TEST manuscript No.
(will be inserted by the editor)

Dating multiple change points in the correlation matrix

Pedro Galeano · **Dominik Wied**

Received: date / Accepted: date

Abstract A nonparametric procedure for detecting and dating multiple change points in the correlation matrix of sequences of random variables is proposed. The procedure is based on a recently proposed test for changes in correlation matrices at an unknown point in time. Although the procedure requires constant expectations and variances, only mild assumptions on the serial dependence structure are assumed. The convergence rate of the change point estimators is derived and the asymptotic validity of the procedure is proved. Moreover, the performance of the proposed algorithm in finite samples is illustrated by means of a simulation study and the analysis of a real data example with financial returns. These examples show that the algorithm has large power in finite samples.

Keywords Binary segmentation algorithm · Correlation matrix · CUSUM statistics · Financial returns · Multiple change point detection · Nonparametric estimation

Mathematics Subject Classification (2000) 62M10 · 62G10 · 91B84

1 Introduction

The problem of detecting change points in a sequence of random variables can be stated as follows: a sequence of random variables has a set of characteristics, such

Financial support by Ministerio de Economía y Competitividad grant ECO2015-66593-P and Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged. We also would like to thanks three anonymous referees for very helpful comments.

P. Galeano
Universidad Carlos III de Madrid, Department of Statistics and UC3M-BS Institute of Financial Big Data, 28903 Getafe, Madrid, Spain
Tel.: +34 91 624 8901
Fax: +34 91 624 9848
E-mail: pedro.galeano@uc3m.es

D. Wied
University of Cologne, Institute for Econometrics and Statistics, 50923 Köln, Germany / TU Dortmund, Fakultät Statistik, 44221 Dortmund, Germany

as the mean and/or the variance, that follow a piecewise constant structure. Then, the goal is to detect the number of times that these characteristics change from a set of values to another, as well as the location of the changes. Additionally, it is of interest to estimate the characteristics in each constant period. The piecewise constant structure can then be taken into account to construct an appropriate model that can be used, for instance, to forecast future values of the sequence.

Many different strategies have been proposed to solve specific change point problems, such as penalized likelihood methods or binary segmentation procedures, see Aue and Horváth (2013) and Jandhyala et al (2013) for two recent and rather complete references on the topic. There are either methods which estimate all change points concurrently or ones that do so hierarchically. For instance, concurrent methods generally optimize a single objective function, i.e., given that there are k change points, Hawkins (2001) estimates change point locations by maximizing a likelihood function, whilst Lavielle and Teyssière (2006) accomplish the same task by minimizing a loss function. Sequential methods generally estimate change points one at a time (Guralnik and Srivastava, 1999), although some have the ability to estimate two or more at any given stage (Olshen and Venkatraman, 2004). Such approaches are often characterized as bisection procedures, which is the case in the proposed method that utilizes a bisection (binary segmentation) approach for its computational efficiency.

The idea of the procedure is the following: first, one searches for a single change point in the whole sequence using, for instance, a likelihood ratio or a cumulative sum (CUSUM) statistic. If a change point is detected, then the sequence is split in two subsequences that are used to search for new change points. This procedure was first proposed by Vostrikova (1981) and posteriorly implemented in various problems by Inclán and Tiao (1994), Bai (1997), Bai and Perron (1998), Andreou and Ghysels (2002), Gooijer (2006), Galeano (2007) and Galeano and Tsay (2010), among many others. See also Fryzlewicz and Rao (2014) and Fryzlewicz (2014) for two recent references on binary segmentation.

Change point problems have been mainly focused on changes in the mean and/or the variance of univariate sequences and in the mean and/or the covariance matrix of multivariate sequences. However, the case of changes in the correlation between sequences of multiple random variables has not been extensively analyzed. Wied et al (2012) proposed a nonparametric CUSUM statistic to test if correlations between two random variables remain constant over time, while Galeano and Wied (2014) proposed an algorithm based on the correlation constancy test to estimate both the number and the timing of possible change points. However, these two papers only consider single correlations between two random sequences. This restricts the applicability of these procedures when more than two variables are of interest. Recently, Wied (2015) proposed a CUSUM statistic that extends the methodology from the test proposed by Wied et al (2012) to higher dimensions, but keeping its nonparametric and model-free approach. A difference is that, in contrast to Wied et al (2012), a bootstrap estimator is used for standardizing purposes. Wied (2015) shows that the matrix-based test outperforms a method based on performing several pairwise tests and using a level correction like Bonferroni-Holm in some situations. Moreover, Berens et al (2015) show that the test is useful for Value at Risk (VaR) forecasting.

The main aim of this paper is twofold. First, we extend the nonparametric procedure for detecting and dating multiple change points in bivariate correlations proposed by Galeano and Wied (2014) to the case of dating multiple change points in the correlation matrix of sequences of random variables based on the test proposed by Wied (2015). In contrast to the bivariate procedure, the multivariate one is bootstrap-based. The procedure proceeds as follows: first, we determine the "dominating" change point and decide if this point is statistically significant. Then, we split the series in two pieces and again test for possible change points in each part of the series. The procedure stops if we do not find any new change point any more. Finally, a refinement step is performed to delete all possible false change points and to estimate their location more accurately. We believe that the proposed extension to the multivariate framework can be useful in several situations. For instance, in portfolio management, the number of assets considered is usually more than two and typically the interest lies in determining where a possible change occurs or how many changes there are. The papers cited previously do not consider both situations simultaneously as it is done in this paper. So, there is a need for such a procedure covering these two important aspects. The second aim is to provide a theoretical justification of the multivariate procedure. Due to its multivariate nature, the theory involved in order to show the main characteristics of the proposed procedure is more involved than that of the procedure in Galeano and Wied (2014) for single correlations. Assuming a finite number of change points, we analytically show that the proposed procedure gives the correct number of change points and that these are consistently estimated. As a by-product, we derive the convergence rate of the change point estimator which is not done in Galeano and Wied (2014). Furthermore, we show that the algorithm yields appealing results in simulated samples and in an empirical application.

As we are interested in the constancy of the correlation and not the covariance matrix, our procedure is based on the correlation constancy test proposed by Wied (2015) and not, for example, on the covariance constancy test proposed by Aue et al (2009). While both approaches are rather similar, there is some evidence in Wied et al (2012) (Table 2) that the correlation test might be powerful in some situations in which both correlation and covariance change, in particular in the case of multiple change points. Note also that our approach is nonparametric and consequently different than those assuming a parametric model for the multivariate time series, e.g., a VARMA model such as in Galeano and Peña (2007), Dvořák and Prášková (2013) and Dvořák (2017), among others.

The following part of the paper is organized as follows. Section 2 introduces the proposed procedure for detecting multiple change points in the correlation matrix of a multivariate random variable. Section 3 derives the asymptotic properties of the procedure (in particular its validity). Sections 4 and 5 present some simulation studies and a real data application that show the appealing behavior of the procedure in finite settings. Finally, Section 6 provides some conclusions. All proofs are presented in the supplementary material of the paper.

2 The procedure

In this section, we present the algorithm for detection of change points in the correlation matrix of a sequence of p-dimensional random vectors. First of all, we introduce some notation. Throughout the paper, $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{p,t})$, $t \in \mathbb{Z}$, denotes a sequence of p-variate random vectors on a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ with finite 4-th moments and (unconditional) correlation matrix $R_t = (\rho_t^{ij})_{1 \leq i,j \leq p}$, where

$$\rho_t^{ij} = \frac{\mathsf{Cov}(X_{i,t}, X_{j,t})}{\sqrt{\mathsf{Var}(X_{i,t})\mathsf{Var}(X_{j,t})}}.$$

We write $A \sim (m,n)$ for a matrix A with m rows and n columns. Moreover, we call $||\cdot||_r$ the L_r-norm of a vector, where $r > 0$, that means, for $A = (a_1, \ldots, a_m) \sim (m,1)$,

$$||A||_r := \left(\sum_{i=1}^{m} |a_i|^r \right)^{1/r}.$$

We denote by $\to_d$ and $\to_p$ convergence in distribution and probability, respectively, of random variables or vectors. The convergence symbols as well as all moment operators like Var are used with respect to P if not denoted otherwise. Moreover, let $\vee$ and $\wedge$ denote maximum and minimum, respectively. We impose the convention that in a vector indexed by $1 \leq i < j \leq p$, the first entries correspond to $i = 1$, the next ones to $i = 2$ and so on.

