Supplementary Material: Testing Constant Cross-Sectional Dependence with Time-Varying Marginal Distributions in Parametric Models

A Non-parametric Test Frameworks

A.1 Testing Mean-Variance-Stability

Using partial sums of sample moments to test for constant correlation has been suggested by Wied et al. [2012b], who derive the limiting distribution of the sequence of partial sums. The same framework can also be adapted to testing parameter constancy at the marginal distributions: Let $\{Z_t\}_{t=1,...,T}$, $Z_t \in \mathbb{R}^m$ denote an i.i.d. sample without assuming a particular parametric model but rather test hypothesis on sample-moments, specified by a function $g : \mathbb{R}^m \to \mathbb{R}^k$, where k denotes the number of moment hypotheses that are imposed. A fluctuation test is then based on the partial sums

$$Q_j = \frac{j}{\sqrt{n}} \sqrt{(S_j - S_n)' \hat{\Omega}^{-1} (S_j - S_n)} \quad \text{with} \quad S_j = \frac{1}{j} \sum_{t=1}^{j} g(Z_t) \tag{A.1}$$

with covariance matrix Ω, in practice replaced by some estimator $\hat{\Omega}$. The limiting process associated with the general fluctuation test (4.1) is closely related to a k-dimensional Brownian Bridge $B_k(\pi)$:

$$Q_j \Rightarrow \sqrt{(\Gamma_k(\pi) - \pi\Gamma_k(1))'(\Gamma_k(\pi) - \pi\Gamma_k(1))} \tag{A.2}$$

with $\Gamma(s)$ being a k-dimensional vector of independent Brownian Motions defined over $\pi \in [0, 1]$. It is now possible to apply different functionals on the limit process, the sup-functional being the most suitable for the alternative of a single regime-change. Consequently

$$\sup Q_j \to_d \sup ||B_k(\pi)|| \tag{A.3}$$

So for the CUSUM of squares test, we have $k = 1$, if means are also subject to change, $k = 2$. Critical values c_α, such that $P(\sup_{\pi\in\Pi} B_k > c_\alpha) = \alpha$, are tabulated for example in Kiefer [1959] or can easily be simulated. Consequently, H_0 will be rejected if $\sup Q_j$ exceeds the quantile of $\sup ||B_2(\pi)||$ associated with the desired significance level. The change-point is estimated by

$$\hat{l}_i = \underset{2\leq j\leq n-2}{\arg\max}\, Q_j \tag{A.4}$$

where $t = 1, n-1, n$ have to be excluded from the sets of potential break points as each subsample needs to contain at least two elements. Testing time series with higher observation frequency for structural changes is usually performed under the assumptions of constant means, as for example in Wied et al. [2012a]. The latter authors develop a fluctuation test framework using a CUSUM of squares process, which we adopt for the case of daily financial return series. Since μ can not assumed to be constant in low-frequency application, the variance test is slightly generalized in the following by allowing μ_i to break simultaneously with σ_i^2. Under the assumption of Gaussian marginal distributions, mean μ and variance σ^2 are the only parameters subject to structural changes. It is also possible to embed the t_ν-distributional assumption into testing mean-variance stability: if degrees of freedom are assumed to be constant, only one location parameter μ_i, one scale parameter ξ_i are subject to change. In a fluctuation test framework, this corresponds to testing constancy of the first and second moment, $(\mu_{1,i}, \mu_{2,i})'$ of X_i, i indexing the dimension under consideration. This follows from the variance definition $\sigma^2 = \mu_2 - \mu_1^2$ and that μ_1 is also constant under the null hypothesis. No data transformation is required, such that one can write $Z_i = X_i$ and it is possible to directly apply a fluctuation test on $(\mu_1, \mu_2)'$ by assuming that we observe samples from $Z_{i,t} \overset{i.i.d.}{\sim} (\mu_{1,i}, \mu_{2,i}, \mu_{3,i}, \mu_{4,i})$. The moment hypothesis is imposed through

$$g(Z_t) = (Z_i, Z_i^2)$$

For bounded third and fourth moments the asymptotic covariance matrix follows from the Central Limit Theorem, applied to the full-sample estimator of $(\mu_{1,i}, \mu_{2,i})'$, namely the sample moments $(\hat{\mu}_{1,i}, \hat{\mu}_{2,i})'$:

$$\sqrt{n} \begin{pmatrix} \hat{\mu}_{1,i} \\ \hat{\mu}_{2,i} \end{pmatrix} \to_d N\left(\begin{pmatrix} \mu_{1,i} \\ \mu_{2,i} \end{pmatrix} ; \mathrm{Var}\left(\begin{pmatrix} \hat{\mu}_{1,i} \\ \hat{\mu}_{2,i} \end{pmatrix} \right) \right)$$

with

$$\mathrm{Var}\left(\begin{pmatrix} m_{1,i} \\ m_{2,i} \end{pmatrix} \right) = \begin{pmatrix} \mathrm{Var}(Z_i) & \mathrm{Cov}(Z_i, Z_i^2) \\ \mathrm{Cov}(Z_i, Z_i^2) & \mathrm{Var}(Z_i^2) \end{pmatrix} = \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix}$$

The fluctuation test statistic Q_t is computed with

$$S_j = \Big(\frac{1}{j} \sum_{t=1}^{j} Z_t, \frac{1}{j} \sum_{t=1}^{j} Z_t^2 \Big)' \quad \text{where} \quad \hat{\Omega} = \frac{1}{n} \begin{pmatrix} \hat{\mu}_2 - \hat{\mu}_1^2 & \hat{\mu}_3 - \hat{\mu}_1\hat{\mu}_2 \\ \hat{\mu}_3 - \hat{\mu}_1\hat{\mu}_2 & \hat{\mu}_4 - \hat{\mu}_2^2 \end{pmatrix} \tag{A.5}$$

In the Gaussian case, the relevant moments are obtained as

$$\begin{aligned} \mu_1 &= \mu \\ \mu_2 &= \mu^2 + \sigma^2 \\ \mu_3 &= \mu^3 + 3\mu\sigma^2 \\ \mu_4 &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \end{aligned}$$

such that the asymptotic covariance matrix of the full-sample estimator is given by

$$\Omega = \frac{1}{n}\begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix}$$

Several extensions of practical interest can be tested, one could for example suspect that skewness and/or kurtosis are also subject to structural changes. Simulation evidence however revealed that this additional flexibility does not improve testing in either framework.

