TESTING SYNCHRONIZATION OF CHANGE-POINT IN MULTIPLE TIME-SERIES

By Sayar Karmakar, Maggie Cheng and Wei Biao Wu

University of Florida, Illinois Institute of Technology, University of Chicago

In this paper, we investigate the problem of detecting a change-point in a multiple time-series for both fixed and high-dimensions. For the fixed dimensional case, we detect change-points for each individual co-ordinates using a moving average technique and focus on testing synchronization of these change-points. The identification of synchronized change-points can often lead to finding an unanimous reason behind such changes. We provide an application of our study in speedy recovery of power grid system. We also build a theoretical framework for testing synchronization in a high-dimensional regime where our testing method is different from the fixed dimension.

1. Introduction. Change-point analysis is an important tool to identify the location/time and impact of a distributional change while studying a stochastic process over time. Page (1955, [20] 1957, [21]) is considered to be the pioneer who initiated this long and well-studied research topic in statistics and electrical engineering. The CUSUM chart was proposed by Hinkley (1971, [12]) and Pettitt (1980, [22]). A bootstrapping idea was suggested by Hinkley and Schechtman, (1987, [13]). These are some of the seminal works towards detecting one change in a single time-series.

Over the past two decades, significant amount of research are also being done in analyzing structural changes in multiple time-series. Such processes where change-point analysis can be meaningful are prominent in the fields of finance, neuro-science, signal processing, biology and medicine among others. See Vert and Bleakley (2010) for a list of such applications. Ombao et al. (2005, [19]) employed the SLEX (smooth localized complex exponentials) whereas Lavielle and Teyssi'ere (2006, [16]) introduced a procedure based on penalized Gaussian log-likelihood function. The change-point detection problem was re-formulated as a penalized regression problem and was solved by the group Lasso (Yuan and Lin (2006, [26]). CUSUM-type statistics have been widely used in time series segmentation. In the context of multivariate time series segmentation, Groen et al. (2013, [10]) and Horváth, Lajos and Hušková, Marie (2012, [15]) studied the average of d CUSUM statistics, each obtained from one component of a d-dimensional time series. The average test statistic was also adopted in for detecting a single change in the mean of a panel data model. Aue et al. (2009, [2]), detection of a single change-point in the covariance structure of multivariate time-series using a CUSUM statistic.

Keywords and phrases: Change-point, Local Linear Average, Gaussian Approximation

Very recently Hoga(2016, [14]) used a Gaussian approximation result by Liu and Lin (2009, [17]) to study asymptotic properties of the CUSUM procedure for a stationary multiple time-series. However, the Gaussian approximation result from [17] was improved in a recent work by Karmakar and Wu (2017) where un-improvable approximation bounds were achieved. We use this new result to detect existence and synchronization of change-points. We choose a MOSUM technique instead of CUSUM as the latter is often criticized to perform well only in the situation where change-point occurs early. Moreover, MOSUM is computationally easier since at any given time point one needs to compute the average of observations falling in a small window.
Consider the multiple series sequence model

$$\mathbf{X}_i = \mu(i/n) + \mathbf{e}_i = (\mu_1(i/n), \ldots, \mu_d(i/n))^T + \mathbf{e}_i, \tag{1.1}$$

where $(\mathbf{e})_{i\in\mathbb{N}}$ denotes the unobserved multivariate error process and $\mu_r(i)$ is the signal for the rth series at time index i. We discuss a very simple structural assumption first and slowly generalize in multiple directions to substantially increase the scope of applicability. Assume, for each co-ordinate r, μ_r is a piecewise constant function from $[0,1] \to \mathbb{R}$. We also assume there are at most two pieces for each such μ_r function. For each $1 \le r \le d$ we denote $0 \le \tau_r \le 1$ to be the change-point and let λ_r denote the jump length at the point τ_r. More formally,

$$\mu_r(x) = \begin{cases} \mu_{r1}, & \text{if } x < \tau_r \\ \mu_{r2}, & \text{if } x \ge \tau_r \end{cases} \tag{1.2}$$

The jump length λ_r at the point τ_r is defined as $\lambda_r = \mu_{r2} - \mu_{r1}$. We will first detect whether there is any change-point in any of these series first. That is same as testing the following hypothesis.

$$H_{10} : \lambda_1 = \ldots = \lambda_d = 0. \tag{1.3}$$

Provided we reject (1.3), we will test the following hypothesis

$$H_{20} : \tau_1 = \ldots = \tau_d. \tag{1.4}$$

The setting in model (1.2) is a very simplistic one and thus require generalization in different directions to enlarge the scope of application. We address two types of generalization: first we allow the different pieces for every co-ordinates to be continuous instead of just constant. Secondly, we also allow more than one change-point. This generalization needs to be handled with some care since if every series is allowed to have multiple change-point,

then the question of synchronization is no longer valid. We discuss the multiple change point scenario in Section **??**.

We do not put any special form in the signal part. Instead, we allow both time and panel dependence in the error process. Some mild and easily verifiable moment conditions for the time-dependence using Wu (2005, [23])'s framework of functional dependence measure are imposed. We show in Section **??** how one can also transfer the problem of detecting existence and synchronization of change-points in covariance process of a multiple time-series in our framework. While there are a lot of work in the literature about the convergence rate of the change point location estimators, little has been done to develop an inferential theory in this direction.

Karmakar and Wu (2018+) obtained an invariance principle for the partial sums of the vector-valued process e_i. In particular, let $S_i = \sum_{j=1}^{i} e_j$ be the partial sum process of (e_i). We approximate the process S_i by a Gaussian process with independent (but not necessarily identically distributed) increments. We will show, under suitable conditions, we can construct a Gaussian process G_i on a richer probability space and a process S_i' such that (S_i) and (S_i') are identically distributed for all $1 \leq i \leq n$ and

$$\max_{1\leq i\leq n} |S_i' - G_i| = o(\alpha_n). \tag{1.5}$$

If the approximation error (α_n) is small enough, then we could use functional involving the Gaussian process to approximate the statistics involving e_i. This will be our key tool for testing the synchronization since otherwise the null distribution of the proposed test statistics for testing (1.4) will be difficult to obtain. We will obtain a Gaussian analogue of the statistic and thus obtain the bootstrap distribution of the analogue to create an inferential framework.

We conclude the introduction with some notations and assumptions that will be used throughout.

1.1. *Notations.* For a matrix $A = (a_{ij})$ we define it Frobenius norm as $|A| = (\sum a_{ij}^2)^{1/2}$. For a random vector Y , write $Y \in \mathcal{L}_p, p > 0$, if $\|Y\|_p := [\mathbb{E}(|Y|^p)]^{1/p} < \infty$. For $\mathcal{L}_2$ norm write $\|\cdot\| = \|\cdot\|_2$. Define the projection operator P_i by

$$P_i Y = \mathbb{E}(Y|\mathcal{F}_i) - \mathbb{E}(Y|\mathcal{F}_{i-1}), \qquad Y \in \mathcal{L}_1. \tag{1.6}$$

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. $N_p(\mu, \Sigma)$ means p-variate normal distribution with mean μ and covariance matrix Σ. $Var(Y)$ or $Cov(Y)$ for a random vector Y stands for the variance-covariance matrix of Y. For a positive semi-definite matrix A, $A^{1/2}$ refers to the usual Grammian square root of A. If $A = QDQ^T$ is the spectral decomposition of the matrix A then $A^{1/2} = QD^{1/2}Q^T$

If two quantities M and N satisfy $M <= cN$ for some $c < \infty$ then we write $A \preceq B$ or $B \preceq A$. If both $A \ll B$ and $B \ll A$ then we write $A \asymp B$. We use the same symbols if such relationships holds for large n, the sample size as our result is anyway asymptotic.

