Change-Point Detection and Synchronization Testing in Multiple Time-Series

Sayar Karmakar Maggie Cheng Wei Biao Wu

Abstract

In this paper, we investigate the problem of detecting a change-point in a multiple time-series. 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.

Keywords: Change-point, Local Linear Average, Gaussian Approximation, Gaussian Multiplier Resampling

I. 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, [15] 1957, [16]) 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, [8]) and Pettitt (1980, [17]). A bootstrapping idea was suggested by Hinkley and Schechtman, (1987, [9]). These are some of the seminal works towards detecting one change in a single time-series.

Sayar Karmakar and Wei Biao Wu are with the University of Chicago, Department of Statistics; Maggie Cheng is with New Jersey Institute of Technology.

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, [14]) employed the SLEX (smooth localized complex exponentials) whereas Lavielle and Teyssi'ere (2006, [12]) 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, [20]). CUSUM-type statistics have been widely used in time series segmentation. In the context of multivariate time series segmentation, Groen et al. (2013, [6]) and Horváth, Lajos and Hušková, Marie (2012, [11]) 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, [1]), detection of a single change-point in the covariance structure of multivariate time-series using a CUSUM statistic.

Very recently Hoga(2016, [10]) used a Gaussian approximation result by Liu and Lin (2009, [13]) to study asymptotic properties of the CUSUM procedure for a stationary multiple time-series. However, the Gaussian approximation result from [13] 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. We assume a constant signal, with a possible jump-point for each co-ordinate and add a multivariate error process to the vector of signals. 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, [18])'s framework of functional dependence measure are imposed. We show in Section V 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 estimators, little has been done to develop an inferential theory in this direction.

Our methods in this paper are aimed for stationary processes to maintain simplicity. However, departure from stationarity is also not uncommon in real life. One can assume a simple piece-wise continuous function for the autocovariance and the cross-covariance for the error process which makes the corresponding time-series roughly stationary within the change-points. Since the Gaussian approximation mentioned in [] is applicable for even more general class of non-stationary time-series, we can naturally extend our methods there. We discuss this generalization as a possible variant of our methods at the end.

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 II. Section III 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 IV we validate our detection methods with some theoretical results. We study an analogous problem in exploring contemporaneous covariance process in Section V. Section VI is used to describe a Gaussian block multiplier technique to practically implement this. Network anomaly detection is discussed in Section IX as a key application of our study. Section VII shows some data simulations to substantiate our theoretical methods.

II. Mathematical Models

Consider the multiple series sequence model

$$\mathbf{X}_i = \mu(i/n) + \mathbf{e}_i = (\mu_1(i), \ldots, \mu_d(i))^T + \mathbf{e}_i, \tag{2.1}$$

where $(\mathbf{e})_{i\in\mathbb{N}}$ denotes the unobserved error process and $\mu_r(i)$ is the signal for the rth series at time index i. We impose some simple structure on μ. For each co-ordinate r, let μ_r be 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 $1 \leq r \leq d$ we denote $0 \leq \tau_r \leq 1$ to be the change-point and let λ_r denote the jump length at the point τ_r. Mathematically speaking, say the r th co-ordinate function μ_r takes the form

$$\mu_r(x) = \begin{cases} \mu_{r1}, & \text{if } x < \tau_r \\ \mu_{r2}, & \text{if } x \geq \tau_r \end{cases} \tag{2.2}$$

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.

$$H_{10} : \lambda_1 = \ldots = \lambda_d = 0. \tag{2.3}$$

Provided we reject (2.3), we will test the following hypothesis

$$H_{20} : \tau_1 = \ldots = \tau_d. \tag{2.4}$$

Karmakar and Wu (2017) 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

$$\max_{1\leq i\leq n} |S_i' - G_i| = o(\alpha_n). \tag{2.5}$$

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 (2.4) 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.

III. Method

A. *Detection*

Instead of using the traditional CUSUM technique to detect the change point, in this paper we will instead use a MOSUM local linear average. CUSUM technique suffer from huge computational burden for updating.

