Testing synchronization of change-points for multiple time series

Abstract

In this paper, we investigate the problem of detecting synchronization of a single change-point across components of a multivariate time series. The identification of synchronized change-points can often lead to finding a unanimous reason behind such changes whereas rejection might consequently prompt further analysis. Our proposed test statistic is simple to perceive, but its null distribution can be highly nontrivial to explicitly characterize. To overcome this, we employ a Gaussian approximation result, assisted by a clever and agnostic (to the existence of change-point) estimation of covariance matrix. We also cover the testing problem where the number of dimensions can grow with n. Extensive simulations are provided to corroborate our theoretical results for both low- and high-dimensional scenarios. We also provide two interesting real-world applications and discuss the implications of our findings based on the statistical tests.

Keywords: Change-point, Multiple time series, Synchronization, Gaussian approximation, Covariance estimation

1 Introduction

Consider a multiple series $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d})^\top \in \mathbb{R}^d$ with the mean-noise structure as

$$\mathbf{X}_i = \boldsymbol{\mu}_i + \mathbf{e}_i = (\mu_{i,1}, \ldots, \mu_{i,d})^T + \mathbf{e}_i, \; i = 1, \ldots, n, \tag{1.1}$$

where $\mathbf{e}_i \in \mathbb{R}^d$ is a stationary time series with $\mathbb{E}[\mathbf{e}_i] = \mathbf{0}$. For each $1 \leq j \leq d$, denote the j-th component of the time series, $(X_{1,j}, \ldots, X_{n,j})^\top$, as $\boldsymbol{X}_{\cdot j}$. Additionally, we assume for each stream/coordinate series j, there is at most one change-point. If $\boldsymbol{X}_{\cdot j}$ has a change-point, namely,

$$\mu_{ij} = \begin{cases} \mu_j^L, & \text{if } i/n \leq \tau_j, \\ \mu_j^R, & \text{if } i/n > \tau_j \end{cases}, \; 1 \leq i \leq n, \tag{1.2}$$

where $\tau_j \in (0,1)$ is the (re-scaled) change-point, then the jump size δ_j at the point τ_j is defined as $\delta_j = \mu_j^R - \mu_j^L$. For notational convenience, if the $\boldsymbol{X}_{\cdot j}$ has no such change-point, we set $\delta_j = 0$. Our focus in this paper lies in testing 'synchronization hypothesis', described as follows

$$H_0 : \tau_1 = \ldots = \tau_d. \tag{1.3}$$

Note that, if $\delta_j = 0$, then τ_j is not well-defined. Assume $k > 0$ out of d coordinate-series have true change-point at indices $r_1, \ldots, r_k$. If the true change-locations are 'synchronized' i.e. $\tau_{r_1} = \ldots = \tau_{r_k} = l \in (0,1)$ (say), then we set the convention that H_0 in (1.3) is true, as one could vacuously think $\tau_j = l$, for any $j \notin \{r_1, \ldots, r_k\}$ with corresponding $\delta_j = 0$.

In defense of starting with a rather simple and yet interpretable model in (1.1), our aim in this paper is to cover a large class of possibly non-linear, stationary multiple time series $(\mathbf{e}_i)$, so as to enlarge scope of applying it in large number of scenarios. Note, when $\mathbf{e}_i$ are i.i.d. normal variables, this is referred as the popular Gaussian sequence model. Our goal is to keep the simple model structure, and generalize substantially from the independence assumption of $\mathbf{e}_i$. Technically speaking, even though we assume stationarity for $\mathbf{e}_i$, the mean-noise structure imposed in (1.1) and the possibly piece-wise nature of the vectors $\boldsymbol{\mu}$'s, make the observed $\mathbf{X}_i$ process non-stationary. Moreover, such non-stationarity is interesting, as it depends on whether an individual component series has a change-point or not. Some components might not have any change-point at all, while the ones that have one, are not guaranteed to have them all synchronized. This translates to the following dynamics; before any change-point occurs at any of the series, the multivariate series $(\mathbf{X}_i)$ is stationary. However, unless all the series have a change-point which is exactly synchronized on their positions, post the first change-point at any series, the multivariate process becomes non-stationary. If one looks at every component individually, they are either completely stationary, or they are piece-wise stationary, depending on whether there is no change-point or there is one true break, respectively. Holistically speaking, we cover an interesting class of non-stationary multivariate series, which one could not simply classify as a piece-wise stationary process. Change-point analysis for multiple time series is not a new topic. However, as outlined below, almost all of the research works in this direction make the simplifying assumption of a synchronized change-point, which makes the models piece-wise stationary. Our paper, to the best of our knowledge, is a first in proposing a statistical test to validate such an assumption of synchronization.

Change-point testing and detection for time series data has a widespread literature spanning over several decades. Early work on change-points started with [84, 85], and then numerous seminal papers [34, 49, 50, 35, 99, 111, 68] etc. discussed the problem of detecting structural breaks in different settings, such as mean-shift or two-stage regression. A great overview of the early progression of this literature can be found in [17]. As [7] points out through multiple references, typically the literature first considers independent settings, for example [52] etc., and then more complex dependent settings are considered in [1, 14, 18, 110, 94] etc. The literature for change-point detection and inference for panel data or multivariate time series, albeit much smaller, also has a long history by now.

The early work in this particular research area dates back to [59, 60], in which they discuss a random break model with the breakpoints having an independent and identical distribution in a Bayesian framework. Later this was extended to autoregressive models in [61]. While they allow for different breaks in different series and put a distribution around it to describe its randomness, their models only allow for stationarity across components, which might be too restrictive for panel data as [12] points out succinctly. In terms of notation, all these works assume that in (1.1), $\boldsymbol{X}_{\cdot j}$ is independent of $\boldsymbol{X}_{\cdot k}$ for $j \neq k$. On the other hand, the assumption of common breaks across different components was slowly becoming popular around this time. [13] proposes construction of confidence interval for this shared change-point, while allowing for only a subset of coordinates to have a proper change. [52] tests for existence of such a shared change-point. Interested readers might take a look at [53] and [5] for developments on this topic. A similar test for existence of change-points was developed in [122] based on scan and segmentation algorithms, and in [51] using adaptation of the CUSUM method to panel data. [12] constructs limiting distribution for such a shared change-point in mean and variance for linear time series. [38] discusses hypothesis testing about the magnitude of change in mean for multivariate data in the non-vanishing difference regime. [64, 15, 109] investigated estimation of the change point in panel data, wherein the cross-sectional dependence is modeled by a common factor model, which effectively makes the cross-sectional dependence low-dimensional. [71] also discusses a problem of similar flavor; they estimate the common breaks, but allow for unobservable fixed effect. [109] provides a consistent estimation technique for an unknown shared change-point in mean. Inference on common change-point for panel data with independence and cross-sectional dependence are discussed at [21] and [22] respectively. Except for the Bayesian treatments at [59, 60, 61] and a very recent Bayesian work [107], we believe that the vast literature of multivariate time series very rarely allowed change-points to be asynchronized across the components and a solid theoretical framework to test such a common assumption is probably due. As mentioned above, this assumption of common change-point induces piece-wise stationarity for regimes without any change-point, and thus it becomes easier to use tools developed for stationary multivariate series even in settings where there is significant spatial or contemporaneous correlations. But unfortunately, such an assumption could turn out quite restrictive, as it is not very rare that abrupt changes occur at different times, at different components or spatial locations of interest.

Next we discuss a few real-data scenarios, spread across different scientific fields, where synchronization of potential change-points can be questionable. Consider a neuroimage data: multiple time series emanating from neighboring voxels. Under certain medication or intervention, whether these series act in a synchronous fashion or not is an important medical question. In the research of Human Activity Recognition (HAR), [63, 3] analyze 561 such time series obtained from different health tracker from smartphones. The change-points or the interventions are introduced when an individual changes their activity. In the world of climate data, such asynchronous behavior is not uncommon either, in say, change-point analysis for hurricane or other adversarial climate events. Alongside a good review of this topic, [40] discusses a rate shift in hurricane incidence and how these are different for overall US and the southern part of Florida. To perceive why synchronization could be questionable, consider a pathway of a hurricane. The related climate variables will show some form of short-term abrupt change; however these changes should pop up not together at all locations, depending on when how far are these from the eye of hurricane and the timeline of the hurricane passing close to them. In different areas of time series econometrics, especially those in the domain of energy and developments, change-points often occur due to external events, political intervention, international relations, etc., and these change-points have interesting spatial flavor in them. The question of synchronization is loosely related to the classical framework of Granger causality (See [45]) as it talks about correlation between two series at some lag being significant. Under this premise of Granger causality, it is easy to perceive why these two series, observed simultaneously, might lead to similar pattern but in an asynchronous fashion. See [75, 101] for excellent reviews on this literature. Granger causality is also well studied for neuroscience [39, 98, 36] etc. and climate analysis, [8, 6, 77] etc. Finally, relevant applications are not rare in epidemiology. For instance, one could analyze incidence rate time series of contagious diseases in different locations. If the synchronization hypothesis fails, then one could proceed to understand the progression based on the spikes or change-points observed. A recent Bayesian work, [107], states that for time-series analysis for different spatial location assumption of shared change-point might be too restrictive. They allow for asynchronized change-points and showed that temperature data across 207 locations in California and Covid count data across all counties of Illinois indeed show different breaks. We also show a couple of interesting applications in Section 9 and discuss the implication of our findings.

We summarize our contributions in this paper as follows. First, we propose a test statistic to test the synchronization hypothesis, which is spelled out at (1.3), and establish its validity and consistency. However, even though the test statistic itself is intuitive, its null distribution does not have a closed form expression, and thus, from the perspective of practicability, this poses a challenge to actually carry this test out. To overcome this, we use a Gaussian approximation result for multiple stationary time series with optimal rate. Although, there have been a few works on this front, namely [73, 116, 62], none of the existing work suffices for our purpose. The best possible rates were obtained in [62] for a general non-stationary process and we adopt this to arrive at a Gaussian approximation with variance directly related to the long run covariance matrix of the error process. One final step remains, in estimating this error (long-run) variance, and given the premise of our problem of possible existence of possibly non-synchronized change-point, it is a non-trivial problem. To this end, we were able to establish consistency of our proposed method of estimating this covariance, agnostic of both whether a particular series has a change-point or the change-points across different series are synchronized or not. This is, to the best of our knowledge, a novel contribution on its own. A very pertinent question arises on whether these results are extendable beyond just the low dimensional case i.e. where the dimension of the mean parameter does not grow with n. Here, we also answer this question in affirmative. We adapt from [80] to arrive at a suitable gaussian approximation in high dimensions and use it to conduct our test. We also provide estimation of the covariance matrix in this high-dimensional case. Finally, we conclude our paper by discussing two interesting real-life datasets where a synchronization testing could yield some interesting insights. In the appendix, we provide extensive simulations to thoroughly address different scenarios based on number of components with true change-points and whether they are synchronized or not. We also provide some simulations for the high-dimensional scenario at the very end of our appendix.

We conclude the introduction with organization and some notations, to be used throughout.

1.1 Organization

We begin Section 2 by rigorously introducing our test statistic for the synchronization problem. We further prove that, under a very general class of alternative settings, a test based on this test statistic will achieve full asymptotic power. Section 3 is devoted to the application of the Gaussian approximation result to the bootstrap approximation for the null distribution of the test statistic. In particular, we include an oracle bootstrap procedure and prove its validity. This oracle bootstrap algorithm motivates us to estimate the long run covariance matrix Σ_∞ of the stationary error process $(\mathbf{e}_i)$. Our estimate is shown to be consistent in an agnostic fashion, i.e. irrespective of the presence or absence of a change point in each dimension. Finally, all of these ideas are combined to yield our final bootstrap Algorithm 3, whose validity is shown in Theorem 4.1. Crucially, our bootstrap algorithm has a "hidden" first stage, where we individually test the existence of change-point at each coordinate. This is discussed in Section 4. In Section 5, we consider the behavior of our test statistic in high-dimensional set-ups, and provide general theory corresponding to the results obtained for the multivariate case. Finally, in Section 6, we briefly summarize our simulation studies, and provide two interesting data examples where synchronization of change-points translates into meaningful hypotheses in corresponding fields, and testing such synchronization reflects statistically valid insights from the data. Details of our simulation studies backing up our methodology and all theoretical proofs are deferred to the Appendix Sections 8-13. At the end of the Appendix, in Section 14, a brief discussion on bootstrapping the test statistic in high dimensions, is also included.

1.2 Notation

For a matrix $A = (a_{ij})$, define the Frobenius norm as $|A| := (\sum a_{ij}^2)^{1/2}$. With slight abuse of notation, when suitable, we use $|\cdot|$ to denote (i) absolute value of a real number, (ii) Euclidean norm of a vector $\in \mathbb{R}^d$ for $d \geq 2$, and, (iii) Frobenius norm of a matrix. Moreover, for a matrix A, we let $\rho^\star(A)$ be the largest singular value of A; further, if A is symmetric and positive semidefinite, $\lambda_{\min}(A)$ will denote its lowest eigen value. For a random vector $\mathbf{Y} \in \mathbb{R}^d$, write $\mathbf{Y} \in \mathcal{L}_p,\ p > 0$, if $\|\mathbf{Y}\|_p := [\mathbb{E}(|\mathbf{Y}|^p)]^{1/p} < \infty$. Throughout the text, $\lfloor x \rfloor$ refers to the greatest integer less than or equal to x. C_p would refer to a constant that depends only on p, but could take different values on different occurrences. If two sequences $\{x_n\}$ and $\{y_n\}$ satisfy $|x_n| \leq c y_n$ for some $c < \infty$ and all sufficiently large n, then we write $x_n \lesssim y_n$. If both $x_n \lesssim y_n$ and $y_n \lesssim x_n$ hold, then we write $x_n \asymp y_n$. We also use $a \wedge b$ for $\min(a, b)$.

2 Methodology

We briefly discuss the motivation behind our test statistic in a very general set-up, and in subsequent sections, we describe our algorithms in detail specific to our model (1.1). For $\mathbf{X} := \mathbf{X}_1^n = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$, $\mathbf{X}_i \in \mathbb{R}^d$, assume the general parametric model

$$\mathbf{X}_i \sim f(\boldsymbol{\lambda}_i), \boldsymbol{\lambda}_i := (\lambda_{i1}, \ldots, \lambda_{i,d}) \in \mathbb{R}^d,\ f : \mathbb{R}^d \to \mathbb{R}^d.$$

For each $1 \leq j \leq d$, we let $\lambda_{ij} = \begin{cases} \lambda_j^L, & i \leq n\tau_j, \\ \lambda_j^R, & i > n\tau_j, \end{cases}$ with $\tau_j \in [0, 1)$ for all j. Suppose the change-points are synchronized, i.e. $\tau_1 = \ldots = \tau_d = \tau$. For each $j \in \{1, \ldots, d\}$, assume the practitioner uses a data-based loss function $\mathcal{L}_j : \mathcal{X}^n \times K \to \mathbb{R}$, to estimate τ_j as

$$\hat{\tau}_j := \arg\max_{\gamma \in K} \mathcal{L}_j(\mathbf{X}, \gamma), \tag{2.1}$$

where $K \subseteq [0, 1)$ is some appropriate measurable set, and $\mathcal{X}$ is the sample space of the random variables X_i. For the validity of our procedure, we require $\hat{\tau}_j \xrightarrow{\mathbb{P}} \tau_j$. The usual choices of the loss function include likelihood-based methods, or more general non-parametric methods such as CUSUM, MOSUM or methods based on U-statistics or M-statistics (see [97, 37, 55]). Since, under null we expect $\hat{\tau}_1 \approx \hat{\tau}_2 \approx \ldots \approx \hat{\tau}_d$, intuitively, the expressions $\sum_{j=1}^{d} \max_\gamma \mathcal{L}_j(\mathbf{X}, \gamma)$, the max and $\sum$ can be (approximately) interchanged. Based on this motivation, our test statistic reads as follows.

$$G_n = \sum_{j=1}^{d} \max_{\gamma} \mathcal{L}_j(\mathbf{X}, \gamma) - \max_{\gamma} \sum_{j=1}^{d} \mathcal{L}_j(\mathbf{X}, \gamma). \tag{2.2}$$

Note that, $G_n \geq 0$ always, and as suggested above, under H_0, we expect $G_n \approx 0$. Therefore, we reject H_0 for large values of G_n.

2.1 Test statistic for model (1.1)

With (1.1) being an additive model, testing (1.3) motivates us to use loss function same as that of the well-studied CUSUM statistic. Mathematically speaking, in (2.1) we employ

$$\mathcal{L}_j(\mathbf{X}, \gamma) = |S_{ij} - i\bar{X}_{\cdot j}|/\sqrt{n} \text{ where } i = \lfloor n\gamma \rfloor, \gamma \in (0,1), \text{ and } S_{ij} = \sum_{k=1}^{i} X_{kj}.$$

Here and onwards, in (1.1) we assume $\mathbf{X}_i \in \mathbb{R}^d$ for $1 \leq i \leq n$. Therefore, for the specific model (1.1), equation (2.2) can be rewritten as

$$T_n := T_n(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = n^{-1/2} \left(\sum_{j=1}^{d} |S_{n\hat{\tau}_j, j} - n\hat{\tau}_j \bar{X}_{\cdot j}| - \sum_{j=1}^{d} |S_{n\hat{\tau}, j} - n\hat{\tau} \bar{X}_{\cdot j}| \right), \tag{2.3}$$

where,

$$\hat{\tau}_j := \frac{1}{n} \arg\max_{1 \leq i \leq n} |S_{ij} - i\bar{X}_{\cdot j}|/\sqrt{n}, \text{ and } \hat{\tau} := \frac{1}{n} \arg\max_{1 \leq i \leq n} \sum_{j=1}^{d} |S_{ij} - i\bar{X}_{\cdot j}|/\sqrt{n}. \tag{2.4}$$

Subsequently in this paper, we will consider T_n as our test statistic. As explained in (2.1), it is crucial that $\hat{\tau}_j$ is a consistent estimator of the individual change-points under suitable conditions. Moreover, for the validity of our test, it is also necessary that under H_0, the common change-point τ is consistently estimated by $\hat{\tau}$. However, in order to discuss such results, we first need to explicitly characterize the dependency structures of the error processes $(\mathbf{e}_i)_{i \in \mathbb{Z}}$. In the following subsection, we provide a very general stationary causal set-up, which enables us to arrive at interpretable and useful theoretical results.

2.2 Dependence structure

To perform some meaningful analysis of our test statistic T_n- in particular, to retrieve the unknown change-points $(\tau_j)_{j=1}^d$ from the observed $(\mathbf{X}_i)_{i=1}^n$, we need to impose some dependence structure on the process $(\mathbf{e}_i)$. We assume the following causal representation:

$$\mathbf{e}_i = H(\varepsilon_i, \varepsilon_{i-1}, \ldots) = (e_{i1}, e_{i2}, \ldots, e_{id})^\top, \tag{2.5}$$

where H is a measurable function $\mathbb{R}^\infty \to \mathbb{R}^d$ and ε_i's are independent and identically distributed random variables. We also assume that $\mathbf{e}_i \in \mathcal{L}_p$ where $p > 2$. This representation is inspired from writing the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ in terms of conditional quantile function of i.i.d. uniform random variables. It allows us to employ the widely used idea of coupling to model the dependence structure. In fact, we will use the framework of functional dependence measure on the underlying process (see [112]). Suppose that $(\varepsilon_i')_{i\in\mathbb{Z}}$ is an independent copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$. Define the functional dependence measure

$$\theta_{i,p} = \|\mathbf{e}_i - \mathbf{e}_{i,\{0\}}\|_p = \|H(\mathcal{F}_i) - H(\mathcal{F}_{i,\{0\}})\|_p, \ i \geq 0, \ p \geq 2, \tag{2.6}$$

where, for $k \leq i$, $\mathcal{F}_{i,\{k\}}$ is the coupled version of $\mathcal{F}_i$ with ε_k in $\mathcal{F}_i$ replaced by ε_k':

$$\mathcal{F}_{i,\{k\}} = (\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{k+1}, \varepsilon_k', \varepsilon_{k-1}, \ldots), \tag{2.7}$$

and $\mathbf{e}_{i,\{k\}} = H(\mathcal{F}_{i,\{k\}})$. In particular, [112] showed that for a linear process $\mathbf{e}_i = \sum_{k=0}^\infty a_k \varepsilon_{i-k}$, $\theta_{k,p} \leq 2\|\varepsilon_0\|_p |a_k|$. Therefore, $\theta_{k,p}$ measures the dependence of e_k on ε_0. We further restrict ourselves to short range dependent processes; i.e. we assume that,

$$\Theta_{0,p} = \sum_{i=0}^\infty \theta_{i,p} < \infty. \tag{2.8}$$

Processes with long-range dependency often involve approximation through a "Non-Central Limit Theorem"(see [20, 125]), and application of standard tools (such as various moment and large deviation bounds [96, 28, 81, 42]) is very different, compared to weak dependent processes. However, (2.8) is not a major restriction, since, almost all popularly used stationary processes (such as ARMA, ARCH, GARCH, Volterra processes, etc.) can be shown to fit into our framework. Further interesting examples can be found in [32, 58, 124], among others. Subsequently, we discuss how we can establish the validity and consistency of our test statistic under this framework.

2.3 Validity and consistency of our test statistic

As discussed before, we start with a consistency result for our individual CUSUM estimates of change-points, $\hat{\tau}_j$ and a consistency result for our common change-point estimator $\hat{\tau}$ under H_0.

Proposition 2.1. *Grant model (1.1) for $(\mathbf{X}_t)$ with the error process $(\mathbf{e}_t)$ satisfying (2.5) and (2.8). Then, for all $1 \leq j \leq d$, $|\hat{\tau}_j - \tau_j| = O_{\mathbb{P}}((n\delta_j^2)^{-1} \wedge 1)$. Further, if $H_0 : \tau_1 = \ldots = \tau_d := \tau$ is true, then it also holds that $|\hat{\tau} - \tau| = O_{\mathbb{P}}((n\sum_{j=1}^d \delta_j^2)^{-1} \wedge 1)$.*

The rate $O_{\mathbb{P}}(1/(n\delta^2))$ has been long studied in the change-point literature, appearing at least in [9], [10] and [11], as well as in recent minimax optimality results (see [108] and [104]). However, to the best of our knowledge, the literature is missing any such results in the general setting of causal stationary process satisfying (2.5) and (2.8). The argument for Proposition 2.1 is standard, involving classical techniques such as Hàjek-Rényi inequality. As discussed immediately after equation (2.1), Proposition 2.1 enables us to argue about the validity of our test statistic T_n in the asymptotic sense. The following result summarizes this, as well as the effectiveness of T_n under the alternative hypothesis H_0^c.

Proposition 2.2. *Grant model (1.1) for $\mathbf{X}_t$ with the error process $\mathbf{e}_t$ satisfying (2.5) and (2.8). Then under the synchronized setting, i.e. under H_0 described in (1.3), $T_n = O_{\mathbb{P}}(1)$. On the other hand, under H_0^c, i.e., if*

$$\mathcal{H} := \{\{j_1, j_2\} : 1 \leq j_1, j_2 \leq d, \tau_{j_1} \neq \tau_{j_2}\}$$

is non-empty, then $T_n \xrightarrow{\mathbb{P}} \infty$ if

$$n \max_{\{j_1,j_2\}\in\mathcal{H}} (\delta_{j_1}^2 \wedge \delta_{j_2}^2) \to \infty. \tag{2.9}$$

Remark 1. *The imposed condition (2.9) can be shown to be optimal, even in a minimax sense. Consider the following toy example. Suppose $d = 2$, $\tau_1 \neq \tau_2$, and $n\delta_1^2 \to \infty$ (say, $\delta_1 = 1$), but $n\delta_2^2 \to 0$ (eg., $\delta_2 = 1/n$). Intuitively speaking, since $\delta_1 \gg \delta_2$, $\hat{\tau} \approx \tau_1$, and since $\hat{\tau}_1 \xrightarrow{\mathbb{P}} \tau_1$, therefore, $|S_{n\hat{\tau}_1,1} - n\hat{\tau}_1\bar{X}_{\cdot 1}| \approx |S_{n\hat{\tau},1} - n\hat{\tau}\bar{X}_{\cdot 1}|$. On the other hand, note that since δ_2 is small, $\hat{\tau}_2$ is no longer a consistent estimate of τ_2. Therefore, from the null behavior of CUSUM estimate of $\hat{\tau}_2$, as well as the fact that $\hat{\tau}$ is not close to τ_2, one can show both $|S_{n\hat{\tau}_2,2} - n\hat{\tau}_2\bar{X}_{\cdot 2}|$ and $|S_{n\hat{\tau},2} - n\hat{\tau}\bar{X}_{\cdot 2}|$ are small. Therefore, T_n can be shown to be $O_{\mathbb{P}}(1)$ even under this alternative. Thus, the condition (2.9) is necessary to distinguish between the null H_0 and alternate H_0^c.*

The proofs of Propositions 2.1 and 2.2 are provided in Section 10. Note that, even though under the null $T_n = O_{\mathbb{P}}(1)$, in general the null distribution of T_n will be extremely complicated, or even intractable. Therefore, we aim to provide a bootstrap approximation for it. In Section 3, we state a KMT-type Gaussian approximation result for the partial sums of the error process S_i^e, which we then use to provide a bootstrap approximation to T_n.

3 Approximation of null distribution of T_n

This section comprises of the two crucial theoretical results, that form the basis of our bootstrap-based algorithm for testing the hypothesis of synchronization. First, in Section 3.1, we mention a Gaussian approximation result that will be used to approximate the null distribution of T_n via bootstrap. This result involves an unknown parameter in the form of Σ_∞, the long-run variance of $(\mathbf{e}_t)$, which is estimated in Section 3.2.

3.1 KMT-type Gaussian approximation

Strong invariance principles, originating as extensions of classical functional central limit theorems (FCLTs), is well-studied in the literature, with the case for i.i.d. random variables settled by [66, 67] with the optimal rate $n^{1/p}$ for $p > 2$. For univariate stationary process, such optimal rate has been achieved in the seminal work by [19]. Recently, [62] extended this to multivariate non-stationary process, albeit without explicit regularization of variance. Following the corresponding argument in [19], for stationary multivariate process $(\mathbf{e}_i)_{i\in\mathbb{Z}}$, the variance of the approximating Gaussian process (G_i) can be regularized to be $i\Sigma_\infty$. We state the complete result here.

Theorem 3.1. *Suppose $(\mathbf{e}_i)_{i\in\mathbb{Z}}$ has the causal representation (2.5), and satisfies (2.8) for some $p > 2$. Let $S_i^e = \sum_{j=1}^i \mathbf{e}_j$, $1 \le i \le n$. Define the long-run variance $\Sigma_\infty = \sum_{k\in\mathbb{Z}} \mathbb{E}[\mathbf{e}_0\mathbf{e}_k^\top]$. Assume it satisfies $\lambda_{\min}(\Sigma_\infty) \ge c > 0$ for some positive constant c. If we further assume*

$$\Theta_{i,p} = O(i^{-A}), \quad \textit{with } A > A_0 = \max\left\{\frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p}, 1\right\}, \tag{3.1}$$

then, there exists a probability space (Ω_c, A_c, P_c) on which we can define random vectors $(\mathbf{e}_i^c)$, with the partial sums $S_i^c = \sum_{j=1}^i \mathbf{e}_j^c$, and a Gaussian process G_i with independent increments, such that $(S_i^c)_{i=1}^n \overset{D}{=} (S_i^e)_{i=1}^n$, and it holds

$$\max_{i\le n} |S_i^c - G_i| = o_{\mathbb{P}}(n^{1/p}) \textit{ where, } G_i = \sum_{j=1}^i \mathbf{Z}_j \textit{ with } (\mathbf{Z}_i)_{i=1}^n \overset{i.i.d.}{\sim} N(\mathbf{0}, \Sigma_\infty). \tag{3.2}$$

The above theorem also appeared in [102, 72]. Let us discuss the implications of this result in our context. We are interested in obtaining Gaussian approximations of functionals of the form $W(t) := \sum_{i=1}^n \mathbf{e}_i w_i(t)$, where $w_i(\cdot) : [0,1] \to \mathbb{R}$ are weight functions, and $(\mathbf{e}_i)_{i=1}^n$ are mean-zero, multivariate stationary process. Such quantities occur frequently in various methodologies of change point estimation, and also in many other applications. One can employ our Theorem 3.1 to deal with $W(t)$. A similar treatment can also be found in [115, 23]. Suppose $\mathbf{Z}_1, \ldots, \mathbf{Z}_n \overset{\text{i.i.d.}}{\sim} N(0, \Sigma_\infty)$ are such that $G_i = \sum_{j=1}^i \mathbf{Z}_i$, and let

$$W^\diamond(t) = \sum_{i=1}^n w_i(t)\mathbf{Z}_i. \tag{3.3}$$

Here, $W^\diamond(t)$ is the Gaussian process that we want to use to approximate $W(t)$. Let $\Omega_n =$

$\sup_{t\in(0,1)}\{|w_1(t)| + \sum_{i=2}^n |w_i(t) - w_{i-1}(t)|\}$. Now, from Theorem 3.1, one obtains

$$\sup_{t\in(0,1)} |W(t) - W^\diamond(t)| \leq \Omega_n \sup_{1\leq i\leq n} |S_i^e - G_i| = o_{\mathbb{P}}(\Omega_n n^{1/p}). \tag{3.4}$$

We can motivate an oracle bootstrap algorithm based on (3.4). By "oracle", we emphasize that at this stage, we assume that Σ_∞ and the means $\boldsymbol{\mu}_i$'s are known; we simply wish to investigate the rate of error if T_n is approximated by its Gaussian analogue, as dictated by (3.3). This is done in the following lemma, whose proof is provided in Section 11.2.

Lemma 3.1. *Assume* (1.1). *Under the assumptions of Theorem 3.1, on a possibly enlarged probability space, there exists independent* $(\mathbf{Z}_i \sim N(\boldsymbol{\mu}_i, \Sigma_\infty))_{i=1}^n$ *such that it holds*

$$|T_n - T_n^Z| = o_{\mathbb{P}}(n^{1/p-1/2}) \text{ where } T_n^Z := T_n(\mathbf{Z}_1, \ldots, \mathbf{Z}_n). \tag{3.5}$$

This lemma is repeatedly used while analyzing the validity of the bootstrap algorithms proposed subsequently. In view of Lemma 1, the aforementioned "oracle" bootstrap algorithm can be motivated naturally. For simplicity, assume $\mu_j^L = 0$ for $1 \leq j \leq d$. As a prelude to our complete bootstrap algorithm in Section 4, we provide this algorithm here.

Algorithm 1: Oracle test of synchronization

---

1 **Input**: $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, bootstrap size B, $\tau \in (0,1)$, sequence of jumps $\{\delta_j\}_{j=1}^d$, long-run covariance Σ_∞.

2 **Goal**: Test if $\tau_1 = \ldots = \tau_d = \tau$.

- Construct Test statistic T_n from (2.3).
- For s= $1, \ldots, B$
  - Generate bootstrap samples $(\mathbf{Z}_i^{(s)})_{i=1}^n \overset{\text{i.i.d}}{\sim} N(0, \Sigma_\infty)$.
  - For $j = 1, \ldots, d$, $X_{ij}^{(s)} \leftarrow Z_{ij}^{(s)} + \delta_j I\{i/n > \tau\}$, $1 \leq i \leq n$.
  - Generate $T_n^{(s)}$ from $(\mathbf{X}_1^{(s)}, \ldots, \mathbf{X}_n^{(s)})$.
- Bootstrap p-value: $p_0 \leftarrow \frac{1}{B+1}(\sum_{s=1}^{B} \mathbb{I}\{T_n > T_n^{(s)}\} + 1)$.

While the Lemma 3.1 emphasizes the efficacy of the oracle algorithm, it is important to take note of what more a practitioner requires in order to obtain a valid, yet completely data-based bootstrap algorithm to test (1.3). In particular, observe that in the input of Algorithm 1, the usually unknown quantities are: common change point τ, the jumps $\{\delta_j\}_{j=1}^d$ and Σ_∞. It will be convenient to have a checklist of the quantities that can be readily estimated, and the quantities that are yet to be estimated. In the following, each of statement holds under the corresponding set of assumptions of the accompanying mathematical results.

- Under null, the common change-point $\hat{\tau}$ is consistently estimated due to Proposition 2.1.
- The jumps δ_j (and in general the means pre and post-change-point) can also be consistently estimated under H_0 as well as under H_0^c; upon consistently estimating $\hat{\tau}_j$'s (or $\hat{\tau}$ under the H_0), we can simply consider $\hat{\delta}_j := \hat{\mu}_j^R - \hat{\mu}_j^L$ as an estimate, where $\hat{\mu}_j$'s are defined in (3.6).
- Therefore, in order to have a consistent, data-based, Gaussian bootstrap algorithm, we require an estimation procedure for Σ_∞. This is addressed in our next section.

3.2 Estimation of Σ_∞

Consider the model (1.1), and recall Σ_∞ from Theorem 3.1 as the long-run variance matrix of the error process $(\mathbf{e}_t)$. Since $\mathbf{e}_i$'s are not directly observed, we have to use the original observations $\mathbf{X}_i$'s and the estimated means pre- and post-change-point. Combining these ideas, in this section we propose a non-parametric estimator of Σ_∞, which is consistent if there is at most one change point (popularly referred as AMOC in the change-point literature) in each time series. In particular, we show that our estimator is consistent agnostic to whether H_0 is true or not, i.e., the change-points do not need to be synchronized. For $1 \le j \le d$, recall $\hat{\tau}_j$ from (2.4) as a CUSUM-based estimate of true change-points τ_j. For $1 \le i \le n$, define the estimated means as,

$$\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{ij})_{j=1}^d, \text{ where } \hat{\mu}_{ij} = \begin{cases} \hat{\mu}_j^L := \frac{1}{\lfloor n\hat{\tau}_j \rfloor} \sum_{i=1}^{\lfloor n\hat{\tau}_j \rfloor} X_{ij}, & \text{if } i \le n\hat{\tau}_j, \\ \hat{\mu}_j^R := \frac{1}{n-\lfloor n\hat{\tau}_j \rfloor} \sum_{i=\lfloor n\hat{\tau}_j \rfloor+1}^{n} X_{ij}, & \text{if } i > n\hat{\tau}_j \end{cases}. \tag{3.6}$$

The lag-k autocovariance matrix is estimated as

$$\hat{\Gamma}_k := \frac{1}{n} \sum_{i=1}^{n-k} (\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{X}_{i+k} - \hat{\boldsymbol{\mu}}_{i+k})^\top.$$

Let $K : [-\omega, \omega] \to \mathbb{R}$ be a continuous kernel with $K(0) = 1$. Then, with a suitable choice of bandwidth B_n, our estimator of Σ_∞ is:

$$\hat{\Sigma}_{n,B_n} := \hat{\Gamma}_0 + \sum_{k=1}^{n-1} K(k/B_n)(\hat{\Gamma}_k + \hat{\Gamma}_k^\top). \tag{3.7}$$

Observe that, this is a multivariate version of a HAC estimator (see [82, 2]). The following result yields the error rate of $\hat{\Sigma}_{n,B_n}$ as an estimator of Σ_∞.

