size G is allowed to grow with the sample size T.

Assumption 3. *The window size* G *and the sample size* T *are assumed to grow as*

$$\frac{T}{G}\to\infty \quad \textit{and} \quad \frac{T^{\frac{2}{2+\nu}}\log(T)}{G}\to 0$$

with ν *as in* (A.3).

A.2 Results

We present here the results in the "fixed p, large T" regime to give the reader a sense of the flavour of the theoretical results which correspond to our numerical analysis. Theoretical results in the "large p, large T" regime, where the dimension p is allowed to grow with the sample size T are

beyond the scope of this work and left for future consideration. We suspect that one can utilize the techniques in Jirak (2015) to prove more general results; see Remark 5 below.

The following proposition states the asymptotic behavior of $\max_{G\le k\le T-G}\widetilde{\mathrm{J}}_k(G)$ in (A.2) and is an extension of Theorem 2.1 in Eichinger and Kirch (2018) in the sense that we consider multivariate observations and allow for temporal and spatial dependence.

Proposition 1. *Suppose Assumptions 1–3 are satisfied. Then, under the hypothesis* (1.2),

$$\lim_{T\to\infty}\mathrm{P}\left(\max_{G\le k\le T-G}|\widetilde{\mathrm{J}}_k(G)|\le \boldsymbol{u}_T(x)\right)\xrightarrow{d}\mathcal{G},$$

where $P(\mathcal{G}\le x)=\exp\left(-e^{-x}\right)$ *and* $\boldsymbol{u}_T(x)=\boldsymbol{u}(x,\frac{T}{G})$ *with* $\boldsymbol{u}(x,y)=\frac{x}{a(y)}+b(y)$ *and*

$$a(y)=\sqrt{2\log(y)},\;\; b(y)=\frac{-\frac{1}{2}\log(2\pi)+\log\left(p\frac{3}{2}\right)}{\sqrt{2\log(y)}}+\sqrt{2\log(y)}. \tag{A.4}$$

Proof of Proposition 1. By the invariance principle in Assumption 2

$$\max_{G\le k\le T-G}\frac{1}{\sqrt{2G}}\left\|\Sigma^{-\frac{1}{2}}\sum_{t=k-G+1}^{k+G}\varepsilon_t-(\mathbb{B}(k+G)-\mathbb{B}(k-G+1))\right\|=O_{\mathrm{P}}\left(\frac{T^{\frac{1}{2+\nu}}}{\sqrt{2G}}\right)=O_{\mathrm{P}}\left(a^{-1}\left(\frac{T}{G}\right)\right)$$

Then, X_t can be replaced by an i.i.d. vector series Z_t with $\mathbb{E}X_tX_t'=\mathbb{E}Z_tZ_t'$. Define

$$L_{T,i}(k)=\frac{1}{\sqrt{2G}}\Big(\sum_{t=k+1}^{k+G}Z_{i,t}-\sum_{t=k-G+1}^{k}Z_{i,t}\Big).$$

Note that $L_{T,i}(k)$ is stationary and $2G$-dependent. Now, following the proof of Theorem 2.1 in Hušková and Slabý (2001), set

$$Y_{T,i}(u)=L_{T,i}([Gu]),\;\; 1\le u\le\frac{T}{G}-1$$

and $Y_T(u)=(Y_{T,1}(u),\dots,Y_{T,p}(u))'$. Then,

$$\{Y_T(u),1\le u\le\frac{T}{G}-1\}\to\{Y(u),1\le u<\infty\}$$

with $Y(u)=(Y_1(u),\dots,Y_p(u))'$ and

$$Y_i(u)=\frac{1}{\sqrt{2}}\left(\mathbb{B}_i(u)-\mathbb{B}_i(u-1)-(\mathbb{B}_i(u+1)-\mathbb{B}_i(u))\right).$$

The process $Y(u)$ is stationary and Gaussian with cross-covariance function

$$\begin{aligned}r_{ij}(v)&:=\mathrm{Cov}(Y_i(u+v),Y_j(u))\\&=\frac{1}{2}\mathbb{E}\Big((\mathbb{B}_i(u+v)-\mathbb{B}_i(u+v-1)-(\mathbb{B}_i(u+v+1)-\mathbb{B}_i(u+v)))\\&\qquad\times(\mathbb{B}_j(u)-\mathbb{B}_j(u-1)-(\mathbb{B}_j(u+1)-\mathbb{B}_j(u)))\Big)\\&=\begin{cases}\frac{1}{2}(-3v+2)\gamma_{ij} & v\in(0,1],\\ \frac{1}{2}(-v)\gamma_{ij} & v\in(1,2],\\ 0 & v>2,\end{cases}\end{aligned} \tag{A.5}$$

since $\mathbb{E}(\mathbb{B}_i(u_1)\mathbb{B}_j(u_2)) = \min\{u_1, u_2\}\gamma_{ij}$. We aim to use Theorem 1 stated in Section A.3 below. In order to do so, we have to verify the conditions (A.19)–(A.21).