For the r th co-ordinate, define $V_{i,r}$ as follows,

$$V_{i,r} = |\frac{1}{G}\sum_{j=i-G+1}^{i} X_{j,r} - \frac{1}{G}\sum_{j=i+1}^{i+G} X_{j,r}|.$$

let

$$M_r = \max_{G\le i\le n-G} |V_{i,r}|^2 \tag{3.1}$$

and the maximizer

$$\hat{k}_r = \arg\max_{G\le i\le n-G} |V_i|, \quad \hat{\tau}_r = \hat{k}_r/n. \tag{3.2}$$

We will show that $\hat{\tau}_r$ is consistent for τ and obtain a rate of convergence for this estimation. Using that rate of convergence, one can obtain an asymptotic null distribution of $\hat{\tau}_r$. Note that, results of similar flavor for the independent errors has been established in the literature for long. However, not much has been done for dependent errors and thus we decide to provide such results in this paper.

Under the null H_{10} that $\lambda_r = 0$ for all r, i.e there is no jump in any of the functions μ_i, one should get a small value of $M_1 + M_2 + \ldots + M_d$. For testing H_{10}, we will use $M_1 + M_2 + \ldots + M_d$ as our test statistic. We will obtain the asymptotic null distribution of M_r and $M_1 + \ldots + M_d$ in the next section.

B. Synchronization

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 first define the synchronized maximizer.

Let

$$M_G = \max_{1 \le i \le n} \sum_{r=1}^{d} |\frac{1}{G} \sum_{j=i-G+1}^{i} X_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} X_{j,r}|^2 \tag{3.3}$$

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 (IV-C) 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.

IV. Main Results

We introduce some notations first. After that, we will discuss some general assumptions and then go to our main theorem and corollaries.

A. Notations and General Assumptions:

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 $\|Y\|_p := [E(|Y|^p)]^{1/p} < \infty$. For $\mathcal{L}_2$ norm write $\|\cdot\| = \|\cdot\|_2$. Define the projection operator P_i by

$$P_i Y = E(Y|\mathcal{F}_i) - E(Y|\mathcal{F}_{i-1}), \qquad Y \in \mathcal{L}_1. \tag{4.1}$$

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.

B. The error process

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).

$$e_i = H(\epsilon_i, \epsilon_{i-1}, \dots,), \tag{4.2}$$

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, [18]) to model the dependence structure. Consider the process in (4.2). 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, [18]] 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

$$\lambda_{i,p} = \|e_i - e_{i,0}\|_p = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,0}\|_p, \tag{4.3}$$

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',

$$\mathcal{F}_{i,k} = (\epsilon_i, \epsilon_{i-1}, \dots, \epsilon_k', \epsilon_{k-1}, \dots,) \tag{4.4}$$

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, [18]) 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,

$$\Theta_{0,p} = \sum_{i=m}^{\infty} \lambda_{i,p} < \infty. \tag{4.5}$$

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 $\Theta_{i,p}$

C. Main Results

From Berkes et al. (2014, [4]), 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

$$\chi > \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p}, \tag{4.6}$$

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

$$\max_{i \leq n} |S_i^c - G_i| = o_P(n^{1/p}) \qquad \text{in } (\Omega_c, A_c, P_c) \tag{4.7}$$

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

Karmakar and Wu (2017, []) 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 (IV-C) holds with $\Sigma = \sum_{k=0}^{\infty} E(e_0 e_k^T)$ being the corresponding long-run covariance matrix of the vector valued (e_i) process.

We first develop a Hàjek-Rènyi type inequality for stationary sequences with the representation (4.2). 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,[7]) obtained similar result for independent random variables, Birnbaum and Marshall ([5]) extended it to martingales. Bai (1994, [2]) and Bai (1997, [3]) generalized this to linear process and mixingales respectively. However, based on some recent works in the framework of Wu (2005, [**?**], it is possible to extend this to a more general stationary processes.

Proposition 4.1. *Assume e_i admits the causal representation in (4.2) with the long-run covariance σ. Then,*

$$Pr(\max_{m\le k\le n}\frac{1}{k}|\sum_{i=1}^{k} e_i| > \alpha) \le C\frac{\sigma^2}{m\alpha^2}. \tag{4.8}$$

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

$$\begin{aligned} P(\sup_{m\le i\le n}|S_i|/i \ge x) &\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} P(\sup_{2^r\le i\le 2^{r+1}-1}|S_i| \ge 2^r x) \\ &\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} \frac{1}{2^{2r}x^2} P(\sup_{1\le i\le 2^{r+1}-1}|S_i| \ge 2^r x) \\ &\le \sum_{\lfloor \log_2 m\rfloor \le r\le \lceil \log_2 n\rceil} \frac{1}{2^{2r}x^2} 2^{r+1} = O(\frac{1}{mx^2}). \end{aligned} \tag{4.9}$$

Next we discuss consistent estimation of the change-points for one series. For this part, we suppress the suffix of $\tau_r = k_r/n$ and $\hat{\tau}_r = \hat{k}_r/n$. Since we are focusing on only one series now, we assume that the jump-length is $\lambda \to 0$.

