SFB 823

Discussion Paper
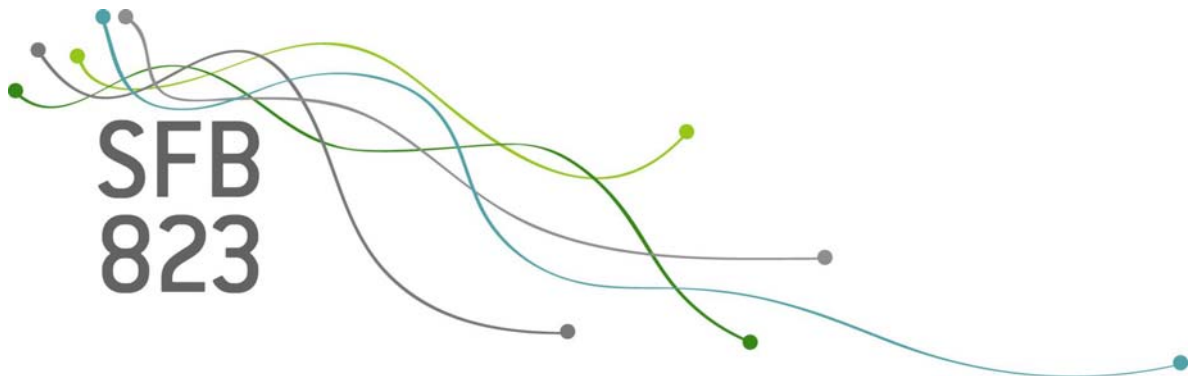
Dating multiple change points in the correlation matrix

Pedro Galeano, Dominik Wied

Nr. 18/2014

Dating multiple change points in the correlation matrix

Pedro Galeano, Dominik Wied[*]

Universidad Carlos III Madrid, TU Dortmund

This Version: May 10, 2014

Abstract

We propose a nonparametric procedure for detecting and dating multiple change points in the correlation matrix of a sequence of random variables. The procedure is based on a test for changes in correlation matrices at an unknown point in time recently proposed by Wied (2014). Although the procedure requires constant expectations and variances, only mild assumptions on the serial dependence structure are assumed. We show the validity of the procedure including the convergence rate of the change point estimators. Moreover, we illustrate its performance in finite samples by means of a simulation study and the analysis of a real data example with financial returns. These examples show that the proposed algorithm has large power in finite samples.

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

1. Introduction

The detection of change points in sequences of random variables is an important problem with a wide range of applications. Briefly, the problem can be stated as follows: a sequence of random variables has a set of characteristics, such as the mean and/or the variance, that follow a piecewise constant structure. Then, we want to detect the number of times that the characteristics change from a set of values to another and the location of the changes. Additionally, we want 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, among others. Although the literature on change point detection is vast, two recent and complete references on the topic can be found in Aue and Horváth (2013) and Jandhyala et al. (2013).

In particular, binary segmentation is one of the most flexible methodologies for change point detection. The idea of the procedure is the following: first, search 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. In particular, Wied et al.

(2012) propose a nonparametric CUSUM statistic to formally test if correlations between two random variables remain constant over time. This approach allows the practitioner to determinate if there is a change or not but he cannot determine where a possible change occurs or how many changes there are. Wied et al. (2012) fill this gap by proposing an algorithm based on the correlation constancy test to estimate both the number and the timing of possible change points. However, the previous papers only consider bivariate correlations that restricts the applicability of these procedures when more than two variables are of interest. For instance, in portfolio management, typically the interest is in more than two assets and constancy of the whole correlation matrix is of interest. Recently, Wied (2014) has 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. Wied (2014) show that the matrix-based test outperforms a method based on performing several pairwise tests and to use a level correction like Bonferroni-Holm in some situations. Moreover, Berens et al. (2013) show that the test is useful for Value at Risk (VaR) forecasting.

The main aim of this paper is to propose a nonparametric procedure for detecting and dating multiple change points in the correlation matrix of sequences of random variables based on the test proposed by Wied (2014). The proposed procedure is a binary segmentation type procedure with a refinement step to avoid the presence of false change points due to multiple testing. The method is an extension of the method proposed by Galeano and Wied (2013) for detecting multiple change points in the correlation between two random variables. 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 parts 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 added to delete all possible false change points and to better estimate their location. We analytically show that the proposed procedure gives the correct number of change points and that, assuming a finite number of change points, these are consistently esti-

mated. In addition to that, we derive the convergence rate of the change point estimator. Furthermore, we show that the algorithm gives good results in simulated samples and in an empirical application.

The rest 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 good behavior of the procedure in finite settings. Finally, Section 6 provides some conclusions. All proofs are presented in the Appendix.

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 call $||\cdot||_r$ the L_r-norm, where $r > 0$. Additionally, we write $A \sim (m, n)$ for a matrix A with m rows and n columns. We denote by $\rightarrow_d$ and $\rightarrow_p$ 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.

