RUHR
ECONOMIC PAPERS

Fang Duan

Forecasting Risk Measures Based on Structural Breaks in the Correlation Matrix

RGS econon

technische universität dortmund

#945

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #945

Responsible Editor: Ludger Linnemann

ISSN 1864-4872 (online) – ISBN 978-3-96973-106-2

Ruhr Economic Papers #945

Fang Duan

Forecasting Risk Measures Based on Structural Breaks in the Correlation Matrix

technische universität dortmund

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973106
ISSN 1864-4872 (online)
ISBN 978-3-96973-106-2

Fang Duan[1]

Forecasting Risk Measures Based on Structural Breaks in the Correlation Matrix

Abstract

Correlation models, such as Constant Conditional Correlation (CCC) GARCH model or Dynamic Conditional Correlation (DCC) GARCH model, play a crucial role in forecasting Value-at-Risk (VaR) or Expected Shortfall (ES). The additional inclusion of constant correlation tests into correlation models has been proven to be helpful in terms of the improvement of the accuracy of VaR or ES forecasts. Galeano & Wied (2017) suggested an algorithms for detecting structural breaks in the correlation matrix whereas Duan & Wied (2018) proposed a residual based testing procedure for constant correlation matrix which allows for time-varying marginal variances. In this chapter, we demonstrate the application of aforementioned correlation testing procedures and compare its performance in backtesting VaR and ES predictions. Portfolios in consideration are constructed from four stock indices DAX30, STOXX50, FTSE100 and S&P500.

JEL-Codes: C12, C32, C53, C58

Keywords: Structural break tests; correlation model; value-at-risk; expected shortfall

March 2022

1 TU Dortmund and RGS Econ. - Fang Duan gratefully acknowledges funding by Ruhr Graduate School in Economics (RGS Econ) – All correspondence to: Fang Duan, Technical University Dortmund, Otto-Hahn-Str.6, 44227, Dortmund, Germany, e-mail: duan@statistik.tu-dortmund.de

1 Introduction

Multivariate GARCH model is appealing in both theoretical and practical usages, due to its parsimonious specification and its feasibility of the estimations in high dimensional portfolio of assets. The correlation-based models, i.e., Constant Conditional Correlation (CCC) model and Dynamic Conditional Correlation (DCC) model are considered in this chapter. Berens et al. (2015) has shown that the usage of structural break testing procedure for correlations is able to significantly improve the forecasting precision of VaR and ES based on CCC model. Adams et al. (2017) argued that a spurious daily correlation dynamics among assets exists and the levels of correlations shift due to correlation breaks detected by the algorithm proposed by Galeano & Wied (2014). This series of structural break tests for constant correlation matrix is able to detect the level shifts of correlations being associated with crucial financial events.

In this chapter, we examine the impact of the integration of constant correlation test by Duan & Wied (2018) into CCC and DCC models. One takes the conditional correlation models without structural break tests of the correlation matrix as a benchmark model. Then, the new testing procedure is compared with the test of Wied (2017) via a series of evaluations of backtesting VaR and ES forecasts. In addition to the cases of single break points in the correlation matrix, the algorithm by Galeano & Wied (2017) allows for the consideration of multiple break points in the correlation matrix. These tests are performed in four equally weighted portfolios from four stock indices STOXX50, DAX30, FTSE100 and S&P500, respectively.

The rest of this chapter is organized as follows. We briefly introduce basic definitions and models in Section 2. In Section 3, we review the structural breaks tests for correlation matrix, after which we discuss popular backtesting procedures in Section 4. In Section 5 we present an empirical application, followed by our conclusions in Section 6.

2 Definitions and Models

In order to fulfill the regulatory requirements from Basel III Accords, see Basel Committee (2010), banks and financial institutes ought to assess the potential losses of their portfolios. The common appealing metrics of market risk measures are Value-at-Risk(VaR) and Expected Shortfall (ES). The latter tends to replace the former as the main risk measure as in Basel accords, see Basel Committee (2016) and Basel Committee (2017).

For starters, one defines the random variable $r_{i,t} = \log p_{i,t} - \log p_{i,t-1}, i = 1, \dots, k, t = 1, \dots, T$ as the daily log return of i-th asset in a portfolio at time point t where $p_{i,t}$ denotes the asset price for the i-th asset at time t. The portfolio return can be written as $r_{p,t} = \sum_{i=1}^{k} \omega_i r_{i,t}$ at time t, where $\omega_i, i = 1 \dots, k$ represents the portfolio weights associated with i-th asset. The extension of original definition in McNeil & Frey (2000) gives the dynamics of the portfolio return $r_{p,t}$ as $r_{p,t} = \mu_{p,t} + \sigma_{p,t} Z_t$ where Z_t follows i.i.d. distribution $F_Z(z)$ with mean zero and unit variance. $\mu_{p,t}$ and $\sigma_{p,t}$ are measurable with respect to $\mathcal{F}_{t-1}$ which contains all information about the return series available up to time $t-1$. The cumulative distribution function (CDF) of $r_{p,t}$ given the information set $\mathcal{F}_{t-1}$ is $F_t(r_p) = P(r_{p,t} \leq r_p | \mathcal{F}_{t-1})$. Accordingly, the quantile function of $r_{p,t}$ given the information set $\mathcal{F}_{t-1}$ at level $\alpha \in (0, 1)$, i.e., the VaR at level α is $Q_t(r_{p,t}|\mathcal{F}_{t-1}) = F_t^{-1}(\alpha) = \inf\{r_p \in \mathbb{R} : F_t(r_p) \geq \alpha\}$, where the function $F^{-1}(\cdot)$ is the inverse CDF function. In Christoffersen (2003), the α-VaR of a portfolio is defined as the largest amount such that the probability that the loss of portfolio return over a specific time horizon is greater than VaR is α. Or equivalently, the VaR of portfolio return at level α is defined as the lower α-quantile of the distribution of the portfolio return at time t: $\text{VaR}_{p,t}^{\alpha} = F_{r_{p,t}}^{-1}(\alpha)$. Indeed, in some context $\text{VaR}_{p,t}^{\alpha}$ can be positive when one defines it with respect to the loss variable $l_{p,t} = -r_{p,t}$. In order to determine the VaR forecast on t based on $\mathcal{F}_{t-1}$, the predicted distribution of return is necessary in the first place. Franke et al. (2015) allows that the CDF function of $r_{p,t}$ incorporates time-varying parameter vector $\boldsymbol{\theta}_t \in \mathbb{R}^p$ and static parameter vector $\boldsymbol{\phi} \in \mathbb{R}^q$. One defines a forecast distribution at time point t as $P_t^{\theta_t}(r_{p,t}|\mathcal{F}_{t-1})$. The possible conditional distributions of $r_{p,t}$ belong to the parameter class $\mathcal{P}_t = \{P_t^{\theta_t} | \theta(t) \in \Theta\}$.

Since VaR does not guarantee the property of subadditivity, Artzner et al. (1999) proposed a coherent risk measure, namely Expected Shortfall (ES) or tail-VaR, which is the expected value of portfolio loss given the VaR exceedance has occurred. If the function $F_t(\cdot)$ defined above is continuous at α quantile, then the ES can be written as

$$\mathrm{ES}_\alpha(r_{p,t}|\mathcal{F}_{t-1}) = E(r_{p,t}|r_{p,t} \leq Q_t(r_{p,t}|\mathcal{F}_{t-1})). \tag{1}$$

To evaluate the correctness of the risk measures forecasting, namely backtesting, Bayer & Dimitriadis (2020) introduced a general and strict definition of the terminology to compare the forecasts for the risk measure ρ and the realized return series: a backtest for the forecasts $\{\hat{\rho}_t, t = 1, \ldots, T\}$ for the d-dimensional risk measure ρ relative to the realized return series $\{r_{p,t}, t = 1, \ldots, T\}$ is a function $f : \mathbb{R}^T \times \mathbb{R}^{T\times d} \to \{0, 1\}$.

Next, we focus on the necessary steps of delivering the final risk measure forecasts. In order to capture the marginal dynamics in random variables, one often resorts to GARCH-type models. For example, a simple univariate GARCH model enables us to predict the marginal volatilities. Let $r_{i,t}$ denotes log-return of i-th asset at time t. The univariate GARCH(1,1) model follows

$$\begin{aligned} r_{i,t} &= \mu_{i,t} + \epsilon_{i,t} \\ \epsilon_{i,t} &= \sigma_{i,t}\eta_{i,t} \\ \sigma^2_{i,t} &= \alpha_{1,i} + \alpha_{2,i}\epsilon^2_{i,t-1} + \beta_{1,i}\sigma^2_{i,t-1}, \end{aligned} \tag{2}$$

where $\eta_{i,t}$ stands for the innovation term, which is commonly assumed to be Student's t or Hansen's skewed-t distribution. One could estimate the vector of dynamic parameters $(\alpha_{1,i}, \alpha_{2,i}, \beta_{1,i})'$ via Maximum Likelihood Estimation (MLE), then one-step-ahead volatility $\hat{\sigma}_{i,t+1}$ is obtained with the plug-in estimated dynamic parameters:

$$\hat{\sigma}^2_{i,t+1} = \hat{\alpha}_{1,i} + \hat{\alpha}_{2,i}\epsilon^2_{i,t} + \hat{\beta}_{1,i}\sigma^2_{i,t}. \tag{3}$$

The multivariate GARCH would be more relevant to the empirical volatility modeling and the risk measure forecasting. The multivariate GARCH model can be expressed as

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{Z}_t, \tag{4}$$

where $\boldsymbol{r}_t$ represents $d \times 1$ vector of the log returns at time t and $\boldsymbol{\mu}_t$ is treated as the constant mean vector of log returns. $\boldsymbol{\Sigma}_t^{1/2}$ is the Cholesky factor of a positive definite conditional covariance matrix $\boldsymbol{\Sigma}_t$. The innovations $\boldsymbol{Z}_t$ could be assumed to follow Student't distribution. There are plenty of possibilities in specifying $\boldsymbol{\Sigma}_t$, e.g., modeling the conditional covariance directly or modeling the conditional variances and correlations instead. According to Bollerslev (1990), the conditional covariance matrix $\boldsymbol{\Sigma}_t$ can be decomposed to a constant conditional correlation matrix and time-varying conditional standard deviations:

$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R}_c \boldsymbol{D}_t, \tag{5}$$

where $\boldsymbol{D}_t = \text{diag}\{\sigma_{i,t}\}_{i=1}^d$ is the diagonal matrix of the standard deviations of i-th assets, $i = 1, \ldots, d$. $\boldsymbol{R}_c$ denotes the constant correlation matrix:

$$\boldsymbol{R}_c = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,d} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{d,1} & \rho_{d,2} & \cdots & 1 \end{pmatrix}. \tag{6}$$