A.2 Testing Constant Cross-Moments

The principle of using partial sums of empirical moments can be extended to testing constant dependency under the assumption that Pearson's correlation coefficient (or correlation matrix in a higher-dimensional system) is the only parameter of the joint-distribution, that changes between sub-samples and that marginal distribution change only in mean and variance. Trivially satisfied by the multivariate Gaussian, the same methods also apply to a t-distribution with constant degrees of freedom as the correlation coefficient for standardized data is obtained as

$$\rho_{12} = \frac{\frac{\nu}{\nu-2}\xi_{12}}{\sqrt{\frac{\nu}{\nu-2}\xi_{11}} \cdot \sqrt{\frac{\nu}{\nu-2}\xi_{22}}} = \frac{\xi_{12}}{\sqrt{\xi_{11}}\sqrt{\xi_{22}}} = \xi_{12} \tag{A.6}$$

and so the cross-dispersion ξ_{12} is the only dependency-shaping parameter subject to change. In order to test for constant correlation, the observed data are now cleaned from possible changes in marginal parameters using the results from the previous section. Therefore we specifically assume that observations are drawn from a latent DGP by

$$X_t = \begin{pmatrix} \mu_{1,1}\mathbb{1}_{t\le l_1} + \mu_{1,2}\mathbb{1}_{t>l_1} \\ \cdots \\ \mu_{m,1}\mathbb{1}_{t\le l_m} + \mu_{2,m}\mathbb{1}_{t>l_m} \end{pmatrix} + \begin{pmatrix} \sqrt{\sigma^2_{1,1}\mathbb{1}_{t\le l_1} + \sigma^2_{1,2}\mathbb{1}_{t>l_1}} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sqrt{\sigma^2_{m,1}\mathbb{1}_{t\le l_m} + \sigma^2_{m,2}\mathbb{1}_{t>l_m}} \end{pmatrix} Z_t$$

Inference on constant marginal distributions is directly based on the observed X_t. Prior to step 2, the data are transformed according to

$$\hat{Z}_{i,t} = \frac{X_{i,t} - \hat{\mu}_{i,1}\mathbb{1}_{t\le\hat{l}_1} - \hat{\mu}_{i,2}\mathbb{1}_{t>\hat{l}_1}}{\sqrt{\hat{\sigma}^2_{i,1}\mathbb{1}_{t\le\hat{l}_1} + \hat{\sigma}^2_{i,2}\mathbb{1}_{t>\hat{l}_1}}} \text{ if a break is detected or } \hat{Z}_t = \frac{X_{i,t} - \hat{\mu}_i}{\hat{\sigma}_i} \text{ else} \tag{A.7}$$

for $i = 1, ..., n$. Using $\hat{Z}_t$, partial sums are computed by stacking the elements of the matrix of standardized cross-moments:

$$\hat{S}_j = \frac{1}{j}\sum_{t=1}^{j}\left(\hat{Z}_1\hat{Z}_2, \quad \cdots, \quad \hat{Z}_1\hat{Z}_m, \quad \hat{Z}_2\hat{Z}_3, \quad \cdots, \quad \hat{Z}_2\hat{Z}_m, \quad \cdots, \quad \hat{Z}_{m-1}\hat{Z}_m\right)' \tag{A.8}$$

Based on the (unobserved) latent DGP Z_t, the test statistic $\sup \hat{Q}_j$ would have the limiting distribution $\sup ||B_{(m-1)m/2}(\pi)||$. It remains to find an estimator for the full-sample covariance matrix Ω. Wied [2017] suggests a block bootstrap estimator of the corresponding covariance matrix in the case of weakly stationary time series, the transformation prior to step 2 however allows to work under the assumption of strict stationarity, so we employ a simple bootstrap scheme to estimate Ω and usual critical values from the Brownian Bridge apply. This way, the effect of stochastic volatility can be absorbed.

Bootstrap approximations of the covariance matrix are however no longer valid if breaks are present in the margins. In this case, data are standard piecewise and asymptotic critical values do not apply. Demetrescu and Wied [2019] therefore suggest to apply a wild bootstrap scheme: Let $X_1^*, ..., X_n^*$ denote a sample from $X_1, ..., X_n$, drawn with replacement. The bootstrap sample is obtained from

$$X_{i,t}^{\circ} = \mu_{1,i} + \frac{(X_{i,t}^* - \mu_{1,i}^*)}{\sigma_{i,t}^{2*}} \sigma_{i,t} \tag{A.9}$$

such that in the bootstrap sub-samples at each margin the sample mean $\mu_{i,1}^b$ and sample variance $\sigma_{i,1}^{2,b}$ match sample mean $\mu_{i,1}$ and sample variance $\sigma_{i,1}^2$ from the original sample. Here $\mu_{1,i}^* = \frac{1}{\hat{\lambda}_i} \sum_{t=1}^{\hat{\lambda}_i} X_t^*$ and $\sigma_{1,i}^{2*} = \frac{1}{\hat{\lambda}_i} \sum_{t=1}^{\hat{\lambda}_i} (X_t^* - \mu_{i,1}^*)^2$. In every bootstrap repetition b the fluctuation test statistic is computed as in (3.10) and denoted by $\sup Q_j^b$ and the p-value approximated by

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} 1_{\{\sup \hat{Q}_j^b > \sup \hat{Q}_j\}}$$

A.3 Testing for Constant Copula

Throughout the previous section it has been implicitly assumed that Pearson's correlation coefficient suffices to describe the dependency in a multivariate system. By imposing this restriction on the dependence structure, many features frequently observed in financial applications are ignored in the testing procedure. Testing for a constant copula is a suitable way to test constant dependency of multivariate random variables beyond the moment hypothesis lined out previously. As before, applying the fluctuation test framework to copulae does not require parametric assumptions but only a test decision regarding the constancy of the marginal distributions.

Based on the (potentially) piecewise residuals, we follow Bücher et al. [2014] to transform the data onto the 'copula-scale' $[0, 1]^d$, by the empirical distribution function either over the full or the partial samples resulting from a split at j:

$$\hat{U}_i(x) = \frac{1}{n} \sum_{t=1}^{n} 1(X_{t,i} \leq x)$$
$$\hat{U}_i^{1:j}(x) = \frac{1}{j} \sum_{t=1}^{j} 1(X_{t,i} \leq x)$$
$$\hat{U}_i^{j+1:n}(x) = \frac{1}{n-j} \sum_{t=j+1}^{n} 1(X_{t,i} \leq x) \ \forall x \in \mathbb{R}$$

Define next the full-sample and partial-sample empirical copula by obtained from dividing the sample at a given j

$$\hat{C}_n(u) := \frac{1}{n}\sum_{t=1}^{n} 1(\hat{U}_t \leq u)$$
$$\hat{C}_{1:j}(u) := \frac{1}{j}\sum_{t=1}^{j} 1(\hat{U}_t^{1:j} \leq u)$$
$$\hat{C}_{j+1:n}(u) := \frac{1}{n-j}\sum_{t=j+1}^{n} 1(\hat{U}_t^{j+1:n} \leq u)$$

