Testing Constant Cross-Sectional Dependence with Time-Varying Marginal Distributions in Parametric Models *

Matthias Kaldorf † Dominik Wied ‡
University of Cologne University of Cologne

September 20, 2020

Abstract

This paper proposes parametric two-step procedures for assessing the stability of cross-sectional dependency measures in the presence of potential breaks in the marginal distributions. The procedures are based on formerly proposed sup-LR tests in which restricted and unrestricted likelihood functions are compared with each other. First, we show theoretically that standard asymptotics do not hold in this situation. We propose a suitable bootstrap scheme and derive test statistics in different commonly used settings. The properties of the test statistics and precision of the associated change-point estimator are analyzed and compared with existing non-parametric methods in various Monte Carlo simulations. These studies reveal advantages in test power for higher-dimensional data and an almost uniform superiority of the sup-LR test in terms of precision of the change-point estimator. We then apply this method to equity returns of European banks during the financial crisis of 2008.

Key words: Cumulated Sums; Empirical Copula; sup-LR Test; Structural Break; Two-Step Procedure

JEL classification: C12 (Hypothesis Testing), C58 (Financial Econometrics)

1 Introduction

Testing stability of cross-sectional dependence in multivariate time series models has received considerable attention over recent years, both in terms of methodological advance and in applications. In financial econometrics, those methods find application to asset price data subject to financial crisis or policy shocks. In the context of financial crisis, this phenomenon is usually called *shift contagion* and has been formally analyzed first by King and Wadhwani [1990] who use recursively calculated sample correlations to assess stability of the correlation

*We are grateful to a referee, Anne-Florence Allard, Julien Chevallier, Hans Manner and participants at the 2018 Paris Financial Management Conference, 2019 Quantitative Finance Workshop (ETH Zurich), 2019 Econometric Society European Meeting (University of Manchester) for useful comments and suggestions.

†Center for Macroeconomic Research, University of Cologne, Germany, email: kaldorf@wiso.uni-koeln.de

‡Institute for Econometrics and Statistics, University of Cologne, Germany, email: dwied@uni-koeln.de.

over the considered sample. In an important contribution, Forbes and Rigobon [2002] stress that in equity markets increases in volatility of some equity market often precede an increase in correlation (or some other dependency measure). Therefore, before testing for constant correlation, potential changes in variances have to be taken into account. These procedures test constancy of the marginal distributions in a first step, eliminate potential structural changes in the margins by suitable transformations and then test constancy of the cross-sectional dependence in step two. We will call such procedures *two-step testing procedures* in the following.

In general, there exist two fundamentally different approaches in a structural change context: likelihood-ratio-type tests that rely on some parametric model and tests imposing hypothesis on moments or quantile exceedance-probabilities, that use cumulated sums of empirical counterparts.

A seminal contribution for the first approach is Andrews [1993] who derived asymptotic tests for (partial) structural changes in a generalized method of moments framework. One class of these tests are supremum likelihood ratio type (sup-LR-type) tests. In a multivariate model, parameters are partitioned into those that change under the null hypothesis of constancy and the alternative and nuisance-parameters that are invariant under null and alternative hypothesis. For any change-point candidate, the sample is divided into two sub-samples and parameter stability is rejected, if the difference between two GMM objective functions becomes too large. This method has first been applied in the context of constant correlation by Dias and Embrechts [2004].

Within the latter class one can distinguish tests imposing constancy on cross-moments of the multivariate system and tests imposing constancy of the copula. Stability is rejected if the fluctuations in the cumulated sums of their empirical counterparts exceed certain critical values. Important contributions in an econometric context include Aue et al. [2009] for covariances, Wied et al. [2012b] for correlations, Bücher et al. [2014] and Rémillard [2017] for copulas.

In both frameworks, two-step procedures have been proposed. While the latter framework was tackled in Demetrescu and Wied [2019], Blatt et al. [2015] worked in the first framework by analyzing shift contagion using a multivariate normal distribution. Our aim is to continue the work in Blatt et al. [2015] by deriving appropriate sup-LR-type statistics in different parametric models, which are typically used for financial time series data. The motivation for using such tests is that, if the assumed model is correct and different from a normal distribution model, a parametric test might have higher power than a non-parametric test. To the best of our knowledge, test statistics for different multivariate distributions have not been explored in the literature before.

Critical values for the tests are obtained by bootstrap approximations. In particular, these bootstrap approximations are used because it cannot be expected that the usage of transformed/standardized data (using piecewise constant variance estimators, GARCH-residuals or empirical cumulative distribution functions) in step 2 leaves the asymptotics unaltered. While Demetrescu and Wied [2019] derived analytic results for the residual effects in non-parametric models (see also Duan and Wied, 2018), the first contribution of this paper is to quantify the impact of transforming the original data by Monte Carlo simulations.

Secondly, after showing that the residual effect matters quantitatively in commonly used models, we move to compare power of the different approaches after correcting for the invalidity of standard asymptotics. Moreover, the Monte Carlo study compares the ability of parametric and non-parametric procedures to detect and date structural changes in the sequential setup we are interested in. Our simulation study extends simulation results in Galeano and Peña [2007] who compare Gaussian sup-LR and fluctuation tests in the case of variance/covariance changes. Apart from Demetrescu and Wied [2019], we also include the recently proposed tests for constant copulas by Bücher et al. [2014] (embedded in the two-step-procedure) in our comparison.

If test and estimation procedures are constructed with applications to finance or macroeconomics in mind, it is natural to study their behaviour in settings that feature typical characteristics of financial data. Therefore we allow for conditionally heteroscedastic data at the margins and adjust our test procedures accordingly. In a first simulation study parametric and non-parametric methods building on the joint distribution function are examined using data generated from multivariate Gaussian distributions that allows us to abstract from issues with non-trivial parameter estimation. Performance of copula-based methods is compared under non-linear dependence by generating data from t-copulas in a second simulation study. Due to its practical relevance, dimensionality effects are also taken into account. Some attention is devoted to the situations where the test procedures are applied under misspecification, such as choosing the wrong joint distribution function or copula within the sup-LR framework.

While we find strong empirical power advantages of sup-LR-type tests over fluctuation tests in when data are sampled from a multivariate Gaussian distribution, results are more mixed in the case of copulas. Here, the sup-LR test does not necessarily provide the strongest results, presumably caused by a large degree of parameter uncertainty. Should one operate in these settings, sup-LR tests under the simplest parametric assumption, namely that of a multivariate Gaussian distribution or copula, provide a suitable alternative to non-parametric methods.[1] Also one might use both tests and use a simple error correction scheme like the Boole-Bonferroni method to correct for multiple testing of the same hypothesis.

Results are stronger when it comes to estimating change-points: in almost every considered case, the parametric sup-LR method yields better estimators in terms of bias and variance, irrespectively of correct or incorrect model specification. If precise knowledge on the timing of regime-shifts is of central importance, for example in a portfolio management situation, one should rather use a parametric method. Even if one does not want to rely on a certain specification, our results strongly suggest the sup-LR framework under a Gaussian assumption to achieve useful results.

The structure of the paper is as follows: First, section 2 introduces the hypotheses pairs used in the two-step procedure and the sup-LR test framework, moreover it contains some analytical high-level background. Section 3 applies these results to several parametric frameworks in simulation studies, while an illustration of the discussed methods is given in section 4, using bank equity returns around the financial crisis of 2008. Section 5 gives a conclusion, while the appendix contains the proofs. In the supplementary material we provide an overview about all non-parametric test frameworks (moment-based fluctuation and empirical copula tests) used in the simulation study. With regard to the moment-based fluctuation test, there are some new analytical derivations. We also present simulation evidence under serial independence, supporting the validity of our bootstrap scheme and confirming the simula-

[1] This has also strong computational advantages relative to the correctly specified sup-LR test.

tion evidence for empirical power and break point estimation from the GARCH-simulations. Moreover, the supplementary material contains a third application using European government bond spreads. In a second application to Oil and EURO STOXX 50 return data, we demonstrate that ignoring the residual effect can actually lead to wrong test decisions.

2 Theoretical Framework

For $t = 1, \dots, n$, let $X_t \sim (\delta_{(t)}, \theta_{(t)}, \eta)$ be a multivariate random variable with dimension m and joint density $f_t(\cdot) = (f_{1,t}(\cdot), \dots, f_{m,t}(\cdot))$. $\delta_{(t)} \in \Theta_\delta \subset \mathbb{R}^{k_\delta}$ denotes a possibly time-varying k_δ-dimensional parameter vector shaping its dependence structure and $\theta_{(t)} = [\theta'_{1,(t)}, ..., \theta'_{m,(t)}]'$ denotes a possibly time-varying vector consisting of the parameters shaping the marginal distributions, indexed by i, with $\theta_{i,(t)} \in \Theta_\theta \subset \mathbb{R}^{k_\theta}$. Moreover, $\eta \in \Theta_\delta \subset \mathbb{R}^{k_\eta}$ is another parameter vector which is assumed to be constant throughout all settings. We are interested in changes in $\delta_{(t)}$ and/or $\theta_{(t)}$ and allow for a time-constant nuisance parameter η.

Let the change-point corresponding to dimension i be denoted by l_i and l_D the change-point of the dependency-structure and further assume $l_1 \leq l_2 \leq ... \leq l_m \leq l_D$. The particular order of change-points merely eases notation and does not lead to loss of generality, since the asymptotics in sequential procedures are not affected from switching the change-point order. Denoting the time index by t, one formally has

$$
\begin{aligned}
&X_t \sim (\theta_{1,1}, \theta_{2,1}, \cdots, \theta_{m,1}, \delta_{D,1}) \quad \text{for } t = 1, ..., l_1 \\
&X_t \sim (\theta_{1,2}, \theta_{2,1}, \cdots, \theta_{m,1}, \delta_{D,1}) \quad \text{for } t = l_1 + 1, ..., l_2 \\
&\cdots\cdots \\
&X_t \sim (\theta_{1,2}, \theta_{2,2}, \cdots, \theta_{m,2}, \delta_{D,1}) \quad \text{for } t = l_m + 1, ..., l_D \\
&X_t \sim (\theta_{1,2}, \theta_{2,2}, \cdots, \theta_{m,2}, \delta_{D,2}) \quad \text{for } t = l_D, ..., n
\end{aligned}
$$

with at most $m+1$ asymptotically distinct break points. Note that, defining $\lambda_i := \frac{l_i}{n}$, two change-points are asymptotically distinct if $\lambda_1 \neq \lambda_2$ as $n \to \infty$. The ordering of the break dates reflects a situation where *shift contagion* is present: at first, there is a change in mean and variance of one variable (e.g. a stock market index of country A), followed by a change in the second (country B), third (country C) variable and so forth. The correlation between both markets changes at some later point in time, l_D. Note that one could, but does not have to assume that the X_t are independent between the change-points, see the discussion in the end of Section 2.

The following hypotheses pairs are relevant for the sequential procedures under consideration and have appeared in this or slightly different forms throughout the existing literature:

Hypotheses Pair 1 (Marginal Distributions)**.** For every margin i, we test:

$$
\begin{aligned}
&H_0 : \theta_{i,(1)} = ... = \theta_{i,(n)} \quad \text{against} \\
&H_1 : \theta_{i,(1)} = ... = \theta_{i,(l_i)} \neq \theta_{i,(l_i+1)} = ... = \theta_{i,(n)} \text{ for some } l_i \in \{2, \dots, n-1\}.
\end{aligned}
$$

Hypotheses Pair 2 (Dependency, Constant Margins)**.**

$$
\begin{aligned}
&H_0 : \delta_{(1)} = ... = \delta_{(n)} \text{ with } \theta_{i,(1)} = ... = \theta_{i,(n)} \quad \forall\, i = 1, ..., m \quad \text{against} \\
&H_1 : \delta_{(1)} = ... = \delta_{(l_D)} \neq \delta_{(l_D+1)} = ... = \delta_{(n)} \text{ for some } l_D \in \{2, \dots, n-1\} \\
&\text{with } \theta_{i,(1)} = ... = \theta_{i,(n)} \quad \forall\, i = 1, ..., m
\end{aligned}
$$

If Pearson's correlation matrix is used to measure cross-sectional dependency, Hypotheses Pair 2 is in line with the test proposed in Wied et al. [2012b] who extend the covariance test from Aue et al. [2009] to moment hypothesis on correlations. In a shift contagion situation it comes natural to extend the first two hypotheses pairs into a joint framework:

Hypotheses Pair 3 (Two-Step Testing Procedure)**.**

$$
\begin{aligned}
H_0 :\ & \delta_{(1)} = ... = \delta_{(n)} \\
& \theta_{1,(1)} = ... = \theta_{1,(l_1)}, \theta_{1,(l_1+1)} = ... = \theta_{1,(n)} \\
& \dots \quad \dots \quad \dots \\
& \theta_{m,(1)} = ... = \theta_{m,(l_m)}, \theta_{m,(l_m+1)} = ... = \theta_{m,(n)} \\
& \text{for some } (l_1, \dots, l_m) \in \{2, \dots, n-1\}^m \text{ against} \\
H_1 :\ & \delta_1 = \delta_{(1)} = ... = \delta_{(l_D)} \neq \delta_{(l_D+1)} = ... = \delta_{(n)} = \delta_2 \text{ for some } l_D \in \{2, \dots, n-1\} \\
& \theta_{1,(1)} = ... = \theta_{1,(l_1)}, \theta_{1,(l_1+1)} = ... = \theta_{1,(n)} \\
& \dots \quad \dots \quad \dots \\
& \theta_{m,(1)} = ... = \theta_{m,(l_m)}, \theta_{m,(l_m+1)} = ... = \theta_{m,(n)} \\
& \text{for some } (l_1, \dots, l_m) \in \{2, \dots, n-1\}^m.
\end{aligned}
$$

Hypotheses Pair 3 allows for changes in the marginal distributions under both the null and alternative hypothesis. In particular, there is no stationarity under the null hypothesis. Under the null, we have constant dependence and under the alternative, we have a two-regime model in the dependence structure.