Theorem 3.2. *Assume model (1.1) for $\mathbf{X}_i$, with $\mathbf{e}_i$ satisfying (2.5) and (2.8) for some $p > 2$. Moreover, let $K : [-\omega, \omega] \to \mathbb{R}$, $K \in \mathcal{C}^1$ be a symmetric bounded kernel function with $K(0) = 1$ and $\sup_x |K'(x)| \leq C$. Suppose $p' = \min\{p, 4\}$. Then, for a bandwidth $B_n \to \infty$ with $B_n n^{2/p'-1} \to 0$, the error rate for the long-run covariance estimate $\hat{\Sigma}_{n,B_n}$ in (3.7) can be summarized as*

$$\rho^\star(\hat{\Sigma}_{n,B_n} - \Sigma_\infty) = O_{\mathbb{P}}(B_n n^{2/p'-1} + B_n^{-1}). \tag{3.8}$$

Here the $B_n n^{2/p'-1}$ is the consistency error, and B_n^{-1} corresponds to bias.

Remark 2 (Agnostic nature of Theorem 3.2)**.** *We would like to point out that, even for those coordinates j for which there is no change-points (i.e. $\delta_j = 0$), we pretend that there is a change-point, estimate it and use it to estimate the left and right means $\hat{\mu}_j^L$ and $\hat{\mu}_j^R$. Interestingly, this still results in a consistent estimate of Σ_∞. This is convenient from the point of view of a practitioner, since they usually have no way to know which coordinates have no change-point. Similar ideas also appear in Remark 3.2 of [33], albeit in the context of studentization in a synchronized change-point setting. As mentioned therein, [54] also mentions similar estimators in a synchronized AMOC setting with univariate random variables, with theoretical guarantees of consistency. The agnostic nature of such a construction in a multivariate set-up is not underpinned in [33], where the methodological contributions therein do not presuppose consistency, and a weaker condition (A6) was sufficient. To the best of our knowledge, the agnostic consistency of $\hat{\Sigma}_{n,B_n}$ as an estimator, in an asynchronized, multivariate setting, is a new contribution to the literature.*

Remark 3 (Choice of the kernel function)**.** *A special class of kernel function is the Rectangular kernel:* $K^{Rec}(u) = I\{|u| \leq 1\}$*. This is a very classical and yet popular choice of kernel and dates back several decades in works of [16] and others. Note that* $K^{Rec} \notin \mathcal{C}^1$*. Nevertheless, almost the entire argument of Theorem 3.2 goes through to yield a bias of* $O(B_n^{-A})$ *where we recall* $A > A_0$ *is the decay exponent of* $\Theta_{i,p}$*. This rate is strictly better than that of (3.8). However, a major disadvantage of* K^{rec} *is that it is not a positive semi-definite kernel, and therefore it is not guaranteed that* $\hat{\Sigma}_{n,B_n} \succeq 0$*. On the other hand, the error rate (3.8) can be improved by assuming* $r > 1$ *continuous derivatives of* K*. A sweet spot, with regards to positive-definiteness and bias reduction, is advocated through the use of Splitted Rectangular Cosine kernels, as in [24, 26] etc. This idea is also related with the infinite-order flat top kernels, suggested by [91, 92, 90] and many others in the context of spectral density estimation. Some other choices include the Bartlett kernel and its convolutions. In view of such a huge literature, and in order not to divert too much from our main topic of discussion, we choose not to delve any deeper into the theory behind the appropriate choice of kernel function (and the corresponding bandwidth). Instead, we take this issue up empirically through some simulation exercises in Section 8.3.*

The proof of Theorem 3.2 is deferred to Section 11.3. A key insight into our proof is that, indifferent to the existence of change-points and even jump-sizes, the estimated mean vector $\hat{\boldsymbol{\mu}}_i$ in (3.6) will always be close, on an average, to the original mean vector $\boldsymbol{\mu}_i$. On first glance, this is not quite obvious, since, for a fixed $1 \leq j \leq d$, some algebra shows that with probability 1,

$$\max_{1 \leq i \leq n} |\hat{\mu}_{ij} - \mu_{ij}| \asymp \delta_j + O_{\mathbb{P}}(1/\sqrt{n}),$$

which can be large for larger jump-sizes. However, we show that the number of indices i on which this maximum occurs, decreases with increasing δ_j, and therefore on an average $|\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i|$ can be proven to be small. This is quantified in the following proposition.

Proposition 3.1. *Recall* $\hat{\boldsymbol{\mu}}_i$ *from (3.6). Then under uniformly for* $0 \leq k \leq n-1$ *and* $1 \leq j, l \leq d$*, it holds that*

$$\frac{1}{n}\sum_{i=1}^{n-k}(\hat{\mu}_{ij}-\mu_{ij})(\hat{\mu}_{i+k,l}-\mu_{i+k,l}) = O_{\mathbb{P}}(1/n), \tag{3.9}$$

where we have assumed (2.5) *and* (2.8) *for our error process* $(\mathbf{e}_t)$, *and* (1.1) *for* $(\mathbf{X}_t)$.

We emphasize that, agnostic to the location of change-points and size of the jump, Proposition 3.1 asserts that $\hat{\boldsymbol{\mu}}_i$ achieves the optimal rate of estimation. Of particular interest is the case, when the jump is small, or zero, which we briefly discuss here. In this case, Proposition 3.1 can be realized in the context of the well-known result, that when $\delta_j = 0$, $\hat{\tau}_j$ will be approximately distributed as $\arg\max_{t\in(0,1)} |\mathbb{B}^{br}(t)|$ where $\mathbb{B}^{br}(t)$ is a standard Brownian Bridge. Therefore, with high probability, the estimated change-point will lie towards the middle of the sequence $\{1,\ldots,n\}$, leading to the optimal rate that we observe in Proposition 3.1.

However, the proof of Proposition 3.1 does not require such asymptotic results for the case when δ_j is small. In particular, when $\delta_j = 0$, our proof assumes a dummy change-point $\tau_j \in (0,1)$, and shows that, the argument used for large δ_j also works for this case. This is, of course consistent with our notion of synchronization, where we have assumed $\tau_{j_1} = \tau_{j_2}$ if $\delta_{j_1} = \delta_{j_2} = 0$. The details of the proof can be found in the Section 11.3.

4 Bootstrap algorithm and theoretical validity

With the estimation of Σ_∞ dealt with, we now move towards describing our complete bootstrap algorithm. In order to conveniently establish the theoretical validity of our bootstrap procedure, we impose a condition on the jump-sizes of each dimension. Suppose $\mathcal{V}_0 = \{1 \le j \le d : \delta_j = 0\}$, and $\mathcal{V}_1 = \{1,\ldots,d\} \setminus \mathcal{V}_0$. We assume the following.

Assumption 4.1. *The jumps* δ_j *satisfy* $\min_{j:\delta_j\in\mathcal{V}_1} |\delta_j| \gg 1/\sqrt{n}$.

Assumption 4.1 resembles the well-known "beta-min" condition from high dimensional regression literature (see [78, 123]). As asserted by [25], such restrictive conditions on the minimum signal strength are necessary in order to achieve asymptotic validity of the corresponding procedure. In our context, it is important to briefly discuss the motivation behind such an assumption.

Along with $\hat{\Sigma}_{n,B_n}$, we aim to use $\hat{\tau}$, and $\hat{\delta}_j = \hat{\mu}_j^R - \hat{\mu}_j^R$ as a plug-in for τ and δ_j respectively, in the oracle algorithm 1. Note that, when $n\delta_j^2 \to \infty$, it can be shown that $|\hat{\delta}_j - \delta_j| = O_{\mathbb{P}}(1/\sqrt{n})$. However, for $\delta_j \ll 1/\sqrt{n}$, the estimate $\hat{\delta}_j$ can be quite large compared to δ_j. This is primarily because, for such a small size of jump, the CUSUM estimate is not enough accurate (cf. Proposition 2.1 entails a rate of only $O_{\mathbb{P}}(1)$). Therefore, it is clear that, for the validity of our procedure, if $|\delta_j|$ is small, we should draw our bootstrap samples $(Z_{1j}, \dots, Z_{nj})$ while pretending that $\delta_j = 0$. On the other hand, for $n\delta_j^2 \to \infty$, $\hat{\delta}_j$ works well enough from the sense of optimality. Importantly, in this case, $\hat{\tau}_j$ is very close to $\hat{\tau}$ under H_0, which ensures validity of our bootstrap procedure.

This immediately results in a thresholded/banded estimation procedure, where we estimate δ_j only if we know $n\delta_j^2 \to \infty$, and otherwise estimate δ_j by zero (see Step 3 of Algorithm 3). In practice, δ_j's would not usually be known, necessitating a "regularized" bootstrap procedure, whereby we first estimate $\mathcal{V}_0$ and $\mathcal{V}_1$. We undertake an individual level CUSUM test, and then conclude the $\delta_j = 0$ if the null hypothesis H_{0j} of existence of change-point in the j-th dimension is not rejected. This approach essentially determines the assignment of dimensions to sets $\hat{\mathcal{V}}_0$ and $\hat{\mathcal{V}}_1$ based on $\hat{\delta}_j I\{|\hat{\delta}_j| \gg 1/\sqrt{n}\}$. It can be interpreted as a "hard-thresholding" (eg. [27, 106]) of the naive estimator $\hat{\delta}_j$ of δ_j.

Therefore, as motivated above, we start off by testing for the existence of change-point for each individual dimension. The detailed procedure for this "hidden" first step of our main algorithm, is given in Algorithm 2. Following up, we briefly discuss the Algorithm 2 from a theoretical

---

Algorithm 2: Test of existence of change-point

---

1 **Input**: $\mathbf{X}_1, \dots, \mathbf{X}_n$, long-run variance estimate $\hat{\Sigma}_{n,B_n}$.

2 **Goal**: For each $j : 1, \dots, d$: Test $H_{0j} : \delta_j = 0$ vs H_{0j}^c.

- For $j = 1, \dots, d$: construct $U_{nj} := U_{nj}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ as in (4.1).
- For $s = 1, \dots, B$
  1. For $i = 1, \dots, n$, generate bootstrap samples $\mathbf{Z}_i^{(s)} \overset{\text{i.i.d}}{\sim} N(0, \hat{\Sigma}_{n,B_n})$.
  2. For $j = 1, \dots, d$: $U_{nj}^{(s)} \leftarrow U_{nj}(\mathbf{Z}_1^{(s)}, \dots, \mathbf{Z}_n^{(s)})$.
- For $j = 1, \dots, d$, $p_j \leftarrow \frac{1}{B+1}(\sum_{s=1}^{B} \mathbb{I}\{U_{nj} \le U_{nj}^{(s)}\} + 1)$.

---

perspective. Let us denote

$$U_{nj}(\mathbf{X}_1, \ldots, \mathbf{X}_n) := \max_{1 \le k \le n} |\sum_{i=1}^{k} (X_{ij} - \bar{X}_{\cdot j})| / \sqrt{n}. \tag{4.1}$$

For some $\alpha \in (0,1)$, observe that the B Monte Carlo bootstrap samples in Algorithm 2 are used essentially to estimate the $(1-\alpha)$-th quantile $a_{\alpha,j}(\hat{\Sigma}_{n,B_n})$ such that

$$a_{\alpha,j}(\Sigma_\infty) := \inf\{a : \mathbb{P}(U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) > a) \le \alpha\} \text{ for } \mathbf{Z}_1, \ldots, \mathbf{Z}_n \overset{\text{i.i.d}}{\sim} N(0, \Sigma_\infty).$$

The following result shows rigorously that under null, the test statistic U_{nj} cannot be too bigger than $a_{\alpha,j}(\hat{\Sigma}_{n,B_n})$ with high probability.

Proposition 4.1. *Assume the model (1.1) and the conditions of Theorem 3.1 for the stationary error process $(\mathbf{e}_i)$. Fix $\alpha \in (0,1)$ and $j \in \{1, \ldots, d\}$. If $c_n \to 0$, $v_n \to 0$ are chosen to be two deterministic positive sequence such that $c_n^2 \gg v_n^{-1}(B_n^{-1} + B_n n^{2/p'-1})$, with $p' = p \wedge 4$, then for U_{nj} as in Algorithm 2, under H_{0j} for every $1 \le j \le d$ it holds that,*

$$\overline{\lim_{n\to\infty}}\, \mathbb{P}(U_{nj} \ge a_{\alpha - v_n, j}(\hat{\Sigma}_{n,B_n}) + c_n) \le \alpha, \tag{4.2}$$

where $\hat{\Sigma}_{n,B_n}$ is constructed as in (3.7), satisfying the conditions of Theorem 3.2.

Proposition 4.1 allows us to confidently discern the sets $\mathcal{V}_0 := \{j : n\delta_j^2 \to 0\}$ and $\mathcal{V}_1 := \{j : n\delta_j^2 \to \infty\}$. In fact, this yields that $\mathbb{P}(\hat{\mathcal{V}}_1 \supseteq \mathcal{V}_1) \to 1$, and $\overline{\lim}\, \mathbb{P}(\hat{\mathcal{V}}_0 \supseteq \mathcal{V}_0) \ge 1 - \alpha$, as $n \to \infty$. With this premise, we now provide a complete algorithm for testing synchronization of change-points.

Since we have already established the validity of our Algorithm 2, it is reasonable to assume that $\mathcal{V}_0$ and $\mathcal{V}_1$ are known for subsequent analysis. We provide a theoretical analysis of showing the efficacy of the bootstrap-based quantile of Algorithm 3. Such a result also appears in [80], Section 4, to justify their bootstrap-based tests.

Theorem 4.1. *For the model (1.1), grant the conditions of Theorem 3.1 for the error process $\mathbf{e}_i$, and the conditions of Theorem 3.2 for the long-run covariance estimate $\hat{\Sigma}_{n,B_n}$. Further suppose that the sets $\mathcal{V}_0$ and $\mathcal{V}_1$ are known in Step 2 of Algorithm 3, and the Assumption 4.1 holds for all $1 \le j \le d$. For a general sequence of vectors $(\boldsymbol{\nu}_i)_{i=1}^n \in \mathbb{R}^d$, and a symmetric positive definite matrix Γ, let a generic Gaussian-based quantile $b_\alpha(\boldsymbol{\nu}, \Gamma)$ be defined as:*

$$b_\alpha(\boldsymbol{\nu}, \Gamma) = \inf\{b : \mathbb{P}(T_n(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \ge b) \le \alpha\},$$

Algorithm 3: Testing synchronization of change-points

1 **Input:** X, bootstrap size B, bandwidth b_n, level α. **Goal:** To test $H_0 : \tau_1 = \ldots = \tau_d$.

1. Construct T_n, $\hat{\tau}$ and $\hat{\Sigma}_{n,B_n}$ based on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ as in (2.3) and (3.7) respectively.
2. Use Algorithm 2 to obtain sets $\hat{\mathcal{V}}_0 := \{j : H_{0j} \text{ was not rejected}\}$, and $\hat{\mathcal{V}}_1 = \{j : H_{0j} \text{ was rejected}\}$.
3. For $s = 1, \ldots, B$,
   - Generate bootstrap samples $(\mathbf{Z}_i^{(s)})_{i=1}^n \overset{\text{i.i.d}}{\sim} N(0, \hat{\Sigma}_{n,B_n})$.
   - If $j \in \hat{\mathcal{V}}_0$: $X_{ij}^{(s)} \leftarrow Z_{ij}^{(s)} + \bar{X}_{\cdot j}$, $1 \le i \le n$.
   - If $j \in \hat{\mathcal{V}}_1$: $X_{ij}^{(s)} \leftarrow Z_{ij}^{(s)} + \frac{1}{n\hat{\tau}} \sum_{k=1}^{n\hat{\tau}} X_{kj} + (\frac{1}{n-n\hat{\tau}} \sum_{k=n\hat{\tau}+1}^{n} X_{kj} - \frac{1}{n\hat{\tau}} \sum_{k=1}^{n\hat{\tau}} X_{kj}) I\{i/n > \hat{\tau}\}$, $1 \le i \le n$.
   - Calculate $T_n^{(s)}$ based on $(\mathbf{X}_i^{(s)})_{i=1}^n$.
4. p-value: $p_0 \leftarrow \frac{1}{B+1}(\sum_{s=1}^{B} \mathbb{I}\{T_n \le T_n^{(s)}\} + 1)$.

where $\boldsymbol{Y}_i := \mathbf{Z}_i + \boldsymbol{\nu}_i$ *and* $\mathbf{Z}_i \overset{i.i.d}{\sim} N(0, \Gamma)$. *Recall* $\hat{\boldsymbol{\mu}}_i$ *from* (3.6). *Suppose* $\{u_n\}$, $\{h_n\}$ *are two positive deterministic sequences such that* $u_n \to 0$, $h_n \to 0$, *and*

$$u_n^2 \gg h_n^{-1}(B_n^{-1} + B_n n^{2/p'-1}) + h_n^{-2} \max_{j \in \mathcal{V}_1} 1/(n\delta_j^2). \tag{4.3}$$

If under H_0 *in* (1.3), $\hat{\tau}_j, \tau \in (c, 1-c)$ *almost surely for some* $0 < c < \tau \wedge 1/2$, *then, it holds that,*

$$\overline{\lim_{n\to\infty}} \, \mathbb{P}(T_n \ge b_{\alpha - h_n}(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) + u_n) \le \alpha, \tag{4.4}$$

for $\tilde{\boldsymbol{\mu}}_i = (\tilde{\mu}_{ij})_{j=1}^d$ *defined as*

$$\tilde{\mu}_{ij} = \begin{cases} \bar{X}_{\cdot j}, \ j \in \mathcal{V}_0, \ 1 \le i \le n \\ \frac{1}{n\hat{\tau}} \sum_{k=1}^{n\hat{\tau}} X_{kj} I\{1 \le i \le n\hat{\tau}\} + \frac{1}{n-n\hat{\tau}} \sum_{k=n\hat{\tau}+1}^{n} X_{kj} I\{n\hat{\tau}+1 \le i \le n\}, \ j \in \mathcal{V}_1. \end{cases}$$

It is instructive to briefly discuss the rather technical condition (4.3) in u_n and h_n. It can be noted that $B_n^{-1} + B_n n^{2/p'-1} \gtrsim n^{1/p'-1/2}$, with $p' \in (2, 4]$ for all choices of B_n. Therefore, for the "strong signal" setting with $\min_{j \in \mathcal{V}_1} |\delta_j| \gg n^{-1/p'}$, a choice satisfying (4.3) is $u_n \asymp h_n \asymp 1/\log n$, and $B_n \asymp n^{1/2-1/p'}$. In particular, this includes the setting where δ_j's are constant. Note that, with this particular choice of u_n and h_n, and for all sufficiently large n, the above restriction on B_n can be generalized to $(\log n)^3 \ll B_n \ll n^{1-2/p'} (\log n)^{-3}$. On the other hand, for the complementary setting with weaker signal strength, a choice of u_n and h_n will crucially depend on $\min_{j \in \mathcal{V}_1} |\delta_j|$. If $c_n := \min_{j \in \mathcal{V}_1} |\delta_j|$ with $n^{-1/2} \ll c_n \ll n^{-1/p'}$, then a conservative choice is given by $u_n \asymp h_n \asymp 1/\log(\sqrt{n} c_n)$ along with $(\log \sqrt{n} c_n)^3 \ll B_n \ll n^{1-2/p'}(\log \sqrt{n} c_n)^{-3}$.

With the validity of the bootstrap-based test of synchronization established, it is important to look at the aspect of power of these tests. In the simulation studies of Section 8, (a summary is provided below), we will see that, in practice, the sizes of these tests not only achieve the level of significance, but they also produce much power under various alternatives. In fact, the following theorem shows that asymptotically, the power of the bootstrap-based test approaches 1 under a general class of alternatives. A finer, finite sample analyses of sizes and powers of these tests would require a case-by-case treatment, and are out of the scope of this paper.

Theorem 4.2. *Recall the notation used in Theorem 4.1. In addition, suppose that $\hat{\tau} \in (c, 1-c)$ almost surely for some $c \in (0, 1/2)$. If (2.9) is satisfied, then $\underline{\lim}_{n\to\infty} \mathbb{P}(T_n \geq b_\alpha(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n})) = 1$.*

The proofs of Proposition 4.1, Theorem 4.1 and Theorem 4.2 are provided in Section 12. Our bootstrap-based results requires an additional restriction on $\hat{\tau}$. However, this condition is fairly mild, and can be guaranteed by restricting the search space for $\hat{\tau}$ in (2.4) to $nc \leq i \leq n(1-c)$ for some small but fixed constant $c > 0$. Boundary removal techniques have been a common feature in change-point literature. [1] suggests using a restricted interval when no knowledge of the change-point is available. The specific form $[nc, n(1-c)]$ of the search space has appeared at least as early as [9, 4] and [37]. The main concern behind such restriction is that CUSUM behaves with volatility when the candidate change-point k is very close to 0 or 1- this is also the primary technical challenge that ensures the imposition of this restriction upon our bootstrap analysis. In the context of bootstrap, such conditions are also present in more recent works such as [100, 118] and [119], among others. Finally, we emphasize that this condition is only required for theoretical validity and consistency of the bootstrap-based test; for the other general results of this paper, this restriction is not needed.

5 Test of synchronization in high dimension

As a natural generalization of multivariate change-point detection, we also investigate the problem of synchronization for high-dimensional stationary time series. Recently, [117] has dealt with estimation of asynchronous change-points, but the testing problem is not addressed. For this test, we will use the same test statistic that we used in the multivariate set-up. Our subsequent analysis and discussion will assume $d \leq cn$ for some constant $c > 0$. We impose an additional condition on the true locations of the change-points.

Assumption 5.1 (Non-vanishing change-points)**.** *There exists a constant* $c \in (0, 1)$*, such that* $c \leq \min_j \tau_j \leq \max_j \tau_j \leq 1 - c$*.*

Assumption 5.1 essentially guards against the complications arising out of having some coordinates with change-points too close to 0 or 1. Note that it trivially holds under null of synchronization. This can be thought of as a high-dimensional generalization of the ubiquitous assumption of change-point away from the boundary. It can be verified that our results hold for a general set of change-points with some routine modifications. To characterize the behavior of our test statistic, we will define some more notations pertaining to the dependence structure in high-dimension. For $q \geq 2$, $r \geq 2$, let

$$\gamma_{i,q,r} = \| |\mathbf{e}_i - \mathbf{e}_{i,\{0\}}|_{\mathcal{L}_r} \|_q ; \Delta_{q,r,\alpha} := \sup_{m \geq 0} (m+1)^\alpha \Gamma_{m,q,r}, \alpha \geq 0, \text{ where } \Gamma_{m,q,r} = \sum_{i=m}^{\infty} \gamma_{i,q,r}, m \geq 0. \tag{5.1}$$

The following results quantify the oracle behavior of our test statistic. In particular, corresponding to Propositions 2.1 and 2.2, the following two results establish (1) the consistency of the common estimate of change-point under null, and (2) the validity and power of the test statistic under the null and the alternate. Both Propositions 5.1 and 5.2 accommodate Propositions 2.1 and 2.2 as special cases; however the proofs get more convoluted due to the dimension d increasing with n. Thus, for the reader's convenience, we present separate proofs of these results building on the corresponding results for the multivariate set-up. These proofs are provided in Section 13.

Proposition 5.1. *Grant model (1.1) for $\mathbf{X}_t$ with the error process $\mathbf{e}_t$ satisfying (2.5). If H_0 : $\tau_1 = \ldots = \tau_d := \tau$ is true, then it holds that*

$$|\hat{\tau} - \tau| = O_{\mathbb{P}}(d^3(n(\sum_{1\le j\le d} |\delta_j|)^2)^{-1}\Delta^2_{p,2,\alpha}). \tag{5.2}$$

Proposition 5.2. *Assume the conditions in Proposition 5.1, and further grant Assumption 5.1. Under the synchronized setting, i.e. under H_0 described in (1.3), $T_n d^{-5/2}\Delta^{-1}_{p,2,\alpha} = O_{\mathbb{P}}(1)$.*

For the behavior under H_0^c, let $\mathcal{J} = \{h_1, h_2, \ldots, h_{|\mathcal{J}|}\}$ be the set of unique change-points, $|\mathcal{J}| \ge 2$, with $D_h := \{1 \le j \le d : \tau_j = h\}$ denoting the coordinates with change-point at $h \in \mathcal{J}$. If the set $\mathcal{J}$ is ordered so that $Q_i := \sum_{j\in D_{h_i}} |\delta_j|$ follows $Q_1 \ge Q_2 \ldots \ge Q_{|\mathcal{J}|} > 0$. Further, if

$$\mathcal{V}(n, d) := \sqrt{n} d^{-5/2}\Delta^{-1}_{p,2,\alpha} \sum_{i=2}^{|\mathcal{J}|} Q_i|h_i - h_{i-1}| \to \infty, \tag{5.3}$$

holds, then $U_n \xrightarrow{\mathbb{P}} \infty$. Here, note that for any $h \in \mathcal{J}$, $j \in \mathcal{D}_h$, it holds $\delta_j^2 > 0$.

Remark 4. *In the result under H_0, the normalizing quantity $d^{-5/2}\Delta^{-1}_{p,2,\alpha}$ depends on the process $(\mathbf{e}_i)_{i\in\mathbb{Z}}$, and will generally be unknown. However, we pursue a bootstrap-based test, and hence, normalizing quantities will not affect the validity or power of the test. Therefore, we can simply perform the test with T_n.*

Remark 5. *The condition (5.3) is worth discussing, especially in light of the corresponding condition (2.9) for the multivariate case. In a high-dimension synchronization problem, clustering of change-points is an innate feature affecting the rate- more so than in the multivariate set-up, where this feature does not show up in the rates. As an example, suppose under alternate there are only two distinct change-points at τ_1 and τ_2. Then the condition (5.3) simplifies to*

$$\sqrt{n} d^{-5/2}\Delta^{-1}_{p,2,\alpha}(\sum_{j\in D_1} |\delta_j| \wedge \sum_{j\in D_2} |\delta_j|) \to \infty.$$

In particular, if all $|\delta_j| \asymp \delta$ and $|D_1| \asymp |D_2| \asymp d$, then we require

$$\sqrt{n} d^{-3/2}\Delta^{-1}_{p,2,\alpha}\delta \to \infty, \tag{5.4}$$

for power of our test to approach 1*. On the other hand, if under alternate* $|\mathcal{J}| = d$ *and* $\delta_j \asymp \delta$ *for all* $j \in [d]$*, then the condition* (5.3) *condenses to* $\sqrt{n}d^{-5/2}\Delta_{p,2,\alpha}^{-1}\delta(\max_j \tau_j - \min_j \tau_j) \to \infty$*. This implies that if* $\delta = O(1)$*,* $\sqrt{n}d^{-5/2}\Delta_{p,2,\alpha}^{-1} \to \infty$ *is a necessary condition for power to approach* 1 *under the latter case. This condition reflects a curious hardness of this particular alternative. In a high-dimensional problem, when all the change-points are distinct, there will be dimensions where the change-point estimates are quite close to each other, purely due to randomness. Since each coordinate corresponds to one unique change-point, the possible clustering of estimates implies a slightly harder problem, leading to a more stringent condition for a consistent test. On the other extreme, if* $|D_1| \asymp |D_2| \asymp d$*, information of these two change-points are pulled from all dimensions, and even with some spurious estimates, we recover the two change-points extremely well, reflected by the weaker condition* (5.4)*.*

5.1 Gaussian Approximation for high-dimensional time-series

This section discusses establishing a high-dimensional counterpart of the Gaussian approximation result of Section 3.1. Our result is inspired heavily from [80], but has somewhat lesser conditions than them. We will describe the theorem and the adjoining notations formally. As before, the stationary process $\mathbf{e}_i$ is modeled as (2.5). Moreover, recalling (5.1), much like the decay condition in Theorem 3.1, we will assume that there exist $\Theta, \beta > 0$, and a power $p \geq 2$, such that for all t,

$$\gamma_{j,p,2} \leq \Theta \cdot (j+1)^{-\beta}, j \geq 0 \text{ and } \|\mathbf{e}_0\| \leq \Theta. \tag{5.5}$$

As [80] mentions, if all d components are iid, then we expect $\Theta \approx \sqrt{d}$. For $q \geq 2$, define,

$$\xi(q,\beta) = (1+c_1)(\frac{1}{4} - \frac{1}{2q}), \text{ where } c_1 = \frac{\frac{1}{2q} - \frac{1}{4}}{\frac{1}{4} - \frac{1}{2q} - \frac{1-\beta}{2} \vee (-\frac{1}{2})}.$$

The quantity ξ essentially pins down our Gaussian approximation rate. Note that, this rate is sharper compared to the one in Theorem 3.1 in [80]. Moreover, following [79], we also relax $\beta > 1$ from [80]'s condition of $\beta > 2$. However, some of the crucial arguments behind our result, which are stated below, follow from [80]. The detailed proof is provided in Section 13.

Theorem 5.1. *For $1 \leq i \leq n$, let $\mathbf{e}_i$ with $\mathbb{E}(\mathbf{e}_i) = 0$ satisfy (2.5), and (5.5) for some $p > 2$ and $\beta > 1$, and suppose $d \leq cn$ for some $c > 0$. Then, on a potentially different probability space, there exist random vectors $(\mathbf{e}'_t)_{t=1}^n \stackrel{d}{=} (\mathbf{e}_t)_{t=1}^n$ and independent, mean zero, Gaussian random vectors $Y_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \Sigma_\infty)$ such that*

$$\left(\mathbb{E}\max_{k\leq n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}(X'_t - Y'_t)\right|^2\right)^{\frac{1}{2}} \leq C\Theta\sqrt{\log(n)}\left(\frac{d}{n}\right)^{\xi(p,\beta)} := \daleth_1(d,n). \tag{5.6}$$

Remark 6. *Compared to [80], we only need $\beta > 1$, an improvement from $\beta > 2$ therein. In fact, for a fixed β, our approximation rates are better. Moreover, due to the stationarity, we no longer require their condition (G.2) to accommodate non-stationarity. Finally, compared to our low-dimensional Gaussian approximation result Theorem 3.1, we do not need $\lambda_{\min}(\Sigma_\infty) \geq c$ or $A > A_0$ therein, as $A > 0$ (or equivalently $\beta > 1$) suffices. As a trade-off, albeit in low-dimension, we have rate-optimality in Theorem 3.1.*

Now, note that the proof of Lemma 3.1 can be adapted to the Theorem 5.1 to obtain

$$d^{-5/2}\Delta_{q,2,\alpha}^{-1}|T_n - T_n^Y| = O_{\mathbb{P}}(\sqrt{\log n}\, d^{\xi(p,\beta)-3/2}n^{-\xi(p,\beta)}) = o_{\mathbb{P}}(1), \text{ if } \Theta \asymp \sqrt{d}, \tag{5.7}$$

for all combinations of $(n,d) = (n,d_n)$ as $n \to \infty$. Equation (5.7) indicates the effectiveness of our bootstrap strategy even in high-dimension.

5.2 Estimation of long-run covariance matrix and bootstrap

In order to apply Theorem 5.1 to obtain a bootstrap algorithm, we are required to estimate Σ_∞. We will define a few new notations which will allow us to explicitly characterize the estimation error. For $1 \leq j \leq d$ and $r \geq 2$, let

$$\theta_{i,r,j} = \|e_{ij} - e_{ij,\{0\}}\|_r;\ \|\mathbf{e}_{\cdot j}\|_{r,\alpha} = \sup_{m\geq 0}(m+1)^\alpha\Theta_{m,r,j}, \alpha \geq 0,\ \text{where } \Theta_{m,r,j} = \sum_{i=m}^{\infty}\theta_{i,r,j}, m \geq 0,$$

be the coordinate-wise functional dependence measures and dependence-adjusted norms, respectively. In order to tackle high-dimensionality, we are further required to define

$$\Psi_{q,\alpha} = \max_{1\le j\le d} \|\mathbf{e}_{.j}\|_{q,\alpha}, \text{ and } \Upsilon_{q,\alpha} = \left(\sum_{j=1}^{d} \|\mathbf{e}_{.j}\|_{q,\alpha}^{q/2}\right)^{2/q} \le d^{2/q}\Psi_{q,\alpha},\ q>2.$$

For example, when all d dimensions are i.i.d, and $\theta_{i,r,j} = O(j^{-\beta})$ for each $1\le j\le d$ and $\beta > \alpha+1$, then for a fixed q, it is easy to show that $\Delta_{q,r,\alpha} = O(d^{1/r})$, $\Psi_{q,\alpha} = O(1)$ and $\Upsilon_{q,\alpha} = O(d^{2/q})$. For more details on the relations between these norms and $\Gamma_{m,q,r}$ defined in (5.1), we refer the readers to [120, 121, 80]. Mimicking the low-dimensional setting in Section 3.2, we use the same estimate, and establish that Σ_∞ is estimated agnostically even for an high-dimensional setting.

Theorem 5.2. *Grant Assumption 5.1 as well as the assumptions of Theorem 3.2 along with* $\Upsilon_{p,\alpha} < \infty$ *for some* $p > 4$ *and* $\alpha > 0$. *Consider the estimate* $\hat{\Sigma}_{n,B_n}$ *in* (3.7), *with* $B_n = \lfloor n^\beta \rfloor$ *for some* $\beta < \min\{1-4p^{-1}, \alpha p 2^{-1}\}$.

$$|\hat{\Sigma}_{n,B_n} - \Sigma_\infty|_\infty = O_{\mathbb{P}}(\daleth_2(d,n))$$
$$\text{where } \daleth_2(d,n) = B_n\left(\frac{(\log d)^2\Delta_{p,\infty,\alpha}^2 + \sqrt{\log dn}\,\Psi_{4,\alpha}^2}{\sqrt{n}} + \frac{1}{n}\Upsilon_{4,\alpha} + \frac{1+B_n^{-\alpha+1}}{n}\Psi_{2,0}\Psi_{2,\alpha}\right). \quad (5.8)$$

If V_0 and V_1 are known, Theorems 5.1 and 5.2 can be applied to yield results similar to Theorems 4.1 and 4.2. This result is included in Section 14, where we also include a discussion on the practical implementation of the algorithm in high-dimension.

6 Applications: Simulation and Real Data analyses

Due to space constraints, here we discuss a brief summary of the simulation studies, and two interesting real-life applications. The details are relegated to Section 8 and Section 9 respectively.

6.1 Simulation studies (Summary)

Section 8.1 explores the distribution of T_n under different synchronized settings in multivariate set-ups. We focus on identifying the effect of jump-sizes δ on the distribution of T_n, and work with a relatively simple VAR model for the stationary errors $\mathbf{e}_i$'s. Increasing the jump-size δ compels T_n to converge towards 0- a phenomena discussed in more detail in Section 8.1. Working under the same setting, in Section 8.2 we move on to numerically inspect the efficacy of our Gaussian approximation result by looking at how the finite sample distributions of the $T_n^{(s)}$'s from the oracle bootstrap Algorithm 1 compare with the null distributions of T_n.

The set of simulations in Section 8.3 aims to numerically showcase the effect of bandwidths B_n and choice of kernel functions K on the estimation accuracy of $\hat{\Sigma}_{n,B_n}$. Here, we see that the choice of kernel does not seem to hugely affect the performance of $\hat{\Sigma}_{n,B_n}$, as long as the choice of B_n is restricted to $\lfloor n^{1/4} \rfloor$ between $\lfloor n^{1/3} \rfloor$. Finally, in Section 8.4, we take up two non-linear and yet popular models for the stationary error processes: a TAR model, and a GJR-GARCH model. For both the models, we compute the empirical type-1 errors and powers of our bootstrap-based Algorithm 3 under null (synchronized) and various alternative (asynchronized) scenarios. See Tables 3, 4 and 5. These simulations highlight that testing procedure via Algorithm 3 maintains empirical size close to the nominal level and yet achieves high power even in the "difficult" scenarios of (1) asynchronized change-points being relatively close to each other, as well as (2) the jumps corresponding to the change-points being small. We also provide additional simulation studies (Tables 6 and 7) to highlight some behaviors unique to the high-dimensional set-up.