Condition (A.19) easily follows by (A.5) since

$$r_{ii}(v) = \mathrm{Cov}(Y_i(u+v), Y_i(u)) = 1 - \frac{3}{2}|v| + o(|v|) \quad \text{as } v \to 0. \tag{A.6}$$

Similarly (A.5) gives

$$r_{ij}(v)\log(v) \to 0, \quad \text{as } v \to \infty, \tag{A.7}$$

which verifies (A.20) and by Assumption 1, (A.21) follows as

$$\max_{i \neq j} \sup_{v} |r_{ij}(v)| < 1. \tag{A.8}$$

Due to Theorem 1, the limiting distribution can then be characterized by the following limit

$$\frac{U}{\sqrt{2\pi}}\frac{3}{2}\exp\left(-\frac{\boldsymbol{u}^2(x,U)}{2}\right) \to \frac{1}{p}e^{-x}, \quad \text{as } U \to 0 \tag{A.9}$$

with $\boldsymbol{u}(x,U)$ as in (A.4). The coefficient sequences in (A.4) are chosen such that (A.9) is satisfied. In order to verify that $\boldsymbol{u}(x,U)$ chosen as in (A.4) indeed satisfies (A.9), we write $\sim$ to indicate asymptotic equivalence. Then, applying the logarithm to (A.9) and, with further explanations given below, we get, as $U \to \infty$,

$$\log(p) - \frac{1}{2}\log(2\pi) + \log\left(\frac{3}{2}\right) + \log(U) - \frac{\boldsymbol{u}^2(x,U)}{2} \sim -x \tag{A.10}$$

$$\Leftrightarrow \log(p) - \frac{1}{2}\log(2\pi) + \log\left(\frac{3}{2}\right) + \log(U) - \frac{\boldsymbol{u}^2(x,U)}{2} \sim -(\boldsymbol{u}(x,U) - b(U))a(U) \tag{A.11}$$

$$\Leftrightarrow \log(p) - \frac{1}{2}\log(2\pi) + \log\left(\frac{3}{2}\right) + \log(U) - \frac{\boldsymbol{u}^2(x,U)}{2} \sim -(\boldsymbol{u}(x,U) - b(U))\sqrt{2\log(U)}$$

$$\Leftrightarrow \frac{\log\left(\frac{1}{2\pi}\right) + 2\log\left(p\frac{3}{2}\right)}{\sqrt{2\log(U)}} + \sqrt{2\log(U)} - \frac{\boldsymbol{u}^2(x,U)}{\sqrt{2\log(U)}} \sim -2(\boldsymbol{u}(x,U) - b(U))$$

$$\Leftrightarrow \frac{-\frac{1}{2}\log(2\pi) + \log\left(p\frac{3}{2}\right)}{\sqrt{2\log(U)}} + \sqrt{2\log(U)} \sim b(U). \tag{A.12}$$

The relation (A.11) follows by replacing x with $(\boldsymbol{u}(x,U) - b(U))a(U)$ using our assumption $\boldsymbol{u}(x,y) = \frac{x}{a(y)} + b(y)$. Note that (A.10) also implies that $\boldsymbol{u}(x,U) \sim \sqrt{2\log(U)}$ as $U \to \infty$. Then, (A.12) follows by using $\boldsymbol{u}(x,U) \sim \sqrt{2\log(U)}$. Given (A.6), (A.7) and (A.8), Theorem 11.2.3. in Leadbetter et al. (2012) implies

$$\mathrm{P}\left(\max_{1\leq i\leq p}\sup_{u\in[1,U]} Y_i(u) \leq \boldsymbol{u}(x,U)\right) = \mathrm{P}\left(\bigcap_{1\leq i\leq p}\left\{\sup_{u\in[1,U]} Y_i(u) \leq \boldsymbol{u}(x,U)\right\}\right) \to \exp\left(-e^{-x}\right).$$

□

Remark 5. In the "large T, large p" regime, we expect the proof techniques proposed in Jirak (2015) to be useful. However, there is a crucial difference. The test statistic in Jirak (2015) does

not consider the maximum over a moving window. Instead, the maximum over the partial sums is considered. Therefore, Jirak (2015) reduces the problem to investigating the asymptotic behavior of the supremum of a Brownian Bridge over a finite interval. In contrast, here, the supremum over an increasing interval needs to be considered. The limit in Jirak (2015) is an extreme value distribution which is due to the maximum over the dimension. In contrast, here, the supremum of a Gaussian process over an increasing interval needs to be considered. For this reason, not only the maximum over the increasing dimension but also the supremum over a Gaussian process would contribute to the limit.

