

This is the final peer-reviewed accepted manuscript published in ANNALS OF FINANCIAL ECONOMICS (World Scientific), EVALUATING THE ACCURACY OF TAIL RISK FORECASTS FOR SYSTEMIC RISK MEASUREMENT, Christian Brownlees, Giuseppe Cavaliere and Alice Monti, 2018

The final published version is available online at: <https://10.1142/S2010495218500094>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

Evaluating the Accuracy of Tail Risk Forecasts for Systemic Risk Measurement

Christian Brownlees, Giuseppe Cavaliere and Alice Monti

Abstract In this paper we address how to evaluate tail risk forecasts for systemic risk measurement. We propose two loss functions, the Tail Tick Loss and the Tail Mean Square Error, to evaluate, respectively, CoVaR and MES forecasts. We then analyse CoVaR and MES forecasts for a panel of top US financial institutions between 2000 and 2012 constructed using a set of bivariate DCC-GARCH-type models. The empirical results highlight the importance of using an appropriate loss function for the evaluation of such forecasts. Among other findings, the analysis confirms that the DCC-GJR specification provides accurate predictions for both CoVaR and MES, in particular for the riskiest group of institutions in the panel (Broker-Dealers).

Keywords: Systemic risk, Conditional Value-at-Risk, Marginal Expected Shortfall, DCC model, Forecasting, Forecast evaluation, Tick loss, Loss functions

1 Introduction

One of the lessons learnt from the 2007–2009 Global Financial Crisis (GFC) is the need for appropriate measures of systemic risk for the financial system. In the aftermath of the crisis several systemic risk measures have been proposed in the literature

Christian Brownlees

Pompeu Fabra University, Department of Economics and Business. Ramon Trias Fargas, 25-27, Barcelona, Spain, e-mail: christian.brownlees@upf.edu

Giuseppe Cavaliere

University of Bologna, Department of Economics. Piazza Scaravilli, 2, Bologna, Italy, e-mail: giuseppe.cavaliere@unibo.it

Alice Monti

UnipolSai Assicurazioni S.p.A., Risk Management Unit. Piazza Costituzione 2/2, Bologna, Italy, e-mail: alice.monti@unipolsai.it

by researchers, regulators and practitioners (e.g. Acharya et al., 2010; Adrian and Brunnermeier, 2011; Brownlees and Engle, 2017; Huang et al., 2009). A comprehensive survey of systemic risk measures and conceptual frameworks is presented in Bisias et al. (2012) and Benoit et al. (2016). The probability-distribution measures, that is quantile-based risk measures that focus on extreme losses, are widespread in the literature and largely applied by banks and financial institutions. They can be constructed using public market data (i.e. stock returns) and are based on statistical methods that require fairly weak assumptions on the data to obtain an estimate of a financial institution's contribution to systemic risk. Among the probability-distribution measures, the most widespread and well-known systemic risk measures are the ΔCoVaR (Adrian and Brunnermeier, 2011) and the SRISK (Brownlees and Engle, 2017) measures, which are constructed using, respectively, CoVaR and MES forecasts. These, in turn, are often constructed on the basis of DCC-GARCH models (Engle, 2002), in particular the DCC-GARCH model (e.g. Girardi and Ergun, 2013; Cabrera et al., 2014) and the DCC-GJR model (e.g. Cao, 2013; Popescu and Turcu, 2014; Yun and Moon, 2014; Brownlees and Engle, 2017; Engle et al., 2015).

In this paper we address how to evaluate tail risk forecasts for systemic risk measurement. In particular, we propose two loss functions, the Tail Tick Loss and the Tail Mean Square Error, to evaluate, respectively, CoVaR and MES forecasts. These loss functions are useful to compare the forecasting performance of different competing models and allow to determine which models produce accurate systemic risk measures.

We carry out a CoVaR and MES forecasting horse-race using a collection of DCC-GARCH specifications in order to determine which models produce the most accurate forecasts. We consider a panel comprising 91 top US financial institutions from 2000 to 2012. The set of GARCH specifications we consider includes the FIGARCH (Baillie et al., 1996), the Component-GARCH (Engle and Lee, 1999) as well as an asymmetric version of this model that we call Asymmetric-Component-GARCH that we introduce in this work. The empirical results show that the loss functions give useful guidance to identify the most successful specifications for predicting CoVaR and MES. In particular, results show that models that capture asymmetry and long-range dependence produce the best out-of-sample forecasts.