Then the test is based on the difference process of the partial-sample empirical copulae:

$$\mathbb{S}(j,u) = \frac{j(n-j)}{n^{3/2}}\left(\frac{1}{j}\sum_{t=1}^{j} 1_{\{\hat{U}_t \leq u\}} - \frac{1}{n-j}\sum_{t=j+1}^{n} 1_{\{\hat{U}_t \leq u\}}\right) \tag{A.10}$$

Constructing the difference process in this way improves the test statistics of Rémillard [2017] and Bücher and Ruppert [2013] who use the empirical copula over the full sample rather than partial-sample empirical copulae, as pointed out by Bücher et al. [2014] in a simulation study. Thus only the more recent method is used in the subsequent Monte Carlo studies.
The test statistic follows from integrating with respect over $[0,1]^d$, where in practice a discretization grid has to be chosen and subsequently take the sup-functional over the set of change point candidates $j \in \{2, ..., n-2\}$:

$$T = \sup_j \left(\int_{u[0,1]^d} \mathbb{S}(j,u)^2 d\hat{C_n}(u) \right) \tag{A.11}$$

In order to approximate critical or p-values, Bücher et al. [2014] suggest a block multiplier bootstrap scheme that works under strong mixing conditions and is also used in Bücher et al. [2014][1]. Under the assumption of a proper transformation prior to testing for a constant copula, a simplified i.i.d.-multiplier bootstrap can be used in the following way: For each bootstrap repetition b one draws $\zeta_{b,t} \overset{i.i.d.}{\sim} N(0,1)$ and computes

$$\begin{aligned} \hat{T} &= \sup_j \left(\int_{u[0,1]^d} \mathbb{S}_b(j,u)^2 d\hat{C_n}(u) \right) \\ \mathbb{S}_b(j,u) &= \mathbb{B}_b(j,u) - j\mathbb{B}_b(1,u) \\ \mathbb{B}_b(j,u) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{j} \zeta_{b,t}\left(1_{\{\hat{U}_t \leq u\}} - \hat{C}_n(u)\right) \end{aligned} \tag{A.12}$$

from where the approximated p-value follows as

$$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} 1_{\{\hat{T}_b > T\}}$$

[1] The authors also suggest a more advanced bootstrap scheme based on the partial-sample copulae, which however is computationally intractable for the Monte Carlo studies in section 4.

B Sup-LR Tests for Piecewise i.i.d. Margins

Complementary to our simulation study in section 3.1, we confirm our central observations for piecewise i.i.d. data. In contrast to before, we allow for $\mu_i > 0$. Such an assumption would be useful in the case of bond spreads. Decomposing the entire covariance matrix into

$$\Sigma = S'PS = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{pmatrix} \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \ddots & 1 & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{pmatrix}$$

enables us to easily separating inference on marginal parameters and correlation matrix by writing

$$\begin{aligned}
&X_t \overset{i.i.d.}{\sim} N(\theta_{1,1}, \theta_{1,1}, \cdots, \theta_{m,1}, P_1) \quad \text{for } t = 1, ..., l_1 \\
&X_t \overset{i.i.d.}{\sim} N(\theta_{1,2}, \theta_{2,1}, \cdots, \theta_{m,1}, P_1) \quad \text{for } t = l_1 + 1, ..., l_2 \\
&\cdots\cdots \\
&X_t \overset{i.i.d.}{\sim} N(\theta_{1,2}, \theta_{2,2}, \cdots, \theta_{m,2}, P_1) \quad \text{for } t = l_m + 1, ..., l_D \\
&X_t \overset{i.i.d.}{\sim} N(\theta_{1,2}, \theta_{2,2}, \cdots, \theta_{m,2}, P_2) \quad \text{for } t = l_D, ..., n
\end{aligned}$$

and testing Hypothesis Pair 3 with $\theta_i = (\sigma_i^2, \mu_i)$ and $\delta_i = P_i$. Following the steps outlined in section 3.1, the difference of the log-likelihood under full-sample and partial-sample estimators yields

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

Note that under i.i.d. sampling there is no need for robust variance-covariance estimators and the limiting process under the null hypothesis is given by eq. (3) with $k = 2$ degrees of freedom. Data are standardized according to

$$\hat{Z}_{i,t} = \frac{X_{i,t} - \hat{\mu}_{i,1} \mathbb{1}_{t \le \hat{l}_1} - \hat{\mu}_{i,2} \mathbb{1}_{t > \hat{l}_1}}{\sqrt{\hat{\sigma}^2_{i,1} \mathbb{1}_{t \le \hat{l}_1} + \hat{\sigma}^2_{i,2} \mathbb{1}_{t > \hat{l}_1}}} \text{ if a break is detected or } \hat{Z}_t = \frac{X_{i,t} - \hat{\mu}_i}{\hat{\sigma}_i} \text{ else} \quad \text{(B.1)}$$

The test for constant correlation is set up in the same way as under volatility clustering.

C Simulation Results for Piecewise i.i.d. Margins

In this section we provide simulation results under the assumptions of piecewise i.i.d. Gaussian marginal distributions. Otherwise, the setup is analogous to section 3.1 in the main text. Using vector notation and the covariance decomposition of the multivariate Gaussian distribution $\Sigma = S'PS$, data is generated according to

$$X_t \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0.05 \\ 0.05 \\ 0.05 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix} \right) \quad \text{for } t = 1, ..., l_1$$

$$X_t \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0.06 - 0.01s \\ 0.05 \\ 0.05 \end{pmatrix}, \begin{pmatrix} s & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix} \right) \quad \text{for } t = l_1, ..., l_2$$

$$X_t \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0.06 - 0.01s \\ 0.06 - 0.01s \\ 0.06 - 0.01s \end{pmatrix}, \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix}, \begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix} \right) \quad \text{for } t = l_2, ..., l_D$$

$$X_t \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0.06 - 0.01s \\ 0.06 - 0.01s \\ 0.06 - 0.01s \end{pmatrix}, \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix}, \begin{pmatrix} 1 & \rho_2 & \rho_2 \\ \rho_2 & 1 & \rho_2 \\ \rho_2 & \rho_2 & 1 \end{pmatrix} \right) \quad \text{for } t = l_D, ..., n$$

In this way we ensure that mean μ and variance σ^2 change simultaneously. Again, empirical rejection rates under the (true) null hypothesis of constant correlation coefficients are presented in table C.1. Using the incorrect asymptotic values would again lead to severe size distortions. These are however corrected with our bootstrap scheme.