The following Proposition 2 gives the asymptotic behavior of the test statistic $\max_{G\le k\le T-G}\widetilde{\mathtt{T}}_k^2(G)$ in (A.2). The proof follows the same ideas as the proof of Proposition 1 and Remark 5 applies as well.

Proposition 2. *Suppose Assumptions 1–3 are satisfied. Then, under the hypothesis* (1.2),

$$\lim_{T\to\infty}\mathrm{P}\left(\max_{G\le k\le T-G}|\widetilde{\mathtt{T}}_k^2(G)|\le \boldsymbol{u}_T(x)\right)\xrightarrow{d}\mathcal{G},$$

where $P(\mathcal{G}\le x)=\exp\left(-e^{-x}\right)$ *and* $\boldsymbol{u}_T(x)=\boldsymbol{u}(x,\frac{T}{G})$ *with* $\boldsymbol{u}(x,y)=\frac{x}{a(y)}+b(y)$ *and*

$$a(y)=\sqrt{2\log(y)},\;\; b(y)=\frac{-\frac{1}{2}\log(2\pi)+\log\left(p\frac{3}{2}\right)}{\sqrt{2\log(y)}}+2\log(y). \tag{A.13}$$

Proof of Proposition 2. Note that

$$\mathrm{P}\left(\max_{G\le k\le T-G}|\widetilde{\mathtt{T}}_k^2(G)|\le \boldsymbol{u}_T(x)\right)=\mathrm{P}\left(\max_{G\le k\le T-G}|\widetilde{\mathtt{J}}_k(G)|\le \boldsymbol{u}_T^{\frac{1}{2}}(x)\right).$$

Then, it remains to follow the proof of Proposition 1 until relation (A.9). The convergence in (A.9) determines the choice of $\boldsymbol{u}(x,y)$. For completeness, we state the relation (A.9) here for $\boldsymbol{u}(x,y)=\frac{x}{a(y)}+b(y)$ as in (A.13), that is,

$$\frac{U}{\sqrt{2\pi}}\frac{3}{2}\exp\left(-\frac{\boldsymbol{u}(x,U)}{2}\right)\to\frac{1}{p}e^{-x},\quad\text{as } U\to 0.$$

Our choices of a and b in (A.13) satisfy this relations since

$$\log(p)-\frac{1}{2}\log(2\pi)+\log\left(\frac{3}{2}\right)+\log(U)-\frac{\boldsymbol{u}(x,U)}{2}\sim -x \tag{A.14}$$

$$\Leftrightarrow \log(p)-\frac{1}{2}\log(2\pi)+\log\left(\frac{3}{2}\right)+\log(U)-\frac{\boldsymbol{u}(x,U)}{2}\sim -(\boldsymbol{u}(x,U)-b(U))a(U) \tag{A.15}$$

$$\Leftrightarrow \log(p)-\frac{1}{2}\log(2\pi)+\log\left(\frac{3}{2}\right)+\log(U)-\frac{\boldsymbol{u}(x,U)}{2}\sim -(\boldsymbol{u}(x,U)-b(U))\sqrt{2\log(U)}$$

$$\Leftrightarrow \frac{\log\left(\frac{1}{2\pi}\right)+2\log\left(p\frac{3}{2}\right)}{\sqrt{2\log(U)}}+\sqrt{2\log(U)}-\frac{\boldsymbol{u}(x,U)}{\sqrt{2\log(U)}}\sim -2(\boldsymbol{u}(x,U)-b(U))$$

$$\Leftrightarrow \frac{-\frac{1}{2}\log(2\pi)+\log\left(p\frac{3}{2}\right)}{\sqrt{2\log(U)}}+2\log(U)\sim b(U). \tag{A.16}$$

The relation (A.15) follows by replacing x with $(\boldsymbol{u}(x,U)-b(U))a(U)$ using our assumption $\boldsymbol{u}(x,y)=\frac{x}{a(y)}+b(y)$. Note that (A.14) also implies that $\boldsymbol{u}(x,U)\sim 2\log(U)$ as $U\to\infty$. Then, (A.16) follows by using $\boldsymbol{u}(x,U)\sim 2\log(U)$. Proceeding as in the proof of Proposition 1 we get the result. □

After studying the asymptotic behavior of the test statistics

$$\max_{G\leq k\leq T-G}|\widetilde{\mathtt{J}}_k(G)| \quad \text{and} \quad \max_{G\leq k\leq T-G}|\widetilde{\mathtt{T}}^2_k(G)| \tag{A.17}$$

in Propositions 1 and 2, it remains to consider $\max_{G\leq k\leq T-G}|\widetilde{\mathtt{V}}_k(G)|$ in (A.2). While it was possible to derive the limiting distributions for the statistics in (A.17) in a "fixed p, large T" regime with a path to possible extensions to a "large p, large T" regime, proving convergence results for $\max_{G\leq k\leq T-G}|\widetilde{\mathtt{V}}_k(G)|$ is more delicate. For this reason, we add a discussion instead of stating a formal result. Further considerations are left for future work.