The conditional variances could follow the GARCH(1,1) process. The conditional covariance is written as

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_{1,t}^2 & \rho_{1,2}\sigma_{1,t}\sigma_{2,t} & \cdots & \rho_{1,d}\sigma_{1,t}\sigma_{d,t} \\ \rho_{2,1}\sigma_{2,t}\sigma_{1,t} & \sigma_{2,t}^2 & \cdots & \rho_{2,d}\sigma_{2,t}\sigma_{d,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d,1}\sigma_{d,t}\sigma_{1,t} & \rho_{d,2}\sigma_{d,t}\sigma_{2,t} & \vdots & \sigma_{d,t}^2 \end{pmatrix}. \tag{7}$$

The CCC-GARCH model can be estimated in two steps: the conditional variances in d margins are firstly fitted by univariate GARCH model and the standardized residuals $\hat{\boldsymbol{\epsilon}}_t$, then the constant conditional correlation of the standardized residuals is estimated in the second step. The estimation of CCC-GARCH is computationally easy but might be too restrictive in practice. Engle & Sheppard (2001) extended the CCC model to the DCC-GARCH framework which allows for a time-varying correlation matrix $\boldsymbol{R}_t$ instead of $\boldsymbol{R}_c$. In details,

$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t, \tag{8}$$

where $\boldsymbol{D}_t$ is the diagonal matrix of the conditional standard deviations and $\boldsymbol{R}_t$ can be written as $\boldsymbol{R}_t = \boldsymbol{Q}_t^* \boldsymbol{Q}_t \boldsymbol{Q}_t^*$. The $\boldsymbol{Q}_t$ can be expressed as

$$\boldsymbol{Q}_t = (1 - a - b)\bar{\boldsymbol{R}} + a\boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + b\boldsymbol{Q}_{t-1} \tag{9}$$

and

$$\boldsymbol{Q}_t^* = (\boldsymbol{Q}_t \odot \boldsymbol{I}_k)^{-1/2}, \tag{10}$$

where $\odot$ denotes the Hadamard multiplication. The standardized returns $\boldsymbol{u}_{t-1}$ can be represented $\boldsymbol{u}_{t-1} = \boldsymbol{R}_{t-1}^{1/2}\boldsymbol{e}_{t-1}$ where $\boldsymbol{e}_{t-1}$ is the i.i.d. innovations with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_d$. $\bar{\boldsymbol{R}}$ is the unconditional covariance matrix of $\boldsymbol{u}_t$. Compared to the estimation of the CCC-GARCH model, the estimation of the DCC-GARCH model requires the estimation of dynamic parameters a and b in equation (9) in the third step. The one-step-ahead VaR and ES forecasts rely on the forecast of the conditional covariance matrix $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{D}_{t+1}\boldsymbol{R}_{t+1}\boldsymbol{D}_{t+1}$ given the constant $\boldsymbol{\mu}_t$. The same GARCH model is utilized in the prediction of the univariate variances, hence the VaR and ES predictions are only determined by the forecasts of the correlation matrix $\boldsymbol{R}_{t+1}$. In CCC-GARCH model, the correlation matrix $\boldsymbol{R}_c$ is set to be constant over time, as a result, the conditional covariance matrix turns to be $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{D}_{t+1}\boldsymbol{R}_c\boldsymbol{D}_{t+1}$. With the conditional covariance matrix $\boldsymbol{\Sigma}_{t+1}$ one can generate the predicted distribution at $t+1$. For example, we resort to parametric Monte Carlo simulation to determine the simulated distribution of returns for VaR and ES forecasts as described in Algorithm 1 in Appendix A.1.

3 Structural Break Tests for the Correlation Matrix

Unlike the approaches dealing with the structural break in bivariate time series, e.g., Wied, Krämer & Dehling (2012), Wied (2017) proposed a fluctuation test for structural break in d-dimensional correlation matrix. Define the vector of pairwise correlations as $\boldsymbol{\rho}_t = \{\rho_t^{i,j}\}_{1\le i<j\le p}$. The null hypothesis H_0 is $\boldsymbol{\rho}_1 = \ldots = \boldsymbol{\rho}_T$ whereas alternative hypothesis $H_1 : \neg H_0$. The test statistics is given by

$$Q_{1,T}^W = \max_{2\le k\le T} \frac{k}{\sqrt{T}} ||\hat{E}_{1,T}^{-1/2} P_{k,1,T}||_1, \tag{11}$$

where $P_{k,1,T} = \{\hat{\rho}_{1,k}^{ij} - \hat{\rho}_{1,T}^{ij}\}_{1\le i\le j\le p} \in \mathbb{R}^{p(p-1)/2}$ and $\hat{\rho}_{1,k}^{ij}$ is the estimated pairwise correlation between variable i and j based on the first k observations and $\hat{E}_{1,T}$ is the bootstrap estimator of asymptotic covariance of $\{\hat{\rho}_{1,T}^{ij}\}_{1\le i<j\le p}$. According to Corollary

1 in Wied (2017), under the null hypothesis and Assumptions 1-5 in Wied (2017) hold, we have

$$\max_{2\leq k\leq T}\frac{k}{\sqrt{T}}||\hat{E}_{1,T}^{-1/2}P_{k,1,T}||_1 \rightarrow_d \sup_{0\leq s\leq 1}||B^{\frac{p(p-1)}{2}}(s)||_1, \tag{12}$$

where $B^{\frac{p(p-1)}{2}}(s)$ is a vector of $\frac{p(p-1)}{2}$-dimensional Brownian bridge. The H_0 will be rejected as long as $\max_{2\leq k\leq T}\frac{k}{\sqrt{T}}||P_{k,1,T}||_1$ is larger than the $1-\alpha$ quantile of $\sup_{0\leq s\leq 1}||B^{\frac{p(p-1)}{2}}(s)||_1$. The factor E can be estimated with the help of block bootstrap estimator. For $b = 1,\ldots,B$, the vector of pairwise correlation coefficient based on b bootstrapped sample is $v_b = \{\hat{\rho}_{b,T}^{ij}\}_{1\leq i<j\leq p}$, then the empirical covariance matrix is $\hat{E}_{1,T} = \frac{1}{B}\sum_{b=1}^{B}(v_b-\hat{v})(v_b-\hat{v})'$, where $\hat{v} = \frac{1}{B}\sum_{b=1}^{B}v_b$. To identify and locate multiple break points in correlations of financial assets, Galeano & Wied (2017) proposed an effective algorithm which is summarized as Algorithm 2 in Appendix A.1. Assuming that $z_1,\cdots,z_l$ are the timing of the changes in the correlation matrix. The test statistics in (11) is the rewritten as

$$Q_{1,T}^W = \sup_{0\leq z\leq 1}\frac{\tau(z)}{\sqrt{T}}||\hat{E}_{1,T}^{-1/2}P_{\tau(z),1,T}||_1, \tag{13}$$

where $P_{\tau(z),1,T} = \{\hat{\rho}_{1,\tau(z)}^{ij} - \hat{\rho}_{1,T}^{ij}\}_{1\leq i\leq j\leq p} \in \mathbb{R}^{p(p-1)/2}$ and $\tau(z) = [2+z(T-2)]$ with floor function $[\cdot]$. The estimator of the change point is determined in terms of the break fraction such that $\hat{z} = \tau(\hat{z}^*)/T$ with $\hat{z}^* = \arg\max_k\frac{\tau(z)}{\sqrt{T}}||P_{\tau(z),1,T}||_1$. The test statistics constructed on the subsamples during the iterations follows

$$Q_{\eta(l_1),\tau(l_2)}^W(z) = \frac{\tau(z)-\eta(l_1)+1}{\sqrt{\tau(l_2)-\eta(l_1)+1}}||\hat{E}_{\eta(l_1),\tau(l_2)}^{-1/2}P_{\tau(z),\eta(l_1),\tau(l_2)}||_1, \tag{14}$$

where $\eta(z) = \tau(z)-1$, $z\in[l_1,l_2], \forall 0\leq l_1\leq l_2\leq 1$, $P_{\tau(z),\eta(l_1),\tau(l_2)} = \{\hat{\rho}_{\eta(l_1),\tau(z)}^{ij} - \hat{\rho}_{\eta(l_1),\tau(l_2)}^{ij}\}_{1\leq i\leq j\leq p}$ and $\hat{E}_{\eta(l_1),\tau(l_2)}^{-1/2}$ is the bootstrap estimator based on the sample from $\eta(l_1)$ to $\tau(l_2)$. The asymptotic critical value for a given upper tail probability is defined as $c_{T,\alpha}$. $Q_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)}^W$ is the test statistics based on sample ranged from $\eta(z_{k-1}+\frac{1}{T})$ to $\tau(z_k), k = 1,\ldots,l+1$ with $z_0 = 0, z_{l+1} = 1$.

However, one crucial assumption of the aforementioned nonparametric test is that the marginal variances are required to be constant. Duan & Wied (2018) proposed a residual-based multivariate constant correlation testing procedure. Once again, the null hypothesis is that the vector of pairwise correlations is constant over time. One assumes the true DGP with a single break point λ_0 in both marginal means and marginal variances. The test statistics is in the form of a multivariate cumulative sum of standardized residuals. The feasible test statistics follows

$$\hat{Q}_n = \max_{1\leq j\leq n} \frac{j}{\sqrt{n}}\sqrt{(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n)'\hat{\Omega}^{-1}(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n)}, \tag{15}$$

where the partial sums based on residuals are defined as $\hat{\boldsymbol{S}}_j = \frac{1}{j}\sum_{t=1}^{j} vech(\hat{\boldsymbol{Z}}_t\hat{\boldsymbol{Z}}_t')$. The standardized residuals are

$$\hat{Z}_{t,i} = \frac{X_{t,i} - \hat{\mu}_{i,1}}{\hat{\sigma}_{i,1}}\mathbb{I}_{t/n<\lambda_0} + \frac{X_{t,i} - \hat{\mu}_{i,2}}{\hat{\sigma}_{i,2}}\mathbb{I}_{t/n\geq\lambda_0}, t = 1,\ldots,n, i = 1,\ldots,d. \tag{16}$$

Under the null hypothesis and necessary assumptions, the test statistics convergence to the non-standard limit process. Hence, a block bootstrap method is necessary for the approximation of the limit process. The procedure is summarized in Algorithm 3 in Appendix A.1 and the details can be found in Duan & Wied (2018).