We move on to propose a framework which uses fully specified parametric models in order to evaluate parameter stability and test the hypotheses laid out before. The framework goes back to the seminal contribution of Andrews [1993] who suggests a method which is essentially applicable for all GMM-type estimators, such as maximum-likelihood and pseudo-maximum-likelihood. The framework is the following: The sample is divided into two sub-samples for any $j = \pi \cdot n,\ \pi \in [\underline{\pi}, \overline{\pi}], 0 < \underline{\pi} < \overline{\pi} < 1$, where parameters are divided into those that change under the null and alternative hypothesis and nuisance parameters that are invariant under null and alternative, denoted by η. Parameter constancy is tested by forming a sequence of likelihood-ratio test statistics for all change point candidates. The testing function is given by the log-likelihood function and the test statistic for a fixed j is given by the difference of the log-likelihood under the restricted and the unrestricted ML- or pseudo-ML-estimator. No restriction here means that the parameter, which is tested for constancy, is calculated based on $X_1, \dots, X_j$ and $X_{j+1}, \dots, X_n$ separately. We note that, while the framework is based on Andrews [1993], that paper does not look directly at these test statistics, but on "supLR-type" statistics which are based on the differences of GMM-objective functions evaluated at the restricted and unrestricted estimator. We use the likelihood functions themselves in order to avoid calculating the scores for each of our parametric models.

In the following, we shortly present the parametric frameworks for the first two hypotheses pairs, which are known from Andrews [1993]. Hypotheses Pair 3 is discussed afterwards.

Testing constancy of marginal distributions, i.e. Hypotheses Pair 1, is performed with a test statistic for dimension i that is based on

$$A_{i,j} := A_j(\hat{\theta}_{i,0}, \hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \hat{\eta}) := 2\big(L(\hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \hat{\eta}) - L(\hat{\theta}_{i,0}, \hat{\eta})\big) \tag{1}$$

with

$$\begin{aligned} L(\hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \hat{\eta}) &= \sum_{t=1}^{j} l_{i,t}(\hat{\theta}_{i,1}, \hat{\eta}) + \sum_{t=j+1}^{n} l_{i,t}(\hat{\theta}_{i,2}, \hat{\eta}) \text{ and} \\ L(\hat{\theta}_{i,0}, \hat{\eta}) &= \sum_{t=1}^{n} l_{i,t}(\hat{\theta}_{i,0}, \hat{\eta}). \end{aligned}$$

Here, $l_{i,t}(\cdot)$ denotes the contribution to the log-likelihood for margin i from observation t conditionally on the observations from 1 to $t-1$ (with the convention that the first contribution $l_{i,1}(\cdot)$ depends on some initial value), i.e., $l_{i,t}(\theta_i, \eta) = f_{i,t}(x_t|x_{t-1}, \ldots, x_1; \theta_i, \eta)$. Moreover, $\hat{\theta}_{i,1} = \hat{\theta}_{i,1,j}$ is the ML-estimator for θ_i based on $X_1, \ldots, X_j$, where $j := [\pi n]$ (the floor function is omitted in the following for brevity) and $\hat{\theta}_{i,2} = \hat{\theta}_{i,2,j}$ the one based on $X_{j+1}, \ldots, X_n$, $\hat{\theta}_{i,0}$ the one based on $X_1, \ldots, X_n$ and $\hat{\eta}$ the ML estimator based on $X_1, \ldots, X_n$ for the constant nuisance parameter η. The estimation of the parameter θ_i is always performed conditionally on $\hat{\eta}$.

Regarding Hypotheses Pair 2, we restrict the analysis to the case in which we can consistently estimate the true dependence parameter δ_0 without having to estimate the marginals. Then, the hypothesis is tested with a test statistic based on

$$A_j(\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2, \hat{\eta}) = 2\big(L(\hat{\delta}_1, \hat{\delta}_2, \hat{\eta}) - L(\hat{\delta}_0, \hat{\eta})\big) \tag{2}$$

with

$$\begin{aligned} L(\hat{\delta}_1, \hat{\delta}_2, \hat{\eta}) &= \sum_{t=1}^{j} l_t(\hat{\delta}_1, \hat{\eta}) + \sum_{t=j+1}^{n} l_t(\hat{\delta}_2, \hat{\eta}) \text{ and} \\ L(\hat{\delta}_0, \hat{\eta}) &= \sum_{t=1}^{n} l_t(\hat{\delta}_0, \hat{\eta}) \end{aligned}$$

and the definitions of $l_t(\cdot)$ and the estimators $\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2$ similar to Hypotheses Pair 1.

We now state a theorem concerning the asymptotic distribution of the sequence of LR-statistics. The result can be indirectly inferred from Andrews [1993] and Andrews and Ploberger [1995]: As stated by Andrews and Ploberger [1995], the asymptotic distribution is identical to the asymptotic distribution of the sup-Wald and the sup-LM statistics introduced by Andrews [1993]. Moreover, they are shown to converge to the same limit in our Theorem 1. On the other hand, we provide a direct proof, which is presented in the Appendix. Before, we impose some regularity assumptions. Here and in the following, Γ_k denotes the stochastic process given by a k-dimensional vector of independent Brownian motions.

Assumption 1. For Hypotheses Pair 1, it holds under the null hypothesis for margin i with the true parameter $\theta_{i,0}$:

1. The third derivatives of $l_{i,t}(\theta_i, \eta)$ with respect to θ_i exist and are uniformly bounded for $t = 1, \dots, n$.

2. The estimators $\hat{\theta}_{i,1,\pi n}$, $\hat{\theta}_{i,2,\pi n}$ and $\hat{\theta}_{i,0}$ fulfill a functional central limit theorem, i.e., the stochastic process

$$\hat{T}_{1,n}(\pi) := \sqrt{n} \begin{pmatrix} \hat{\theta}_{i,1,\pi n} - \theta_{i,0} \\ \hat{\theta}_{i,2,\pi n} - \theta_{i,0} \\ \hat{\theta}_{i,0} - \theta_{i,0} \end{pmatrix}$$

converges in distribution to the stochastic process

$$T_1(\pi) := \begin{pmatrix} \frac{1}{\pi} H_i^{-1/2} \cdot \Gamma_{k_\theta}(\pi) \\ \frac{1}{1-\pi} H_i^{-1/2} \cdot \Gamma_{k_\theta}(1-\pi) \\ H_i^{-1/2} \cdot \Gamma_{k_\theta}(1) \end{pmatrix}$$

in $D[0,1]^{3k_\theta}$, the space of $3k_\theta$-dimensional cádlág functions on the unit interval with

$$H_i = -\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial\theta_i \partial\theta_i'} l_{i,t}(\theta_{i,0}, \hat{\eta}).$$

The limit H_i remains the same if $\theta_{i,0}$ is replaced by a consistent estimator.

For Hypotheses Pair 2, it holds under the null hypothesis with the true parameter δ_0:

1. The third derivatives of $l_t(\delta, \theta, \eta)$ with respect to δ exist and are uniformly bounded for $t = 1, \dots, n$.

2. The estimators $\hat{\delta}_{1,\pi n}$, $\hat{\delta}_{2,\pi n}$ and $\hat{\delta}_0$ fulfill a functional central limit theorem, i.e., the stochastic process

$$\hat{T}_{2,n}(\pi) := \sqrt{n} \begin{pmatrix} \hat{\delta}_{1,\pi n} - \delta_0 \\ \hat{\delta}_{2,\pi n} - \delta_0 \\ \hat{\delta}_0 - \delta_0 \end{pmatrix}$$

converges in distribution to the stochastic process

$$T_2(\pi) := \begin{pmatrix} \frac{1}{\pi} H^{-1/2} \cdot \Gamma_{k_\delta}(\pi) \\ \frac{1}{1-\pi} H^{-1/2} \cdot \Gamma_{k_\delta}(1-\pi) \\ H^{-1/2} \cdot \Gamma_{k_\delta}(1) \end{pmatrix}$$

in $D[0,1]^{3k_\delta}$, the space of $3k_\delta$-dimensional cádlág functions on the unit interval with

$$H = -\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial\delta \partial\delta'} l_t(\delta_0, \hat{\eta}).$$

The limit H remains the same if δ_0 is replaced by a consistent estimator.

Note that the expression of the limits appears to be natural if it can be assumed that (e.g.) $\hat{\delta}_{1,j}-\delta_0$ can be linearized by a Taylor approximation as $\left(\sum_{t=1}^{j}\frac{\partial^2}{\partial\delta\partial\delta'}l_t(\delta_0,\hat{\eta})\right)^{-1}\sum_{t=1}^{j}\frac{\partial}{\partial\delta}l_t(\delta_0,\hat{\eta})+o_p(\sqrt{n})$ and $\hat{\delta}_{2,j}-\delta_0$ can be linearized as $\left(\sum_{t=j+1}^{n}\frac{\partial^2}{\partial\delta\partial\delta'}l_t(\delta_0,\hat{\eta})\right)^{-1}\sum_{t=j+1}^{n}\frac{\partial}{\partial\delta}l_t(\delta_0,\hat{\eta})+o_p(\sqrt{n})$. This follows from standard asymptotic theory for maximum likelihood estimators, see e.g. van der Vaart [1998].

Definition 1. Consider a k-dimensional Brownian motion Γ_k. Then, the process $\mathcal{B}_k$ with

$$\mathcal{B}_k(\pi)=\frac{(\Gamma_{k_\theta}(\pi)-\pi\Gamma_{k_\theta}(1))'(\Gamma_{k_\theta}(\pi)-\pi\Gamma_{k_\theta}(1))}{\pi(1-\pi)} \tag{3}$$

is called standardized tied-down Bessel process of order k.

Theorem 1. *Let Assumption 1 be true. Under Hypotheses Pair 1, it holds that $A_{i,\pi n},\underline{\pi}\leq\pi\leq\overline{\pi}$, converges to a $\mathcal{B}_{k_\theta}$-process. Under Hypotheses Pair 2, it holds that $A_{\pi n},\underline{\pi}\leq\pi\leq\overline{\pi}$, converges to a $\mathcal{B}_{k_\delta}$-process. Both convergences take place in $D[\underline{\pi},\overline{\pi}]$, the space of cádlág-functions over $[\underline{\pi},\overline{\pi}]$.*

As we have the factor $\pi(1-\pi)$ in the denominator, it is clear that $\Pi=[\underline{\pi},\overline{\pi}]$ has to be a strict subset of the unit interval. To test the null hypothesis of parameter constancy against a single unknown change point, the sup-functional is applied to the components. Then, for (2) (and similarly for (1)) and based on the continuous mapping theorem, in the framework of Andrews [1993],

$$\sup_{\underline{\pi}\cdot n\leq j\leq\overline{\pi}\cdot n} A_j \to_d \sup_{\Pi}\mathcal{B}_{k\delta}(\pi) \tag{4}$$

So the null hypothesis is rejected when the $(1-\alpha)$-quantile associated with the limiting process from Theorem 1, defined by $c_\alpha=P(\sup_{\pi\in\Pi}\mathcal{B}_{k\delta}>c_\alpha)=\alpha$ is exceeded. Critical values are tabulated in Andrews [1993] and depend on the degrees of freedom of the limiting process and the considered interval Π of candidate change points. In every situation considered in the following, the supremum of $\{A_j\}$ is also used to estimate the change-point by

$$\hat{l}=\operatorname*{arg\,sup}_{\underline{\pi}n\leq j\leq\overline{\pi}n} A_j \tag{5}$$

In practical applications sup- and argsup-functional are replaced by the max- and argmax-functional, respectively. Following the suggestion of Andrews [1993] the set of potential change points is chosen to be $\Pi=[0.15,0.85]$, the general case will however be maintained in the notation.