Given an observed time series $\mathbf{X}_1, \ldots, \mathbf{X}_T$, the algorithm is based on the test statistic[1]

$$A_{1,T} := \max_{2 \leq k \leq T} \frac{k}{\sqrt{T}} \left|\left| \hat{E}_{1,T}^{-1/2} P_{k,1,T} \right|\right|_1, \tag{1}$$

for the hypothesis pair $H_0 : R_1 = \ldots = R_T$ versus $H_1 : \neg H_0$. Here, $P_{k,1,T} := \left(\hat{\rho}_{1,k}^{ij} - \hat{\rho}_{1,T}^{ij} \right)_{1 \leq i < j \leq p} \in \mathbb{R}^{\frac{p(p-1)}{2}}$,

$$\hat{\rho}_{1,k}^{ij} := \frac{\sum_{t=1}^{k} (X_{i,t} - \bar{X}_{i,1,k})(X_{j,t} - \bar{X}_{j,1,k})}{\sqrt{\sum_{t=1}^{k} (X_{i,t} - \bar{X}_{i,1,k})^2} \sqrt{\sum_{t=1}^{k} (X_{j,t} - \bar{X}_{j,1,k})^2}},$$

$\bar{X}_{i,1,k} := \frac{1}{k} \sum_{t=1}^{k} X_{i,t}$, $\bar{X}_{j,1,k} := \frac{1}{k} \sum_{t=1}^{k} X_{j,t}$, and $\hat{E}_{1,T}$ is a bootstrap estimator which serves for approximating

$$E := \lim_{T \to \infty} \mathsf{Cov}\left(\sqrt{T} \left(\hat{\rho}_{1,T}^{ij} \right)_{1 \leq i < j \leq p} \right) \sim \left(\frac{p(p-1)}{2} \times \frac{p(p-1)}{2} \right) \tag{2}$$

for large T and large B, where B is the number of bootstrap replications. The bootstrap estimator $\hat{E}_{1,T}$ of E is computed as follows. First, we divide the observed time series $\mathbf{X}_1, \ldots, \mathbf{X}_T$ into $T - l_T - 1$ overlapping blocks O_i, $i = 1, \ldots, T - l_t - 1$, where l_T is a

[1] Note that we compare successively calculated correlation coefficients in the spirit of fluctuation tests like Ploberger et al (1989) here. The weighting factor $\frac{k}{\sqrt{T}}$ downweights imprecise estimations at the beginning of the sample and serves for obtaining weak convergence to a Brownian bridge.

block length, such that $O_1 = (\mathbf{X}_1, \ldots, \mathbf{X}_{l_T})$, $O_2 = (\mathbf{X}_2, \ldots, \mathbf{X}_{l_T+1}), \ldots$ Then, for some large B, we sample $\left\lceil \frac{T}{l_T} \right\rceil$ times with replacement one of the $T - l_T - 1$ blocks and stick the blocks together, obtaining B p-dimensional time series with length $\left\lceil \frac{T}{l_T} \right\rceil \cdot l_T$. Now, we calculate the vector $v_b := \sqrt{T}\left(\hat{\rho}^{ij}_{b,1,T}\right)_{1\leq i<j\leq p}$, where $\rho^{ij}_{b,1,T}$ is the sample correlation of the bootstrapped time series $b = 1, \ldots, B$. Finally, the estimator $\hat{E}_{1,T}$ is the empirical covariance matrix of these B vectors, i.e.,

$$\hat{E}_{1,T} = \frac{1}{B}\sum_{b=1}^{B}(v_b - \bar{v})(v_b - \bar{v})' \tag{3}$$

where $\bar{v} := \frac{1}{B}\sum_{b=1}^{B} v_b$.

To show the validity of the procedure, we will need several assumptions.

Assumption 1. For $0 \leq l_1 < l_2 \leq 1$,

$$U_t := \begin{pmatrix} X_{1,t}^2 & - & \mathsf{E}(X_{1,t}^2) \\ \vdots & & \vdots \\ X_{p,t}^2 & - & \mathsf{E}(X_{p,t}^2) \\ X_{1,t} & - & \mathsf{E}(X_{1,t}) \\ \vdots & & \vdots \\ X_{p,t} & - & \mathsf{E}(X_{p,t}) \\ X_{1,t}X_{2,t} & - & \mathsf{E}(X_{1,t}X_{2,t}) \\ X_{1,t}X_{3,t} & - & \mathsf{E}(X_{1,t}X_{3,t}) \\ \vdots & & \vdots \\ X_{p-1,t}X_{p,t} & - & \mathsf{E}(X_{p-1,t}X_{p,t}) \end{pmatrix}$$

and $S_j := \sum_{t=1\vee[l_1T]}^{j} U_t$, the probability limit as $T \to \infty$ of

$$\left(\frac{1}{(l_2 - l_1)T} S_{[l_2T]} S'_{[l_2T]}\right)$$

is a finite and positive definite matrix with dimension $\left(2p + \frac{p(p-1)}{2}, 2p + \frac{p(p-1)}{2}\right)$, denoted by D_1.

Assumption 2. For some $r > 2$, the r-th absolute moments of the components of U_t are uniformly bounded, that means, $\sup_{t\in\mathbb{Z}} \mathsf{E}||U_t||_r < \infty$.

Assumption 3. For the r from Assumption 2, the vector $(X_{1,t}, \ldots, X_{p,t})$ is L_2-NED (near-epoch dependent) with size $-\frac{r-1}{r-2}$ and constants $(c_t), t \in \mathbb{Z}$, on a sequence $(V_t), t \in \mathbb{Z}$, which is α-mixing of size $\phi^* := -\frac{r}{r-2}$, i.e.

$$\left|\left|(X_{1,t}, \ldots, X_{p,t}) - ((X_{1,t}, \ldots, X_{p,t}) \,|\sigma(V_{t-l}, \ldots, V_{t+l}))\right|\right|_2 \leq c_t w_l$$

with $\lim_{l\to\infty} w_l = 0$. The constants $(c_t), t \in \mathbb{Z}$ fulfill $c_t \leq 2||U_t||_2$ with U_t from Assumption 1.

Assumption 4. $(X_{1,t},\ldots,X_{p,t}), t \in \mathbb{Z}$, has constant expectation and variances, that means, $\mathsf{E}(X_{i,t})$ and $\mathsf{E}(X_{i,t}^2) > 0$, for $i = 1,\ldots,p$, do not depend on t.

Assumption 5. For $T \to \infty$, $l_T \to \infty$ and $l_T \sim T^{\alpha}$ for $\alpha \in (0,1)$.

Assumptions 1, 2 and 3 concern moments and serial dependencies of the components of $\mathbf{X}_t$. In particular, Assumption 1 is a regularity condition which rules out trending random variables. This situation is however not relevant, for instance, in the case of financial returns. Assumption 2 requires finite $(4+\gamma)$-th moments of $\mathbf{X}_t$ with $\gamma > 0$ arbitrary. Assumption 3 is a very general serial dependence assumption which holds, for instance, in most relevant econometric models such as GARCH models under certain conditions (cf. Carrasco and Chen, 2002). More precisely, Assumption 3 guarantees that the vector

$$(X_{1,t}^2,\ldots,X_{p,t}^2,X_{1,t},\ldots,X_{p,t},X_{1,t}X_{2,t},X_{1,t}X_{3,t},\ldots,X_{p-1,t}X_{p,t})$$

is L_2-NED (near-epoch dependent) with size $-\frac{1}{2}$, cf. Davidson (1994), p. 273. Assumption 4 consider that the first and second moments are constant which is a condition in line with Aue et al (2009) and can be slightly relaxed to allow for some fluctuations in the first and second moments. However, we do not consider this situation for ease of exposition and because the algorithm would not change. Nevertheless, with the observation that standard GARCH models are covered by this assumption, we investigate in our simulation study in Section 4 how the procedure behaves in finite samples in the presence of GARCH effects (volatility clustering).[2] Finally, Assumption 5 is similar as in Calhoun (2016), Theorem 1, and guarantees that the block length becomes large but not too large compared to T.