Given an observed time series $\mathbf{X}_1, \ldots, \mathbf{X}_T$, Wied (2014) proposed a statistic to test for the null hypothesis $H_0 : R_1 = \ldots = R_T$ versus $H_1 : \neg H_0$. The statistic is given by:

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

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

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

$\bar{X}_{i,1,k} = \frac{1}{k}\sum_{t=1}^{k}X_{i,t}$, $\bar{X}_{j,1,k} = \frac{1}{k}\sum_{t=1}^{k}X_{j,t}$, and $\hat{E}_{1,T}$ is a bootstrap estimate of

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

The bootstrap estimate $\hat{E}_{1,T}$ 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 B_i, $i = 1, \ldots, T - l_t - 1$, where l_T is a block length, such that $B_1 = (\mathbf{X}_1, \ldots, \mathbf{X}_{l_T})$, $B_2 = (\mathbf{X}_2, \ldots, \mathbf{X}_{l_T+1})$, ... Then, for some large B, we sample $\left[\frac{T}{l_T}\right]$ times with replacement one of the $T - l_T - 1$ blocks and stick the blocks together, obtaining B p-dimensional time series with length $\left[\frac{T}{l_T}\right] \cdot l_T$. Now, we calculate the vector $v_b := \sqrt{T}\left(\hat{\rho}_{b,1,T}^{ij}\right)_{1 \le i,j \le p, i < j}$, where $\rho_{b,1,T}^{ij}$ 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 derive the limiting null distribution of (1) and to obtain local power results, Wied (2014) considered some additional assumptions that are also required for our latter developments. Similar as in Galeano and Wied (2013), Assumption 1 is slightly extended by considering an arbitrary subinterval of $[0,1]$.

Assumption 1. *For an arbitrary interval $[l_1, l_2] \subseteq [0,1]$ with $l_1 < l_2$,*

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

and $S_j := \sum_{t=1\vee [l_1 T]}^{j} U_t$, we have

$$
\lim_{T\to\infty} \mathsf{E}\left(\frac{1}{(l_2 - l_1)T} S_{[l_2 T]} S'_{[l_2 T]} \right) =: D_1 \sim \left(2p + \frac{p(p-1)}{2}, 2p + \frac{p(p-1)}{2} \right),
$$

where D_1 is a finite and positive definite matrix.

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 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.*

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

with $\lim_{l\to\infty} v_l = 0$. 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 holds in many models, except perhaps the case of 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. Although this condition may be more critical, our simulation evidence shows that the proposed procedure still works in models not fulfilling this assumption (see Section 4 below). 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 is a stationarity condition which is in line with Aue et al. (2009) and that 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 procedure would remain exactly the same. Nevertheless, we investigate in our simulation study in Section 4, how the procedure behaves in finite samples in the presence of GARCH effects (volatility clustering). Finally, Assumption 5 is similar as in Calhoun (2013), Corollary 2, and guarantees that the block length becomes large but not too large compared to T. Under H_0 and assumptions 1, 2, 3, 4 and 5, Corollary 1.b in Wied (2014) shows that the statistic $A_{1,T}$ converges in distribution to the supremum of the sum of the absolute value of independent Brownian Bridges. More precisely, H_0 si 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 assumed for easiness in exposition that expectations and variances are constant. Additionally, we assume that under the alternative hypothesis, we have a piecewise constant second-order cross moment matrix $\mathsf{E}\left(X_t X_t'\right)$. The formal assumption is:

Assumption 6. *Under the 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 $i = 1, \ldots, p$, while the second-order cross moments matrix changes from $\mathsf{E}\left(X_t X_t'\right) = m_{XX'}$ to $\mathsf{E}\left(X_t X_t'\right) = m_{XX'} + g\left(\frac{t}{T}\right)$. The function $g(z)$, $z \in [0,1]$ 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 matrices $a_0, \ldots, a_\ell$ 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.

Next, we present the proposed procedure with the goal of estimating ℓ, $z_1, \ldots, z_\ell$ and $a_0, \ldots, a_\ell$. The procedure is divided in four steps. The first and second steps of the algorithm are basically the steps taken by the usual binary segmentation algorithm. The idea of these steps is to isolate each change point in different time intervals by splitting the series into two parts once a change point is found. Then, the search of a new change point is repeated in both sections. These two steps are iterated until no more change points

are detected. The third step of the algorithm is a refinement step in which the detected change points are tested in subintervals containing a single change point. Additionally, this step allows to better estimate the true locations of the change points. The final step computes the correlation matrix in subintervals with constant (unconditional) correlation. To establish the asymptotic results, it is better to present the procedure in terms of the estimator of the change point fraction. To that purpose, we rewrite the test statistic (1) as

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

with $\tau(z) = [2+z(T-2)]$ (where $[\cdot]$ is the floor function), $P_{\tau(z),1,T} = \left(\hat{\rho}_{1,\tau(z)}^{ij} - \hat{\rho}_{1,T}^{ij} \right)_{1 \leq i,j \leq p, i<j} \in \mathbb{R}^{\frac{p(p-1)}{2}}$, and estimate the timing of the break by $\hat{z} := \tau(\hat{z}^*)/T$ with $\hat{z}^* := \mathsf{argmax}_z B_{1,T}(z)$ and $B_{1,T}(z) := \frac{\tau(z)}{T} \left\| P_{\tau(z),1,T} \right\|_1$. Note that we exclude 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"

$$A_{\eta(l_1),\xi(l_2)}(z) := \frac{\xi(z) - \eta(l_1) + 1}{\sqrt{\xi(l_2) - \eta(l_1) + 1}} \left\| \hat{E}_{\eta(l_1),\xi(l_2)}^{-1/2} P_{\tau(z),\eta(l_1),\xi(l_2)} \right\|_1, \tag{4}$$