1.2. *The error process and assumptions.* We will assume e_i is a very general non-gaussian dependent process. In order to do some meaningful analysis to retrieve the unknown or the hidden function μ from the observed X_i we need to order some dependence structure on the process (e_i). We assume the following causal representation for the process (e_i).

$$e_i = H(\epsilon_i, \epsilon_{i-1}, \dots,), \tag{1.7}$$

where H is a measurable function taking values in $\mathbb{R}^d$ and ϵ_i's are independent and identically distributed innovations. This representation allows us to use the widely used idea of coupling, first introduced by Wu(2005, [23]) to model the dependence structure.
Consider the process in (1.7). To regularize it with a Gaussian process, we first introduce uniform functional dependence measure on the underlying process. We will use the idea of coupling as done in Wu[2005, [23]] Suppose $(\epsilon_i')_{i\in\mathbb{Z}}$ is an independent copy of $(\epsilon_i)_{i\in\mathbb{Z}}$. Assume that X_i has mean 0 and $e_j \in \mathcal{L}^p, p > 0$. For $j \geq 0$, define the functional dependence measure

$$\theta_{i,p} = \|e_i - e_{i,0}\|_p = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,0}\|_p, \tag{1.8}$$

where $\mathcal{F}_{i,k}$ is the coupled version of $\mathcal{F}_i$ with ϵ_k in $\mathcal{F}_i$ replaced by an i. i. d copy ϵ_k',

$$\mathcal{F}_{i,k} = (\epsilon_i, \epsilon_{i-1}, \dots, \epsilon_k', \epsilon_{k-1}, \dots,) \tag{1.9}$$

and $e_{i,\{i-j\}} = H(\mathcal{F}_{i,\{i-j\}})$. Clearly, $\mathcal{F}_{i,k} = \mathcal{F}_i$ is $k > i$. As Wu(2005, [23]) suggests, $\|H(\mathcal{F}_i) - H(\mathcal{F}_{i,(i-j)})\|_p$ measures the dependence of X_i on ϵ_{i-j}. We now write down the assumptions. For completeness we again mention the causal representation of the (e_i) process.

(2.A) e_i is a $d-$dimensional mean 0, non-stationary random sequence that assumes the following representation

$$e_i = H(\mathcal{F}_i) = (X_{i1}, X_{i2}, ...X_{id})^T,$$

where $\mathcal{F}_i = (..., \epsilon_{i-1}, \epsilon_i)$, H is a measurable function such that X_i is a well-defined random vector, and ^{T} denotes matrix transpose. The ϵ_i random variables are i. i. d.. We also assume that, $e_i \in \mathcal{L}^p$ where $p > 2$. In other words, $\sup_i \|X_i\|_p < \infty$.

(2.B) We assume short range dependency,

$$\Theta_{0,p} = \sum_{i=m}^{\infty} \theta_{i,p} < \infty. \tag{1.10}$$

This condition implies the cumulative dependence of $(X_j)_{j\geq k}$ on ϵ_k is finite. If this fails, then the X_i process can be long-range dependent, and the partial sum process cannot be regularized by Gaussian process. In the next section, while we state our main theorem and corollaries we will further impose some restrictions on the rate of $\Theta_{i,p}$

From Berkes et al. (2014, [5]), we have the following result for scalar stationary process (e_i).

Result Suppose (2.A)-(2.C) holds. In addition to that, we assume

$$\Theta_{i,p} = O(i^{-\chi}),$$

with χ satisfying

$$\chi > \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2+20p+4}}{8p}, \tag{1.11}$$

there exists a probability space (Ω_c, A_c, P_c) on which we can define random vectors Y_i^c with the partial sum process $S_n^c = \sum_{i=1}^n Y_i^c$ and a Gaussian process G_i with independent increments such that $S_i^c \stackrel{D}{=} (S_i)_{i\in N}$ and

$$\max_{i\leq n} |S_i^c - G_i| = o_{\mathbb{P}}(n^{1/p}) \qquad \text{in } (\Omega_c, A_c, P_c) \tag{1.12}$$

where $G_i = \Sigma^{1/2} \sum_{j=1}^i Z_i$ with Z_i being i.i.d. standard normal random variables and $\Sigma = \sum_{k=0}^{\infty} \mathbb{E}(e_0 e_k)$ is the long-run covariance of the e_i process.

Karmakar and Wu (2018+) extended this result to a multi-variate and non-stationary process. As a special case, if the error process is assumed to be stationary, a similar result as (1.2) holds with $\Sigma = \sum_{k=0}^{\infty} \mathbb{E}(e_0 e_k^T)$ being the corresponding long-run covariance matrix of the vector valued (e_i) process.

1.3. *Organization.* The rest of the paper is organized as follows. We will first specify the mathematical model and the specifications for the error process in Section **??**. Section **??** describe our methods of detecting the change-points and if there are change-points how to detect whether they occur simultaneously or not. Next, in Section **??** we validate our detection methods with some theoretical results. We study an analogous problem in exploring contemporaneous covariance process in Section **??**. Section 3.1 is used to describe a Gaussian block multiplier technique to practically implement this. Network anomaly detection is discussed in Section 5.2 as a key application of our study. Section 6 shows some data simulations to substantiate our theoretical methods.

2. At most one change point. In this section we discuss the set-up where we can have at most one change point (AMOC hereafter) in each series. Starting from the simplest piecewise constant set-up we build the local linear estimation for the more general piecewise continuous case.

Assume (1.2). Consider the following left and right average process for the r th coordinate

$$\theta_r^L(i) = \frac{1}{G}\sum_{k=i-G+1}^{i} X_{i,r} \quad \theta_r^R(i) = \frac{1}{G}\sum_{k=i-G+1}^{i} X_{i,r} \tag{2.1}$$

The change-point location rescaled to (0,1) can then be estimated by the following

$$\hat{\tau}_r = \frac{1}{n}\arg\max_{G\le i\le n-G} |\theta_r^L(i) - \theta_r^R(i)|. \tag{2.2}$$

2.1. *Piecewise continuous signals.* Viewing (2.2) as a primer one can generalize the left and right simple average to a more general local constant average technique that can accommodate piecewise continuous signals. Reference lists.

We consider the following assumption about the signal functions μ_r for $1 \le r \le d$ for the rest of this section:

$$\mu_r(x) = \begin{cases} \mu_{r1}(x), & \text{if } x < \tau_r \\ \mu_{r2}(x), & \text{if } x \ge \tau_r, \end{cases} \tag{2.3}$$

where μ_{r1} and μ_{r2} are two piece wise continuous functions with the possibility of a jump at τ_r i.e. $\lambda_r = \mu_{r2}(\tau_r) - \mu_{r1}(\tau_r)$ is possibly non-zero. With a slight abuse of the notation of $\theta_r^L(\cdot)$ and $\theta_r^R(\cdot)$ we can proceed with

$$\begin{aligned} \theta_r^L(t) &= \sum_{i-m+1}^{i} K\left(\frac{i/n - t}{b_n}\right) X_{i,r} \\ \theta_r^R(t) &= \sum_{i+1}^{i+m} K\left(\frac{i/n - t}{b_n}\right) X_{i,r} \end{aligned} \tag{2.4}$$

Then the estimator is defined by

$$\hat{\tau_r} = \arg\max_{b_n<t<1-b_n} |\theta_r^L(t) - \theta_r^R(t)| \tag{2.5}$$

Since local-constant approaches such as Pristley-Chao estimate or Nadaraya- Watson estimates suffer from boundary conditions, we settle with a local-linear estimation. Incidentally, one can motivate the estimators in (2.4) as inspired by the following optimization problem

$$\hat{\theta}_t = \arg\max_{\eta_1} \sum_{i=1}^{n} (y_{i,r} - \eta_1)^2 K\left(\frac{i/n - t}{b_n}\right).$$