First we show that the estimated change-point cannot be too far from the true change-point. It is easy to see that $P(|\hat{k} - k_0| = o_P(G)$

Theorem 4.2. *Assume λ and G satisfy the following*

$$\frac{n^{1/p}}{\lambda_n G} \to 0, \frac{\log\log n + \log(n/G)}{\lambda_n \sqrt{G}} \to 0. \tag{4.10}$$

$$\hat{\tau} - \tau \to 0. \tag{4.11}$$

Lemma 4.3. *An alternative objective function for estimating the change point location is as follows*

$$\arg \max_{|k-k_0| \leq G} |V_k| = \arg \max_{|k-k_0| \leq G} V_k - V_{k_0}. \tag{4.12}$$

In the next theorem, we provide an improved rate for the convergence of $\hat{\tau}$. This improved rate is necessary to perform inferential asymptotics of the distribution of $\hat{\tau}$

Theorem 4.4. *Assume* $\to 0$

$$\hat{\tau} - \tau = O_p(n^{-1}\lambda^{-2}). \tag{4.13}$$

Define two-sided Brownian motion $W(t)$ as follows

$$W(t) = \begin{cases} W_1(t) & \text{if } t \geq 0, \\ -W_2(t) & \text{otherwise,} \end{cases} \tag{4.14}$$

where W_1 and W_2 are independent Brownian motion on $(0, \infty)$

Theorem 4.5. *Assume λ_n and G satisfy the following*

$$\frac{n^{1/p}}{\lambda_n G} \to 0, \frac{\log\log n + \log(n/G)}{\lambda_n \sqrt{G}} \to 0. \tag{4.15}$$

Then

$$c_n(\hat{\tau} - \tau) \xrightarrow{D} \arg\max(\sigma W(t) - |t|), \tag{4.16}$$

where $c_n =$.

In the following theorem, we show that each individually detected change-point is close to the true change-point under suitable conditions. Refer to the definition of $\hat{\tau}_r$ from 3.2.

Theorem 4.6. *For* $1 \leq r \leq d$*, we assume that the* r*th co-ordinate function* μ_r *has a change-point at* τ_r *with the jump* $\lambda_r > 0$*. Then if* G *satisfy*

$$\frac{1}{\sqrt{G \log n}} + \frac{n^{1/p}}{G} \to 0 \tag{4.17}$$

then

$$\hat{\tau}_r - \tau_r = o_P(G/n)$$

Proof. Let us assume that $k_0/n \leq \tau_r < (k_0 + 1)/n$. Note that, $|\hat{\tau}_r - \tau_r| > G/n$ can only happen if $\hat{\tau}_r < k_0 - G$ or $\hat{\tau}_r > k_0 + G$. From (4.7), note that if $\hat{\tau}_r < k_0 - G$ or $\hat{\tau}_r > k_0 + G$, we have

$$\begin{aligned}
\sqrt{M_r} &= \frac{1}{\sqrt{G \log n}} \frac{\sqrt{\log n}}{\sqrt{G}} \{ \max_{G \leq k \leq n-G} |2G_k - G_{k+G} - G_{k-G}| + o_P(n^{1/p}) \} \qquad (4.18) \\
&= \frac{1}{\sqrt{G \log n}} \{ O_P(1) + \frac{\sqrt{\log n}}{\sqrt{G}} o_P(n^{1/p}) \} \\
&= O_P(\frac{1}{\sqrt{G \log n}}) + o_P(n^{1/p}/G) = o_p(1).
\end{aligned}$$

whereas using the same logic if $|\hat{\tau}_r - \tau_r| \leq 1/n$, then

$$\sqrt{M_r} = \lambda_r + o_p(1),$$

due to the form of μ_r in (2.2). □

The next theorem concerns consistently estimating the jump-length at these change-points. If the jump length λ_r is positive then

Theorem 4.7. *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 (4.17) then*

$$\sqrt{M_r} - \lambda_r = o_P(1).$$

Proof. Assume μ_r has a true change-point at $k_0/n < \tau_r \leq (k_0+1)/n$. Then we have,

$$\arg \max_{k_0 - G \leq k \leq k_0 + G} |2S_k - S_{k+G} - S_{k-G}| = \arg \max_{k_0 - G \leq k \leq k_0 + G} |2S_k - S_{k+G} - S_{k-G}| + o_P(n^{1/p}),$$

using the form of μ_r in (2.2), the gaussian invariance principle in (4.7) and the property of gaussian random variables as done in (4.18).