As discussed in Remark 5, the interplay of maximizing across the dimensions and over a certain window in order to detect multiple change-points in (A.17), makes deriving theoretical results particularly challenging. However, by taking the maximum twice, we remain in the same class of distributions by utilizing max-stable properties.

In contrast, the statistic $\max_{G\leq k\leq T-G}|\widetilde{\mathtt{V}}_k(G)|$, averages over the dimensions and takes the maximum over a certain window. Extreme value distributions are not sum-stable. To illustarte that in more detail, consider

$$\mathrm{P}\left(\max_{G\leq k\leq T-G}|\widetilde{\mathtt{V}}_k(G)|\leq u_T(x)\right)$$

Following the same steps as in the proofs of Propositions 1 and 2, we get by the invariance principle in Assumption 2

$$\max_{G\leq k\leq T-G}\frac{1}{\sqrt{2G}}\left\|\Sigma^{-\frac{1}{2}}\sum_{t=k-G+1}^{k+G}\varepsilon_t-(\mathbb{B}(k+G)-\mathbb{B}(k-G+1))\right\|=O_{\mathrm{P}}\left(\frac{T^{\frac{1}{2+\nu}}}{\sqrt{2G}}\right)=O_{\mathrm{P}}\left(a^{-1}\left(\frac{T}{G}\right)\right)$$

Then, X_t can be replaced by an i.i.d. vector series Z_t with $\mathbb{E}X_tX_t'=\mathbb{E}Z_tZ_t'$. Define

$$L_{T,i}(k)=\frac{1}{\sqrt{2G}}\Big(\sum_{t=k+1}^{k+G}Z_{i,t}-\sum_{t=k-G+1}^{k}Z_{i,t}\Big).$$

Note that $L_{T,i}(k)$ is stationary and $2G$-dependent. Now, following the proof of Theorem 2.1 in Hušková and Slabý (2001), set

$$Y_T(u)=L_T([Gu]),\ \ 1\leq u\leq\frac{T}{G}-1.$$

Then,

$$\{Y_T(u),1\leq u\leq\frac{T}{G}-1\}\to\{Y(u),1\leq u<\infty\}$$

with

$$Y_i(u)=\frac{1}{\sqrt{2}}\left(\mathbb{B}_i(u)-\mathbb{B}_i(u-1)-(\mathbb{B}_i(u+1)-\mathbb{B}_i(u))\right).$$

The process $Y(u) = (Y_1(u), \ldots, Y_p(u))$ is stationary and Gaussian with cross-covariance function

$$\begin{aligned} r_{ij}(v) &:= \mathrm{Cov}(Y_i(u+v), Y_j(u)) \\ &= \frac{1}{2}\mathbb{E}\Big((\mathbb{B}_i(u+v) - \mathbb{B}_i(u+v-1) - (\mathbb{B}_i(u+v+1) - \mathbb{B}_i(u+v))) \\ &\qquad \times (\mathbb{B}_j(u) - \mathbb{B}_j(u-1) - (\mathbb{B}_j(u+1) - \mathbb{B}_j(u)))\Big) \\ &= \begin{cases} \frac{1}{2}\rho_{ij}(-3v+2) & v \in (0,1], \\ \frac{1}{2}\rho_{ij}(-v) & v \in (1,2], \\ 0 & v > 2. \end{cases} \end{aligned}$$

Then, choosing $a(y)$ and $b(y)$ as in (A.13)

$$\mathrm{P}\left(\frac{1}{p}\sum_{i=1}^{p} \sup_{u\in[1,U]} Y_i(u) \leq \boldsymbol{u}_T^{\frac{1}{2}}(x) \right) = \mathrm{P}\left(\sum_{i=1}^{p} \{ \sup_{u\in[1,U]} Y_i(u) \leq p\boldsymbol{u}_T^{\frac{1}{2}}(x) \} \right).$$

While $\{\sup_{u\in[1,U]} Y_i(u)\}$ (with the appropriate normalization through $a(y)$ and $b(y)$) follows an extreme value distribution and so does the maximum over all $i = 1, \ldots, p$ one cannot make an immediate statement about the sum. With that we finish our discussion.