Under the assumption of smooth derivatives of all the piecewise signals for all the series one can instead use the following minimization

$$(\hat{\theta}(t), \hat{\theta}'(t)) = \arg\max_{\eta_1,\eta_2} \sum_{i=1}^{n} (y_{i,r} - \eta_1 - \eta_2(i/n - t))^2 K\left(\frac{i/n - t}{b_n}\right)$$

and its restriction to the left and right averages. Define the local linear averages ([8]) as

$$(\hat{\theta}_L(t), \quad \hat{\theta}_R(t) = \sum_{i=1}^{[nt]} X_i w_i, \quad \hat{\theta}_R(t) = \sum_{i=[nt]+1}^{n} X_i w_i, \quad w_i = K\left(\frac{t - i/n}{b_n}\right) \frac{S_2(t) - (t - i/n)S_1(t)}{S_2(t)S_0(t) - S_1(t)^2},$$

where $S_i(t) = \sum_{i=1}^{n} K((i/n - t)/b_n)(t - i/n)^j$.The estimator of the change-point location is defined by (2.5).

We propose some consistency and distributional result for τ_r. While the consistency results, since our d is finite reduces to stating consistency results for only one series and has been well-studied under literature, the distributional theory of the location point for the multiple process is, to the best of our knowledge is a new contribution. Also our results allow a large class of dependent processes that generalize the previously stated similar results in literature for independent or restrictive class of dependent process.

We first develop a Hàjek-Rènyi type inequality for stationary sequences with the representation (1.7). Since this may be of independent interest we state this result here. This is a key tool in not only change-point analysis but also in different limit theorems. After Hàjek and Rényi (1955,[11]) obtained similar result for independent random variables, Birnbaum and Marshall ([6]) extended it to martingales. Bai (1994, [3]) and Bai (1997, [4]) generalized this to linear process and mixingales respectively. However, based on some recent works in the framework of Wu (2005, [24], it is possible to extend this to a more general class of stationary processes.

PROPOSITION 2.1. *Assume e_i admits the causal representation in (1.7) with the long-run covariance σ. Then,*

$$\mathbb{P}(\max_{m \le k \le n} \frac{1}{k} |\sum_{i=1}^{k} e_i| > \alpha) \le C \frac{\sigma^2}{m\alpha^2}. \tag{2.7}$$

Next we discuss consistent estimation of the change-points for a single series. Since for the scope of this paper, we restrict ourselves to finite dimensions, it suffices to prove consistency for the process observed in a single co-ordinate. Thus, we suppress the suffix of $\tau_r = k_r/n$, $\hat{\tau}_r = \hat{k}_r/n$, λ_r etc. Also, note that, the jump length $\lambda = \lambda(n)$ and the window-length $G = G(n)$ are functions of sample-size n but we will, in the sequel, call it just λ and G.

First, as is expected, we show that the estimated change-point cannot be too far from the true change-point. Since we use our window size is G, one would naturally expect

$$|\hat{k} - k_0| = o_{\mathbb{P}}(G).$$

In fact, it is possible to achieve much sharper consistency result with some mild assumption on the jump-length. Without loss of generality, we assume $\lambda > 0$. Using the fact that, for $G \leq k \leq n - G$,

$$0 \leq |V_k| - |V_{k_0}| \leq \sup_{G \leq k \leq n-G} |V_k - \mathbb{E}(V_k)| + |\mathbb{E}(V_k)| - |\mathbb{E}(V_{k_0})|$$

and

$$\mathbb{E}(V_k) = \frac{G - |k - k_0|}{G} \lambda \mathbf{1}_{k_0 - G < k < k_0 + G},$$

one can see

$$|\hat{\tau} - \tau| \leq 2 \frac{G}{n\lambda} \sup_{G \leq k \leq n-G} |V_k - \mathbb{E}(V_k)|.$$

Using invariance principle from Berkes, Liu and Wu ([5]) and Theorem 1.2.1 from Csörgö and Révész ([7]) we have,

$$\begin{aligned} \sup_{G \leq k \leq n-G} |V_k - \mathbb{E}(V_k)| &= \sup_{G \leq k \leq n-G} \frac{1}{G} \left| \sum_{i=k+1}^{k+G} e_i - \sum_{i=k-G+1}^{k} e_i \right| \\ &= o_{\mathbb{P}}\left(\frac{n^{1/p}}{G} \right) + \frac{2}{G} \sup_{0 \leq t \leq n-G} \sup_{0 \leq s \leq G} |W(t+s) - W(t)|) \\ &= o_{\mathbb{P}}\left(\frac{n^{1/p}}{G} \right) + O_p\left(\frac{\log(n/G)^{1/2} + (\log\log n)^{1/2}}{G^{1/2}} \right). \end{aligned} \tag{2.8}$$

Thus the conditions on G and λ

$$\frac{n^{1/p}}{\lambda G} \to 0, \frac{(\log\log n)^{1/2} + \log(n/G)^{1/2}}{\lambda G^{1/2}} \to 0 \tag{2.9}$$

ensures consistency of $\hat{\tau}$. We move on to provide a sharper convergence rate adapted for a single series in the following theorem. This improved rate is also necessary to perform inferential asymptotics of the distribution of $\hat{\tau}$.

THEOREM 2.2. *Assume λ and G satisfy (2.9) Then*

$$\hat{\tau} - \tau = O_{\mathbb{P}}(n^{-1}\lambda^{-2}). \tag{2.10}$$

Next, we provide a distributional theory for the locations of the change-point. This can be used to construct confidence intervals for the location. This line of work, to the best of our knowledge, is not discussed very much in the literature and thus call for an important exploration. Similar type of work has been done for univariate series for quite some time. But extending the same to multiple dimension comes with its own challenge and possibly that is why this problem is not explored for a general scenario. Define two-sided Brownian motion $W(t)$ as follows

$$W(t) = \begin{cases} W_1(t) & \text{if } t \geq 0, \\ -W_2(t) & \text{otherwise,} \end{cases} \tag{2.11}$$

where W_1 and W_2 are independent Brownian motion on $(0, \infty)$. Next, we state a distributional result on the change point location estimator $\hat{\tau}$. This can be used to test point hypotheses about the location of the process.

THEOREM 2.3. *Under the conditions of Theorem 2.2, we have*

$$n\lambda^2(\hat{\tau} - \tau) \xrightarrow{D} \arg\max(\sigma W(t) - |t|). \tag{2.12}$$

where σ is the long-run variance of the error process (e_i).The next theorems concern consistently estimating the jump-length at the change-points and then obtain distributional result about the statistic. If the jump length λ_r is positive then

THEOREM 2.4. *For $1 \leq r \leq d$, we assume that the r^{th} co-ordinate function μ_r has a jump of length $\lambda_r > 0$ at τ_r. If G satisfies the condition in (2.9) then*

$$\sqrt{M_r} - \lambda_r = o_{\mathbb{P}}(1).$$

Next we obtain the distributions of individual $\sqrt{M_r}$ using the extreme value theory for Gaussian random variables.

THEOREM 2.5. *If λ and G satisfies (2.9), we have,*

$$\sqrt{2\log(n/G)}\left(\sqrt{G}\sigma^{-1}\sqrt{M_r} - [2\log(n/G) + \frac{1}{2}\log\{\log(n/G)\}] - \log(3)\right) \xrightarrow{D} V, \tag{2.13}$$

where σ is the long-run covariance for the r^{th} error series $\{e_r\}$ and V has extreme value distribution $\mathbb{P}(V \leq u) = \exp\{-\pi^{-1/2}\exp(-u)\}$.