□

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

Theorem 4.8. *If G satisfies (4.15), we have,*

$$\sqrt{2\log(n/G)}\left(\sqrt{G}\sigma^{-1}\sqrt{M_r} - [2\log(n/G) + \frac{1}{2}\log\{\log(n/G)\}] - \log(3)\right) \xrightarrow{D} V \quad (4.19)$$

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

Remark: One can easily obtain the asymptotic distribution of M_r from Theorem 4.8. 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 4.9.

$$\sqrt{2\log(n/G)}\left(\sqrt{\frac{G}{\phi_0}}\sqrt{M_G} - B_d\left(\frac{n}{G}\right)\right) \xrightarrow{D} V, \tag{4.20}$$

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

V. 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 $W_i = (X_{ir}X_{is})_{1\leq r\leq s\leq d}$. 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

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

where μ_W and e_W play analogous role to μ and e described in the context of (2.1). 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,*

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

where χ_0 *is defined at (4.6). 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* **??** *and* **??**, *we have*

$$\max_{i \leq n} |i\bar{W}_i - iE(W_1) - \Sigma^{1/2} IB(i)| = o_P(n^{1/p}), \tag{5.3}$$

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.

VI. 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.

A. Algorithm for detection and synchronization

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 $1 \leq r \leq d$,

$$M_r^Z = \max_{1 \leq i \leq n} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2, \tag{6.1}$$

$$\hat{\tau}_r^Z = \arg\max_{1 \leq i \leq n} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2. \tag{6.2}$$

Let

$$M_G^Z = \max_{1\le i\le n} \sum_{r=1}^{d} |\frac{1}{G} \sum_{j=i-G+1}^{i} g_{j,r} - \frac{1}{G} \sum_{j=i+1}^{i+G} g_{j,r}|^2. \tag{6.3}$$

Theorem 6.1. *For any* $1 \le r \le d$ *we have*

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

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

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

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

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 (3.1) and (3.3) respectively.
2. Use the input data X_{ij} for estimation of $\hat{\tau}_r$, as in (3.2). 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 (6.5).

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 (6.1) and (6.3). 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 $G = n^{1/3}$.

B. 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 (IV-C). 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 (3.2). 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 $\hat{\Sigma}$.

$$\hat{\Sigma} = \frac{1}{G(n-G+1)} \sum_{j=0}^{n-G} (S^*_{j+G} - S^*_j)(S^*_{j+G} - S^*_j)^T, \tag{6.4}$$

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, []). We first introduce a non-overlapping estimate of block means. For $1 \leq r < s \leq d$, $m = [n/G]$,

$A_{i,r} = \sum_{j=1}^{G} X_{j+iG,r}$,

$$\begin{aligned}
\hat{\Sigma}^A_{r,r} &= \frac{G}{2u_{1/4}} \text{median}(|A_{i,r} - A_{i,r-1}|^2), \\
\hat{\Sigma}^A_{r,s} &= \frac{G}{2u_{1/4}} \text{median}((A_{i,r} - A_{i,r-1})(A_{i,r} - A_{i,r-1})), \\
\hat{\Sigma}^B_{r,r} &= \frac{G}{2(m-1)} (\sum_{i=1}^{m-1} |A_{i,r} - A_{i,r-1}|^2), \\
\hat{\Sigma}^B_{r,s} &= \frac{G}{2(m-1)} (\sum_{i=1}^{m-1} (A_{i,r} - A_{i,r-1})(A_{i,s} - A_{i,s-1})).
\end{aligned} \tag{6.5}$$

Similar to Theorem 3 in Wu and Zhao (2007, [19]), we have the following result for a Lipschitz-continuous μ.

Theorem 6.2. *1) If $G \asymp n^{5/8}$, then $|\Sigma^A - \Sigma| = O_P(n^{-1/16} \log n)$.*
2) If $G \asymp n^{1/3}$, then $|\Sigma^B - \Sigma| = O_P(n^{-1/3})$.
3) If $G \asymp n^{1/3}$, then $E(|\Sigma^B - \Sigma|^2) = O(n^{-2/3})$.