According to Wied (2015), Corollary 1, under H_0 and assumptions 1, 2, 3, 4 and 5, it holds

$$A_{1,T} := \max_{2\leq k\leq T} \frac{k}{\sqrt{T}} \left|\left|P_{k,1,T}\right|\right|_1 \to_d \sup_{0\leq s\leq 1} \left|\left|E^{-1/2}B^{\frac{p(p-1)}{2}}(s)\right|\right|_1 .$$

This implies the statistic $A_{1,T}$ can be approximated by the supremum of the sum of the absolute value of independent Brownian Bridges for large B and T. More precisely, H_0 is rejected whenever $A_{1,T}$ is larger than the $1-\alpha$ quantile of $A := \sup_{0\leq s\leq 1} \left|\left|B^{\frac{p(p-1)}{2}}(s)\right|\right|_1$. Quantiles of A can be easily estimated by Monte Carlo simulations by approximating the path of Brownian Bridges on fine grids, as seen in Section 4.

The proposed procedure sequentially employs the test statistic in (1) to estimate the timings and the number of multiple change points. In particular, we assume that there is a finite number of change points. However, the number, location and size of the change points are unknown. In Assumption 4 we assume that expectations and

[2] Given the similarity of the null hypotheses in Aue et al (2009) and Wied (2015) under Assumption 4, one could think about replacing test statistic (1) with the ones from Aue et al (2009) and work with their different assumption regarding serial dependencies. Aue et al (2009) do not use a bootstrap approximation so that such a procedure would be faster. However, we would not directly measure correlations in this case and finite sample simulations indicate that the Wied (2015)-test is more powerful in the case of multiple change points in both correlation and covariance.

variances are constant. Additionally, we assume that, under the alternative hypothesis, we have a piecewise constant second-order cross moment matrix $\mathsf{E}(X_t X_t')$. The formal assumption is:

Assumption 6. Under the (fixed) alternative, expectations and variances of $X_{1,t},\ldots,X_{p,t}$ are constant and equal to finite numbers, μ_i and σ_i^2, respectively, for $i=1,\ldots,p$. Moreover, the vector of covariances changes from $(\mathsf{E}(X_{i,t}X_{j,t})_{i,j})_{1\le i<j\le p} = ((m_{XX'})_{i,j})_{1\le i<j\le p}$ to

$$((\mathsf{E}\left(X_t X_t'\right))_{i,j})_{1\le i<j\le p} = ((m_{XX'})_{i,j})_{1\le i<j\le p} + g\left(\frac{t}{T}\right). \tag{4}$$

The function $g(z)\in\mathbb{R}^{\frac{p(p-1)}{2}}$, $z\in[0,1]$, is a step function with a finite number of steps ℓ, i.e. there is a partition $0=z_0<z_1<\ldots<z_\ell<z_{\ell+1}=1$ and there are second cross moment level vectors $a_0,\ldots,a_\ell$ such that

$$g(z)=\sum_{i=0}^{\ell} a_i \mathbf{1}_{\{z\in[z_i,z_{i+1})\}}$$

and $g(1)=a_l$. The quantities ℓ, $z_1,\ldots,z_\ell$ and $a_0,\ldots,a_\ell$ do not depend on T.

The function g specifies the timing and the size of the changes in the correlation matrix. Since this is a step function, we consider sudden changes in the correlation and do not consider smooth changes. The setting allows for a change in only a subset of the dimensions. Therefore, different change points can affect different components of the sequence.

Next, we present the proposed procedure with the goal of estimating ℓ, $z_1,\ldots,z_\ell$ and $a_0,\ldots,a_\ell$. To establish the asymptotic results, it is more convenient to present the procedure in terms of the estimator of the change point fraction. To that purpose, we rewrite the test statistic (1) as

$$A_{1,T} := \sup_{z\in[0,1]} \frac{\tau(z)}{\sqrt{T}} \left|\left|\hat{E}_{1,T}^{-1/2} P_{\tau(z),1,T}\right|\right|_1,$$

with $\tau(z)=[2+z(T-2)]$ (where $[\cdot]$ is the floor function) and

$$P_{\tau(z),1,T} = \left(\hat{\rho}_{1,\tau(z)}^{ij} - \hat{\rho}_{1,T}^{ij}\right)_{1\le i<j\le p} \in \mathbb{R}^{\frac{p(p-1)}{2}}.$$

Then, we estimate the timing of the break by $\hat{z} := \tau(\hat{z}^*)/T$ with $\hat{z}^* := \mathsf{argmax}_z B_{1,T}(z)$ and $B_{1,T}(z) := \frac{\tau(z)}{T}\left|\left|P_{\tau(z),1,T}\right|\right|_1$. Note that we do not consider the bootstrap estimator from the argmax estimator as it would disturb the information about the location of the change points given by the correlation differences. In fact, we will see later on that $B_{1,T}(z)$ converges to a function that essentially depends only on the function g. Here and in the following, we restrict the values z for which the argmax is calculated to multiples of $1/T$. In case of multiple solutions, we choose the smallest one. Note that in the first step of the procedure, $B_{1,T}(z)$ is calculated from all observations. In subsequent iterations, if needed, we just consider the observations in the relevant part of the sample and call the corresponding "target function"

$$A_{\eta(l_1),\tau(l_2)}(z) := \frac{\tau(z)-\eta(l_1)+1}{\sqrt{\tau(l_2)-\eta(l_1)+1}} \left|\left|\hat{E}_{\eta(l_1),\tau(l_2)}^{-1/2} P_{\tau(z),\eta(l_1),\tau(l_2)}\right|\right|_1, \tag{5}$$

where, $z \in [l_1, l_2]$ for $0 \leq l_1 < l_2 \leq 1$, $\eta(z) = \tau(z) - 1$, $\hat{E}_{\eta(l_1),\tau(l_2)}$ denotes the bootstrap estimate of E in (2) using data from $\eta(l_1)$ to $\tau(l_2)$ and

$$P_{\tau(z),\eta(l_1),\tau(l_2)} = \left(\hat{\rho}^{ij}_{\eta(l_1),\tau(z)} - \hat{\rho}^{ij}_{\eta(l_1),\tau(l_2)}\right)_{1\leq i<j\leq p} \in \mathbb{R}^{\frac{p(p-1)}{2}}.$$

Then the timing of break is estimated by

$$\hat{z}_{l_1,l_2} := \hat{z} := \tau(\hat{z}^*_{l_1,l_2})/T \tag{6}$$

with $\hat{z}^*_{l_1,l_2} = \mathsf{argmax}_{l_1\leq z\leq l_2} B_{\eta(l_1),\tau(l_2)}(z)$ and

$$B_{\eta(l_1),\tau(l_2)}(z) := \frac{\tau(z) - \eta(l_1) + 1}{\tau(l_2) - \eta(l_1) + 1} \left|\left|P_{\tau(z),\eta(l_1),\tau(l_2)}\right|\right|_1,$$

Basically, this means that we always look for the time point at which the test statistic (5) (calculated from data in a particular interval) takes its maximum and divide by T.

Under the null hypothesis of no correlation change, it can be shown by Markov's inequality $\hat{E}_{1,T}$ converges to a positive definite matrix with respect to the product measure $\mathsf{P}^{\times}$ which combines randomness from the data as well from the bootstrap. So, it is reasonable to assume that the matrix $\hat{E}_{\eta(l_1),\tau(l_2)}$ is invertible. It is an interesting task for further research to prove what happens with this matrix under fixed alternatives as considered below. In order to ensure consistency of the test statistic, we impose the convention that $\hat{E}_{\eta(l_1),\tau(l_2)}$ if not invertible is perturbed slightly such that it becomes invertible. This does not affect the asymptotic properties neither under the null hypothesis nor under alternatives. The only potential drawback is that there might be some overrejections in finite samples if p is large compared to T. Nevertheless, in the following, we always assume without loss of generality that $\hat{E}^{-1/2}_{\eta(l_1),\tau(l_2)}$ exists.