First this paper contributes to the literature on the econometric analysis of systemic risk. It also builds up on the literature on the evaluation of VaR forecasts; see, e.g., Christoffersen (1998), Engle and Manganelli (2004), Giacomini and Komunjer (2005). Our results extend and complement the ones found in, among others, Glosten et al. (1993), Engle and Lee (1999), Cheong et al. (2007), Engle and Rangel (2008).

The remainder of the paper is organized as follows. Section 2 introduces the CoVaR and MES measures. Section 3 presents the loss functions designed to evaluate CoVaR and MES forecasts. Section 4 contains the empirical application. Section 5 concludes.

2 CoVaR and MES measures

Given the returns r_t^i of an institution i and a confidence level α , the Value-at-Risk (VaR) risk measure $\text{VaR}_{\alpha,t}^i$ is defined as the α -quantile of the return distribution $\Pr_{t-1}[r_t^i \leq \text{VaR}_{\alpha,t}^i] = \alpha$. The Conditional-Value-at-Risk (CoVaR), firstly introduced by Adrian and Brunnermeier (2011), measures direct and indirect spillover effects in order to capture externalities that an individual institution imposes on the system. The CoVaR is defined as the VaR of the system returns r_t^s conditional on some event $\mathbb{C}(r_t^i)$ of institution i . More precisely, it is the α -quantile of the conditional probability distribution:

$$\Pr_{t-1} \left[r_t^s \leq \text{CoVaR}_{\alpha,t}^{s|i} \mid \mathbb{C}(r_t^i) \right] = \alpha$$

Adrian and Brunnermeier (2011) define the conditioning set event as $\mathbb{C}(r_t^i) = \{r_t^i = \text{VaR}_{\alpha,t}^i\}$. However, the latter CoVaR conditioning set does not consider severe losses which are further in the tail and does not allow CoVaR to be backtestable using standard procedures. As a consequence, Girardi and Ergun (2013) propose a conditioning financial distress event that refers to the institution i being at most at its VaR, i.e. $\mathbb{C}(r_t^i) = \{r_t^i \leq \text{VaR}_{\alpha,t}^i\}$:

$$\Pr_{t-1} \left[r_t^s \leq \text{CoVaR}_{\alpha,t}^{s|i} \mid r_t^i \leq \text{VaR}_{\alpha,t}^i \right] = \alpha \quad (1)$$

The CoVaR risk measure is the main component of the Δ -Conditional Value-at-Risk (ΔCoVaR) systemic risk measure which evaluates the systemic spillover of an individual institution to the system. Hence, the systemic risk contribution of an institution to the system is the (percentage) relative distance between the VaR of the financial system s conditional on the distressed state of institution i and the VaR of the financial system s conditional on the benchmark state of institution i :

$$\Delta\text{CoVaR}_{\alpha,t}^{s|i}(\%) = 100 \left[\frac{\text{CoVaR}_{\alpha,t}^{s|r_t^i \leq \text{VaR}_{\alpha,t}^i} - \text{CoVaR}_{\alpha,t}^{s|b^i}}{\text{CoVaR}_{\alpha,t}^{s|b^i}} \right]$$

where b^i indicates the benchmark state defined as a one-standard deviation about the mean event $\mu_t^i - \sigma_t^i \leq r_t^i \leq \mu_t^i + \sigma_t^i$ where μ_t^i and σ_t^i are, respectively, the conditional mean and the standard deviation of institution i . Due to time-varying correlations, the CoVaR of an institution here has a time-varying exposure to its VaR and this feature enables us to detect and incorporate in the systemic risk measurement possible changes over time in the linkage between the institution and the financial system.

The second systemic risk measure is the SRISK proposed by Brownlees and Engle (2017), whose main component is the Marginal Expected Shortfall (MES). MES, firstly introduced by Acharya et al. (2010), defines the systemic risk contribution as the expected return of an individual institution conditional on the system being distressed:

$$\text{MES}_{\alpha,t}^{i,s}(C) = \mathbb{E}_t [r_t^i | r_t^s \leq C] \quad (2)$$