6.2 Real data analyses (Summary)

We now summarize findings for two real-world datasets. In the first one, (See Figure 1) we test for synchronization of two time series of water discharge of Mississippi river in two spatial locations. Although it is hard to discern in naked eyes, we statistically reject synchronization. Towards explaining this behavior, we find that one of the location is further downstream than the other location, resulting in a crucial delay in flood onset- a delay that can help with policy designing for flash floods. See the details in Section 9.1.

In our second analysis, we analyze the data on cardio-respiratory response of a pilot. Monitoring three time series on their heart rate (HR), partial pressure of end-tidal CO_2 (petCO_2) and respi-

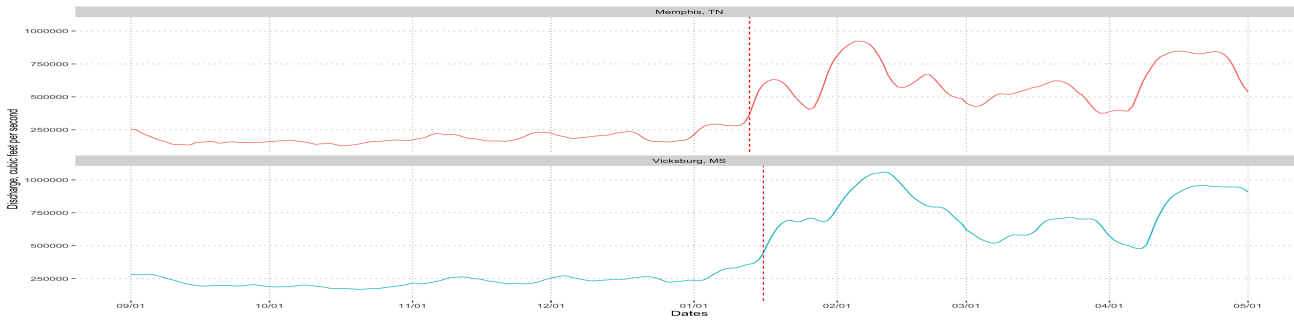

Figure 1: Water discharge data of Mississippi river at two different locations from Sept'23 to May'24. The vertical red line indicates the individual change points detected by CUSUM.

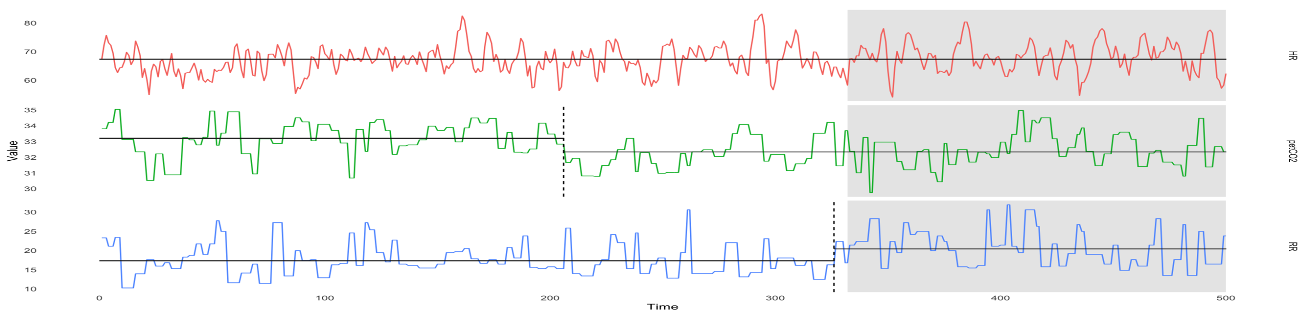

Figure 2: Time series plot for early stage (Time 1-500s) of pilot mental load dataset.The red, green and blue plots indicate the time series corresponding to heart rate, PET CO_2, and respiratory rate respectively. The white and shaded regions indicate 'Resting Baseline' phase and 'Vanilla Baseline' phase. The black dashed lines mark the estimated change-points by our method, and the solid horizontal black lines denote the estimated means of the corresponding segments.

ratory rate (RR), (See Figure 2) at early stage of the experiment, we obtain a pvalue of 0.0362 implying asynchronized change-point. We also estimate no breaks in HR, and two seemingly distant change-points in the other two and provide some reasoning why this is the case. All details can be found in Section 9.2 along with a follow-up analyses for a later stage of the experiment.

7 Conclusion

In the literature of change-point analyses of multiple time series, it is almost unanimously assumed that all coordinates exhibit the change-points simultaneously at the same time-stamp. Citing reasons and motivations why this might be too restrictive, in this paper, we propose a statistical test for this synchronization assumption. Although, we discuss only the synchronization of means, our methods are general enough to perform similar testing for other moments such as variance, correlations, kurtosis etc. In the financial econometrics literature, volatility plays a crucial role, and our method could be instrumental to test whether multiple stocks/indices show similar changes in their (possibly estimated) volatility. Moreover, sometimes irregularity in time series can be observed due to the errors being non-stationary. This can be easily handled by a uniform notion of functional dependence measure and using suitable Gaussian approximation such as [62, 23]. Since this does not require any technical novelty, we decided to restrict ourselves to a stationary case here. It is important to reemphasize that the final observed process can potentially be non-stationary when change-points occur at different times for different components.

Finally, one natural extension could be testing for synchronization when multiple change-points could be present in the multivariate stream. Note that, when we reject the synchronization hypothesis, it could be automatically thought of as if we have detected multiple change-points, as the ones not synchronized with others form another change-point. Therefore, our problem faces some identifiability issues if we allow for more than one change-point in each coordinate, as one has to put a restriction that change-point time-stamp in any of the co ordinates has to differ from each other by a significant time-delay. If such an assumption is made, we could possibly replace our CUSUM based method by moving sum (MOSUM) restricting to a window with at-most one change-point, and execute a similar statistical test.

References

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

[2] Donald W. K. Andrews and J. Christopher Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966, 1992.

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, page 3, 2013.

[4] Jaromír Antoch and Marie Hu ˇ ková. Change-point problem and bootstrap. *J. Nonparametr. Statist.*, 5(2):123–144, 1995.

[5] Manuel Arellano. *Panel data econometrics*. OUP Oxford, 2003.

[6] Alessandro Attanasio, Antonello Pasini, and Umberto Triacca. Granger causality analyses for climatic attribution. *Atmospheric and Climate Sciences*, 2013, 2013.

[7] Alexander Aue and Lajos Horváth. Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16, 2013.

[8] Mohammad Taha Bahadori and Yan Liu. Granger causality analysis with hidden variables in climate science applications. In *Climate Informatics workshop (CI 2011)*. Citeseer, 2011.

[9] Jushan Bai. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.*, 15(5):453–472, 1994.

[10] Jushan Bai. Estimating multiple breaks one at a time. *Econometric Theory*, 13(3):315–352, 1997.

[11] Jushan Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, 1997.

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

[13] Jushan Bai, Robin L. Lumsdaine, and James H. Stock. Testing for and dating common breaks in multivariate time series. *Rev. Econom. Stud.*, 65(3):395–432, 1998.

[14] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

[15] Matteo Barigozzi, Haeran Cho, and Piotr Fryzlewicz. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometrics*, 206(1):187–225, 2018.

[16] M. S. Bartlett. On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl. J. Roy. Statist. Soc.*, 8:27–41, 1946.

[17] Michèle Basseville and Igor V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice Hall Information and System Sciences Series. Prentice Hall, Inc., 1993.

[18] István Berkes, Edit Gombay, and Lajos Horváth. Testing for changes in the covariance structure of linear processes. *J. Statist. Plann. Inference*, 139(6):2044–2063, 2009.

[19] István Berkes, Weidong Liu, and Wei Biao Wu. Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 2014.

[20] Eric Beutner, Wei Biao Wu, and Henryk Zähle. Asymptotics for statistical functionals of long-memory sequences. *Stochastic Process. Appl.*, 122(3):910–929, 2012.

[21] Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Common change point estimation in panel data from the least squares and maximum likelihood viewpoints. *arXiv preprint arXiv:1708.05836*, 2017.

[22] Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in panel data with temporal and cross-sectional dependence. *arXiv preprint arXiv:1904.11101*, 2019.

[23] Soham Bonnerjee, Sayar Karmakar, and Wei Biao Wu. Gaussian approximation for nonstationary time series with optimal rate and explicit construction. *Ann. Statist.*, 52(5):2293–2317, 2024.

[24] Peter Bühlmann. Locally adaptive lag-window spectral estimation. *J. Time Ser. Anal.*, 17(3):247–270, 1996.

[25] Peter Bühlmann. Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242, 2013.

[26] Peter Bühlmann and Hans R. Künsch. Block length selection in the bootstrap for time series. *Comp. Statist. D. Anal.*, 31(3):295–310, 1999.

[27] Peter Bühlmann and Sara van de Geer. *Statistics for high-dimensional data*. Springer Series in Statistics. Springer, Heidelberg, 2011. Methods, theory and applications.

[28] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability*, 1:19–42, 1973.

[29] Jedelyn Cabrieto, Kristof Meers, Evelien Schat, Janne Adolf, Peter Kuppens, Francis Tuerlinckx, and Eva Ceulemans. `kcpRS`: An r package for performing kernel change point detection on the running statistics of multivariate time series. *Behavior Research Methods*, pages 1–22, 2022.

[30] Jedelyn Cabrieto, Francis Tuerlinckx, Peter Kuppens, Mariel Grassmann, and Eva Ceulemans. Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior research methods*, 49:988–1005, 2017.

[31] K. S. Chan, Joseph D. Petruccelli, H. Tong, and Samuel W. Woolford. A multiple-threshold AR(1) model. *J. Appl. Probab.*, 22(2):267–279, 1985.

[32] Xiaohui Chen, Mengyu Xu, and Wei Biao Wu. Covariance and precision matrix estimation for

high-dimensional time series. *Ann. Statist.*, 41(6):2994–3021, 2013.
[33] Haeran Cho. Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.*, 10(2):2000–2038, 2016.
[34] Gregory C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28:591–605, 1960.
[35] George W. Cobb. The problem of the nile: Conditional solution to a changepoint problem. *Biometrika*, 65(2):243–251, 1978.
[36] Robert Coben and Iman Mohammad-Rezazadeh. Neural connectivity in epilepsy as measured by granger causality. *Frontiers in human neuroscience*, 9:194, 2015.
[37] Miklós Csörgő and Lajos Horváth. *Limit theorems in change-point analysis.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1997.
[38] Holger Dette and Dominik Wied. Detecting relevant changes in time series models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(2):371–394, 2016.
[39] Mingzhou Ding, Yonghong Chen, and Steven L. Bressler. Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis: recent theoretical developments and applications*, pages 437–460, 2006.
[40] James B. Elsner, Xufeng Niu, and Thomas H. Jagger. Detecting shifts in hurricane rates using a markov chain monte carlo approach. *Journal of Climate*, 17(13):2652–2666, 2004.
[41] John F/ England Jr, Timothy A. Cohn, Beth A. Faber, Jery R Stedinger, Wilbert O. Thomas Jr, Andrea G. Veilleux, Julie E. Kiang, and Robert R. Mason Jr. Guidelines for determining flood flow frequency—bulletin 17c. Technical report, US Geological Survey, 2019.
[42] D. Kh. Fuk and S. V. Nagaev. Probability inequalities for sums of independent random variables. *Theory of Probability & Its Applications*, 16(4):643–660, 1971.
[43] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.
[44] F. Götze and A. Yu. Zaitsev. Bounds for the rate of strong approximation in the multidimensional invariance principle. *Theory of Probability & Its Applications*, 53(1):59–80, 2009.
[45] Clive W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, pages 424–438, 1969.
[46] Mariel Grassmann, Elke Vlemincx, Andreas von Leupoldt, and Omer Van den Bergh. The role of respiratory measures to assess mental load in pilot selection. *Ergonomics*, 59(6):745–753, 2016.
[47] J. Hájek and A. Rényi. Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.*, 6:281–283, 1955.
[48] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, Inc., New York-London, 1980.
[49] David V. Hinkley. Inference about the change-point in a sequence of random variables. 1970.
[50] David V. Hinkley and Edna Schechtman. Conditional bootstrap methods in the mean-shift model. *Biometrika*, 74(1):85–93, 1987.
[51] Lajos Horváth and Marie Hušková. Change-point detection in panel data. *J. Time Series Anal.*, 33(4):631–648, 2012.
[52] Lajos Horváth, Piotr Kokoszka, and Josef Steinebach. Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.*, 68(1):96–119, 1999.
[53] Cheng Hsiao. *Analysis of panel data.* Number 64. Cambridge university press, 2003.
[54] Marie Hu ˇ sková and Claudia Kirch. A note on Studentized confidence intervals for the change-point. *Comput. Statist.*, 25(2):269–289, 2010.
[55] Marie Hušková. Tests and estimators for the change point problem based on M-statistics. *Statist. Decisions*, 14(2):115–136, 1996.

[56] Dorcas Idowu and Wendy Zhou. Spatiotemporal evaluation of flood potential indices for watershed flood prediction in the mississippi river basin, usa. *Environmental & Engineering Geoscience*, 27(3):319–330, 2021.

[57] J. Richard Jennings, Thomas Kamarck, Christopher Stewart, Michael Eddy, and Paul Johnson. Alternate cardiovascular baseline assessment techniques: Vanilla or resting baseline. *Psychophysiology*, 29(6):742–750, 1992.

[58] Moritz Jirak. Uniform change point tests in high dimension. *Ann. Statist.*, 43(6):2451–2483, 2015.

[59] Lawrence Joseph and David B. Wolfson. Estimation in multi-path change-point problems. *Communications in Statistics-Theory and Methods*, 21(4):897–913, 1992.

[60] Lawrence Joseph and David B. Wolfson. Maximum likelihood estimation in the multi-path change-point problem. *Ann. Inst. Statist. Math.*, 45(3):511–530, 1993.

[61] Lawrence Joseph, David B. Wolfson, Roxane du Berger, and Roseann M. Lyle. Analysis of panel data with change-points. *Statistica Sinica*, pages 687–703, 1997.

[62] Sayar Karmakar and Wei Biao Wu. Optimal gaussian approximation for multiple time series. *Statist. Sinica*, 30(3):1399–1417, 2020.

[63] Abhishek Kaul, Stergios B. Fotopoulos, Venkata K. Jandhyala, and Abolfazl Safikhani. Inference on the change point under a high dimensional sparse mean shift. *Electron. J. Stat.*, 15(1):71–134, 2021.

[64] Dukpa Kim. Common breaks in time trends for large panel data with a factor structure. *Econom. J.*, 17(3):301–337, 2014.

[65] Piotr Kokoszka and Remigijus Leipus. Change-point estimation in ARCH models. *Bernoulli*, 6(3):513–539, 2000.

[66] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.

[67] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976.

[68] Walter Krämer, Werner Ploberger, and Raimund Alt. Testing for structural change in dynamic models. *Econometrica*, 56(6):1355–1369, 1988.

[69] Zbigniew W. Kundzewicz and Alice J. Robson. Change detection in hydrological records—a review of the methodology/revue méthodologique de la détection de changements dans les chroniques hydrologiques. *Hydrological Sciences Journal*, 49(1):7–19, 2004.

[70] Marc Lavielle and Eric Moulines. Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.*, 21(1):33–59, 2000.

[71] Degui Li, Junhui Qian, and Liangjun Su. Panel data models with interactive fixed effects and multiple structural breaks. *J. Amer. Statist. Assoc.*, 111(516):1804–1819, 2016.

[72] Jiaqi Li, Likai Chen, Weining Wang, and Wei Biao Wu. $\ell 2$ inference for change points in high-dimensional time series via a Two-Way MOSUM. *Ann. Statist.*, 52(2):602–627, 2024.

[73] Weidong Liu and Zhengyan Lin. Strong approximation for a class of stationary processes. *Stochastic Process. Appl.*, 119(1):249–280, 2009.

[74] Weidong Liu and Wei Biao Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4):1218–1245, 2010.

[75] Mariusz Maziarz. A review of the granger-causality fallacy. *The journal of philosophical economics: Reflections on economic and social issues*, 8(2):86–105, 2015.

[76] Gregory J. McCabe and David M. Wolock. A step increase in streamflow in the conterminous united states. *Geophysical Research Letters*, 29(24):38–1, 2002.

[77] Marie C McGraw and Elizabeth A. Barnes. Memory matters: A case for granger causality in climate variability studies. *Journal of Climate*, 31(8):3289–3300, 2018.

[78] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

[79] Fabian Mies. Strong gaussian approximations with random multipliers. *arXiv preprint arXiv:2412.14346*, 2024.

[80] Fabian Mies and Ansgar Steland. Sequential gaussian approximation for nonstationary time series in high dimensions. *Bernoulli*, 29(4):3114–3140, 2023.

[81] S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.*, 7(5):745–789, 1979.

[82] Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

[83] Jeffrey A. Nittrouer, John Shaw, Michael P. Lamb, and David Mohrig. Spatial and temporal trends for water-flow velocity and bed-material sediment transport in the lower mississippi river. *Bulletin*, 124(3-4):400–414, 2012.

[84] Ewan S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[85] Ewan S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527, 1955.

[86] Sandip Pal, Temple R. Lee, and Nicholas E. Clark. The 2019 mississippi and missouri river flooding and its impact on atmospheric boundary layer dynamics. *Geophysical Research Letters*, 47(6):e2019GL086933, 2020.

[87] Emanuel Parzen. On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.*, 28:329–348, 1957.

[88] A. N. Pettitt. A nonparametric approach to the change-point problem. *J. Roy. Statist. Soc. Ser. C*, 28(2):126–135, 1979.

[89] Milana Pokorná, Emanuel Nečas, Jaroslav Kratochvíl, Roman Skřipskỳ, Michal Andrlík, and Ondrej Franěk. A sudden increase in partial pressure end-tidal carbon dioxide (petco2) at the moment of return of spontaneous circulation. *The Journal of emergency medicine*, 38(5):614–621, 2010.

[90] Dimitris N. Politis. Adaptive bandwidth choice. *J. Nonparametr. Stat.*, 15(4-5):517–533, 2003.

[91] Dimitris N. Politis and Joseph P. Romano. Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.*, 16(1):67–103, 1995.

[92] Dimitris N. Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

[93] Associated Press. Flood damage will cost more than $2 billion for mississippi river towns. *Chicago Tribune*.

[94] Philip Preuss, Ruprecht Puchstein, and Holger Dette. Detection of multiple structural breaks in multivariate time series. *J. Amer. Statist. Assoc.*, 110(510):654–668, 2015.

[95] M. B. Priestley. *Spectral analysis and time series. Vol. 1.* Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1981. Univariate series.

[96] Haskell P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303, 1970.

[97] Pranab Kumar Sen. Tests for change-points based on recursive U-statistics. *Comm. Statist. C—Sequential Anal.*, 1(4):263–284, 1982/83.

[98] Anil K. Seth. Granger causality. *Scholarpedia*, 2(7):1667, 2007.

[99] SA Shaban. Change point problem and two-phase regression: an annotated bibliography. *International Statistical Review/Revue Internationale de Statistique*, pages 83–93, 1980.

[100] Xuesheng Shi, Colin Gallagher, Robert Lund, and Rebecca Killick. A comparison of single and multiple changepoint techniques for time series data. *Comput. Statist. Data Anal.*, 170:Paper No. 107433, 19, 2022.

[101] Ali Shojaie and Emily B Fox. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9:289–319, 2022.

[102] Ansgar Steland. Flexible nonlinear inference and change-point testing of high-dimensional spectral density matrices. *J. Multivariate Anal.*, 199:Paper No. 105245, 25, 2024.
[103] Howell Tong. *Threshold models in nonlinear time series analysis*, volume 21 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1983.
[104] Nicolas Verzelen, Magalie Fromont, Matthieu Lerasle, and Patricia Reynaud-Bouret. Optimal change-point detection and localization. *Ann. Statist.*, 51(4):1586–1610, 2023.
[105] Gabriele Villarini, Francesco Serinaldi, James A Smith, and Witold F Krajewski. On the stationarity of annual flood peaks in the continental united states during the 20th century. *Water Resources Research*, 45(8), 2009.
[106] Martin J. Wainwright. *High-dimensional statistics*, volume 48. Cambridge University Press, Cambridge, 2019. A non-asymptotic viewpoint.
[107] Mengchen Wang, Trevor Harris, and Bo Li. Asynchronous changepoint estimation for spatially correlated functional time series. *J. Agric. Biol. Environ. Stat.*, 28(1):157–176, 2023.
[108] Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(1):57–83, 2018.
[109] Joakim Westerlund. Common breaks in means for cross-correlated fixed-T panel data. *J. Time Series Anal.*, 40(2):248–255, 2019.
[110] Dominik Wied, Walter Krämer, and Herold Dehling. Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3):570–589, 2012.
[111] Keith J. Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73(1):91–104, 1986.
[112] Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005.
[113] Wei Biao Wu. On false discovery control under dependence. *Ann. Statist.*, 36(1):364–380, 2008.
[114] Wei Biao Wu and Mohsen Pourahmadi. Banding sample autocovariance matrices of stationary processes. *Statist. Sinica*, 19(4):1755–1768, 2009.
[115] Wei Biao Wu and Zhibiao Zhao. Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):391–410, 2007.
[116] Wei Biao Wu and Zhou Zhou. Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, 21(3):1397–1413, 2011.
[117] Weichi Wu, David Veitch, and Zhou Zhou. Asynchronous jump testing and estimation in high dimensions under complex temporal dynamics. 2024.
[118] Mengjia Yu and Xiaohui Chen. Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 83(2):247–270, 2021.
[119] Mengjia Yu and Xiaohui Chen. A robust bootstrap change point test for high-dimensional location parameter. *Electron. J. Stat.*, 16(1):1096–1152, 2022.
[120] D. Zhang and W. B. Wu. Gaussian Approximation for High Dimensional Time Series. *Ann. Statist.*, 45(5):1895–1919, 2017.
[121] Danna Zhang and Wei Biao Wu. Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.*, 49(1):233–254, 2021.
[122] Nancy R. Zhang, David O. Siegmund, Hanlee Ji, and Jun Z. Li. Detecting simultaneous change-points in multiple sequences. *Biometrika*, 97(3):631–645, 2010.
[123] Peng Zhao and Bin Yu. On model selection consistency of Lasso. *J. Mach. Learn. Res.*, 7:2541–2563, 2006.
[124] Zhou Zhou. Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.*, 33(3):438–457, 2012.
[125] Zhou Zhou and Wei Biao Wu. On linear models with long memory and heavy-tailed errors. *J. Multivariate Anal.*, 102(2):349–362, 2011.

8 Simulation results

In this section, we present detailed simulation studies justifying the theoretical excursions of Sections 2, 3 and 4. Some further simulation results related to the performance of Algorithm 3 in high-dimensional regime can be found in Section 14.

8.1 Behavior of test statistic under H_0

Proposition 2.2 instructs that under H_0, the test statistic $T_n = O_{\mathbb{P}}(1)$. In this subsection, we aim to empirically investigate the distribution of T_n under different type of null behavior, i.e. under $\tau_1 = \ldots = \tau_d$. For our numerical studies, we consider $d = 4$, and look at the following five settings of synchronized change-points. Let us consider some $\delta(n)$ such that $n\delta^2(n) \to \infty$, and denote

- Model 1: (No jumps) $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$.
- Model 2: (One jump) $\delta_1 = \delta(n)$, $\delta_2 = \delta_3 = \delta_4 = 0$.
- Model 3: (Two jumps) $\delta_1 = \delta_2 = \delta(n)$, $\delta_3 = \delta_4 = 0$.
- Model 4: (Three jumps) $\delta_1 = \delta_2 = \delta_3 = \delta(n)$, $\delta_4 = 0$.
- Model 5: (Four jumps) $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta(n)$.

We consider two values of n : 500 and 1000 For the model (1.1), let the errors $(\mathbf{e}_i)_{i\in\mathbb{Z}}$ follow a Vector Autoregressive (VAR) model of lag 1:

$$\mathbf{e}_i = A\mathbf{e}_{i-1} + \boldsymbol{\varepsilon}_i, \text{ where } (\boldsymbol{\varepsilon}_i)_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{RQ}^{5,1}). \tag{8.1}$$

Here $\boldsymbol{\Sigma}_{RQ}^{a,k}$ is the Rational Quadratic covariance matrix, i.e,

$$\boldsymbol{\Sigma}_{RQ}^{a,k}(j_1, j_2) = (1 + \frac{|j_1 - j_2|^2}{2ak^2})^{-a} \text{ with } a > 0, k > 0.$$

The A matrix is taken so that $A_{ij} = 0.3\exp(-|i-j|)$. Since we are working under null, the common change-point is taken to be 0.5. Finally, in order to properly investigate the effect of large jumps on T_n, we consider $\delta(n) = 0.5$. For each of the five models, the null distribution of T_n has been empirically estimated based on 5000 independent Monte Carlo draws, and is shown in Figure 3. Even if Proposition 2.2 instructs $T_n = O_{\mathbb{P}}(1)$, the asymptotic distribution of T_n is markedly different for each of the models. In particular, T_n is small if no dimensions have change-point. As more and more dimensions have a large enough jump, the distribution of T_n seems to become more and more spread out, until the number of dimensions with change-points is no longer greater than the number of dimensions without change-points. Subsequently, as we continue increasing the number of coordinates with large jumps, T_n puts more and more mass on zero. This behavior is, of course, natural, since if dimension j has a large jump, we expect $\hat{\tau}_j \approx \hat{\tau}$ under null, and in turn, $|S_{n\hat{\tau}_j,j} - n\hat{\tau}_j\bar{X}_{\cdot j}| \approx |S_{n\hat{\tau},j} - n\hat{\tau}\bar{X}_{\cdot j}|$. Therefore, T_n will have smaller values with increasing probability, as more and more dimensions have a significant jump. In fact, if $n\min_j \delta_j^2 \to \infty$, then following (10.17), one can show $T_n \xrightarrow{\mathbb{P}} 0$ under H_0. This behavior is indeed verified in Figure 3.

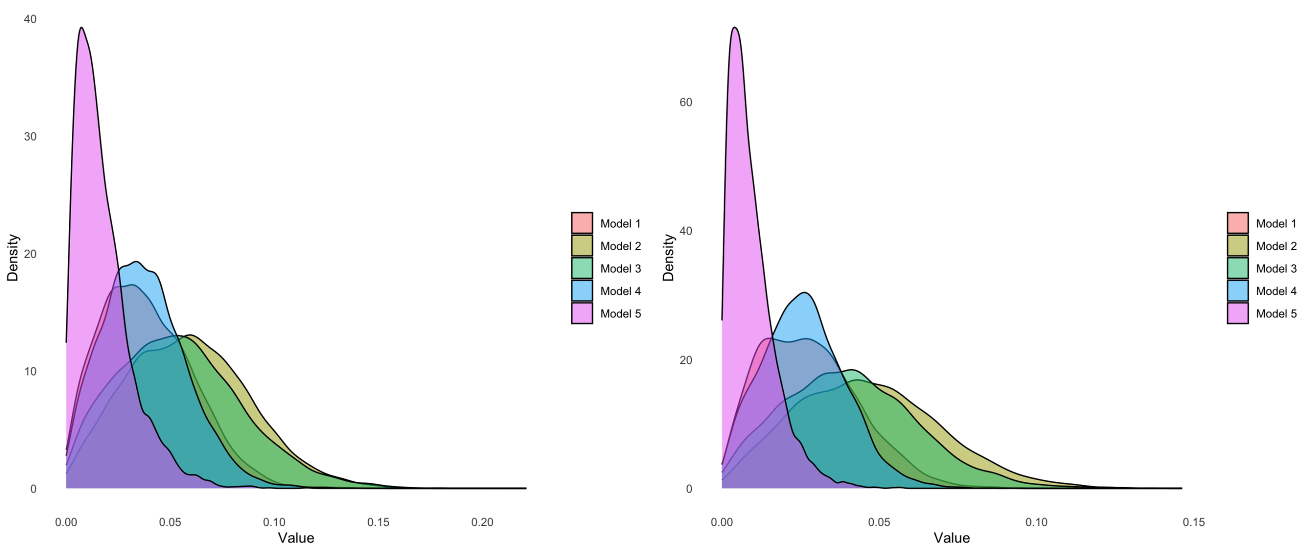

Figure 3: Distribution of T_n for the five models in Section 8.1 for $n = 500$(left) and $n = 1000$(right).

8.2 Performance of oracle bootstrap

This section is devoted to the efficacy of our Gaussian approximation theorem 3.1. Here, we look through the lens of our oracle bootstrap algorithm 1, and will explore how well the distribution of oracle bootstrap test statistic T_n^Z approximates that of T_n. Consider the Models 1-5 from Section 8.1, and let $n = 1000$. For each model, the distribution of T_n is estimated based on 5000 iterations. Similarly, 5000 "oracle" bootstrap samples of $T_n^{(s)}$ are drawn as in Algorithm 1. Figure 4 justifies the validity of the oracle bootstrap, in turn showcasing the effectiveness of the asymptotic approximation of Theorem 3.1.

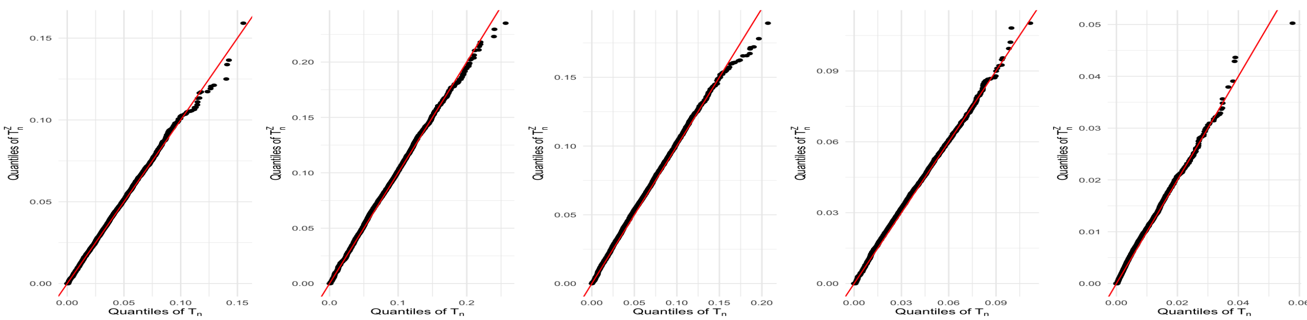

Figure 4: QQ plot of T_n with the oracle bootstrap samples $T_n^{(s)}$ for Models 1(left-most)-5(right-most).

8.3 Choice of kernel function and bandwidth for estimation of Σ_∞

In this section, we focus on the performance of $\hat{\Sigma}_{n,B_n}$ as an estimator of long-run variance Σ_∞ for different choices of bandwidths B_n, and also different choices of the Kernel function $K \in \mathcal{C}^1$. The long-run covariance matrix Σ_∞ for the innovations $(\mathbf{e}_i)$ from (8.1) can be computed explicitly, and has spectral norm $\rho^\star(\Sigma_\infty) = 9.534$. To estimate Σ_∞, consider three popular kernel functions.

- Parzen Kernel: $K_1(x) = (1 - 6|x|^2 + 6|x|^3)I\{0 \leq |x| \leq \frac{1}{2}\} + 2(1 - |x|)^3 I\{\frac{1}{2} \leq |x| \leq 1\}$.
- Tukey-Hanning Kernel: $K_2(x) = 0.5(1 + \cos(\pi x))I\{|x| < 1\}$.
- A split Rectangular Cosine kernel. $K_3(x) = I\{|x| < 0.95\} + 0.5(1 + \cos(20(x - 0.95)\pi))I\{0.95 \leq |x| \leq 1\}$.

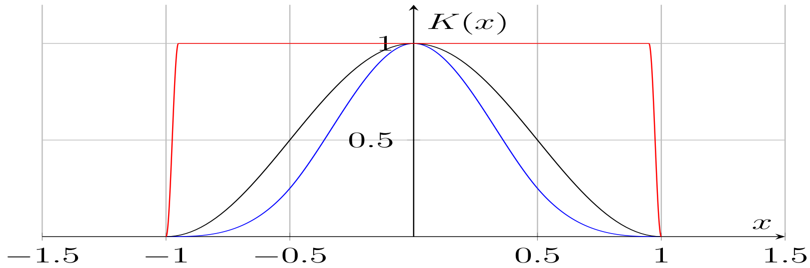

Figure 5: Plot of $K_1(x)$ (in blue), $K_2(x)$ (in **black**), and $K_3(x)$ (in red).

Note that $K_3(x)$ can be viewed as smoothed version of rectangular window. On the other hand, K_1 and K_2 are standard examples of $\mathcal{C}^1$ kernel functions [87, 24, 95]. In the following simulation studies, we again consider the five different models, and two values of $n = 500$ and 1000 as in Section 8.1. For each setting, the empirical mean and SD of $\rho^\star(\hat{\Sigma}_{n,B_n} - \Sigma_\infty)$ is estimated via 5000 independent Monte Carlo draws. Regarding the choice of bandwidth B_n, we note that for $p \geq 4$, (3.8) is minimized for $B_n \asymp n^{1/4}$. Some other popular choices include $n^{1/3}$ ([24], [26]) and $n^{1/(2r+1)}$ for $\mathcal{C}^r$ kernels ([91]). In light of this, we let B_n vary from $\lfloor n^{1/5} \rfloor$ to $\lfloor n^{1/3} \rfloor$ for each n. Tables 1 and 2 show that, on an average, K_3 consistently achieves the least estimation error, in line with its reduced bias, as discussed in Remark 3. As a trade-off, it also has slightly higher variation compared to Parzen or Tukey-Hamming kernel functions. The bandwidths from $\lfloor n^{1/4} \rfloor$ to $\lfloor n^{1/3} \rfloor$ seem to yield better accuracy in estimation with regards to both bias and variance; the different choices of kernel function for these bandwidths do not translate into any striking differences in performance in terms of MSE. Therefore, for our subsequent simulations and real-data exercises, we work with $B_n = \lfloor n^{1/4} \rfloor$.

	Kernel	$B_n = 3$	$B_n = 4$	$B_n = 5$	$B_n = 6$	$B_n = 7$
Model 1	Parzen	3.093(0.481)	3.494(0.798)	2.686(0.328)	2.976(0.461)	2.87(0.653)
	Tukey-Hanning	4.117(1.437)	3.386(0.736)	2.901(0.332)	2.591(0.271)	2.407(0.38)
	Splitted Rectangular Cosine	2.425(0.461)	2.003(0.798)	1.986(0.871)	2.11(0.848)	2.276(0.832)
Model 2	Parzen	3.027(0.483)	3.429(0.802)	2.605(0.337)	2.88(0.451)	2.763(0.633)
	Tukey-Hanning	4.072(1.454)	3.33(0.739)	2.833(0.334)	2.517(0.281)	2.329(0.39)
	Splitted Rectangular Cosine	2.358(0.46)	1.941(0.791)	1.931(0.862)	2.049(0.851)	2.206(0.839)
Model 3	Parzen	2.984(0.488)	3.378(0.802)	2.539(0.338)	2.817(0.446)	2.697(0.619)
	Tukey-Hanning	4.045(1.471)	3.293(0.748)	2.788(0.341)	2.465(0.289)	2.276(0.396)
	Splitted Rectangular Cosine	2.315(0.454)	1.893(0.792)	1.89(0.868)	2.014(0.855)	2.171(0.846)
Model 4	Parzen	2.868(0.506)	3.256(0.812)	2.391(0.357)	2.645(0.442)	2.53(0.607)
	Tukey-Hanning	3.97(1.507)	3.196(0.771)	2.673(0.366)	2.34(0.316)	2.146(0.421)
	Splitted Rectangular Cosine	2.209(0.455)	1.8(0.781)	1.81(0.868)	1.937(0.884)	2.093(0.899)
Model 5	Parzen	2.829(0.524)	3.211(0.821)	2.335(0.369)	2.574(0.44)	2.455(0.616)
	Tukey-Hanning	3.943(1.528)	3.161(0.788)	2.63(0.383)	2.291(0.331)	2.094(0.429)
	Splitted Rectangular Cosine	2.167(0.463)	1.755(0.78)	1.772(0.873)	1.905(0.901)	2.063(0.93)

Table 1: Empirical mean (standard deviation) of $\rho^\star(\hat{\Sigma}_{n,B_n} - \Sigma_\infty)$ for different choices of K and B_n. Here $n = 500$. Results have been rounded to three decimals.