Table C.1: Gaussian Distribution, 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	0.027	0.028	0.044	0.038	0.038	0.038	0.039	0.036	0.046	0.051	0.044	0.048
1.5	0.108	0.110	0.124	0.122	0.696	0.662	0.665	0.609	0.996	0.997	0.994	0.991
2	0.297	0.342	0.366	0.363	0.995	0.994	0.993	0.990	1	1	1	1
2.5	0.540	0.594	0.652	0.640	1	1	1	1	1	1	1	1
3	0.706	0.747	0.833	0.808	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	0.026	0.044	0.052	0.035	0.028	0.038	0.044	0.043	0.036	0.051	0.054	0.043
1.5	0.058	0.085	0.083	0.046	0.015	0.035	0.100	0.098	0.019	0.048	0.060	0.049
2	0.040	0.063	0.114	0.053	0.011	0.050	0.069	0.058	0.019	0.051	0.057	0.039
2.5	0.033	0.051	0.111	0.054	0.013	0.051	0.059	0.048	0.019	0.048	0.054	0.039
3	0.030	0.043	0.099	0.057	0.014	0.051	0.056	0.046	0.020	0.048	0.054	0.042
	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	0.031	0.050	0.120	0.057	0.044	0.048	0.070	0.057	0.045	0.052	0.062	0.049
1.5	0.043	0.071	0.152	0.062	0.021	0.048	0.138	0.111	0.021	0.047	0.073	0.061
2	0.039	0.064	0.179	0.083	0.019	0.047	0.078	0.063	0.023	0.047	0.069	0.055
2.5	0.031	0.045	0.178	0.079	0.019	0.051	0.071	0.058	0.024	0.049	0.065	0.052
3	0.023	0.042	0.164	0.077	0.021	0.057	0.071	0.049	0.024	0.050	0.065	0.047
	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	0.016	0.039	0.130	0.034	0.049	0.061	0.086	0.051	0.059	0.065	0.086	0.067
1.5	0.022	0.042	0.166	0.038	0.032	0.055	0.174	0.126	0.031	0.070	0.099	0.075
2	0.029	0.044	0.218	0.055	0.021	0.058	0.108	0.077	0.033	0.066	0.097	0.066
2.5	0.030	0.039	0.222	0.056	0.025	0.061	0.093	0.063	0.031	0.067	0.094	0.067
3	0.020	0.023	0.194	0.046	0.027	0.065	0.093	0.065	0.032	0.067	0.095	0.065

We proceed to empirical power which is computed the same values of ρ, varying from -0.1 to

0.9 in steps of 0.1. Similar to before, the residual effect works into the same direction as the power is increased by the bootstrap scheme in the fluctuation test framework and corrected downwards in the sup-LR test framework.

Figure C.1: Gaussian Distribution, n=100, Scenario 2: Empirical Power

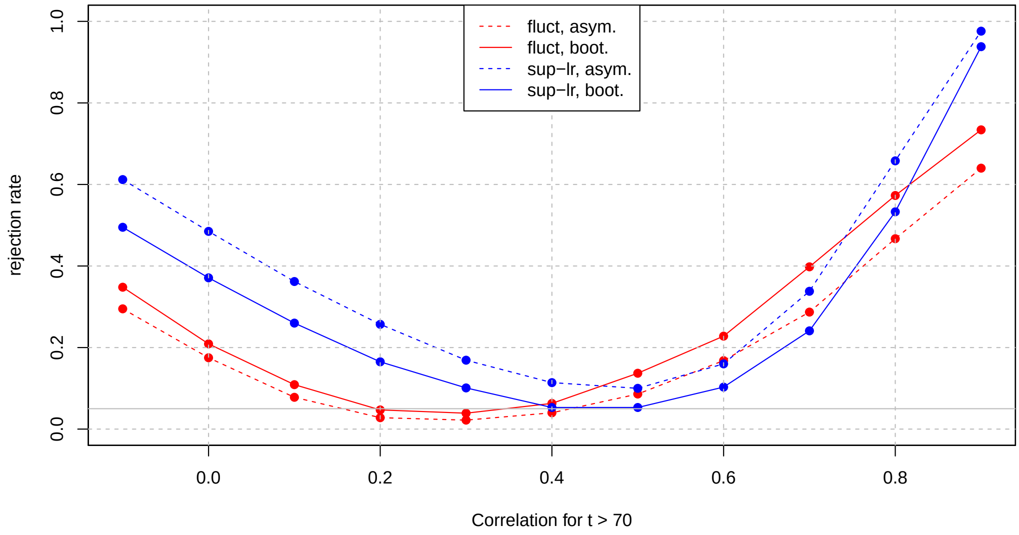


Figure C.2: Gaussian Distribution, n=500, Scenario 2: Empirical Power

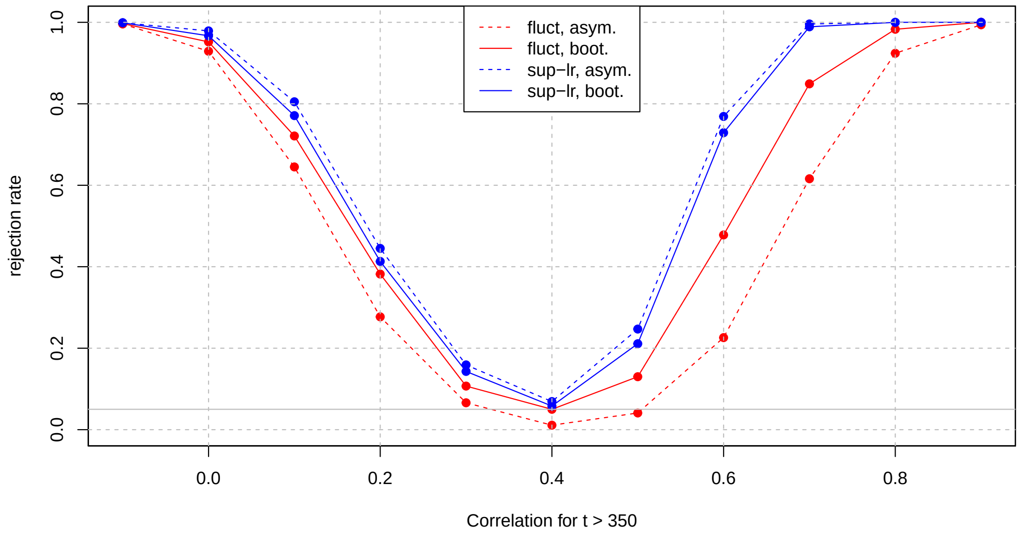
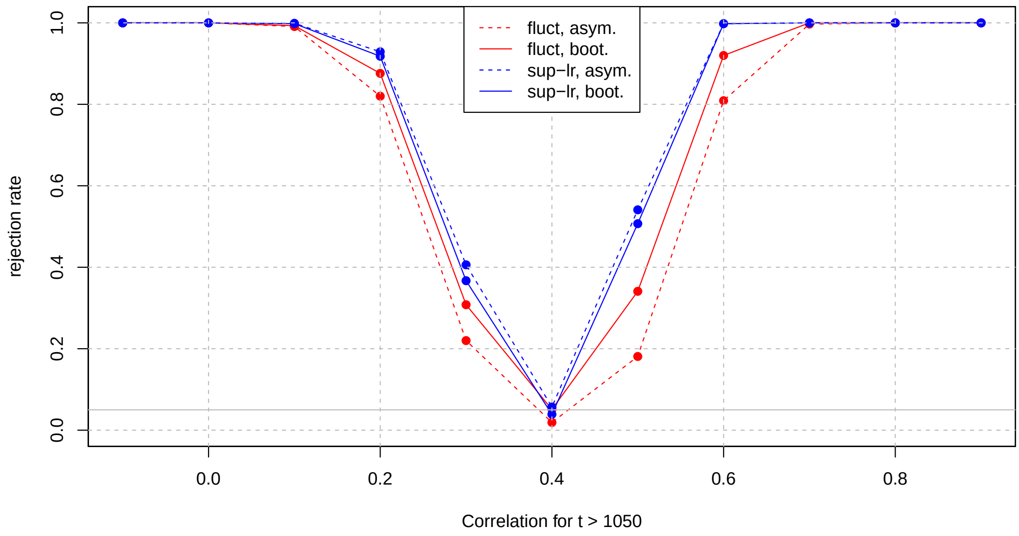