where, $z \in [l_1, l_2]$ for $0 \leq l_1 < l_2 \leq 1$, $\eta(z) = ([zT] \vee 1) \wedge (T-1)$, $\xi(z) = \eta(z) \vee (\eta(l_1)+1)$, $\hat{E}_{\eta(l_1),\xi(l_2)}$ denotes the bootstrap estimate of E in (2) using data from $\eta(l_1)$ to $\xi(l_2)$ and $P_{\tau(z),\eta(l_1),\xi(l_2)} = \left(\hat{\rho}_{\eta(l_1),\tau(z)}^{ij} - \hat{\rho}_{\eta(l_1),\xi(l_2)}^{ij} \right)_{1 \leq i,j \leq p, i<j} \in \mathbb{R}^{\frac{p(p-1)}{2}}$. Then the timing of break is estimated by

$$\hat{z} := \xi(\hat{z}^*)/T \tag{5}$$

with $\hat{z}^* = \text{argmax}_{l_1 \le z \le l_2} B_{\eta(l_1),\xi(l_2)}(z)$ and

$$B_{\eta(l_1),\xi(l_2)}(z) := \frac{\xi(z) - \eta(l_1) + 1}{\xi(l_2) - \eta(l_1) + 1} \left|\left|P_{\tau(z),\eta(l_1),\xi(l_2)}\right|\right|_1 , \tag{6}$$

Basically, this means that we always look for the time point at which the test statistic (4) (calculated from data in a particular interval) takes its maximum and divide by T. Under the null hypothesis of no correlation change, $\hat{E}_{1,T}$ converges to a positively definite matrix with respect to the product measure $\mathsf{P}^\times$ which combines randomness from the data as well from the bootstrap (see the proof of Corollary 1.b in Wied, 2014). So, it is reasonable to assume that the matrix $\hat{E}_{\eta(l_1),\xi(l_2)}$ is invertible. It is unclear what happens with this matrix under fixed alternatives as considered below. However, in order to ensure consistency of the test statistic, we impose the convention that $\hat{E}_{\eta(l_1),\xi(l_2)}$ 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 $\hat{E}_{\eta(l_1),\xi(l_2)}^{-1/2}$ 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 (5) 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

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

where $A_{\eta(z_{k-1}+\frac{1}{T}),\xi(z_k)}$ is the value of the statistic calculated from the data from $\eta(z_{k-1} + \frac{1}{T})$ to $\xi(z_k)$, 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 $A_{\eta(z_{k_{\max}-1}+\frac{1}{T}),\xi(z_{k_{\max}})}$ is attained, where:

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

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 $\xi(z_{k+1})$, 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 taken by 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 a single change point, which is not guaranteed in step 2.

A key issue in applying the procedure to real data is the selection of the critical level used in the algorithm. A possibility is to use always the same critical value in each step of the procedure. However, the use of the same critical level in steps 2 and 3 may lead

to over-estimation of the number of change points, because 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 $1 - \alpha_0 = (1 - \alpha_k)^{k+1}$, leading to $\alpha_k = 1 - (1 - \alpha_0)^{\frac{1}{k+1}}$, that keeps the same significance level constant for all tests. For instance, if $\alpha_0 = 0.05$, $\alpha_1 \approx 0.025$, $\alpha_2 \approx 0.017$ and so on. In fact, the initial type I error in the asymptotic result concerning the number of break points (Theorem 2) would have to converge to zero. However, in finite samples, $\alpha_0 = 0.05$ seems to be an acceptable choice. 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

As in Galeano and Wied (2013), we assume that there are dominating change points (if any) in order to obtain analytic results.

Assumption 7. *For any $0 \le l_1 < l_2 \le 1$, the function*

$$P_{l_1,l_2}(z) := \left|\left| P^*_{l_1,l_2}(z) \right|\right|_1$$

12

with

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

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.

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. The ideas of the proofs are similar to those in Galeano and Wied (2013) and Wied (2014), whereas we additionally have to ensure that the bootstrap variance estimator $\hat{E}^{-1/2}_{\eta(l_1),\xi(l_2)}$ is positively definite if there is a dominating change point between l_1 and l_2.

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* (5) *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, as in Galeano and Wied (2013), 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 is excluded from 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* $\epsilon > 0$, *there is a* $M > 0$ *such that*

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

Interestingly, the convergence rate is not $\sqrt{T}$, which would hold within e.g. the central limit theorem. This is true despite the fact that the change point estimator is based on the argmax which is also true for the usual maximum likelihood estimator. An intuition behind this can be obtained in a simple bivariate model with independent, bivariately normally distributed random variables and one change point in which the change point fraction z is the only parameter. The likelihood in this model is given by

$$l(z|(X_1, Y_1), \ldots, (X_T, Y_T)) = \prod_{i=1}^{T}(f_{1,i}((X_i, Y_i))\mathbf{1}_{i \leq [Tz]} + f_{1,i}((X_i, Y_i))\mathbf{1}_{i > [Tz]}),$$

where $f_{1,i}$ and $f_{2,i}$ are the densities of (X_i, Y_i) before and after the break, respectively. Here, one directly sees that the likelihood is not differentiable in the change point such that standard regularity conditions for a $\sqrt{T}$-asymptotic do not hold.

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 3, especially for the fact that the number of change points is not overestimated asympotically. In fact, one also needs such a condition in the proof of Theorem 2 in Galeano and Wied (2013) (although it is not explicitly stated there).

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.

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* $k \in \mathbb{N}_0$.