	Kernel	$B_n = 3$	$B_n = 4$	$B_n = 5$	$B_n = 6$	$B_n = 7$	$B_n = 8$	$B_n = 9$
Model 1	Parzen	2.811 (0.675)	3.215 (1.035)	2.178 (0.291)	2.49 (0.415)	2.188 (0.402)	2.32 (0.46)	2.3 (0.602)
	Tukey-Hanning	3.954 (1.778)	3.165 (1.003)	2.616 (0.495)	2.238 (0.262)	1.986 (0.317)	1.827 (0.425)	1.734 (0.504)
	Splitted Rectangular Cosine	2.111 (0.342)	1.531 (0.726)	1.417 (0.855)	1.467 (0.849)	1.566 (0.807)	1.677 (0.772)	1.791 (0.757)
Model 2	Parzen	2.794 (0.686)	3.191 (1.04)	2.141 (0.29)	2.447 (0.412)	2.14 (0.413)	2.266 (0.469)	2.242 (0.609)
	Tukey-Hanning	3.943 (1.795)	3.15 (1.017)	2.596 (0.505)	2.213 (0.268)	1.957 (0.321)	1.795 (0.429)	1.699 (0.511)
	Splitted Rectangular Cosine	2.094 (0.343)	1.508 (0.722)	1.396 (0.851)	1.448 (0.847)	1.547 (0.808)	1.657 (0.772)	1.767 (0.757)
Model 3	Parzen	2.746 (0.694)	3.144 (1.048)	2.079 (0.302)	2.378 (0.413)	2.065 (0.419)	2.185 (0.47)	2.165 (0.603)
	Tukey-Hanning	3.914 (1.814)	3.112 (1.029)	2.549 (0.516)	2.161 (0.284)	1.902 (0.333)	1.739 (0.438)	1.643 (0.52)
	Splitted Rectangular Cosine	2.049 (0.35)	1.465 (0.718)	1.357 (0.848)	1.412 (0.849)	1.517 (0.814)	1.631 (0.785)	1.74 (0.78)
Model 4	Parzen	2.706 (0.7)	3.102 (1.053)	2.025 (0.312)	2.315 (0.413)	1.999 (0.421)	2.114 (0.469)	2.101 (0.603)
	Tukey-Hanning	3.887 (1.829)	3.077 (1.038)	2.508 (0.522)	2.114 (0.291)	1.852 (0.34)	1.687 (0.448)	1.591 (0.53)
	Splitted Rectangular Cosine	2.007 (0.352)	1.419 (0.718)	1.321 (0.842)	1.382 (0.845)	1.489 (0.816)	1.609 (0.79)	1.727 (0.787)
Model 5	Parzen	2.678 (0.712)	3.065 (1.057)	1.967 (0.329)	2.256 (0.418)	1.938 (0.416)	2.045 (0.456)	2.017 (0.581)
	Tukey-Hanning	3.872 (1.852)	3.055 (1.054)	2.478 (0.532)	2.077 (0.302)	1.81 (0.351)	1.643 (0.456)	1.547 (0.534)
	Splitted Rectangular Cosine	1.982 (0.349)	1.39 (0.708)	1.297 (0.834)	1.363 (0.844)	1.466 (0.825)	1.573 (0.81)	1.676 (0.812)

Table 2: Empirical mean (standard deviation) of $\rho^\star(\hat{\Sigma}_{n,B_n} - \Sigma_\infty)$ for different choices of K and B_n. Here $n = 1000$. Results have been rounded to three decimals.

8.4 Simulation for Algorithm 3

In this section, we carry out an extensive simulation study that numerically justifies the asymptotic validity of our bootstrap procedure as proved in Theorem 4.1. We consider two separate models below.

8.4.1 Threshold autoregressive models

In this section, we consider observations from Model (1.1) with the stationary errors $(\mathbf{e}_i)_{i\in\mathbb{Z}}$ following a Threshold Auto-Regressive (TAR) process [103, 31]. Mathematically, borrowing the notation of (1.1) and (1.2), we write

$$e_{ij} = -\rho|e_{i-1}, j| + \varepsilon_{ij}, 1 \leq i \leq n, 1 \leq j \leq d, \tag{8.2}$$

and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{id}) \in \mathbb{R}^d$ are the innovations such that $(\boldsymbol{\varepsilon}_i)_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, 0.75\boldsymbol{\Sigma}_{RQ}^{5,1})$. We work with $d = 4$, $n = 500$ and 1000, and $\rho = 0.5$ in (8.2). For each $1 \leq j \leq d$, let $\mu_j^L = 0$. Let us consider the following scenarios.

- **Setting 1**. $(\tau_1, \tau_2, \tau_3, \tau_4) = (0.5, 0.5 - r_1, 0.5 + r_2, 0.5)$, where $r_1, r_2 \in \{0, 0.01, 0.02, \ldots, 0.1\}$. The jumps δ_j are taken as $(6/\log n, -6/\log n, 6/\log n, 0)$. Note that, the null H_0 corresponds to $r_1 = r_2 = 0$. When exactly one of r_1 and r_2 is zero, then this setting has asynchronized change-points at only two component series. When $\min\{r_1, r_2\} > 0$, one finds asynchronized change-points at three component series.

- **Setting 2**. $(\tau_1, \tau_2, \tau_3, \tau_4) = (0.5, 0.5 - r, 0.5, 0.5)$ with $r = 0.01, 0.02, \ldots, 0.1$. The jumps are $(\delta_1, \delta_2, \delta_3, \delta_4) = (6/\log n, -6/\log n, 0, 0)$. Under the alternate, Setting 2 has asynchronized change-points at only two component series.

For each n, we use the bandwidth $B_n = \lfloor n^{1/4} \rfloor$ while estimating Σ_∞ by $\hat{\Sigma}_{n,B_n}$. The bootstrap quantile $b_\alpha(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n})$ is empirically estimated based on 5000 bootstrap samples. Finally, for each particular simulation setting in each of the model, we have used 1000 independently sampled Monte Carlo draws to empirically estimate the Type-1-error or power (at 5% level of significance) for that corresponding setting. Figure 6 shows that, under Models 1 and 2, the distinct change-points are difficult to spot in the asynchronized case.

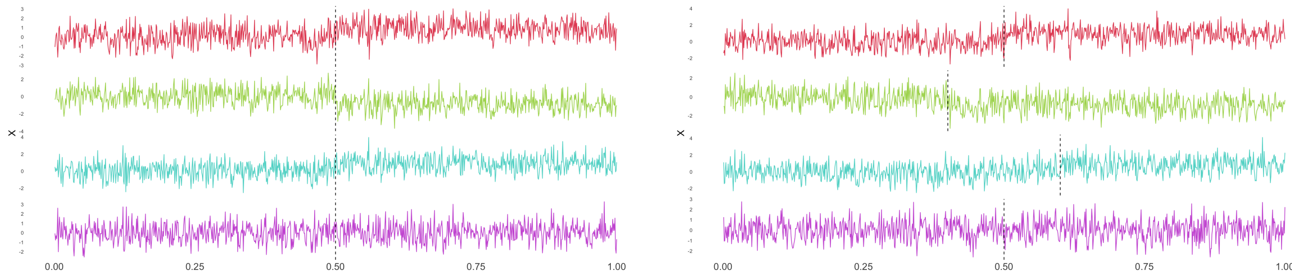

Figure 6: A random draw of $\mathbf{X}(n = 1000)$ from Setting 1 with $r_1 = r_2 = 0$ (left), and $r_1 = r_2 = 0.1$ (right).

		r_2										
n	r_1	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
500	0	0.057	0.123	0.178	0.268	0.403	0.508	0.573	0.67	0.729	0.794	0.855
	0.01	0.1	0.181	0.302	0.362	0.502	0.607	0.674	0.755	0.768	0.812	0.889
	0.02	0.191	0.267	0.383	0.512	0.63	0.718	0.732	0.829	0.852	0.882	0.908
	0.03	0.298	0.377	0.491	0.617	0.707	0.75	0.824	0.867	0.895	0.925	0.94
	0.04	0.4	0.481	0.615	0.67	0.756	0.826	0.88	0.884	0.923	0.938	0.965
	0.05	0.496	0.615	0.682	0.757	0.81	0.885	0.899	0.934	0.955	0.967	0.982
	0.06	0.583	0.691	0.746	0.81	0.883	0.898	0.924	0.948	0.97	0.974	0.973
	0.07	0.663	0.747	0.817	0.882	0.889	0.929	0.96	0.954	0.962	0.977	0.982
	0.08	0.724	0.799	0.849	0.912	0.907	0.946	0.961	0.964	0.979	0.987	0.99
	0.09	0.788	0.837	0.885	0.92	0.952	0.966	0.977	0.979	0.988	0.986	0.994
	0.1	0.814	0.863	0.913	0.943	0.95	0.969	0.975	0.982	0.989	0.995	0.993
1000	0	0.062	0.132	0.232	0.381	0.527	0.667	0.768	0.861	0.894	0.924	0.939
	0.01	0.123	0.24	0.368	0.529	0.66	0.758	0.851	0.886	0.928	0.962	0.963
	0.02	0.218	0.387	0.523	0.683	0.776	0.839	0.923	0.939	0.966	0.977	0.978
	0.03	0.358	0.502	0.665	0.767	0.86	0.917	0.935	0.963	0.98	0.988	0.988
	0.04	0.554	0.663	0.763	0.854	0.912	0.944	0.972	0.981	0.995	0.989	0.994
	0.05	0.659	0.771	0.861	0.92	0.935	0.964	0.985	0.997	0.995	0.996	0.998
	0.06	0.764	0.842	0.88	0.948	0.966	0.984	0.991	0.991	0.996	1	0.998
	0.07	0.824	0.894	0.929	0.959	0.98	0.981	0.995	0.998	0.996	0.999	0.999
	0.08	0.877	0.925	0.946	0.977	0.99	0.994	0.998	1	0.998	1	0.999
	0.09	0.908	0.956	0.963	0.979	0.989	0.992	0.997	1	0.997	1	1
	0.1	0.951	0.962	0.978	0.992	0.991	0.996	1	0.998	0.999	0.999	1

Table 3: Type-I error (when $r_1 = r_2 = 0$), and power of Algorithm 3 for Setting 1.

r	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$n = 500$	0.093	0.111	0.141	0.186	0.219	0.302	0.322	0.355	0.411	0.489
$n = 1000$	0.099	0.109	0.17	0.225	0.287	0.382	0.422	0.512	0.598	0.672

Table 4: Power of Algorithm 3 for Setting 2.

8.4.2 GJR-GARCH models

Next, we also apply our bootstrap algorithm to the case when the error process $(\mathbf{e}_i)$ follows a GJR-GARCH(1,1) model ([43]):

$$e_{i,j} = \sigma_{i,j}\varepsilon_{i,j}\ ; \sigma_{i,j}^2 = 0.01 + 0.7\sigma_{i-1,j}^2 + 0.1e_{i-1,j}^2 + 0.2e_{i-1,j}^2 I\{e_{i-1,j} \leq 0\}. \tag{8.3}$$

As in Section 8.4.1, we work with innovations $(\boldsymbol{\varepsilon}_i)_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, 0.75\mathbf{\Sigma}_{RQ}^{5,1})$. We let $d = 4$, $\mu_j^L = 0$ for all $1 \leq j \leq d$, and for each particular setting, consider $n = 500$ and $n = 1000$. We focus on the following model.

- **Setting 3**. $(\tau_1, \tau_2, \tau_3, \tau_4) = (0.5, 0.5 - r_1, 0.5 + r_2, 0.5)$, where $r_1, r_2 \in \{0, 0.01, \ldots, 0.1\}$. The jumps are $(\delta_1, \delta_2, \delta_3, \delta_4) = (1/(\log n), 1/(\log n), -1/(\log n), 0)$.

For each particular setting, we compute the empirical type-1-error and power via exactly the same mechanism as described in Section 8.4.1. Tables 3, 4, and 5 show the empirical type-

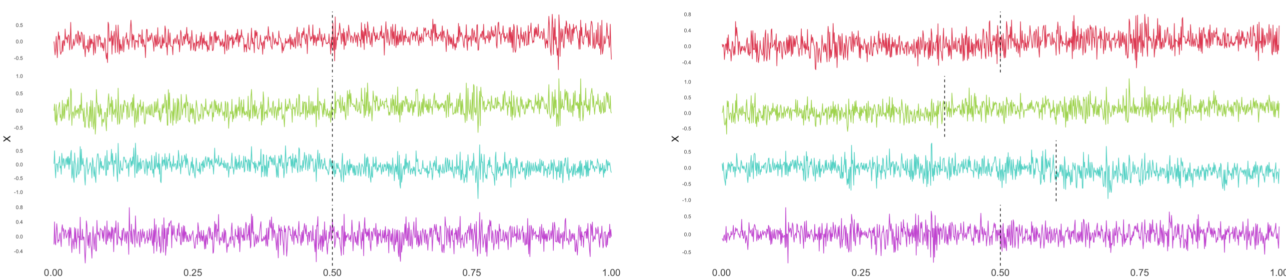

Figure 7: A random draw of $\mathbf{X}(n = 1000)$ from Setting 3 with $r_1 = r_2 = 0$ (left), and $r_1 = r_2 = 0.1$ (right).

1 error and powers of the models in Sections 8.4.1 and 8.4.2. The simulation results are as expected, based on the theory of the preceding sections. In particular, in Tables 3 and 5, as the sample size n increases from 500 to 1000, the empirical type-1 error stabilizes to around 5% for both TAR and GJR-GARCH errors. This justifies the asymptotic result of Theorem 4.1. Obviously, the empirical powers under different alternative settings increase with increasing n. Moreover, Table 4, and the entries corresponding to $r_1 = 0$ or $r_2 = 0$ in Table 3 show that, the power is comparatively lesser when there are only two distinct change-points under the

		r_2										
n	r_1	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
500	0	0.083	0.102	0.112	0.169	0.257	0.326	0.419	0.515	0.54	0.628	0.692
	0.01	0.083	0.114	0.174	0.248	0.329	0.41	0.474	0.563	0.619	0.669	0.715
	0.02	0.133	0.184	0.234	0.32	0.383	0.477	0.566	0.628	0.685	0.704	0.76
	0.03	0.177	0.256	0.322	0.385	0.487	0.548	0.617	0.686	0.749	0.775	0.826
	0.04	0.254	0.327	0.399	0.498	0.554	0.617	0.673	0.732	0.797	0.829	0.847
	0.05	0.339	0.39	0.502	0.554	0.626	0.69	0.755	0.777	0.811	0.851	0.867
	0.06	0.414	0.483	0.542	0.616	0.65	0.762	0.796	0.82	0.84	0.868	0.886
	0.07	0.519	0.532	0.623	0.688	0.728	0.777	0.841	0.851	0.895	0.916	0.912
	0.08	0.561	0.618	0.665	0.741	0.763	0.825	0.852	0.887	0.898	0.915	0.935
	0.09	0.618	0.636	0.723	0.788	0.811	0.842	0.869	0.893	0.905	0.93	0.942
	0.1	0.676	0.736	0.75	0.799	0.845	0.857	0.898	0.91	0.934	0.932	0.961
1000	0	0.06	0.088	0.164	0.272	0.366	0.515	0.592	0.679	0.751	0.809	0.834
	0.01	0.1	0.172	0.226	0.347	0.47	0.559	0.675	0.738	0.815	0.863	0.892
	0.02	0.169	0.249	0.367	0.47	0.571	0.651	0.763	0.813	0.853	0.895	0.904
	0.03	0.268	0.377	0.413	0.565	0.673	0.761	0.803	0.856	0.909	0.929	0.944
	0.04	0.375	0.484	0.587	0.661	0.758	0.817	0.861	0.885	0.923	0.944	0.964
	0.05	0.475	0.56	0.69	0.728	0.83	0.859	0.912	0.93	0.953	0.961	0.98
	0.06	0.587	0.665	0.719	0.806	0.858	0.91	0.923	0.949	0.95	0.972	0.98
	0.07	0.669	0.726	0.794	0.878	0.885	0.943	0.948	0.963	0.974	0.982	0.982
	0.08	0.742	0.822	0.85	0.889	0.93	0.937	0.972	0.979	0.979	0.985	0.983
	0.09	0.787	0.868	0.873	0.934	0.939	0.953	0.966	0.982	0.981	0.995	0.991
	0.1	0.83	0.894	0.907	0.918	0.95	0.964	0.977	0.987	0.987	0.99	0.993

Table 5: Type-1 error ($r_1 = r_2 = 0$) and Power of Algorithm 3 for Setting 3.

alternative; the other columns and rows in Table 3 correspond to three distinct change-points, and in conjunction, to an increased power. This is corroborated by the proof of Proposition 2.2, where, with an increase in the number of distinct τ_j's with jumps $\delta_j \gg 1/\sqrt{n}$, $T_n \xrightarrow{\mathbb{P}} \infty$ faster. Interestingly, in Table 5, even for the conditionally heteroscadastic set-ups such as GARCH, our Gaussian bootstrap-based test performs really well, and achieves around 80% power for $n = 1000$ when the two distinct change-points are separated by only 0.09. These results augur well for the performance and robustness of Algorithm 3 in real-life scenarios; in particular, based on our theoretical excursion and extensive simulation studies, we expect the test to remain valid and yield considerable power in very general settings.

9 Applications: Real data analyses

In this section we present two interesting real-life applications. In the first one, we test for synchronization of two time series in two spatial locations, recovering interesting connotations behind the asynchronization. On the other hand, for the second dataset, we show that blindly assuming synchronized change-points across the panel results in missing potentially interesting disturbances or shifts.

9.1 Onset of winter floods in Mississippi river

Change-point analysis is often employed to detect various climate-influenced or man-made changes in hydrological data [69]. With regards to flood statistics, Pettitt's test[88] and CUSUM-based methods have been applied in detecting change-points in annual flood peaks in the mainland United States [76, 105]. However, often such an analysis is limited by an i.i.d. assumption, or an adaptation of any particular stationary parametric model such as Log-Pearson type III [41]. Several works, such as [83, 56, 105], also analyze the lower-Mississippi water levels using spatio-temporal modelling. On the other hand, the upper Mississippi basin already suffered a record catastrophic flood [86] in Dec 2018-19, causing an estimated \$2 billion dollars in damages [93]. Thus, analyzing onset of flood, particularly in winter, is also necessary. In particular, we would attempt to understand how the onset dates of winter surge in water level have varied (or stayed the same) in two different locations ~ 200 miles apart. This problem can be conveniently posed in our test of synchronization framework, where the change-points signify the flood onset at the corresponding location. Note that, in most of the works on spatio-temporal modeling of water-levels, usually Gaussianity and a suitable parametric form of the covariance structure is assumed in order to incorporate the spatial effects. On the contrary, under a mild set of assumptions, our methodologies allow us to draw meaningful statistical inferences about this problem, without resorting to sophisticated modeling exercises involving stringent and un-testable assumptions. We will discuss more about the potential usefulness of our results after having looked into the dataset and the statistical results.

The data is taken from USGS Water Data for the Nation. Figure 8 shows the time series plots of the daily water discharge (in ft^3 per second) from 1st September 2023 to 1st May 2024 at Memphis and Vicksburg, along with their individual change-points. In particular, we have $d = 2$ corresponding to the two locations, and $n = 243$ observations for each locations. The

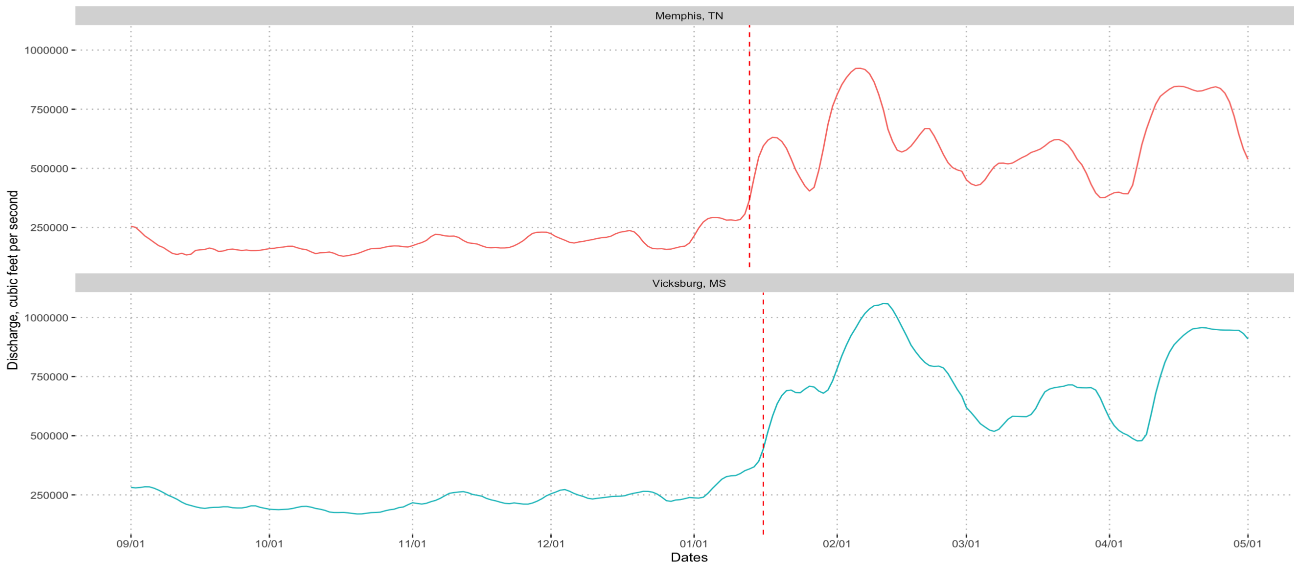

Figure 8: Water discharge data of Mississippi river at two different locations from Sept'23 to May'24. The vertical red line indicates the individual change points detected by CUSUM.

locations on the same river obviously have spatial interaction, further justifying our bootstrap procedure based on simultaneous Gaussian approximation of stationary multivariate processes. To perform the test of synchronization, we use $B = 5000$ bootstrap samples. For covariance matrix estimation we take $B_n = \lfloor n^{1/4} \rfloor$. The p-value comes out to be 0.0264, which implies we reject the null hypothesis of synchronized change-point at 5% level of significance. The conclusion of asynchronized flood onset dates for Memphis and Vicksburg have important connotations for policy planning, preparation for flooding events, and much more. In particular, Memphis saw a sudden increase in the amount of water discharged on 13-th January of 2024, whereas this increased volume of water reached downstream at Vicksburg only three days later, i.e. on 16-th January, 2024. This difference can be interpreted as the additional time available for Vicksburg to prepare for a flood event, after Memphis (around 220 miles away) has witnessed a surge in river discharge. Our test statistically validates this difference in flood onset dates, and makes way for further detailed research to understand how distance affect the flood onset dates in different locations.

9.2 Mental load of aviation pilots

In this section, we analyze the data on cardio-respiratory response of pilots, collected by [46]. We describe the data briefly. 61 pilots underwent four phases of increasing mental and physical demand, whose start and end-time are indicated in parenthesis that follows

1. "Resting Baseline"(0-332s) phase of simply focusing on a cross;
2. "Vanilla Baseline"(333-673s) phase of a minimally demanding vigilance task;
3. "Multiple Tasks"(674-1053s) phase of performing three demanding, cognition-related activities simultaneously, and,
4. "Recovery"(1054-1393s) phase of relaxation by watching a movie.

For more details and context, readers are referred to [46]. For each pilot, there are three time series on their heart rate (HR), partial pressure of end-tidal CO_2 (petCO_2) and respiratory rate (RR) respectively. We work with the dataset of a randomly selected pilot as provided in R package `kcpRS`. The common change-point between "Vanilla Baseline" and "Multiple Task" can be easily spotted; it is intuitive and well-documented in [46, 30, 29]. In that light, we first focus on the change-points occurring during the shift between "Resting Baseline" and "Vanilla Baseline". We do this by analyzing these 3-dimensional time-series for the first 500 time points (Time 1-500s). We plot this data at Figure 9. Note that, [29] found no change-points in variance, and only found change in auto-correlations for specific choices of hyper-parameters while performing non-parametric change-point detection methods. Thus, we also assume the multivariate time series to be stationary, and focus on mean-based change points. Both [30, 29] assume synchronized change-points for this dataset, with their common change-point estimated at exactly the location of shift between phases-i.e, at $t = 332$s.

We will employ our Algorithm 3 to test synchronization at level 5%. For estimating Σ_∞, we specify $B_n = \lfloor n^{1/4} \rfloor$ with $n = 500$. The corresponding p-value for the test of synchronization is 0.0362. Therefore, for the early stage comprising of the first shift between the first two phases, our test result implies asynchronized change-point. Since this finding is inconsistent with what was assumed in literature, some follow-up analyses and explanations are in order.

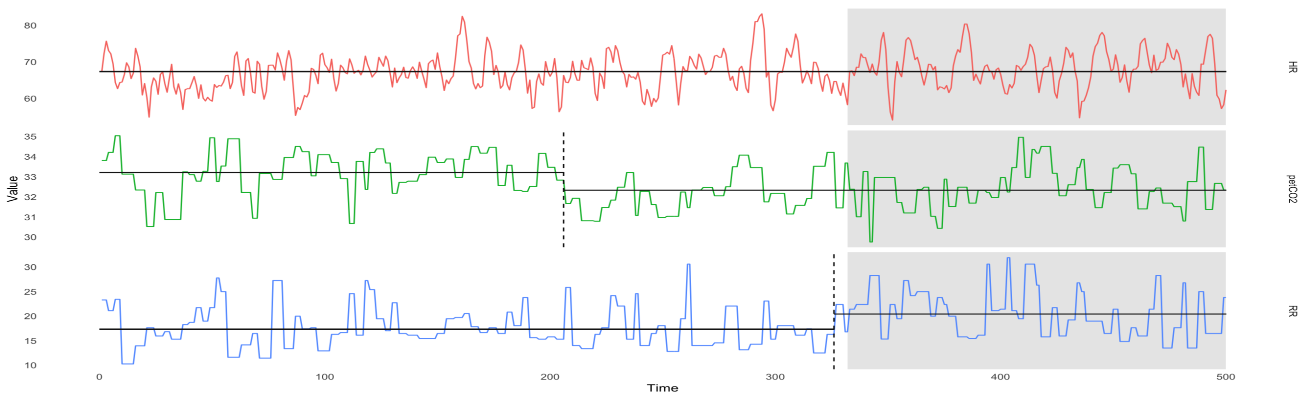

Figure 9: Time series plot for early stage (Time 1-500s) of the pilot mental load dataset. The red, green and blue plots indicate the time series corresponding to heart rate, PET CO_2, and respiratory rate respectively. The white and shaded regions indicate the phases "Resting Baseline" and "Vanilla Baseline". The black dashed lines indicate the estimated change-points by our method, and the solid horizontal black lines denote the estimated means of the corresponding piece-wise segments.

There seems to be no change-point detected in the heart-rate time series. [57] hypothesized the heart-rate to decrease during vanilla baseline. On the other hand, the petCO_2 time series displays a change-point in between the "Resting Baseline" phase (estimated at $t = 206$s), and then the mean level stays the same through the initial "Vanilla Baseline" period. One possible explanation could be that, after the start of the experiment, the level of stress recedes leading to decreased CO_2 circulation, and intensity of the body's metabolism improves regulated midway through the resting phase [89]; this stays the same even through the "Vanilla Baseline", the task in the second phase being only minimally demanding. The respiratory rate displays a clear change point near the boundary between "Resting" and "Vanilla Baseline" phases (estimated at $t = 325$s). In fact, we see that consistent with our intuition, breathing increases slightly in performing the vigilance task at the Vanilla Baseline stage. It is important to note that, these follow up analyses of introspecting into individual components and the subsequent findings are results of questioning the 'popular' assumption of synchronized change-point through our statistical testing procedure.

We also employ our algorithm separately to the later stage of this dataset i.e. a period that comprises of the shift between "Multiple Tasks" and "Recovery". We fail to reject the null hypothesis here, as the p-value according to our test for this part of the data comes out at 0.1088. The common change-point is detected exactly when the phase shift happens at $t = 1053$. This echoes the assumption in [29]. The corresponding plot for this dataset is shown in Figure 10.

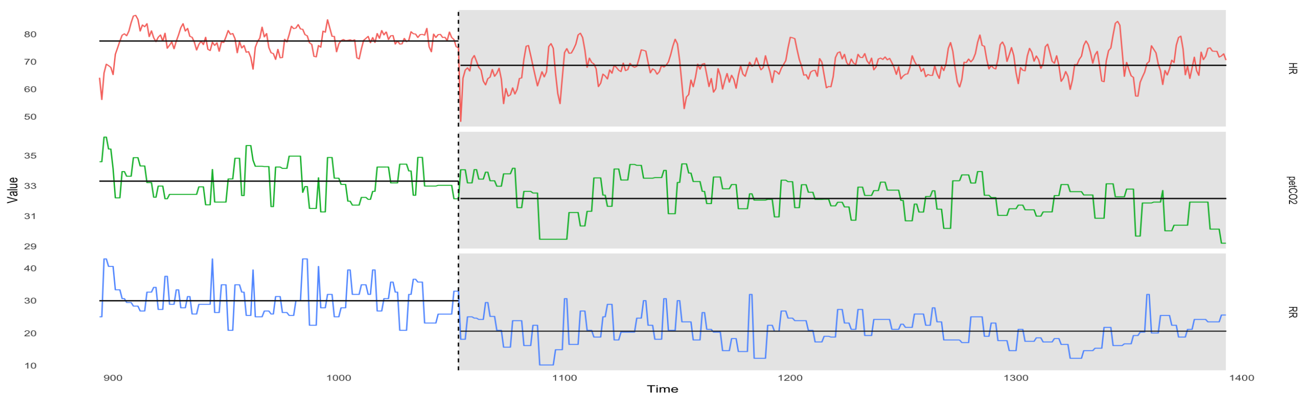

Figure 10: Time-series plot, corresponding to Figure 9, for the later stage of the pilot mental load dataset (Time 894-1393s). The white and shaded regions in the figure indicate the phases "Multiple Tasks" and "Recovery".

10 Proofs of Section 2: Behavior of test statistic

In this section, our main aim is to prove Propositions 2.1 and 2.2. Proposition 2.1 plays a crucial role in guaranteeing that our test statistic has a small value under the null of synchronization, thereby leading to the statement of Proposition 2.2. In fact, Proposition 2.2 characterize both the validity and consistency of our test statistic.

The main technical tool we require in order to analyze the behavior of CUSUM statistic under various scenarios, is a variation of the well-known Hàjek-Rényi type inequality ([47]). In the context of time series, [9] proved such inequalities for linear stochastic processes of the form of $X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$. It has since been extended further for more general processes, for example in Theorem 1 of [70] and Theorem 4.1 of [65]. For the sake of completeness, we provide a version of Hàjek-Rényi inequality for processes satisfying (2.5) along with a simple proof. We also use a well-known Rosenthal-type inequality controlling the maximum of block sums of stationary processes. To state both the results in a general setting, we invoke the univariate stationary process $\{Y_i\}_{i\in\mathbb{Z}}$. In particular, let $Y_i \in \mathbb{R}$ have the causal representation $Y_i = g(\varepsilon_i, \varepsilon_{i-1}, \ldots)$ for i.i.d. innovations ε_i's, and a measurable function $g : \mathbb{R}^\infty \to \mathbb{R}$. Further suppose $(Y_i)_{i\in\mathbb{Z}}$ satisfy (2.8), where $\theta_{i,p}$'s are defined as in (2.6) with Y_i's. Now we are ready to state the results discussed above.

Lemma 10.1 (Theorem 2.(i) of [112])**.** *Consider stationary processes $(Y_i)_{i\in\mathbb{Z}}$ with $\mathbb{E}(Y_i) = 0$, satisfying (2.5) and (2.8) for some $p \geq 2$. Then, for $1 \leq m \leq n$, it holds that*

$$\max_a \| \max_{1\leq k\leq m} |Y_{a+1} + \ldots + Y_{a+k}| \|_p \leq \frac{p}{\sqrt{p-1}} m^{1/2} \Theta_{0,p}. \tag{10.1}$$

Lemma 10.2 (A Hàjek-Rényi-type inequality)**.** *Under the assumptions of Lemma 10.1, we have*

$$\mathbb{P}\left(\max_{0\leq k\leq m} \frac{1}{n-k} \left| \sum_{i=k+1}^{n} Y_i \right| \geq \alpha \right) \leq \frac{C_p \Theta_{0,p}^2}{\alpha^2 (n-m)} \text{ for } m \leq n-2, \tag{10.2}$$

where C_p denotes a constant depending only on p.

Proof of Lemma 10.2. Let $L_0 := \lfloor \log_2(n-m) \rfloor \geq 0$, and $L_1 := \lfloor \log_2 n \rfloor$. Therefore, using Lemma 10.1 and Markov's inequality

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq m} \frac{1}{n-k} \left| \sum_{i=k+1}^{n} Y_i \right| \geq \alpha \right) &\leq \sum_{l=L_0}^{L_1} \mathbb{P}\left(\max_{(n-2^{l+1})\vee 0\leq k\leq (n-2^l)\wedge m} \frac{1}{n-k} \left| \sum_{i=k+1}^{n} Y_i \right| \geq \alpha \right) \\
&\leq \sum_{l=L_0}^{L_1} \mathbb{P}\left(\max_{(n-2^{l+1})\vee 0\leq k\leq (n-2^l)\wedge m} \left| \sum_{i=k+1}^{n} Y_i \right| \geq \alpha (2^l \wedge (n-m)) \right) \\
&\leq \sum_{l=L_0}^{L_1} C_p \frac{\Theta_{0,p}^2}{\alpha^2 2^{l-1}} \lesssim \frac{C_p \Theta_{0,p}^2}{\alpha^2 2^{L_0}} \leq \frac{C_p \Theta_{0,p}^2}{\alpha^2 (n-m)},
\end{aligned}$$

which completes the proof. $\square$

The Hàjek-Rényi type inequality enables us to tackle the behavior of sample mean to the left and right of the estimated change-point. The use of Lemma 10.2 underpins much of the probabilistic arguments in the proof of Propositions 2.1 and 2.2, which are provided below sequentially.