Figure C.3: Gaussian Distribution, n=1500, Scenario 2: Empirical Power



Again these results point into the same direction as section 3.1. Also, empirical power exhibits the same dimensionality effects in both testing frameworks we already saw under volatility clustering (figure C.4 - C.6).

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

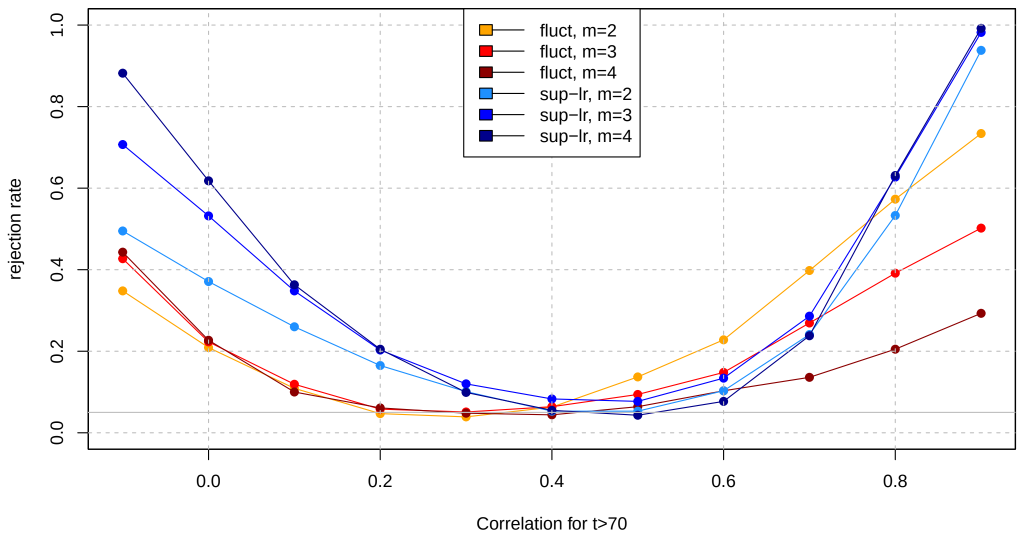


Figure C.5: Multivariate Gaussian, n=500, Scenario 2: Empirical Power

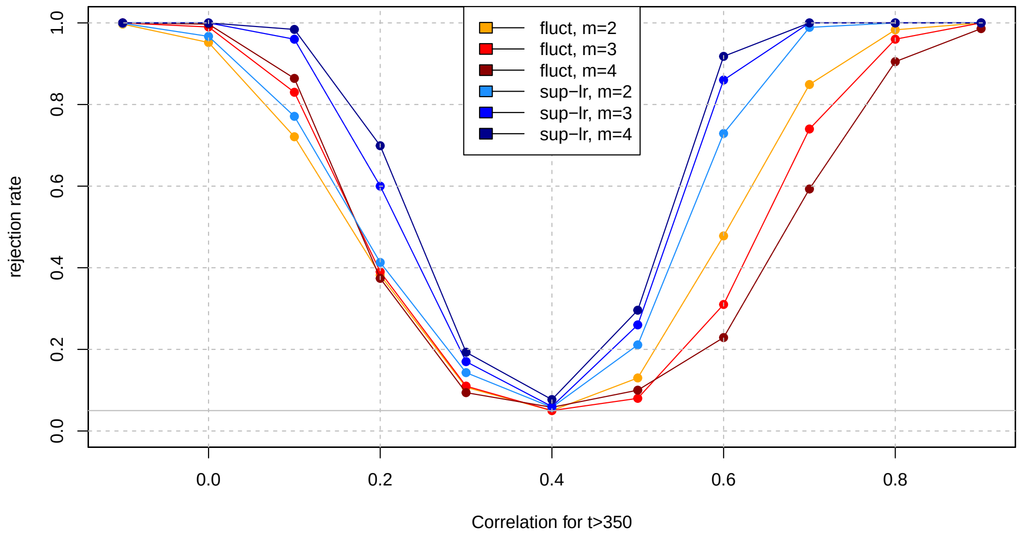
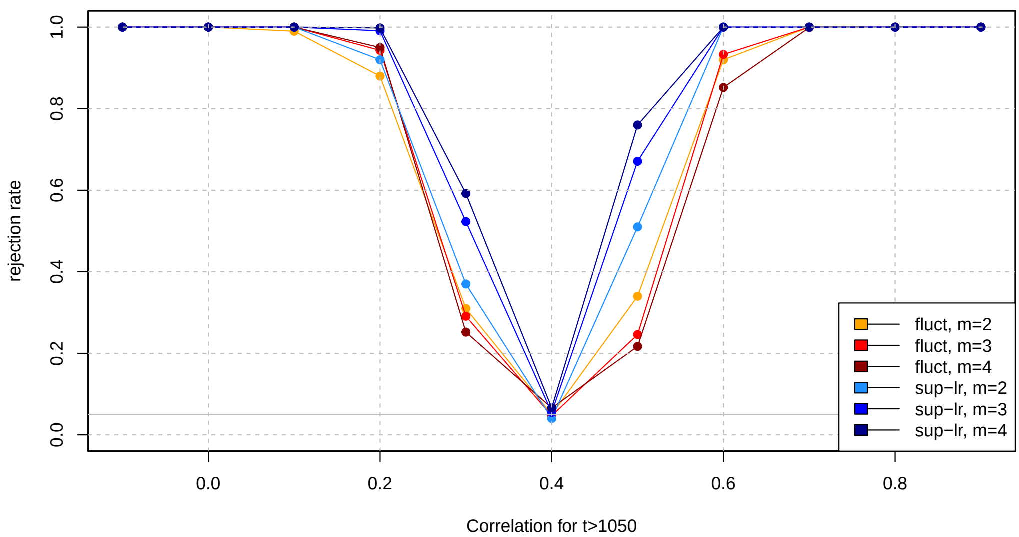