Formally, the algorithm proceeds as follows:

1. Let $\mathbf{X}_1, \ldots, \mathbf{X}_T$ be the observed series. Obtain the test statistic $A_{1,T}$. There are two possibilities:
   (a) If the test statistic is statistically significant, i.e., if $A_{1,T} > c_{T,\alpha}$, where $c_{T,\alpha}$ is the asymptotic critical value for a given upper tail probability, then a change in the correlation matrix is announced. Let z_1 be the break point estimator from (6) and go to Step 2.
   (b) If the test statistic is not statistically significant, the algorithm stops, and no change points are detected.
2. Let $z_1, \ldots, z_\ell$ be the ℓ change points in increasing order already found in previous iterations. If

$$\max_k \left\{A_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)},\ k = 1, \ldots, \ell+1\right\} > c_{T,\alpha},$$

   where $A_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)}$ is the value of the statistic calculated from the data from $\eta(z_{k-1} + \frac{1}{T})$ to $\tau(z_k)$, for $k = 1, \ldots, \ell+1$, taking $z_0 = 0$ and $z_{\ell+1} = 1$, then a new

change point is detected at the point fraction at which the value $A_{\eta(z_{k_{\max}-1}+\frac{1}{T}),\tau(z_{k_{\max}})}$ is attained, where:

$$k_{\max} = \arg\max_k \left\{ A_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)}, k = 1,\ldots,\ell \right\}.$$

Repeat this step until no more change points are found.

3. Let $(z_1 < \ldots < z_\ell)$ be the detected change points. If $\ell > 1$, refine the estimates of the change point locations by calculating the statistic from the data from $\eta(z_{k-1} + \frac{1}{T})$ to $\tau(z_{k+1})$, for $k = 1,\ldots,\ell$, where $z_0 = 0$ and $z_{\ell+1} = 1$. If any of the change points is not statistically significant, delete it from the list, and repeat this step.
4. Finally, estimate the correlation matrix of $\mathbf{X}_1,\ldots,\mathbf{X}_T$ in each segment separately with the usual sample correlation matrix where individual correlations are computed using the Bravais-Pearson correlation coefficient.

As shown in Section 3, the proposed procedure consistently detects the true change points. Steps 1 and 2 are, essentially, the steps performed within the usual binary segmentation procedure. Step 3 is meant to refine the estimation of the change points as in this step the algorithm computes the value of the statistic in intervals only affected by the presence of one single change point, which is not guaranteed in Step 2. The final step computes the correlation matrix in subintervals with constant (unconditional) correlation.

A key issue in applying the procedure to real data is the selection of the critical value used in the algorithm. A possibility is to use always the same critical value in each step of the procedure. However, this may lead to over-estimation of the number of change points due to the fact that, the larger the number of detected change points, the larger the accumulation of type I errors. Although we later prove that we can consistently estimate the correct number of change points even if this is the selected strategy, in practice, we require that the type I errors used depend on the number of change points already detected by the algorithm. More precisely, let α_0 be the type I error for Step 1. Then, we use the critical value c_{T,α_k} after detecting the k-th change point, where α_k is such that $1-\alpha_0 = (1-\alpha_k)^{k+1}$. This leads to $\alpha_k = 1-(1-\alpha_0)^{\frac{1}{k+1}}$ and the overall significance level (the probability of at least one false detection) is bounded by α_0 for all tests. For instance, if $\alpha_0 = 0.05$, $\alpha_1 \approx 0.025$, $\alpha_2 \approx 0.017$ and so on.[3]

For consistently estimating the number of break points the initial type I error in the asymptotic result concerning the number of break points (Theorem 3) would have to converge to zero with a certain restriction on the convergence rate, see Assumption 8. The Bonferroni-type correction outlined in the paragraph above fits into this framework if α_k is constant for all k larger than a threshold k_0. In finite samples it is not necessary to fix such a threshold and it is acceptable to set α_0 constantly equal to 0.05. Then, one can expect that the probability of an over-estimated number of

[3] This way, the significance level becomes very small for larger k, but one should keep in mind that a piecewise constant model might in general not be appropriate in the case of many change points. Under such circumstances, a model with continuously changing correlations could be better.

change points is bounded by 5%. Indeed, the simulations in Section 4 clearly suggest that this is the case. As noted before, we use the quantiles of the distribution of the supremum of the sum of the absolute value of independent Brownian Bridges estimated by Monte Carlo simulations that can be approximated numerically as explained in Section 4.

The proposed procedure works reasonably well even in small samples in terms of detection of the true number of changes, as shown in the Monte Carlo experiments of Section 4. However, note that, if the number of change points detected is large compared to the sample size, then a piecewise constant correlation matrix may not be a good description of the true correlation of the series.

3 Analytic results

In this section, we derive analytic results for the change point estimators and detection algorithm. In order to do this, we assume that all change-points that exist are "dominating" ones.

Assumption 7. For any $0 \leq l_1 < l_2 \leq 1$, define the vector

$$P^*_{l_1,l_2}(z) := \int_{l_1}^{z} g(t)dt - \frac{z-l_1}{l_2-l_1}\int_{l_1}^{l_2} g(t)dt.$$

Moreover, let C be a diagonal matrix with the diagonal vector given by $\left(\sqrt{\sigma_i^2\sigma_j^2}\right)_{1\leq i<j\leq p}$ and define $P^{**}_{l_1,l_2}(z) = CP^*_{l_1,l_2}(z)$. Then $P_{l_1,l_2}(z) := \left|\left|P^{**}_{l_1,l_2}(z)\right|\right|_1$ is either constant or has a unique maximum for a $z \in [l_1, l_2]$.

This is fulfilled for example if there is one dominating break in one component, whereas the correlations in the other components stay constant. The condition is also fulfilled if there is a dominating break in all components at the same time point. A detailed discussion on dominating change points including examples in which Assumption 7 holds or not (in the case of $p = 2$) can be found in Galeano and Wied (2014), p. 265.

Based on this assumption, we can show consistency of the estimator, validity of the algorithm (in the sense that the number of change points is detected correctly asymptotically) and a local power result.

Theorem 1 *Let Assumptions 2, 3, 4, 6 and 7 be true and let there be at least one break point in a given interval* $[l_1, l_2] \subseteq [0,1]$ *with* $l_1 < l_2$. *Then the change point estimator* (6) *is consistent for the dominating change point.*

Note that, for Theorem 1, we need not apply a functional central limit theorem so that we do not need Assumption 1. Moreover, one could relax Assumption 2 by only assuming the existence of finite q-th moments for a $q > 1$. Finally, we do not need Assumption 5 on a block length as the bootstrap estimator does not appear within the argmax estimator.

In addition to a consistency result, a statistician is also interested in the convergence rate of the change point estimator. Such a result is given in Theorem 2.

Theorem 2 *Let the assumptions from Theorem 1 be true. Then, for every $i = 1, \ldots, \ell$ and $\varepsilon > 0$, there is a $M > 0$ and a $T_0 > 0$ such that*

$$\mathsf{P}(|T(\hat{z}_i - z_i)| > M) < \varepsilon.$$

for all $T > T_0$.

Interestingly, the convergence rate is not $\sqrt{T}$, which would hold within e.g. the central limit theorem, see also Dümbgen (1991).

The preceding theorem is of potential own interest, but is also needed in order to ensure that the asymptotic behavior of the test statistic calculated from $\tau(\hat{z}_i)$ to $\tau(\hat{z}_{i+1})$ is similar to that calculated from $\tau(z_i)$ to $\tau(z_{i+1})$. This will be important in the proof of Theorem 2, especially for the fact that the number of change points is not overestimated asymptotically.