Proof of Proposition 2.1. We will first show that $|\hat{\tau}_j - \tau_j| = O_{\mathbb{P}}((n\delta_j^2)^{-1} \wedge 1)$ for each $1 \le j \le d$. Fix $1 \le j \le d$ and $\varepsilon > 0$. Let $C_\varepsilon > 1$ denote a large enough constant depending on ε, whose choice will be made explicitly clear in the appropriate part of our argument. Moreover, let M_ε be another large constant such that $\mathbb{P}(|\hat{\tau}_j - \tau_j| > M_\varepsilon/(n\delta_j^2)) \le \varepsilon$. Our argument will necessarily hinge on finding an appropriate M_ε. Note that, if $n\delta_j^2 \le C_\varepsilon$, then the conclusion follows trivially by choosing $M_\varepsilon \ge C_\varepsilon$. Henceforth, it is assumed that $n\delta_j^2 > C_\varepsilon$. Let $\hat{k}_j = n\hat{\tau}_j$, and $k_{0,j} = \lfloor n\tau_j \rfloor$. Clearly, $n - k_{0,j} \asymp n(1-\tau_j)$. Observe that $\hat{k}_j$ can be written as $\arg\max_{1\le i\le n} |V_{i,j}^X|$, where

$$V_{i,j}^X = S_{ij} - i\bar{X}_{\cdot j}. \tag{10.3}$$

Further, let $V_{i,j}^e := S_{ij}^e - i\bar{e}_{\cdot j}$. A crucial observation for our subsequent arguments is that,

$$V_{i,j}^X - \mathbb{E}[V_{i,j}^X] = V_{i,j}^e, \quad \text{with } \mathbb{E}[V_{i,j}^X] = \begin{cases} -i(1-\frac{k_{0,j}}{n})\delta_j, & i/n \le \tau_j, \\ -\frac{k_{0,j}}{n}(n-i)\delta_j, & i/n > \tau_j, \end{cases} \quad \text{for all } 1 \le i \le n,\ 1 \le j \le d. \tag{10.4}$$

Let $\boldsymbol{X}_{\cdot j}$ denote the vector $(X_{1j}, \ldots, X_{nj})^\top$ for $1 \le j \le d$. Since $(V_{i,j}^X)_{i=1}^n$ are invariant with respect to μ_j^L, hence without loss of generality, we can let $\mu_j^L = 0$ for $1 \le j \le d$. Further, without loss of generality, assume that $\delta_j > 0$ (otherwise consider $-\boldsymbol{X}_{\cdot j}$). Observe that for all sufficiently large n,

$$\mathbb{P}\left(|\hat{\tau}_j - \tau_j| > \frac{M_\varepsilon}{n\delta_j^2}\right) \le \mathbb{P}\left(\max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} |V_{k,j}^X| \ge |V_{k_{0,j},j}^X|\right)$$

$$
\leq \mathbb{P}\left(\max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} V_{k,j}^X + V_{k_{0,j},j}^X \geq 0\right) + \mathbb{P}\left(\max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} V_{k,j}^X - V_{k_{0,j},j}^X \leq 0\right)
$$

$$
:= P_1 + P_2. \tag{10.5}
$$

Now, for the first term, suppose $k_1 := \arg\max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} V_{k,j}^X + V_{k_{0,j},j}^X$. Consider the following sequence of implications

$$
\begin{aligned}
& V_{k_1,j}^X + V_{k_{0,j},j}^X \geq 0 \\
\implies & (V_{k_1,j}^X - \mathbb{E}[V_{k_1,j}^X]) + (V_{k_{0,j},j}^X - \mathbb{E}[V_{k_{0,j},j}^X]) \geq -(\mathbb{E}[V_{k_1,j}^X] + \mathbb{E}[V_{k_{0,j},j}^X]) \\
\implies & 2\max_{1\leq k\leq n} |V_{k,j}^X - \mathbb{E}[V_{k,j}^X]| \geq -\mathbb{E}[V_{k_{0,j},j}^X], \quad (\text{as } \mathbb{E}[V_{k_1,j}^X] < 0 \text{ for } \delta_j > 0) \\
\implies & 2\max_{1\leq i\leq n} |V_{i,j}^e| \geq -\mathbb{E}[V_{k_{0,j},j}^X] = k_{0,j}(1-k_{0,j}/n)\delta_j.
\end{aligned}
$$

Therefore, in view of these assertions and applying Lemma 10.1 and Markov's inequality, one obtains for sufficiently large n,

$$
P_1 \leq C_p \frac{n\Theta_{0,p}^2}{k_{0,j}^2(1-k_{0,j}/n)^2\delta_j^2} \leq C_p \frac{\Theta_{0,p}^2}{\tau_j^2(1-\tau_j)^2 n\delta_j^2} < \frac{C_p\Theta_{0,p}^2}{\tau_j^2(1-\tau_j)^2 C_\varepsilon} < \varepsilon,
$$

where the last inequality can be guaranteed by choosing C_ε large enough. On the other hand, for P_2, suppose $k_2 := \arg\max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} V_{k,j}^X - V_{k_{0,j},j}^X$. Again, note the following sequence of implications:

$$
\begin{aligned}
& V_{k_2,j}^X - V_{k_{0,j},j}^X \leq 0 \\
\implies & (V_{k_2,j}^X - \mathbb{E}[V_{k_2,j}^X]) - (V_{k_{0,j},j}^X - \mathbb{E}[V_{k_{0,j},j}^X]) \leq -(\mathbb{E}[V_{k_2,j}^X] - \mathbb{E}[V_{k_{0,j},j}^X]) \\
\implies & |S_{k_{0,j}j}^e - S_{k_2j}^e - (k_{0,j}-k_2)\bar{e}_{\cdot j}| \geq \begin{cases} (k_{0,j}-k_2)(1-k_{0,j}/n)\delta_j, & k_2 \leq k_{0,j}, \\ (k_2-k_{0,j})(k_{0,j}/n)\delta_j, & k_2 > k_{0,j} \end{cases} \\
\implies & \max_{k:|k-k_{0,j}|>M_\varepsilon/\delta_j^2} \left| \frac{1}{k_{0,j}-k}(V_{k_{0,j},j}^e - V_{k,j}^e)\right| \geq \min\{k_{0,j}/n, 1-k_{0,j}/n\}\delta_j.
\end{aligned}
$$

By virtue of Lemmas 10.1 and 10.2, and in view of $|\hat{\tau}_j - \tau_j| > M_\varepsilon/(n\delta_j^2)$, from the above implications we obtain

$$P_2 \leq \frac{C_p \Theta_{0,p}^2}{\min\{\tau_j^2, (1-\tau_j)^2\} M_\varepsilon} < \varepsilon,$$

where, as in the case for P_1, the last inequality is guaranteed by a choice of large enough M_ε. Combining the analysis of P_1 and P_2, from (10.5) we obtain the conclusion.

Next we prove consistency of the $\hat{\tau}$ under the null of synchronization. Let $\tilde{k} = \lfloor n\tau \rfloor$. As $\tilde{k}/n \asymp \tau$, for ease of exposition and to avoid cumbersome notation, we assume $\tilde{k} = n\tau \in \mathbb{N}$. Recall $\boldsymbol{X}_{\cdot j} = (X_{1,j}, \ldots, X_{n,j})^\top$, j-th component of the time series. Observe that, for any set $A \subseteq \{1, 2, \ldots, d\}$, replacing $\boldsymbol{X}_{\cdot j}$ by $-\boldsymbol{X}_{\cdot j}$ for all $j \in A$ does not change $\hat{\tau}$. Therefore, without loss of generality, we assume that $\delta_1 \geq \ldots \geq \delta_d \geq 0$. Given $\varepsilon > 0$, we aim to find a large enough L_ε such that $\mathbb{P}(|\hat{\tau} - \tau| > L_\varepsilon/(n\delta_1^2)) < \varepsilon$ for all sufficiently large n. Fix a constant $D_\varepsilon > 1$, whose value will be specified later. If $n\delta_1^2 \leq D_\varepsilon$, then the conclusion follows trivially by choosing $L_\varepsilon \geq D_\varepsilon$. Henceforth it is assumed that $n\delta_1^2 > D_\varepsilon$. As we go along in our proof, we will indicate the choices to be made for L_ε and D_ε.

We are interested in the following probability:

$$\mathbb{P}(|\hat{\tau} - \tau| > L_\varepsilon/(n\delta_1^2)) \leq \mathbb{P}\left(\max_{i:|i-\tilde{k}|>L_\varepsilon/\delta_1^2} \sum_{j=1}^d |V_{i,j}^X| > \sum_{j=1}^d |V_{\tilde{k},j}^X| \right). \tag{10.6}$$

We will bound (10.6) with a similar, but much more general argument compared to the first part of the proposition. In particular, the presence of $\sum_{j=1}^d$ on both sides of the inequality of the event

$$\mathcal{X} := \sum_{j=1}^d |V_{i,j}^X| > \sum_{j=1}^d |V_{\tilde{k},j}^X|,$$

necessitates the introduction of some notations. Let $k_3 := \arg\max_{i:|i-\tilde{k}|>L_\varepsilon/\delta_1^2} \sum_{j=1}^d |V_{i,j}^X|$. Define the random variables

$$\alpha_j := I\{V_{k_3,j}^X \geq 0\} - I\{V_{k_3,j}^X < 0\} \ ; \ \beta_j := I\{V_{\tilde{k},j}^X \geq 0\} - I\{V_{\tilde{k},j}^X < 0\}. \tag{10.7}$$

Obviously both $\alpha_j, \beta_j \in \{-1, 1\}$ for $1 \le j \le d$. Suppose $1 \le j^\star \le d$ be such that $\delta_{j^\star} \ge \delta_1/d > \delta_{j^\star+1}$. In particular, $j^\star = d$ if $\delta_d \ge \delta_1/d$. Write (10.6) as

$$\mathbb{P}(\mathcal{X}) = \mathbb{P}(\underbrace{\mathcal{X}, \exists j_0 \le j^\star \text{ such that } \alpha_{j_0} = 1}_{\mathcal{X}_1}) + \mathbb{P}(\underbrace{\mathcal{X}, \alpha_1 = \ldots = \alpha_{j^\star} = -1}_{\mathcal{X}_2}) := \mathbb{P}(\mathcal{X}_1) + \mathbb{P}(\mathcal{X}_2). \quad (10.8)$$

We tackle the two terms $\mathbb{P}(\mathcal{X}_1)$ and $\mathbb{P}(\mathcal{X}_2)$ in (10.8) one-by-one. For a fixed $1 \le i \le n$, define the function $f_{n,j}(i) = \mathbb{E}[V^X_{i,j}]$. Further, for $1 \le j \le d$, let us denote $A_j := V^e_{k_3,j}$, $B_j := -f_{n,j}(k_3)$, $C_j := V^e_{\tilde{k},j}$, and $D_j := -f_{n,j}(\tilde{k})$. From (10.4) and $\delta_1 \ge \ldots \ge \delta_d \ge 0$, it follows

$$D_1 \ge \ldots \ge D_d \ge 0,\ B_1 \ge \ldots \ge B_d \ge 0, \text{ and } D_j \ge \gamma B_j \text{ for all } 1 \le j \le d \text{ and } \gamma \in \{-1, 1\}. \quad (10.9)$$

Moreover, (10.4) instructs $V^X_{k_3,j} = A_j - B_j$, and $V_{\tilde{k},j} = C_j - D_j$. We will leverage (10.9) to make $\mathbb{P}(\mathcal{X}_1)$ and $\mathbb{P}(\mathcal{X}_2)$ amenable to Lemmas 10.1 and 10.2. To begin with, observe that $\mathcal{X}$ can be written as

$$\sum_{j=1}^d \alpha_j(A_j - B_j) - \sum_{j=1}^d \beta_j(C_j - D_j) > 0. \quad (10.10)$$

For $\mathbb{P}(\mathcal{X}_1)$, we note that, (10.10) implies,

$$\mathcal{X}_1 \implies \sum_{j=1}^d (\alpha_j A_j + C_j) > \sum_{j=1}^d (D_j + \alpha_j B_j) = D_{j_0} + B_{j_0} + \sum_{j \ne j_0} (D_j + \alpha_j B_j) > D_{j^\star} \ge n\tau(1-\tau)\delta_1/d, \quad (10.11)$$

where the first implication is due to $\sum_{j=1}^d \beta_j(C_j - D_j) > \sum_{j=1}^d (C_j - D_j)$, the second inequality follows from (10.9), and the final inequality holds by definition of $j^\star$ and D_j. Noting that

$$\sum_{j=1}^d (\alpha_j A_j + C_j) \le 2d \max_{1 \le j \le d} \max_{1 \le i \le n} |V^e_{ij}|,$$

from (10.11), we finally have

$$\mathcal{X}_1 \implies \max_{1\le j\le d}\max_{1\le i\le n}|V_{ij}^e| > n\delta_1 C_d,$$

where $C_d = \tau(1-\tau)/(2d^2)$. Applying Lemma 10.1 and Markov's inequality, for a constant $C_{p,d}$ depending on p and d we obtain

$$\mathbb{P}(\mathcal{X}_1) \le C_{p,d}\frac{\Theta_{0,p}^2}{n\delta_1^2} \le C_{p,d}\frac{\Theta_{0,p}^2}{D_\varepsilon} < \varepsilon, \tag{10.12}$$

where the final inequality is guaranteed by choosing δ_1 large enough.

Now we focus on tackling $\mathbb{P}(\mathcal{X}_2)$. Let us define the random sets $\mathcal{A} := \{1 \le j \le d : \beta_j = 1\}$, and $\mathcal{B} := \{1 \le j \le d : \alpha_j = -1\}$. Observe that

$$\begin{aligned}
\sum_{j=1}^d \beta_j(C_j - D_j) &= \sum_{j\in\mathcal{A}}(C_j - D_j) - \sum_{j\notin\mathcal{A}}(C_j - D_j)\\
&\ge -\sum_{j\in\mathcal{A}\cap\mathcal{B}}(C_j - D_j) + \sum_{j\in\mathcal{A}\cap\mathcal{B}^c}(C_j - D_j) - \sum_{j\in\mathcal{A}^c\cap\mathcal{B}}(C_j - D_j) + \sum_{j\in\mathcal{A}^c\cap\mathcal{B}^c}(C_j - D_j)\\
&= -\sum_{j\in\mathcal{B}}(C_j - D_j) + \sum_{j\in\mathcal{B}^c}(C_j - D_j),
\end{aligned}$$

where for the inequality we have used $C_j - D_j \ge 0$ for $j \in \mathcal{A}$, and $C_j - D_j < 0$ for $j \in \mathcal{A}^c$. Therefore, from (10.10) one obtains

$$\begin{aligned}
&-\sum_{j\in\mathcal{B}}(A_j - B_j) + \sum_{j\in\mathcal{B}^c}(A_j - B_j) + \sum_{j\in\mathcal{B}}(C_j - D_j) - \sum_{j\in\mathcal{B}^c}(C_j - D_j) > 0\\
\implies &\sum_{j\in\mathcal{B}}(-A_j + C_j) + \sum_{j\in\mathcal{B}^c}(A_j - C_j) > \sum_{j\in\mathcal{B}}(D_j - B_j) - \sum_{j\in\mathcal{B}^c}(D_j - B_j)\\
&\qquad\qquad \ge \begin{cases}(1-\tau)(\tilde{k} - k_3)(\sum_{j\in\mathcal{B}}\delta_j - \sum_{j\in\mathcal{B}^c}\delta_j), & k_3 \le k,\\ \tau(k_3 - \tilde{k})(\sum_{j\in\mathcal{B}}\delta_j - \sum_{j\in\mathcal{B}^c}\delta_j), & k_3 > k.\end{cases}
\end{aligned} \tag{10.13}$$

Now, for $\mathbb{P}(\mathcal{X}_2)$, note that since $\alpha_1 = \ldots = \alpha_{j^\star} = -1$, therefore $\{1, \ldots, j^\star\} \subseteq \mathcal{B}$. Moreover, by definition of $j^\star$, $\sum_{j=j^\star+1}^{d}\delta_j \leq (1-j^\star/d)\delta_1$, and hence,

$$\sum_{j\in\mathcal{B}}\delta_j - \sum_{j\in\mathcal{B}^c}\delta_j \geq \sum_{j=1}^{j^\star}\delta_j - \sum_{j=j^\star+1}^{d}\delta_j \geq \frac{j^\star}{d}\delta_1 + \ldots + \delta_{j^\star} \geq \frac{\delta_1}{d}.$$

In view of this, (10.13) entails

$$\mathcal{X}_2 \implies \max_{k:|k-\tilde{k}|>L_\varepsilon/\delta_1^2} \frac{\sum_{j=1}^{d}|V_{k,j}^e - V_{\tilde{k},j}^e|}{|k-\tilde{k}|} \geq \frac{\delta_1}{d}\min\{\tau, 1-\tau\}.$$

Therefore, using Lemma 10.2 and $|k-\tilde{k}| > L_\varepsilon/\delta_1^2$, we have

$$\mathbb{P}(\mathcal{X}_2) \leq C_{p,d}\frac{\Theta_{0,p}^2}{\min\{\tau^2,(1-\tau)^2\}L_\varepsilon} < \varepsilon, \tag{10.14}$$

where the last inequality is guaranteed by choosing large enough L_ε. The proof is complete in light of (10.12) and (10.14). □

Now we proceed towards the proof of Proposition 2.2.

Proof of Proposition 2.2. Write $T_n = \sum_{j=1}^{d}(|V_{n\hat{\tau}_j,j}^X| - |V_{n\hat{\tau},j}^X|)/\sqrt{n}$. We tackle the validity and consistency of our test separately.

10.1 Behavior under H_0: Validity

For each $1 \leq j \leq d$, we will show that $|V_{n\hat{\tau}_j,j}^X| - |V_{n\hat{\tau},j}^X| = O_{\mathbb{P}}(\sqrt{n})$. Henceforth in this subsection, we will fix j. In light of (10.4) and d being fixed, we have

$$||V_{n\hat{\tau}_j,j}^X| - |V_{n\hat{\tau},j}^X|| \leq (|V_{n\hat{\tau}_j,j}^e| + |V_{n\hat{\tau},j}^e|) + |f_{n,j}(n\hat{\tau}_j) - f_{n,j}(n\hat{\tau})|. \tag{10.15}$$

Lemma 10.1 instructs that

$$\|\max_{1\leq i\leq n} V_{i,j}^e\|_p = O(\sqrt{n}\Theta_{0,p}), \tag{10.16}$$

which takes care of the first term in the RHS of (10.15). The second term is tackled as follows. Recall the convention that change-points are synchronized for dimensions with $\delta_j = 0$. If $\tau_1 = \ldots = \tau_d = \tau$, then Proposition 2.1 implies that $|\hat{\tau}_j - \hat{\tau}| = O_{\mathbb{P}}(\min\{1/(n\delta_j^2), 1\})$. Following (10.4), this assertion further yields that

$$\Big||f_{n,j}(n\hat{\tau}_j)| - |f_{n,j}(n\hat{\tau})|\Big| \leq 2(\tau \vee (1-\tau))n|\delta_j|O_{\mathbb{P}}(\frac{1}{1 \vee n\delta_j^2}) = O_{\mathbb{P}}(\sqrt{n} \wedge |\delta_j|^{-1}). \tag{10.17}$$

This completes the proof of validity under H_0 in light of (10.15) and (10.16).

10.2 Behavior under H_0^c: Consistency

Recall $\mathcal{H}$ from the statement of Proposition 2.2, and in view of (2.9), consider j_1, j_2 such that $n\delta_j^2 \to \infty$ for $j \in \{j_1, j_2\}$. Observe that for all $1 \leq j \leq d$, $|V^X_{n\hat{\tau}_j,j}| \geq |V^X_{n\hat{\tau},j}|$. Therefore, from (10.4) it is enough to show that

$$n^{-1/2} \sum_{j \in \{j_1, j_2\}} (|V^e_{n\hat{\tau}_j,j} + f_{n,j}(n\hat{\tau}_j)| - |V^e_{n\hat{\tau},j} + f_{n,j}(n\hat{\tau})|) \xrightarrow{\mathbb{P}} \infty. \tag{10.18}$$

Note that, (10.16) implies that $|n^{-1/2}V^e_{n\hat{\tau}_j,j}| = O_{\mathbb{P}}(1)$, and $|n^{-1/2}V^e_{n\hat{\tau},j}| = O_{\mathbb{P}}(1)$. Therefore, we focus on characterizing how far off $f_{n,j}(n\hat{\tau})$ can be from $f_{n,j}(n\hat{\tau}_j)$. Note that $|f_{n,j}(n\tau_j)| \geq |f_{n,j}(n\hat{\tau})|$ always. Moreover, for $j \in \{j_1, j_2\}$, it holds always that

$$|f_{n,j}(n\tau_j)| - |f_{n,j}(n\hat{\tau})|/\sqrt{n} \geq \sqrt{n}|\delta_j|\mathsf{C}_j|\tau_j - \hat{\tau}|, \tag{10.19}$$

where $\mathsf{C}_j := \min\{\tau_j(1-\tau_j), (1-\tau_j)\tau_j\}$. Therefore, from (10.19) it follows almost surely

$$\begin{aligned} n^{-1/2} \sum_{j \in \{j_1, j_2\}} (|f_{n,j}(n\tau_j)| - |f_{n,j}(n\hat{\tau})|) &\geq \sqrt{n} \sum_{j \in \{j_1, j_2\}} |\delta_j|\mathsf{C}_j|\tau_j - \hat{\tau}| \\ &\geq \mathsf{C}\sqrt{n} \min_{j \in \{j_1, j_2\}} \mathsf{C}|\delta_j| \to \infty, \end{aligned} \tag{10.20}$$

where $\mathsf{C} := \min\{|\tau_{j_1} - \tau_{j_2}| : \{j_1, j_2\} \in \mathcal{H}, n(\delta_{j_1}^2 \wedge \delta_{j_2}^2) \to \infty\} > 0$ is a constant, and the limiting assertion follows from (2.9). Moreover, noting that an argument similar to (10.17) along with $|\hat{\tau}_j - \tau_j| = O_{\mathbb{P}}((n\delta_j^2)^{-1} \wedge 1)$, we have

$$n^{-1/2} \sum_{j \in \{j_1, j_2\}} (|f_{n,j}(n\hat{\tau}_j)| - |f_{n,j}(n\tau_j)|) = O_{\mathbb{P}}(1). \tag{10.21}$$

From (10.20), along with (10.21), we obtain

$$n^{-1/2} \sum_{j \in \{j_1, j_2\}} (|f_{n,j}(n\hat{\tau}_j)| - |f_{n,j}(n\hat{\tau})|) \xrightarrow{\mathbb{P}} \infty, \tag{10.22}$$

which completes the proof of (10.18). $\square$

11 Proofs of Section 3

This section is devoted to the proofs of the results appearing in Section 3.

11.1 Proof of Theorem 3.1

The proof of this theorem is similar to [19], [62] and [23]. While [19] deals with univariate data, there is a technical challenge to extend these to multivariate scenario. A later work [62] achieves that. However, they work on a non-stationary set-up leaving an opportunity to streamline the proof and relax some conditions therein. Another recent work [23] on univariate non-stationary processes relaxes some of the conditions for the non-stationary processes. Since we are re-purposing the proof, we will be using some technical results from [23] and avoid unnecessary technical details. For better read, we divide the proof in the following two stages.

- In Section 11.1.1, we first obtain an optimal (rate-wise) Gaussian approximation that might not be regularized and granular as $G_i = \sum_{j=1}^{i} Z_j$ with each Z_j having iid multivariate normal distribution. However, it maintains the same optimal rate $n^{1/p}$ matching the final rate at (3.2).

- Then in Section 11.1.2, we regularize the approximating Gaussian process in the correct granular structure while keeping the optimal rate intact.

These main steps are similar to the 6 steps outlined in Section 2.6 of [23] towards the proof of Theorem 2.5 therein. We remind the readers thar we use $|x|$ to denote the Euclidean norm of $x \in \mathbb{R}^d$ for $d \geq 1$.

Proof of Theorem 3.1. **11.1.1 Possibly unregularized Gaussian Approximation**

Recall A_0 from (3.1). With $A > A_0$, we invoke the following equations from [23].

$$L = \frac{f_1 - f_2 + A\sqrt{(p-2)(f_3 - 3p)}}{Af_4}, \qquad (11.1)$$
$$\gamma = \frac{(2p+p^2)A + p^2 + 3p + 2 + \sqrt{f_5}}{2 + 2p + 4A},$$

with $f_1(p, A) = Ap^2(A + 1)$, $f_2(p, A) = A(2pA + 3p - 2)$, $f_3(p, A) = p^3(1 + A)^2 + 6f_1 + 4pA - 2$, $f_4(p, A) = 2p(2pA^2 + 3pA + p - 2)$ and $f_5(p, A) = p^2(p^2 + 4p - 12)A^2 + 2p(p^3 + p^2 - 4p - 4)A + (p^2 - p - 2)^2$. Our choice of L and γ satisfies the following relations:

$$\frac{1}{2} - \frac{1}{p} - LA < 0, \qquad (11.2)$$

$$L\left(\frac{\gamma}{2} - 1\right) + 1 - \frac{\gamma}{p} < 0, \qquad (11.3)$$

$$p < \gamma < 2(1 + p + pA)/3, \qquad (11.4)$$

$$1/p - 1/\gamma + L - L(A + 1)p/\gamma = 0. \qquad (11.5)$$

For $w \in \mathbb{R}^d$, define the truncation operator $T_b(w) = (T_b(w_1), T_b(w_2), \cdots, T_b(w_d))$ where $T_b(x) = \max\{\min\{x, b\}, -b\}$. The following series of results, versions of analogous results in [62, 23], are instrumental towards our argument.

Lemma 11.1 (Lemma 7.1 of [62]). *For the truncated process, it holds that*

$$\mathbb{E}(|T_{n^{1/p}}(\mathbf{e}_i)|^\gamma) = o(n^{\gamma/p-1}).$$

Proposition 11.1 (Lemma 7.3 of [62]). *Assume 3.1, along with* (11.2), (11.3), (11.4) *and* (11.5) *for A, L and γ. Let $m = \lfloor n^L \rfloor$ and let*

$$\tilde{R}_{s,t} = \tilde{\mathbf{e}}_s + \cdots + \tilde{\mathbf{e}}_{s+t},$$

where $\tilde{X}_i$ is as defined in (11.9). *Then*

$$\max_s \mathbb{E}\left[\max_{1\le t\le m} |\tilde{R}_{s,t}|^\gamma\right] = O(m^{\gamma/2}). \tag{11.6}$$

Lemma 11.2 (Lemma 9.1 of [23]). *Under the assumption of Theorem 3.1,*

$$\min_{l\ge 1}\{\Theta_{l,p} + ln^{2/p-1}\} = o\left(\frac{n^{1/p-1/2}}{\sqrt{\log\log n}}\right). \tag{11.7}$$

Define the truncated partial sum process $\{S_i^\oplus\}_{i=1}^n$ as $S_i^\oplus := \sum_{j=1}^i \left(\mathbf{e}_i^\oplus - \mathbb{E}(\mathbf{e}_i^\oplus)\right)$. From the stationary causal representation (2.5), it is easy to observe that

$$\max_{1\le i\le n} |S_i^e - S_i^\oplus| = o_{\mathbb{P}}(n^{1/p}). \tag{11.8}$$

We further define an m-dependent process $(\tilde{\mathbf{e}}_i)_{i=1}^n$ as

$$\tilde{\mathbf{e}}_i = \mathbb{E}(\tilde{\mathbf{e}}_i^\oplus|\varepsilon_j, \cdots, \varepsilon_{j-m}) - \mathbb{E}(\tilde{\mathbf{e}}_i). \tag{11.9}$$

Let $\tilde{S}_i = \sum_{j=1}^i \tilde{\mathbf{e}}_j$. Using Lemma A1 of [73] and (11.2), we have

$$\|\max_{1\le i\le n} |S_i^\oplus - \tilde{S}_i|\|_p \le c_p n^{1/2}\Theta_{1+m,p} = o(n^{1/p}), \tag{11.10}$$

which implies $\max_{1\le i\le n}|S_i^{\oplus}-\tilde{S}_i| = o_{\mathbb{P}}(n^{1/p})$. Let $q_n \asymp n/m$ and $l_i = \lfloor \frac{q_i}{3} \rfloor$. Denote

$$K = \lceil \frac{q_n}{3} \rceil,\ \tilde{B}_k = \sum_{j=(k-1)m+1}^{km\wedge n} \tilde{\mathbf{e}}_j \text{ for } 1 \le k \le q_n. \tag{11.11}$$

For the blocking approximation we will approximate the partial sum process $\tilde{S}_i$ by $S_i^{\diamond} = \sum_{l=1}^{l_i}(\tilde{B}_{3l-2} + \tilde{B}_{3l-1} + \tilde{B}_{3l})$. Let $\tilde{S}_{k,l} = \sum_{j=k+1}^{l} \tilde{\mathbf{e}}_j$. Following an argument exactly same as Proposition 8.3 of [23], one obtains

$$\max_{1\le i\le n} |\tilde{S}_i - S_i^{\diamond}| = o_{\mathbb{P}}(n^{1/p}). \tag{11.12}$$

Denote by $\boldsymbol{a}$ the sequence $(\cdots, \boldsymbol{a}_0, \boldsymbol{a}_3, \cdots)$ with $\boldsymbol{a}_{3k} = (a_{(3k-1)m+1}, \cdots, a_{3km})$. Let $\boldsymbol{\eta} = (\cdots, \boldsymbol{\eta}_0, \boldsymbol{\eta}_3, \cdots)$, where $\boldsymbol{\eta}_k = (\varepsilon_{(k-1)m+1}, \cdots, \varepsilon_{km})$. Let g be a measurable function such that $\tilde{\mathbf{e}}_i = g(\varepsilon_{i-m}, \cdots, \varepsilon_i)$. For $1 \le k \le K$, define the random functions,

$$\begin{aligned}
\tilde{B}_{3k-2}(\boldsymbol{a}_{3k-3}) &= \sum_{i=(3k-3)m+1}^{(3k-2)m} g(a_{i-m}, \cdots, a_{(3k-3)m}, \varepsilon_{(3k-3)m+1}, \cdots, \varepsilon_i),\\
\tilde{B}_{3k-1} &= \sum_{i=(3k-2)m+1}^{(3k-1)m} g(\varepsilon_{i-m}, \cdots, \varepsilon_{(3k-2)m}, \cdots, \varepsilon_i),\\
\tilde{B}_{3k}(\boldsymbol{a}_{3k}) &= \sum_{i=(3k-1)m+1}^{3km} g(\varepsilon_{i-m}, \cdots, \varepsilon_{(3k-1)m}, a_{(3k-1)m+1}, \cdots, a_i).
\end{aligned}$$

Let $M_{3l}(\boldsymbol{a}_{3l}) = \mathbb{E}(\tilde{B}_{3l}(\boldsymbol{a}_{3l}))$, and $M_{3l-2}(\boldsymbol{a}_{3l-3}) = \mathbb{E}(\tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}))$. For $1 \le l \le K$, let

$$Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}) := \tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}) - M_{3l-2}(\boldsymbol{a}_{3l-3}) + \tilde{B}_{3l-1} + \tilde{B}_{3l}(\boldsymbol{a}_{3l}) - M_{3k}(\boldsymbol{a}_{3l}). \tag{11.13}$$

Subsequently, we will denote $Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})$ by $Y_l^{\boldsymbol{a}}$ to emphasize the dependency on $\boldsymbol{a}$. Due to our conditioning, Y_l's are independent. The corresponding mean and variance functionals, for

$1 \le k \le K$, are

$$M_k(\boldsymbol{a}) = \sum_{l=1}^{k} [M_{3l-2}(\boldsymbol{a}_{3l-3}) + M_{3l}(\boldsymbol{a}_{3l})], \tag{11.14}$$

$$Q_k(\boldsymbol{a}) = \sum_{l=1}^{k} V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}), \tag{11.15}$$

where $V_l^{\boldsymbol{a}} := V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}) = \mathbb{E}[Y_l^{\boldsymbol{a}} Y_l^{\boldsymbol{a}^\top}]$. Let $C_{3l-1}(\boldsymbol{\eta}_{3l-2}) = \mathbb{E}[\tilde{B}_{3l-1} | \boldsymbol{\eta}_{3l-2}]$. We will decompose V_l as follows:

$$\begin{aligned}
V_l^{\boldsymbol{a}} &= \mathbb{E}(Y_l^{\boldsymbol{a}} Y_l^{\boldsymbol{a}^\top}) \\
&= \mathbb{E}\left[\left(\mathbb{E}[Y_l^{\boldsymbol{a}} | \boldsymbol{\eta}_{3l-2}, \boldsymbol{\eta}_{3l-1}] - \mathbb{E}[Y_l^{\boldsymbol{a}} | \boldsymbol{\eta}_{3l-2}] \right) \left(\mathbb{E}[Y_l^{\boldsymbol{a}} | \boldsymbol{\eta}_{3l-2}, \boldsymbol{\eta}_{3l-1}] - \mathbb{E}[Y_l^{\boldsymbol{a}} | \boldsymbol{\eta}_{3l-2}] \right)^\top \right] \\
&\quad + \mathbb{E}\left[\left(\mathbb{E}[Y_l^{\boldsymbol{a}} Y_l^{\boldsymbol{a}^\top} | \boldsymbol{\eta}_{3l-2}] \right)^2 \right] \\
&= \mathbb{E}\left[\left(\tilde{B}_{3l-1} - C_{3l-1}(\boldsymbol{\eta}_{3l-2}) + \tilde{B}_{3l}(\boldsymbol{a}_{3l}) - M_{3l}(\boldsymbol{a}_{3l}) \right) \left(\tilde{B}_{3l-1} - C_{3l-1}(\boldsymbol{\eta}_{3l-2}) + \tilde{B}_{3l}(\boldsymbol{a}_{3l}) - M_{3l}(\boldsymbol{a}_{3l}) \right)^\top \right] \\
&\quad + \mathbb{E}\left[\left(\tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}) - M_{3l-2}(\boldsymbol{a}_{3l-3}) + C_{3l-1}(\boldsymbol{\eta}_{3l-2}) \right) \left(\tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}) - M_{3l-2}(\boldsymbol{a}_{3l-3}) + C_{3l-1}(\boldsymbol{\eta}_{3l-2}) \right)^\top \right] \\
&:= \tilde{V}_{2l}(\boldsymbol{a}_{3l}) + \tilde{V}_{2l-1}(\boldsymbol{a}_{3l-3}).
\end{aligned} \tag{11.16}$$

Let us accumulate $V_l^0(\boldsymbol{a}_{3l}) = \tilde{V}_{2l}(\boldsymbol{a}_{3l}) + \tilde{V}_{2l+1}(\boldsymbol{a}_{3l})$. Then, for all $t \in \mathbb{N}$,

$$\sum_{l=1}^{t} V_l(\boldsymbol{a}_{3l}, \boldsymbol{a}_{3l+3}) = \tilde{V}_1(\boldsymbol{a}_0) + \sum_{l=1}^{t-1} V_l^0(\boldsymbol{a}_{3l}) + \tilde{V}_{2t}(\boldsymbol{a}_{3t}). \tag{11.17}$$

We will invoke Proposition 11.6 with $\nu_l = lJ$, and $s = K/J$, where $J = \lfloor K^{2/\gamma} / \log^2 K \rfloor$, which allows us to apply Theorem 4 of [44] on a set $\mathcal{A}$ such that $\mathbb{P}(\boldsymbol{a} \in \mathcal{A}) \to 1$. Therefore, we obtain

$$\mathbb{P}_{\boldsymbol{a}} \left(\max_{1 \le i \le n} |\Gamma_i^{\boldsymbol{a}} - D_i^{\boldsymbol{a}}| \ge cx \right) \le C_0 \frac{L_\gamma^{\boldsymbol{a}}}{x^\gamma}, \tag{11.18}$$

where $\Gamma_i^{\boldsymbol{a}} := \sum_{l=1}^{l_i} Y_l^{\boldsymbol{a}}$, and $D_i^{\boldsymbol{a}} := \sum_{l=1}^{l_i} W_l^{\boldsymbol{a}}$, where $W_l^{\boldsymbol{a}} \overset{ind}{\sim} N(\mathbf{0}, V_l^{\boldsymbol{a}})$. Observe that, $\mathbb{E}[L_\gamma(\boldsymbol{\eta}] \le$