Figure C.6: Multivariate Gaussian, n=1500, Scenario 2: Empirical Power



As in the GARCH-case, except for small samples and small ρ_2, bias and RMSE, shown in table C.2, are considerably smaller for the sup-LR test. Note that for small samples all tests perform better in a piecewise i.i.d. setting than under volatility clustering, presumably since parameter estimators and testing power at the margins are less precise here.

Table C.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	-2.99	13.15	-4.96	16.87	-14.29	31.03	-0.49	24.78	-18.39	41.73	1.76	20.40
0.1	-5.86	19.13	-10.23	23.07	-26.68	57.01	-12.77	62.56	-38.63	83.59	0.23	62.21
0.3	-9.77	22.73	-15.25	27.36	-85.00	137.44	-76.94	138.01	-164.40	305.72	-121.79	308.35
0.5	-9.48	20.22	-16.10	29.00	-77.34	131.78	-59.34	119.72	-181.78	327.76	-97.57	259.20
0.7	-7.21	15.12	-8.63	22.72	-27.59	61.15	-4.31	27.92	-42.10	101.06	-2.43	20.76
0.9	-4.92	11.04	0.53	5.21	-12.70	32.69	0.11	4.56	-14.87	42.38	-0.08	3.25
	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	-2.55	9.78	-3.57	12.62	-10.31	19.48	0.92	9.65	-13.47	25.87	1.54	7.93
0.1	-5.27	16.69	-9.94	21.59	-22.96	45.92	-3.34	37.48	-32.87	65.28	3.13	24.49
0.3	-8.58	19.48	-16.41	27.98	-82.95	125.35	-71.20	132.78	-167.12	287.54	-87.16	256.22
0.5	-9.42	17.62	-17.55	29.68	-96.40	134.51	-57.88	118.46	-205.71	320.80	-68.59	212.62
0.7	-8.28	14.64	-7.89	22.36	-37.18	70.05	-0.41	11.78	-33.52	84.18	-1.36	9.92
0.9	-6.60	11.68	1.27	3.03	-12.03	31.38	0.71	1.87	-11.35	37.17	0.55	1.91
	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	-1.98	7.38	-1.73	9.05	-9.12	17.52	0.69	5.00	-11.90	21.67	0.72	1.35
0.1	-3.62	12.91	-8.35	19.82	-22.04	43.03	-3.77	27.17	-28.06	53.14	-0.03	6.91
0.3	-7.70	16.60	-14.84	27.33	-79.97	115.49	-67.37	126.15	-171.65	277.71	-66.74	175.89
0.5	-9.54	16.18	-17.67	30.64	-102.77	133.12	-46.35	107.56	-230.12	326.45	-79.45	235.21
0.7	-9.87	15.23	-7.51	22.41	-45.46	76.43	0.19	7.07	-44.56	104.33	1.57	16.64
0.9	-9.45	14.07	1.40	2.86	-17.84	37.67	0.87	1.31	-10.53	35.65	0.77	4.58

D Sup-LR Test for Gaussian Copula

As alternative to using t-Copula as in section 3.2 in the main paper, step 2 can also be based on the copula associated with the Gaussian distribution assumption. Step 1 remains unchanged, however the data are now (piecewise) transformed onto the copula scale by

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

The pseudo-observations are then used to estimate the dependency parameter (i.e. the correlation matrix) of the Gaussian copula under the null and alternative hypothesis. Consider next the density of the Gaussian copula

$$f(\hat{U}_t; P) = |R|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \hat{U}_t'(R^{-1} - I)\hat{U}_t \right)$$

from where the full-sample log-likelihood

$$L(\hat{U}; P_0) = -\frac{n}{2}|R| - \frac{1}{2} \sum_{t=1}^{n} \hat{U}_t'(R_0^{-1} - I)\hat{U}_t$$

and the partial-sample log-likelihood

$$L(\hat{U}; P_1, P_2) = -\frac{j}{2}|R_1| - \frac{n-j}{2}|R_2| - \frac{1}{2} \sum_{t=1}^{j} \hat{U}_t'(R_1^{-1} - I)\hat{U}_t - \frac{1}{2} \sum_{t=j+1}^{n} \hat{U}_t'(R_2^{-1} - I)\hat{U}_t$$

are obtained. Let $\hat{R}_0$, $\hat{R}_1$ and $\hat{R}_2$ denote the ML-estimators for the correlation matrix of the full sample and each sub-sample. Evaluating the log-likelihood at the respective parameter estimates gives the test statistic for a fixed j as

$$A_j = 2\big(L(\hat{U}; \hat{R}_1, \hat{R}_2) - L(\hat{U}; \hat{R}_0)\big).$$

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

E Application to Commodity and Equity Index Data

A third empirical application uses the methods subject to the simulation studies section 3.1: We test 3 for Crude Oil spot market returns [2] and the European equity sector, which we proxy by the EUROSTOXX50[3] over the time-period 1991-04-17 to 2003-03-05. Foreign involvement in petrol-exporting countries has been fairly low following the early 1990s until 2003. Additionally events in the late 1980s and later technological changes in oil production, the financial crises and relaxed monetary policy probably did not influence the fundamental market environment over the sample period. However markets experienced a period of increased volatility around 2000, associated with events such as the burst of the dotcom-bubble among others. This can be observed in table F.2 and F.2. Rolling correlations in figure F.1 however indicate a rather stable correlation pattern over the test period and thus making the sample a plausible candidate to test for Hypothesis Pair 3. Correlations and (annualized) volatilities are expressed in percentage points.