While the convergence results above are important, our main interest lies in consistently estimating the number of change points. For this, we need the assumptions for applying a functional central limit theorem and an additional assumption on the critical values which goes back to Bai (1997). Compare also the discussion in the end of Section 2 regarding the choice of the critical values and see also the comments in Section 2 in Galeano and Wied (2014) about the choice of c_{T,α_k} in finite sample settings.

Assumption 8. The critical values c_{T,α_k} used in the algorithm obey the condition $\lim_{T\to\infty} c_{T,\alpha_k} = \infty$ and $c_{T,\alpha_k} = o(\sqrt{T})$ for $k \in \mathbb{N}_0$.

Moreover, we need an assumption regarding the bootstrap estimator.

Assumption 9. Consider $0 \leq l_1 < l_2 \leq 1$.

a) If there is no correlation change in the interval $[l_1, l_2]$, $\hat{E}_{\eta(l_1),\tau(l_2)}$ consistently estimates the matrix E, uniformly over l_1 and l_2.

b) The eigenvalues of $\hat{E}_{\eta(l_1),\tau(l_2)}$ are stochastically bounded, uniformly over l_1 and l_2.

Assumption 9.a) ensures that the bootstrap approximation works. A sufficient condition would be to assume, for some $\delta > 0$ and $b = 1, \ldots, B$, that the random variable

$$C_T := \mathsf{E}\left(\left| \sqrt{T}(\hat{\rho}^{ij}_{b,[l_1T],[l_2T]} - \hat{\rho}^{ij}_{[l_1T],[l_2T]}) \right|^{2+\delta} | \mathbf{X}_1, \ldots, \mathbf{X}_T \right)$$

is stochastically bounded uniformly over l_1 and l_2. In particular, this means that the bootstrap correlation coefficients are sufficiently close to the correlation coefficients obtained from the data. With such an assumption, we can infer moment from distribution convergence based on arguments given in Cheng (2015) and Kato (2011). It remains an open research question if there are sharp low-level assumptions which guarantee that this assumption holds.

Theorem 3 *Let Assumptions 1, 2, 3, 4, 5, 6, 7, 8 and 9 be true. Let $\hat{\ell}$ the estimated number of change points from the algorithm proposed in Section 2. Then, for any $B \in \mathbb{N}$, $\hat{\ell} \to_p \ell$, $\hat{z}_i \to_p z_i$ for $i = 1, \ldots, \ell$ and $\hat{a}_i \to_p a_i$ for $i = 0, \ldots, \ell$. Here, $\to_p$ means convergence in probability with respect to the product measure $P^{\times}$.*

Note that we need the product measure $P^{\times}$ in the previous theorem as the estimators $\hat{\ell}$,$\hat{z}_i, i = 1,\ldots,\ell$,$\hat{a}_i, i = 0,\ldots,\ell$ always depend on the randomness from the original data as well as from the outcomes from the bootstrap repetitions.

Moreover, note that Theorem 3 holds for any value of B. The requirement $B \to \infty$ is only needed for obtaining an asymptotic level-α-test in the case of no change point.

Finally, in this section, we want to address the case in which the correlation shifts tend to zero with rate $\frac{1}{\sqrt{T}}$ as the sample size increases such that in Assumption 6 we replace (4) by

$$((\mathsf{E}\left(X_t X_t'\right))_{i,j})_{1\leq i<j\leq p} = ((m_{XX'})_{i,j})_{1\leq i<j\leq p} + \frac{1}{\sqrt{T}} g\left(\frac{t}{T}\right). \tag{7}$$

In this setting, we do not have consistency to the true break point any more, but the change point estimator converges to a non-degenerated random variable as the next theorem shows.

Theorem 4 *Let Assumptions 2, 3, 4, 6 (with* (4) *replaced by* (7)*) and 7 be true and let there be at least one break point in a given interval* $[l_1, l_2] \subseteq [0,1]$ *with* $l_1 < l_2$. *Then it holds for the change point estimator* (6) *that*

$$\hat{z} \to_d \underset{l_1\leq z\leq l_2}{\operatorname{argmax}} \left|\left| E^{1/2} W^{\frac{p(p-1)}{2}}(z) - E^{1/2} W^{\frac{p(p-1)}{2}}(l_1) \right.\right.$$
$$\left.\left. - \frac{z-l_1}{l_2-l_1} E^{1/2}\left(W^{\frac{p(p-1)}{2}}(l_2) - W^{\frac{p(p-1)}{2}}(l_1)\right) + P^{**}_{l_1,l_2}(z) \right|\right|_1,$$

where $P^{**}_{l_1,l_2}(z)$ *is from Assumption 7,* E *is from* (2) *and* $W^{\frac{p(p-1)}{2}}(z)$ *is a* $\frac{p(p-1)}{2}$*-dimensional standard Brownian motion.*

Note that we use the convergence rate $T^{-1/2}$ for local power analysis that has been previously used in other papers, see Wied et al (2012) and Wied (2015), among others, and that are confirmed by results in Dümbgen (1991). If we choose a different convergence rate for the local alternatives, as done in Csörgö and Horváth (1997), for instance, one would get a different rate for the change point estimator.

4 Simulation evidence

In this section, we present several Monte Carlo experiments to illustrate the performance of the proposed algorithm in finite samples. We focus on three important aspects: first, the empirical size of the procedure, second, its power in correct detection of changes, and third, its ability to accurately identify the location of the change points. In all the Monte Carlo experiments in this section and the real data example in Section 5, the critical values used are the estimated quantiles of the distribution of the supremum of the sum of the absolute value of independent Brownian Bridges. In particular, as the dimension of the series in the simulations and the real data example in Section 5 is $p = 4$, we obtain the estimated quantiles by generating 100000 sets

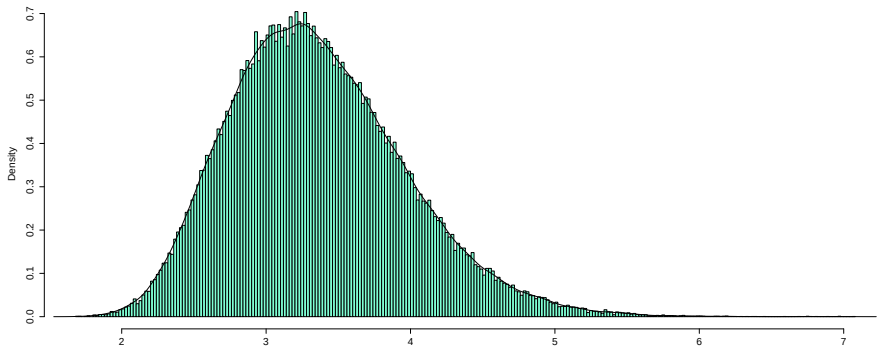

Fig. 1 Histogram of 100000 generated values from the asymptotic distribution of the $A_{1,T}$ statistic with a kernel density estimate

of 6 independent Brownian Bridges in a fine grid of 1000 points in the interval $[0,1]$. Then, for each set, we take the absolute values of the observed Brownian Bridges, add the six of them and obtain the maximum of the sums over the generated 1000 points. In this way, we obtain a sample of 100000 random values of the required distribution, from which we can estimate the quantiles. For instance, the first five quantiles used in Steps 1 and 2 of the procedure, if needed, are 4.4366, 4.6890, 4.8298, 4.9230 and 4.9907, respectively. An histogram of the 100000 random values and a kernel estimate of their density function are shown in Figure 1 that suggests that the asymptotic distribution is slightly positive skewed.

In the simulations, we consider several variants of the scalar BEKK model proposed by Ding and Engle (2001) for two main reasons. First, our main fields of application are financial returns and BEKK models are ones of the most widely used models to analyze these kind of time series. Second, unlike many other multivariate GARCH models, it is possible to derive the unconditional covariance and correlation matrices of the series that allow us to simulate series with a changing unconditional correlation matrix.