Remark: One can easily obtain the asymptotic distribution of M_r from Theorem 2.5. The next theorem discusses the asymptotic distribution of $\sqrt{M_G}$ under the null $H_{20} : \tau_1 = \tau_2 = \ldots = \tau_d$
Let us define

$$B_d(x) = \sqrt{2\log(x)} + \frac{\log(C_K) + (d/2 - 1/2)\log(\log r) - \log(2)}{\sqrt{2\log(x)}},$$

with $C_K = \frac{(\int_{-1}^{1} |K'(u)|^2 du)^{1/2}}{\gamma s/2}$.

THEOREM 2.6.

$$\sqrt{2\log(n/G)}\left(\sqrt{\frac{G}{\phi_0}}\sqrt{M_G} - B_d\left(\frac{n}{G}\right)\right) \xrightarrow{D} V, \tag{2.14}$$

where V has extreme value distribution $\mathbb{P}(V \le u) = \exp\{-2\exp(-u)\}$.

For more discussion about the quantiles of the process on the right hand side in (2.14), see [9] and the references therein. Next we discuss, under the assumption of one existing change-point location, the joint distribution of the estimated location parameter. Even though, the literature of change-point analysis is very rich, to the best of our knowledge this is the first treatment that allows dependence in the error process and possibly asynchronized change-points across different co-ordinates. For ease of notation, we state the following result for $d = 2$ but it is obvious how to extend this for higher but finite d

THEOREM 2.7. *Under conditions on b_n, we have, for the estimated location vector*

$$\begin{pmatrix} c_n(\hat{\tau}_1 - \tau_1) \\ c_n(\hat{\tau}_1 - \tau_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \arg\max_{0\le t\le 1}(\mathbb{B}(t+\tau_1) - \mathbb{B}(t)) - 1/2|t|) \\ \arg\max_{0\le t\le 1}(\mathbb{B}(t+\tau_2) - \mathbb{B}(t)) - 1/2|t| \end{pmatrix}. \tag{2.15}$$

This result can be used for inference on the location of the change-points in different co-ordinates. Specifically this can also be used to identify the synchronization of the locations. However we resort to a different approach based on the maximum values of the individual and overall objective functions.

3. Testing for synchronization. Before discussing our method to answer the question of synchronization, we provide a motivation behind our proposal. Consider $d = 2$. Intuitively if $\tau_1 = \tau_2$, provided we have a consistent estimation procedure for τ_r, one would expect $\hat{\tau}_1 - \hat{\tau}_2$ to be small. However, such a notion of distance will be complicated if $d > 2$. Drawing analogy between a t-test of equality for two treatments and ANOVA for multiple, we focus on the maximized values of the objective functions to judge closeness of $\tau_i's$. To

test the combined hypothesis, $H_{20} : \tau_1 = \tau_2 = \ldots = \tau_r$ we define $M_r = |\theta_r^L(n\hat{\tau}_r) - \theta_r^R(n\hat{\tau}_r)|$. Define the synchronized maximizer

$$M_G = \max_{1\le i\le n} \sum_{r=1}^{d} |\frac{1}{G} \sum_{j=i-G+1}^{i} X_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} X_{j,r}|^2 \tag{3.1}$$

The following inequality is always true.

$$M_G \le M_1 + M_2 + \ldots + M_d.$$

Under the null that the change-points synchronize

$$\tau_1 = \ldots = \tau_d$$

we should have a small value of $M_1 + M_2 + \ldots + M_d - M_G$. As mentioned before, we will obtain the distribution of M_i for all i and M_G. But, obtaining the theoretical asymptotic distribution of $M_1+M_2+\ldots+M_d$ or $M_1+M_2+\ldots+M_d-M_G$ under the null hypothesis is not an easy task. The distribution probably does not have a well-known name. But it is crucial to obtain the null distribution to perform the tests. We will adopt a Bootstrap technique along with our gaussian approximation result, Theorem (1.2) to circumvent that. We use summation by parts argument we can instead use the corresponding gaussian versions of these test statistics and use a bootstrap method to obtain the theoretical distribution and consequently the threshold for testing perspective.

3.1. *Bootstrap method to find the proper threshold.* In this section, we discuss a bootstrap method to approximate the asymptotic null distribution of our test statistics. This involves consistent estimation of the long-run covariance matrix which is discussed later. Using the atomic gaussian process $g_i = \hat{\Sigma}^{1/2} Z_i$, we obtain the gaussian analogues $M_1^Z, M_2^Z, \ldots, M_d^Z$ and M_G^Z of the corresponding quantities $M_1, M_2, \ldots, M_d$ and M_G respectively. They are derived from the following equations.
For $1 \le r \le d$,

$$M_r^Z = \max_{1\le i\le n} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2, \tag{3.2}$$

$$\hat{\tau}_r^Z = \arg\max_{1\le i\le n} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2. \tag{3.3}$$

Let

$$M_G^Z = \max_{1\le i\le n} \sum_{r=1}^{d} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2. \tag{3.4}$$

THEOREM 3.1. *For any $1 \leq r \leq d$ we have*

$$|M_r^Z - M_r| = o_{\mathbb{P}}(\frac{n^{2/p}}{G^2}).$$

$$|M_G^Z - M_G| = o_{\mathbb{P}}(\frac{n^{2/p}}{G^2}).$$

In other words

$$\frac{G^2}{n^{2/p}}(M_r^Z - M_r) \xrightarrow{P} 0.$$

$$\frac{G^2}{n^{2/p}}(M_G^Z - M_G) \xrightarrow{P} 0.$$

PROOF. From the invariance principle in Karmakar, Wu (2016) we have

$$\begin{aligned}
\sqrt{M_G} &= \max_{G \leq i \leq n-G} |\frac{1}{G}\{S_{i+G} - 2S_i + S_{i-G}\}| + O(\frac{1}{G}) \\
&= \max_{G \leq i \leq n-G} |\frac{1}{G}\{G^*_{i+G} - 2G^*_i + G^*_{i-G}\}| + o_{\mathbb{P}}(\frac{n^{1/p}}{G}) \\
&= \sqrt{M_G^Z} + o_{\mathbb{P}}(\frac{n^{1/p}}{G}).
\end{aligned}$$

Similar result holds for individual co-ordinates with M and M_r replaced by M_G^Z and M_r^Z. □

In this section, we provide algorithm description for the detection of change points and testing of synchronized changes.

1. Compute $M_1, \ldots, M_d$ and M_G using (**??**) and (3.1) respectively.
2. Use the input data X_{ij} for estimation of $\hat{\tau}_r$, as in (**??**). Use the estimated τ to decide X^{mean1} and X^{mean2}, and use them to de-mean X_{ij} and get X^*_{ij}. Calculate dispersion estimate $\hat{\Sigma}$ using X^*_{ij} as in (3.6).
3. Obtain the gaussian analogues $M_1^Z, M_2^Z, \ldots, M_d^Z$ and M_G^Z by using the atomic gaussian process g_i as in (3.2) and (3.4). g_i is generated by using the estimated $\hat{\Sigma}$ and i.i.d. standard normal random variables Z_i: $g_i = \hat{\Sigma}^{1/2} Z_i$.
4. Repeat Step 3 for a large number of times (e.g., 1,000 times). Get the bootstrapped distribution of $M_1^Z + M_2^Z + \ldots + M_d^Z$ and $M_1^Z + M_2^Z + \ldots + M_d^Z - M_G^Z$).
5. Let $c1$ and $c2$ be the 95-th percentiles for the above distributions, respectively.
6. Reject the null of no change-point H_{10} at 5% level of significance if

$$M_1 + M_2 + \ldots + M_d > c_1.$$

7. If we have rejected H_{10}, reject the null of synchronization H_{20} at 5% level of significance if

$$M_1 + M_2 + \ldots + M_d - M_G > c_2.$$

Remark: To start one can take $G = n^{1/3}$.