The following lemma ensures that replacing σ_i by the Bartlett long-run variance estimator (2.2) maintains consistent estimation. Recall the Bartlett estimator in (2.2)

$$\widehat{\sigma}_i^2 = \sum_{j=-q_i}^{q_i} \left(1 - \frac{|j|}{2q_i+1}\right) \widehat{\gamma}_i(j)$$

with

$$\widehat{\gamma}_i(j) = \frac{1}{2G} \sum_{t=k-G+1}^{k+G-j} Y_t Y_{t+j}$$

with the mean-adjusted sequence defined in (2.3) and reaclled here

$$Y_{i,t} = X_{i,t} - \overline{X}_{i,L(k)} 1_{\{t\leq k\}} - \overline{X}_{i,R(k)} 1_{\{t>k\}}.$$

Lemma 1. *The estimator for the long-run variance in* (2.2) *satisfies*

$$\max_{G\leq k\leq T-G} \left| \frac{1}{\widehat{\sigma}_i} - \frac{1}{\sigma_i} \right| = o_P\left(\left(\log\left(\frac{T}{G}\right)\right)^{-1} \right).$$

Proof. By the continuous mapping theorem it suffices to prove

$$\max_{G\leq k\leq T-G} \left| \widehat{\sigma}_i^2 - \sigma_i^2 \right| = O_P\left(\frac{(2q_i+1)T^{\frac{1}{2}}}{G} + r_T \right).$$

The result then follows by (2.4) in Theorem 2.1 in Eichinger and Kirch (2018). $r_T = \sum_{j\in\mathbb{Z}} \frac{|j|}{2q_i+1} |\gamma(h)|$ $q_i^2 G/T = O(1)$

□

A.3 Technical results and their proofs

In this section, we generalize a result in Leadbetter et al. (2012). Before we formally state our result, the following paragraph attempts to clarify how our result complements those in Leadbetter et al. (2012). Leadbetter et al. (2012) considers stationary Gaussian processes $\{\xi(t)\}_{0\le t\le T}$ with covariance function $r(v) = \mathbb{E}(\xi(t)\xi(t+v))$ which is assumed to satisfy

$$r(v) = 1 - \lambda|v|^\alpha + o(|v|^\alpha) \quad \text{as} \quad v \to 0, \tag{A.18}$$

for some $\alpha \in (0, 2]$. Chapter 12.3 in Leadbetter et al. (2012) establishes convergence results for the supremum over an increasing interval $M(T) = \sup_{0\le t\le T} \xi(t)$. One aims to find functions $a(T), b(T)$ such that

$$\mathrm{P}(a(T)(M(T) - b(T)) \le x) \to \exp\left(-e^{-x}\right), \quad \text{as} \quad T \to \infty.$$

However, results for multivariate Gaussian processes are only established for the case $\alpha = 2$ in (A.18); see Chapter 11.2 in Leadbetter et al. (2012). For our results, we need to consider multivariate Gaussian processes in the case $\alpha = 1$.

Let $\{\boldsymbol{\xi}(t) = (\xi_1(t), \ldots, \xi_p(t))\}_{0\le t\le T}$ be a jointly Gaussian process with zero means, variances one and covariance function $r_{ij}(v) = \mathrm{Cov}(\xi_i(t), \xi_j(t+v))$. In order to establish results for the joint probability of the suprema $M_k(T) = \sup_{0\le t\le T} \xi_k(t)$, we suppose

$$r_{ii}(v) = 1 - \lambda_k|v| + o(|v|) \quad \text{as} \quad v \to 0, \tag{A.19}$$

$$r_{ij}(v)\log(v) \to 0, \quad \text{as} \quad v \to \infty, \tag{A.20}$$

$$\max_{i\ne j} \sup_v |r_{ij}(v)| < 1. \tag{A.21}$$

The final result can then be stated as follows.

Theorem 1. *Let $u_k = u_k(T) \to \infty$ as $T \to \infty$, so that*

$$T\mu_k := T\lambda_k \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_k^2}{2}\right) \to \tau_k > 0, \;\; 1 \le k \le p \tag{A.22}$$

and suppose that the jointly stationary Gaussian process $\{\boldsymbol{\xi}(t)\}_{0\le t\le T}$ satisfies (A.19)–(A.21)*. Then,*

$$\mathrm{P}(M_k(T) \le u_k, 1 \le k \le p) \to \exp\left(-\sum_{k=1}^{p} \tau_k\right), \quad \textit{as} \quad T \to \infty.$$

Most of the required results and techniques to prove Theorem 1 are already provided in Leadbetter et al. (2012). For completeness, we present its proof here. The proof is based on a series of lemmas which provide an approximation of the probability of suprema over an increasing interval. Following the notation in Leadbetter et al. (2012), we introduce for a fixed $h > 0$, $n = [T/h]$.