Our testing idea for analyzing Hypotheses Pair 3 is similar to that of Hypotheses Pair 2. The test statistic is given by

$$\sup_{\underline{\pi}\cdot n\leq j\leq\overline{\pi}\cdot n} A_j \tag{6}$$

with A_j given in (2), but with the original data $X_t,t=1,\ldots,n$, replaced by appropriate residuals $\hat{X}_t,t=1,\ldots,n$. Here, "residuals" implies that we transform marginal time series such that they do not exhibit breaks any more, while the dependence structure is not affected by the transformation. In fact, we are interested in the dependence parameters of the multivariate random vector X_t, but cannot make inference for it directly and need a proper transformation instead. One simple example would the model $X_t = \Sigma_t R$, where $\Sigma_t = \mathsf{diag}(\sigma_{t1}, \sigma_{t2})$ and R is bivariate centered normally distributed with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Here, inference about ρ would be based on $\hat{X}_t = \Sigma_t^{-1} X_t$.

The type of transformations as well as the particular test statistics depend on the respective parametric specification, which are derived in the following subsections. For example, the setting allows for time-varying marginal variances if $\theta = \theta_1$ for $t \leq j = \pi_\theta n$ and $\theta = \theta_2$ for $t > j = \pi_\theta n$. Demetrescu and Wied [2019] discuss such models in detail and also argue analytically and with numerical evidence why it is not possible to test Hypotheses Pair 3 with the standard method of Andrews [1993] who assumes stationarity under the null hypothesis. Allowing for unknown marginal parameters, which have to be estimated, introduces complications concerning the limit distribution. As pointed out in many studies on that matter, using estimated parameters and change-point locations in the first step potentially affects estimation of parameters and change-point locations in the second step, see Qu and Perron [2007], Chan et al. [2009] and Demetrescu and Wied [2019].

In our setting, the intuition for getting a residual effect is the following: Define with $\tilde{\delta}_{\pi n} = (\hat{\delta}_{1,\pi n}, \hat{\delta}_{2,\pi n})$ the vector of unrestricted ML-estimators and $\hat{\theta}_t$ the estimator for the nuisance parameter θ_t such that $l_t(\tilde{\delta}_{\pi n}, \hat{\theta}_t) := l_t(\hat{\delta}_{1,\pi n}, \hat{\theta}_t, \hat{\eta})$ for $t \leq j = \pi n$ and $l_t(\tilde{\delta}_{\pi n}, \hat{\theta}_t) := l_t(\hat{\delta}_{2,\pi n}, \hat{\theta}_t, \hat{\eta})$ for $t > j = \pi n$. We consider the special case that the whole vector θ_t breaks at exactly one point $\pi_\theta n, \pi_\theta \in (0, 1)$, such that we have two estimators $\hat{\theta}_{1,\pi_\theta n}$ and $\hat{\theta}_{2,\pi_\theta n}$ for θ_1 and θ_2, respectively. To ease notation, we omit the dependency of the likelihood contributions on the nuisance parameter $\hat{\eta}$. We impose an additional assumption, which appears to be natural given that both $\tilde{\delta}_\pi$ and $\hat{\delta}$ are consistent for δ under the null hypothesis. We leave it as a task for further research to validate these assumptions in the examples considered below; the theorem should mainly serve as an illustration how the residual effect looks like in general.

Assumption 2. For Hypotheses Pair 3, it holds under the null hypothesis for the true parameters $\theta_1, \theta_2, \delta_0$:

1. The third derivatives of $l_t(\cdot, \theta_t)$ with respect to θ_t exist and are uniformly bounded for $t = 1, \ldots, n$.

2. In the case of known θ_1, θ_2, the estimators $\hat{\delta}_{1,\pi n}$, $\hat{\delta}_{2,\pi n}$ and $\hat{\delta}_0$ satisfy a functional central limit theorem, i.e., the stochastic process

$$\hat{T}_{3,n}(\pi) := \sqrt{n} \begin{pmatrix} \hat{\delta}_{1,\pi n} - \delta_0 \\ \hat{\delta}_{2,\pi n} - \delta_0 \\ \hat{\delta}_0 - \delta_0 \end{pmatrix}$$

converges in distribution to the stochastic process

$$T_3(\pi) := \begin{pmatrix} \frac{1}{\pi} H_*^{-1/2} \cdot \Gamma_{k_\delta}(\pi) \\ \frac{1}{1-\pi} H_*^{-1/2} \cdot \Gamma_{k_\delta}(1 - \pi) \\ H_*^{-1/2} \cdot \Gamma_{k_\delta}(1) \end{pmatrix}$$

in $D[0,1]^{3k_\delta}$, the space of $3k_\delta$-dimensional cádlág functions on the unit interval with

$$H_* = -\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{\pi_\theta n} \frac{\partial^2}{\partial\theta\partial\theta'} l_t(\tilde{\delta}_0, \theta_1) - \lim_{n\to\infty} \frac{1}{n}\sum_{t=\pi_\theta n+1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'} l_t(\tilde{\delta}_0, \theta_2).$$

and Γ_{k_δ} denoting a k_δ-dimensional vector of independent Brownian motions. The limit $H*$ remains the same if δ_0 is replaced by a consistent estimators.

3. The process

$$\begin{aligned} B_{\pi n} \quad &:= \quad \sqrt{n}(\hat{\theta}_{1,\pi_\theta n} - \theta_1)' \frac{1}{\sqrt{n}} \sum_{t=1}^{\pi_\theta n} \frac{\partial}{\partial\theta}\left(l_t(\tilde{\delta}_{\pi n}, \theta_1) - l_t(\hat{\delta}, \theta_1)\right) \\ &+ \quad \sqrt{n}(\hat{\theta}_{2,\pi_\theta n} - \theta_2)' \frac{1}{\sqrt{n}} \sum_{t=\pi_\theta n}^{n} \frac{\partial}{\partial\theta}\left(l_t(\tilde{\delta}_{\pi n}, \theta_2) - l_t(\hat{\delta}, \theta_2)\right) \end{aligned}$$

converges to some limit process $R(\pi)$ in $D[\underline{\pi}, \overline{\pi}]$.

4. The process $C_{\pi n} := \frac{\partial^2}{\partial\theta\partial\theta'}\frac{1}{n}\sum_{t=1}^{n}\left(l_t(\tilde{\delta}(\pi), \theta_t) - l_t(\hat{\delta}, \theta_t)\right)$ converges to zero in probability.

Then, we have the following theorem, whose proof is given in the appendix:

Theorem 2. *Under Assumption 2, it holds for $A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \hat{\theta}_{\pi n}) := A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \hat{\theta}_{\pi n}, \hat{\eta})$ that*

$$A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \hat{\theta}_{\pi n}) \Rightarrow_d \mathcal{B}_l(\pi) + R(\pi),$$

where the residual effect is given by $R(\pi)$.

Simulation evidence supports using a residual bootstrap scheme, which leads to correctly sized tests. Under the assumption of proper transformation prior to step two, we can use a simple residual bootstrap scheme, which is now briefly lined out: Let a sample $\hat{X}_1^*, ..., \hat{X}_n^*$ be drawn from $\hat{X}_1, ..., \hat{X}_n$. For any bootstrap repetition b, let the sup-LR test statistic from (1) or (2) be denoted by $\sup A_j^b$, such that the p-value follows as

$$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} 1_{\{\sup A_j^b > \sup A_j\}} \tag{7}$$

If the estimation error in the first step could be ignored, it would be reasonable to use the same critical values as in Hypotheses Pair 2, because the difference of estimated parameters under alternative and null remains exactly the same.

The way in which the sample $\hat{X}_1^*, ..., \hat{X}_n^*$ is drawn, depends on the actual assumption on the serial dependence of the X_t. If one assumes that they are independent between the change-points, one obtains independent draws with replacement. While this strong assumption is often violated in practice, there is a certain consensus in financial econometrics that serial dependence takes the form of (conditional) heteroscedasticity which can be captured by GARCH or related processes. Some empirical evidence for asset returns can be found in Cont [2001]. Chuang and Lee [2014] and Bartram et al. [2007] use for example a GJR-GARCH(1,1)-t model while Candelon and Manner [2010] employ a SWARCH-approach and Andreou and Ghysels [2003] use GARCH(1,1)-t and GARCH(1,1)-normal models. In order

to focus on the relevant aspects in two-step testing procedures, we restrict ourselves to the simplest case of an univariate GARCH(1,1)-structure from Bollerslev [1986] for all *marginal* distributions and apply appropriate *univariate* filters, e.g. we consider GARCH residuals. When testing constant marginal distribution, the serial dependence is taken into account by using an appropriate covariance-matrix as in Blatt et al. [2015]. Before testing for constant cross-sectional dependence, we estimate piecewise or full-sample GARCH-residuals, depending on the test result in the first step. This way, we deal with *almost i.i.d.* data, as Bücher et al. [2015] do with the rationale that the dependence structure is not affected by this kind of filtering. Consequently we can draw independent draws with replacement from the GARCH residuals in the second step. If one assumes other forms of serial dependence, one might have to resort to a block-bootstrap scheme. Our numerical results however indicate validity of the bootstrap with replacement in the presence of volatility clustering, consistent with evidence from the references given above.
With these preliminary remarks out of the way, the sup-LR framework obviously requires assumption of a particular parametric model, while the particular moments which are subject to structural changes need to be specified in the fluctuation test framework. We therefore turn to a description of models that we use in our simulation study and the way they enter the testing procedures.

3 Simulation Evidence for Selected Parametric Models

This section uses several commonly used parametric settings, to evaluate finite sample properties of the parametric and non-parametric tests. We assume GARCH(1,1)-margins with mean zero and at most one change in the *unconditional* variance σ^2[2]. We denote the conditional variance by $h_{i,t}^2$. The state and volatility memory factors (α, β) are regarded as time-invariant nuisance parameters. In both testing frameworks we therefore employ the HAC-consistent estimator proposed by Andrews [1991] to obtain correctly sized tests in the first step. GARCH-residuals are computed the usual way, either over the full sample or piece-wise, if the test for constant variance rejects Hypotheses Pair 1.
The main part of interest is step two: first, empirical rejection rates under H_0 are reported, i.e. the correlation coefficient is kept constant while there are changes in the unconditional variance (**scenario 1**). The extent to which the unconditional variance changes is controlled by a tuning parameter s. We report the asymptotic and bootstrapped critical values to demonstrate the failure of standard asymptotics. In a second study we compute empirical power under changing dependency parameters (**scenario 2**). Additionally the Monte Carlo bias and Mean Squared Error of each break point estimator reveal a superiority of the sup-LR framework that is robust to model misspecification.

3.1 Gaussian Distribution

The easiest way to model cross-sectional dependence is a multivariate Normal distribution, parametrized in terms of correlations. Although most likely not the best choice in many cases, it provides a good starting point and offers some useful insight into more complicated models. We impose the regularity condition that variances are bounded away from zero. In

[2]Often these methods are applied to daily equity returns, where it is reasonable to assume zero means.

such a situation, one can test Hypotheses Pair 3 by setting $\theta_i = \sigma_i^2$ and $\delta_i = P_i$. In the fluctuation test framework, this specification boils down to the CUSUM of squares test from Wied et al. [2012a].