3.2. *Estimating Σ for a real data application.* In order to use the Bootstrap technique mentioned in section 4, we will need to get an usable version of the G_i process mentioned in Theorem (1.2). For this purpose one needs to estimate the long-run covariance matrix Σ.
We first define the demeaned series X^* based on the estimated $\tau_i's$. Since we are assuming our μ function is piece-wise constant, we have at most two different values of the function μ. The demeaning process is motivated from that fact. For $1 \leq j \leq d$, let $\hat{\tau}_j$ be the estimated change-point location as mentioned in (**??**). Denote,

$$X_j^{mean1} = \frac{1}{\hat{\tau}_j} \sum_{i=1}^{\hat{\tau}_j} X_{ij} \text{ and } X_j^{mean2} = \frac{1}{n - \hat{\tau}_j} \sum_{i=\hat{\tau}_j+1}^{n} X_{ij}.$$

$$X_{ij}^* = \begin{cases} X_{ij} - X_j^{mean1}, & \text{for } 1 \leq i \leq \hat{\tau}_j \\ X_{ij} - X_j^{mean2}, & \text{for } \hat{\tau}_j + 1 \leq i \leq n \end{cases}$$

Let us define the following estimate of the long-run covariance matrix $\hat{\Sigma}$.

$$\hat{\Sigma} = \frac{1}{G(n - G + 1)} \sum_{j=0}^{n-G} (S_{j+G}^* - S_j^*)(S_{j+G}^* - S_j^*)^T, \tag{3.5}$$

where S_i^* is the related partial sum process generated from the vector process X_i^*.
We propose some other methods of estimating the long-run variance Σ consistently. These are natural generalizations of two estimates proposed in Wu and Zhao (2007, [25]).
We first introduce a non-overlapping estimate of block means. For $1 \leq r < s \leq d$, $m = [n/G]$,
$A_{i,r} = \sum_{j=1}^{G} X_{j+iG,r}$,

$$
\begin{aligned}
\hat{\Sigma}^A_{r,r} &= \frac{G}{2u_{1/4}}\text{median}(|A_{i,r} - A_{i,r-1}|^2), \\
\hat{\Sigma}^A_{r,s} &= \frac{G}{2u_{1/4}}\text{median}((A_{i,r} - A_{i,r-1})(A_{i,r} - A_{i,r-1})), \\
\hat{\Sigma}^B_{r,r} &= \frac{G}{2(m-1)}(\sum_{i=1}^{m-1} |A_{i,r} - A_{i,r-1}|^2), \\
\hat{\Sigma}^B_{r,s} &= \frac{G}{2(m-1)}(\sum_{i=1}^{m-1} (A_{i,r} - A_{i,r-1})(A_{i,s} - A_{i,s-1})).
\end{aligned} \tag{3.6}
$$

Similar to Theorem 3 in Wu and Zhao (2007, [25]), we have the following result for a Lipschitz-continuous μ.

THEOREM 3.2. *1. If $G \asymp n^{5/8}$, then $|\Sigma^A - \Sigma| = O_{\mathbb{P}}(n^{-1/16}\log n)$.*
2. If $G \asymp n^{1/3}$, then $|\Sigma^B - \Sigma| = O_{\mathbb{P}}(n^{-1/3})$.
3. If $G \asymp n^{1/3}$, then $\mathbb{E}(|\Sigma^B - \Sigma|^2) = O(n^{-2/3})$.

4. Testing synchronization in high dimension. For the high-dimensional scenario where we keep the time horizon n fixed and allow the dimension d grow to ∞ the Gaussian approximation of the type (1.12) fails. Portnoy (1986) showed the central limit theorem fails if $d = n^{1/2}$. However, in today's world big data is often the rule than exception and many fields are increasingly demanding analyzing many datasets simultaneously. In this section we allow d to grow and often much faster than n but still be able to perform the hypothesis test of synchronization. For simplicity of exposition, as we mentioned before we assume that for the entire series only one changepoint has been detected.

4.1. *Problem description.* Consider the model in (1.1) in conjunction with the piecewise model specification in (1.2). Note that, for some of the series it is possible to have no changepoint i.e.

$$\mu_{r,1} = \mu_{r,2}$$

for some $1 \le r \le d$. Define

$$\mathbb{A} = \{1 \le r \le d : \mu_{r,1} \ne \mu_{r,2}\}.$$

We refine the hypothesis resting problem in (1.4) as follows: Let $\mathbb{A} = i_1, , i_l$

$$H_0 : \tau_{i,1} = \ldots = \tau_{i,l} \tag{4.1}$$

For the high-dimensional scenario, we can no longer work with the original proposal of $M_1 + \ldots_+ M_d - M_G$ for finite d.

Following some of the recent development in high dimensional statistics we will focus on the $\|.\|_\infty$ norm.

4.2. *Method.*

5. Examples. In this example, we give a few applications of our results. Specifically, we collate one result aimed towards building a general theoretical framework and another one motivated from a real-world application.

5.1. *Change-point in covariance process.* In this section, we discuss how we can extend our ideas to test for existence and synchronization of a covariance process. Let $X_i \in \mathbb{R}^d$ is a non-stationary sequence that has finite qth moment for some $q > 4$. Let the $d(d+1)/2$ dimensional vector $W_i = (X_{ir}X_{is})_{1\le r\le s\le d}$. Then $\bar{W}_n := \sum_{i=1}^n W_i/n$ gives sample contemporaneous covariances of $(X_i)_{i=1}^n$. We view W_i as

$$W_i = \mathbb{E}(W_i) + W_i - \mathbb{E}(W_i) = \mu_W(i/n) + e_W(i),$$

where μ_W and e_W play analogous role to μ and e described in the context of (1.1). This allows us to not only detect the change-points for the possible $d(d+1)/2$ contemporaneous covariances it allows us to test for their synchronization. From a practical purpose, if we do not observe the data itself and have restricted availability to only their covariance estimates this can help us recover any structural change happening in the covariance process which might lead to further investigation for some specific co-ordinates of the multiple time-series.

We need an optimal Gaussian approximation result for the covariance process. Karmakar and Wu (2017) proved in their paper the following result for the specific case of vector-linear process.

PROPOSITION 5.1. *Assume that X_i is a vector linear process*

$$X_i = \sum_{j=0}^{\infty} B_j \epsilon_{i-j}, \tag{5.1}$$

where B_j are $d\times d$ coefficient matrix, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{id})^T$, ϵ_{ir} are i.i.d. random variables with mean 0 and finite qth moment, $q > 4$. Moreover assume,

$$\sum_{j=t}^{\infty} |B_j| = O(t^{-\chi}), \chi \ge \chi_0 \tag{5.2}$$

where χ_0 is defined at (1.11). Write $p = q/2$. Let $\Sigma = \sum_{k=-\infty}^{\infty} Cov(W_0, W_k)$ be the long-run covariance matrix of (W_i). By Theorems 1 and 2 therein, we have

$$\max_{i\le n} |i\bar{W}_i - i\mathbb{E}(W_1) - \Sigma^{1/2} IB(i)| = o_{\mathbb{P}}(n^{1/p}), \tag{5.3}$$

where IB is a centered standard Brownian motion.

However, through careful inspection of their proof it is easy to extend them for a more general stationary sequence with proper functional dependence measure.

5.2. *Application: Network Anomaly Detection.* In this subsection we discuss two possible applications of our results. In the first of them, we use simulated data from different power grade system which closely emulate the real data and in the second one we compare pricing of oil and gas and test for synchronization within the data.