Theorem 3. *Under Assumptions 1, 2, 3, 4, 5, 6, 7 and 8, the change point algorithm asymptotically gives the correct number of change points ℓ and the change points are consistently estimated.*

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

14

$\mathsf{E}(X_t X_t') = m_{XX'} + g\left(\frac{t}{T}\right)$ by $\mathsf{E}(X_t X_t') = m_{XX'} + \frac{1}{\sqrt{T}}g\left(\frac{t}{T}\right)$. In this setting, Wied (2014) provides local power results (compare his Corollary 2). 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 $\mathsf{E}(X_t X_t') = m_{XX'} + g\left(\frac{t}{T}\right)$ replaced by $\mathsf{E}(X_t X_t') = m_{XX'} + \frac{1}{\sqrt{T}}g\left(\frac{t}{T}\right)$) 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 (5) that*

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

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

4. SIMULATION EVIDENCE

In this section, we present several Monte Carlo experiments to illustrate the performance of the proposed algorithm in finite samples. In particular, we focus our attention to 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 by the proposed procedure 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 of this section and the real data example of Section 5 is $p = 4$, we obtain the estimated quantiles by generating 100000 sets 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, finally, 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 easily estimate the quantiles. For instance, the first five quantiles used in steps 1 and 2 of the procedure, if needed, are given by 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. The plot suggests that the asymptotic distribution is slightly positive skewed.

Figure 1 about here

In the simulations that we report below, we consider several variants of the scalar BEKK model proposed by Ding and Engle (2001). There are two main reasons to use this model. First, our main fields of application of the proposed procedure are financial returns and BEKK models are one 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 easily 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:

$$\mathbf{X}_t = \mathbf{H}_t^{1/2}\mathbf{E}_t$$

$$\mathbf{H}_t = (1 - \alpha - \beta)\,\mathbf{H} + \alpha\mathbf{X}_{t-1}\mathbf{X}_{t-1}' + \beta\mathbf{H}_{t-1}$$

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 $\alpha + \beta < 1$, 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 of $\mathbf{X}_t$, that are the elements in

16

the main diagonal of $\mathbf{H}$. In particular, we take $\alpha = 0.14$ and $\beta = 0.85$ to reflect volatility persistence, and $\mathbf{D} = \mathbf{I}_4$, so that $\mathbf{H} = \mathbf{R}$, with:

$$
\mathbf{R} = \begin{pmatrix}
1 & 0.5 & 0.6 & 0.7 \\
0.5 & 1 & 0.5 & 0.6 \\
0.6 & 0.5 & 1 & 0.5 \\
0.7 & 0.6 & 0.5 & 1
\end{pmatrix}.
\tag{7}
$$

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 3 degrees of freedom. The latter distribution represents a rather extreme situation not covered by our assumptions but we analyze this case to see the performance of the procedure in settings which can hold 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 $l_{3000} = 7$, respectively. The number of bootstrap replications is $B = 1000$. Table 1 gives the results based on 1000 replications 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 distorsions for the standardized Student-t with 3 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.

Table 1 about here

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 (7) and then changes at z_1 to:

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0.7 & 0.6 & 0.5 \\ 0.5 & 1 & 0.7 & 0.6 \\ 0.6 & 0.5 & 1 & 0.7 \\ 0.7 & 0.6 & 0.5 & 1 \end{pmatrix}. \tag{8}$$

Note that the largest correlation change is of magnitude .2, while two of the correlations do not change at all. We believe that this setting seems quite reasonable in practice. Table 2 shows the relative frequency detection of zero, one and more than one change points. It is seen 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 3 degrees of freedom. 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 detection frequency of the procedure unless the sample size is small. In this latter case, the procedure detects more frequently the change point at the middle of the series. This is in coincidence with other procedures relying in CUSUM statistics as the one used here. Finally, in most cases, the percentage of false detection is always quite close to the nominal 5%, specially 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 in each case. 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.

Table 2 about here

18

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 (7), then changes to the correlation matrix in (8), and, finally, changes again to the correlation matrix in (7) at the second change point. Note that in this scenario there is no a dominant change point but we prefer to consider this situation to show that the procedure works well also in this case. Table 4 shows the relative frequency detection of zero, one, two and more than two changes. As in the case of a single change point, the proposed procedure works reasonably well, especially when the sample size gets larger. In addition, the frequency of over-detection is small. It may underestimate the number of change points, however, 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. 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.

5. EMPIRICAL APPLICATION

In this section, we illustrate the performance of the proposed procedure with a financial time series. For this, we look for changes in the correlation structure of the daily simple return series of four stocks. Specifically, as in Wied (2014), we consider four European

companies, Total, Sanofi, Siemens and BASF from January 1, 2007 to June 1, 2012 consisting of $T = 1414$ data points. The data was obtained from the database Datastream. The four return series are plotted in Figure 1, 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 2 show 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.

Figure 1 about here

Figure 2 about here

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 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. Then, 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 were found in the three subintervals $[1, 134]$, $[135, 443]$, and $[444, 1414]$. Then, 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.

Table 6 about here

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, and finally 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, which is very close to the second estimated change point.

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 increases

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 phenomenon kwnon as "correlation meltdown" that affirms that in times of crisis, the correlations between financial returns often increases.

Table 7 about here

6. CONCLUSIONS

This paper has proposed a procedure for detecting change points in the correlation matrix of a sequence of multiple random variables. The procedure is based on a CUSUM test statistic proposed by Wied (2014). The asymptotic distribution of the test statistic is the one of the supremum of the maximum of the absolute value of independent Brownian bridges. We have shown that, under certain circumstances, the procedure consistently detects the true number and location of the change points. The finite sample behavior of the procedure has been analyzed via several simulation studies and illustrated with the analysis of a four dimensional time series of simple returns of four European companies. The real data example suggests that the procedure detects changes at points that fare well with external events affecting the financial markets.