Test Statistics The sup-LR-analogue follows from the probability density of a Gaussian random variable

$$f(X_{i,t}; \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big(- \frac{X_{i,t}^2}{2\sigma_i^2}\Big)$$

the log-Likelihood for full-sample estimation is given by

$$L(X_i, \sigma_0^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_0^2) - \sum_{t=1}^{n} \frac{X_{t,i}^2}{2\sigma_0^2}$$

Dividing the sample at any j yields

$$\begin{aligned} L(X_i, \sigma_{i,1}^2, \sigma_{i,2}^2) = &- \frac{j}{2}\log(2\pi) - \frac{j}{2}\log(\sigma_{i,1}^2) - \sum_{t=1}^{j} \frac{X_{i,t}^2}{2\sigma_{i,1}^2} \\ &- \frac{n-j}{2}\log(2\pi) - \frac{n-j}{2}\log(\sigma_{i,2}^2) - \sum_{t=j+1}^{n} \frac{X_{i,t}^2}{2\sigma_{i,2}^2} \end{aligned}$$

for the log-likelihood, where $\sigma_{i_1}^2$ and $\sigma_{i,2}^2$ are the partial-sample estimators. Evaluating the difference of the log-likelihood under full-sample and partial-sample estimators gives after some simplifications the test statistic for a fixed j:

$$A_j(X_i; \hat{\sigma}_{i,0}^2, \hat{\sigma}_{i,1}^2, \hat{\sigma}_{i,2}^2) = n\log(\hat{\sigma}_{i,0}^2) - j\log(\hat{\sigma}_{i,1}^2) - (n-j)\log(\hat{\sigma}_{i,2}^2)$$

The limiting process of $\{A_j\}$ is of the form of equation (3) and has $k = 1$ degrees of freedom. Since we allow for volatility clustering, critical values depend on the HAC-consistent variance estimator (Qu and Perron [2007] and Blatt et al. [2015]). Conditional on the test decision, the data are standardized by

$$\begin{aligned} &\hat{h}_{i,t}^2 = \hat{c}_{i,0} + \hat{\alpha}_{i,0}X_{i,t-1} + \hat{\beta}_{i,0}h_{i,t-1}^2 \\ &\hat{Z}_{i,t} = \frac{X_{i,t}}{\hat{h}_{i,t}} \\ &\qquad \text{if no break was detected or by} \\ &\hat{h}_{i,t}^2 = \hat{c}_{i,1} + \hat{\alpha}_{i,0}X_{i,t-1} + \hat{\beta}_{i,0}h_{i,t-1}^2 \\ &\hat{Z}_{i,t} = \frac{X_{i,t}}{\hat{h}_{i,t}} \quad \text{for } t = 1, ..., l_i \\ &\hat{h}_{i,t}^2 = \hat{c}_{i,2} + \hat{\alpha}_{i,0}X_{i,t-1} + \hat{\beta}_{i,0}h_{i,t-1}^2 \\ &\hat{Z}_{i,t} = \frac{X_{i,t}}{\hat{h}_{i,t}} \quad \text{for } t = l_i, ..., n \end{aligned} \tag{8}$$

For the piecewise standardized data, full-sample and partial-sample ML-estimators follow from the simplified log-likelihood, now given by

$$
\begin{aligned}
L(\hat{Z}; P_0) &= -\frac{n}{2}\log|P_0| - \frac{1}{2}\sum_{t=1}^{n}\hat{Z}_t' P_0^{-1}\hat{Z}_t \\
L(\hat{Z}; P_1, P_2) &= -\frac{j}{2}\log|P_1| - \frac{1}{2}\sum_{t=1}^{j}\hat{Z}_t' P_1^{-1}\hat{Z}_t - \frac{n-j}{2}\log|P_2| - \frac{1}{2}\sum_{t=j+1}^{n}\hat{Z}_t' P_2^{-1}\hat{Z}_t
\end{aligned}
$$

yielding

$$
\begin{aligned}
\hat{P}_0 &= \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}\hat{Z}_{1,t}\hat{Z}_{2,t} & \cdots \\ \cdots & \hat{Z}_{m-1,t}\hat{Z}_{m,t}\end{pmatrix}, \text{ and} \\
\hat{P}_1 &= \frac{1}{j}\sum_{t=1}^{j}\begin{pmatrix}\hat{Z}_{1,t}\hat{Z}_{2,t} & \cdots \\ \cdots & \hat{Z}_{m-1,t}\hat{Z}_{m,t}\end{pmatrix}, \quad \hat{P}_2 = \frac{1}{n-j}\sum_{t=j+1}^{n}\begin{pmatrix}\hat{Z}_{1,t}\hat{Z}_{2,t} & \cdots \\ \cdots & \hat{Z}_{m-1,t}\hat{Z}_{m,t}\end{pmatrix}
\end{aligned}
$$

where it was used that $\sum_{t=1}^{j}\hat{Z}_{i,t}^2 = 1$ and $\sum_{t=j+1}^{n}\hat{Z}_{i,t}^2 = 1$ for every dimension i by construction of $\hat{Z}$. Given j, the likelihood-ratio test statistic for centered and standardized Gaussian data is obtained as

$$
A_j = n\cdot\log(|\hat{P}_0|) - j\cdot\log(|\hat{P}_1|) - (n-j)\cdot\log(|\hat{P}_2|)
$$

Had one based the test statistic on the unobserved Z_t, a reasonable approximation for the critical value associated with the sup-functional $\sup_{\underline{\pi}\cdot n\le j\le\overline{\pi}\cdot n} A_j$ would be given by the appropriate quantile of $\sup_{\pi\in\Pi}\mathcal{B}_{(m-1)m/2}(\pi)$.

Finite-Sample Properties In the first Monte Carlo exercise we apply both sequential testing procedures to data generated according to

$$
\begin{aligned}
&Z_t \overset{i.i.d.}{\sim} N(0,P_1) && \text{for } t = 1,...,l_D \\
&Z_t \overset{i.i.d.}{\sim} Nt(0,P_2) && \text{for } t = l_D,...,n \\
&h_{1,1}^2 = 1\ , \quad h_{1,t}^2 = \alpha Z_{1,t-1}^2 + \beta h_{1,t-1}^2 && \text{for } t = 2,...,n \\
&X_{1,t} = h_{1,t}Z_{1,t} && \text{for } t = 1,...,l_1 \\
&X_{1,t} = sh_{1,t}Z_{1,t} && \text{for } t = l_1,...,n \\
&h_{2,1}^2 = 1\ , \quad h_{2,t}^2 = \alpha Z_{2,t-1}^2 + \beta h_{2,t-1}^2 && \text{for } t = 2,...,n \\
&X_{2,t} = h_{2,t}Z_{2,t} && \text{for } t = 1,...,l_2 \\
&X_{2,t} = sh_{2,t}Z_{2,t} && \text{for } t = l_2,...,n \\
&\cdots \quad \cdots && \cdots \\
&h_{m,1}^2 = 1\ , \quad h_{m,t}^2 = \alpha Z_{m,t-1}^2 + \beta h_{m,t-1}^2 && \text{for } t = 2,...,n \\
&X_{m,t} = h_{m,t}Z_{m,t} && \text{for } t = 1,...,l_2 \\
&X_{m,t} = sh_{m,t}Z_{m,t} && \text{for } t = l_2,...,n
\end{aligned}
$$

with $l_1 < l_2 < l_D$. The change points are chosen to be distinct for the first margin, all other margins, and the joint distribution, consistent with the interpretation that dimension one corresponds to the 'origin country' of a financial crisis which spreads to other countries (dimensions $i > 1$) with a certain time lag. To ensure comparability of results across different sample sizes, we set $\lambda_1 \equiv \frac{l_1}{n} = 0.5$, $\lambda_2 \equiv \frac{l_2}{n} = 0.6$, and $\lambda_D \equiv \frac{l_D}{n} = 0.7$. The sample sizes are set to 100, which seems reasonable for quarterly data, 500 which should be reached either in long macroeconomic time series or (daily) financial market data and 1500 to approximate asymptotic behavior. Depending on the sample size, change-points therefore are given by

n	l_1	l_2	l_D
100	50	60	70
500	250	300	350
1500	750	900	1050

For scenario 1 the correlation is kept constant by setting $\rho_2 = 0.4$. In order to focus on the important aspects, the magnitude of parameter changes in each marginal distribution is identical and ranges over $s \in [1, 1.5, 2, 2.5, 3]$. The case of $s = 1$ corresponds to testing Hypotheses Pair 2 while all other cases $s_1 = s_2 \neq 1$ test Hypotheses Pair 3 where H_0 is true. The nominal significance level is set to 5 %, the corresponding are either taken from Kiefer [1959] for the fluctuation tests and Andrews [1993] for the sup-LR test or simulated using 1000 Monte Carlo repetitions on a discrete grid with 10.000 elements. We use 1000 Monte Carlo repetitions for each parameter constellation.

Results are shown in table 1: the test for constant marginal distributions has higher power for X_1 than for X_2, which is consistent with both theory and previous studies, as we set $\lambda_1 = 0.5$ and $\lambda_2 = 0.6$.

Table 1: Multivariate Gaussian, Scenario 1: Rejection Rates under H_0

s	$n = 100$				$n = 500$				$n = 1500$			
	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
Margins	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
1	0.213	0.197	0.127	0.149	0.063	0.052	0.041	0.043	0.032	0.025	0.034	0.025
1.5	0.774	0.716	0.577	0.489	0.981	0.952	0.874	0.767	1	1	0.998	0.997
2	0.975	0.950	0.917	0.857	1	1	1	0.995	1	1	1	1
2.5	0.997	0.997	0.992	0.974	1	1	1	1	1	1	1	1
3	1	1	1	0.998	1	1	1	1	1	1	1	1

	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 2$	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.
1	0.028	0.040	0.065	0.038	0.021	0.030	0.052	0.044	0.035	0.042	0.048	0.036
1.5	0.022	0.035	0.101	0.053	0.014	0.024	0.078	0.070	0.022	0.030	0.068	0.054
2	0.016	0.026	0.101	0.058	0.020	0.025	0.068	0.052	0.026	0.035	0.061	0.051
2.5	0.020	0.040	0.092	0.050	0.014	0.025	0.064	0.053	0.028	0.035	0.062	0.053
3	0.019	0.049	0.088	0.052	0.014	0.026	0.067	0.051	0.029	0.037	0.063	0.052

	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 3$	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.
1	0.026	0.048	0.142	0.058	0.037	0.047	0.061	0.046	0.036	0.045	0.057	0.045
1.5	0.019	0.041	0.164	0.072	0.015	0.021	0.111	0.081	0.029	0.033	0.083	0.071
2	0.023	0.038	0.143	0.054	0.019	0.024	0.094	0.067	0.034	0.040	0.086	0.065
2.5	0.020	0.041	0.151	0.048	0.018	0.029	0.090	0.065	0.033	0.043	0.088	0.067
3	0.010	0.040	0.161	0.051	0.022	0.028	0.082	0.059	0.029	0.038	0.089	0.064

	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 4$	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.	asym.	boot.
1	0.014	0.031	0.164	0.037	0.040	0.049	0.079	0.055	0.051	0.056	0.089	0.064
1.5	0.018	0.032	0.177	0.050	0.029	0.034	0.137	0.095	0.041	0.046	0.119	0.096
2	0.017	0.029	0.185	0.047	0.030	0.032	0.126	0.079	0.041	0.048	0.119	0.088
2.5	0.020	0.038	0.181	0.048	0.032	0.034	0.117	0.074	0.041	0.046	0.113	0.088
3	0.011	0.029	0.188	0.040	0.026	0.033	0.119	0.076	0.044	0.047	0.117	0.093

The simulations reveal presence of the residual effect for every sample size, which appears as soon as $s \neq 1$. In this case the bootstrap corrections increase the rejection rate of the fluctuation close to the nominal significance level of 5 %, while the sup-LR test is corrected for the increased rejection rates under H_0. As n increases, we observe correctly sized test decisions at each margin in both test frameworks. To obtain results on empirical power, the correlation of the first regime is set to $\rho_1 = 0.4$ and ρ_2 varies symmetrically around ρ_1 from -0.1 to 0.9 in steps of 0.1. The dashed lines in figure 1 - 3 represent empirical power, if the incorrect asymptotic critical values are used, in this way one can quantify the residual effect on empirical power.