4 Backtesting Procedures

4.1 Backtesting Value-at-Risk

This subsection shortly review a series of tests for VaR and ES backtesting. The backtesting procedure is defined as that the risk measure forecasts are compared to the actual financial losses within a particular time horizon. The backtesting of the risk measures investigates whether the risk measure forecasts are correctly specified. Following the definition in Christoffersen (2003), based on the *ex ante* VaR forecasts and *ex post* returns, a hit sequence of VaR violations can be defined as

$$I_{t+1} = \begin{cases} 1, & r_{t+1} < VaR_{t+1}^{\alpha} \\ 0, & r_{t+1} > VaR_{t+1}^{\alpha}. \end{cases} \tag{17}$$

The hit sequence gives a 1 if the return on $t+1$ is smaller than the VaR forecast VaR_{t+1}^{α} at time $t+1$ whereas the hit sequence gives a 0 if the return outperforms the predicted VaR at time $t+1$. Under a perfect VaR forecasting model, the hit sequence of VaR violations should not be completely predictable and follows Bernoulli distribution over time, i.e., $I_{t+1} \sim_{i.i.d.} Bernoulli(\alpha)$. In the VaR backtesting, when the hit sequence is obtained from a correctly specified VaR model, VaR forecasts at level α violate the realized returns only in α of the days in the time horizon.

Kupiec (1995) proposed an unconditional coverage test which enables us to test if the empirical VaR violation rate from a particular risk model is statistically significant from the expected level α implied by the VaR confidence level. The likelihood of i.i.d. hit sequence is

$$\mathcal{L}(\pi) = (1-\pi)^{T_0}\pi^{T_1}, \tag{18}$$

where T_0 and T_1 are the number of non-violations and violations, respectively. π is the VaR violation rate from a specific VaR model, which could be simply estimated by the empirical fraction of violations $\hat{\pi} = \frac{T_1}{T_0+T_1}$. The likelihood function with estimated violation rates turns to be

$$\mathcal{L}(\hat{\pi}) = (1 - \frac{T_1}{T_0+T_1})^{T_0}(\frac{T_1}{T_0+T_1})^{T_1}. \tag{19}$$

Under the null hypothesis that $\pi = \alpha$, the likelihood function follows

$$\mathcal{L}(\alpha) = (1-\alpha)^{T_0}\alpha^{T_1}. \tag{20}$$

In order to test the unconditional coverage hypothesis, it is sufficient to employ a likelihood ratio test,

$$Q_K = -2\log\left(\frac{(1-\alpha)^{T_0}\alpha^{T_1}}{(1-\frac{T_1}{T_0+T_1})^{T_0}(\frac{T_1}{T_0+T_1})^{T_1}}\right). \tag{21}$$

This test statistics is asymptotically χ^2-distributed with one degree of freedom. If the p-value $p = 1 - F_{\chi_1^2}(Q_K)$ is smaller than the desired significance level, the H_0 is rejected.

Since the arrivals of VaR violations are expected to randomly spread over the time horizon, the VaR models which generate clustered VaR violations over time would be rejected in the application. Christoffersen (1998) proposed a test for the assumption of i.i.d. distribution of VaR violations. If the sequence of violation is dependent over time, then it can be assumed as a first order Markov sequence with the transition probability matrix. Christoffersen (2003) defined an indicator variable I_t such that $I_t = 0$ corresponds to no violation at time t whereas $I_t = 1$ indicates the violation occurs at t. In addition, π_{ij} is the transition probability from state i at time t to state j at time $t+1$. The likelihood function of the first order Markov process is

$$\mathcal{L}(\Pi_1) = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}, \tag{22}$$

where $T_{ij}, i, j = 0, 1$ is the number of observations from state i to state j. Under the independence hypothesis, the probability that the violation occurs tomorrow does not depend on the state today, i.e., $\pi = \pi_{01} = \pi_{11}$. Next, one defines the estimated probability of the occurance of the violation: $\hat{\pi}_{01} = \frac{T_{01}}{T_{01}+T_{00}}$ and $\hat{\pi}_{11} = \frac{T_{11}}{T_{11}+T_{10}}$. A likelihood ratio test can be used to test the null hypothesis of independence.

Engle & Manganelli (2004) proposed a dynamic quantile test for the distribution of VaR violations, or so-called CAViaR test. One could evaluate the performance of VaR forecasts with such a testing procedure. Similarly, as shown in Berens et al. (2015), the following autoregression based on the VaR violations can be considered:

$$I_t = \alpha + \beta_1 I_{t-1} + \beta_2 VaR_t + u_t, \tag{23}$$

where u_t can be assumed to follow a logistic distribution. Under the null hypothesis that the model is correctly specified, the statistical significance of coefficients $\beta_1 = \beta_2 = 0$ can be tested with a likelihood ratio test. The test statistics follows χ^2_2 asymptotically.

Diebold & Mariano (1995) proposed a test for relative comparison forecasting performance between two VaR forecasting models in terms of the forecasting accuracy. $L(r_{t+1}, VaR^k_{t+1})$ is a generic loss function for the actual return r_{t+1} and the α-VaR forecast obtained from k-th model $VaR^k_{t+1}, k = A, B$ at time $t+1$. The test statistics is

$$Q_{DM} = \frac{R^{-1}\sum_{t=M+1}^{M+R}(L(r_{t+1}, VaR^A_{t+1}) - L(r_{t+1}, VaR^B_{t+1}))}{\sqrt{\hat{\Sigma}(d_t)}} \tag{24}$$

where $\hat{\Sigma}(d_t)$ denotes the HAC estimator of the long run variance of $d_t := L(r_{t+1}, VaR^A_{t+1}) - L(r_{t+1}, VaR^B_{t+1})$. M is the number of observations used in estimation period whereas R is the number of observations in the prediction period. VaR^A_{t+1} and VaR^B_{t+1} are predicted VaR at time $t+1$ from VaR model A and B, respectively. The null hypothesis is $E(d_t) = 0$ and the alternative hypothesis is $E(d_t) < 0$ and $E(d_t) > 0$. Large negative (positive) test statistics indicates VaR model A outperforms (underperforms) VaR model B. Test statistics being close to zero implies the prediction performance of two models are equivalent. An example of the loss function is

$$L(r_{t+1}, VaR_{t+1}) = \alpha(r_{t+1} - VaR_{t+1})(1 - \mathbb{I}_{\{r_{t+1} < VaR_{t+1}\}}) + (1-\alpha)(VaR_{t+1} - r_{t+1})\mathbb{I}_{\{r_{t+1} < VaR_{t+1}\}}. \tag{25}$$

An alternative approach for comparing the predictive performance between VaR models is Conditional Predictive Ability (CPA) test from Giacomini & White (2006). In addition to the standard squared error or the absolute error loss function, as assumed in Santos et al. (2013), an asymmetric linear loss function of order α is given as:

$$L_\alpha(e_{t+1}) = (\alpha - \mathbb{I}_{\{e_{t+1}<0\}})e_{t+1}, \text{ for } \alpha \in (0,1), \tag{26}$$

where $e_{t+1} = r_{t+1} - VaR^\alpha_{t+1}$ is the prediction error. The out-of-sample loss difference from predictive model A and B is defined as $\Delta L_\alpha(e_{t+1}) = L^A_\alpha(e_{t+1}) - L^B_\alpha(e_{t+1})$. Under the null hypothesis of equivalent predictive abilities in two models, the moment restriction holds: $E(\Delta L_\alpha(e_{t+1})|\mathcal{F}_t) = 0$ almost surely $t = 1, 2, \ldots$, or equivalently $E(h_t \Delta L_\alpha(e_{t+1})) = 0$, where h_t is a $\mathcal{F}_t$-measurable test function of instruments which helps predicting the difference between two models in terms of the forecast performance, e.g., $h_t = (1, \Delta L_\alpha(e_t))$ as in Giacomini & White (2006). Under the null hypothesis, the difference of loss between two models follows a martingale difference process. The test statistics is in the Wald-type:

$$Q_{CPA} = T(T^{-1}\sum_{t=1}^{T-1} h_t \Delta L_\alpha(e_{t+1}))'\hat{\Omega}^{-1}(T^{-1}\sum_{t=1}^{T-1} h_t \Delta L_\alpha(e_{t+1})), \tag{27}$$

where $\hat{\Omega}$ is the consistent estimator of the variance of $h_t \Delta L_\alpha(e_{t+1})$. The null hypothesis is rejected when $Q_{CPA} > \chi^2_{T,1-\alpha}$.

4.2 Backtesting Expected Shortfall

To backtest the expected shortfall, we are interested in the discrepancy between actual loss of return r_{t+1} and predicted ES e^α_t when the VaR exceedance occurs on $t+1$ such that the condition $E((r_{t+1} - e^\alpha_t)I_{t+1}|\mathcal{F}_t) = 0$ is fulfilled under the null hypothesis, see McNeil & Frey (2000). Recall that the dynamics of log returns r_t can be written as $r_t = \mu_t + \sigma_t Z_t$, where $Z_t \sim_{i.i.d.} F_Z(z)$ with mean zero and unit variance. Then the residuals are defined as

$$er_{t+1} = \frac{r_{t+1} - e^\alpha_t}{\sigma_{t+1}} = Z_{t+1} - E(Z|Z < z_\alpha). \tag{28}$$

These residuals are i.i.d. and have zero conditional expectation on $r_{t+1} < e^\alpha_t$. The residuals can be standardized as $er^*_{t+1} = \frac{er_{t+1}}{\sqrt{\text{Var}(Z|Z<z_\alpha)}}$. The empirical counterpart

of the standardized residuals, namely exceedance residuals are expressed as $\hat{\text{er}}^*_{t+1} = \frac{r_{t+1} - \hat{e}^\alpha_t}{\hat{\sigma}_{t+1}\sqrt{\hat{\text{Var}}(Z|Z<z_\alpha)}}$. $\hat{e}^\alpha_t$ is the estimate of the expected shortfall and $\hat{\text{Var}}(Z|Z < z_\alpha)$ is the estimate of the variance of the random variance Z being truncated at z_α. Under the null hypothesis that the distribution of Z_t is correctly specified and the expected shortfall is correctly estimated, the exceedance residuals should be i.i.d. with mean zero and unit variance. This hypothesis can be tested with the help of the bootstrap procedure proposed by Efron & Tibshirani (1993) without assuming the underlying distribution of the residuals. Furthermore, it applies to both the standardized residuals and unstandardized ones.