We initially focus on the size of the procedure, i.e., the accuracy of the procedure in estimating the number of change points if the true value is zero. For that, we consider the scalar BEKK model given by:

$$\mathbf{X}_t = \mathbf{H}_t^{1/2}\mathbf{E}_t$$
$$\mathbf{H}_t = \left(1-\alpha^2-\beta^2\right)\mathbf{H}+\alpha^2\mathbf{X}_{t-1}\mathbf{X}_{t-1}'+\beta^2\mathbf{H}_{t-1}$$

where $\mathbf{H}_t$ is the conditional covariance matrix of $\mathbf{X}_t$, $\mathbf{E}_t$ are iid random vectors with mean $\mathbf{0}_4$ and covariance matrix $\mathbf{I}_4$, and α and β are positive numbers such that $\alpha^2+\beta^2<1$, to ensure covariance stationary. Under these assumptions, it is not difficult to show that $\mathbf{H}$ is the unconditional covariance matrix of $\mathbf{X}_t$. Therefore, the unconditional correlation matrix of $\mathbf{X}_t$ can be written as $\mathbf{R}=\mathbf{D}^{-1/2}\mathbf{H}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is a diagonal matrix with elements the unconditional variances of the components

Table 1 Relative frequency detection of 0 and more than 0 change points given by the procedure based on 500 generated time series and an initial nominal significant level of $\alpha_0 = 0.05$.

	Gaussian		Student-t_5	
T	0	≥ 1	0	≥ 1
500	.938	.062	.896	.104
1000	.940	.060	.918	.082
2000	.938	.062	.930	.070
3000	.942	.058	.938	.062
4000	.948	.052	.944	.056

of $\mathbf{X}_t$, that are the elements in the main diagonal of $\mathbf{H}$. In particular, we take $\alpha = 0.1$ and $\beta = 0.8$, and $\mathbf{D} = \mathbf{I}_4$, so that $\mathbf{H} = \mathbf{R}$, with:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.5 & 0.6 & 0.7 \\ 0.5 & 1 & 0.5 & 0.6 \\ 0.6 & 0.5 & 1 & 0.5 \\ 0.7 & 0.6 & 0.5 & 1 \end{pmatrix}. \tag{8}$$

The random errors, $\mathbf{E}_t$, are assumed to be, first, a four dimensional standard Gaussian distribution, and, second, a four dimensional standardized Student-t with 5 degrees of freedom. The latter distribution represents an extreme situation that can be realistic in financial applications. The sample sizes considered are $T = 500$, 1000, 2000, 3000 and 4000, which are usual sample sizes of financial returns, while the block lengths are $l_T = \left[T^{1/4}\right]$, i.e., $l_{500} = 4$, $l_{1000} = 5$, $l_{2000} = 6$, $l_{3000} = 7$ and $l_{4000} = 7$, respectively. The number of bootstrap replications is $B = 1000$. Table 1 gives the relative frequency detection of 0 and more than 0 change points given by the procedure based on 500 generated series and an initial nominal significant level of $\alpha_0 = 0.05$. From this table, it seems that the type I error of the proposed procedure is very close to the initial nominal level for the Gaussian even with the smallest sample size, while there are some small size distortions for the standardized Student-t with 5 degrees of freedom, although the level seems to converge to the initial nominal significant level of $\alpha_0 = 0.05$ for higher T. Therefore, overestimation does not appear to be an issue for the proposed procedure if there are no changes in the correlation.

Next, we analyze the power of our procedure when there is a single change point in the series. The Monte Carlo setup is similar to the one described above, but the series are generated with a single change point in the unconditional correlation matrix. Three locations of the change point are considered, $z_1 = 0.25$, 0.50 and 0.75, respectively. The change is such that $\mathbf{R}$ is initially as in (8) and then changes at z_1 to:

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0.7 & 0.6 & 0.5 \\ 0.7 & 1 & 0.7 & 0.6 \\ 0.6 & 0.7 & 1 & 0.7 \\ 0.5 & 0.6 & 0.7 & 1 \end{pmatrix}. \tag{9}$$

Then, the largest correlation change is of magnitude .2, while two of the correlations do not change at all. This setting seems quite reasonable in practice. Indeed, in the real data example shown in Section 5, we found a correlation change in a system with 4 time series in which the largest correlation change found is of magnitude

Table 2 Relative frequency detection of 0, 1 and more than 1 change points given by the procedure based on 500 generated time series and an initial nominal significant level of $\alpha_0 = 0.05$.

		$z_1 = .25$			$z_1 = .50$			$z_1 = .75$		
	T	0	1	≥ 2	0	1	≥ 2	0	1	≥ 2
Gaussian	500	0.034	0.910	0.056	0.000	0.930	0.070	0.048	0.892	0.060
	1000	0.004	0.940	0.056	0.000	0.928	0.072	0.000	0.952	0.048
	2000	0.006	0.934	0.060	0.000	0.942	0.058	0.002	0.922	0.076
	3000	0.000	0.948	0.052	0.000	0.946	0.054	0.000	0.938	0.062
	4000	0.000	0.946	0.054	0.000	0.946	0.054	0.002	0.942	0.056
	T	0	1	≥ 2	0	1	≥ 2	0	1	≥ 2
Student-t_5	500	0.452	0.518	0.030	0.328	0.620	0.052	0.618	0.370	0.012
	1000	0.276	0.694	0.030	0.156	0.800	0.044	0.304	0.658	0.038
	2000	0.130	0.808	0.062	0.042	0.894	0.064	0.100	0.852	0.048
	3000	0.080	0.870	0.050	0.014	0.924	0.062	0.048	0.908	0.044
	4000	0.028	0.928	0.044	0.000	0.942	0.058	0.010	0.936	0.054

0.3644, while others have magnitudes 0.2673, 0.2343, or 0.1873. Table 2 shows the relative frequency detection of zero, one and more than one change points given by the procedure based on 500 generated time series and an initial nominal significant level of $\alpha_0 = 0.05$. First, it seems that the procedure performs quite well in detecting a single change point, with many cases over 90% correct detection. Second, as expected, as the sample size increases the procedure works better. Third, when the sample size is small, the probability of under-detection may be large only if the errors are Student-t with 5 degrees of freedom. A reason for the poor behavior with this distribution might be that the kurtosis of the marginal distributions is very high. However, in practice, one does not expect to have many change points if the length of the series is small. Fourth, the location of the change point does not strongly affect the behavior of the procedure unless the sample size is small. In this latter case, the power of the procedure is higher if the change point lies in the middle of the series. This is in coincidence with other procedures relying on CUSUM statistics as the one used here. Finally, in most cases, the percentage of false detection is always quite close to the nominal 5%, especially in the Gaussian case. Specifically, the frequency of over-detection is small. Regarding estimation of the location of the change point, Table 3 shows the median and mean absolute deviation of the change point estimators when the number of change points detected is equal to 1. The table shows that the medians of the estimates are reasonable close to the true change point locations. Indeed, the larger the sample size, the smaller the empirical mean absolute deviation. In particular, the bias appear to be smaller in the Gaussian case.

Next, we conduct another Monte Carlo experiment to study the power of the proposed procedure for detecting two change points. In this case, the location of the change points are $z_1 = 0.35$ and $z_2 = 0.7$, respectively. The changes are such that the correlation matrix of the series before the first change point is the correlation matrix in (8), then changes to the correlation matrix in (9), and, finally, changes again to the correlation matrix in (8) at the second change point. Note that in this scenario there is no dominant change point but we prefer to consider this situation to show that the procedure works well also in this case if the sample size is sufficiently large.

Table 3 Median and MAD of the change point estimators when the number of change points detected is equal to 1.