Proof. By Lemma 2 below

$$\mathrm{P}(M_k(nh) \le u_k, k = 1, \ldots, p) - \mathrm{P}^n(M_k(h) \le u_k, k = 1, \ldots, p) \to 0.$$

Then, since $nh \le T \le (n+1)h$, it follows

$$\begin{aligned}
\mathrm{P}(M_k(T) \le u_k, k = 1, \dots, p) &= \mathrm{P}^n(M_k(h) \le u_k, k = 1, \dots, p) + o(1) \\
&= \left(1 - \sum_{k=1}^{p} \mathrm{P}(M_k(h) > u_k) + o\left(\sum_{k=1}^{p} \mu_k\right)\right)^n && \text{(A.23)} \\
&= \left(1 - \sum_{k=1}^{p} (\mu_k h + o(\mu_k)) + o\left(\sum_{k=1}^{p} \mu_k\right)\right)^n && \text{(A.24)} \\
&\to \exp\left(-\sum_{k=1}^{p} \tau_k\right), && \text{(A.25)}
\end{aligned}$$

where (A.23) follows by (A.40) in Lemma 5 and (A.24) by Theorem 12.2.9 in Leadbetter et al. (2012). Furthermore, the last relation (A.25) is valid since by (A.22) and $n = [T/h]$ we get $\mu_k \sim \frac{\tau_k}{T}$ $n \sim \frac{T}{h}$. □

The following Lemma 2 is crucial in proving Theorem 1. Its proof will be given by a series of lemmas below the statement.

Lemma 2. *Suppose* (A.19)–(A.21). *Then,*

$$\mathrm{P}(M_k(nh) \le u_k, k = 1, \dots, p) - \mathrm{P}^n(M_k(h) \le u_k, k = 1, \dots, p) \to 0.$$

Proof. The result follows by Lemmas 3 and 4 below. □

Recall that for a fixed $h > 0$, $n = [T/h]$, and divide $[0, nh]$ into n intervals of length h. We further split those intervals into subintervals I_r and I_r^* of length $h - \varepsilon$ and ε, respectively. The following lemma corresponds to Lemma 12.3.2 in Leadbetter et al. (2012).

Lemma 3. *Suppose* $u \to \infty$, $q \to 0$, $u^2 q \to a > 0$, (A.19) *and* (A.22) *hold. Then,*

$$0 \le \mathrm{P}\left(M_k\left(\bigcup_{r=1}^{n} I_r\right) \le u_k, k = 1, \dots, p\right) - \mathrm{P}(M_k(nh) \le u_k, k = 1, \dots, p) \le \frac{\varepsilon}{h} \sum_{k=1}^{p} \tau_k, \tag{A.26}$$

$$\begin{aligned}
0 &\le \mathrm{P}\left(\xi_k(jq) \le u_k, jq \in \bigcup_{r=1}^{n} I_r, k = 1, \dots, p\right) - \mathrm{P}\left(M_k\left(\bigcup_{r=1}^{n} I_r\right) \le u_k, k = 1, \dots, p\right) \\
&\le \rho(a) \sum_{k=1}^{p} \tau_k
\end{aligned} \tag{A.27}$$

with τ_k *as defined in* (A.22) *and* $\rho(a) \to 0$ *as* $a \to 0$.

Proof. We consider (A.26) and (A.27) separately.

Proof of (A.26)*:* By Boole's inequality

$$
\begin{aligned}
0 &\le \mathrm{P}\left(M_k\left(\bigcup_{r=1}^{n} I_r\right) \le u_k, k=1,\dots,p\right) - \mathrm{P}(M_k(nh) \le u_k, k=1,\dots,p) \\
&\le n\mathrm{P}(M_k(I_1^*) > u_k, k=1,\dots,p) \\
&\sim n\sum_{k=1}^{p} \mathrm{P}(M_k(I_1^*) > u_k) \\
&= n\sum_{k=1}^{p} \mu_k \varepsilon \to \sum_{k=1}^{p} \tau_k \frac{\varepsilon}{h}, \qquad \text{(A.28)}
\end{aligned}
$$

where (A.28) follows under assumption (A.19) by Theorem 12.2.9 in Leadbetter et al. (2012) and since $n\mu_k \sim T\frac{\mu_k}{h} \to \frac{\tau_k}{T}$.