As pointed out by Weber (2006), Fissler et al. (2015) and Fissler & Ziegel (2016), the main challenges of ES backtesing is the non-elicitability and non-identifiability. Many existing approaches perform the ES backtesting based on the entire or tail distribution of the return series, or the cumulative violation process, e.g. Löser et al. (2018). Some tests requires the auxiliary quantities, for example the information of the VaR forecasts as additional inputs for the ES backtesting in addition to the ES forecasts themselves, see Nolde & Ziegel (2017). Alternatively, Bayer & Dimitriadis (2020) proposes new regression-based ES backtests which only require ES forecasts as input variables. The log returns are regressed on the ES forecasts and an intercept term. If the ES forecasting model is correctly specified, the intercept term should be zero and the slope should be one. The testing procedure follows Mincer & Zarnowitz (1969),

$$r_t = \gamma_1 + \gamma_2 \hat{e}^\alpha_t + u^e_t. \tag{29}$$

The functional ES of r_t at α can be written as $ES_\alpha(r_t|\mathcal{F}_{t-1}) = \gamma_1 + \gamma_2 \hat{e}^\alpha_t$ given that $ES_\alpha(u^e_t|\mathcal{F}_{t-1}) = 0$ almost surely. The test hypothesis is

$$H_0 : (\gamma_1, \gamma_2) = (0, 1) \quad \text{vs.} \qquad H_1 : (\gamma_1, \gamma_2) \neq (0, 1). \tag{30}$$

Under the null hypothesis, $\hat{e}^\alpha_t = ES_\alpha(r_t|\mathcal{F}_{t-1})$ holds almost surely when the model for ES forecasts is correctly specified. As pointed out by Gneiting (2011), the ES is not elicitable and the regression model can not be estimated for the ES on a standalone basis since the strictly consistent loss and identification functions are not available for the functional ES. Based on the joint loss and identification functions for VaR and ES in Fissler & Ziegel (2016), Patton et al. (2019) and Dimitriadis & Bayer (2019) propose a joint regression framework to estimate the regression parameters:

$$r_t = V_t'\beta + u_t^v \quad \text{and} \quad r_t = W_t'\gamma + u_t^e, \tag{31}$$

where V_t and W_t are k-dimensional, $\mathcal{F}_{t-1}$ measurable covariate vectors and $Q_t(u_t^v|\mathcal{F}_{t-1}) = 0$ and $ES_\alpha(u_t^e|\mathcal{F}_{t-1}) = 0$ almost surely. First, they employ the auxiliary VaR forecasts $\hat{v}_t^\alpha$ at level $\alpha \in (0,1)$ as the explanatory variable in the quantile equation, i.e., the $V_t = (1, \hat{v}_t^\alpha)$ and $W_t = (1, \hat{e}_t^\alpha)$ are selected. The auxiliary ESR backtest is presented as

$$r_t = \beta_1 + \beta_2 \hat{v}_t^\alpha + u_t^v \quad \text{and} \quad r_t = \gamma_1 + \gamma_2 \hat{e}_t^\alpha + u_t^e. \tag{32}$$

The hypotheses are tested via the Wald-type test statistics:

$$T_{A-ESR} = T(\hat{\gamma}_T - (0,1))\hat{\Omega}_\gamma^{-1} T(\hat{\gamma}_T - (0,1))', \tag{33}$$

where $\hat{\Omega}_\gamma^{-1}$ is the consistent covariance estimator for the covariance of the parameter vector γ. The choice $V_t = W_t = (1, \hat{e}_t^\alpha)$, i.e., only the information of ES forecasts being utilized, gives us the second ES regression test, namely the strict ESR backtest. It is the first backtesting for the ES standalone, but it potentially suffers from the model misspecification, see Bayer & Dimitriadis (2020) for greater details.

The banks and financial institutions have incentives to report less conservative risk forecasts which leads to lower capital requirements. The concern of the regulators only focuses on the prevention of underestimation of financial risks, hence an ES backtesting procedure for one-sided hypotheses would be sufficient. The third version of the regression backtesting is proposed, i.e., the forecast error $r_t - \hat{e}_t^\alpha$ is only regressed on an intercept term in the ES-specific regression:

$$r_t - \hat{e}_t^\alpha = \beta_1 + \beta_2 \hat{e}_t^\alpha + u_t^v \quad \text{and} \quad r_t - \hat{e}_t^\alpha = \gamma_1 + u_t^e. \tag{34}$$

The one-sided hypothesis is defined as

$$H_0 : \gamma_1 \geq 0 \quad \text{vs.} \quad \gamma_1 < 0. \tag{35}$$

The testing is accomplished by t-tests with the estimated asymptotic covariance matrix, please see Bayer & Dimitriadis (2020) for further details.

5 Empirical Illustration

We consider the log return series of eight stocks in STOXX50, DAX30, FTSE100 and S&P500 indices, respectively. In each index, the eight stocks with highest market values on January 1st 2010 with complete data history are selected into the portfolio under consideration. Equally weighted assets constitute of each portfolio. The log returns are computed from the sample ranged from January 1st 2005 to December 29th 2017 with 3390 observations and the non-trading days are excluded. The raw data is extracted from Thomson Reuters Financial Datastream. The descriptive statistics can be found in Table 1. The annualized volatilities of the log-returns range from 19.492% to 23.155%. The excess kurtosis also shows that the fat tailed-ness exist in each of the four portfolios.

Table 1: Summary Statistics

	STOXX50	DAX30	FTSE100	S&P500
Minimum	-8.782%	-10.705%	-8.269%	-12.359%
5% Quantile	-1.881%	-2.232%	-1.570%	-2.085%
Median	0.050%	0.080%	0.039%	0.090%
95% Quantile	1.887%	2.056%	1.538%	2.029%
Mean	0.037%	0.031%	0.025%	0.059%
Maximum	9.611%	13.826%	9.042%	11.444%
Volatility	1.228%	1.422%	1.036%	1.459%
Skewness	0.028	-0.015	-0.002	-0.026
Excess Kurtosis	6.445	10.472	7.619	11.764
Annualized Volatility	19.492%	22.580%	16.449%	23.155%

Note: The data set consists of log returns of four portfolios from January 2005 to December 2017. The annualized volatility is calculated based on 252 days.

The univariate modeling of volatility is performed by fitting the GARCH(1,1) model to the log return of the asset in a rolling window of 1000 observations up to time t. In order to make the one-step-ahead forecast of the volatility for the i-th asset $\hat{\sigma}_{i,t+1}$, the equation (3) with the estimated α_1, α_2 and β_1 are employed to compute. The innovations are fixed to be Student's t distributed with mean 0, scale parameter 1 and $\nu = 15$ degrees of freedom in all margins.

Two different structural break tests for correlations are performed prior to the estimation of two conditional correlation models. The standardized residuals $\hat{z}_{i,t}$ being filtered from the aforementioned univariate GARCH model are fed into the structural break tests. The standardized residuals are computed by $\hat{z}_{i,t} = \hat{\epsilon}_{t,i}\hat{\sigma}_{i,t}^{-1}$ where $\hat{\sigma}_{i,t}$ is the plug-in estimator of the conditional volatility according to (2) and $\hat{\epsilon}_{i,t}$ is obtained by subtracting $\hat{\mu}_i$ from $r_{i,t}$. The structural break tests are applied to the residuals in the rolling window with the length $T = 1000$ up to time t. Two distinct versions of each structural break test are considered here, i.e., the tests could detect either an unique break point or multiple break points. For each test, the data point at which the test statistic takes the maximum value is firstly determined based on the test sample. If the maximum of the test statistics is equal to or larger than the corresponding critical value, the null hypothesis of the constant correlation is rejected, which means that the point associated with the maximum test statistic is decided as the single break point. To obtain the multiple change points, the full sample is segmented into two subsamples at the single change point and one searches for any significant change points in the subsamples. This procedure continues until no significant points can be found according to Galeano & Wied (2017). The sample starting from the latest identified break point to the last point in the sample is utilized to estimate the constant correlation matrix $\boldsymbol{R}_c$ and time-varying correlation matrix $\boldsymbol{R}_t$. The predicted correlations at $t+1$ are accomplished by the CCC and DCC models as described in Section 2. The one-step-ahead prediction of correlations is chosen to rely on the rolling window since the estimated parameters based on the sample within the rolling window are able to account for the most recent information.

Next, in order to calculate the VaR and ES forecasts at $t+1$, it is necessary to obtain the predicted distributions of the portfolio returns which are determined through simulations, i.e., for i-th asset in the portfolio at day $t+1$, $K = 10000$ random observations following Student's t distribution with the degree of freedom $\nu = 15$ are generated as innovations and then transformed into simulated returns $\{\tilde{r}_{t+1}\}_{k=1}^{K}$

with the help of the estimated mean $\hat{\mu}$ and predicted volatility $\hat{\sigma}_{i,t+1}$. The simulated individual log returns construct the corresponding portfolio. The α-quantile of the simulated portfolio log-returns is the VaR at the α level. The parameters $\alpha = 0.05$ and $\alpha = 0.01$ are selected as two significance levels for VaR and ES forecasts at which the VaR and ES forecasts are compared with the realized portfolio returns to assess the model performance, respectively. The ES forecasts at day $t + 1$ are obtained by the mean of the simulated log returns below the estimated VaR_α at the day.

An acceptable VaR forecasting model should generate an appropriate VaR violation ratio which is defined as the actual number of VaR violations divided by the total number of VaR forecasts, i.e., the VaR violation ratio from a VaR forecasting model is close enough to the VaR quantile α. Panel A in Table 2 presents the VaR violation ratios for 5%-VaR and 1%-VaR from all six VaR prediction models for four portfolios. The correlation models integrated with structural break tests generally deliver the VaR violation ratios which are closer to the nominal VaR percentages than that from the correlation models without any tests. The CCC models dominate DCC models in terms of the VaR violation ratios. Among all correlation models, the CCC model with Wied (2017) test (hereinafter referred to as the 'wied17' or the 'non-parametric' test) detecting multiple break points outperforms other models. Furthermore, the results of Kupiec (1995) unconditional coverage test are shown in Panel B in Table 2. The null hypothesis is not rejected for all cases of 5%-VaR forecasts. When it comes to the 1%-VaR, the test results are mostly rejected at 5% or 1% significance level and the null hypothesis is not rejected for few portfolios based on the correlation models with the *wied17* test at all significance levels, which confirms the findings in the VaR violation ratios.