Figure 1: Bivariate Gaussian, n=100, Scenario 2: Empirical Power

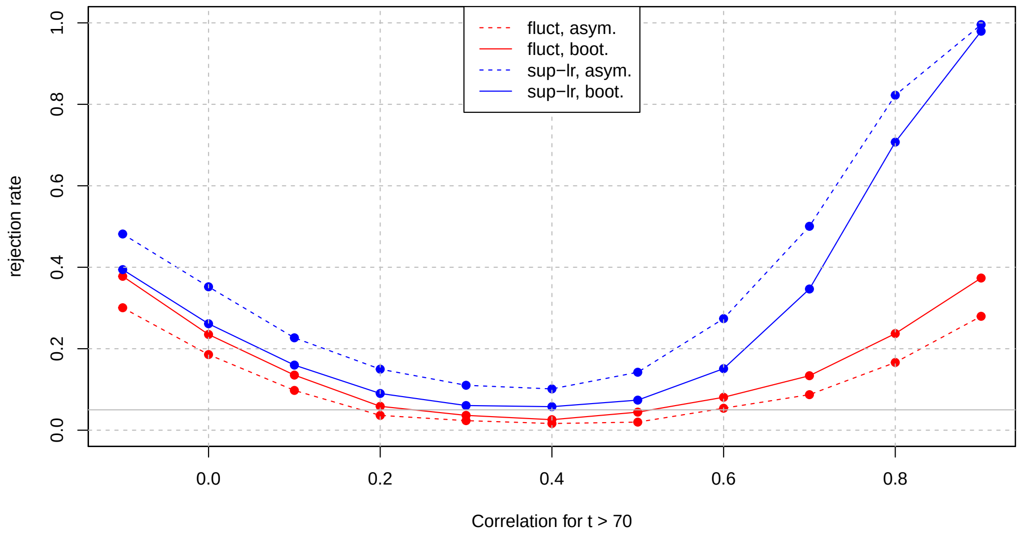


Figure 2: Bivariate Gaussian, n=500, Scenario 2: Empirical Power

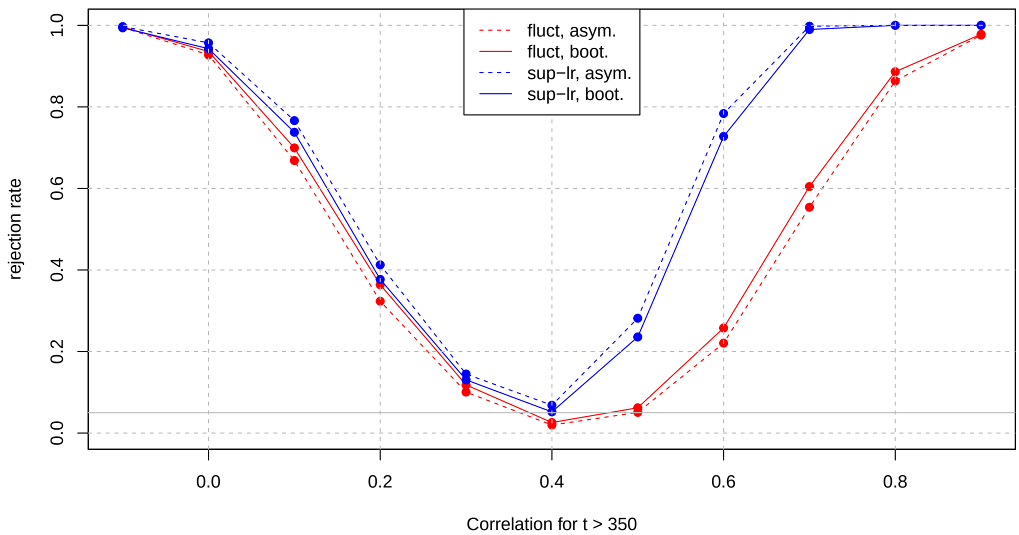
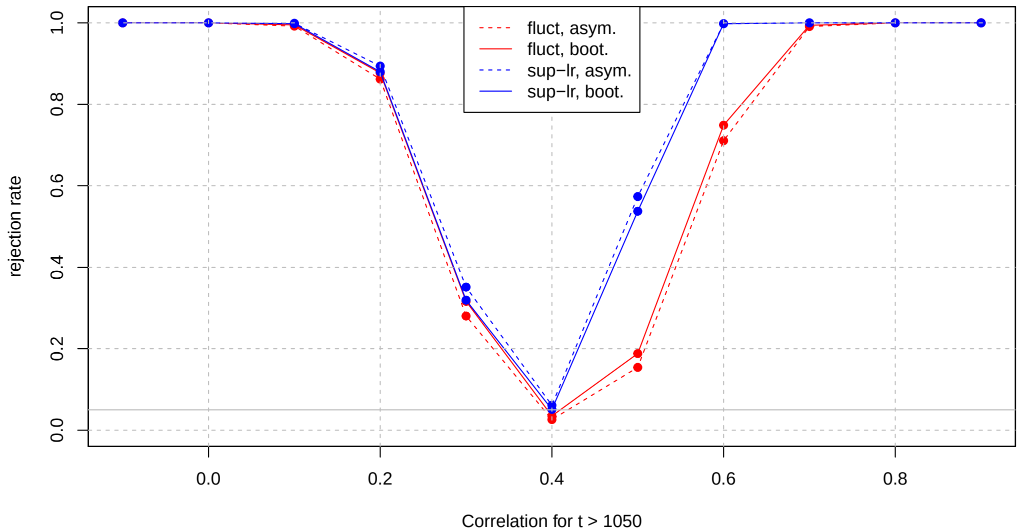


Figure 3: Bivariate Gaussian, n=1500, Scenario 2: Empirical Power



If one compares the bootstrap-corrected versions indicated by solid lines, the results for testing constant correlation uniformly favor the sup-LR test. Although the fluctuation test for constant *marginal* distributions outperforms the sup-LR test across all sample sizes (see the upper panel of table 1), this result does not carry over to the second step of the procedure, where the parametric framework delivers significantly higher power when testing constant correlation. For example, at $\rho_2 = 0.6$ and $n = 500$, the 95 %-confidence intervals are $[0.700, 0.755]$ for the sup-LR test and $[0.231, 0.285]$ for the fluctuation test.
Next we consider dimensionality effects for different sample sizes. As can be seen from the lower two panels of table 1, the residual effect slowly declines with dimension m in the fluctuation test framework and even increases with m in the sup-LR-test framework. Hence, empirical power in figure 4 - 6 is compared only using the respective bootstrap schemes.

Figure 4: Multivariate Gaussian, n=100, Scenario 2: Empirical Power

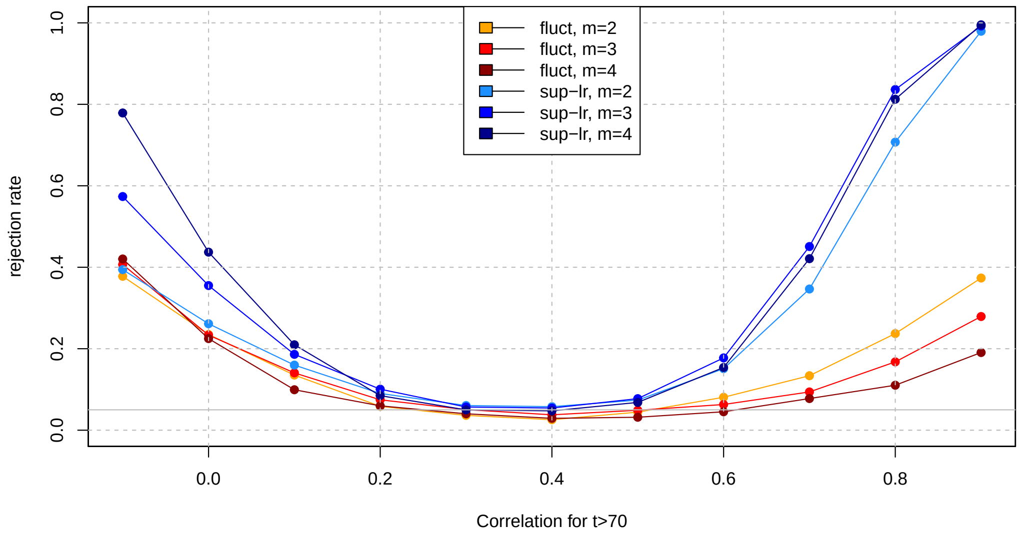


Figure 5: Multivariate Gaussian, n=500, Scenario 2: Empirical Power

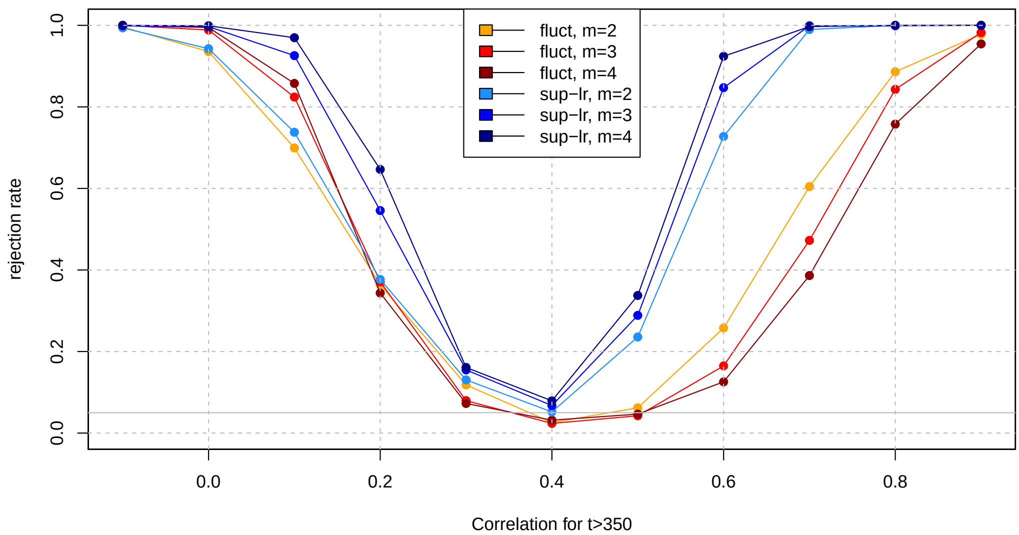
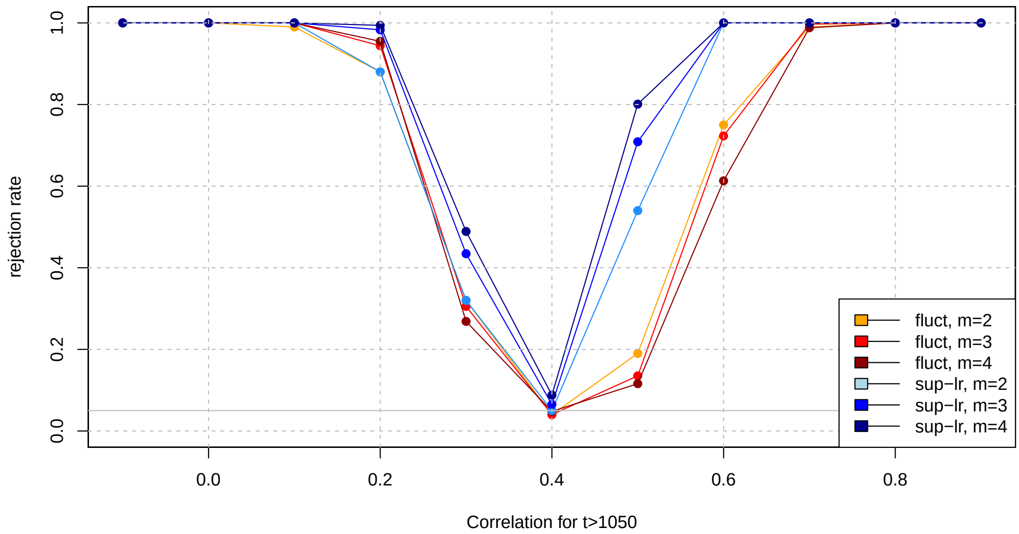


Figure 6: Multivariate Gaussian, n=1500, Scenario 2: Empirical Power



Across all sample sizes, the performance of the sup-LR test increases with the dimension m. For $n = 500$ and $\rho_2 = 0.6$, empirical power increases to 85 % (90 %) in the three (four)-dimensional case, from around 65 % in the bivariate case. The dimensionality effect is largely absent in the fluctuation test framework, where in the case of increasing correlation empirical power even declines with the dimension m. Consequently the size advantage is now even wider, the confidence intervals being $[0.825, 0.870]$ for the sup-LR test and $[0.142, 0.188]$ for the fluctuation test in the three-dimensional case. Finally, in table 2, we consider Monte Carlo bias and root mean-squared error of the break point estimator.

Table 2: Multivariate Gaussian, Scenario 2: Break Point Estimation

ρ_2	$n = 100$				$n = 500$				$n = 1500$			
	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 2$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$
-0.1	-7.12	15.52	-4.99	18.50	-15.90	31.19	0.30	28.67	-24.86	48.19	3.63	20.10
0.1	-11.85	22.58	-10.63	24.28	-27.49	54.98	-13.31	65.87	-48.15	92.74	5.31	61.23
0.3	-17.07	27.62	-15.91	28.64	-70.63	124.07	-78.39	139.11	-136.95	268.24	-141.85	334.66
0.5	-17.94	28.12	-15.42	28.41	-81.37	136.53	-58.51	119.45	-178.67	341.36	-108.79	271.80
0.7	-13.42	22.41	-6.22	20.00	-33.39	72.49	-3.29	28.50	-39.56	109.57	-3.06	23.60
0.9	-8.77	16.08	1.28	5.61	-14.38	40.12	1.24	5.07	-12.41	50.33	0.90	3.84
	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 3$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$
-0.1	-6.04	12.92	-2.60	14.39	-11.48	11.04	2.17	11.26	-19.44	45.56	3.03	9.31
0.1	-11.85	20.36	-10.73	23.95	-21.94	41.99	-3.38	42.12	-38.49	99.70	6.80	27.92
0.3	-16.50	24.74	-17.27	29.60	-73.02	114.19	-75.60	136.95	-146.71	340.15	-97.70	279.38
0.5	-18.13	25.37	-14.83	28.50	-96.84	134.16	-57.00	118.38	-211.63	254.20	-65.65	207.47
0.7	-15.77	22.30	-3.51	17.40	-43.67	77.62	-0.62	17.49	-36.49	67.02	-1.23	12.75
0.9	-11.36	17.70	2.21	3.80	-16.55	39.81	1.64	2.72	-10.29	34.16	1.54	2.59
	Fluctuation test		sup-LR test		Fluctuation test		sup-LR test		Fluctuation test		sup-LR test	
$m = 4$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$
-0.1	-6.14	12.07	-0.16	9.59	-10.38	18.78	1.94	6.28	-15.29	26.51	2.04	5.10
0.1	-10.29	18.14	-6.78	20.63	-22.20	43.03	-2.99	35.95	-34.07	57.30	4.05	19.06
0.3	-15.78	23.42	-15.73	29.20	-72.64	109.31	-71.08	130.44	-145.14	240.08	-87.12	262.16
0.5	-17.56	24.13	-14.98	28.89	-101.63	131.12	-46.20	109.92	-231.42	332.23	-47.67	180.04
0.7	-17.54	23.20	-3.43	17.85	-77.96	83.47	1.08	11.46	-48.84	118.42	0.41	9.11
0.9	-14.79	20.39	2.63	4.09	-24.42	44.31	1.86	2.37	-8.22	41.38	1.74	2.22

Table 2 paints a very clear picture: except for $n = 100$, $m = 2$ and small ρ_2 both bias and variance are considerably smaller for the sup-LR test. The fluctuation test underestimates l_d even for a change from $\rho_1 = 0.4$ to $\rho_2 = 0.9$ and $n = 1500$ as compared to the sup-LR test, which has a negligible bias even for $\rho_2 = 0.7$ (second panel). Results in the higher-dimensional case (the bottom two panels) also favour the parametric framework: while using the sup-LR framework the regime shift is estimated very accurately in the 4-dimensional case compared to the (also precise) estimates in the two and three-dimensional set-up, break-point estimation is not considerably improved with m in the fluctuation test framework. This is especially true for situations of shift contagion, namely for increases in P. Since scenarios with potential shift contagion are usually associated with increasing correlation, the preceding findings suggest to use the sup-LR test, in particular when a precise estimation of the change-point is desired.
Although there is some inconclusiveness for small samples we summarize from the simulation results laid out in this section, that in moderate to large samples the sup-LR test with the residual-bootstrap scheme has acceptable size properties under H_0 and outperforms the fluctuation test with a wild bootstrap scheme both in terms of detecting and estimating regime-shifts.