An alternative method to detect changes in the correlation matrix of a sequence of random variables is to consider each entry in a higher dimensional correlation matrix separately to determine whether there have been changes in the individual correlations. However, if the dimension is large enough, the number of comparisons between correlations can be huge. Consequently, we think that the use of the whole correlation matrix is a more elegant and compact way to detect these kind of changes.

Finally, it might be interesting to consider a more sophisticated algorithm that includes modifications of the standard binary segmentation procedure such as the ones introduced

22

in Fryzlewicz (2014) to increase the power of the procedure in small samples. However, the modification made in Fryzlewicz (2014) is not straightforward in the case of changes in the correlation matrix.

References

ANDREOU, E. AND E. GHYSELS (2002): "Detecting multiple breaks in financial market volatility dynamics," *Journal of Applied Econometrics*, 17(5), 579–600.

AUE, A., S. HÖRMANN, L. HORVATH, AND M. REIMHERR (2009): "Break detection in the covariance structure of multivariate time series models," *Annals of Statistics*, 37(6B), 4046–4087.

AUE, A. AND L. HORVÁTH (2013): "Structural breaks in time series," *Journal of Time Series Analysis*, 34(1), 1–16.

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

BAI, J. AND P. PERRON (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66(1), 47–78.

BERENS, T., G. WEISS, AND D. WIED (2013): "Testing for structural breaks in correlations: Does it improve Value-at-Risk forecasting?" *SSRN working paper, online: http://ssrn.com/abstract=2265488*.

CALHOUN, G. (2013): "Block bootstrap consistency under weak assumptions," *Iowa State Working Paper*, version 25.03.2013, http://www.econ.iastate.edu/sites/default/files/publications/papers/p14313-2011-09-23.pdf.

CARRASCO, M. AND X. CHEN (2002): "Mixing and moment properties of various GARCH and stochastic volatility models," *Econometric Theory*, 18(1), 17–39.

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

DING, Z. AND R. ENGLE (2001): "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing," *Academia Economic Papers*, 29(2), 157–184.

FRYZLEWICZ, P. (2014): "Wild binary segmentation for multiple change-point detection," *Working paper, London School of Economics.*

FRYZLEWICZ, P. AND S. S. RAO (2014): "Multiple-change-point detection for autoregressive conditional heteroscedastic processes," *to appear in: Journal of the Royal Statistical Society, Series B, doi: 10.1111/rssb.12054.*

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. AND R. TSAY (2010): "Shifts in individual parameters of a GARCH model," *Journal of Financial Econometrics*, 8(1), 122–153.

GALEANO, P. AND D. WIED (2013): "Multiple break detection in the correlation structure of random variables," *to appear in: Computational Statistics and Data Analysis, online: http://dx.doi.org/10.1016/j.csda.2013.02.031.*

GOOIJER, J. D. (2006): "Detecting change-points in multidimensional stochastic processes," *Computational Statistics and Data Analysis*, 51(3), 1892–1903.

INCLÁN, C. AND G. TIAO (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., S. FOTOPOULOS, I. MCNEILL, AND P. LIU (2013): "Inference for single and multiple change-points in time series," *to appear in: Journal of Time Series Analysis, doi: 10.1111/jtsa12035.*

VOSTRIKOVA, L. (1981): "Detecting disorder in multidimensional random processes," *Soviet Mathematics Doklady*, 24, 55–59.

WIED, D. (2014): "A nonparametric test for a constant correlation matrix," *SSRN working paper, online: http://ssrn.com/abstract=2372748.*

WIED, D., W. KRÄMER, AND H. DEHLING (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.

A. APPENDIX

A.1. Proofs

Proof of Theorem 1

Applying the proof of Theorem 1 in Galeano and Wied (2013) component by component and making use of the fact that a variance estimator like $\hat{E}_{\eta(l_1),\xi(l_2)}^{-1/2}$ is excluded, one shows that, uniformly on $[l_1, l_2]$,

$$B_{\eta(l_1),\xi(l_2)}(z) \rightarrow_{a.s.} CP_{l_1,l_2}(z)$$

for a certain constant C. Then, the theorem follows by applying the argmax continuous mapping theorem. ∎

Proof of Theorem 2

We assume without loss of generality that there is only one change point in $k_0 = [Tz_0]$ and that $[l_1, l_2] = [0, 1]$. Denote $P_{0,1}(z) =: P(z)$. Then $P(z)$ has a unique maximum in z_0. Similarly as in the proof of Proposition 2 in Bai (1997), we show that

$$\mathsf{P}_{M,T} := \mathsf{P}\left(\max_{|k-k_0|>M} B_{1,T}\left(\frac{k}{T}\right) - B_{1,T}\left(\frac{k_0}{T}\right) \geq 0\right)$$

becomes small for large M and T. That means that, for every $\epsilon > 0$, there is a $M > 0$ and a $T_0 > 0$ such that, for all $T > T_0$, $\mathsf{P}_{M,T} < \epsilon$.