To investigate the existence of VaR violation clusters, the Christoffern test and the CaViaR test are employed to evaluate the model performance from the perspective of the distribution of the VaR violations. In Panel C in Table 2, the null hypothesis of Christoffern test is frequently rejected for the 5%-VaR forecasts in FTSE portfolio at 10% and 5% significant level, respectively. The CCC model combined with non-parametric test with multiple change points and the DCC model with both types of the non-parametric test improve the model performance being compared to the correlation

Table 2: The results of VaR forecasts in four portfolios

			5%-VaR				1%-VaR			
Model	Test	cp	STOXX	DAX	FTSE	S&P	STOXX	DAX	FTSE	S&P
Panel A: VaR Violation Ratio										
CCC	No Test	-	0.052	0.053	0.050	0.051	0.016	0.015	0.014	0.016
	Wied 2017	Single	0.051	0.050	0.049	0.048	0.015	0.013	0.014	0.016
		Multiple	0.050	0.050	0.049	0.046	0.014	0.012	0.012	0.013
	DW 2018	Single	0.054	0.054	0.053	0.050	0.016	0.014	0.015	0.017
		Multiple	0.052	0.056	0.053	0.052	0.015	0.015	0.015	0.021
DCC	No Test	-	0.056	0.057	0.057	0.054	0.017	0.019	0.014	0.018
	Wied 2017	Single	0.052	0.055	0.053	0.053	0.015	0.015	0.013	0.018
		Multiple	0.051	0.055	0.052	0.050	0.014	0.015	0.013	0.015
	DW 2018	Single	0.055	0.056	0.057	0.054	0.017	0.017	0.015	0.018
		Multiple	0.053	0.057	0.056	0.054	0.015	0.016	0.014	0.020
Panel B: Kupiec Test										
CCC	No Test	-	0.660	0.458	0.985	0.889	0.008***	0.020**	0.104	0.010**
	Wied 2017	Single	0.876	0.989	0.908	0.659	0.036**	0.163	0.069*	0.006***
		Multiple	0.973	0.989	0.908	0.405	0.056*	0.418	0.295	0.143
	DW 2018	Single	0.416	0.403	0.581	0.965	0.008***	0.050*	0.029**	0.002***
		Multiple	0.730	0.194	0.581	0.670	0.014**	0.020**	0.029**	0.000***
DCC	No Test	-	0.202	0.115	0.137	0.421	0.003***	0.000***	0.069*	0.001***
	Wied 2017	Single	0.594	0.265	0.459	0.538	0.023**	0.020**	0.151	0.001***
		Multiple	0.802	0.307	0.647	0.958	0.056*	0.032**	0.151	0.016**
	DW 2018	Single	0.237	0.227	0.137	0.369	0.003***	0.004***	0.029**	0.001***
		Multiple	0.471	0.138	0.164	0.421	0.014**	0.007***	0.045**	0.000***
Panel C: Christoffern Test										
CCC	No Test	-	0.282	0.186	0.083*	0.618	0.619	0.299	0.441	0.132
	Wied 2017	Single	0.050*	0.347	0.074*	0.914	0.523	0.373	0.471	0.146
		Multiple	0.040**	0.183	0.160	0.946	0.330	0.423	0.355	0.063*
	DW 2018	Single	0.209	0.202	0.012**	0.904	0.140	0.327	0.104	0.179
		Multiple	0.461	0.083*	0.031**	0.718	0.287	0.299	0.104	0.089*
DCC	No Test	-	0.040**	0.212	0.097*	0.816	0.686	0.201	0.471	0.042**
	Wied 2017	Single	0.078*	0.067*	0.178	0.885	0.301	0.299	0.411	0.215
		Multiple	0.057*	0.125	0.136	0.778	0.330	0.313	0.063*	0.118
	DW 2018	Single	0.077*	0.074*	0.097*	0.478	0.686	0.259	0.104	0.196
		Multiple	0.192	0.195	0.007***	0.856	0.287	0.272	0.092*	0.071*
Panel D: CaViaR Test										
CCC	No Test	-	0.542	0.190	0.136	0.773	0.876	0.559	0.741	0.157
	Wied 2017	Single	0.144	0.266	0.047**	0.938	0.801	0.261	0.440	0.094*
		Multiple	0.120	0.244	0.201	0.922	0.609	0.505	0.446	0.117
	DW 2018	Single	0.431	0.420	0.013**	0.976	0.302	0.405	0.265	0.137
		Multiple	0.762	0.221	0.031**	0.898	0.567	0.471	0.209	0.052*
DCC	No Test	-	0.116	0.336	0.243	0.816	0.901	0.426	0.683	0.070*
	Wied 2017	Single	0.211	0.140	0.247	0.931	0.567	0.367	0.594	0.189
		Multiple	0.152	0.231	0.165	0.908	0.586	0.407	0.139	0.179
	DW 2018	Single	0.200	0.159	0.129	0.772	0.798	0.415	0.198	0.231
		Multiple	0.426	0.429	0.009***	0.961	0.562	0.385	0.159	0.049**

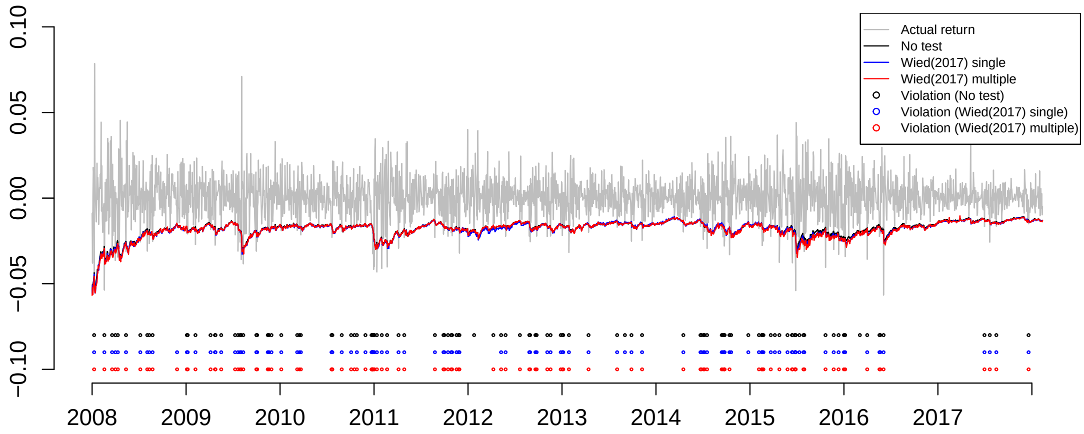
Note: Two types of change points (denoted by 'cp') are allowed: single and multiple change points. The notations $*, **$ and $***$ indicate the 10%, 5%, and 1% statistical significance levels, respectively.

models without any structural break tests, respectively. In addition, the DCC model with Duan & Wied (2018) test (hereinafter referred to as the 'dw18' or the 'residual-based' test) with multiple change points show the improvement in STOXX portfolio. With the respect to 1%-VaR prediction, the DCC models with both versions of the *wied17* test and the *dw18* test outperform the plain DCC model in S&P portfolio. The results of CaViaR test show no significant improvement in the correlation models with structural break tests.

In Figure 1, it is hard to distinguish the CCC model and its counterparts including the break points identified by Wied (2017) test except that the model with multiple break points deliver more conservative VaR forecasts during the volatile period between 2015 and 2016 but less conservative risk forecasts during the calm period in 2012. As Figure 2 presented, the result of the CCC model with its counterparts with Duan & Wied (2018) follows the same pattern except the larger difference between the plain CCC model and the CCC models with structural break tests is observed at the end of the sample period. The same findings can be identified in the comparison between the plain DCC model and its variants with structural break points as illustrated in Figure 3 and Figure 4.

To compare the VaR prediction performance between different VaR forecasting models, the CPA test proposed by Giacomini & White (2006) is used to compare different VaR models. The counting numbers in Table 4 in Appendix A.2 gives the frequency that a specific model candidate is preferred over another model in four portfolios. With respect to 5%-VaR forecasts, the plain CCC model has the comparable performance with the CCC model combined with structural break tests in four portfolios except that the CCC model with the *wied17* test with single change point outperforms the plain CCC model. The plain DCC model generally has higher conditional predictive ability than the CCC and DCC counterparts with different tests. The CCC models being integrated with residual-based tests have better forecasting performance than the CCC model combined with the non-parametric test with multiple change points. The DCC model with the residual-based test with a single change point outperforms the DCC model with the two non-parametric tests. When it comes to 1%-VaR, the comparison of the performance of the CCC model and its counterparts is ambiguous.

Figure 1: The CCC model with the Wied (2017) test for the 5%-VaR forecast in STOXX portfolio

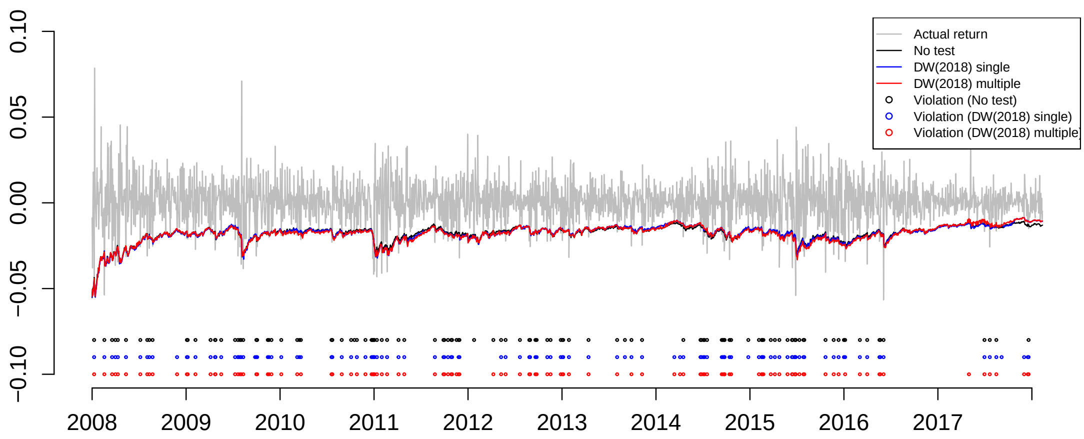


Note: This figure shows the actual returns, 5%-VaR forecasts and 5%-VaR violations based on the CCC model being combied with the Wied (2017) test for the STOXX portfolio with respective lines and dots in different colors, the circles at the bottom of the figure indicate the data points where the respective violations of the VaR forecasts occur.

The plain DCC model outperforms the CCC and DCC models including change points except the CCC model with the residual-based test including multiple change points. The statistically significant results in the last three columns in the Table 4 indicates that the test is generally indifferent between the CCC and DCC models and their counterparts including the structural break tests. The comparison between different VaR forecasting models with respect to the individual portfolio can be found in Table 5 to Table 8 in Appendix A.2.