The threshold value C , which represents the systemic event, may be set to $C = \text{VaR}_{\alpha,t}^s$, i.e. the VaR of the financial system. This risk measure is based on the Expected Shortfall (ES) defined as the expected loss of the system conditional on the loss being greater than the VaR calculated at a given level of confidence $1 - \alpha$. SRISK is an alternative dynamic reduced form estimate of capital shortages to measure the systemic risk contribution of a financial firm. Brownlees and Engle (2017) define SRISK of institution i on day t as the prediction of the capital shortfall when system declines below a threshold C over a time horizon h :

$$\text{SRISK}_t^i = \mathbb{E}_t [\text{CS}_{t+h}^i | r_{t+1:t+h}^s < C] = W_t^i (k\text{LVG}_t^i - (1-k)\text{LRMES}_t^i - 1)$$

where

- CS_{t+h}^i is the capital shortfall of firm i over a time horizon h , defined as:

$$\text{CS}_t^i = kA_t^i - W_t^i = k(D_t^i + W_t^i) - W_t^i$$

where W_t^i is the market value of equity, D_t^i is the book value of debt, A_t^i is the value of quasi assets and k is the prudential capital fraction;

- LVG_t^i denotes the quasi-leverage ratio $(D_t^i + W_t^i)/W_t^i$;
- LRMES_t^i is Long Run MES, defined as the expectation of the firm equity multi-period return conditional on the systemic event:

$$\text{LRMES}_t^i = \mathbb{E}_t [r_{t+1:t+h}^i | r_{t+1:t+h}^s < C]$$

where $r_{t+1:t+h}$ is the multi-period equity return between period $t + 1$ and $t + h$, of firm and system, respectively.

3 Measuring forecast accuracy

3.1 Background

The accuracy and efficiency tests have been primarily developed in the literature to evaluate VaR models. The assessment of the performance of the VaR forecasts is usually carried out on the basis of a two-step procedure (see, Brownlees and Gallo, 2008). The first step assesses the adequacy of the VaR forecasting methods using VaR specification tests (Kupiec, 1995; Christoffersen, 1998). The second step assesses the accuracy of those forecasting methods not rejected by the backtests using a VaR loss function that measures the VaR closeness to their nominal coverage. Regarding the first step, the most common adequacy tests are the unconditional coverage test of Kupiec (1995) and the conditional coverage test of Christoffersen (1998), where the null hypothesis is rejected if the VaR model considered generates

too many or too few or too clustered exceptions. CoVaR is a VaR-based measure, hence it is possible to adapt these tests, used for VaR, to the CoVaR framework using appropriate modifications (see Girard and Ergun, 2013). Regarding the second step, a widely used loss function in assessing VaR accuracy is the tick loss function proposed by Komunjer (2005) (see Brownlees and Gallo, 2008).

Considering the VaR violation indicator I_t^i of an institution i at time $t \in \mathcal{T} = \{1, \dots, T\}$ ($I_t^i = 1$ when $r_t^i \leq \text{VaR}_{\alpha,t}^i$ and 0 otherwise), the CoVaR violation indicator $I_t^{s|i}$ at time $t \in \mathcal{T}_i$ is defined as:

$$I_t^{s|i} = \begin{cases} 1 & \text{if } r_t^s \leq \text{CoVaR}_{\alpha,t}^{s|i} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\mathcal{T}_i \subset \mathcal{T}$ denotes the set of time periods when the VaR violations of institution i occur (i.e. when the institution i is in financial distress), in particular when the VaR violations sequence I_{t+1}^i gives 1 as result. According to Christoffersen (1998), the problem of determining the adequacy of CoVaR can be reduced to the problem of determining whether the resulting CoVaR violation passes the unconditional coverage (UC) test (Kupiec, 1995) and the independence test. The former examines whether the number of exceptions over a specific number of observations in the backtesting window is consistent with the corresponding confidence level (the percentage of violations should not significantly differ from $\alpha \times 100\%$). The latter examines whether the probability of an exception on any day depends on the outcome of the previous day. Only a CoVaR violations sequence that satisfies both these properties can be described as evidence of an adequate CoVaR model. Another test to verify the validity of the forecasting model is the dynamic quantile test proposed by Engle and Manganelli (2004) which examines whether the CoVaR violations sequence is an iid sequence of Bernoulli random variables with probability α .