3.2 t-copula

Test Statistics An increasingly popular way to model marginal and joint distribution in separate steps is the use of copulas. When the copula is specified in some parametric form, our testing framework directly can be directly applied. More specifically we explore a setting where the cross-sectional dependence structure follows a t-copula. We still impose GARCH-processes at each margin, such that step 1 of the sequential testing procedure remains unchanged. Given the test result at step 1, data are standardized and subsequently transformed by the cumulative distribution function of the Gaussian distribution, denoted $F(X_{i,t}, \theta_{i,t})$ evaluated at the ML-estimator $\hat{\sigma}_i$:

$$\begin{aligned}
&\hat{U}_{i,t} = F(X_i, \hat{\sigma}_{i,1}) \quad \text{for } i = 1, ..., \hat{l}_i \\
&\hat{U}_{i,t} = F(X_i, \hat{\sigma}_{i,2}) \quad \text{for } i = \hat{l}_i + 1, ..., n \quad \text{if the test rejects} \\
&\hat{U}_{i,t} = F(X_i, \hat{\sigma}_{i,0}) \quad \text{for } i = 1, ..., \hat{n} \quad \text{if not}
\end{aligned} \tag{9}$$

The test for a constant t-copula is now based on $\hat{U}$. From the probability density of the t-copula

$$c(\hat{U}_t; P, \nu) = \frac{\Gamma(\frac{\nu+m}{2})\big(\Gamma(\frac{\nu}{2})\big)^{m-1}}{\big(\Gamma(\frac{\nu+1}{2})\big)^m |P|^{0.5}} \left(\prod_{i=1}^{m} \big(1 + \frac{\hat{Y}_{i,t}^2}{\nu}\big)^{\frac{\nu+1}{2}} \right) \left(1 + \frac{1}{\nu} \hat{Y}_t' P^{-1} \hat{Y}_t \right)$$

with $\hat{Y}_{i,t} = F^{-1}(\hat{U}_{i,t}, \nu)$ denoting the quantile function of a standardized t-distribution and $l\Gamma$ the $\log - \Gamma$-function, the log-likelihood follows as

$$\begin{aligned}
L(\hat{U}; P_0, \nu) = & n \cdot \left(l\Gamma(\frac{\nu+m}{2}) + (m-1) l\Gamma(\frac{\nu}{2}) - m \cdot l\Gamma(\frac{\nu+1}{2}) - 0.5 \log |P_0| \right) \\
& + \sum_{t=1}^{j} \left(\frac{\nu+1}{2} \sum_{i=1}^{m} \log \left(1 + \frac{\hat{Y}_{i,t}^2}{\nu}\right) - \frac{\nu+m}{2} \log \left(1 + \frac{\hat{Y}_t' P_0^{-1} \hat{Y}_t}{\nu_0}\right) \right)
\end{aligned}$$

for the full-sample and

$$
\begin{aligned}
L(\hat{U};P_1,P_2,\nu) =& j \cdot \Big(l\Gamma(\frac{\nu+m}{2}) + (m-1)l\Gamma(\frac{\nu}{2}) - m \cdot l\Gamma(\frac{\nu+1}{2}) - 0.5\log|P_1| \Big) \\
&+ \sum_{t=1}^{j} \left(\frac{\nu_0+1}{2} \sum_{i=1}^{m} \log\left(1+\frac{\hat{Y}_{i,t}^2}{\nu_0}\right) - \frac{\nu_1+m}{2}\log\left(1+\frac{\hat{Y}_t' P_1^{-1}\hat{Y}_t}{\nu_0}\right)\right) \\
&(n-j) \cdot \Big(l\Gamma(\frac{\nu+m}{2}) + (m-1)l\Gamma(\frac{\nu}{2}) - m \cdot l\Gamma(\frac{\nu+1}{2}) - 0.5\log|P_2| \Big) \\
&+ \sum_{t=j+1}^{n} \left(\frac{\nu+1}{2} \sum_{i=1}^{m} \log\left(1+\frac{\hat{Y}_{i,t}^2}{\nu}\right) - \frac{\nu+m}{2}\log\left(1+\frac{\hat{Y}_t' P_2^{-1}\hat{Y}_t}{\nu}\right)\right)
\end{aligned}
$$

for the partial samples. ML-estimation requires numerical methods, such as EM-algorithms. Let $(\hat{P}_0,\hat{\nu})$ and $(\hat{P}_1,\hat{P}_2)$ denote the full-sample and partial-sample ML-estimator, we have for a fixed j:

$$
A_j = 2\left(L(\hat{U};\hat{P}_1,\hat{P}_2,\hat{\nu}) - L(\hat{U};\hat{P}_0,\hat{\nu}) \right)
$$

It is reasonable to approximate the distribution of the corresponding sup-LR test statistic by the distribution $\sup_{\pi\in\Pi} \mathcal{B}_{(m-1)m/2}(\pi)$ under the null hypothesis, if the residual effect could be ignored. Full ML-estimation of the t-copula is extremely time-consuming, particularly in higher dimensions, Demarta and McNeil [2005] therefore suggest a semi-parametric pseudo-ML procedure sharing the asymptotic properties of full ML-estimation. In a first step, the empirical Kendall's tau matrix $\hat{P}^\tau$ of the residuals is calculated as

$$
\hat{P}^\tau = \begin{pmatrix} \hat{\rho}_\tau(\hat{Z}_1,\hat{Z}_1) & \cdots & \hat{\rho}_\tau(\hat{Z}_1,\hat{Z}_n) \\ \vdots & \ddots & \vdots \\ \hat{\rho}_\tau(\hat{Z}_n,\hat{Z}_1) & \cdots & \hat{\rho}_\tau(\hat{Z}_n,\hat{Z}_n) \end{pmatrix}
$$

where each element is given as the empirical pairwise Kendall's tau coefficient.

$$
\hat{\rho}_\tau(\hat{Z}_n,\hat{Z}_n) = \binom{n}{2}^{-1} \sum_{1\le t_1<t_2\le n} \operatorname{sign}\left((\hat{Z}_{t_1,i}-\hat{Z}_{t_2,i})(\hat{Z}_{t_1,j}-\hat{Z}_{t_2,j}) \right)
$$

The empirical Kendall's tau matrix serves to construct a method-of-moments estimator for P by $P^* = \sin(\frac{\pi}{2}\hat{P}^\tau)$ and subsequently estimate ν_C, holding P^* fixed. Since the estimator $\hat{\nu}_C$ does not have a closed-form solution, we follow Mashal and Zeevi [2002] and perform a simple bisection algorithm over the first-order condition of the log-likelihood with respect to ν_C. As pointed out by Mashal and Zeevi [2002] using Pseudo-ML-estimators affects the limit distribution. This is unproblematic in our case, because a bootstrap scheme that could also be used to approximate the appropriate limit distribution in small samples, is already at hand.

In our simulations, we also consider the effects of misspecification of the t-copula. More precisely, we assume that the marginal distributions are correctly specified and tested but that the underlying copula is mistakenly assumed to be Gaussian, which (as a by-product) reduces the computational effort in higher dimensions. The deviation of the test statistic is relegated to appendix D. One could of course also discuss effects of misspecification at the margins. As this section is concerned with the performance of non-parametric and parametric copula tests, the issue is omitted here.

Finite-Sample Properties Finite-sample properties are evaluated with data generated from a t_4-copula and subsequently transformed using the same GARCH(1,1)-process from as before:

$$\begin{aligned}
&U_t \overset{i.i.d.}{\sim} C_t(P_1, 4) && \text{for } t = 1, ..., l_D \\
&U_t \overset{i.i.d.}{\sim} C_t(P_2, 4) && \text{for } t = l_D, ..., n \\
&Z_{1,t} = \Phi^{-1}(U_{1,t}) && \\
&Z_{2,t} = \Phi^{-1}(U_{2,t}) && \\
&h_{1,1}^2 = 1\ , \quad h_{1,t}^2 = \alpha Z_{1,t-1}^2 + \beta h_{1,t-1}^2 && \text{for } t = 2, ..., n \\
&X_{1,t} = h_{1,t} Z_{1,t} && \text{for } t = 1, ..., l_1 \\
&X_{1,t} = s h_{1,t} Z_{1,t} && \text{for } t = l_1, ..., n \\
&h_{2,1}^2 = 1\ , \quad h_{2,t}^2 = \alpha Z_{2,t-1}^2 + \beta h_{2,t-1}^2 && \text{for } t = 2, ..., n \\
&X_{2,t} = h_{2,t} Z_{2,t} && \text{for } t = 1, ..., l_2 \\
&X_{2,t} = s h_{2,t} Z_{2,t} && \text{for } t = l_2, ..., n
\end{aligned}$$

with

$$P_1 = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix}$$

Generalizations to higher-dimensional cases are obtained by extending the correlation matrix and subsequently transform data with the quantile function accordingly. Change-points are set as before, however we restrict our attention to the cases of $n \in \{100, 500\}$ and $m \in \{2, 3\}$. Under this DGP we compare the non-parametric benchmark-test based on the empirical copula from Bücher et al. [2014], lined out in appendix A, with the sup-LR test under correct specification and under misspecification as Gaussian copula.

Table 3: t-copula, Scenario 1: Rejection Rates under H_0

s	$n = 100$					$n = 500$				
$m = 2$	EC-test	t-Cop		Gauss		EC-test	t-Cop		Gauss	
		asym	boot	asym	boot		asym	boot	asym	boot
1	0.043	0.031	0.032	0.247	0.043	0.051	0.053	0.040	0.316	0.046
1.5	0.043	0.045	0.055	0.261	0.054	0.048	0.076	0.054	0.322	0.050
2	0.034	0.044	0.053	0.247	0.068	0.050	0.067	0.048	0.310	0.054
2.5	0.037	0.042	0.047	0.249	0.054	0.051	0.069	0.046	0.316	0.053
3	0.042	0.039	0.045	0.242	0.051	0.052	0.066	0.044	0.311	0.051
$m = 3$	EC-test	t-Cop		Gauss		EC-test	t-Cop		Gauss	
		asym	boot	asym	boot		asym	boot	asym	boot
1	0.051	0.020	0.040	0.249	0.037	0.053	0.049	0.039	0.391	0.058
1.5	0.037	0.039	0.052	0.248	0.046	0.058	0.077	0.064	0.406	0.058
2	0.033	0.047	0.062	0.257	0.052	0.053	0.078	0.054	0.402	0.055
2.5	0.035	0.028	0.055	0.246	0.045	0.057	0.077	0.053	0.395	0.053
3	0.037	0.032	0.046	0.236	0.050	0.061	0.074	0.056	0.392	0.053

The empirical copula test keeps its size when testing for a constant copula, indicating that piecewise standardization appropriately accounts changes in the margins and the i.i.d. multiplier process can be applied. Except for the case of moderate changes at the margins, where Hypotheses Pair 2 is rejected too frequently, this also holds for the correctly specified sup-LR test, suggesting that the residual effect is present but less prominent than in the previous section. Unsurprisingly, the misspecified sup-LR test does not keep its size given the nominal level of 5 %, this is however appropriately corrected by the residual bootstrap. Using the scheme lined out in section 2, empirical rejection rates are similar to the bootstrapped sup-LR test under correct specification.