To evaluate the accuracy of the ES forecasting of different risk forecasting models, three backtests are employed. Panel A in Table 3 shows that the null hypothesis of the exceedance residuals backtest proposed by McNeil & Frey (2000), i.e., the mean of the exceedance residuals is zero or intuitively the risk measurement procedure correctly estimates the ES, is rejected for the 5%-ES forecasts being obtained from all models (including all CCC and DCC models) for DAX, STOXX and S&P portfolios at 0.05 statistical significance levels. It indicates that all variants of the CCC model are ap-

Figure 2: The CCC model with the Duan & Wied (2018) test for the 5%-VaR forecast in STOXX portfolio



Note: This figure shows the actual returns, 5%-VaR forecasts and 5%-VaR violations based on the CCC model being combied with the Duan & Wied (2018) test for the STOXX portfolio with respective lines and dots in different colors, the circles at the bottom of the figure indicate the data points where the respective violations of the VaR forecasts occur.

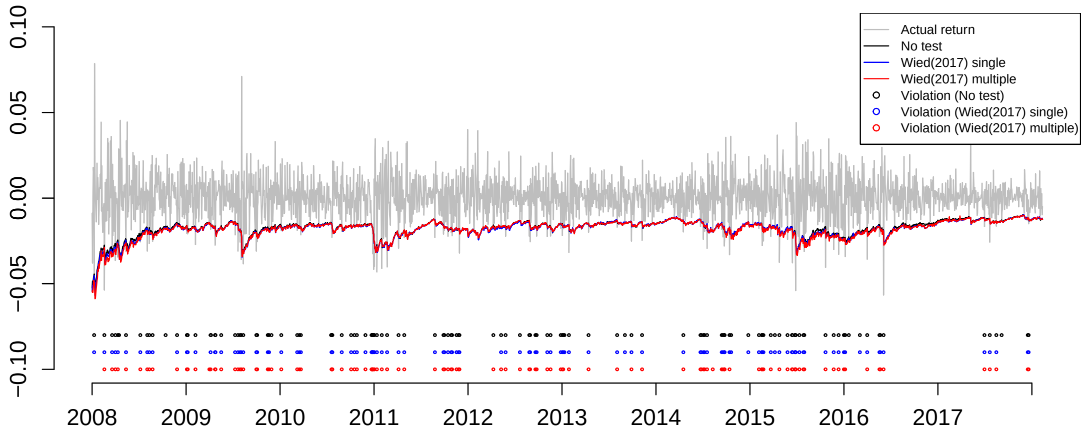
propriate for 1%-ES forecasts in STOXX and FTSE portfolios regardless of statistical significance levels whereas the null hypothesis is rejected for the remaining two portfolios at either 5% or 1% significance levels. All DCC models are not rejected for the 1%-ES forecasts in the STOXX portfolio and the DCC model with the non-parametric test allowing multiple change points is only not rejected for the FTSE portfolio. The result of the conditional calibration backtest proposed by Nolde & Ziegel (2017) is presented in Panel B in Table 3. Since the conservative risk estimation is generally not the main concern of the regulators, i.e., the more capitals reserved than minimally required are allowed, the one-sided calibration test is considered in this context. The inclusion of the non-parametric tests (with single and multiple break points) into the CCC model generally gives less rejection frequency of the null hypothesis across the four portfolios being compared to the cases without any tests or with the residual-based test, which holds for both 5%-ES and 1%-ES forecasts. The conditional calibration backtest delivers non-conclusive result for the DCC models.

Table 3: The results of Expected Shortfall Backtesting

			5%-ES				1%-ES			
Model	Test	cp	STOXX	DAX	FTSE	S&P	STOXX	DAX	FTSE	S&P
Panel A: Exceedance Residuals Backtest p-value										
CCC	No Test	-	0.000***	0.000***	0.002***	0.000***	0.140	0.047**	0.140	0.004***
	Wied 2017	Single	0.009***	0.012**	0.009***	0.000***	0.337	0.022**	0.141	0.017**
		Multiple	0.011**	0.009***	0.070*	0.002***	0.358	0.003***	0.196	0.006***
	DW 2018	Single	0.002***	0.010**	0.006***	0.000***	0.141	0.047**	0.330	0.016**
		Multiple	0.009***	0.003***	0.002***	0.000***	0.284	0.041**	0.143	0.029**
DCC	No Test	-	0.004***	0.000***	0.013**	0.000***	0.314	0.054*	0.016**	0.005***
	Wied 2017	Single	0.009***	0.001***	0.007***	0.000***	0.293	0.006***	0.021**	0.009***
		Multiple	0.004***	0.001***	0.013**	0.000***	0.277	0.009***	0.127	0.008***
	DW 2018	Single	0.002***	0.000***	0.012**	0.000***	0.406	0.026**	0.055**	0.004***
		Multiple	0.003***	0.000***	0.008***	0.000***	0.227	0.011**	0.042**	0.003***
Panel B: Conditional Calibration Backtest p-value										
CCC	No Test	-	0.057*	0.038**	0.241	0.046**	0.068*	0.080*	0.157	0.039**
	Wied 2017	Single	0.198	0.179	0.469	0.091*	0.170	0.179	0.237	0.056*
		Multiple	0.255	0.167	0.681	0.295	0.219	0.170	0.381	0.115
	DW 2018	Single	0.074*	0.093*	0.184	0.084*	0.076*	0.130	0.168	0.042**
		Multiple	0.164	0.044**	0.139	0.024**	0.122	0.089*	0.127	0.013**
DCC	No Test	-	0.031**	0.003***	0.052*	0.010**	0.072*	0.023**	0.075*	0.016**
	Wied 2017	Single	0.119	0.021**	0.138	0.017**	0.111	0.052*	0.105	0.022**
		Multiple	0.181	0.029**	0.259	0.078*	0.173	0.061*	0.205	0.045**
	DW 2018	Single	0.036**	0.014**	0.072*	0.016**	0.068*	0.044**	0.086*	0.020**
		Multiple	0.074*	0.013**	0.069*	0.010**	0.098*	0.048**	0.103	0.010**
Panel C: Intercept ES Regression Backtest p-value										
CCC	No Test	-	0.022**	0.024**	0.089*	0.060*	0.029**	0.088*	0.097*	0.196
	Wied 2017	Single	0.065*	0.084*	0.190	0.059*	0.079*	0.131	0.091*	0.168
		Multiple	0.105	0.072*	0.286	0.194	0.116	0.118	0.174	0.071*
	DW 2018	Single	0.026**	0.041**	0.084*	0.075*	0.053*	0.051*	0.081*	0.079*
		Multiple	0.073*	0.019**	0.059*	0.016**	0.040**	0.069*	0.053*	0.144
DCC	No Test	-	0.008***	0.001***	0.020**	0.011**	0.107	0.041**	0.049**	0.078*
	Wied 2017	Single	0.051*	0.012**	0.055*	0.008***	0.060*	0.072*	0.075*	0.135
		Multiple	0.072*	0.017**	0.101	0.066*	0.066*	0.072*	0.105	0.135
	DW 2018	Single	0.012**	0.006***	0.026**	0.010**	0.021**	0.081*	0.075*	0.046**
		Multiple	0.046**	0.005***	0.028**	0.004***	0.018**	0.104	0.037**	0.004***

Note: Two types of change points (denoted by 'cp') are allowed: single and multiple change points. The exceedance residuals backtest is proposed by McNeil & Frey (2000). The conditional calibration backtest proposed by the Nolde & Ziegel (2017) is one-sided test. The intercept ES regression test proposed Bayer & Dimitriadis (2020) is also an one-sided test. The notations $*, **$ and $***$ indicate the 10%, 5%, and 1% statistical significance levels, respectively.

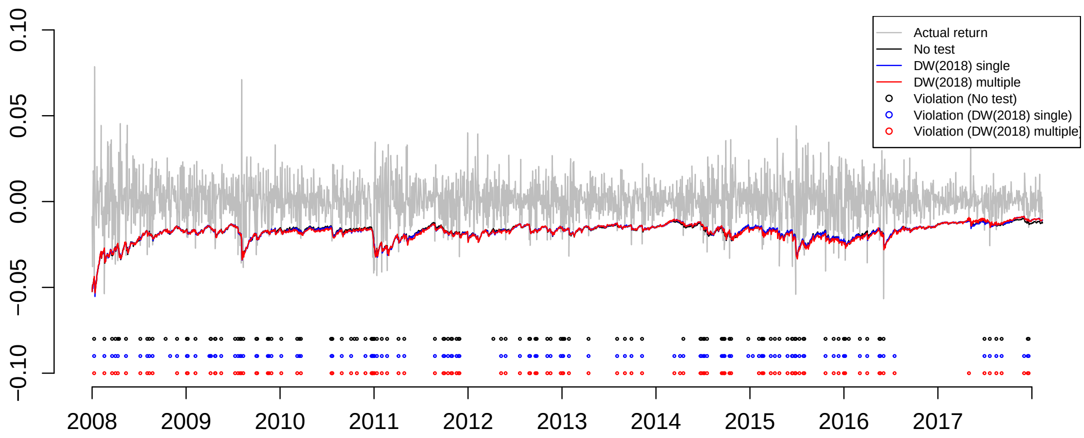
Figure 3: The DCC model with the Wied (2017) test for the 5%-VaR forecast in STOXX portfolio



Note: This figure shows the actual returns, 5%-VaR forecasts and 5%-VaR violations based on the DCC model being combied with the Wied (2017) test for the STOXX portfolio with respective lines and dots in different colors, the circles at the bottom of the figure indicate the data points where the respective violations of the VaR forecasts occur.

Bayer & Dimitriadis (2020) argues that most of backtests require additional input variables such as the VaR forecasts at α level, in response they propose the first strict ES backtests which only require ES forecasts in addition to realized returns as the input variables. Among them, the intercept ES regression test is selected to backtest the ES forecasts against the one-sided alternative due to the incentive of the banks and financial institutions to report too risky forecasts and the mission of regulators to only prevent the underestimation of the risks. The result of the intercept ESR backtest is presented in Panel C in Table 3. Similar to the result of the calibration test, when the CCC models being integrated with the *wied17* test with single and multiple change points are used, the null hypothesis is rejected at a lower frequency across all portfolios, i.e., H_0 is not rejected for the 5%-ES and the 1%-ES forecasts in all portfolios at 0.05 significance level. The null hypothesis is rejected for both 5%-ES and 1%-ES forecasts generated by the DCC model without any structural break tests at 10% statistical significance level for all portfolios except the 1%-ES forecast in the

Figure 4: The DCC model with the Duan & Wied (2018) test for the 5%-VaR forecast in STOXX portfolio



Note: This figure shows the actual returns, 5%-VaR forecasts and 5%-VaR violations based on the DCC model being combied with the Duan & Wied (2018) test for the STOXX portfolio with respective lines and dots in different colors, the circles at the bottom of the figure indicate the data points where the respective violations of the VaR forecasts occur.