In communication networks, many anomaly detection problems can be treated as a change-point detection problem. Among many others, malicious attacks aiming to cause disruption of normal network operations are a kind of anomaly that requires immediate attention. Detection of such anomalies in the first time is very important, as further diagnosis and effective countermeasures following the detection are needed in order to restore the normal operation.

In the following, we present a case of jamming attack in a wireless network. A group of nodes as shown in Fig. 1 are communicating with each other: $0 \to 1$, $2 \to 3$, and $4 \to 5$. At some point, a jammer starts to send jamming signals into the wireless channel. Jamming signals are large blocks of continuous signals, often transmitted with high power to maximize the jamming effect. A normal wireless transmitter would perform carrier sense before transmitting to avoid interrupting the ongoing traffic. Once a transmitter finds the channel is busy, it can only wait until the channel is clear to transmit. However, the jammer would not follow the protocol and can jam the channel at any time. The presence of a jammer will directly cause the performance deterioration of normal nodes. The performance deterioration includes multi-folds: the packet queueing delay will increase as they wait for the channel to clear, and the throughput between a pair of transmitter-receiver will drop as fewer packets are successfully received. The detection of jamming attack is from the observation of these network performance measures. In this work, we use the delay and received packets interframe space (IFS) time series for change point detection. When the packet size is uniform, increased IFS is an indicator of decreased throughput. Samples are taken from a widely-used network simulator ns-3 [18], and the time series are extracted from the trace file.

When a jammer starts to jam the channel, all nodes hearing the jamming signal will be impacted. It is expected that change-points will appear in many time series. We put time series in two groups: same measurements from different nodes, and different measurements from the same node, and we perform change point detection on multiple time series within a group. Although the change points are results of the same causing event—- jamming, we cannot expect that the change points to occur at exactly the same time tick due to the sampling issue, therefore the change points are only loosely synchronized. Considering the sampling rate and node transmission rate, a tolerance band of λ_t is used such that change points occurring within λ_t of each other are considered synchronized.

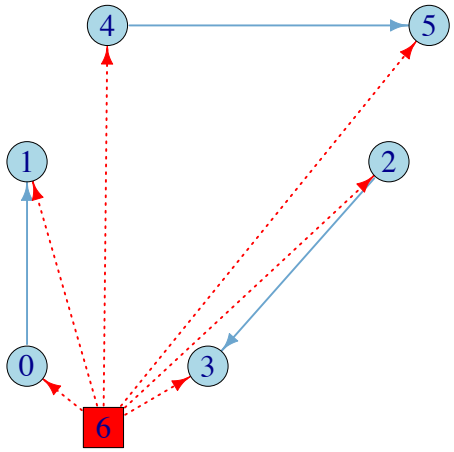


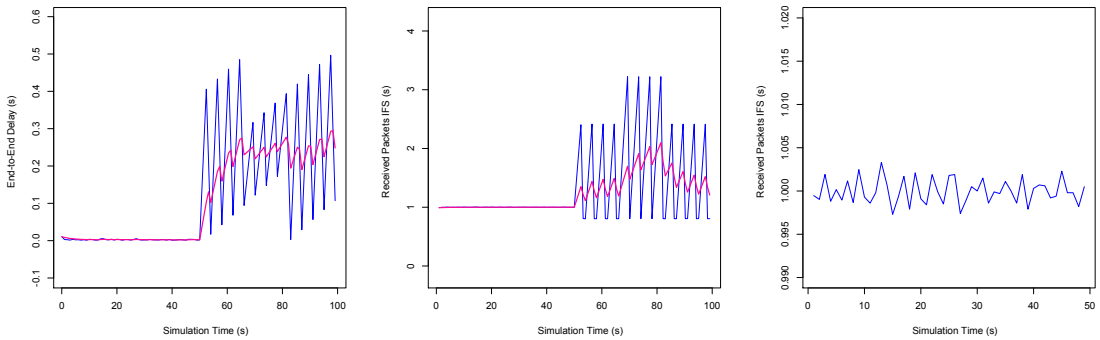
FIG 1. *network*

Fig. 2 shows the measurements from node 3. Other nodes show similar patterns. In Fig. 2 (a) and (b), the blue series shows the original sample values, and the pink series shows the smoothed values using Exponential Weighted Moving Average (EWMA). Using EWMA of samples is a commonly used technique in communication networks, since individual packets may experience large delay or large jitter even without any anomaly. This is mainly due to some bursty traffic and contention for channel access. Only when the increased delay or decreased throughput becomes a general trend for a sequence of packets, it is considered as a change point. Note that the delay and IFS before the attack are not uniform, but have very small variance. Picture (c) shows the microscopic IFS before the attack.

The simulation runs for 100 seconds, and the jamming attack starts at 50 second. The jamming attack will impact the packets that are sent but not received at this time, as well as the packets that are scheduled to send after this point. We apply the change point detection algorithm and synchronization testing algorithm on both the original samples and the smoothed series. Multiple time series include delay (series 1) and IFS (series 2). The results are summarized in table 1. The smoothed data reduce the number of false positives, but also incur additional detection delay.

TABLE 1
Detection of Jamming Attack

	Change Point Detected?	Detection Time $\hat{\tau}$ (s)		Synchronization	
Samples (Blue)	Yes	$\hat{\tau}_1$=52.406	$\hat{\tau}_2$=52.406	Yes	$\lambda_t = 0$
EWMA (Pink)	Yes	$\hat{\tau}_1$=54.017	$\hat{\tau}_2$=58.042	Yes	$\lambda_t = 3$



(a) (b) (c)

FIG 2. *(a) Delay, (b) Received packet IFS, (c) Microscopic IFS before the attack.*

6. Simulation results. For the simulated results, we stick to bivariate data. Moreover, we use the true long-run variance for constructing the Gaussian analogue of the observed data. Although we have Theorem 3.2 to ensure the consistency of the estimated covariance matrix, we use the true long-run variance to evaluate our methods. Moreover, the estimated block covariances are close to the true covariances and thus using the true long-run covariance does not affect our result much. Moreover, we use μ_r function to be 0 before the change-point appears since this problem is location invariant with respect to the range of μ.

6.1. *Choice of the error process.* We make following choices of the error process

- IID Normal (0,$\sigma^2 = 1$).,
- An AR(1) process which has no co-integration. The two co-ordinates evolve independently with AR coefficient 0.6 and -0.3 ,
- An VAR(1) process with the first order coefficient matrix being

$$\begin{bmatrix} 0.6 & -0.2 \\ 0.3 & -0.3 \end{bmatrix}$$

.

6.2. *Choice of G.* The condition (2.9) describes the rate of G and k_n for asymptotic consistency of our method. For practical implementation we choose $G = O(n^{1/3})$. However, all our results are asymptotic and since $O(n^{1/3})$ is free to change up to a constant factor, it leads to a challenge for a small sample size n.
The results presented here are for $n = 2000$ and $G = 100$. We show in Table 2 that the choice of $G = 50$ will not make a huge difference in the observed rejection probabilities

under null. Thus we decide to stick to $G = 100$ for the rest of the simulations presented here.

We perform our simulation in mainly three different set-ups.

1. No change-point exists.
2. Change-point exists and they are synchronized.
3. Change-points exists and they are asynchronized.

The first one is tabulated in Table 2 where we exhibit the performance for three different choice of error process and two different choices of window-size G. The second two set-ups are collectively tabulated in Tables 3 4 5 where $\lambda_\tau = 0$ stands for synchronized change-points.

TABLE 2

Performance for the case of null hypotheses. Table shows number of times change-point is not detected/ detected out of 500 iteration with n=2000, kn=100, error=Independent gaussian. Table shows no. of not detected, no. of detected, no. of synchronized, no. of asynchronized

	Gaussian	AR(1)	VAR(1)
G=100	474, 26, 19, 7	484, 16, 15, 1	483, 17, 11, 6
$G = 50$	473, 27, 13, 14	491, 9, 8, 1	492, 8, 5, 3

Remark A few remarks are in order.