Figure 7: bivariate t-copula, n=100, Scenario 2: Empirical Power

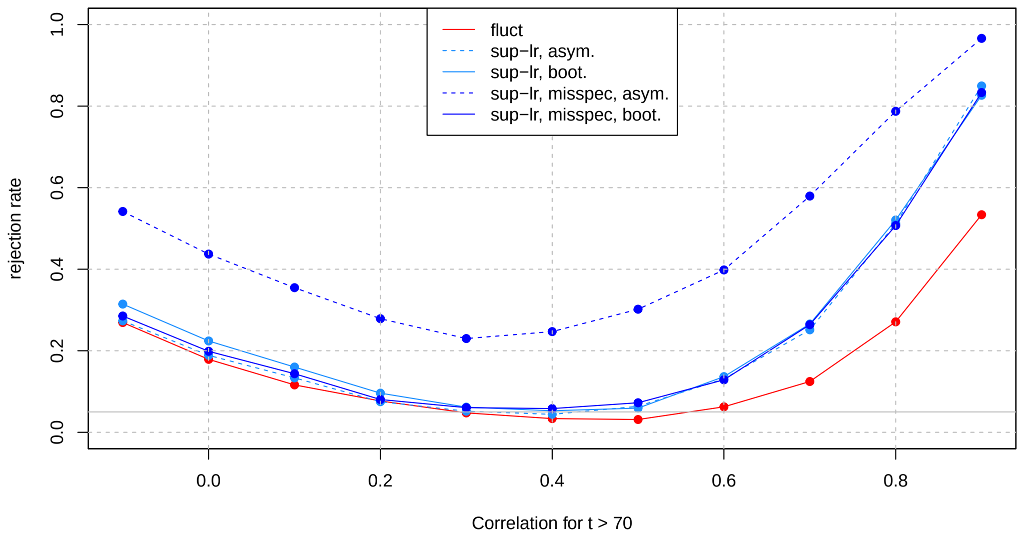


Figure 8: bivariate t-copula, n=500, Scenario 2: Empirical Power

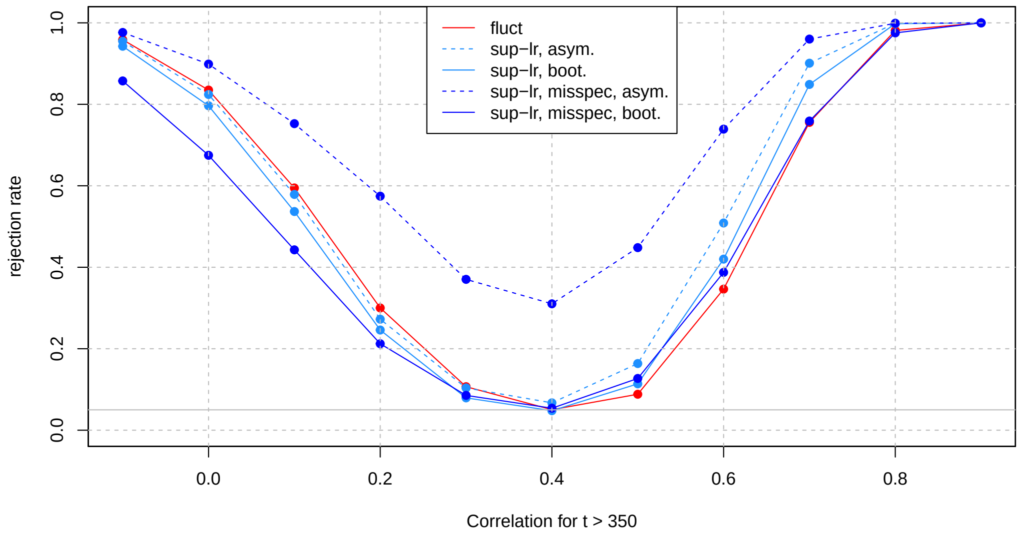


Figure 9: 3-variate t-copula, n=100, Scenario 2: Empirical Power

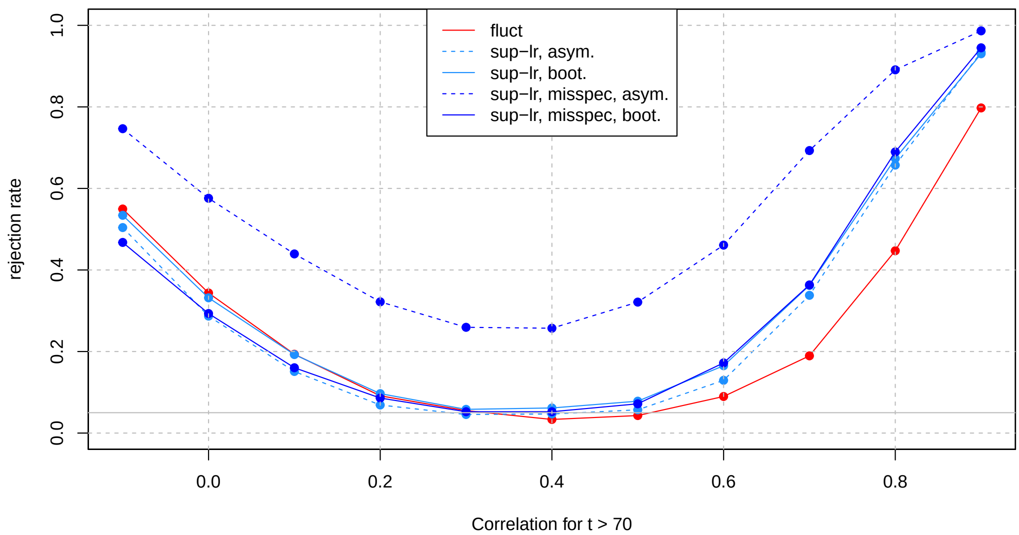
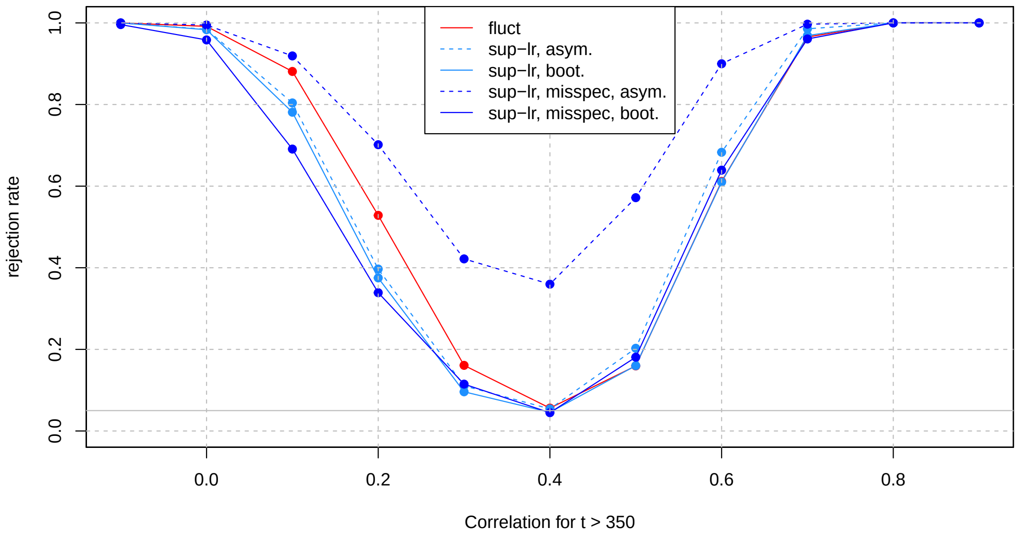


Figure 10: 3-variate t-copula, n=500, Scenario 2: Empirical Power



In contrast to before, results are mixed as no test is strictly superior. It appears that the fluctuation test has slightly higher power if $\rho_2 < \rho_1$ while the result is flipped for $\rho_2 > \rho_1$. This would be a prime situation for using both tests together with the Boole-Bonferroni correction. Remarkably, the sup-LR test under misspecification does not perform consistently worse than its correctly specified counterpart.

Table 4 however reveals that the sup-LR test yields considerably better results in estimating the change-point, irrespective of whether the model is correctly specified or not. Even for $n = 500$ and $\rho_2 = 0.9$, l_D is visibly biased in the fluctuation test framework, while l_D is already precisely estimated for $\rho_2 = 0.7$. We also observe similar or even better results of the misspecified sup-LR test compared to its correctly specified counterpart for samples that might be too small for reliably estimating a t-copula. In addition, using the potentially misspecified sup-LR test with a Gaussian copula has also large computational advantages. Therefore we suggest using the sup-LR test whenever precise estimation of the change-point is required. Within the sup-LR framework usage of any advanced model is only advantageous if one has high confidence on the model's appropriateness and the sample size (relative to dimension) permits reliable estimation.

Table 4: Multivariate t-copula, Scenario 2: Copula-Break Point Estimation

	$n = 100$						$n = 500$					
ρ_2	EC-test		sup-LR		sup-LR, mis.		EC-test		sup-LR		sup-LR, mis.	
$m = 2$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$
-0.1	-10.09	18.02	-8.22	20.84	-8.76	21.06	-16.99	36.51	-2.30	36.03	-6.59	36.59
0.1	-14.06	21.95	-13.29	25.60	-12.81	25.36	-33.93	67.26	-24.15	76.52	-33.71	83.62
0.3	-17.93	25.32	-17.88	29.47	-16.17	28.46	-80.81	117.94	-81.27	113.03	-88.56	143.59
0.5	-20.52	27.01	-17.55	28.95	-15.80	27.93	-90.09	124.42	-79.34	112.89	-85.13	141.39
0.7	-16.82	23.06	-10.04	22.56	-9.71	22.42	-37.46	64.75	-4.99	38.94	-13.56	41.27
0.9	-10.41	15.85	-1.47	10.76	-0.37	10.67	-14.43	25.56	0.00	9.19	2.12	9.43

ρ_2	EC-test		sup-LR		sup-LR, mis.		EC-test		sup-LR		sup-LR, mis.	
$m = 3$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$	$bias(l_D)$	$rmse(l_D)$
-0.1	-8.06	13.18	-4.21	14.26	-3.50	14.06	-12.22	22.26	0.72	12.88	0.94	12.90
0.1	-12.32	19.03	-8.62	21.14	-10.62	22.03	-23.45	45.57	-10.17	52.72	-17.04	54.46
0.3	-17.32	24.03	-14.49	26.63	-15.27	27.06	-74.73	107.64	-79.06	137.33	-88.67	143.08
0.5	-18.94	25.09	-16.16	28.12	-14.22	27.05	-79.08	79.84	-69.52	127.49	-73.79	129.87
0.7	-13.36	19.72	-8.18	20.32	-6.05	19.56	-22.87	38.31	-2.97	26.25	-1.85	26.15
0.9	-6.70	11.65	-0.18	7.18	0.98	7.24	-8.21	13.93	0.40	4.20	2.12	4.69

4 Application to European Financial Sector Stocks

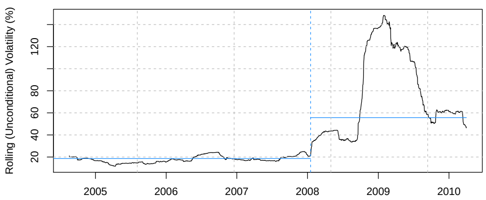
In a second practical example, we apply the non-parametric methods testing constant copula as lined out in section 3.2 to EURO STOXX 50 bank equity data around the financial crisis following the Lehman Brothers insolvency. Therefore we take daily log-returns of BNP Paribas, Santander, ING Group, BBVA, Intesa Sanpaolo, Société Générale, and Deutsche Bank from April-01-2004 to April-01-2010. All margins are assumed to follow a GARCH(1,1)-process. Based on critical values 12.35 for the demeaned sup-LR test on constant unconditional variances and 1.628 for the CUSUM of squares test (at 99 % nominal level respectively), Hypotheses Pair 1 is rejected at each margin. Although there are some minor differences in the break point estimates, this can also be seen graphically from rolling (annualized) volatilities in section 4. Here, change-point and partial-sample estimators for $\sigma_{i,1}$ and $\sigma_{i,2}$ are based on the results of the sup-LR test. Before moving to the second step, we perform two diagnostic tests on the piecewise GARCH-residuals, reported in the leftmost columns of section 4: the Ljung-Box test for autocorrelation does not reject the null hypothesis for any dimension at the 10 % level, indicating that serial dependence is eliminated by computing GARCH-residuals sufficiently well. We report the test result for a lag-size of four, results are robust to different lags. The Kolmogorov-Smirnov test against normality rejects in four of seven dimensions at the 1 % level and always at the 10 % level, such that we operate the sup-LR test under misspecification.