The information contained in the CoVaR violations sequence refers only to whether or not an exceedance occurred, and does not provide the magnitude of the exceedance. Statistical adequacy is, in fact, a necessary requirement that CoVaR forecasts must satisfy, but it does not provide information on the accuracy of such predictions and does not always help to discriminate among different CoVaR forecasting methods. Hence, the assessment of the CoVaR forecasting methods adequacy is not sufficient to assess the CoVaR forecasts performance. Moreover, as anticipated in the introduction section, these tests have low power, as pointed out by Kupiec (1995), Berkowitz (2001), Escanciano and Pei (2012). As a consequence, using a CoVaR loss function is necessary to assess the accuracy of the forecasting methods (see for VaR framework Brownlees and Gallo, 2008; Escanciano and Pei, 2012). Loss functions, in fact, may be extremely useful for determining whether a model provides a better risk assessment than another competing model, and more suited to discriminate among different competing models and to judge the accuracy of a single model. Loss functions could be tailored to address specific concerns about the evaluation of accuracy of systemic risk forecasts. Since CoVaR is a

VaR-based measure, the VaR loss functions used in the literature can be adapted to CoVaR framework, as for VaR backtests.

3.2 A loss functions for CoVaR

As pointed out, the adoption of a more fine-grained loss function is fundamental to help in assessing the CoVaR accuracy and selecting the proper model, since backtests do not discriminate among different competing models. The CoVaR measure is defined as a conditional quantile, hence a straightforward loss function suitable to this framework is the tick loss function, introduced by Komunjer (2005). Komunjer, in fact, develops an approach to conditional quantile estimation based on quasi-maximum likelihood. Hence, we propose a modified version of the tick loss function called the Tail Tick Loss (TTL) function defined as:

$$\text{TTL}_{\alpha}^{s|i} = \frac{1}{N_i} \sum_{t \in \mathcal{T}_i} (\alpha - I_{(r_t^s \leq \eta_t)}) (r_t^s - \eta_t), \quad (4)$$

where

- $\eta_t = \text{CoVaR}_{\alpha,t}^{s|i}$ is the optimal predictor of TTL function;
- \mathcal{T}_i is the time set when the conditioning event of CoVaR occurs, in particular when $\mathbb{C}(r_t^i) = \{r_t^i \leq \text{VaR}_{\alpha,t}^i\}$ holds (see eq. (1)). The letter i indicates the dependence of the time set by the VaR of institution i ;
- $N_i = \#(\mathcal{T}_i)$ is the time set sample size, in particular the number of observations when the financial institution is in financial distress;
- $I_{(r_t^s \leq \eta_t)} \equiv I_t^{s|i}$ is the CoVaR violations sequence ($I_t^{s|i} = 1$ when $r_t^s \leq \text{CoVaR}_{\alpha,t}^{s|i}$ and 0 otherwise). It is important to underline that this sequence is constructed on the VaR violations sequence, as highlighted by $s|i$. Consequently, it is defined only on the time set \mathcal{T}_i , in particular when the VaR violations of institution i occur.

Hence, in order to correctly evaluate CoVaR forecast accuracy, we apply to the tick loss function the same conditioning set of the CoVaR measure. More precisely, we condition the tick loss function on the \mathcal{T}_i time periods set, where the conditioning event occurs $\{r_t^i \leq \text{VaR}_{\alpha,t}^i\}$, instead of the overall time period \mathcal{T} as in the standard tick loss function.

The proposed Tail Tick Loss function is an appropriate loss function to evaluate the CoVaR accuracy. The CoVaR measure, in fact, is the optimal predictor associated to the TTL loss function since it minimizes the expected value of the loss function:

$$\text{CoVaR}_{\alpha,t}^{s|i} = \underset{\eta_t \in M_{t-1}}{\text{argmin}} \mathbb{E}_{t-1} \left[(\alpha - I_{(r_t^s \leq \eta_t)}) (r_t^s - \eta_t) \mid \mathbb{C}(r_t^i), \mathcal{F}_{t-1} \right]$$

where $M_{t-1} \subset \mathbb{R}$, $\mathbb{C}(r_t^i) = \{r_t^i \leq \text{VaR}_{\alpha,t}^i\}$ and \mathcal{F}_{t-1} is the information set available until time $t-1$. The focus is particularly on the extreme events, because the financial institution failure can occur more probably on these days.

Comparing the values of the average sample TTL function for the same institution i among all competing models used to compute the CoVaR measure, the most accurate model for CoVaR corresponds to the smallest average sample loss reached. Thus, we are able to identify which is the model that predicts more accurately the CoVaR.