	$z_1 = .25$		$z_1 = .50$		$z_1 = .75$	
	Gaussian	Student-t_5	Gaussian	Student-t_5	Gaussian	Student-t_5
T	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)
500	0.2760 (0.0622)	0.3660 (0.1690)	0.5040 (0.0444)	0.5160 (0.1275)	0.7340 (0.0593)	0.6860 (0.1482)
1000	0.2610 (0.0252)	0.3010 (0.1186)	0.5040 (0.0192)	0.5140 (0.0874)	0.7470 (0.0281)	0.7025 (0.1230)
2000	0.2560 (0.0155)	0.2665 (0.0489)	0.5015 (0.0103)	0.5005 (0.0555)	0.7490 (0.0111)	0.7382 (0.0685)
3000	0.2533 (0.0098)	0.2600 (0.0434)	0.5006 (0.0079)	0.5036 (0.0311)	0.7493 (0.0079)	0.7420 (0.0365)
4000	0.2527 (0.0063)	0.2525 (0.0363)	0.5007 (0.0055)	0.5027 (0.0303)	0.7497 (0.0066)	0.7465 (0.0303)

Table 4 Relative frequency detection of 0, 1, 2 and more than 2 change points given by the procedure based on 500 generated time series and an initial nominal significant level of $\alpha_0 = 0.05$.

	$(z_1, z_2) = (.35, .7)$							
	Gaussian				Student-t_5			
T	0	1	2	≥ 3	0	1	2	≥ 3
500	0.586	0.084	0.310	0.020	0.754	0.134	0.108	0.004
1000	0.116	0.006	0.804	0.074	0.482	0.096	0.386	0.036
2000	0.000	0.000	0.938	0.062	0.086	0.010	0.824	0.080
3000	0.000	0.000	0.932	0.068	0.004	0.004	0.914	0.078
4000	0.000	0.000	0.940	0.060	0.004	0.002	0.925	0.069

Table 4 shows the relative frequency detection of zero, one, two and more than two change points given by the procedure based on 500 generated time series and an initial nominal significant level of $\alpha_0 = 0.05$. As in the case of a single change point, the proposed procedure works reasonably well, especially when the sample size gets larger. However, there are problems for $T = 500$ and the procedure underestimates the number of change points in the case of Student-t errors. On the other hand, Table 5 shows the median and mean absolute deviation of the estimates of the two change point locations when the number of change points detected is equal to 2. As expected in view of the results in Section 3, the medians of the estimates are reasonable close to the true ones. Again, it appears that the larger is the sample size, the better are the locations estimated.

Finally, note that these simulations are carried out for $p = 4$. When the dimension of the system is higher, the size and power properties become slightly worse, unless the sizes of the changes are large.

5 Empirical application

In this section, we illustrate the performance of the proposed procedure with empirical financial time series. For this, we look for changes in the correlation structure of the daily simple return series of four stocks. Specifically, we consider four European companies, Total, Sanofi, Siemens and BASF from January 1, 2007 to June 1,

Table 5 Median and MAD of the change point estimators when the number of change points detected is equal to 2.

	$(z_1, z_2) = (.35, .7)$			
	Gaussian		Student-t_5	
T	Median $(\widehat{z}_1)$ (Mad)	Median $(\widehat{z}_2)$ (Mad)	Median $(\widehat{z}_1)$ (Mad)	Median $(\widehat{z}_2)$ (Mad)
500	0.3540 (0.0415)	0.7060 (0.0296)	0.3570 (0.0415)	0.699 (0.0415)
1000	0.3535 (0.0222)	0.7010 (0.0192)	0.3530 (0.0326)	0.704 (0.0415)
2000	0.3515 (0.0111)	0.7005 (0.0111)	0.3525 (0.0229)	0.700 (0.0207)
3000	0.3510 (0.0079)	0.7006 (0.0069)	0.3510 (0.0168)	0.701 (0.0153)
4000	0.3510 (0.0055)	0.7005 (0.0044)	0.3521 (0.0109)	0.700 (0.0090)

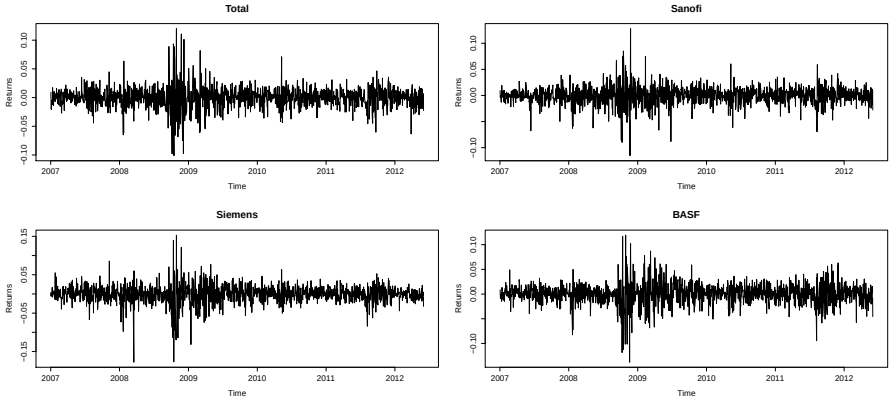

Fig. 2 Daily simple returns of Total, Sanofi, Siemens and BASF

2012 consisting of $T = 1414$ data points. The data was obtained from the database Datastream. The four return series are plotted in Figure 2, which shows very similar patterns. The autocorrelation functions of the simple returns show some minor serial dependence, while the autocorrelation functions of the squared simple returns reveal considerable serial dependence, as usual in stock market returns.

The empirical full sample correlation matrix is given by:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.5483 & 0.6460 & 0.6734 \\ 0.5483 & 1 & 0.4821 & 0.4998 \\ 0.6460 & 0.4821 & 1 & 0.7208 \\ 0.6734 & 0.4998 & 0.7208 & 1 \end{pmatrix}.$$

Figure 3 shows rolling windows for the six pairwise correlations of the simple return series with window length 120 that roughly corresponds to a trading time of about half a year. The plots show time-varying correlations. It is interesting to see several correlation-ups and -downs.

Next, we apply the proposed segmentation procedure of Section 2 to detect correlation changes for the simple returns of the Total, Sanofi, Siemens and BASF stock

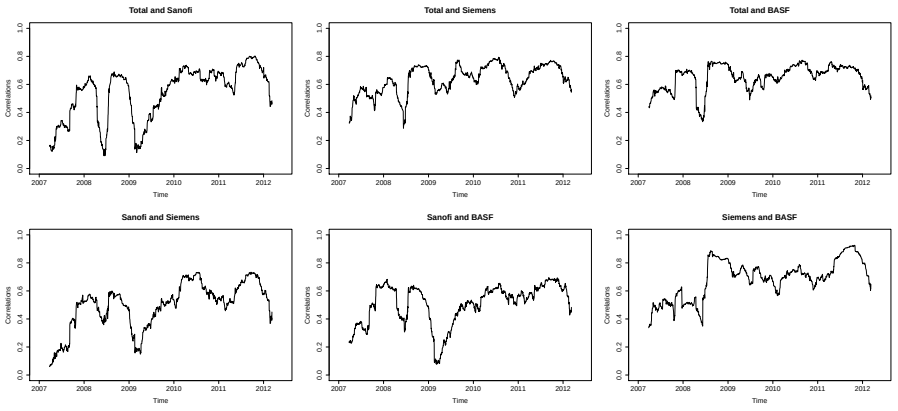

Fig. 3 Rolling correlations for the daily simple returns of Total, Sanofi, Siemens and BASF

assets. Table 6 shows the iterations taken by the procedure. In the first step, we start with the asymptotic critical value at the 5% significance level and the procedure detects a change in the correlation at time point $t = 443$ (September 11, 2008). The value of the test statistic (1) is 6.3280 which is statistically significant at the 5% level. Then, we split the series into two subperiods and look for changes in the subintervals $[1,443]$ and $[444,1414]$, respectively. In the first subinterval, the procedure detects a change point at time point $t = 134$ (July 6, 2007). The value of the test statistic is 4.8159. Afterwards, we split the subinterval $[1,443]$ into two subintervals, $[1,134]$ and $[135,443]$, respectively, and look again for new change points. No more changes are found in the three subintervals $[1,134]$, $[135,443]$, and $[444,1414]$. Next, we pass to Step 3 (the refinement step) and compute the statistic in the subintervals $[1,443]$, and $[135,1414]$, respectively. In the first subinterval, the procedure detects a change point at time point $t = 134$ (July 6, 2007) and the value of the test statistic is 4.7438. In the second subinterval, the procedure detects a change point at time point $t = 443$ (September 11, 2008) and the value of the test statistic is 5.5399. As the detected change points are the same as in the previous iterations, the algorithm stops and the time points located at $t = 134$ (July 6, 2007) and 443 (September 11, 2008) are the final detected change points.