Proof of (A.27)*:* Following the proof of Lemma 12.3.2 (ii) in Leadbetter et al. (2012), we get

$$
\begin{aligned}
0 &\le \mathrm{P}\left(\xi_k(jq) \le u_k, jq \in \bigcup_{r=1}^{n} I_r, k=1,\dots,p\right) - \mathrm{P}\left(M\left(\bigcup_{r=1}^{n} I_r\right) \le u_k, k=1,\dots,p\right) \\
&\le n \max_{r=1,\dots,n} \left(\mathrm{P}\left(\xi_k(jq) \le u_k, jq \in I_r, k=1,\dots,p\right) - \mathrm{P}\left(M_k\left(I_r\right) \le u_k, k=1,\dots,p\right)\right) \\
&\le n \max_{r=1,\dots,n} \big(\mathrm{P}(\xi_k(0) > u_k, k=1,\dots,p) + \mathrm{P}\left(\xi_k(jq) \le u_k, jq \in [0,h], k=1,\dots,p\right) \\
&\qquad - \mathrm{P}\left(M_k(h) \le u_k, k=1,\dots,p\right)\big) \\
&= n \max_{r=1,\dots,n} \big(\mathrm{P}(\xi_k(0) > u_k, k=1,\dots,p) - \mathrm{P}\left(\xi_k(jq) > u_k, jq \in [0,h], k=1,\dots,p\right) \\
&\qquad + \mathrm{P}\left(M_k(h) > u_k, k=1,\dots,p\right)\big), \qquad \text{(A.29)} \\
&\le n\left((1-H_1(a))h\sum_{k=1}^{p}\mu_k + o\left(\sum_{k=1}^{p}\mu_k\right)\right), \qquad \text{(A.30)}
\end{aligned}
$$

where $H_1(a)$ is a constant depending on the limit $u^2 q \to a > 0$ and satisfying $\rho(a) := 1 - H_1(a) \to 0$ as $a \to 0$; see the proof of Lemma 12.3.2 in Leadbetter et al. (2012). Each probability in (A.29) can be dealt with separately to get (A.30). With explanations given below, we get for the first probability in (A.29)

$$
\mathrm{P}(\xi_k(0) > u_k, k=1,\dots,p) \le \sum_{k=1}^{p} \mathrm{P}(\xi_k(0) > u_k) \le \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{p} \frac{1}{u_k} \exp\left(-\frac{u_k^2}{2}\right) = o\left(\sum_{k=1}^{p}\mu_k\right), \qquad \text{(A.31)}
$$

where the second inequality in (A.31) is a consequence of $1 - \Phi(u) \sim \varphi(u)/u$, where Φ and φ denote the Gaussian distribution and density functions respectively. The last relation in (A.31) follows by (12.2.18) in Leadbetter et al. (2012).

For the second probability in (A.29), one can write

$$\begin{aligned}\mathrm{P}\left(\xi_k(jq) > u_k, jq \in [0,h], k = 1,\ldots,p\right) &\leq \mathrm{P}\left(\max_{jq\in[0,h]} \xi_k(jq) > u_k, k = 1,\ldots,p\right) \\ &\leq \sum_{k=1}^{p} \mathrm{P}\left(\max_{jq\in[0,h]} \xi_k(jq) > u_k\right) \\ &= H_1(a)h\sum_{k=1}^{p}\mu_k + o\left(\sum_{k=1}^{p}\mu_k\right), \qquad \text{(A.32)}\end{aligned}$$

where (A.32) follows by Lemma 12.2.4 (i) in Leadbetter et al. (2012) with $\alpha = 1$.

Similarly, the third probability in (A.29) satisfies,

$$\mathrm{P}\left(M_k(h) > u_k, k = 1,\ldots,p\right) \leq \sum_{k=1}^{p}\mathrm{P}\left(M_k(h) > u_k\right) = h\sum_{k=1}^{p}\mu_k + o\left(\sum_{k=1}^{p}\mu_k\right),$$

by (12.2.18) in Leadbetter et al. (2012). □

The following lemma corresponds to Lemma 12.3.3 in Leadbetter et al. (2012).

Lemma 4. *Let $r(v) \to 0$ as $v \to \infty$, $u^2 q \to a > 0$,* (A.19)*,* (A.20) *and* (A.22) *be satisfied. Then, as $T \to \infty$,*

$$\mathrm{P}\left(\xi_k(jq) \leq u_k, jq \in \bigcup_{r=1}^{n} I_r, k = 1,\ldots,p\right) - \prod_{r=1}^{n}\mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1,\ldots,p) \to 0 \qquad \text{(A.33)}$$

$$\begin{aligned}&\limsup_{u\to\infty}\left|\prod_{r=1}^{n}\mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1,\ldots,p) - \mathrm{P}^n(M(h) \leq u_k, k = 1,\ldots,p)\right| \\ &\leq \left(\rho(a) + \frac{\varepsilon}{h}\right)\sum_{k=1}^{p}\tau_k.\end{aligned} \qquad \text{(A.34)}$$

Proof. We prove (A.34) and (A.34) separately.

Proof of (A.33)*:* Follows by the proof of relation (11.2.4) in Leadbetter et al. (2012) which is part of the proof of Lemma 11.2.1.