Figure E.1: Estimation of European Crude Oil and Equity Data

	Fluctuation Test		sup-LR test	
	Crude Oil	Equity	Crude Oil	Equity
$\hat{l}_i$	1998-01-19	1997-07-16	1996-03-15	1997-06-27
test statistic (volatility)	88.76	7.49	378.04	903.76
test statistic (correlation)	1.094		8.92	
p-value (boot)	0.18		0.23	
ρ_0	1.25		0.008	
ρ_1	-5.81		-5.81	
ρ_2	5.44		5.46	

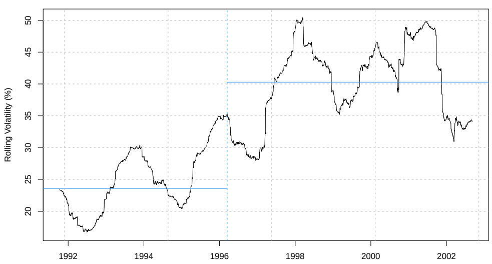
All procedures strongly reject the hypothesis of constant margins, the critical values at 95 % for the fluctuation test being 1.358 and for the sup-LR test 8.68. When it comes to testing constant correlation, our empirical findings from section 3.1 directly carry over to this particular example. Following table 1, it is crucial to apply a suitable bootstrap here. Using the bootstrapped p-values around or larger than 0.2, neither fluctuation test and not the sup-LR tests reject Hypothesis Pair 3. Had one used the incorrect asymptotic value for the sup-LR test, which is 8.68 at 95 % confidence level, one might incorrectly reject at

[2] Europe Brent, Data is taken from the U.S. Energy Information Administration: https://www.eia.gov/dnav/pet/hist/

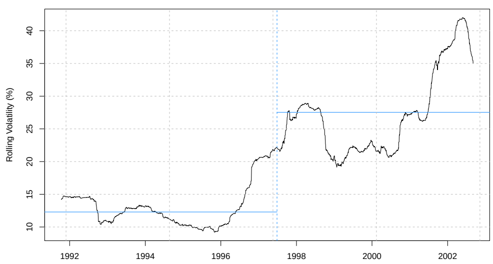
[3] ISIN: EU0009658145, returns are calculated from the closing price of the last trading day each month.

the second step. As the fluctuation test exhibits rejection rates below the nominal level in the simulation, it is not surprising that the test does not reject, even when the incorrect asymptotic value is used.
It has been previously established that incorrectly assuming constant variances when testing for constant correlation - implicitly by considering covariances as Aue et al. [2009] or explicitly by directly using the procedure of Wied et al. [2012b] - leads to flawed inference. But, as the preceding application points out, even if changes at the marginal distributions are taken into account correctly, applying invalid standard asymptotics may lead to incorrectly rejecting constant cross-sectional dependence.

(a) Rolling Correlations, Equity and Crude Oil



(b) Crude Oil, Rolling Volatility



(c) Equity, Rolling Volatility

F Application to European Bond Data

This section takes the sup-LR method to a Government Bond spreads around the Euro area debt crisis. We use par yields from France, Germany, Italy and Spain with a maturity of 5 years. Spreads are computed against the 5y-EURIBOR-Swap rate and reported in basis points. Our sample starts on 2010-01-01, when negative news on the fiscal capacity of several Euro area members started to accrue and goes until 2012-09-06, when the ECB announced the Outright Monetary Transactions programme that eventually led to a fall in

credit spreads. Since our framework allows for at most one structural change at each margin and the dependency structure, extending the sample to cover a period of three regimes is not possible with the method subject to our simulation study in section 3.
Since our methods are particularly suited for dating structural changes, even under misspecification, we maintain the assumption of a multivariate Gaussian distribution. Results for the marginal distributions are reported in table F.1.

Table F.1: Estimation of European Bond Data: Margins

France	Fluctuation	sup-LR, Gaussian
$\hat{l}_i$	2011-10-14	2011-10-13
$\sup A_j$	10.83	1341.93
$\hat{\mu}_1$	-15.28	-15.32
$\hat{\mu}_2$	26.52	26.41
$\hat{\sigma}_1$	5.94	5.91
$\hat{\sigma}_2$	21.59	21.58

Germany	Fluctuation	sup-LR, Gaussian
$\hat{l}_i$	2011-07-12	2011-07-11
$\sup A_j$	11.34	958.52
$\hat{\mu}_1$	-39.76	-39.69
$\hat{\mu}_2$	-74.98	-74.93
$\hat{\sigma}_1$	9.91	9.84
$\hat{\sigma}_2$	10.69	10.69

Italy	Fluctuation	sup-LR, Gaussian
$\hat{l}_i$	2011-07-07	2011-07-08
$\sup A_j$	12.23	1456.68
$\hat{\mu}_1$	84.42	84.62
$\hat{\mu}_2$	344.51	345.11
$\hat{\sigma}_1$	33.44	33.63
$\hat{\sigma}_2$	38.58	84.08

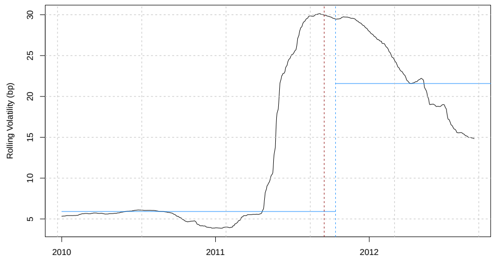
Spain	Fluctuation	sup-LR, Gaussian
$\hat{l}_i$	2011-06-13	2011-07-08
$\sup A_j$	10.45	748.84
$\hat{\mu}_1$	125.36	129.15
$\hat{\mu}_2$	317.63	324.71
$\hat{\sigma}_1$	54.62	55.99
$\hat{\sigma}_2$	110.11	109.59

Both procedures strongly reject the hypothesis of constant margins, the critical values at 99 % for the sup-LR tests being 15.51. For France, Italy and Spain, both the level and volatility of bond spreads increased substantially, while for Germany the level of spreads decreased from around -40bp to -73bp, while the volatility remained roughly constant. Thus our methods support the occurrence of a strong flight-to-quality event, that also let French spreads rise substantially, although the fiscal capacities of France were never a primary concern during the European debt crisis.
In figure F.1 we display spreads and their rolling volatilities over a 252-day window with

the respective partial-sample means and standard deviations obtained from the sup-LR test. The dashed blue line indicates the change-point for each margin, the dashed red line the change point of the correlation matrix.