$c_\gamma K \max_l \mathbb{E}[|\tilde{S}_{l,m+l}|^\gamma] = O(nm^{\gamma/2-1})$, which is $o(n^{\gamma/p})$ using (11.3). Thus, putting $x = n^{1/p}$ yields

$$\max_{1\le i\le n} |\Gamma_i^{\boldsymbol{\eta}} - D_i^{\boldsymbol{\eta}}| = o_{\mathbb{P}}(n^{1/p}). \tag{11.19}$$

The probability space for the above convergence is exactly same as that in [19], as well as [23]. Observe that we can write $D_i^{\boldsymbol{a}} \overset{d}{=} \omega_i^{\boldsymbol{a}} + R_i^{\boldsymbol{a}}$, with $\omega_k^{\boldsymbol{a}} = \sum_{l=1}^{k-1} V_l^{0^{1/2}}(\boldsymbol{a}_{3l}) Z_l^{\boldsymbol{a}}$, and let $R_k^{\boldsymbol{a}} = \tilde{V}_1^{1/2}(\boldsymbol{a}_0) Z_0^a + \tilde{V}_{2k}^{1/2}(\boldsymbol{a}_{3k}) Z_k^a$, and $(Z_l^a)_{l=1}^K$ are i.i.d standard Gaussian random variables. Clearly,

$$\mathbb{P}(\max_{1\le k\le K} |\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| \ge cn^{2/p}) \le \sum_{k=1}^K \mathbb{P}(|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| \ge cn^{2/p}) \le c^{-\gamma/2} \frac{n}{3m} n^{-\gamma/p} \|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})\|^{\gamma/2} = o(1). \tag{11.20}$$

Similarly, $|\tilde{V}_1(\boldsymbol{\eta}_0)| = o_{\mathbb{P}}(n^{2/p})$. Therefore, we can construct i.i.d. $Z_l^{\boldsymbol{\eta}} \sim N(\mathbf{0}, I_d)$, independent of ε's, such that

$$\max_{1\le i\le n} |D_i^{\boldsymbol{\eta}} - \Phi_i^{\boldsymbol{\eta}}| = o_{\mathbb{P}}(n^{1/p}), \tag{11.21}$$

where $\Phi_i^{\boldsymbol{\eta}} = \sum_{l=1}^{l_i-1} V_l^{0^{1/2}}(\boldsymbol{\eta}_{3l}) Z_l^{\boldsymbol{\eta}}$. Together with (11.19), (11.21) implies that

$$\max_{1\le i\le n} |\Gamma_i^{\boldsymbol{\eta}} - \Phi_i^{\boldsymbol{\eta}}| = o_{\mathbb{P}}(n^{1/p}). \tag{11.22}$$

In view of $(\Gamma_i(\boldsymbol{\eta}) + M_{l_i}(\boldsymbol{\eta}))_{1\le i\le n} \overset{d}{=} (S_i^\diamond)_{1\le i\le n}$ and (11.22), we need to prove strong invariance for $\Phi_i + M_{l_i}^{\boldsymbol{\eta}}$. Let for $1 \le l \le K$,

$$\tilde{A}_l = V_l^0(\boldsymbol{\eta}_{3l})^{1/2} Z_l + M_{3l}(\boldsymbol{\eta}_{3l}) + M_{3l+1}(\boldsymbol{\eta}_{3l}). \tag{11.23}$$

Let $S_i^\natural = \sum_{l=1}^{l_i} \tilde{A}_l$. Note that by the same argument as in (11.20), we have

$$\max_{1\le i\le n} |\Phi_i + M_{l_i}(\boldsymbol{\eta}) - S_i^\natural| = \max_{1\le i\le n} |V_{l_i}^0(\boldsymbol{\eta}_{3l_i})^{1/2} Z_{l_i} - M_{3l_i+1}\left(\boldsymbol{\eta}_{3l_i}\right) + M_1\left(\boldsymbol{\eta}_0\right)| = o_{\mathbb{P}}(n^{1/p}). \tag{11.24}$$

One can easily verify that with our choice of ν_k and s, conditions (11.67) and (11.68) are satisfied by this new process $S_i^\natural$. Hence, by Theorem 4 of [44], we have a Gaussian process $G_i \sim N(\mathbf{0}, \sum_{l=1}^{l_i} \mathrm{Var}(\tilde{A}_l\tilde{A}_l^\top))$ such that

$$\max_{1\le i\le n} |S_i^\natural - G_i| = o_{\mathbb{P}}(n^{1/p}). \tag{11.25}$$

11.1.2 Regularizing Gaussian Approximation

Observe that

$$\begin{aligned} V_l^0(\boldsymbol{a}_{3l}) =& \mathbb{E}[\tilde{B}_{3l-1}\tilde{B}_{3l-1}^\top] + \mathbb{E}[\tilde{B}_{3l}(\boldsymbol{a}_{3l})\tilde{B}_{3l}(\boldsymbol{a}_{3l})^\top] - M_{3l}(\boldsymbol{a}_{3l})M_{3l}(\boldsymbol{a}_{3l})^\top + \|\tilde{B}_{3l+1}(\boldsymbol{a}_{3l})\|^2 \\ & - M_{3l+1}(\boldsymbol{a}_{3l})M_{3l+1}(\boldsymbol{a}_{3l})^\top - \mathbb{E}[C_{3l-1}(\boldsymbol{\eta}_{3l-2})C_{3l-1}(\boldsymbol{\eta}_{3l-2})^\top] + \mathbb{E}[C_{3l+2}(\boldsymbol{\eta}_{3l+1})C_{3l+2}(\boldsymbol{\eta}_{3l+1})^\top] \\ & + \mathbb{E}\left[\tilde{B}_{3l-1}\tilde{B}_{3l}(\boldsymbol{a}_{3l})^\top\right] + \mathbb{E}\left[\tilde{B}_{3l}(\boldsymbol{a}_{3l})\tilde{B}_{3l-1}^\top\right] \\ & + \mathbb{E}\left[\tilde{B}_{3l+1}(\boldsymbol{a}_{3l})C_{3l+2}(\boldsymbol{\eta}_{3l+1})^\top\right] + \mathbb{E}\left[C_{3l+2}(\boldsymbol{\eta}_{3l+1})\tilde{B}_{3l+1}(\boldsymbol{a}_{3l})^\top\right] \end{aligned} \tag{11.26}$$

Therefore,

$$\begin{aligned} \tilde{v}_l := \mathbb{E}[\tilde{A}_l\tilde{A}_l^\top] =& \mathbb{E}[\tilde{B}_{3l-1}\tilde{B}_{3l-1}^\top] + \mathbb{E}[\tilde{B}_{3l}\tilde{B}_{3l}^\top] + \mathbb{E}[\tilde{B}_{3l+1}\tilde{B}_{3l+1}^\top] + \mathbb{E}[\tilde{B}_{3l-1}\tilde{B}_{3l}^\top] + \mathbb{E}[\tilde{B}_{3l}\tilde{B}_{3l-1}^\top] + \mathbb{E}[\tilde{B}_{3l}\tilde{B}_{3l+1}^\top] \\ & + \mathbb{E}[\tilde{B}_{3l+1}\tilde{B}_{3l}^\top] + \mathbb{E}[\tilde{B}_{3l+1}\tilde{B}_{3l+2}^\top] + \mathbb{E}[\tilde{B}_{3l+2}\tilde{B}_{3l+1}^\top] - \mathbb{E}[C_{3l-1}(\boldsymbol{\eta}_{3l-2})C_{3l-1}(\boldsymbol{\eta}_{3l-2})^\top] \\ & + \mathbb{E}[C_{3l+2}(\boldsymbol{\eta}_{3l+1})C_{3l+2}(\boldsymbol{\eta}_{3l+1})^\top]. \end{aligned} \tag{11.27}$$

We will try to approximate $\tilde{v}_l$ by the long-run covariance Σ_∞. Rest of the proof is also heavily influenced by the arguments of Theorem 2.5 in [23]; however, for the sake of convenience, we present the complete proof with necessary modifications due to the multivariate structure of the theorem. To that end, let $B_k^\oplus = \sum_{j=(k-1)m+1}^{km} \left(\mathbf{e}_j^\oplus - \mathbb{E}(\mathbf{e}_j^\oplus)\right)$. Note that for all $j \ge 1$, $1 \le s \le d$, $\|\tilde{B}_{js} - B_{js}^\oplus\| \le \sqrt{m}\Theta_{m,2} \le \sqrt{m}\Theta_{m,p}$. Thus, using Cauchy-Schwarz, and $B_j^\oplus = \sum_{k=-\infty}^{j} P_k B_j^\oplus$, for $j \ge 1$, $1 \le s, t \le d$,

$$|\mathbb{E}[\tilde{B}_{js}\tilde{B}_{jt}] - \mathbb{E}[B_{js}^\oplus B_{jt}^\oplus| \le \|B_{js}^\oplus\|\|\tilde{B}_{jt} - B_{jt}^\oplus\| + \|\tilde{B}_{jt}\|\tilde{B}_{js} - B_{js}^\oplus\| \le 4m\Theta_{m,p}\Theta_{0,p}. \tag{11.28}$$

Therefore, $|\mathbb{E}[\tilde{B}_j\tilde{B}_j^\top] - \mathbb{E}[B_j^\oplus B_{jt}^{\oplus\top}| = O(m\Theta_{m,p})$. Further, note that for $k,l \ge 1$ and $1 \le s,t \le d$, we obtain

$$\begin{aligned}|\mathbb{E}(e_{ks}e_{lt} - e_{ks}^\oplus e_{lt}^\oplus)| =& |\mathbb{E}(e_{ks}e_{lt}\mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}\le n^{1/p}}) - \mathbb{E}(e_{ki}^\oplus e_{lj}^\oplus) + \mathbb{E}(e_{ki}e_{lj}\mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}}| \\ =& | - \mathbb{E}(e_{ks}^\oplus e_{lt}^\oplus \mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}}) + \mathbb{E}(e_{ks}e_{lt}\mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}}| \\ \le& |\mathbb{E}(e_{ks}^\oplus e_{lt}^\oplus \mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}})| + |\mathbb{E}(e_{ks}e_{lt}\mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}}| \\ \le& \mathbb{E}\left((|e_{ks}|^2 + |e_{lt}|^2)\mathbf{1}_{\max\{|e_{ks}|,|e_{lt}|\}>n^{1/p}}\right) = o(n^{2/p-1}).\end{aligned}$$

Next, observe that, for $l > k$,

$$\begin{aligned}|\mathbb{E}(e_{ks}e_{lt})| = |\sum_{i\in\mathbf{Z}}\sum_{j\in\mathbf{Z}}\mathbb{E}\left[(P_ie_{ks})(P_je_{lt})\right]| \le \sum_{i\in\mathbf{Z}}\|P_i(e_{ks})\|\|P_i(e_{lt})\| \\ \le \sum_{i=\infty}^{k}\delta_p(k-i)\delta_p(l-i) = \sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+l-k). \quad (11.29)\end{aligned}$$

Noting that $|\mathbb{E}(\mathbf{e}_i^\oplus)| = o(n^{1/p-1})$, we have, for a fixed $0 \le j \le m-1$ and $l \ge 0$,

$$\begin{aligned}|\mathbb{E}(B_{j+1}B_{j+1}^\top - B_{j+1}^\oplus B_{j+1}^{\oplus\top})| = \Bigg|&\sum_{k=1}^{m}\mathbb{E}(\mathbf{e}_{jm+k}\mathbf{e}_{jm+k}^\top - \mathbf{e}_{jm+k}^\oplus\mathbf{e}_{jm+k}^{\oplus\top}) + \sum_{s\ne t}^{m}\mathbb{E}(\mathbf{e}_{jm+s}\mathbf{e}_{jm+t}^\top - \mathbf{e}_{jm+s}^\oplus\mathbf{e}_{jm+t}^{\oplus\top}) \\ &- \left(\mathbb{E}[\sum_{k=1}^{m}\mathbf{e}_{jm+k}^\oplus]\right)\left(\mathbb{E}[\sum_{k=1}^{m}\mathbf{e}_{jm+k}^\oplus]\right)^\top\Bigg| \\ &\le o(mn^{2/p-1}) + O(lmn^{2/p-1} + m\sum_{s=l+1}^{\infty}\sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+s)^\top),\end{aligned}$$

where the last line follows from using the fact that there are $\le m$ terms of the form $\mathbb{E}(\mathbf{e}_k\mathbf{e}_{k+s}^\top - \mathbf{e}_k^\oplus\mathbf{e}_{k+s}^{\oplus\top})$ for a fixed $s \le m$ and applying (11.29). Note that $\sum_{j=l}^\infty\sum_{i=0}^\infty\delta_p(i)\delta_p(i+j) \le \Theta_{0,p}\Theta_{l,p}$. Further, using (11.29) repeatedly,

$$\begin{aligned}&|\mathbb{E}(B_kB_{k+1}^\top)| \\ &\le \sum_{j=0}^{2m}(m - |m-j|)\sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+j)\end{aligned}$$

$$= \sum_{i=0}^{\infty} \delta_p(i) \left(\delta_p(i+1) + 2\delta_p(i+2) + \cdots + m\delta_p(i+m) + (m-1)\delta_p(i+m+1) + \cdots + \delta_p(i+2m-1) \right)$$
$$\leq \sum_{i=0}^{\infty} \delta_p(i) \left(\Theta_{i+1,p} + \Theta_{i+2,p} + \cdots + \Theta_{i+2m-1,p} \right)$$
$$\leq \sum_{i=0}^{\infty} \delta_p(i) \sum_{j=1}^{2m-1} \Theta_{j,p} = \Theta_{0,p} \sum_{j=1}^{2m-1} \Theta_{j,p} = O\left(\sum_{j=1}^{2m-1} (j+1)^{-A} \right) = O(1), \tag{11.30}$$

since $A > 1$. Therefore,

$$\frac{\left| \tilde{v}_l - \mathbb{E}[B_{3l-1}B_{3l-1}^\top + B_{3l}B_{3l}^\top + B_{3l+1}B_{3l+1}^\top] \right|}{m} = O(\Theta_{m,p} + \min_{l \geq 1}\{ln^{2/p-1} + \Theta_{l,p}\}). \tag{11.31}$$

Observe that $\lim_{m\to\infty} \mathbb{E}[B_{3l}B_{3l}^\top]/m = \Sigma_\infty$. Thus in view of (11.31),

$$|\tilde{v}_l/m - \Sigma_\infty| = O(\Theta_{m,p} + \min_{l \geq 1}\{ln^{2/p-1} + \Theta_{l,p}\}) = o(1). \tag{11.32}$$

Recall G_i from (11.25). Clearly, in view of Lemma 11.3, we can write $G_i = \sum_{l=1}^{l_i} \sum_{k=(l-1)m+1}^{lm} \frac{\tilde{v}_l^{1/2}}{m^{1/2}} Z_k$ for some i.i.d. standard Gaussian random vectors Z_k, $1 \leq l \leq K$. Consider

$$\tilde{\mathbb{B}}_i = \sum_{l=1}^{l_i} \sum_{k=(l-1)m+1}^{lm} \frac{\tilde{v}_l^{1/2}}{m^{1/2}} Z_k + \sum_{k=l_i m+1}^{i} \frac{\tilde{v}_{l_i+1}^{1/2}}{m^{1/2}} Z_k,$$

and $\mathbb{B}_i = \sum_{k=1}^{i} \Sigma^{1/2} Z_k$. Firstly, using $\sup_l |\tilde{v}_l| = O(m)$, and (11.3),

$$\max_{1 \leq i \leq n} |\tilde{v}_{l_i+1}| = O(m) = o(n^{(\gamma/p-1)/(\gamma/2-1)}),$$

which immediately yields

$$\max_{1 \leq i \leq n} |G_i - \tilde{\mathbb{B}}_i| = o_{\mathbb{P}}(n^{(\gamma/p-1)/(\gamma-2)} \log n) = o_{\mathbb{P}}(n^{1/p}). \tag{11.33}$$

Now, in our final approximation step we will show $\tilde{\mathbb{B}}_i$ is close to $\mathbb{B}_i$. To that end, note that

$$\Psi_n := \mathrm{Var}(\tilde{\mathbb{B}}_n - \mathbb{B}_n) = \sum_{l=1}^{l_n} m\left(\frac{\tilde{v}_l^{1/2}}{m^{1/2}} - \Sigma_\infty^{1/2}\right)^2 + (n - l_n m)\left(\frac{\tilde{v}_{l_n+1}^{1/2}}{m^{1/2}} - \Sigma_\infty^{1/2}\right)^2. \tag{11.34}$$

Clearly, the condition $\lambda_{\min}(\Sigma_\infty) \geq c > 0$ and (11.32) imply that for all sufficiently large m, $\lambda_{\min}\left(\frac{\tilde{v}_l^{1/2}}{m^{1/2}} + \Sigma_\infty^{1/2}\right) > c/2$. Therefore, in view of (11.32),

$$\rho^\star\left(\frac{\tilde{v}_l^{1/2}}{m^{1/2}} - \Sigma_\infty^{1/2}\right)^2 \lesssim \frac{\rho^\star\left(\frac{\tilde{v}_l}{m} - \Sigma_\infty\right)^2}{\lambda_{\min}\left(\frac{\tilde{v}_l^{1/2}}{m^{1/2}} + \Sigma_\infty^{1/2}\right)^2} = O\left(m^{-2A} + \min_{l\geq 1}\{ln^{2/p-1} + \Theta_{l,p}\}^2\right),$$

which, together with Lemma 11.2 immediately yields,

$$\Psi_n = O\left(nm^{-2A} + n\left(\min_{l\geq 1}\{ln^{2/p-1} + \Theta_{l,p}\}\right)^2\right) = o(n^{2/p}/\log\log n). \tag{11.35}$$

Observing that $\tilde{\mathbb{B}}_n - \mathbb{B}_n$ is a Gaussian process with independent increments and applying the law of iterated logarithm, we obtain

$$\overline{\lim_{n\to\infty}}\left|\frac{\tilde{\mathbb{B}}_n - \mathbb{B}_n}{\sqrt{2\|\Psi_n\|\log\log\|\Psi_n\|}}\right| \leq 1 \text{ almost surely,}$$

which shows

$$\max_{1\leq i\leq n}|\tilde{\mathbb{B}}_i - \mathbb{B}_i| = o_{\mathbb{P}}(n^{1/p}) \tag{11.36}$$

in view of (11.35). Of course (11.36) is true if $\lim \Psi_n^2 < \infty$. The equations (11.8), (11.10), (11.12), (11.25), (11.33) and (11.36) combined completes the proof. $\square$

11.2 Proof of Lemma 3.1

Proof. Let $S_i = \sum_{j=1}^{i} X_j$ denote the partial sums of X_j's. Since the quantities $(S_i - i\bar{X}_{\cdot j})_{i=1}^{n}$ are invariant with respect to μ_j^L, therefore, without loss of generality, suppose $\mu_j^L = 0$ for $1 \leq j \leq d$. In light of Theorem 3.1, there exists independent random variables $\mathbf{Z}_i \sim N(\boldsymbol{\mu}_i, \Sigma_\infty)$ such that

$$\max_{1\leq i\leq n} |S_i - \sum_{k=1}^{i} \mathbf{Z}_i| = o_{\mathbb{P}}(n^{1/p}). \tag{11.37}$$

Write $|T_n - T_n^Z| = |W_1 - W_2|/\sqrt{n}$ where

$$\begin{aligned} W_1 &:= \sum_{j=1}^{d} \left[|S_{n\hat{\tau}_j,j} - n\hat{\tau}_j \bar{X}_{\cdot j}| - |S^Z_{n\tilde{\tau}_j^Z,j} - n\tilde{\tau}_j^Z \bar{Z}_{\cdot j}| \right], \\ W_2 &:= \sum_{j=1}^{d} \left[|S_{n\hat{\tau},j} - n\hat{\tau} \bar{X}_{\cdot j}| - S^Z_{n\tilde{\tau}^Z,j} - n\tilde{\tau}^Z \bar{Z}_{\cdot j}| \right]. \end{aligned}$$

For W_1, observe that by definition of $\hat{\tau}_j$ and $\tilde{\tau}_j$,

$$||S_{n\hat{\tau}_j,j} - n\hat{\tau}_j \bar{X}_{\cdot j}| - |S^Z_{n\tilde{\tau}_j^Z,j} - n\tilde{\tau}_j^Z \bar{Z}_{\cdot j}|| \leq \max_{1\leq i\leq n} |S_{ij} - i\bar{X}_{\cdot j} - S^Z_{ij} + i\bar{Z}_{\cdot j}|,$$

which entails, in light of (3.4) and (11.37), that

$$|S_{n\hat{\tau}_j,j} - n\hat{\tau}_j \bar{X}_{\cdot j}| - |S^Z_{n\tilde{\tau}_j^Z,j} - n\tilde{\tau}_j^Z \bar{Z}_{\cdot j}| = o_{\mathbb{P}}(n^{1/p}), \tag{11.38}$$

where we have used that $\Omega_n = 2 - 1/n$. Now we focus on W_2. Let $\boldsymbol{V}_i^X := (V_{i,1}^X, \ldots, V_{i,d}^X)^\top$, where $V_{i,j}^X$'s are defined as in (10.3). Similarly define $\boldsymbol{V}_i^Z$ based on $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. We use the notation $|\cdot|_{\mathbb{L}_1}$ for the vector L_1 norm. Note that $n\hat{\tau} := \arg\max_{1\leq i\leq n} |\boldsymbol{V}_i^X|_{\mathbb{L}_1}$, and similarly, $n\tilde{\tau}^Z := \arg\max_{1\leq i\leq n} |\boldsymbol{V}_i^Z|_{\mathbb{L}_1}$. Therefore, a similar treatment to (11.38) yields,

$$W_2 = \left| |\boldsymbol{V}^X_{n\hat{\tau}}|_{\mathbb{L}_1} - |\boldsymbol{V}^Z_{n\tilde{\tau}^Z}|_{\mathbb{L}_1} \right| \leq \max_{1\leq i\leq n} |\boldsymbol{V}_i^X - \boldsymbol{V}_i^Z|_{\mathbb{L}_1} = o_{\mathbb{P}}(n^{1/p}), \tag{11.39}$$

where we have used the uniform over $1 \le j \le d$ bound from Theorem 3.1 to obtain the $o_{\mathbb{P}}(n^{1/p})$ term. The proof is complete in light of (11.38) and (11.39). $\square$

11.3 Proofs of Section 3.2

In this section, we will prove Theorem 3.2, establishing the consistency of our estimate of long-run covariance matrix. We begin with a proof for the Proposition 3.1, which might be of independent interest.

Proof of Proposition 3.1. Let $\mathcal{B} := \{1 \le j \le d : \delta_j > 0\}$. To facilitate an intermediary oracle estimate, we define

$$\tilde{\mu}_j^L = \frac{1}{\lfloor n\tau_j \rfloor} \sum_{i=1}^{\lfloor n\tau_j \rfloor} X_{ij},\ \tilde{\mu}_j^R = \frac{1}{n - \lfloor n\tau_j \rfloor} \sum_{i=\lfloor n\tau_j \rfloor+1}^{n} X_{ij}$$

for all $1 \le j \le d$. Note that in particular for $j \notin \mathcal{B}$, we pick a dummy $\tau_j \in (0, 1)$. This is consistent with our notion of synchronization, where we assume the change-points to be synchronized if the jump is zero. For $1 \le i \le n$, $1 \le j \le d$, define $\tilde{\boldsymbol{\mu}}_i = (\tilde{\mu}_{i1}, \ldots, \tilde{\mu}_{id})$ with $\tilde{\mu}_{ij} = \tilde{\mu}_j^L + (\tilde{\mu}_j^R - \tilde{\mu}_j^L) I\{i/n > \tau_j\}$. Since $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i$ is invariant with respect to the mean left of change-point, without loss of generality we can assume $\mu_j^L = 0$ for $\delta_j > 0$, and $\mu_j = 0$ otherwise. In light of Cauchy-Schwarz inequality, it is enough to upper bound

$$\sum_{i=1}^{n} (\hat{\mu}_{ij} - \mu_{ij})^2 \tag{11.40}$$

for $1 \le j \le d$. We start with an upper bound on $\sum_{i=1}^{n} (\hat{\mu}_{ij} - \tilde{\mu}_{ij})^2$. For ease of exposition, let us introduce some more notations. For $1 \le j \le d$, let

$$\begin{aligned}
\mathcal{D}_j^{LL} &= |\hat{\mu}_j^L - \tilde{\mu}_j^L|, \\
\mathcal{D}_j^{RR} &= |\hat{\mu}_j^R - \tilde{\mu}_j^R|, \text{ and,} \\
\mathcal{D}_j^{LR} &= |\hat{\mu}_j^L - \tilde{\mu}_j^R| I\{\hat{\tau}_j > \tau_j\} + |\hat{\mu}_j^R - \tilde{\mu}_j^L| I\{\tau_j > \hat{\tau}_j\}.
\end{aligned} \tag{11.41}$$

Let us further denote $\varsigma_{jj} = |\hat{\tau}_j - \tau_j|$ for $1 \le j \le d$. Observe that

$$\begin{aligned}
\mathcal{D}_j^{LR} \le \max\Bigg\{&\Bigg|\frac{1}{n\hat{\tau}_j}\sum_{i=1}^{n\hat{\tau}_j} e_{ij} - \frac{1}{n(1-\tau_j)}\sum_{i=n\tau_j+1}^{n} e_{ij}\Bigg| I\{\hat{\tau}_j > \tau_j\}, \\
&\Bigg|\frac{1}{n-n\hat{\tau}_j}\sum_{i=n\hat{\tau}_j+1}^{n} e_{ij} - \frac{1}{n\tau_j}\sum_{i=1}^{n\tau_j} e_{ij}\Bigg| I\{\tau_j > \hat{\tau}_j\}\Bigg\} \\
&+ |\delta_j| + C|\delta_j\varsigma_{jj}| \\
= O_{\mathbb{P}}(|\delta_j| + 1/\sqrt{n}),& \qquad (11.42)
\end{aligned}$$

uniformly in j, where the $O_{\mathbb{P}}(\cdot)$ rate follows from Lemma 10.2 and Proposition 2.1. Note that in particular for $j \notin \mathcal{B}$, the tactic of choosing an arbitrary $\tau_j \in (0,1)$ is crucial; otherwise, say for $\tau_j = 0$, we would end up with

$$\Bigg|\frac{1}{n\hat{\tau}_j}\sum_{i=1}^{n\hat{\tau}_j} e_{ij} - \frac{1}{n}\sum_{i=1}^{n} e_{ij}\Bigg| \le \max_{1\le k\le n}\Bigg|\frac{1}{k}\sum_{i=1}^{k} e_{ij} - \frac{1}{n}\sum_{i=1}^{n} e_{ij}\Bigg| = O_{\mathbb{P}}(1),$$

which is worse than the $O_{\mathbb{P}}(1/\sqrt{n})$ rate we obtain in (11.42). Similar to (11.42) one can show

$$\mathcal{D}_j^{LL} \vee \mathcal{D}_j^{RR} = O_{\mathbb{P}}\Big(\frac{1}{\sqrt{n}} \wedge \frac{1}{n|\delta_j|}\Big) \text{ uniformly in } j. \qquad (11.43)$$

It can be verified by some elementary algebra that

$$\sum_{i=1}^{n}(\hat{\mu}_{ij} - \tilde{\mu}_{ij})^2 \le n\varsigma_{jj}(\mathcal{D}_j^{LR})^2 + n\tau_j(\mathcal{D}_j^{LL})^2 + n(1-\tau_j)(\mathcal{D}_j^{RR})^2, \qquad (11.44)$$

which in light of Proposition 2.1, (11.42) and (11.43), immediately yields

$$\sum_{i=1}^{n}(\hat{\mu}_{ij} - \tilde{\mu}_{ij})^2 = O_{\mathbb{P}}(1), \text{ for } 1 \le j \le d. \qquad (11.45)$$

We would like to point out that in (11.44), only $n\varsigma_{jj}(\mathcal{D}_j^{LR})^2$ contributes exactly to the $O_{\mathbb{P}}(1)$ rate, and rest of the terms are at most $O_{\mathbb{P}}((n\delta_j)^2)^{-1} \wedge 1)$. Moreover, noting that $|S_{nj}^e| = O_{\mathbb{P}}(\sqrt{n})$ for each $1 \le j \le d$, it is easy to see that $\sum_{i=1}^{n}(\tilde{\mu}_{ij} - \mu_{ij})^2 = O_{\mathbb{P}}(1)$. Therefore, preceding discussion along with (11.45) and Cauchy-Schwarz inequality completes the proof of the Lemma. □

Now we move towards proving the main result of this section.

Proof of Theorem 3.2. We first show that for $0 \le k \le B_n$

$$\rho^{\star}(\hat{\Gamma}_k - \Gamma_k) = O_{\mathbb{P}}(1/\sqrt{n}), \tag{11.46}$$

where $\Gamma_k = \mathbb{E}[\mathbf{e}_0 \mathbf{e}_k^{\top}]$. Using Gershgorin circle theorem and since d is fixed, it is enough to show that $\mathcal{R}_{n,j,l} := |(\hat{\Gamma}_k)_{j,l} - (\Gamma_k)_{j,l}|$ is small for all $1 \le j, l \le d$. Observe that, from (1.1)

$$|\mathcal{R}_{n,j,l}| \le \left|\frac{1}{n}\sum_{i=1}^{n-k} e_{ij}e_{i+k,l} - (\Gamma_k)_{j,l}\right| + \left|\frac{1}{n}\sum_{i=1}^{n-k}(\hat{\mu}_{ij} - \mu_{ij})(\hat{\mu}_{i+k,l} - \mu_{i+k,l})\right| + \\ \left|\frac{1}{n}\sum_{i=1}^{n-k}\Big(e_{ij}(\hat{\mu}_{i+k,l} - \mu_{i+k,l}) + e_{i+k,l}(\hat{\mu}_{ij} - \mu_{ij})\Big)\right|. \tag{11.47}$$

The second term in (11.47) yields a bound of $O_{\mathbb{P}}(1/n)$ from Proposition 3.1. On the other hand, for the first term in (11.47), an argument same as Lemma 1 of [114] yields that

$$\left|\frac{1}{n}\sum_{i=1}^{n-k} e_{ij}e_{i+k,l} - (\Gamma_k)_{j,l}\right| = O_{\mathbb{P}}(n^{2/p'-1} + B_n n^{-1}),$$

where $p' = p \wedge 4$. Note that $\frac{1}{n}\sum_{i=1}^{n-k} e_{ij}^2 = O_{\mathbb{P}}(1)$. Since $1/\sqrt{n} < n^{2/p'-1}$, Therefore, by Cauchy-Schwarz inequality, the third term is also $O_{\mathbb{P}}(n^{2/p'-1})$. Thus we have established (11.46). Combining this with $\sum_k |K(k/B_n)| = O(B_n)$, the consistency of our estimator can be characterized as

$$\rho^{\star}(\hat{\Sigma}_{n,B_n} - \mathbb{E}[\hat{\Sigma}_{n,B_n}]) = O_{\mathbb{P}}\Big(B_n n^{2/p'-1} + B_n^2 n^{-1}\Big) = O_{\mathbb{P}}(B_n n^{2/p'-1}). \tag{11.48}$$

The last equality in (11.48) follows from $B_n n^{2/p'-1} \to 0$. On the other hand, the bias term can be tackled as follows. Note that $\max_{1 \le j,l \le d} |\gamma_{k,j,l}| \le \sum_{s \in \mathbb{Z}} \theta_{s,p}\theta_{k+s,p}$. Hence, for a constant C_d

depending on d,

$$\sum_{k=B_n+1}^{n} \rho^\star(\Gamma_k) \le \mathsf{C}_d \sum_{s\in\mathbb{Z}} \theta_{s,p} \sum_{k=B_n+1}^{\infty} \theta_{k+s,p} \le \mathsf{C}_d \Theta_{0,p}\Theta_{B_n+1,p} = O(B_n^{-A}),$$

and

$$\sum_{k=1}^{B_n} |K(k/B_n)-1|\rho^\star(\Gamma_k) \le \mathsf{C}_d \sum_{s\in\mathbb{Z}} \theta_{s,p} \sum_{k=1}^{B_n} (k/B_n)\theta_{k+s,p} = O(1/B_n),$$

which together yield

$$\rho^\star(\mathbb{E}[\hat{\Sigma}_{n,B_n}] - \Sigma_\infty) = O(B_n^{-1}). \tag{11.49}$$

Hence, (11.48) along with (11.49) jointly implies (3.8), thereby completing the proof. $\square$

11.4 Auxiliary Results for Theorem 3.1

In the following, Lemma 11.3 enables us to represent a Gaussian vector as sums of some other i.i.d. Gaussian vectors, which in turns helps us recover finer decompositions. This result may be of independent interest. The subsequent series of propositions, leading up to Proposition 11.6, enables us to use Theorem 4 of [44] on our conditionally independent processes $Y_l^{\boldsymbol{a}}$ as well as their unconditional counterparts $S_i^\natural$. Results similar to Propositions 11.2-11.6 also appear in [62], however our results do not directly follow from those results due to the different variance decomposition we pursued in (11.16)-(11.17), which also necessitates many new technical novelties.

Lemma 11.3. *Let $A, B, C \in \mathbb{R}^{d\times d}$ be positive semi-definite matrices such that $A = B + C$. Suppose $Z \sim N(0, A)$ be given. Then there exists Z_1, Z_2 independent, such that $Z_1 \sim N(0, B)$ and $Z_2 \sim N(0, C)$.*

Proof. Let $D = A^{-1/2}BA^{-1/2}$, where we use the fact that A is symmetric and positive semi-definite to deduce uniqueness of D. Let $H = D(I - D)$. Draw $W \sim N(0, I)$ independent of Z.