Now, $B_{1,T}\left(\frac{k}{T}\right) - B_{1,T}(\frac{k_0}{T}) \geq 0$ is equivalent to

$$\left(B_{1,T}\left(\frac{k}{T}\right) - P\left(\frac{k}{T}\right)\right) - \left(B_{1,T}\left(\frac{k_0}{T}\right) - P\left(\frac{k_0}{T}\right)\right) + \left(P\left(\frac{k}{T}\right) - P\left(\frac{k_0}{T}\right)\right) \geq 0.$$

Assume for the moment $k > k_0$ and that the standard deviations of all random variables are equal to 1. (Divide each component of $P(\cdot)$ by the standard deviations if the latter assumption is not fulfilled.) We multiply the whole equation with $T/(k - k_0)$ and denote

$$A_1(k, k_0, T) = \frac{T}{k - k_0}\left(\left(B_{1,T}\left(\frac{k}{T}\right) - P\left(\frac{k}{T}\right)\right) - \left(B_{1,T}\left(\frac{k_0}{T}\right) - P\left(\frac{k_0}{T}\right)\right)\right)$$

$$A_2(k, k_0, T) = P\left(\frac{k}{T}\right) - P\left(\frac{k_0}{T}\right).$$

Now, we use several observations in order to argue that the asymptotic behavior of $A_1(k, k_0, T)$ can be reduced to the behavior of

$$\sum_{1 \leq i < j \leq p} \frac{1}{k - k_0} \sum_{t=k_0}^{k} (X_{t,i}X_{t,j} - \mathsf{E}(X_{t,i}X_{t,j})).$$

This quantity is then arbitrarily small by the law of large numbers for sufficiently large M. The observations are the following:

1. $A_1(k, k_0, T)$ can be regarded as the sum of $p(p-1)/2$ components and each component can be treated separately.

2. For large T, with high probability and uniformly in $z \in \{k/T, k_0/T\}$, all components of $TP_{\tau(z),1,T}$ and $TP_{0,1}^*(z)$ have the same sign so that we do not need the absolute values.

3. For large T, with high probability and uniformly in $z \in \{k/T, k_0/T\}$, the successive variances in the denominators of the components of $TP_{\tau(z),1,T}$ are equal to their theoretical counterparts which are the same for $z = k/T$ and $z = k_0/T$, respectively.

4. In the nominators, we have expressions like

$$\frac{1}{k - k_0} \left(\left(\sum_{t=k_0}^{k} X_{t,i} X_{t,j} \right) - (k - k_0) \left(\sum_{t=1}^{T} X_{t,i} X_{t,j} \right) - \left(P_{i,j}\left(\frac{k}{T}\right) - P_{i,j}\left(\frac{k_0}{T}\right) \right) \right.$$

$$\left. - \left(\frac{1}{k} \sum_{t=1}^{k} X_{t,i} \sum_{t=1}^{k} X_{t,j} - \frac{1}{k_0} \sum_{t=1}^{k_0} X_{t,i} \sum_{t=1}^{k_0} X_{t,j} - (k - k_0)\frac{1}{T} \sum_{t=1}^{T} X_{t,i} \frac{1}{T} \sum_{t=1}^{T} X_{t,j} \right) \right).$$

Here, $P_{i,j}(\cdot)$ are the components of $P(\cdot)$.

5. One can uniformly approximate $\int_0^z g(t)dt$, that means,

$$\lim_{T \to \infty} \sup_{z \in [0,1]} \left\| \int_0^z g(t)dt - \frac{1}{T} \sum_{t=1}^{[zT]} X_t X_t' \right\|_1 = 0.$$

Then, after some tedious calculations, one sees that $A_1(k, k_0, T)$ is a random variable such that, for all $\epsilon > 0$ and all $\eta > 0$, there is a $M > 0$ such that $\mathsf{P}(|A_4(k, k_0, T)| > \epsilon) < \eta$ for $k > k_0 + M$ and $T > T_0$. This means that $A_4(k, k_0, T)$ is arbitrarily small whenever T and M are large. On the other hand, $A_2(k, k_0, T)$ does not converge to zero: $P\left(\frac{k}{T}\right) - P\left(\frac{k_0}{T}\right)$ is a finite sum of linear functions in k with negative slope (see Figure 1 in Galeano and Wied (2013)) so that it is equal to $C\left(\frac{k}{T} - \frac{k_0}{T}\right)$ for a $C < 0$ by Taylor's formula. Multiplied with $T/(k - k_0)$, the expression is equal to C. Then, with large probability, $A_1(k, k_0, T) + A_2(k, k_0, T)$ is strictly negative. For $k < k_0$, the argument is similar and the theorem is proven. ∎

Proof of Theorem 3

Denote $Q_T^{l_1, l_2} := \sup_{z \in [l_1, l_2]} A_{\eta(l_1), \xi(l_2)}(z)$ the test statistic calculated from data from $\eta(l_1)$ to $\xi(l_2)$. Now, by Theorem 1, it holds

$$\frac{1}{a_T^k} B_{\eta(l_1), \xi(l_2)}(z) \to_p \infty$$

for any sequence $a_T^k = o\left(\sqrt{T}\right)$ if there is a change point in the interval $[l_1, l_2]$. Moreover, due to Assumption 2, the eigenvalues of $\hat{E}_{\eta(l_1), \xi(l_2)}$ are bounded away from ∞. Conse-

quently, the eigenvalues of $\hat{E}_{\eta(l_1),\xi(l_2)}^{-1/2}$ (remember that we assume its existence) are bounded away from zero and the matrix is positively definite. Therefore,

$$\frac{1}{a_T^k} Q_T^{l_1,l_2} \to_p \infty$$

(with respect to the measure $\mathsf{P}^{\times}$) for any sequence $a_T^k = o\left(\sqrt{T}\right)$ if there is a change point in the interval $[l_1, l_2]$. By Theorem 2, we moreover have

$$\frac{1}{a_T^k} Q_T^{\hat{z}_i,\hat{z}_{i+1}} \to_p \infty$$

(with respect to the measure $\mathsf{P}^{\times}$), where $\hat{z}_i$ and $\hat{z}_{i+1}$ for $i \in \mathbb{N}_0$ are two estimated change points in one of the iterations of the algorithms, as long as there is a change point in the interval $[z_i, z_{i+1}]$. This follows from the fact, that, by Theorem 2,

$$Q_T^{\hat{z}_i,\hat{z}_{i+1}} - Q_T^{z_i,z_{i+1}} = o_{\mathsf{P}^{\times}}\left(\frac{1}{\sqrt{T}}\right).$$