Table 5: Estimation of EURO STOXX Financial Sector Stocks

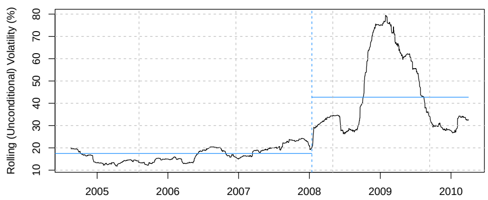
	Empirical copula test				sup-LR test, Gauss-Copula				Diagnostic tests	
	Q_i	$\hat{\lambda}_i$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	A_i	$\hat{\lambda}_i$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	LB	KS
BNP Paribas	4.92	2008-01-18	20.58	59.09	825.17	2008-01-17	20.48	59.10	0.48	0.09
Santander	5.41	2008-01-15	17.59	48.17	759.58	2008-01-14	17.48	48.20	0.89	1.4e-4
ING Group	5.47	2008-09-15	23.76	98.88	1579.87	2008-01-17	18.62	85.97	0.53	1.1e-7
BBVA	5.51	2008-01-15	17.11	44.92	698.93	2008-01-14	17.03	44.94	0.11	4.8e-5
Intesa Sanpaolo	4.83	2008-07-11	20.13	57.15	805.61	2008-07-10	20.00	57.20	0.76	0.02
Société Générale	5.93	2008-01-18	21.95	62.42	855.56	2007-07-25	19.23	58.40	0.11	0.04
Deutsche Bank	5.54	2008-07-01	22.06	72.15	1028.61	2008-06-24	21.86	71.94	0.84	2e-04
Copula (p-value)	0.0025	2007-06-05			0	2008-01-15				

As for the correlation matrix P, the null hypothesis of parameter stability is overwhelmingly rejected; almost all 21 pairwise correlations increase after the change-point. From the simulation results on $\hat{l}_D$ in section 4.4, 2008-01-15 might be the more reliable change-point estimate, largely coinciding with variance change-points of the single stocks. In addition, the change-point estimate delivered by the non-parametric test is somewhat peculiar, since it falls into a period before the actual extent of the financial turmoil became publicly known. Using a sequential procedure again avoids biased change-point estimates, something that could well occur using non-sequential procedures.

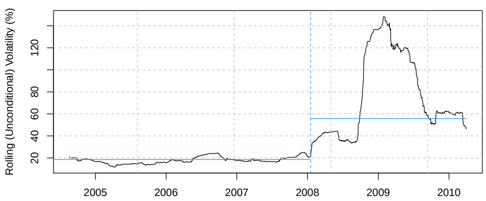
Figure 11: EURO STOXX Banks, Rolling Volatilities



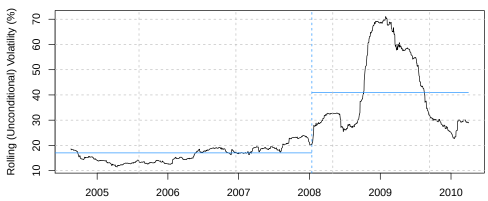
(a) BNP Paribas



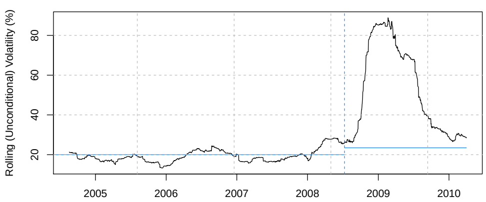
(b) Santander



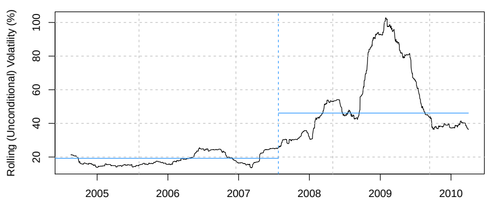
(c) ING



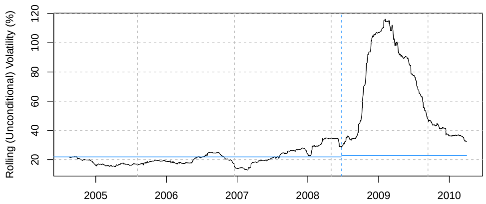
(d) BBVA



(e) Intensa Sanpaolo



(f) Société Générale



(g) Deutsche Bank

5 Concluding Remarks

We have proposed and analyzed parametric two-step procedures for assessing the stability of cross-sectional dependency measures in the presence of potential breaks in the marginal distributions. We have focused on sup-LR tests and it could be interesting or further research to also look at sup-Wald, sup-LM or exponentially weighted test statistics in the spirit of Andrews and Ploberger [1994]. Moreover, while we have tackled the case of serial dependence when discussing volatility filtering, it might be interesting to also investigate e.g. changes in VAR-filtering in the first step of the procedure.

Proofs

Proof of Theorem 1

We sketch the proof for Hypotheses Pair 2 with the proof for Hypotheses Pair 1 being similar. The proof is based on Taylor approximations of $l_{i,t}(\theta_i, \hat{\eta})$ around the first component. First, it holds

$$\frac{1}{2}A_{i,\pi n} = \sum_{t=1}^{\pi n}\left(l_t(\hat{\delta}_1, \hat{\theta}_0, \hat{\eta}) - l_t(\hat{\delta}_0, \hat{\theta}_0, \hat{\eta})\right) + \sum_{t=\pi n+1}^{n}\left(l_t(\hat{\delta}_2, \hat{\theta}_0, \hat{\eta}) - l_t(\hat{\delta}_0, \hat{\theta}_0, \hat{\eta})\right).$$

As we consider ML estimators, the first derivatives cancel so that we obtain with the assumption on the third derivatives

$$A_{i,\pi n} = \left(\hat{\delta}_0 - \hat{\delta}_1\right)' \sum_{t=1}^{\pi n} \frac{\partial^2}{\partial\delta\partial\delta'} l_t(\hat{\delta}_1, \hat{\theta}_0, \hat{\eta}) \left(\hat{\delta}_0 - \hat{\delta}_1\right) + \left(\hat{\delta}_0 - \hat{\delta}_2\right) \sum_{t=\pi n+1}^{n} \frac{\partial^2}{\partial\delta\partial\delta'} l_t(\hat{\delta}_2, \hat{\theta}_0, \hat{\eta}) \left(\hat{\delta}_0 - \hat{\delta}_2\right) + o_p(1).$$

With the assumption on the process convergence of the estimators and the continuous mapping theorem, we obtain that this term converges in process distribution to

$$\begin{aligned} & \left(\sqrt{\pi}\Gamma_k(1) - \frac{1}{\sqrt{\pi}}\Gamma_k(\pi)\right)' \left(\sqrt{\pi}\Gamma_k(1) - \frac{1}{\sqrt{\pi}}\Gamma_k(\pi)\right) \\ + & \left(\sqrt{1-\pi}\Gamma_k(1) - \frac{1}{\sqrt{1-\pi}}\Gamma_k(1-\pi)\right)' \left(\sqrt{1-\pi}\Gamma_k(1) - \frac{1}{\sqrt{1-\pi}}\Gamma_k(1-\pi)\right). \end{aligned}$$

Some algebra using that it holds for a Brownian motion that $\mathsf{Cov}(\Gamma_{k_\delta}(s), \Gamma_{k_\delta}(t)) = \mathsf{min}(s,t) I_{k_\delta}$ (see Webel and Wied [2016]) reveals that this has the same distribution as

$$\frac{(\Gamma_k(\pi) - \pi\Gamma_k(1))'(\Gamma_k(\pi) - \pi\Gamma_k(1))}{\pi(1-\pi)}.$$

□

Proof of Theorem 2

A Taylor approximation of $A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \hat{\theta}_{\pi n}, \hat{\eta})$ in the third component around $\theta_{\pi n}$ yields

$$\begin{aligned}
A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \hat{\theta}_{\pi n}) &= \sum_{t=1}^{n} \left(l_t(\tilde{\delta}_{\pi n}, \hat{\theta}_t) - l_t(\hat{\delta}, \hat{\theta}_t) \right) \\
&= \sum_{t=1}^{n} \left(l_t(\tilde{\delta}_{\pi n}, \theta_t) - l_t(\hat{\delta}, \theta_t) \right) \\
&+ \sum_{t=1}^{n} \left(\frac{\partial}{\partial \theta_t} \left(l_t(\tilde{\delta}_{\pi n}, \theta_t) - l_t(\hat{\delta}, \theta_t) \right) (\hat{\theta}_t - \theta_t) \right) \\
&+ \frac{1}{2} \sum_{t=1}^{n} (\hat{\theta}_t - \theta_t)' \frac{\partial^2}{\partial \theta_t \partial \theta_t'} \left(l_t(\tilde{\delta}_{\pi n}, \theta_t) - l_t(\hat{\delta}, \theta_t) \right) (\hat{\theta}_t - \theta_t) \\
&+ o_p(1) \\
&= A_{\pi n}(\tilde{\delta}_{\pi n}, \hat{\delta}, \theta_t) + B_{\pi n} + \frac{1}{2} C_{\pi n} + o_p(1).
\end{aligned}$$

It holds that

$$C_{\pi n} = \sqrt{n}(\hat{\theta} - \theta)' \frac{\partial^2}{\partial \theta \partial \theta'} \frac{1}{n} \sum_{t=1}^{n} \left(l_t(\tilde{\delta}(\pi), \theta) - l_t(\hat{\delta}, \theta) \right) \sqrt{n}(\hat{\theta} - \theta)$$

Then, $C_{\pi n} \Rightarrow_p 0$ and $B_{\pi n} \Rightarrow_d R(\pi)$. So, $A_{\pi n} \Rightarrow \mathcal{B}_k(\pi) + R(\pi)$. □

References

E. Andreou and E. Ghysels. Tests for breaks in the conditional co-movements of asset returns. *Statistica Sinica*, 13:1045–1074, 2003.

D. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414, 1994.

D. Andrews and W. Ploberger. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Annals of Statistics*, 23:1609–1629, 1995.

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

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

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

S. Bartram, S. Taylor, and Y.-H. Wang. The euro and european financial market integration. *Journal of Banking and Finance*, 31:1461–1481, 2007.

A. Bücher, I. Kojadinovic, T. Rohmer, and J. Segers. Detecting changes in cross-sectional dependence in multivariate time series. *Journal of Multivariate Analysis*, 132:111–128, 2014.

A. Bücher, S. Jäschke, and D. Wied. Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics*, 187(1):154–168, 2015.

D. Blatt, B. Candelon, and H. Manner. Detecting contagion in a multivariate time series system: An application to sovereign bond markets in europe. *Journal of Banking and Finance*, 59(C):1–13, 2015.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

B. Candelon and H. Manner. Testing for asset market linkages: A new approach based on time-varying copulas. *Pacific Economic Review*, 15(3):364–384, 2010.

N. Chan, J. Chen, X. Chen, Y. Fan, and L. Peng. Statistical inference for multivariate resiudal copula of garch models. *Statistica Sinica*, 19:53–70, 2009.

C. Chuang and J. Lee. Economic integration and structure change in stock market dependence: Empirical evidences of cepa. *Journal of Applied Finance and Banking*, 4(2):33–45, 2014.

R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

S. Demarta and A. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.

M. Demetrescu and D. Wied. Residual-based inference on moment hypotheses, with an application to testing for constant correlation. *Econometrics Journal*, 22(1):10–33, 2019.

A. Dias and P. Embrechts. Change-point analysis for dependence structures in finance and insurance. In C. Szegoe, editor, *Risk Measures for the 21st Century*. Wiley Finance Series, Chichester, 2004.

F. Duan and D. Wied. A residual-based multivariate constant correlation test. *Metrika*, 81(6):653–687, 2018.

K. Forbes and R. Rigobon. No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, 57(5):2223–2261, 2002.

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

J. Kiefer. K-sample analogues of the kolmogorov-smirnov and cramér-v.mises tests. *Annals of Mathematical Statistics*, 30(2):420–447, 1959.

M. King and S. Wadhwani. Transmission of volatility between stock markets. *Review of Financial Studies*, 3(1):281–307, 1990.

R. Mashal and A. Zeevi. Beyond correlation: extreme co-movements between financial assets. *Unpublished*, 2002.

Z. Qu and P. Perron. Estimating and testing multipile structural changes in multivariate regressions. *Econometrica*, 75(2):459–502, 2007.

B. Rémillard. Goodness-of-fit tests for copulas of multivariate time series. *Econometrics*, 5: 13, 2017.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

K. Webel and D. Wied. *Stochastische Prozesse*. Springer Gabler, 2016.

D. Wied, M. Arnold, N. Bissantz, and D. Ziggel. A new fluctuation test for constant variances with applications to finance. *Metrika*, 75(8):1111–1127, 2012a.

D. Wied, W. Krämer, and H. Dehling. Testing for a change in correlation at unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3):570–589, 2012b.