Define

$$U_1 = DU + HW;\ U_2 = (I - D)U - HW,$$

and $Z_1 = A^{1/2}U_1$, $Z_2 = A^{1/2}U_2$. Note that U_1, U_2 are jointly Gaussian by virtue of being linear combinations of two Gaussian random variables. It can be verified by elementary calculation that Z_1, Z_2 satisfies the requirements. $\square$

Proposition 11.2. *Recall B_j from* (11.11). *Then it holds that*

$$\Omega(\lambda m) = \lambda_{\min}(Var(\tilde{B}_j)) \leq \rho^\star(Var(\tilde{B}_j)) = O(m\Theta_{0,2}^2). \tag{11.50}$$

Proof. Without loss of generality assume $j = 1$. Observe that $\lim_{m\to\infty} \frac{\mathrm{Var}(B_1)}{m} = \Sigma$. Therefore it holds that

$$\Omega(\lambda m) = \lambda_{\min}(\mathrm{Var}(B_1)) \leq \rho^\star(\mathrm{Var}(B_1)) \leq \|B_1\|^2 = O(m\Theta_{0,2}^2), \tag{11.51}$$

where the first and second equality follows from $\lambda_{\min}(\Sigma_\infty) = \Omega(1)$ and Burkholder's inequality respectively. Moreover, $\|S_m^\oplus - B_1\| = o(m)$ and in view of [74], $\|S_m^\oplus - \tilde{B}_1\| = O(\sqrt{m}\Theta_{m,2}) = o(\sqrt{m})$. This along with (11.51) completes the proof. $\square$

Proposition 11.3. *For a sequence $\boldsymbol{a}$, recall $Y_j^{\boldsymbol{a}}$ from* (11.13).*Let $\boldsymbol{\eta} = (\cdots, \boldsymbol{\eta}_0, \boldsymbol{\eta}_3, \cdots)$, where $\boldsymbol{\eta}_k = (\varepsilon_{(k-1)m+1}, \cdots, \varepsilon_{km})$. Then*

$$\Omega(m) = \lambda_{\min}(Var(Y_j^{\boldsymbol{\eta}})) \leq \rho^\star(Var(Y_j^{\boldsymbol{\eta}})) = O(m). \tag{11.52}$$

Proof. (11.52) follows directly from (11.51) in view of

$$|\|Y_j^{\boldsymbol{\eta}}\|^2 - \|\tilde{B}_{3j-2} + \tilde{B}_{3j-1} + \tilde{B}_{3j}\|^2| = \|M_{3j-2}(\boldsymbol{\eta})\|^2 + \|M_{3j}(\boldsymbol{\eta})\|^2 = O(m\Theta_{0,2}^2).$$

$\square$

Proposition 11.4. *Let $J = K^{2/\gamma}/\log^2 K$. Then, there exists a constant c such that*

$$\mathbb{P}\left(\boldsymbol{a} : \max_{1\le t\le K/J}\Big|Var\Big(\sum_{l=(t-1)J}^{tJ-1} Y_l^{\boldsymbol{a}}\Big) - \mathbb{E}_{\boldsymbol{a}}[Var\Big(\sum_{l=(t-1)J}^{tJ-1} Y_l^{\boldsymbol{a}}\Big)]\Big| \ge cJm\right) \to 0 \text{ as } n\to\infty. \quad (11.53)$$

Proof. Without loss of generality assume that $V_l^{\boldsymbol{a}}$ are independent for different l, otherwise we can break the sums inside the probability statement into the even and odd sums. Further we assume $d = 1$. The proof easily generalizes for the multivariate case. Therefore, in view of (11.16), it is enough to show that

$$K\max_{1\le t\le K/J}\max_{(t-1)J\le l\le tJ-1}\left[\mathbb{P}\Big(|\tilde{V}_{2l}(\boldsymbol{a}_{3l}) - \mathbb{E}(\tilde{V}_{2l}(\boldsymbol{a}_{3l}))| \ge clm\Big) + \mathbb{P}\Big(|\tilde{V}_{2l-1}(\boldsymbol{a}_{3l-3}) - \mathbb{E}(\tilde{V}_{2l-1}(\boldsymbol{a}_{3l-3}))| \ge cJm\Big)\right] \to 0. \quad (11.54)$$

Assume without loss of generality $l = 1$. Observe that ,

$$\begin{aligned}|\tilde{V}_1(\boldsymbol{a}_0) - \mathbb{E}[\tilde{V}_1(\boldsymbol{a}_0)]| \le{}& ||\|\tilde{B}_1(\boldsymbol{a}_0)\|^2 - \mathbb{E}[\|\tilde{B}_1(\boldsymbol{a}_0)\|^2]| + ||\|M_1(\boldsymbol{a}_0)\|^2 - \mathbb{E}[\|M_1(\boldsymbol{a}_0)\|^2]| \\ &+ 2|\mathbb{E}[\tilde{B}_1 C_2(\boldsymbol{\eta}_1)|a_{1-m},\cdots,a_0] - \mathbb{E}[\tilde{B}_1 C_2(\boldsymbol{\eta}_1)]|. \end{aligned} \quad (11.55)$$

For the first term in (11.55), note that $\|\tilde{B}_1(\boldsymbol{a}_0)\|^2 = \mathbb{E}[\tilde{S}_m^2|a_0,\cdots,a_{1-m}]$. Therefore, Burkholder's inequality yields

$$\mathbb{E}\left[||\|\tilde{B}_1(\boldsymbol{a}_0)\|^2 - \mathbb{E}[\|\tilde{B}_1(\boldsymbol{a}_0)\|^2]|^{\gamma/2}\right] = \|\sum_{j=-m}^{0} P_j\tilde{S}_m^2\|_{\gamma/2}^{\gamma/2} \le C_\gamma\Big(\sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2\Big)^{\gamma/4}. \quad (11.56)$$

Using $\|P_jX_i\|_\gamma \le \delta_{i-j,\gamma}$, we obtain

$$\|P_j\tilde{S}_m^2\|_{\gamma/2} = O(m^{1/2})\sum_{r=1}^{m}\tilde{\delta}_{r-j,\gamma} = O(m)n^{1/p-1/\gamma}\sum_{r=1}^{m}\delta_{r-j,p}^{p/\gamma}. \quad (11.57)$$

In view of the fact that there exists $A^\star > A_0$ such that $3 - 2(A^\star+1)p/\gamma = 0$, observe that $\Theta_{i,p} = O(i^{-A}) = O(i^{-A'}(\log i)^{-B})$ for some $A_0 < A' < \min\{A, A^\star\}$ and $B > 2\gamma/p$. The entire proof of the main theorem goes through with A' instead of A. Therefore, without loss of generality we assume $3-2(A+1)p/\gamma>0$. Putting (11.57) back in (11.56),

$$\begin{aligned}\sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2 &= O(m)n^{2/p-2/\gamma}\sum_{j=0}^{m}\left(\sum_{r=1}^{m}\delta_{r+j,\gamma}^{p/\gamma}\right)^2\\ &= O(m)n^{2/p-2/\gamma}\sum_{j=0}^{m}\sum_{l=0}^{\log_2 m}2^{2l(1-p/\gamma)}\Theta_{2^l+j,p}^{2p/\gamma}\\ &= O(m)n^{2/p-2/\gamma}m^{3-2(A+1)p/\gamma}(\log n)^{-2p/\gamma},\end{aligned}$$

which immediately yields,

$$\mathbb{E}\left[|\|\tilde{B}_1(\boldsymbol{a}_0)\|^2-\mathbb{E}[\|\tilde{B}_1(\boldsymbol{a}_0)\|^2]|^{\gamma/2}\right]=O(1)m^{\gamma-(A+1)\frac{p}{2\gamma}}n^{\frac{\gamma}{2p}-1/2}(\log n)^{-2p/\gamma}=o((Jm)^{\gamma/2}),\quad (11.58)$$

where the last equality follows from (11.5) and $\log J\asymp\log m\asymp\log n$. For the second and third terms of (11.55), note that using Cauchy-Schwarz and Jensen's inequality,

$$|\|M_1(\boldsymbol{a}_0)\|^2-\mathbb{E}[\|M_1(\boldsymbol{a}_0)\|^2]|\le\|\tilde{B}_1(\boldsymbol{a}_0)\|^2\text{ and }\|\tilde{S}_m\|^2=O(m),\quad (11.59)$$

and

$$\mathbb{E}\left|\mathbb{E}[\tilde{B}_1C_2(\boldsymbol{\eta}_1)|a_0,\cdots,a_{1-m}]\right|^{\gamma/2}=O(m^{\gamma/2})\|\mathbb{E}[\tilde{B}_1(\boldsymbol{a}_0)\|^2.\quad (11.60)$$

For $|\tilde{V}_2(\boldsymbol{a}_3)-\mathbb{E}(\tilde{V}_2(\boldsymbol{a}_3))|$, a treatment similar to (11.59) and (11.60) reveals that we only need to bound $\mathbb{P}\left(|\|\tilde{B}_3(\boldsymbol{a}_3)\|^2-\mathbb{E}[\|\tilde{B}_3(\boldsymbol{a}_3)\|^2]|>CJm\right)$. Proceeding as in (11.58)

$$\|\tilde{S}_m^2-\mathbb{E}[\tilde{S}_m^2|a_1,\cdots,a_m]\|_{\gamma/2}^{\gamma/2}=o((Jm)^{\gamma/2}).\quad (11.61)$$

Finally, using Nagaev Inequality and in view of the fact $\mathbb{E}[\tilde{S}_m^2]=O(m)$, we have

$$\mathbb{P}\left(|\tilde{S}_m^2-\mathbb{E}[\tilde{S}_m^2]|>CJm\right)\le\mathbb{P}\left(|\tilde{S}_m|>C\sqrt{Jm}\right)\le C\frac{m}{(Jm)^{\gamma/2}}\Theta_{0,p}^p+\exp(-C_1J).\quad (11.62)$$

For the second term in (11.62), clearly $K\exp(-C_1 J) \to 0$. For the first term, $K\frac{m}{(Jm)^{\gamma/2}} = \frac{\log^{\gamma} K}{m^{\gamma/2-1}} \to 0$. This completes the proof. $\square$

Proposition 11.5. *Suppose $L_{\gamma}^{\boldsymbol{a}} = \sum_{l=1}^{K} \mathbb{E}[|Y_l^{\boldsymbol{a}}|^{\gamma}]$ with $Y_l^{\boldsymbol{a}}$ defined in (11.13). Then for some constants c and C it holds that*

$$\mathbb{P}(cKm^{\gamma/2} \le L_{\gamma}^{\boldsymbol{a}} \le CKm^{\gamma/2}) \to 1. \tag{11.63}$$

Proof. Observe that $\mathbb{E}[L_{\gamma}^{\boldsymbol{a}}] = O(Km^{\gamma/2})$, and $\mathbb{E}[|\tilde{S}_m - M_1\boldsymbol{a}|^{\gamma}|\boldsymbol{a}] \le \mathbb{E}[|\tilde{S}_m|^{\gamma}|\boldsymbol{a}]$. Therefore, it is enough to show that

$$\mathbb{P}\left(|\sum_{l=1}^{K} Q_l - \mathbb{E}[Q_l]| > CK\right) \to 0, \tag{11.64}$$

where $Q_l = m^{-\gamma/2}\mathbb{E}[|\tilde{S}_{3ml} - \tilde{S}_{3m(l-1)}|^{\gamma}|\boldsymbol{a}_{3(l-1)}, \boldsymbol{a}_{3l}]$. Without loss of generality we can assume Q_l to be independent, otherwise we can consider even and odd sums separately. A treatment similar to (11.59) and (11.62) yields,

$$\mathbb{P}\left(Q_j > J^{\gamma/2}\right) = o(K^{-1}). \tag{11.65}$$

In light of (11.65), we will be done if we show

$$\mathbb{P}\left(\sum_{l=1}^{K}[T_{J^{\gamma/2}}(Q_l) - \mathbb{E}[T_{J^{\gamma/2}}(Q_l)]\right) \to 0. \tag{11.66}$$

(11.66) is immediate using Markov Inequality upon realizing that $\mathbb{E}[Q_l] = O(1)$, which implies that $\mathbb{E}[T_{J^{\gamma/2}}(Q_l)^2] = O(J^{\gamma/2})$. $\square$

Proposition 11.6. *Choose $\nu_l = lJ$ and $s \asymp K/J$ with $J = \lfloor K^{2/\gamma}/\log^2 K \rfloor$. Further let $\Gamma_k^{\boldsymbol{a}} = Var(\sum_{l=\nu_{k-1}+1}^{\nu_k} Y_l^{\boldsymbol{a}})$. Then for some constants c_1, c_2 we have, with probability going to 1,*

$$c_1 w^2 \le \lambda_{\min}(\Gamma_k^{\boldsymbol{a}}) \le \rho^{\star}(\Gamma_k^{\boldsymbol{a}}) \le c_2 w^2, \tag{11.67}$$

where $w = (L_{\gamma}^{\boldsymbol{a}})^{1/\gamma}/\log s$. Further, if $\zeta_{k,\gamma}^{\boldsymbol{a}} = \sum_{l=\nu_{k-1}+1}^{\nu_k} \mathbb{E}[|Y_l^{\boldsymbol{a}}|^{\gamma}]$, then for some $0 < \varepsilon < 1$ and *constant c_3, it holds with probability going to 1,*

$$c_3 d^{\gamma/2} s^{\varepsilon} (\log s)^{\gamma+3} \max_{1 \le k \le s} \zeta^{\boldsymbol{a}}_{k,\gamma} \le L^{\boldsymbol{a}}_{\gamma}. \tag{11.68}$$

Proof. (11.67) is immediate from Propositions 11.3, 11.4 and 11.5 along with our choice J. Moreover, exactly as in Proposition 11.5, we can show $\zeta^{\boldsymbol{a}}_k \asymp J m^{\gamma/2}$ with probability going to 1. Then (11.68) follows as $n \to \infty$. □

12 Proofs of Section 4

Here, we establish the theoretical results concerning our bootstrap procedure.

Proof of Proposition 4.1. Our proof will follow along similar lines to Proposition 4.3 of [80]. Recall that $\Sigma_\infty = \sum_{k \in \mathbb{Z}} \mathbb{E}[\mathbf{e}_0 \mathbf{e}_k^\top]$ is the long-run variance of the error process $(\mathbf{e}_i)$. Observe that, by Theorem 3.1, there exists $\mathbf{Z}_1, \ldots, \mathbf{Z}_n \overset{\text{i.i.d}}{\sim} N(0, \Sigma_\infty)$ such that

$$\begin{aligned}\mathbb{P}(U_{nj} \ge a_{\alpha - v_n, j}(\hat{\Sigma}_{n,B_n}) + c_n) \le \alpha + \mathbb{P}(a_{\alpha,j}(\Sigma_\infty) \ge a_{\alpha - v_n, j}(\hat{\Sigma}_{n,B_n}) + c_n/2) + \\ \mathbb{P}(|U_{nj} - U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)| > c_n/2).\end{aligned} \tag{12.1}$$

Let Σ_2 be a symmetric, positive definite matrix, and denote by $\mathcal{R} := \rho^\star(\Sigma_\infty - \Sigma_2)$. Then by Lemma 5.2 of [80], there exists i.i.d Gaussian variables $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$ such that $(\tilde{\mathbf{Z}}_i := \mathbf{Z}_i + \boldsymbol{\eta}_i)_{i=1}^n \overset{\text{i.i.d}}{\sim} N(0, \Sigma_2)$. Using Doob's $\mathcal{L}_p$ maximal inequality (Theorem 2.2 of [48]) for $p = 2$, along with the fact that $\rho^\star(A) \ge \max_{j,l} |A_{jl}|$ for any square matrix A, implies that, for every $1 \le j \le d$,

$$\mathbb{E}[|U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) - U_{nj}(\tilde{\mathbf{Z}}_1, \ldots, \tilde{\mathbf{Z}}_n)|^2] \le C\mathcal{R}$$

for some constant $C > 0$ possibly depending upon d. Therefore,

$$
\begin{aligned}
&\mathbb{P}(U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \geq a_{\alpha - v_n, j}(\Sigma_2) + c_n/2) \\
\leq &\alpha - v_n + \mathbb{P}(|U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) - U_{nj}(\tilde{\mathbf{Z}}_1, \ldots, \tilde{\mathbf{Z}}_n)| \geq c_n/2) \\
\leq &\alpha - (v_n - 2C\frac{\mathcal{R}}{c_n^2}) \leq \alpha,
\end{aligned} \tag{12.2}
$$

if $\mathcal{R} \leq c_n^2 v_n/(2C)$. Therefore, if $\mathcal{R} \leq c_n^2 v_n/(2C)$, then $a_{\alpha,j}(\Sigma_\infty) \leq a_{\alpha - v_n, j}(\Sigma_2) + c_n/2$. Now we replace Σ_2 by $\hat{\Sigma}_{n,B_n}$. This being a random quantity, the corresponding random error be denoted as $\hat{\mathcal{R}}$. Thus, in view of (12.2), from (12.1) we have

$$
\mathbb{P}(a_{\alpha,j}(\Sigma_\infty) \geq a_{\alpha - v_n, j}(\hat{\Sigma}_{n,B_n}) + c_n/2) \leq \mathbb{P}(\hat{\mathcal{R}} > c_n^2 v_n/(2C)).
$$

Observe that by our choice of c_n and v_n, from Theorem 3.2 we have

$$
\overline{\lim_{n \to \infty}}\, \mathbb{P}(\hat{\mathcal{R}} > c_n^2 v_n/(2C)) + \mathbb{P}(|U_{nj} - U_{nj}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)| > c_n/2) = 0.
$$

Hence in light of (12.1), the proof of (4.2) is complete. $\square$

Proof of Theorem 4.1. In the following, C will denote a constant depending on d, τ and $\Theta_{0,p}$ whose value may change from line-to-line. Sometimes we also use $\gtrsim$ or $\lesssim$ to hide these constant. We will follow the proof strategy of Proposition 4.1, with some additional but significant modification necessitated by the form of our test statistic T_n. By Theorem 3.1, there exists $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n \overset{\text{i.i.d}}{\sim} N(0, \Sigma_\infty)$ such that $|S_i^e - S_i^\eta| = o_{\mathbb{P}}(n^{1/p})$. Write

$$
\begin{aligned}
&\mathbb{P}(T_n \geq b_{\alpha - h_n}(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) + u_n) \\
\leq &\alpha + \mathbb{P}(b_\alpha(\boldsymbol{\mu}, \Sigma_\infty) \geq b_{\alpha - h_n}(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) + u_n/2) + \mathbb{P}(|T_n - T_n(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)| > u_n/2)
\end{aligned} \tag{12.3}
$$

where $\mathbf{Z}_i = \boldsymbol{\eta}_i + \boldsymbol{\mu}_i$. For the third term in (12.3), Lemma 3.1 along with our choice of u_n implies that $\overline{\lim}_{n\to\infty}\mathbb{P}(|T_n - T_n(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)| > u_n/2) = 0$. Thus we focus on bounding $\mathbb{P}(b_\alpha(\boldsymbol{\mu},\Sigma_\infty) \geq b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}},\hat{\Sigma}_{n,B_n}) + u_n/2)$.

Hereafter, we write $T_n(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)$ as $T_n^{\boldsymbol{\mu},\Sigma_\infty}$. Let $\Sigma^\dagger$ be a symmetric positive definite matrix, with $\mathcal{R} := \rho^\star(\Sigma_\infty - \Sigma^\dagger)$. Moreover, let $\tau^\dagger \in (0,1)$, $(\gamma_j^\dagger)_{j=1}^d \in \mathbb{R}$, $(\nu_j^L)_{j=1}^d \in \mathbb{R}$ and $(\nu_j^R)_{j=1}^d \in \mathbb{R}$ be given, with these quantities satisfying certain conditions mentioned subsequently in appropriate places. Consider a sequence of possibly random vectors $\boldsymbol{\mu}_1^\dagger,\ldots,\boldsymbol{\mu}_n^\dagger \in \mathbb{R}^d$ such that

$$\mu_{ij}^\dagger = \begin{cases} \gamma_j, & \text{if } j \in \mathcal{V}_0, \\ \nu_{ij}^\dagger, & \text{if } j \in \mathcal{V}_1. \end{cases}$$

Here

$$\nu_{ij}^\dagger = \begin{cases} \upsilon_j^L, \text{ if } i \leq n\tau^\dagger \\ \upsilon_j^R, \text{ if } i > n\tau^\dagger. \end{cases}$$

Denote $\mathcal{G}_j^{LL} = \upsilon_j^L - \mu_j^L$, $\mathcal{G}_j^{RR} = \upsilon_j^R - \mu_j^R$ and $\mathcal{G}_j^{LR} = (\upsilon_j^L - \mu_j^R)I\{\tau^\dagger > \tau\} + (\upsilon_j^R - \mu_j^L)I\{\tau > \tau^\dagger\}$. Further, let $\psi := |\tau^\dagger - \tau|$. We point out that above definitions are motivated from the definition of $\mathcal{D}_j$'s from (11.41). Indeed, we will pursue an argument conditional on $\mathcal{G}_j^{LL}$, $\mathcal{G}_j^{LR}$, $\mathcal{G}_j^{RR}$ and $\tau_j^\dagger$. Note that, by Lemma 5.2 of [80], there exists independent Gaussian random variable $W_1,\ldots,W_n$ such that $\mathbf{\Lambda}_i := \mathbf{Z}_i + W_i \sim N(\boldsymbol{\mu}_i,\Sigma^\dagger)$. Let $T_n^{\boldsymbol{\mu},\Sigma^\dagger} = T_n(\mathbf{\Lambda}_1,\ldots,\mathbf{\Lambda}_n)$. Further denote $T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger} = T_n(\mathbf{Z}_1^\dagger,\ldots,\mathbf{Z}_n^\dagger)$, where $\mathbf{Z}_i^\dagger = \mathbf{\Lambda}_i + \boldsymbol{\mu}_i^\dagger - \boldsymbol{\mu}_i$. Note that

$$\begin{aligned} &\mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \geq b_{\alpha-h_n}(\boldsymbol{\mu}^\dagger,\Sigma^\dagger) + u_n/2) \\ \leq &\alpha - h_n + \mathbb{P}(|T_n^{\boldsymbol{\mu},\Sigma_\infty} - T_n^{\boldsymbol{\mu},\Sigma^\dagger}| > u_n/4) + \mathbb{P}(|T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger} - T_n^{\boldsymbol{\mu},\Sigma^\dagger}| > u_n/4). \end{aligned} \tag{12.4}$$

Similar to Lemma 3.1 and (12.2), $\mathbb{P}(|T_n^{\boldsymbol{\mu},\Sigma_\infty} - T_n^{\boldsymbol{\mu},\Sigma^\dagger}| > u_n/4) \leq C\mathcal{R}/u_n^2$. To tackle $|T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger} - T_n^{\boldsymbol{\mu},\Sigma^\dagger}|$, we introduce some notation. Let $V_{i,j}^\dagger = \sum_{k=1}^i (Z_{kj}^\dagger - \bar{Z}_{\cdot j}^\dagger)$, $\hat{\tau}_j^\dagger = (\arg\max_{1\leq i\leq n}|V_{i,j}^\dagger|)/n$, and $\hat{\tau}^\dagger = (\arg\max_{1\leq i\leq n}\sum_{j=1}^d |V_{i,j}^\dagger|)/n$. Likewise, define $V_{i,j}^\Lambda$, $\hat{\tau}_j^\Lambda$ and $\hat{\tau}^\Lambda$ with $\mathbf{\Lambda}_1,\ldots,\mathbf{\Lambda}_n$. Now, simplify

$|T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger} - T_n^{\boldsymbol{\mu},\Sigma^\dagger}|$ as

$$\sum_{j=1}^{d} \Big| \Big(|V^\dagger_{n\hat{\tau}^\dagger_j,j}| - |V^\Lambda_{n\hat{\tau}^\Lambda_j,j}|\Big) - \Big(|V^\dagger_{n\hat{\tau}^\dagger,j}| - |V^\Lambda_{n\hat{\tau}^\Lambda,j}|\Big) \Big| / \sqrt{n}. \tag{12.5}$$

For the first term in (12.5), note that

$$\begin{aligned} \Big|V^\dagger_{n\hat{\tau}^\dagger_j,j} - V^\Lambda_{n\hat{\tau}^\Lambda_j,j}\Big| &\leq \max_{1\leq i\leq n} \Big| \sum_{k=1}^{i} (Z^\dagger_{kj} - \bar{Z}^\dagger_{\cdot j} - \Lambda_{kj} + \bar{\Lambda}_{\cdot j}) \Big| \\ &= \max_{1\leq i\leq n} \Big| \sum_{k=1}^{i} \mu^\dagger_{kj} - \frac{i}{n}\sum_{k=1}^{n} \mu^\dagger_{kj} - \sum_{k=1}^{i} \mu_{kj} + \frac{i}{n}\sum_{k=1}^{n} \mu_{kj} \Big| \end{aligned} \tag{12.6}$$

For $j \in \mathcal{V}_0$, (12.6) immediately vanishes to 0. For $j \in \mathcal{V}_1$, we pursue a more careful approximation. Let $\kappa_j = v^R_j - v^L_j$, and assume that $|\kappa_j| >> n^{-1/2}$ for $j \in \mathcal{V}_1$. Moreover, assume that $\mathcal{G}^{LL}_j = O(n^{-1/2})$, and $\psi \leq \frac{1}{n\delta_1^2 \vee n\kappa_1^2}$. Subsequently, these conditions will be called as $\mathcal{A}$, and $\mathbf{1}_{\mathcal{A}} = 1$ iff $\mathcal{A}$ is satisfied, and 0 otherwise. The motivation behind these assumptions will become clear once we move towards the un-conditioning step of our argument. Similar to the proof of Proposition 2.1, we can be excused for assuming $\delta_j \geq 0$ for each $j \in \mathcal{V}_1$. Moreover, since $\hat{\tau}_j, \hat{\tau} \in (c, 1-c)$, hence,

$$\begin{aligned} &\mathbb{P}(\mathcal{D}) = 1, \text{ for } c \leq \tau \leq 1-c,\, j \in \mathcal{V}_1 \text{ where } c > 0 \text{ is some constant,} \\ &\text{with } \mathcal{D} := \{\arg\max_{1\leq i\leq n} |V^\Lambda_{ij}| = \arg\max_{nc\leq i\leq n(1-c)} |V^\Lambda_{ij}|, \arg\max_{1\leq i\leq n} |V^\dagger_{ij}| = \arg\max_{nc\leq i\leq n(1-c)} |V^\dagger_{ij}|\}. \end{aligned} \tag{12.7}$$

Rest of our arguments will proceed conditional on $\mathcal{D}$. In the next step, we will show that the event $\mathcal{E} = \{\hat{\tau}^\dagger_j = \hat{\tau}^\Lambda_j, j \in \mathcal{V}_1\}$ occurs with very high probability. To see this, note that, since $\delta_j \geq 0$ and $j \in \mathcal{V}_1$, in light of $\hat{\tau}^\lambda_j \in (c, 1-c)$ from (12.7), as well as the decomposition of V_{ij}'s from (10.4),

$$\begin{aligned} \mathbb{P}(V^\Lambda_{n\hat{\tau}^\Lambda_j} < 0\,|\mathcal{D}) &= \mathbb{P}(V^G_{n\hat{\tau}_j^\Lambda} < n\hat{\tau}^\lambda_j(1-\tau_j)\delta_j I\{\hat{\tau}^\lambda_j < \tau_j\} + n\tau_j(1-\hat{\tau}^\lambda_j)\delta_j I\{\hat{\tau}^\lambda_j < \tau_j\}\,|\mathcal{D}) \\ &\to 1, \text{ as } n \to \infty \end{aligned} \tag{12.8}$$

where $G \sim N(0, \Sigma^\dagger)$; the limiting assertion follows from noting $\max_i |V^{\Lambda}_{i,j}| = O_{\mathbb{P}}(\sqrt{n})$, and $j \in \mathcal{V}_1$ along with event $\mathcal{D}$ implies that $n(\hat{\tau}^{\lambda}_j \wedge 1 - \hat{\tau}^{\lambda}_j)\delta_j >> \sqrt{n}$. Call the event $V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j} < 0$ in (12.8) as $\mathcal{F}$. Employing the fact that $|\tau^\dagger - \tau| \le \frac{1}{n\delta_1^2 \vee n\kappa_1^2}$ as well as Proposition 2.1, one further obtains $\mathbb{P}(\mathcal{H}|\mathcal{D}) \to 1$, where $\mathcal{H} := \{V^{\Lambda}_{n\hat{\tau}^{\dagger}_j} < 0\}$. Let $h_{ij} = \mu^\dagger_{ij} - \mu_{ij}$, and let $\lambda_{i,j} = \sum_{k=1}^{i}(h_{kj} - \bar{h}_{\cdot j})$, where $\bar{h}_{\cdot j} = n^{-1}\sum_{k=1}^{n} h_{kj}$. Moreover, denote $\tau^L = \tau^\dagger \wedge \tau$, and $\tau^R = \tau^\dagger \vee \tau$. The following series of implications, conditional on $\mathcal{D}, \mathcal{F}$ and $\mathcal{H}$, follow through from the definition of the event $\mathcal{E}$.

$$
\begin{aligned}
\mathcal{E}^c &\iff \underset{nc \le i \le n(1-c)}{\arg\max} |V^\dagger_{ij}| \neq \underset{nc \le i \le n(1-c)}{\arg\max} |V^\Lambda_{ij}| \\
&\iff \{|V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j}| > |V^{\Lambda}_{n\hat{\tau}^{\dagger}_j}|,\ |V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j} + \lambda_{n\hat{\tau}^{\Lambda}_j, j}| < |V^{\Lambda}_{n\hat{\tau}^{\dagger}_j} + \lambda_{n\hat{\tau}^{\dagger}_j, j}|\}. \\
&\implies |\lambda_{n\hat{\tau}^{\Lambda}_j, j} - \lambda_{n\hat{\tau}^{\dagger}_j, j}| > |V^{\Lambda}_{n\hat{\tau}^{\dagger}_j} - V^{\Lambda}_{n\hat{\tau}^{\dagger}_j}| \\
&\implies n|\hat{\tau}^{\dagger}_j - \hat{\tau}^{\Lambda}_j|(\tau^L|\mathcal{G}^{LL}_j| + \psi|\mathcal{G}^{LR}_j| + \tau^R|\mathcal{G}^{RR}_j|) \gtrsim |V^{\Lambda}_{n\hat{\tau}^{\dagger}_j} - V^{\Lambda}_{n\hat{\tau}^{\dagger}_j}|. \qquad (12.9)
\end{aligned}
$$

Now, for the right hand side of (12.9), observe that noting $|\delta_j| >> n^{-1/2}$ along with (10.4) and (10.19) implies that

$$\mathbb{P}(|V^{\Lambda}_{n\hat{\tau}^{\dagger}_j} - V^{\Lambda}_{n\hat{\tau}^{\dagger}_j}| \ge n|\hat{\tau}^{\dagger}_j - \hat{\tau}^{\Lambda}_j|\delta_j) \to 1.$$

Therefore, from (12.9), and our assumption $\mathcal{G}^{LL}_j = O(n^{-1/2})$, and $\psi \le \frac{1}{n\delta_1^2 \vee n\kappa_1^2}$, one has $\mathbb{P}(\mathcal{E}|\mathcal{D}, \mathcal{F}, \mathcal{H}) \to 1$ as $n \to \infty$. Using $\mathbb{P}(\mathcal{F}|\mathcal{D}) \to 1$ and $\mathbb{P}(\mathcal{H}|\mathcal{D}) \to 1$ as shown above, we finally arrive at

$$\mathbb{P}(\mathcal{E}|\mathcal{D}) \to 1, \text{ as } n \to \infty. \qquad (12.10)$$

Now we are ready for the coup de grâce for the case $j \in \mathcal{V}_1$. Clearly, under $\mathcal{F}$, and keeping in mind (10.4), it holds that $\mathbb{P}(V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j} < \lambda_{n\hat{\tau}^{\Lambda}_j, j}) \to 1$, and consequently, a further conditioning by $\mathcal{E}$ and $\mathcal{F}$ implies

$$|V^{\dagger}_{n\hat{\tau}^{\dagger}_j, j}| - |V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j, j}| = |V^{\Lambda}_{n\hat{\tau}^{\dagger}_j, j} + \lambda_{n\hat{\tau}^{\dagger}_j, j}| - |V^{\Lambda}_{n\hat{\tau}^{\Lambda}_j, j}| = -\lambda_{n\hat{\tau}^{\dagger}_j, j}. \qquad (12.11)$$

Here we have made use of the fact that $|a+b|-|a|=-b$ if $a<-b$. An analogous argument can also be carried out for $\hat{\tau}^\dagger$. Therefore, conditional on $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{F}$, for $j \in \mathcal{V}_1$,

$$\begin{aligned}\left(|V^\dagger_{n\hat{\tau}^\dagger_j,j}|-|V^\Lambda_{n\hat{\tau}^\Lambda_j,j}|\right)-\left(|V^\dagger_{n\hat{\tau}^\dagger,j}|-|V^\Lambda_{n\hat{\tau}^\Lambda,j}|\right) &= \lambda_{n\hat{\tau}^\dagger,j}-\lambda_{n\hat{\tau}^\dagger_j,j} \\ &\lesssim \sqrt{n}|\hat{\tau}^\dagger_j-\hat{\tau}^\dagger|. \end{aligned} \tag{12.12}$$

Combining the analysis for $\mathcal{V}_0$, (12.6), and (12.12), one has, conditional on $\mathcal{D},\mathcal{E}$, and $\mathcal{F}$, that

$$|T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger}-T_n^{\boldsymbol{\mu},\Sigma^\dagger}| \lesssim \sum_{j\in\mathcal{V}_1}|\hat{\tau}^\dagger_j-\hat{\tau}^\dagger|,$$

which implies that

$$\mathbb{P}(|T_n^{\boldsymbol{\mu}^\dagger,\Sigma^\dagger}-T_n^{\boldsymbol{\mu},\Sigma^\dagger}|>u_n/4) \le \mathbb{P}(\sum_{j\in\mathcal{V}_1}|\hat{\tau}^\dagger_j-\hat{\tau}^\dagger| \gtrsim u_n)+\mathbb{P}(\mathcal{E}^c\cup\mathcal{F}^c) \lesssim u_n^{-1}\max_{j\in\mathcal{V}_1}(n\kappa_j^2)^{-1}+\max_{j\in\mathcal{V}_1}(\sqrt{n}\kappa_j)^{-1}. \tag{12.13}$$

Hence, from (12.13) and (12.13), we arrive at $\mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \ge b_{\alpha-h_n}(\boldsymbol{\mu}^\dagger,\Sigma^\dagger)+u_n/2) \le \alpha$ if

$$h_n - C\mathcal{R}u_n^{-2}-Cu_n^{-1}\max_{j\in\mathcal{V}_1}(n\kappa_j^2)^{-1}-\max_{j\in\mathcal{V}_1}(\sqrt{n}\kappa_j)^{-1} \ge 0. \tag{12.14}$$

Clearly, $\mathbb{P}(T_n \ge b_{\alpha-h_n}(\boldsymbol{\mu}^\dagger,\Sigma^\dagger)+u_n/2)\le\alpha$ implies that $b_\alpha(\boldsymbol{\mu},\Sigma_\infty)\le b_{\alpha-h_n}(\boldsymbol{\mu}^\dagger,\Sigma^\dagger)+u_n/2$. We now apply the implication (12.14) with $\mu^\dagger_{ij}=\tilde{\mu}_{ij}$, $\tau^\dagger=\hat{\tau}$ and $\Sigma^\dagger=\hat{\Sigma}_{n,B_n}$. The corresponding $\mathcal{R}$, ψ, and $\mathcal{G}_j$'s are denoted by $\hat{\mathcal{R}}$, $\hat{\psi}$, and $\hat{\mathcal{G}}_j$ respectively. All of these are random variables. In particular, an argument similar to (11.42) and (11.43) shows that, under H_0 in (1.3),

$$\max_{j\in\mathcal{V}_1}\hat{\mathcal{G}}^{LL}_{j1}\vee\hat{\mathcal{G}}^{RR}_{j1}=O_{\mathbb{P}}(\frac{1}{\sqrt{n}}), \text{ and } \max_{j\in\mathcal{V}_1}\hat{\mathcal{G}}^{LR}_{j1}/|\delta_j|=O_{\mathbb{P}}(1).$$