Moreover, with the same argument,

$$Q_T^{\hat{z}_i,\hat{z}_{i+1}} = O_{\mathsf{P}^{\times}}(1)$$

if there is no change point in the interval $[z_i, z_{i+1}]$.

With this result, we can proceed as in the proof of Theorem 2 in Galeano and Wied (2013). ∎

Proof of Theorem 4

This follows similary to the Proof of Theorem 3 in Galeano and Wied (2013), again making use of the fact that a variance estimator like $\hat{E}_{\eta(l_1),\xi(l_2)}^{-1/2}$ is excluded. ∎

Table 1: Relative frequency detection of 0 and more than 0 change points with the scalar BEKK model with a initial nominal significant level of $\alpha_0 = 0.05$.

	Gaussian		Student-t_3	
T	0	≥ 1	0	≥ 1
500	.944	.056	.927	.073
1000	.952	.048	.955	.045
2000	.943	.057	.941	.059
3000	.952	.048	.940	.060
4000	.957	.043	.954	.046

Table 2: Relative frequency detection of 0, 1 and more than 1 change points with the scalar BEKK model with a single change point and with a 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
	500	.044	.913	.021	.001	.961	.038	.045	.915	.040
	1000	.004	.935	.044	.000	.949	.051	.000	.944	.056
Gaussian	2000	.000	.939	.041	.000	.937	.063	.000	.934	.066
	3000	.002	.941	.057	.000	.939	.061	.001	.934	.065
	4000	.002	.945	.053	.000	.949	.051	.000	.946	.054
	T	0	1	≥ 2	0	1	≥ 2	0	1	≥ 2
	500	.558	.426	.016	.346	.630	.024	.629	.365	.006
	1000	.255	.715	.030	.051	.909	.040	.271	.705	.024
Student-t_3	2000	.020	.922	.058	.002	.936	.062	.018	.925	.057
	3000	.001	.949	.050	.001	.940	.059	.003	.942	.055
	4000	.000	.941	.059	.001	.944	.057	.003	.941	.056

Table 3: Median and MAD of the change point estimators for the results in Table 2.

	$z_1 = .25$		$z_1 = .50$		$z_1 = .75$	
	Gaussian	Student-t_3	Gaussian	Student-t_3	Gaussian	Student-t_3
T	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)	Median (Mad)
500	.282 (.0593)	.342 (.1660)	.502 (.0326)	.504 (.1097)	.724 (.0593)	.628 (.1957)
1000	.257 (.0237)	.303 (.1008)	.500 (.0163)	.502 (.0607)	.746 (.0222)	.711 (.0830)
2000	.253 (.0118)	.270 (.0500)	.500 (.0088)	.503 (.0300)	.748 (.0107)	.740 (.0429)
3000	.252 (.0074)	.259 (.0286)	.500 (.0064)	.501 (.0227)	.748 (.0074)	.745 (.0289)
4000	.251 (.0063)	.254 (.0203)	.500 (.0048)	.501 (.0187)	.749 (.0063)	.746 (.0196)

Table 4: Relative frequency detection of 0, 1, 2 and more than 2 change points with the scalar BEKK model with two change points and with a initial nominal significant level of $\alpha_0 = 0.05$.

			$(z_1, z_2) = (.35, .7)$					
	Gaussian				Student-t_3			
T	0	1	2	≥ 3	0	1	2	≥ 3
500	.604	.076	.305	.015	.856	.125	.019	.000
1000	.095	.006	.836	.063	.730	.144	.121	.005
2000	.000	.000	.927	.069	.369	.053	.527	.051
3000	.000	.004	.937	.063	.145	.014	.779	.062
4000	.000	.001	.942	.057	.031	.004	.903	.062

Table 5: Median and MAD of the change point estimators for the results in Table 4.

			$(z_1, z_2) = (.35, .7)$	
	Gaussian		Student-t_3	
T	Mean $(\widehat{z}_1)$ (Mad)	Mean $(\widehat{z}_2)$ (Mad)	Mean $(\widehat{z}_1)$ (Mad)	Mean $(\widehat{z}_2)$ (Mad)
500	.348 (.0296)	.700 (.0326)	.348 (.0652)	.728 (.0593)
1000	.351 (.0222)	.699 (.0177)	.351 (.0489)	.702 (.0429)
2000	.350 (.0096)	.699 (.0096)	.353 (.0303)	.701 (.0363)
3000	.350 (.0069)	.700 (.0064)	.351 (.0247)	.700 (.0207)
4000	.350 (.0044)	.700 (.0048)	.351 (.0177)	.700 (.0170)