Proof of (A.34)*:* Simple application of triangular inequality yields

$$\begin{aligned}&\left|\prod_{r=1}^{n}\mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1,\ldots,p) - \mathrm{P}^n(M(h) \leq u_k, k = 1,\ldots,p)\right| \\ &\leq \left|\prod_{r=1}^{n}\mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1,\ldots,p) - \prod_{r=1}^{n}\mathrm{P}(M_k(I_r) \leq u_k, k = 1,\ldots,p)\right| \\ &\qquad + \left|\prod_{r=1}^{n}\mathrm{P}(M_k(I_r) \leq u_k, k = 1,\ldots,p) - \mathrm{P}^n(M(h) \leq u_k, k = 1,\ldots,p)\right|. \qquad \text{(A.35)}\end{aligned}$$

The first probability in (A.35) can be bounded as

$$
\begin{aligned}
0 &\leq \prod_{r=1}^{n} \mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1, \ldots, p) - \prod_{r=1}^{n} \mathrm{P}(M_k(I_r) \leq u_k, k = 1, \ldots, p) \\
&\leq n \max_{r=1,\ldots,n} (\mathrm{P}(\xi_k(jq) \leq u_k, jq \in I_r, k = 1, \ldots, p) - \mathrm{P}(M_k(I_r) \leq u_k, k = 1, \ldots, p)) \\
&\leq n \left((1 - H_1(a)) h \sum_{k=1}^{p} \mu_k + o\left(\sum_{k=1}^{p} \mu_k \right) \right), \qquad \text{(A.36)}
\end{aligned}
$$

where we used for (A.36) the same arguments as in the proof of (A.27). The second probability in (A.35) satisfies

$$
\begin{aligned}
0 &\leq \prod_{r=1}^{n} \mathrm{P}(M_k(I_r) \leq u_k, k = 1, \ldots, p) - \mathrm{P}^n(M(h) \leq u_k, k = 1, \ldots, p) \\
&= \mathrm{P}^n(M_k(I_1) \leq u_k, k = 1, \ldots, p) - \mathrm{P}^n(M(h) \leq u_k, k = 1, \ldots, p) \qquad \text{(A.37)} \\
&\leq n \left(\mathrm{P}(M_k(I_1) \leq u_k, k = 1, \ldots, p) - \mathrm{P}(M(h) \leq u_k, k = 1, \ldots, p) \right) \\
&\leq n \mathrm{P}(M_k(I_1) > u_k, k = 1, \ldots, p) \\
&\leq n \sum_{k=1}^{p} \mathrm{P}(M_k(I_1) > u_k) \\
&\sim n \sum_{k=1}^{p} \mu_k \varepsilon \to \sum_{k=1}^{p} \tau_k \frac{\varepsilon}{h}, \qquad \text{(A.38)}
\end{aligned}
$$

where (A.37) is due to stationarity and the relation (A.38) follows under assumption (A.19) by Theorem 12.2.9 in Leadbetter et al. (2012). □

Lemma 5. *Suppose* (A.21). *Then,*

$$
\mathrm{P}(M_k(h) > u_k, M_l(h) > u_l) = o(\mu_k + \mu_l) \quad \textit{for} \quad k \neq l, \qquad \text{(A.39)}
$$

$$
\mathrm{P}(M_k(h) \leq u_k, k = 1, \ldots, p) = 1 - \sum_{k=1}^{p} \mathrm{P}(M_k(h) > u_k) + o\left(\sum_{k=1}^{p} \mu_k \right). \qquad \text{(A.40)}
$$

Proof. The statement coincides with Lemma 11.2.2 in Leadbetter et al. (2012) and can be proven without assuming (A.19). Therefore, we omit the details. Note that (A.39) is only necessary to prove (A.40). □

References

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

Baek, C., Gates, K. M., Leinwand, B., and Pipiras, V. (2021). Two sample tests for high-dimensional autocovariances. *Computational Statistics & Data Analysis*, 153:107067.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78–92.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2019). Change point estimation in panel data with temporal and cross-sectional dependence. *arXiv preprint arXiv:1904.11101*.

Chang, J., Yao, Q., and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika*, 104(1):111–127.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038.

Cho, H. and Fryzlewicz, P. (2018). hdbinseg: Change-point analysis of high-dimensional time series via binary segmentation.

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

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

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648.

Hušková, M. and Slabỳ, A. (2001). Permutation tests for multiple changes. *Kybernetika*, 37(5):605–622.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483.

Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing b-valued random variables. *Annals of Probability*, 8(6):1003–1036.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Processing*, 85(8):1501–1510.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Lee, T. and Baek, C. (2020). Block wild bootstrap-based CUSUM tests robust to high persistence and misspecification. *Computational Statistics & Data Analysis*, 150:106996.

Li, H., Munk, A., and Sieling, H. (2016). FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics*, 10(1):918–959.