Therefore, all these random variables satisfy the conditions we put on the sequences $\boldsymbol{\mu}^\dagger_i$ with probability tending to 1. Moreover, Theorem 3.2 instructs $\hat{\mathcal{R}}=O_{\mathbb{P}}(B_n n^{2/p'-1}+B_n^{-1})$, which implies, in light of (12.14),

$$\begin{aligned}
&\varlimsup_{n\to\infty} \mathbb{P}(b_\alpha(\boldsymbol{\mu}, \Sigma_\infty) \geq b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) + u_n/2) \\
\leq &\varlimsup_{n\to\infty} \mathbb{P}\Big(\hat{\mathcal{R}} \gtrsim u_n^2 h_n\Big) + \varlimsup_{n\to\infty} \mathbb{P}\Big(\max_{j\in\mathcal{V}_1}(n\hat{\delta}_j^2)^{-1} \gtrsim u_n h_n)\Big) + \varlimsup_{n\to\infty} \mathbb{P}\Big(\max_{j\in\mathcal{V}_1}(\sqrt{n}\hat{\delta}_j)^{-1} \gtrsim h_n)\Big)
\end{aligned} \tag{12.15}$$

where, the final equality follows from our choice of u_n and h_n. This completes the proof of our theorem. $\square$

Proof of Theorem 4.2. For a generic sequence of independent Gaussian random vectors $G_1, \ldots, G_n$ with $G_i = \boldsymbol{\nu}_i + Z_i$ with $Z_i \overset{i.i.d.}{\sim} N(0, \Gamma)$ for a sequence of vectors $(\boldsymbol{\nu}_i)_{i=1}^n \in \mathbb{R}^d$ and a covariance matrix $\Gamma \in \mathbb{R}^{d\times d}$, we denote $T_n(G_1, \ldots, G_n)$ as in (2.3), by $T_n^{\boldsymbol{\nu},\Gamma}$. Consider u_n and h_n same as in Theorem 4.1. Note that

$$\begin{aligned}
&\mathbb{P}(T_n \leq b_\alpha(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n})) \\
\leq &\mathbb{P}(|T_n - T_n^{\boldsymbol{\mu},\Sigma_\infty}| > u_n) + \mathbb{P}(b_\alpha(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) - b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) > u_n) + \mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \leq b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) + 2u_n),
\end{aligned} \tag{12.16}$$

and, invoking Lemma 3.1 and our choice of u_n, $\mathbb{P}(|T_n - T_n^{\boldsymbol{\mu},\Sigma_\infty}| > u_n) \to 0$ as $n \to \infty$. Next, consider two generic covariance matrices Σ_1 and Σ_2. Let $\boldsymbol{\Lambda}_i \overset{ind}{\sim} N(\boldsymbol{\nu}_i, \Sigma_1)$ for $i \in [n]$. By Lemma 5.2 of [80], there exist i.i.d. mean-zero Gaussian random vectors $\boldsymbol{\eta}_i, \ldots, \boldsymbol{\eta}_n$ such that $\boldsymbol{Q}_i := \boldsymbol{\Lambda}_i + \boldsymbol{\eta}_i \overset{ind}{\sim} N(\boldsymbol{\nu}_i, \Sigma_2)$. The corresponding T_n's for $\boldsymbol{\Lambda}$'s and $\boldsymbol{Q}$'s are denoted as $T_n^{\boldsymbol{\nu},\Sigma_1}$ and $T_n^{\boldsymbol{\nu},\Sigma_2}$ respectively. The corresponding mean zero errors associated with $T_n^{\boldsymbol{\nu},\Sigma}$ be denoted by Z. The notations $V_{i,j}^Z$ and $f_{n,j}$ are same as in Section 10. Further, let $\mathcal{R} = \rho^\star(\Sigma_1 - \Sigma_2)$. Write

$$\mathbb{P}(T_n^{\boldsymbol{\nu},\Sigma_1} \geq b_{\alpha-h_n}(\boldsymbol{\nu}, \Sigma_2) + u_n) \leq \alpha - h_n + \mathbb{P}(|T_n^{\boldsymbol{\nu},\Sigma_1} - T_n^{\boldsymbol{\nu},\Sigma_2}| > u_n). \tag{12.17}$$

Additionally, for each $1 \leq j \leq d$,

$$\max_{1\leq i\leq n} |S_{ij}^{\boldsymbol{\Lambda}} - i\bar{\boldsymbol{\Lambda}}_{\cdot j}| - \max_{1\leq i\leq n} |S_{ij}^{\boldsymbol{Q}} - i\bar{\boldsymbol{Q}}_{\cdot j}| \leq \max_{1\leq i\leq n} |S_{ij}^{\boldsymbol{\eta}} - i\bar{\boldsymbol{\eta}}_{\cdot j}|,$$

and therefore, similar to the arguments preceding (12.2), $\mathbb{E}[|T_n^{\boldsymbol{\nu},\Sigma_1} - T_n^{\boldsymbol{\nu},\Sigma_2}|^2] \leq C\mathcal{R}$. Hence, from (12.17), if $\mathcal{R} \leq C^{-1}u_n^2 h_n$, then $\mathbb{P}(T_n^{\boldsymbol{\nu},\Sigma_1} \geq b_{\alpha-h_n}(\boldsymbol{\nu}, \Sigma_2) + u_n) \leq \alpha$, which immediately implies that

$$b_\alpha(\boldsymbol{\nu}, \Sigma_1) \leq b_{\alpha-h_n}(\boldsymbol{\nu}, \Sigma_2) + u_n.$$

Consequently,

$$\mathbb{P}(b_\alpha(\tilde{\boldsymbol{\mu}}, \hat{\Sigma}_{n,B_n}) - b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) > u_n) \leq \mathbb{P}(\rho^\star(\hat{\Sigma}_{n,B_n} - \Sigma_\infty) > C^{-1}u_n^2 h_n)) \to 0 \text{ as } n \to \infty, \tag{12.18}$$

where the limiting assertion follows from Theorem 3.2 and the choice of u_n and h_n. Finally, we deal with the third term in (12.16). Since (2.9) is satisfied, let $\{j_1, j_2\} = \arg\max_{\{k_1,k_2\}\in\mathcal{H}}(|\delta_{k_1}| \wedge |\delta_{k_2}|)$, and $\delta_\star = |\delta_{j_1}| \wedge |\delta_{j_2}|$. For a constant $C_0 = \mathsf{C}/2$ where C is as in (10.20), write

$$\mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \leq b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) + 2u_n) \leq \mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \leq C_0\sqrt{n}\delta_\star) + \mathbb{P}(b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) > C_0\sqrt{n}\delta_\star - 2u_n). \tag{12.19}$$

We analyze (12.19) term-by-term. Let the individual and common change-point estimates in $T_n^{\boldsymbol{\mu},\Sigma_\infty}$ be denoted by $\tilde{\tau}_j$, $j \in [d]$, and $\tilde{\tau}$ respectively. Note that

$$\begin{aligned}
\mathbb{P}(T_n^{\boldsymbol{\mu},\Sigma_\infty} \leq C_0\sqrt{n}\delta_\star) &\leq \mathbb{P}\Bigg(\sum_{j\in\{j_1,j_2\}} (|V_{n\tilde{\tau}_j,j}^Z + f_{n,j}(n\tilde{\tau}_j)| - |V_{n\tilde{\tau},j}^Z + f_{n,j}(n\tilde{\tau})|) \leq C_0 n\delta_\star\Bigg) \\
&\leq \mathbb{P}\Bigg(\sum_{j\in\{j_1,j_2\}} (|f_{n,j}(n\tilde{\tau}_j)| - |f_{n,j}(n\tilde{\tau})|) \leq C_0 n\delta_\star + \sum_{j\in\{j_1,j_2\}} (|V_{n\tilde{\tau}_j,j}^Z| + |V_{n\tilde{\tau},j}^Z|)\Bigg) \\
&\leq \mathbb{P}((\mathsf{C} - C_0)n\delta_\star \leq 2\sum_{j\in\{j_1,j_2\}} \max_{1\leq i\leq n} |V_{i,j}^Z| - \sum_{j\in\{j_1,j_2\}} (|f_{n,j}(n\tilde{\tau}_j)| - |f_{n,j}(n\tau_j)|))
\end{aligned}$$

$$\to 0, \tag{12.20}$$

where the third inequality follows from (10.20), and the limiting assertion follows from (10.16) and (10.21) in lieu of (2.9). On the other hand, for the second term in (12.19) we pursue yet another conditional argument. For a generic $\boldsymbol{\nu}$ with a common change-point at $\lfloor n\tau^{\boldsymbol{\nu}} \rfloor$ (in the sample level) with $\tau^{\boldsymbol{\nu}} \in (c, 1-c)$, and jumps at $\delta_j^{\boldsymbol{\nu}}$, consider $T_n^{\boldsymbol{\nu},\Sigma_\infty}$ and the corresponding individual and common change-point estimates as $\hat{\tau}_j^{\boldsymbol{\nu}}$, $j \in [d]$ and $\hat{\tau}^{\boldsymbol{\nu}}$, respectively. We clarify that in this particular case we let $\tau^{\boldsymbol{\nu}}$ vary with n. Consider bounding the following probability:

$$\begin{aligned}
&\mathbb{P}(T_n^{\boldsymbol{\nu},\Sigma_\infty} \geq C_0\sqrt{n}\delta_\star - 2u_n) \\
&\leq \sum_{j=1}^{d} \mathbb{P}(|V^Z_{n\hat{\tau}_j^{\boldsymbol{\nu}},j} + f_{n,j}(n\hat{\tau}_j^{\boldsymbol{\nu}})| - |V^Z_{n\hat{\tau}^{\boldsymbol{\nu}},j} + f_{n,j}(n\hat{\tau}^{\boldsymbol{\nu}})| \geq d^{-1}(C_0 n\delta_\star - 2\sqrt{n}u_n)) \\
&\leq \sum_{j=1}^{d} \mathbb{P}(|f_{n,j}(n\hat{\tau}_j^{\boldsymbol{\nu}})| - |f_{n,j}(n\hat{\tau}^{\boldsymbol{\nu}})| \geq d^{-1}(C_0 n\delta_\star - 2u_n\sqrt{n}) - |V^Z_{n\hat{\tau}_j^{\boldsymbol{\nu}},j}| - |V^Z_{n\hat{\tau}^{\boldsymbol{\nu}},j}|) \\
&\leq \sum_{j=1}^{d} \Big(\mathbb{P}(C_{\tau^{\boldsymbol{\nu}}} n|\delta_j^{\boldsymbol{\nu}}||\hat{\tau}_j^{\boldsymbol{\nu}} - \hat{\tau}^{\boldsymbol{\nu}}| \geq (2d)^{-1}(C_0 n\delta_\star - 2u_n\sqrt{n})) + \mathbb{P}(2\max_{1\leq i\leq n}|V^Z_{i,j}| \geq (2d)^{-1}(C_0 n\delta_\star - 2u_n\sqrt{n})) \Big).
\end{aligned} \tag{12.21}$$

Due to (10.16), the second term in (12.21) yields

$$\sum_{j=1}^{d} \mathbb{P}(2\max_{1\leq i\leq n}|V^Z_{i,j}| \geq (2d)^{-1}(C_0 n\delta_\star - 2u_n\sqrt{n})) \leq C_d(C_0\sqrt{n}\delta_\star - 2u_n)^{-1}.$$

For the first term we employ the proof of Proposition 2.1 along with $\tau^{\boldsymbol{\nu}} \in (c, 1-c)$ to deduce

$$\mathbb{P}(C_\tau n|\delta_j^{\boldsymbol{\nu}}||\hat{\tau}_j^{\boldsymbol{\nu}} - \hat{\tau}^{\boldsymbol{\nu}}| \geq (2d)^{-1}(C_0 n\delta_\star - 2u_n\sqrt{n})) \leq C_d(n\delta_\star|\delta_j^{\boldsymbol{\nu}}| - 2u_n|\delta_j^{\boldsymbol{\nu}}|\sqrt{n})^{-1} I\{C_0 n\delta_\star - 2u_n\sqrt{n} < n|\delta_j^{\boldsymbol{\nu}}|\}. \tag{12.22}$$

Recall $\tilde{\mu}$ from Theorem 4.1. Let $\tilde{\delta}_j = \tilde{\mu}_{n\hat{\tau}+1,j} - \tilde{\mu}_{n\hat{\tau}+1,j}$ for $j \in [d]$. Here, $\tilde{\delta}_j = 0$ for $j \in \mathcal{V}_0$. From

(12.21) and (12.22), one obtains in (12.19) that,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\, \mathbb{P}(b_{\alpha-h_n}(\tilde{\boldsymbol{\mu}}, \Sigma_\infty) > C_0\sqrt{n}\delta_\star - 2u_n) \\
&\leq \overline{\lim_{n\to\infty}}\, \mathbb{P}(C_d \sum_{j=1}^{d} (n\delta_\star|\tilde{\delta}_j| - 2u_n|\tilde{\delta}_j|\sqrt{n})^{-1} I\{|\tilde{\delta}_j| > C_0\delta_\star - 2u_n n^{-1/2}\} + C_d(C_0\sqrt{n}\delta_\star - 2u_n)^{-1} > \alpha - h_n) \\
&\leq \overline{\lim_{n\to\infty}}\, \mathbb{P}(C_d \sum_{j:|\tilde{\delta}_j| > C_0\delta_\star - 2u_n n^{-1/2}} (n\delta_\star|\tilde{\delta}_j| - 2u_n|\tilde{\delta}_j|\sqrt{n})^{-1} > \alpha/2 - h_n/2) \\
&\leq \overline{\lim_{n\to\infty}}\, \mathbb{P}((C_0\delta_\star - 2u_n n^{-1/2})^{-1}(n\delta_\star - 2u_n\sqrt{n})^{-1} > (\alpha/2 - h_n/2)(C_d d)^{-1}) = 0 \qquad (12.23)
\end{aligned}$$

where in the second inequality we use that, $C_d(C_0\sqrt{n}\delta_\star - 2u_n) \to \infty$, and in the third one we invoke $(C_0\delta_\star - 2u_n n^{-1/2})(n\delta_\star - 2u_n\sqrt{n}) \to \infty$. Both these assertions are direct implications of (2.9) and our choice of u_n and h_n. Equation (12.23) completes the proof in light of (12.16) and (12.19). □

13 Proofs of Section 5

Proof of Proposition 5.1. The proof follow essentially along the lines of the proof of Proposition 2.1, but with necessary modifications designed to tackle the intricacy of increasing dimension. Again, without loss of generality, assume that $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_d \geq 0$. For any non-negative ordered sequence $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_d$, it is not difficult to realize that the sequence must satisfy at least one of the following two properties:

- $\mathcal{P}_1$: There exists $1 \leq j \leq d$ such that $\delta_j \geq \frac{\sum_{k=1}^{j}\delta_k}{d} \geq \delta_{j+1}$. Call j_0 to be the largest such j.
- $\mathcal{P}_2$: There exists j such that $\delta_j \geq \frac{\sum_{i=1}^{j-1}\delta_i}{d}$. Call j_1 to be the largest such j.

In other words, the sequence of jumps $(\delta_1, \delta_2, \ldots, \delta_d)$ satisfies $\mathcal{P}_1$ if $\delta_2 \leq d^{-1}\delta_1$; otherwise it satisfies $\mathcal{P}_2$. Note that the two properties are not mutually exclusive, and there will be sequences of jumps which satisfy both the properties. For our proof, we need either of these two property to hold to get a correct rate.

Rest of the notations of this proof are same as in the proof of Proposition 2.1. Suppose the sequence of jumps satisfies $\mathcal{P}_1$. Then we will take $j^\star = j_0$, and echo the analysis in Proposition 2.1. Note that, by definition of $\mathcal{P}_1$,

$$\sum_{j=1}^{j_0} \delta_j \leq \sum_{j=1}^{d} \delta_j \leq \sum_{j=1}^{j_0} \delta_j + (1 - j_0 d^{-1}) \sum_{j=1}^{j_0} \delta_j \leq 2 \sum_{j=1}^{j_0} \delta_j. \tag{13.1}$$

The first point of difference from the afore-mentioned proof occurs after (10.11), where we obtain

$$\mathcal{X}_1 \implies 2 \max_{1 \leq i \leq n} \sum_{j=1}^{d} |V_{ij}|^e \geq d^{-1} n \tau (1 - \tau) \sum_{j=1}^{j_0} \delta_j \geq C_\tau d^{-1} n \sum_{j=1}^{d} \delta_j. \tag{13.2}$$

From (13.2), one deduces

$$\begin{aligned}
\mathbb{P}(\mathcal{X}_1) &\leq \mathbb{P}(\max_{1 \leq i \leq n} \sum_{j=1}^{d} |S_{ij}| \geq 2^{-1} C_\tau d^{-1} n \sum_{j=1}^{d} \delta_j) + \mathbb{P}(\sum_{j=1}^{d} |S_{nj}| \geq 2^{-1} C_\tau d^{-1} n \sum_{j=1}^{d} \delta_j) \\
&\leq 2\mathbb{P}(\max_{1 \leq i \leq n} \|S_i\|_{\mathcal{L}_2} \geq C_\tau d^{-3/2} n \sum_{j=1}^{d} \delta_j) \\
&\leq d^3 (n \sum_{j=1}^{d} \delta_j)^{-2} \Delta_{q,2,\alpha}^2,
\end{aligned}$$

where the final inequality follows from an application of Markov's inequality to Theorem 5.6 of [80], along with noting that $\Delta_{2,2,\alpha} \leq \Delta_{p,2,\alpha}$ for $q \geq 2$. On the other hand,

$$\sum_{j=1}^{j_0} \delta_j - \sum_{j=j_0+1}^{d} \delta_j \geq j_0 d^{-1} \sum_{j=1}^{j_0} \delta_j \geq 2^{-1} d^{-1} \sum_{j=1}^{d} \delta_j. \tag{13.3}$$

Therefore, an analysis same as (10.13) implies that

$$\mathcal{X}_2 \implies \max_{k: |k - \tilde{k}| > L_\varepsilon d^3 \Delta_{p,2,\alpha}^2 (\sum_{j=1}^{d} \delta_j)^{-2}} \frac{\|V_k^e - V_{\tilde{k}}^e\|_{\mathcal{L}_2}}{|k - \tilde{k}|} \geq C_\tau d^{-3/2} \sum_{j=1}^{d} \delta_j.$$

Thus, a treatment similar to Lemma 10.2 entails $\mathbb{P}(\mathcal{X}_2) \leq C_\tau L_\varepsilon^{-1}$, which can be made arbitrarily small for large enough L_ε.

For the case when the sequence of jumps does not satisfy $\mathcal{P}_1$, it must satisfy $\mathcal{P}_2$. Let $j^\star = j_1$. Since $\mathcal{P}_1$ is not satisfied, and by definition of j_1, it must be true that

$$\frac{\sum_{j=1}^{j_1-1}\delta_j}{d} \le \delta_{j_1} \le \frac{\sum_{j=1}^{j_1}\delta_j}{d}. \tag{13.4}$$

We need to verify (13.1) and (13.3) in this case. For the first one, the second inequality in (13.4) reveals

$$\sum_{j=1}^{d}\delta_j \le \sum_{j=1}^{j_1-1}\delta_j + (d-j_1+1)\frac{\sum_{j=1}^{j_1}\delta_j}{d} \le 2\sum_{j=1}^{j_1-1}\delta_j + \delta_{j_1} \le 3\sum_{j=1}^{j_1-1}\delta_j.$$

Therefore, by the first inequality of (13.4), we can write an analogue of (13.2):

$$\mathcal{X}_1 \implies 2\max_{1\le i\le n}\sum_{j=1}^{d}|V_{ij}|^e \ge d^{-1}n\tau(1-\tau)\delta_{j_1} \ge C_\tau d^{-1}n\sum_{j=1}^{d}\delta_j.$$

On the other hand,

$$\sum_{j=1}^{j_0}\delta_j - \sum_{j=j_0+1}^{d}\delta_j \ge j_0 d^{-1}\sum_{j=1}^{j_0}\delta_j \ge 3^{-1}d^{-1}\sum_{j=1}^{d}\delta_j.$$

Hence, the analysis of $\mathcal{X}_2$ can be carried forward verbatim. This completes the proof. $\square$

Proof of Proposition 5.2. The proof for the H_0 follows trivially from Section 10.1. For the case under alternate, note that from (10.19), we write

$$\begin{aligned}
n^{-1/2}d^{-5/2}\Delta_{p,2,\alpha}^{-1}\sum_{j=1}^{d}(|f_{n,j}(n\tau_j)| - |f_{n,j}(n\hat\tau)|) &\ge \sqrt{n}d^{-5/2}\Delta_{p,2,\alpha}^{-1}\sum_{i=1}^{|\mathcal{J}|}Q_i|h_i-\hat\tau| \\
&\ge 2^{-1}\sqrt{n}d^{-5/2}\Delta_{p,2,\alpha}^{-1}\sum_{i=2}^{|\mathcal{J}|}Q_i(|h_i-\hat\tau| + |h_{i-1}-\hat\tau|) \\
&\ge 2^{-1}\mathcal{V}(n,d).
\end{aligned}$$

Here the second inequality is due to J being ordered, and the third assertion follows trivially from triangle inequality. The error term in (10.18) is controlled as follows

$$n^{-1/2}d^{-5/2}\Delta_{p,2,\alpha}^{-1}\sum_{j=1}^{d}\max_{1\le i\le n}|V_{ij}^{e}| = O_{\mathbb{P}}(d^{-2}) = O_{\mathbb{P}}(1).$$

This completes the proof in light of (5.3). $\square$

Proof of Theorem 5.1. We will follow the notations of [80], except X_t and X_t' therein are changed to $\mathbf{e}_t$ and $\mathbf{e}_t'$ to maintain our notational consistency. Consider Y_t' defined therein right after (A.4) of the supplementary and $\Delta_t = \Sigma_\infty - \mathrm{Cov}(Y_t')$. Per [80], there exist independent Gaussian random vectors $Y_t^* \sim \mathcal{N}(0, \Sigma_\infty)$ such that $Y_t^* - Y_t'$, $t = 1, \ldots, n$, are also independent random vectors with $(Y_t^* - Y_t') \sim \mathcal{N}(0, |\Delta_t|)$. Thus, by Doob's maximal inequality,

$$\begin{aligned}\left(\mathbb{E}\max_{k\le n}\left\|\sum_{t=1}^{k}(Y_t^* - Y_t')\right\|^2\right)^{\frac{1}{2}} \le C\left(\sum_{t=1}^{n}\mathbb{E}\|Y_t^* - Y_t'\|^2\right)^{\frac{1}{2}} &= C\left(\sum_{t=1}^{n}\mathrm{tr}(|\Delta_t|)\right)^{\frac{1}{2}} \\ &= C\left(\sum_{t=1}^{n}\|\Delta_t\|_{\mathrm{tr}}\right)^{\frac{1}{2}}.\end{aligned}$$

Let $t_i = iL$ with L being specified later. Now we make use of Proposition 1 in [79] to obtain the following improved estimate of $\gamma(h)$: $\|\gamma(h)\|_{\mathrm{tr}} \le \Theta^2(h+1)^{-\beta}$. This yields

$$\begin{aligned}\|\mathrm{Cov}(Y_t') - \Sigma\|_{\mathrm{tr}} &= \left\|\frac{1}{L}\sum_{s,s'=1}^{L}\gamma(s-s') - \sum_{h\in\mathbb{Z}}\gamma(h)\right\|_{\mathrm{tr}} \\ &\le 2\sum_{h\in\mathbb{N}}\|\gamma(h)\|_{\mathrm{tr}}\left(\frac{|h|\wedge L}{L}\right) \\ &\le C\Theta^2\sum_{h\in\mathbb{N}}|h+1|^{-\beta}\left(\frac{|h|\wedge L}{L}\right) \\ &\le C\Theta^2(L^{1-\beta} + L^{-1}). \qquad (13.5)\end{aligned}$$

Combining equation (19) of [80] with (13.5), we obtain

$$\left(\mathbb{E}\max_{k\le n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}(\mathbf{e}_t'-Y_t^*)\right\|^2\right)^{\frac{1}{2}} \le C\Theta\sqrt{\log(n)}\left(\frac{dL}{n}\right)^{\frac{1}{4}-\frac{1}{2q}} + C\Theta L^{\frac{1-\beta}{2}\vee(-\frac{1}{2})} \tag{13.6}$$

Minimizing the rates with a choice of $L=\lceil (d/n)^{c_1}\rceil$, with $c_1=\frac{\frac{1}{2p}-\frac{1}{4}}{\frac{1}{4}-\frac{1}{2p}-\frac{1-\beta}{2}\vee(-\frac{1}{2})}$ we obtain

$$(13.6)\le C\Theta\sqrt{\log(n)}\left(\frac{d}{n}\right)^{\xi(p,\beta)},. \tag{13.7}$$

where $\xi(q,\beta)=(1+c_1)(\frac{1}{4}-\frac{1}{2q})$. This establishes the claim (5.6). $\square$

Proof of Theorem 5.2. Consider the notations as in the Section 11.3. Write

$$\begin{aligned}
&\max_{1\le k\le B_n}|\hat{\Gamma}_k-\Gamma_k|_\infty \\
&\le \max_{1\le k\le B_n}\left|\frac{1}{n}\sum_{i=1}^{n-k}\mathbf{e}_i\mathbf{e}_{i+k}^\top-\Gamma_k\right|_\infty + \max_{1\le k\le B_n}\left|\frac{1}{n}\sum_{i=1}^{n-k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)(\hat{\boldsymbol{\mu}}_{i+k}-\boldsymbol{\mu}_{i+k})^\top\right|_\infty + \\
&\qquad \max_{1\le k\le B_n}\left|\frac{1}{n}\sum_{i=1}^{n-k}\left((\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)\mathbf{e}_{i+k}^\top+\mathbf{e}_i(\hat{\boldsymbol{\mu}}_{i+k}-\boldsymbol{\mu}_{i+k})^\top)\right)\right|_\infty.
\end{aligned} \tag{13.8}$$

We need to bound only the first two terms in (13.8), since the third term follows immediately via Cauchy-Schwarz inequality. Moreover, it follows from equation (3.13) of [121] that

$$\max_{1\le k\le B_n}\left|\frac{1}{n}\sum_{i=1}^{n-k}\mathbf{e}_i\mathbf{e}_{i+k}^\top-\Gamma_k\right|_\infty = O_{\mathbb{P}}\left(\frac{(\log d)^2\Delta_{q,\infty,\alpha}^2+\sqrt{\log dn}\Psi_{4,\alpha}^2}{\sqrt{n}}+\frac{1+B_n^{-\alpha+1}}{n}\Psi_{2,0}\Psi_{2,\alpha}\right). \tag{13.9}$$

On the other hand, note that

$$\left|\frac{1}{n}\sum_{i=1}^{n-k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)(\hat{\boldsymbol{\mu}}_{i+k}-\boldsymbol{\mu}_{i+k})^\top\right|_\infty \le \max_{1\le j\le d}\frac{1}{n}\sum_{i=1}^{n-k}(\hat{\mu}_{ij}-\mu_{ij})^2.$$

We will borrow the notation of Section 11.3. From (11.44),

$$\max_{1\le j\le d} \varsigma_{jj}(\mathcal{D}_j^{LR})^2 = O_{\mathbb{P}}(\sum_{j=1}^{d} \|X_{\cdot j}\|_{2,\alpha}^2) = O_{\mathbb{P}}(\Upsilon_{4,\alpha}).$$

Clearly, as discussed in Section 11.3, rest of the terms in (11.44) can be shown to be smaller. This completes the proof of (5.8). $\square$

14 Bootstrap in high dimension: a discussion

In a high-dimensional setting, if we have access to the knowledge of V_0 and V_1, the bootstrap algorithm theoretically works along the same lines as Algorithm 3. In fact, the following result can be derived exactly along the same lines as Theorems 4.1 and 4.2.

Theorem 14.1. *Grant Assumption 5.1. Impose the conditions of Theorems 5.1 and 5.2 onto the error process* $(\mathbf{e}_i)_{i\in\mathbb{Z}}$ *and the bandwidth* B_n*. Consider the Algorithm 3 with* V_0 *and* V_1 *known. Recall* $\daleth_1(d,n)$ *and* $\daleth_2(d,n)$ *from* (5.6) *and* (5.8) *respectively. Consider* $d = d_n$ *such that* $d\max_i \daleth_i(d,n) \to 0$ *as* $n\to\infty$*. Then, with the definition of* $b_\alpha(\nu,\Gamma)$ *as in Theorem 4.1, for any deterministic sequences* $u(n,d)\to 0$, $h(n,d)\to 0$ *satisfying*

$$u(n,d)^2 \gg h(n,d)^{-1}\daleth_2(d,n) + h(n,d)^{-2}\sum_{j\in\mathcal{V}_1}\frac{\Theta_{0,p,j}}{n\delta_j^2},$$

it holds that, under H_0*,*

$$\overline{\lim_{n\to\infty}}\, \mathbb{P}(T_n \ge b_{\alpha-h(n,d)}(\tilde{\boldsymbol{\mu}},\hat{\Sigma}_{n,B_n}) + u(n,d)) \le \alpha.$$

Moreover, if (5.3) *holds, then*

$$\underline{\lim}_{n\to\infty}\, \mathbb{P}(T_n \ge b_\alpha(\tilde{\boldsymbol{\mu}},\hat{\Sigma}_{n,B_n})) = 1.$$

The dependence of both d and n is rather implicit in Theorem 14.1 due to the general form of dependence that we assume. We believe that for a particular known structure of dependence, finer results with explicit restrictions on d and n might be achievable. More importantly, in high-dimension, the knowledge of $\mathcal{V}_0$ and $\mathcal{V}_1$ is a severely impractical assumption; our solution of using Algorithm 2 to discern between the two sets suffers from a serious issue of multiple testing error. One possible solution is to use the framework of [113]; however, due to the intricate nature of our problems, the dependence between the p-values is complicated, and the corresponding central limit theory might not hold (Condition 1 on [113]). If $|\mathcal{V}_0| \to \infty$ as $n \to \infty$, Bonferroni correction is a valid, but a very conservative solution, with sacrifices on finite sample power. A valid algorithm more tuned towards the vagaries of high-dimension, preferably much less conservative, should implement some form of false discovery control in order to estimate $\mathcal{V}_0$ and $\mathcal{V}_1$. We delegate this topic for future research purposes.

Instead, we showcase that in moderately high-dimensional set ups, when $|\mathcal{V}_0| << |\mathcal{V}_1|$ and $d << n$, our Algorithm 3 retains validity as well as consistency. We define the following two settings.

- **Setting A:** Consider the same TAR set-up as in (8.2), with $n = 300$, $d = 30$. B_n is taken to be 10. For each $1 \leq j \leq d$, we consider $\mu_j^L = 0$, and mimic the simulation exercise in Section 8.4.1 for the following scenario.

$$\delta_j = \begin{cases} \frac{6}{\log n}, & 1 \leq j \leq d/2, \\ -\frac{6}{\log n}, & d/2 < j \leq 5d/6, \\ 0, & 5d/6 < j \leq d, \end{cases} \text{ and } \tau_j = \begin{cases} 1/2, & j \in \{1, d\}, \\ 1/2 - r_1, & 2 \leq j \leq d/2, \\ 1/2 - r_2, & d/2 + 1 \leq j \leq d - 1, \end{cases}$$

  where we will let r_1, r_2 vary in $\{0, 0.01, \ldots, 0.1\}$.

- **Setting B:** Consider the GJR-GARCH(1,1) model in (8.3), but with innovations innovations $(\boldsymbol{\varepsilon}_i)_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, 1.25\boldsymbol{\Sigma}_{RQ}^{5,1})$. The jumps and change-points, δ_j's and τ_j's, are as follows.

$$\delta_j = \begin{cases} \frac{6}{\log n}, & 1 \le j \le d/2, \\ -\frac{6}{\log n}, & d/2 < j \le d - \lfloor d^{1/3} \rfloor, \\ 0, & d - \lfloor d^{1/3} \rfloor < j \le d, \end{cases} \text{ and } \tau_j = \begin{cases} 1/2, & j \in \{1, d\}, \\ 1/2 - r_1, & 2 \le j \le d/2, \\ 1/2 - r_2, & d/2 + 1 \le j \le d - 1, \end{cases}$$

with $r_1, r_2 \in \{0, 0.01, \ldots, 0.1\}$

Similar to Sections 8.4.1 and 8.4.2, the bootstrap quantiles are estimated based on 5000 bootstrap samples. The type-1 error and power are estimated based on 1000 independent Monte Carlo samples of $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. Tables 6 and 7 showcase the results. Even though the results point towards both the validity and consistency of the bootstrap-based tests, one peculiar pattern present in the tables, is worth discussing. Note that, the values of the diagonal elements of this table are particularly small compared to the other values in the corresponding rows. This can be explained in light of (5.3), recalling the notations therein. For example, in Setting A, when $r_1 = r_2$, there are only two distinct change-points in the alternative with $|D_{1/2}| = 1$, $|D_{1/2-r_1}| = 5d/6 - 1$. Therefore, $Q_2 = 6/\log n$, and (5.3) can be simplified to

$$\frac{\sqrt{n}}{d^{5/2} \log n} r_1 \Delta_{p,2,\alpha}^{-1} \to \infty. \tag{14.1}$$

On the other hand, for off-diagonal entries $r_1 \neq r_2$, a similar calculation condenses (5.3) to

$$\frac{\sqrt{n}}{d^{5/2} \log n} (r_2 + \frac{d}{3} |r_1 - r_2|) \to \infty. \tag{14.2}$$

Clearly, (14.2) is a much weaker condition that (14.1); thus the diagonal case is a decidedly harder alternative, resulting in lesser power. An exactly similar calculation also holds in Setting B.

r_1	r_2										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
0	0.044	0.353	0.916	0.997	1	1	1	1	1	1	1
0.01	0.398	0.083	0.511	0.935	0.997	0.999	1	1	1	1	1
0.02	0.887	0.47	0.116	0.612	0.967	0.999	0.999	1	1	1	1
0.03	0.984	0.91	0.599	0.19	0.712	0.977	0.997	1	1	1	1
0.04	1	0.989	0.948	0.72	0.289	0.804	0.984	0.999	1	1	1
0.05	1	0.999	0.997	0.961	0.772	0.362	0.869	0.994	0.999	1	1
0.06	1	1	0.998	0.997	0.974	0.819	0.445	0.901	0.988	1	1
0.07	1	1	1	0.999	0.998	0.981	0.873	0.594	0.944	0.992	1
0.08	1	1	1	1	1	0.999	0.985	0.904	0.672	0.947	0.998
0.09	1	1	1	1	1	1	0.999	0.991	0.932	0.73	0.976
0.1	1	1	1	1	1	1	0.999	0.996	0.996	0.951	0.811

Table 6: Type-1 error ($r_1 = r_2 = 0$) and Power of Algorithm 3 for the Setting A.

r_1	r_2										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
0	0.08	0.188	0.67	1	1	1	1	1	1	1	1
0.01	0.452	0.114	0.164	1	1	1	1	1	1	1	1
0.02	0.784	0.392	0.2	0.164	0.759	0.99	1	1	1	1	1
0.03	1	0.85	0.388	0.262	0.198	0.802	1	1	1	1	1
0.04	1	1	0.858	0.361	0.331	0.181	0.801	1	1	1	1
0.05	1	1	0.996	0.856	0.326	0.405	0.192	0.86	0.997	1	1
0.06	1	1	1	0.996	0.878	0.364	0.493	0.215	0.874	0.997	1
0.07	1	1	1	1	0.993	0.878	0.308	0.581	0.278	0.874	0.994
0.08	1	1	1	1	1	0.997	0.914	0.34	0.66	0.254	0.888
0.09	1	1	1	1	1	1	1	0.918	0.356	0.726	0.293
0.1	1	1	1	1	1	1	1	1	0.922	0.342	0.784

Table 7: Type-1 error ($r_1 = r_2 = 0$) and Power of Algorithm 3 for Setting B.