STOXX portfolio. The intercept ESR backtest shows that the non-parametric test with a single change point utilized in DCC model is only inappropriate for the 5%-ES forecasts in DAX and S&P portfolios at 5% and 1% statistical significance levels, respectively. The null hypothesis is rejected for the DAX portfolio at 5% significance level when the non-parametric test with multiple change points is used in DCC model. The DCC models combined with the residual tests are rejected for the 5%-ES forecasts in all portfolios. The DCC model with the single change point and with multiple change points being detected in *wied17* test are appropriate for the FTSE portfolio and the STOXX portfolio at 5% significance level, respectively.

6 Conclusion

The impact of the inclusion of the residual-based constant correlation test by Duan & Wied (2018) and the non-parametric correlation test by Wied (2017) into the CCC and DCC models is examined in terms of the potential improvement in the accuracy of VaR and ES forecasting from the plain CCC and DCC models. The implementation of the Galeano & Wied (2017) algorithm enables the detection of both single and multiple change points in the constant correlation tests. We perform several backtesting procedure for VaR and ES forecasts to evaluate the appropriateness of the considered correlation models in a standalone or comparative manner.

To enumerate a few main findings: the coverage rate and the unconditional coverage test shows that the CCC model accounting for the multiple change points identified by the Wied (2017) test outperforms other models whereas the consideration of Duan & Wied (2018) test does not lead better forecasting accuracy in comparison with the plain CCC or DCC models especially for the 1%-VaR forecasts. The CPA test by Giacomini & White (2006) indicates that the CCC model including the Duan & Wied (2018) test leads to a better predictive ability in 5%-VaR forecasting than the CCC model combined with the Wied (2017) test which allows the multiple change points. In addition, the plain DCC model is generally hard to be beaten by the CCC and DCC counterparts accounting for structural break points in either the 5%-VaR or the 1%-VaR forecasts, which seems to be coincided with the results in Berens et al. (2015). The statistically significant results in CPA test also imply that the inclusion of the structural break tests would not be able to deliver a significantly different prediction accuracy in the comparative sense. Among the three tests being employed to backtest the ES forecasting precision, the intercept ES regression test by Bayer & Dimitriadis (2020) concludes that the extension of the Wied (2017) test can improve the ES forecasting of the plain CCC or DCC model while the plain models with or without the Duan & Wied (2018) test shows the similar performance in terms of the ES forecasting accuracy.

A Appendix

A.1 Algorithm

Algorithm 1 VaR and ES predictions based on the parametric Monte Carlo simulations

for $i = 1, \ldots, d$ **do**
  a) Specifying the innovation term $\eta_{i,t+1}$ in the GARCH(1,1) process as a particular distribution, e.g., $N(0, 1)$ or t_ν.
  b) Estimate the dynamic parameter vector $\boldsymbol{\theta}_i = (\alpha_{1,i}, \alpha_{2,i}\beta_{1,i})'$ in i-th variate, e.g. described in (2).
  c) Compute conditional mean $\hat{\mu}_{i,t+1}$ and conditional variance $\hat{\sigma}_{i,t+1}$ according to GARCH process parameterized by $\boldsymbol{\theta}_i$ with the help of equation (3).
  d) Compute the predicted i-th predicted return $\tilde{r}_{i,t+1}$ with the estimated $\hat{\mu}_{i,t+1}$, $\hat{\sigma}_{i,t+1}$.
end for
e) Determine the covariance matrix $\boldsymbol{\Sigma}_{t+1}$ based on CCC or DCC model and generate a large number B of simulated returns $\tilde{\boldsymbol{r}}_{t+1}$ based on the covariance matrix $\boldsymbol{\Sigma}_{t+1}$ and the mean vector $\boldsymbol{\mu}_{t+1}$.
f) The $t + 1$ predicted VaR at level α is $VaR_\alpha = \hat{F}_{t+1}(\alpha)^{-1}$, where $\hat{F}_{t+1}(\alpha)^{-1}$ is the empirical α-quantile of the distribution of the portfolio return $\tilde{r}_{p,t+1}$ being constructed on the simulated returns $\tilde{\boldsymbol{r}}_{t+1}$. The estimated ES is calculated as $ES_\alpha(\tilde{r}_{p,t+1}) = E(\tilde{r}_{p,t+1}|\tilde{r}_{p,t+1} \leq VaR_\alpha)$.

Algorithm 2 The detection of multiple change points in the correlation matrix proposed by Galeano & Wied (2017)

if $Q^W_{1,T} \leq c_{T,\alpha}$ **then**
  the algorithm stops, no change points are detected.
else a break in correlation matrix is announced, z_1 is the break point estimator
  while $\max_k\{Q^W_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)}, k = 1, \ldots, l+1\} > c_{T,\alpha}$, given l detected change points $z_1, \ldots, z_l$ in ascending order **do**
    A new change point is detected: $k_{\max} = \arg\max_k\{Q^W_{\eta(z_{k-1}+\frac{1}{T}),\tau(z_k)}, k = 1, \ldots, l\}$.
  end while
  if $l > 1$ **then**
    for $k = 1, \ldots, l$ **do**
      Calculate test statistics from sample from $\eta(z_{k-1} + \frac{1}{T})$ to $\tau(z_{k+1})$, delete the change points which are not statistically significant.
    end for
  end if
end if

Algorithm 3 Duan & Wied (2018)

for $i = 1, \ldots, d$ **do**

 a) Estimate single break point $\hat{\lambda}_i$ in variance of X_i, according to Wied, Arnold, Bissantz & Ziggel (2012).

 b) Estimate the vector of sample mean $\hat{\boldsymbol{\mu}}_{\hat{\lambda}_i,1}$ and $\hat{\boldsymbol{\mu}}_{\hat{\lambda}_i,2}$, the vector of sample variance $\hat{\boldsymbol{\sigma}}_{\hat{\lambda}_i,1}$ and $\hat{\boldsymbol{\sigma}}_{\hat{\lambda}_i,2}$ before and after the estimated break points from step a).

end for

c) Calculate residuals $\hat{\boldsymbol{Z}}_t$ according to (16) and estimate its covariance matrix $\hat{\Omega}$.

d) Calculate test statistics $\hat{\boldsymbol{Q}}_n$ according to (15).

e) Compute the p-value based on bootstrap approximation.

if time dependent sample **then**

 f) Use nonoverlapping block bootstrap $\mathcal{X}_b^*$.

else Draw B random samples with replacement with i.i.d. bootstrap $\mathcal{X}_b$.

end if

g) The bootstrap p-value is given by $p = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}_{\hat{\boldsymbol{Q}}_n \leq \boldsymbol{Q}_j^*}$.

A.2 Tables

Table 4: The results of Conditional Predictive Ability test for VaR forecasts

				Statistically significant results					Statistically significant results		
Model I	Model II	Model I preferred	Model II preferred	Model I preferred	Model II preferred	Indifferent	Model I preferred	Model II preferred	Model I preferred	Model II preferred	Indifferent
		Panel A: 5%-VaR					Panel B: 1%-VaR				
CCC	CCC + Wied 17 (single)	1	3	0	0	4	2	2	1	0	3
CCC	CCC + Wied 17 (multiple)	2	2	0	0	4	1	3	0	1	3
CCC	CCC + DW 18 (single)	2	2	0	0	4	3	1	1	0	3
CCC	CCC + DW 18 (multiple)	2	2	2	1	1	0	4	0	0	4
CCC	DCC	1	3	0	1	3	1	3	0	1	3
CCC	DCC + Wied 17 (single)	2	2	0	1	3	0	4	0	0	4
CCC	DCC + Wied 17 (multiple)	2	2	0	0	4	3	1	0	0	4
CCC	DCC + DW 18 (single)	2	2	1	0	3	1	3	0	0	4
CCC	DCC + DW 18 (multiple)	3	1	1	0	3	2	2	0	0	4
CCC + Wied 17 (single)	CCC + Wied 17 (multiple)	3	1	0	0	4	2	2	1	1	2
CCC + Wied 17 (single)	CCC + DW 18 (single)	3	1	1	0	3	2	2	0	1	3
CCC + Wied 17 (single)	CCC + DW 18 (multiple)	2	2	0	0	4	0	4	0	1	3
CCC + Wied 17 (multiple)	CCC + DW 18 (single)	1	3	0	1	3	4	0	2	1	1
CCC + Wied 17 (multiple)	CCC + DW 18 (multiple)	1	3	0	0	4	1	3	0	0	4
CCC + DW 18 (single)	CCC + DW 18 (multiple)	1	3	0	0	4	0	4	0	0	4
DCC	CCC + Wied 17 (single)	3	1	0	0	4	4	0	0	1	3
DCC	CCC + Wied 17 (multiple)	3	1	0	0	4	3	1	0	0	4
DCC	CCC + DW 18 (single)	4	0	0	0	4	3	1	0	0	4
DCC	CCC + DW 18 (multiple)	4	0	0	0	4	1	3	0	0	4
DCC	DCC + Wied 17 (single)	3	1	0	0	4	3	1	0	0	4
DCC	DCC + Wied 17 (multiple)	3	1	0	0	4	4	0	0	0	4
DCC	DCC + DW 18 (single)	2	2	0	0	4	3	1	0	0	4
DCC	DCC + DW 18 (multiple)	3	1	0	0	4	2	2	1	0	3
DCC + Wied 17 (single)	DCC + Wied 17 (multiple)	1	3	0	0	4	3	1	2	0	2
DCC + Wied 17 (single)	DCC + DW 18 (single)	1	3	0	0	4	1	3	0	1	3
DCC + Wied 17 (single)	DCC + DW 18 (multiple)	2	2	0	0	4	2	2	1	0	3
DCC + Wied 17 (multiple)	DCC + DW 18 (single)	1	3	1	0	3	2	2	0	1	3
DCC + Wied 17 (multiple)	DCC + DW 18 (multiple)	2	2	1	0	3	1	3	0	0	4
DCC + DW 18 (single)	DCC + DW 18 (multiple)	2	2	1	0	3	1	3	0	1	3

Note: this table presents the results of the comparison between the CCC and DCC models and their variates with the additional inclusion of structural break tests in four portfolios based on the CPA tests.