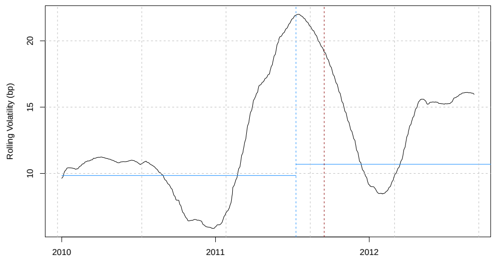
Figure F.1: Bond-spreads (5y) in basis points

(a) France, Bond-Spreads



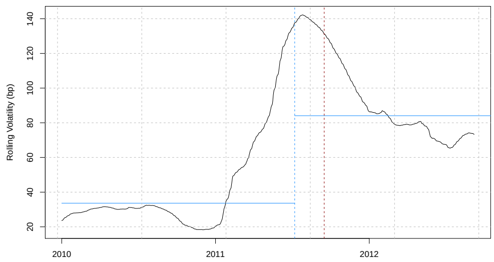
(b) France, Rolling Volatilities

(c) Germany, Bond-Spreads



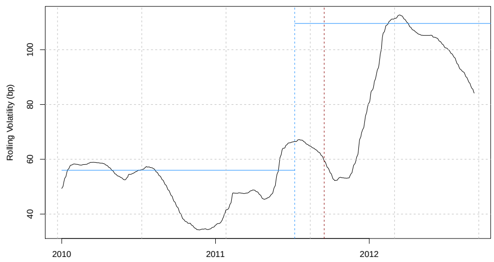
(d) Germany, Rolling Volatilities

(e) Italy, Bond-Spreads



(f) Italy, Rolling Volatilities

(g) Spain, Bond-Spreads



(h) Spain, Rolling Volatilities

In table F.2, we present results of testing and estimating step 2 with the fluctuation and sup-LR framework from section 3.1, respectively. All partial-sample correlations are denoted in percentage points.

Table F.2: Estimation of European Bond Data: Dependency Structure

	Fluct.	sup-LR, Gauss
$\hat{l}_D$	2011-06-13	2011-09-16
$\sup A_j$	7.49	784.76
p-val.	0	0
$\hat{\rho}_1(s_{FR}, s_{GER})$	61.15	27.91
$\hat{\rho}_2(s_{FR}, s_{GER})$	-30.96	-63.69
$\hat{\rho}_1(s_{FR}, s_{IT})$	-48.72	7.38
$\hat{\rho}_2(s_{FR}, s_{IT})$	29.29	-50.14
$\hat{\rho}_1(s_{FR}, s_{ES})$	-42.29	-8.12
$\hat{\rho}_2(s_{FR}, s_{ES})$	-17.16	-68.83
$\hat{\rho}_1(s_{GER}, s_{IT})$	-85.42	-72.10
$\hat{\rho}_2(s_{GER}, s_{IT})$	-32.91	9.80
$\hat{\rho}_1(s_{GER}, s_{ES})$	-82.98	-60.83
$\hat{\rho}_2(s_{GER}, s_{ES})$	30.89	32.18
$\hat{\rho}_1(s_{IT}, s_{ES})$	96.49	83.05
$\hat{\rho}_2(s_{IT}, s_{ES})$	49.16	93.57

As opposed to the marginal distributions, estimated correlations differ considerably, which is largely due the difference in $\hat{l}_D$. Figure F.2 shows rolling correlations of French and German bond spreads, once without correction at the margins (solid black), once using the fluctuation test (dashed red) and once using the sup-LR test (dashed blue). Due to the different change point estimate, the respective sub-samples vary considerably in magnitude. Since our simulation study reveals a higher precision for the sup-LR test, we will use the resulting partial sample estimates in the remainder.

Figure F.2: Bond-spreads (5y) in basis points

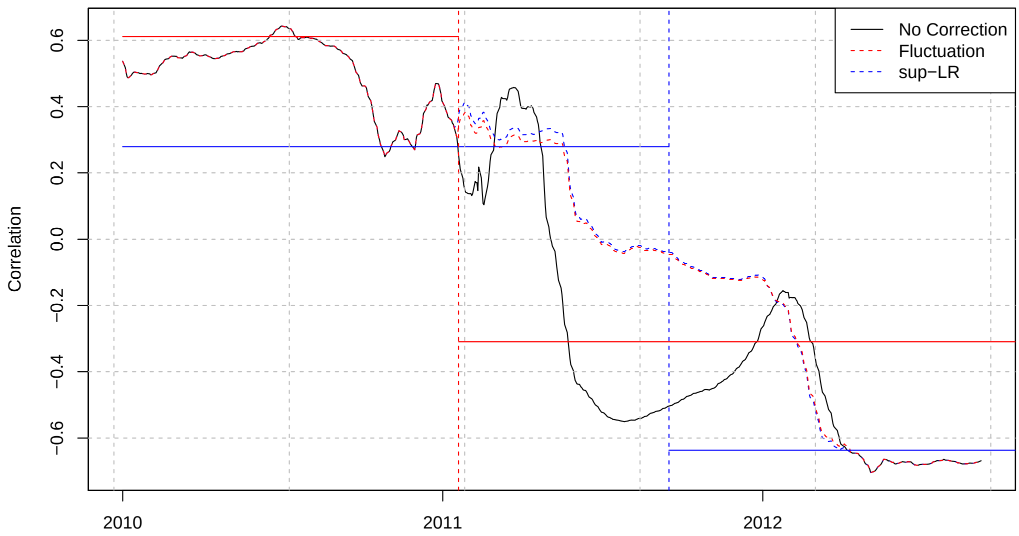


Figure F.3: France and Germany, Rolling Correlations

The partial sample estimates reveal some interesting dynamics, especially with regard to French spreads. As can be seen in figure F.1, Italian and Spanish spreads co-moved even stronger in the second sub-sample, a typical indicator of contagion. Both were co-moving negatively against German spreads in the run-up to the crisis, but only experience a weak correlation with German spreads once they settled at their respective new levels. This finding confirms, that German bonds were regarded as a safe haven throughout the crisis.
The time path and co-movement of French spreads does not yield a similarly clear-cut result. During the first sub-sample, French spreads experienced a weakly positive correlation with German spreads, but were not regarded as a safe haven once the fiscal situation within the Euro area became more critical. After a sharp increase in early 2012, spreads however were moving back to the pre-crisis level in the second half of 2012, as indicated by the strongly negative correlation with German *and* Italian/Spanish spreads. This permits the conclusion, that after initial uncertainty on the soundness of public finances, French spreads returned to pre-crisis level before large-scale interventions by the ECB. The preceding application strengthens, that incorrectly assuming constant variances when testing for constant correlation - implicitly by considering covariances as Aue et al. [2009] or explicitly by directly using the procedure of Wied et al. [2012b] - can lead to imprecise conclusions on partial sample dependency structures.

References

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.

A. Bücher and J. Ruppert. Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis*, 116: 208–229, 2013.

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.

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.

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

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

D. Wied. A nonparametric test for a constant correlation matrix. *Econometric Reviews*, 36 (10):1157–1172, 2017.

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.