It is interesting to see that the dates of the detected change points fare well with well known financial facts. The first estimated change point corresponds to the beginning of the Global Financial Crisis around the middle of 2007. The reduction of interest rates leads to several consequent issues starting with the easiness of obtaining credit, leading to sub-prime lending, so that an increased debt burden arised. Finally, there was a liquidity shortfall in the banking system that resulted in the collapse of important financial institutions such as Lehman Brothers and Merrill Lynch, among others, and the bailout of banks by national governments such as Bear Stearns, Bank of America and Bankia, among others. Specifically, the bankruptcy of Lehman Brothers was formally announced at on September 15, 2008, after a week of rumours. This date is very close to the second estimated change point.

Table 6 Iterations taken by the procedure in the real data example, (*) means statistically significant change point. The initial nominal significant level is $\alpha_0 = 0.05$.

Step 1				
Interval	A	Change point	Time point	Date
$[1, 1414]$	6.3280 (*)	0.3132	443	September 11, 2008
Step 2				
Interval	A	Change point	Time point	Date
$[1, 443]$	4.8159 (*)	0.0947	134	July 6, 2007
$[444, 1414]$	2.1415	0.8437	1193	July 28, 2011
$[1, 134]$	4.2863	0.0827	117	July 13, 2007
$[135, 443]$	3.6897	0.2220	314	March 4, 2008
$[444, 1414]$	2.1415	0.8437	1193	July 28, 2011
Step 3				
Interval	A	Change point	Time point	Date
$[1, 443]$	4.7438 (*)	0.0947	134	July 6, 2007
$[135, 1414]$	5.5399 (*)	0.3132	443	September 11, 2008

Table 7 Correlation matrices in each period.

Period	Empirical correlation matrix			
First	1	0.1564	0.3275	0.4917
	0.1564	1	0.0708	0.2261
	0.3275	0.0708	1	0.3604
	0.4917	0.2261	0.3604	1
Second	1	0.3907	0.5148	0.5677
	0.3907	1	0.4352	0.4924
	0.5148	0.4352	1	0.5110
	0.5677	0.4924	0.5110	1
Third	1	0.5990	0.6924	0.6986
	0.5990	1	0.5159	0.5168
	0.6924	0.5159	1	0.7857
	0.6986	0.5168	0.7857	1

Next, Table 7 shows the correlation matrices of the three simple returns of the Total, Sanofi, Siemens and BASF stock assets for the three periods of constant unconditional correlation provided by the procedure. Note how all the pairwise correlations increase after each detected change point. For instance, the correlation between Total and Sanofi pass from 0.1564 to 0.3907 at the first change point and then from 0.3907 to 0.5990 at the second change point. This is in accordance with the phenomenon kwown as "Diversification Meltdown" according to which correlations between financial returns often increase in times of crises.

6 Conclusions

This paper proposes a procedure for detecting change points in the correlation matrix of a sequence of multiple random variables. The procedure is based on a recently proposed CUSUM test statistic. Under certain assumptions, the procedure consistently detects the true number and location of the change points. The finite sample behavior

of the procedure is analyzed via several simulation studies and illustrated by means of the analysis of a four dimensional time series of simple returns of four European companies. In this real data example, the procedure detects changes at points that fare well with external events affecting the financial markets. Regarding future research, it might be interesting to consider a more sophisticated algorithm that includes modifications of the standard binary segmentation procedure to increase the power of the procedure in small samples. Moreover, one could consider robust dependence measures as alternatives to the usual correlation matrix.

References

Andreou E, Ghysels E (2002) Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17(5):579–600

Aue A, Horváth L (2013) Structural breaks in time series. Journal of Time Series Analysis 34(1):1–16

Aue A, Hörmann S, Horvath L, Reimherr M (2009) Break detection in the covariance structure of multivariate time series models. Annals of Statistics 37(6B):4046–4087

Bai J (1997) Estimating multiple breaks one at a time. Econometric Theory 13(3):315–352

Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. Econometrica 66(1):47–78

Berens T, Weiß G, Wied D (2015) Testing for structural breaks in correlations: Does it improve Value-at-Risk forecasting? Journal of Empirical Finance 32:135–152

Calhoun G (2016) Block bootstrap consistency under weak assumptions. Iowa State Working Paper version September 2016, online: http://gray.clhn.org/dl/calhoun_bootstrap_2016-09-28.pdf

Carrasco M, Chen X (2002) Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18(1):17–39

Cheng G (2015) Moment consistency of the exchangeably weighted bootstrap for semiparametric m-estimation. Scandinavian Journal of Statistics 42(3):665–684

Csörgö M, Horváth L (1997) Limit theorems in change-point analysis. John Wiley & Sons

Davidson J (1994) Stochastic limit theory: An introduction for econometricians. Oxford University Press

Ding Z, Engle R (2001) Large scale conditional covariance matrix modeling, estimation and testing. Academia Economic Papers 29(2):157–184

Dümbgen L (1991) The asymptotic behavior of some nonparametric change-point estimators. Annals of Statistics 19:1471–1491

Dvořák M (2017) Darling-erdös-type test for change detection in parameters and variance for stationary var models. Communications in Statistics - Theory and Methods 46(1):465–484

Dvořák M, Prášková Z (2013) On testing changes in autoregressive parameters of a var model. Communications in Statistics - Theory and Methods 42(7):1208–1226

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

Fryzlewicz P, Rao SS (2014) Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. Journal of the Royal Statistical Society, Series B 76(5):903–924

Galeano P (2007) The use of cumulative sums for detection of changepoints in the rate parameter of a poisson process. Computational Statistics and Data Analysis 51(12):6151–6165

Galeano P, Peña D (2007) Covariance changes detection in multivariate time series. Journal of Statistical Planning and Inference 137(1):194–211

Galeano P, Tsay R (2010) Shifts in individual parameters of a garch model. Journal of Financial Econometrics 8(1):122–153

Galeano P, Wied D (2014) Multiple break detection in the correlation structure of random variables. Computational Statistics and Data Analysis 76:262–282

Gooijer JD (2006) Detecting change-points in multidimensional stochastic processes. Computational Statistics and Data Analysis 51(3):1892–1903

Guralnik V, Srivastava J (1999) Event detection from time series data. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining pp 33–42

Hawkins DW (2001) Fitting multiple change-point models to data. Computational Statistics and Data Analysis 37:323–341

Inclán C, Tiao G (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association 89(427):913–923

Jandhyala V, Fotopoulos S, McNeill I, Liu P (2013) Inference for single and multiple change-points in time series. to appear in: Journal of Time Series Analysis, doi: 101111/jtsa12035

Kato K (2011) A note on moment convergence of bootstrap m-estimators. Statistics and Decisions 28:51–61

Lavielle M, Teyssière G (2006) Detection of multiple change-points in multivariate time series. Lithuanian Mathematical Journal 46:287–306

Olshen A, Venkatraman E (2004) Circular binary segmentation for the analysis of array-based dna copy number data. Biostatistics 5:557–572

Ploberger W, Krämer W, Kontrus K (1989) A new test for structural stability in the linear regression model. Journal of Econometrics 40:307–318

Vostrikova L (1981) Detecting disorder in multidimensional random processes. Soviet Mathematics Doklady 24:55–59

Wied D (2015) A nonparametric test for a constant correlation matrix. Econometric Reviews, forthcoming, doi: 101080/074749382014998152

Wied D, Krämer W, Dehling H (2012) Testing for a change in correlation at an unknown point in time using an extended functional delta method. Econometric Theory 68(3):570–589