VII. 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 6.2 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 invriant with respect to the range of μ.

A. *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

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

.

B. *Choice of G*

The condition (4.17) describes the rate of G 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 I 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 I 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 **??** II III where $\lambda_\tau = 0$ stands for synchronized change-points.

TABLE I

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 **??**, 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.

TABLE II

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

VIII. EXTENSIONS AND VARIANTS

A. *Piecewise continuous signal*

The procedure of estimating the change-point can be further generalized by using a symmetric kernel-based objective function

$$\hat{\tau}_r = \frac{1}{n} \arg \max_{G \le i \le n-G} | \sum_{j=i-G+1}^{i} K(\frac{j-i}{nb_n}) X_{j,r} - \sum_{j=i+1}^{i+G} K(\frac{j-i}{nb_n}) X_{j,r}|^2, \tag{8.1}$$

for a suitable choice of the bandwidth b_n. Note that, the change-point estimates proposed in this paper can also be viewed as an estimator of the type (8.1) with a uniform/rectangular kernel. For a general symmetric kernel, one can capture a larger class of μ functions.

TABLE III

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 IV

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

It is well-known in literature that the Pristley Chao estimate suffers from boundary problem. Since our estimate at (8.1) is related to the same, we use local linear correction (See Fan and Gijbels (1996, []) to detect the change-points consistently. Let

$$w_n(t,i) = K(\frac{t-i/n}{b_n})\frac{S_2(t)-(t-i/n)S_1(t)}{S_2(t)S_0(t)-S_1(t)^2},$$

where

$$S_j(t) = \sum_{i=1}^{n}(t-i/n)^j K(\frac{t-i/n}{b_n}),$$

and the automatic bandwidth selector of Ruppert et al. (1995, []).

B. Non-stationary errors

In this paper, we discussed all the results with stationary error innovations. However, since the Gaussian approximation discussed in [] is suitable for non-stationary processes too, one can naturally extend all the results in this paper to this more general case. Define the functional dependence measure (cf. (4.3)) in the following uniform manner

$$\delta_{j,p} = \sup_i \|e_i - e_{i,(i-j)}\|_p = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)}\|_p. \tag{8.2}$$

All the proofs go through with some obvious modifications which we skip here for brevity.

C. Sequential monitoring

It is also possible to adopt a sequential monitoring procedure that can be computationally much easier in case there is a humongous amount of data. This mostly is related to the fact that for stationary error process, the long run covariance estimate does not change too much due to inclusion of small number of new entries.

In practice, one can store the estimated covariance matrix Σ upto a certain time-point and not recompute it based on arrival of every new data point. This can lead to a much faster detection of existence and synchronization as the cut off values can be stored a priori.

D. More than one change point

Consider the situation where there are potentially more than one change point in one of the coordinate series. Two ways are there to detect them for a single series

1. wild binary segmentation: Fryzlewicz

2. Fix a pre-determinded cutoff t. Look at the top t highest values of max simultaneously with the constraint that the break points are separated by some amount.

IX. 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.

A. Network signal

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: $0 \rightarrow 1$, $2 \rightarrow 3$, and $4 \rightarrow 5$. 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 [**?**], and the time series are extracted from the trace file.

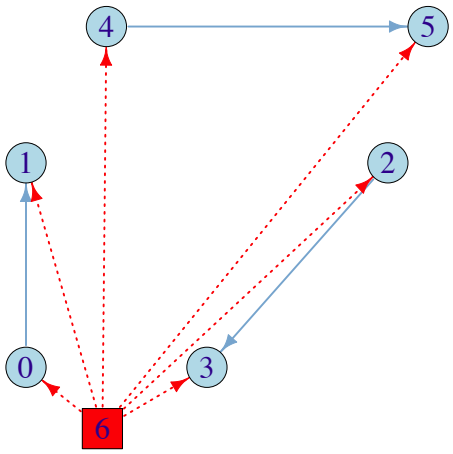
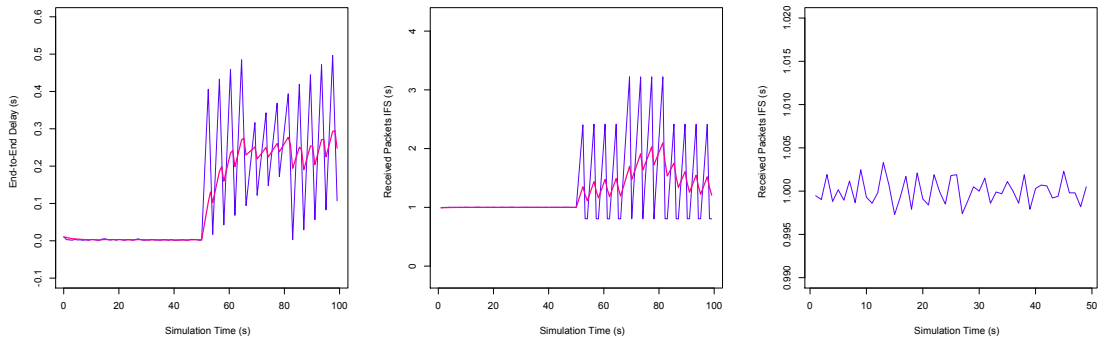