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	$\begin{pmatrix} 1 & 0.1564 & 0.3275 & 0.4917 \\ 0.1564 & 1 & 0.0708 & 0.2261 \\ 0.3275 & 0.0708 & 1 & 0.3604 \\ 0.4917 & 0.2261 & 0.3604 & 1 \end{pmatrix}$			
Second	$\begin{pmatrix} 1 & 0.3907 & 0.5148 & 0.5677 \\ 0.3907 & 1 & 0.4352 & 0.4924 \\ 0.5148 & 0.4352 & 1 & 0.5110 \\ 0.5677 & 0.4924 & 0.5110 & 1 \end{pmatrix}$			
Third	$\begin{pmatrix} 1 & 0.5990 & 0.6924 & 0.6986 \\ 0.5990 & 1 & 0.5159 & 0.5168 \\ 0.6924 & 0.5159 & 1 & 0.7857 \\ 0.6986 & 0.5168 & 0.7857 & 1 \end{pmatrix}$			

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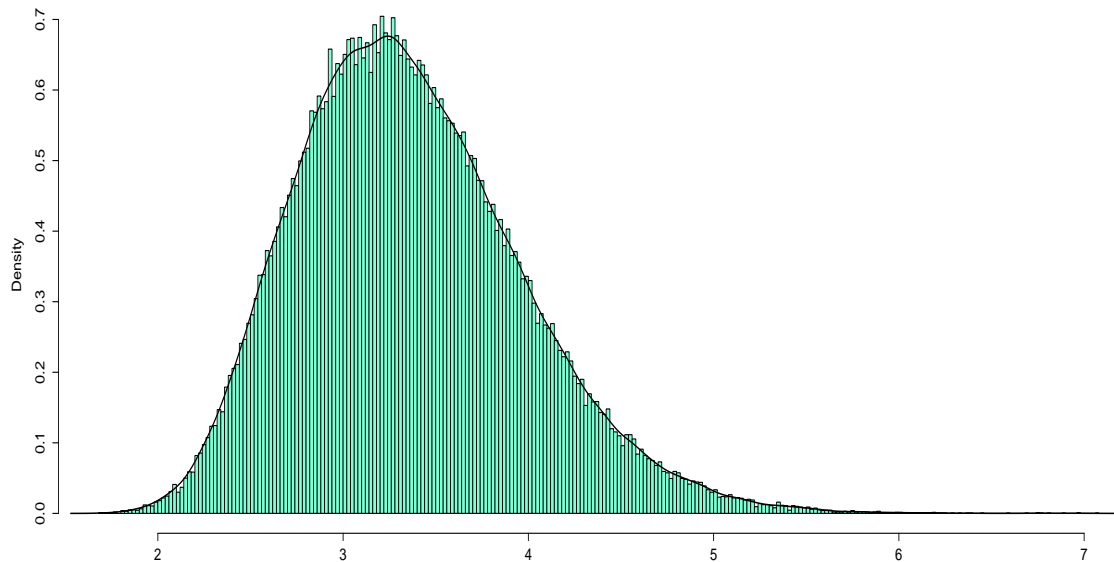

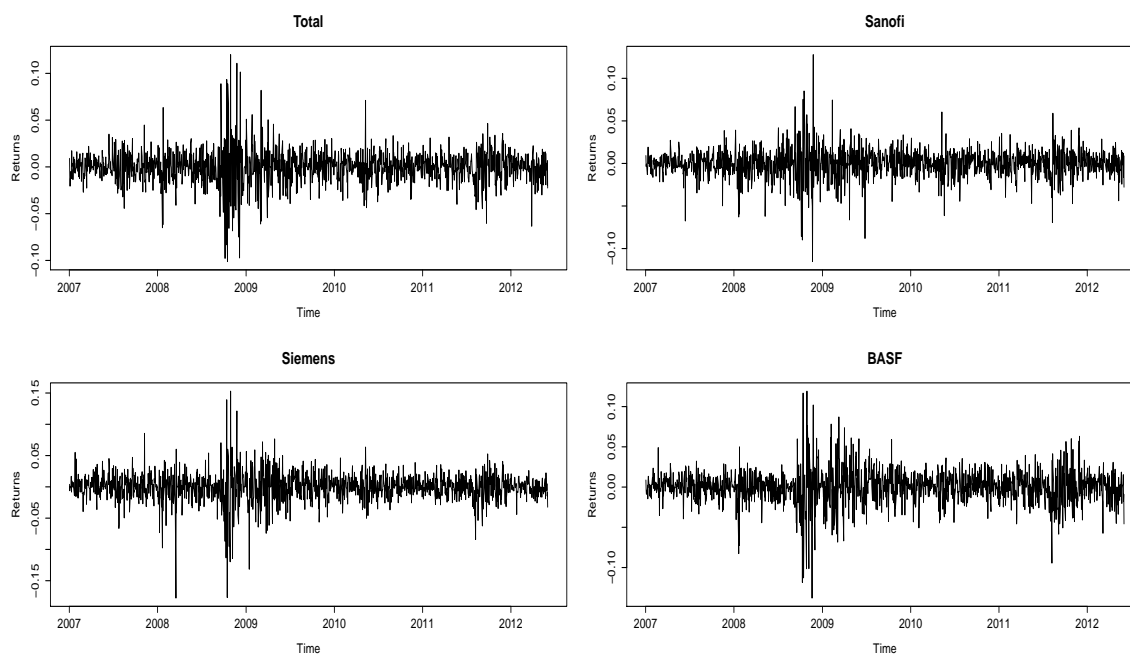Figure 2: Daily simple returns of Total, Sanofi, Siemens and BASF

Figure 3: Rolling correlations for the daily simple returns of Total, Sanofi, Siemens and BASF