- Note that, we see similar results with the two choices of G and hence we will use $G = 100$ for the rest of the simulations. The dependent case has coverage more than 95% in order to allow for the effect of dependence in the long-run covariance.
- For the Gaussian independent error Table 3, we only computed six cases as the effect of the change will be symmetric on the two co-ordinates.
- One can see, if $\lambda_1 = \lambda_2 = 0.5$, our algorithm is not able to detect existence of the change-points in a dependent case. The method still works pretty well for the independent Gaussian errors but does not if λ is as small as 0.5. However, if one of λ_1 and λ_2 is bigger it performs somewhat better.
- If one of λ_1 or λ_2 is as big as 1.2 or 2, the algorithm has 100 % success rate in detecting the change-point. Also note that, if λ_τ increase from 0.001 to 0.01 the ability to detect asynchronization increases. We emphasize that, for $n = 2000$, a distance of $\lambda_\tau = 0.001$ is a mere gap of only 2 index in the locations of the two change-points. This substantiates the accuracy of our method.
- Difference between AR(1) and VAR(1) cases.

7. Real Data Analysis.

7.1. *Oil, Gas and Dow pricing ratios.* As is usual with many pricing data, we will look at the logged and differenced data for this purpose to obtain a somewhat stationary

TABLE 3
Performance for the case $\tau_1 = / \neq \tau_2$ for Gaussian. Table shows no. of not detected, no. of detected, no. of synchronized, no. of asynchronized

	$\lambda_\tau = 0$	$\lambda_\tau = 0.001$	$\lambda_\tau = 0.002$	$\lambda_\tau = 0.005$	$\lambda_\tau = 0.01$
$\lambda = 0.5$	52, 448, 417, 31	69, 431, 397, 34	62, 438, 406, 32	80, 420, 383, 37	76, 424, 343, 81
$\lambda_1, \lambda_2 = (0.5, 1.2)$	0, 500, 468, 32	0, 500, 466, 34	0, 500, 459, 41	0, 500, 423, 77	0, 500, 348, 152
$\lambda_1, \lambda_2 = (0.5, 2)$	0, 500, 468, 32	0, 500, 465, 35	0, 500, 449, 51	0, 500, 408, 92	0, 500, 327, 173
$\lambda = 1.2$	0, 500, 478, 22	0, 500, 456, 44	0, 500, 414, 86	0, 500, 213, 287	0, 500, 63, 437
$\lambda_1, \lambda_2 = (1.2, 2)$	0, 500, 476, 24	0, 500, 447, 53	0, 500, 347, 153	0, 500, 126, 374	0, 500, 27, 473
$\lambda = 2$	0, 500, 476, 24	0, 500, 390, 110	0, 500, 193, 307	0, 500, 32, 468	0 500, 0, 500

TABLE 4
Performance for AR(1) innovations. Table shows no. of not detected, no. of detected, no. of synchronized, no. of asynchronized

	$\lambda_\tau = 0$	$\lambda_\tau = 0.001$	$\lambda_\tau = 0.002$	$\lambda_\tau = 0.005$	$\lambda_\tau = 0.01$
$\lambda = 0.5$	448, 52, 17, 35	449, 51, 18, 33	449, 51, 18, 33	447, 53, 17, 36	446, 54, 15, 39
$\lambda_1, \lambda_2 = (1.2, 0.5)$	183, 317, 253, 64	181, 319, 255, 64	181, 319, 250, 69	183, 317, 230, 87	190, 310, 185, 125
$\lambda_1, \lambda_2 = (0.5, 1.2)$	15, 485, 52, 433	19, 481, 50, 431	23, 477, 44, 433	23, 477, 37, 440	23, 477. 28, 449
$\lambda_1, \lambda_2 = (2, 0.5)$	3, 497, 459, 38	3, 497, 457, 40	3, 497, 451, 46	3, 497, 419, 78	3, 497, 319, 178
$\lambda_1, \lambda_2 = (0.5, 2)$	0, 500, 46, 454	0, 500, 39, 461	0, 500, 33, 467	0, 500, 23,477	0, 500, 18, 482
$\lambda = 1.2$	0, 500, 310, 190	1, 499, 220, 279	1, 499, 192, 307	1, 499, 112, 387	0, 500, 45, 455
$\lambda_1, \lambda_2 = (2, 1.2)$	0, 500, 358, 142	0, 500, 335, 165	0, 500, 279, 221	0, 500, 127, 373	0, 500, 36, 464
$\lambda_1, \lambda_2 = (1.2, 2)$	0, 500, 193, 307	0, 500, 162, 338	0, 500, 114, 386	0, 500, 63, 437	0, 500, 26, 474
$\lambda = 2$	1, 499, 230, 269	0, 500, 226, 274	0, 500, 136, 364	0, 500, 51, 449	0, 500, 12, 488

TABLE 5
Performance for VAR(1) innovations. Table shows no. of not detected, no. of detected, no. of synchronized, no. of asynchronized

	$\lambda_\tau = 0$	$\lambda_\tau = 0.001$	$\lambda_\tau = 0.002$	$\lambda_\tau = 0.005$	$\lambda_\tau = 0.01$
$\lambda = 0.5$	424, 76, 28, 48	425, 75, 26, 49	426, 74, 24, 50	426, 74, 19, 55	422, 78, 19, 59
$\lambda_1, \lambda_2 = (1.2, 0.5)$	134, 366, 260, 106	30, 470, 209, 261	30, 470, 205, 265	30, 470, 194, 276	30, 470, 169, 301
$\lambda_1, \lambda_2 = (0.5, 1.2)$	6, 494, 46, 448	33, 467, 208, 259	35, 465, 209, 256	36, 464, 200, 264	34, 466, 174, 292
$\lambda_1, \lambda_2 = (2, 0.5)$	1, 499, 361, 138	0, 500, 96, 404	0, 500, 93, 407	0, 500, 79, 421	0, 500, 57, 443
$\lambda_1, \lambda_2 = (0.5, 2)$	0, 500, 39, 461	0, 500, 92, 408	0, 500, 91, 409	0, 500, 79, 421	0, 500, 65, 435
$\lambda = 1.2$	1, 499, 207, 292	16, 484, 484, 0	16, 484, 458, 26	14, 486, 350, 136	13, 387, 236, 351
$\lambda_1, \lambda_2 = (2, 1.2)$	0, 500, 318, 182	0, 500, 342, 158	0, 500, 295, 205	0, 500, 183, 317	0, 500, 96, 404
$\lambda_1, \lambda_2 = (1.2, 2)$	0, 500, 176, 324	0, 500, 347, 153	0, 500, 309, 191	0, 500, 189, 311	34, 466, 174, 292
$\lambda = 2$	0, 500, 276, 224	0, 500, 416, 84	0, 500, 311, 189	0, 500, 143, 317	0, 500, 47, 453

distribution. Without this, usually price data show a unit root behavior and cannot be used to fit in our setting. This differenced and logged data are usually called the log-return data. We first plot the log-returns for both the series and then provide analysis by splitting the time horizon in two parts. In the first part, one can see the change-points are synchronized but in the second part the change points are not.

References.