Fig. 1. network

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. 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.



(a) (b) (c)

Fig. 2. (a) Delay, (b) Received packet IFS, (c) 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 V. The smoothed data reduce the number of false positives, but also incur additional detection delay.

TABLE V

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	$\lambda_t = 3$

B. 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 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.

split the data in separate parts.. in one part synchronized.. in the other not synchronized.

Data plots

Splitting scheme

Synchronized change-points

Asynchronized change-points

Remarks

X. PROOFS

Proof of Lemma 4.3. Proof of this part is similar to the results proved in Lemma 2.3, 2.4 and 2.5 in **?**]. But one can have a much shorter proof of the arguments presented there since we have a simpler MOSUM method. For completeness we present the short proof here. Let $Q_{k,1} = V_k - E(V_k)$ and $Q_{k_2} = E(V_k)$. First note that

$$\begin{aligned} \arg \max_{|k-k_0|\leq G} |V_k| &= \arg \max_{|k-k_0|\leq G} V_k^2 - V_{k_0}^2 \\ &= \arg \max_{|k-k_0|\leq G} \sum_{i=1}^{7} R_{k,i}, \end{aligned} \tag{10.1}$$

where $R_{k,1} = (Q_{k,1} - Q_{k_0,1})^2$, $R_{k,2} = 2Q_{k_0,1}(Q_{k,1} - Q_{k_0,1})$, $R_{k,3} = 2(Q_{k,1} - Q_{k_0,1})(Q_{k,2} - Q_{k_0,2})$, $R_{k,4} = 2Q_{k_0,2}(Q_{k,2} - Q_{k_0,2})$, $R_{k,5} = (Q_{k,2} - Q_{k_0,2})^2$, $R_{k,6} = 2Q_{k_0,2}(Q_{k,1} - Q_{k_0,1})$, $R_{k,7} = 2Q_{k,2}(Q_{k,2} - Q_{k_0,2})$. One can show according to similar derivations in Grabovsky, Horváth and Hušková (2000, [**?**]), that (10.1) can be further reduced to

$$\arg \max_{|k-k_0|\leq G} |V_k| = \arg \max_{|k-k_0|\leq G} (R_{k,6} + R_{k,7}).$$

□

Proof of Theorem 4.2. Recall $V_{i,r}$ from (3.2). As mentioned before, we skip the second suffix and call it simply V_i for convenience. Using the invariance principle from Berkes, Liu and Wu ([**?**]) and Theorem 1.2.1 from Csörgö and Révész ([**?**]), we have

$$\begin{aligned} \sup_k |V_k - E(V_k)| &= o_p(n^{1/p}) + \sup_{0\leq t\leq n-G} \sup_{0\leq s\leq G} |W(t+s) - W(t)| \\ &= O_p(n^{1/p} + G\sqrt{\log(n/G)} + G^{1/2}\sqrt{\log\log n}). \end{aligned}$$

First note that, from condition 4.10, we have $P(|k - k_0| > G) = 0$ Also,

$$E(V_k) - E(V_{k_0}) = \lambda \min(1, \frac{k - k_0}{G}).$$

In the set $\{k : |k - k_0| \leq G\}$,

$$
\begin{aligned}
|V_k| - |V_{k_0}| &\leq 2\sup_k |V_k - E(V_k)| + |E(V_k)| - |E(V_{k_0})| \qquad (10.2)\\
|\hat{k} - k_0| &\leq 2\lambda^{-1} G \sup_k |V_k - E(V_k)|.
\end{aligned}
$$

This completes the proof in the view of (4.10).