Table 5: The result of Conditional Predictive Ability test for the STOXX50 portfolio

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Panel A: 5%-VaR										
CCC no test (Model 1)		2	2	2	2	2	2	2	2	2
CCC wied17s (Model 2)			1	1	1	2	1	2	2	2
CCC wied17m (Model 3)				2	2	2	1	2	2	2
CCC dw18s (Model 4)					2	2	1	2	2	2
CCC dw18m (Model 5)						2	1	2	2	2
DCC no test (Model 6)							1	2	2	2
DCC wied17s (Model 7)								2	2	2
DCC wied17m (Model 8)									2	2
DCC dw18s (Model 9)										2
DCC dw18m (Model 10)										
Panel B: 1%-VaR										
CCC no test (Model 1)		2	2	2	2	2	2	2	2	2
CCC wied17s (Model 2)			1	1	2	2	1	1	1	2
CCC wied17m (Model 3)				1	2	2	1	1	1	2
CCC dw18s (Model 4)					2	2	1	2	1	2
CCC dw18m (Model 5)						1	1	1	1	2
DCC no test (Model 6)							1	1	1	2
DCC wied17s (Model 7)								2	2	2
DCC wied17m (Model 8)									1	2
DCC dw18s (Model 9)										2
DCC dw18m (Model 10)										

Note: this table presents the results of the comparison between the CCC and DCC models and their variates with the additional inclusion of structural break tests in STOXX portfolio based on the CPA tests. 1 denotes the row model outperforms the column model whereas 2 denotes the row model underperforms the column model.

Table 6: The result of Conditional Predictive Ability for the DAX30 portfolio

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					Panel A: 5%-VaR					
CCC no test (Model 1)		1	2	1	1	1	1	1	2	1
CCC wied17s (Model 2)			2	2	2	2	2	2	2	2
CCC wied17m (Model 3)				1	1	1	1	1	2	1
CCC dw18s (Model 4)					2	2	2	1	2	2
CCC dw18m (Model 5)						2	2	1	2	1
DCC no test (Model 6)							2	1	2	1
DCC wied17s (Model 7)								1	2	1
DCC wied17m (Model 8)									2	2
DCC dw18s (Model 9)										1
DCC dw18m (Model 10)										
					Panel A: 1%-VaR					
CCC no test (Model 1)		1	2	1	2	1	2	1	2	1
CCC wied17s (Model 2)			2	2	2	2	2	2	2	2
CCC wied17m (Model 3)				1	1	1	1	1	1	1
CCC dw18s (Model 4)					2	1	2	1	2	2
CCC dw18m (Model 5)						1	1	1	1	1
DCC no test (Model 6)							2	1	2	2
DCC wied17s (Model 7)								1	2	1
DCC wied17m (Model 8)									2	2
DCC dw18s (Model 9)										1
DCC dw18m (Model 10)										

Note: this table presents the results of the comparison between the CCC and DCC models and their variates with the additional inclusion of structural break tests in DAX portfolio based on the CPA tests. 1 denotes the row model outperforms the column model whereas 2 denotes the row model underperforms the column model.

Table 7: The result of Conditional Predictive Ability test for the FTSE100 portfolio

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					Panel A: 5%-VaR					
CCC no test (Model 1)		2	1	2	1	2	1	1	1	1
CCC wied17s (Model 2)			1	1	1	1	1	1	1	1
CCC wied17m (Model 3)				2	2	2	1	2	2	2
CCC dw18s (Model 4)					1	2	1	1	1	1
CCC dw18m (Model 5)						2	1	1	1	1
DCC no test (Model 6)							1	1	1	1
DCC wied17s (Model 7)								2	2	2
DCC wied17m (Model 8)									2	1
DCC dw18s (Model 9)										1
DCC dw18m (Model 10)										
					Panel A: 1%-VaR					
CCC no test (Model 1)		2	1	1	2	2	2	1	2	2
CCC wied17s (Model 2)			1	1	2	2	1	1	1	2
CCC wied17m (Model 3)				1	2	2	2	1	2	2
CCC dw18s (Model 4)					2	2	2	1	2	2
CCC dw18m (Model 5)						2	1	1	1	1
DCC no test (Model 6)							1	1	1	1
DCC wied17s (Model 7)								1	2	2
DCC wied17m (Model 8)									2	2
DCC dw18s (Model 9)										2
DCC dw18m (Model 10)										

Note: this table presents the results of the comparison between the CCC and DCC models and their variates with the additional inclusion of structural break tests in FTSE portfolio based on the CPA tests. 1 denotes the row model outperforms the column model whereas 2 denotes the row model underperforms the column model.

Table 8: The result of Conditional Predictive Ability for the S&P500 portfolio

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
					Panel A: 5%-VaR					
CCC no test (Model 1)		2	1	1	2	2	2	2	1	1
CCC wied17s (Model 2)			1	1	2	2	1	1	1	1
CCC wied17m (Model 3)				2	2	2	2	2	1	2
CCC dw18s (Model 4)					2	2	1	2	1	1
CCC dw18m (Model 5)						2	1	1	1	1
DCC no test (Model 6)							1	1	1	1
DCC wied17s (Model 7)								2	1	1
DCC wied17m (Model 8)									1	1
DCC dw18s (Model 9)										2
DCC dw18m (Model 10)										
					Panel A: 1%-VaR					
CCC no test (Model 1)		1	2	1	2	2	2	1	1	1
CCC wied17s (Model 2)			2	2	2	2	2	2	1	1
CCC wied17m (Model 3)				1	2	2	1	1	1	1
CCC dw18s (Model 4)					2	2	2	1	1	1
CCC dw18m (Model 5)						1	1	1	1	1
DCC no test (Model 6)							1	1	1	1
DCC wied17s (Model 7)								1	1	1
DCC wied17m (Model 8)									1	1
DCC dw18s (Model 9)										2
DCC dw18m (Model 10)										

Note: this table presents the results of the comparison between the CCC and DCC models and their variates with the additional inclusion of structural break tests in S&P portfolio based on the CPA tests. 1 denotes the row model outperforms the column model whereas 2 denotes the row model underperforms the column model.

References

Adams, Z., Füss, R. & Glück, T. (2017), 'Are correlations constant? Empirical and theoretical results on popular correlation models in finance', *Journal of Banking & Finance* **84**(C), 9–24.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Basel Committee (2010), Basel III: A global regulatory framework for more resilient banks and banking systems, Technical report, Bank for International Settlements.

Basel Committee (2016), Minimum Capital Requirements for Market Risk, Technical report, Bank for International Settlements.

Basel Committee (2017), Pillar 3 Disclosure Requirements – Consolidated and Enhanced Framework, Technical report, Basel Committee on Banking Supervisions.

Bayer, S. & Dimitriadis, T. (2020), 'Regression-Based Expected Shortfall Backtesting', *Journal of Financial Econometrics* . nbaa013.
URL: *https://doi.org/10.1093/jjfinec/nbaa013*

Berens, T., Weiß, G. N. & Wied, D. (2015), 'Testing for structural breaks in correlations: Does it improve Value-at-Risk forecasting?', *Journal of Empirical Finance* **32**(C), 135 – 152.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model', *The Review of Economics and Statistics* **72**(3), 498–505.

Christoffersen, P. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–62.

Christoffersen, P. (2003), *Elements of Financial Risk Management*, Academic Press, 1 edn, Elsevier Science.

Demetrescu, M. & Wied, D. (2018+), 'Residual-based inference on moment hypotheses, with an application to testing for constant correlation', *Econometrics Journal, forthcoming* .

Diebold, F. & Mariano, R. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–63.

Dimitriadis, T. & Bayer, S. (2019), 'A joint quantile and expected shortfall regression framework', *Electronic Journal of Statistics* **13**(1), 1823–1871.

Duan, F. & Wied, D. (2018), 'A residual-based multivariate constant correlation test', *Metrika* pp. 653–687.

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, number 57 *in* 'Monographs on Statistics and Applied Probability', Chapman & Hall/CRC.

Engle, R. F. & Manganelli, S. (2004), 'CAViaR: Conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**(4), 367–381.

Engle, R. F. & Sheppard, K. (2001), Theoretical and empirical properties of dynamic conditional correlation multivariate garch, Working Paper 8554, National Bureau of Economic Research.

Fissler, T. & Ziegel, J. F. (2016), 'Higher order elicitability and osband's principle', *Annals of Statistics* **44**(4), 1680–1707.

Fissler, T., Ziegel, J. F. & Gneiting, T. (2015), Expected shortfall is jointly elicitable with value at risk - implications for backtesting, Papers, arXiv.org.

Franke, J., Härdle, W. K. & Hafner, C. M. (2015), *Statistics of Financial Markets: An Introduction*, Universitext, 4 edn, Springer-Verlag.

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

Galeano, P. & Wied, D. (2017), 'Dating multiple change points in the correlation matrix', *TEST* **26**(2), 331–352.

Giacomini, R. & White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* **74**(6), 1545–1578.

Gneiting, T. (2011), 'Making and evaluating point forecasts', *Journal of the American Statistical Association* **106**(494), 746–762.

Kupiec, P. H. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The Journal of Derivatives* **3**(2), 73–84.

Longin, F. & Solnik, B. (1995), 'Is the correlation in international equity returns constant: 1960-1990?', *Journal of International Money and Finance* **14**(1), 3–26.

Löser, R., Wied, D. & Ziggel, D. (2018), 'New backtests for unconditional coverage of expected shortfall', *Journal of Risk* **21**, 1–21.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**(3-4), 271–300.

Mincer, J. A. & Zarnowitz, V. (1969), The Evaluation of Economic Forecasts, *in* 'Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance', NBER Chapters, National Bureau of Economic Research, Inc, pp. 3–46.

Nolde, N. & Ziegel, J. F. (2017), 'Elicitability and backtesting: Perspectives for banking regulation', *Annals of Applied Statistics* **11**(4), 1833–1874.

Patton, A. J., Ziegel, J. F. & Chen, R. (2019), 'Dynamic semiparametric models for expected shortfall (and Value-at-Risk)', *Journal of Econometrics* **211**(2), 388–413.

Santos, A., Nogales, F. J. & Ruiz, E. (2013), 'Comparing univariate and multivariate models to forecast portfolio value-at-risk', *Journal of Financial Econometrics* **11**(2), 400–441.

Weber, S. (2006), 'Distribution-invariant risk measures, information, and dynamic consistency', *Mathematical Finance* **16**(2), 419–441.

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

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

Wied, D., Krämer, W. & Dehling, H. (2012), 'Testing for a change in correlation at an unknown point in time using an extended functional delta method', *Econometric Theory* **28**(03), 570–589.