[1] Probability and moment inequalities under dependence. *Statistica Sinica*, 23(3):1257–1272, 2013.
[2] Alexander Aue, Siegfried Hörmann, Lajos Horváth, and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 2009.
[3] Jushan Bai. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.*, 15(5):453–472, 1994.
[4] Jushan Bai. Estimating multiple breaks one at a time. *Econometric Theory*, 13(3):315–352, 1997.
[5] István Berkes, Weidong Liu, and Wei Biao Wu. Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 2014.
[6] Z. W. Birnbaum and Albert W. Marshall. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.*, 32:687–703, 1961.
[7] M. Csörg˝ o and P. Révész. *Strong approximations in probability and statistics*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.
[8] J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1996.
[9] Irina Grabovsky, Lajos Horváth, and Marie Hu˘ sková. Limit theorems for kernel-type estimators for the time of change. *J. Statist. Plann. Inference*, 89(1-2):25–56, 2000.
[10] Jan J. J. Groen, George Kapetanios, and Simon Price. Multivariate methods for monitoring structural change. *J. Appl. Econometrics*, 28(2):250–274, 2013.
[11] J. Hájek and A. Rényi. Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.*, 6:281–283, 1955.
[12] D. V. Hinkley. Inference about the change-point from cumulative sum tests. *Biometrika*, 58:509–523, 1971.
[13] David Hinkley and Edna Schechtman. Conditional bootstrap methods in the mean-shift model. *Biometrika*, 74(1):85–93, 1987.
[14] Yannick Hoga. Monitoring multivariate time series. *J. Multivariate Anal.*, 155:105–121, 2017.
[15] Lajos Horváth and Marie Hu˘ sková. Change-point detection in panel data. *J. Time Series Anal.*, 33(4):631–648, 2012.
[16] M. Lavielle and G. Teyssière. Détection de ruptures multiples dans des séries temporelles multivariées. *Liet. Mat. Rink.*, 46(3):351–376, 2006.
[17] Weidong Liu and Zhengyan Lin. Strong approximation for a class of stationary processes. *Stochastic Process. Appl.*, 119(1):249–280, 2009.
[18] ns3 project. Network simulator, 2016. https://www.nsnam.org.
[19] Hernando Ombao, Jonathan Raz, Rainer von Sachs, and Wensheng Guo. The SLEX model of a non-stationary random process. *Ann. Inst. Statist. Math.*, 54(1):171–200, 2002. Nonparametric approach to time series analysis (Nara, 2000).
[20] E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–527, 1955.
[21] E. S. PAGE. On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1-2):248, 1957.
[22] A. N. Pettitt. A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, 67(1):79–84, 1980.
[23] Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005.
[24] Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154, 2005.
[25] 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.
[26] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(1):49–67, 2006.

8. Proofs.

PROOF OF PROPOSITION 2.1. Let $S_i = \sum_{j=1}^{i} e_i$. Using a Rosenthal-type inequality from Liu, Xiao and Wu (2013, [1]), we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{m\le i\le n}|S_i|/i \ge x\right) &\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} \mathbb{P}\left(\sup_{2^r\le i\le 2^{r+1}-1}|S_i|\ge 2^r x\right) \\
&\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} \mathbb{P}\left(\sup_{1\le i\le 2^{r+1}-1}|S_i|\ge 2^r x\right) \\
&\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} \frac{1}{2^{2r}x^2}2^{r+1} = O\left(\frac{1}{mx^2}\right). \qquad (8.1)
\end{aligned}$$

□

PROOF OF THEOREM 2.2. Without loss of generality, assume $k > k_0$. Note that,

$$\mathbb{P}\Big(\sup_{k:|k-k_0|>M/\lambda^2}|V_k|\ge|V_{k_0}|\Big) = \mathbb{P}\Big(\sup_{k:|k-k_0|>M/\lambda^2}V_k+V_{k_0}\le 0\Big)+\mathbb{P}\Big(\sup_{k:|k-k_0|>M/\lambda^2}V_k-V_{k_0}\ge 0\Big) \qquad (8.2)$$

For the first term, we proceed as follows. Since we assumed $\lambda > 0$, $\mathbb{E}(V_k) > 0$ for all k

$$\mathbb{P}\Big(\sup_{k:|k-k_0|>M/\lambda^2}V_k+V_{k_0}\ge 0\Big) \le \mathbb{P}\Big(\sup_{k:|k-k_0|>M/\lambda^2}V_k-\mathbb{E}(V_k)+V_{k_0}-\mathbb{E}(V_{k_0})\ge -\mathbb{E}(V_{k_0})\Big).$$

It follows from conditions 2.9 that

$$P(\sup_k|X_k-Y_k|>\lambda/4)\to 0.$$

We finish the proof by showing that,

$$\lim_{M\to\infty}\limsup_{n\to\infty}\Big\{\sup_{M\lambda^2\le|k-k_0|}(V_k-V_{k_0})>-c\Big\}=0. \qquad (8.3)$$

We decompose $V_k - V_{k_0} - \mathbb{E}(V_k - V_{k_0})$ in three parts. For $k_0 < k \le k_0 + G$,

$$\begin{aligned}
V_k - V_{k_0} - (\mathbb{E}(V_k) - \mathbb{E}(V_{k_0})) &= M_{1,k} + M_{2,k} + M_{3,k} \\
&= \sum_{i=k_0-G+1}^{k-G} e_i - 2\sum_{i=k_0+1}^{k} e_i + \sum_{i=k_0+G+1}^{k+G} e_i. \qquad (8.4)
\end{aligned}$$

Since, $\mathbb{E}(V_k) - \mathbb{E}(V_{k_0}) = \lambda \min(1, |k - k_0|/G)$, we show it for the three parts from the decomposition of $V_k - V_{k_0} - \mathbb{E}(V_k - V_{k_0})$ that (8.4), i.e.

$$\mathbb{P}(\sup_{|k-k_0|>M/\lambda^2} |\frac{1}{k-k_0} \sum_{k_0-G<i\leq k-G} e_i + \lambda/G| > -\frac{c}{k-k_0}) \quad \to \quad 0,$$

$$\mathbb{P}(\sup_{|k-k_0|>M/\lambda^2} |\frac{1}{k-k_0} \sum_{k_0<i\leq k} e_i + \lambda/G| > -\frac{c}{k-k_0}) \quad \to \quad 0,$$

$$\mathbb{P}(\sup_{|k-k_0|>M/\lambda^2} |\frac{1}{k-k_0} \sum_{k_0+G<i\leq k+G} e_i + \lambda/G| > -\frac{c}{k-k_0}) \quad \to \quad 0.$$

Then (8.3) follows from Result 2.1.

□

Proof of Theorem 2.3. We will first prove that in order to find the distribution of τ, it suffices to consider the quantity $G\lambda(V_k - V_{k_0}) = G\lambda((V_k - \mathbb{E}(V_k)) - (V_{k_0} - \mathbb{E}(V_{k_0})) + G\lambda(\mathbb{E}(V_k) - \mathbb{E}(V_{k_0}))$ in the set $\{k : |k - k_0| \leq Mn\lambda^{-2}\}$.

Write $k = k_0 + \lfloor v\lambda^{-2} \rfloor$. Then $G\lambda(\mathbb{E}(V_k) - \mathbb{E}(V_{k_0})) = |v|$. For the stochastic part, we use the invariance principle from Berkes, Liu and Wu (2014, [5]) and the decomposition in (8.4) to obtain, for $k > k_0$

$$V_k - V_{k_0} = V_k^Z - V_{k_0}^Z + o_{\mathbb{P}}(n^{1/p}) + cW_1(t).$$

A similar result holds for $k < k_0$. □

Proof of Theorem **??**. □

Department of Statistics
University of Florida
Griffin-Floyd Hall
Gainesville, FL 32611, USA.
E-mail: sayarkarmakar@ufl.edu

Illinois Institute of technology
E-mail: chengm@mst.edu

Department of Statistics
University of Chicago
5747 S Ellis Ave
Chicago, IL 60615, USA.
E-mail: wbwu@galton.uchicago.edu