□

Proof of Theorem 4.4. We decompose $V_k - V_{k_0} - E(V_k - V_{k_0})$ in 3 parts.

For $k_0 < k \leq k_0 + G$,

$$
\begin{aligned}
V_k - V_{k_0} - (E(V_k) - E(V_{k_0})) &= M_{1,k} + M_{2,k} + M_{3,k}\\
&= \sum_{i=k_0-G+1}^{k-G} e_i - 2\sum_{i=k_0+1}^{k} e_i + \sum_{i=k_0+G+1}^{k+G} e_i. \qquad (10.3)
\end{aligned}
$$

We show that,

$$
\lim_{M\to\infty} \limsup_{n\to\infty} \{ \sup_{M\lambda^2 \leq |k-k_0|} (V_k - V_{k_0}) > -c \} = 0. \qquad (10.4)
$$

Since, $E(V_k) - E(V_{k_0}) = \lambda \min(1, |k - k_0|/G)$, we show it for the three parts from the decomposition of $V_k - V_{k_0} - E(V_k - V_{k_0})$ that (10.3), i.e.

$$
\begin{aligned}
P(\sup_{|k-k_0|>b_n} |\frac{1}{k-k_0} \sum_{k_0-G<i\leq k-G} e_i + \lambda_n/G| > -\frac{c}{k-k_0}) &\to 0,\\
P(\sup_{|k-k_0|>b_n} |\frac{1}{k-k_0} \sum_{k_0<i\leq k} e_i + \lambda_n/G| > -\frac{c}{k-k_0}) &\to 0,\\
P(\sup_{|k-k_0|>b_n} |\frac{1}{k-k_0} \sum_{k_0+G<i\leq k+G} e_i + \lambda_n/G| > -\frac{c}{k-k_0}) &\to 0.
\end{aligned}
$$

Then (10.4) follows from Result 4.1.

□

Proof of Theorem 4.5. We use Lemma 4.3 to look at the stochastic behaviour of $G\lambda(V_k - V_{k_0}) = G\lambda((V_k - E(V_k)) - (V_{k_0} - E(V_{k_0})) + G\lambda(E(V_k) - E(V_{k_0}))$ in the set $\{k : |k - k_0| \leq Mn\lambda^{-2}\}$.

Write $k = k_0 + \lfloor v\lambda^{-2} \rfloor$. Then $G\lambda(E(V_k) - E(V_{k_0})) = |v|$. For the stochastic part, we use the invariance principle from Berkes, Liu and Wu (2014, [**?**]) and the decomposition in (10.3) to obtain, for $k > k_0$

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

A similar result holds for $k < k_0$ □

A. *Proof of Theorem 4.8*

Follows directly from Theorem 1 in Wu, Zhao (2007, [19])

B. *Proof of Theorem 4.9*

Follows directly from Theorem 1 in Zhou, Wu (2010, [21])

REFERENCES

[1] 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.

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

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

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

[5] Z. W. Birnbaum and Albert W. Marshall. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.*, 32:687–703, 1961.

[6] Jan J. J. Groen, George Kapetanios, and Simon Price. Multivariate methods for monitoring structural change. *J. Appl. Econometrics*, 28(2):250–274, 2013.

[7] J. Hájek and A. Rényi. Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.*, 6:281–283, 1955.

[8] D. V. Hinkley. Inference about the change-point from cumulative sum tests. *Biometrika*, 58:509–523, 1971.

[9] David Hinkley and Edna Schechtman. Conditional bootstrap methods in the mean-shift model. *Biometrika*, 74(1):85–93, 1987.

[10] Yannick Hoga. Monitoring multivariate time series. *J. Multivariate Anal.*, 155:105–121, 2017.

[11] Lajos Horváth and Marie Huˇ sková. Change-point detection in panel data. *J. Time Series Anal.*, 33(4):631–648, 2012.

[12] M. Lavielle and G. Teyssière. Détection de ruptures multiples dans des séries temporelles multivariées. *Liet. Mat. Rink.*, 46(3):351–376, 2006.

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

[14] 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).

[15] E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–527, 1955.

[16] E. S. PAGE. On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1-2):248, 1957.

[17] A. N. Pettitt. A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, 67(1):79–84, 1980.

[18] Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005.

[19] 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.

[20] 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.

[21] Zhou Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(4):513–531, 2010.