3.3 A loss functions for MES

To the best of our knowledge, statistical tools with the purpose to test and compare MES forecasts have not been properly developed in the literature and a deep analysis on their accuracy is largely unexplored. The Mean Square Error loss function is widely used in the literature to assess the forecasting accuracy. MES is a conditional tail expectation, therefore in order to evaluate MES accuracy and its forecasting ability, we propose the Tail Mean Square Error (TMSE) defined as:

$$\text{TMSE}_{\alpha}^{i|s} = \frac{1}{N_s} \sum_{t \in \mathcal{T}_s} \left[\frac{r_t^i - \eta_t}{\sigma_t^s} \right]^2$$

where

- $\eta_t = \text{MES}_{\alpha,t}^{i|s}$ is the optimal predictor of TMSE function;
- \mathcal{T}_s is the time set when the conditioning event of MES occurs, in particular when $\{r_t^s \leq \text{VaR}_{\alpha,t}^s\}$ holds (see eq. (2)). The letter s indicates the dependence of the time set by the VaR of system s ;
- $N_s = \#(\mathcal{T}_s)$ is the time set sample size, in particular the number of observations when the system is in financial distress;
- σ_t^s is the standard deviation of the financial system estimated by the GARCH(1, 1) model for all competing models.

As done for CoVaR framework, we have to condition TMSE on the same conditioning set of MES, i.e. $\{r_t^s \leq \text{VaR}_{\alpha,t}^s\}$. More precisely, in order to guarantee a correct evaluation of MES accuracy, we condition TMSE on the N_s realizations of the observed financial system VaR, instead of the overall time period \mathcal{T} as in the standard TMSE. Furthermore, we standardize TMSE values by σ_t^s , with the purpose of avoiding volatile periods influences on the entire value.

The proposed Tail Mean Square Error is an appropriate loss function to evaluate the MES accuracy. The MES measure, in fact, is the optimal predictor of this TMSE loss function since it minimizes the expected value of the loss function:

$$\text{MES}_{\alpha,t}^{i|s} = \underset{\eta_t \in M_{t-1}}{\text{argmin}} \mathbb{E}_{t-1} \left[\left(\frac{r_t^i - \eta_t}{\sigma_t^s} \right)^2 \middle| \mathbb{C}(r_t^s), \mathcal{F}_{t-1} \right]$$

where $M_{t-1} \subset \mathbb{R}$, $\mathbb{C}(r_t^s) = \{r_t^s \leq \text{VaR}_{\alpha,t}^s\}$ and \mathcal{F}_{t-1} is the information set available until time $t - 1$. The focus is particularly on the extreme events, because the financial institution failure can occur more probably on these days.

Comparing the TMSE values for the same institution i among all models, we are able to detect the model that forecasts more precisely MES.

4 Empirical analysis of the 2007-2009 financial crisis

In this section we provide an empirical analysis of 2007 – 2009 financial crisis according to the different systemic risk approaches presented in Appendix A. In particular we analyze the accuracy of the standard DCC-GARCH-type models that are present in the literature in measuring systemic risk using the novel loss functions proposed in Section 3. We also investigate how the results should be interpreted in order to detect the best prediction models.

4.1 Data

The sample is composed by the daily returns of 91 US financial institutions, with a market capitalization greater than 5 bln USD as of end of June 2007, the same panel used by Brownlees and Engle (2017). We exclude 4 institutions from systemic risk analysis due to the limited length of their in-sample return series used for the models estimation: Mastercard (MA), NYMEX (NMX), NYSE Euronext (NYX), Western Union (WU). At least 100 observations are, in fact, required to estimate a GARCH-type model. The daily CRSP market value weighted index is used as a proxy for the system. We obtained data from CRSP. The panel spans from January 3, 2000 to December 31, 2012 for a total of $T = 3269$ observations. The sample is unbalanced since not all companies have been trading continuously during the sample period. The financial institutions are grouped by financial industry group based on their SIC codes¹. The 4 subindustry groups are: Depositories Institutions (Dep.), that contains banks and counts 28 institutions; Insurance (Ins.), that contains insurance companies and counts 34 institutions; Securities Dealers and Commodity Brokers (Bro.), that contains, for example, Bear Stearns and Lehman Brothers, and counts 9 institutions; Others (Oth.), that contains non-depository institutions and real estate, and counts 20 institutions.

The sample is divided into two different sub-samples in order to allow the out-of-sample forecasting validation and to evaluate the risk measures performances during the forecasting period. In particular, the in-sample period consists of 1611 daily observations from January 3, 2000 to May 31, 2006; while the out-of-sample consists of 1658 daily observations from June 1, 2006 to December 31, 2012. The

¹ The only exception is Goldman Sachs (GS) which is included in Broker-Dealers, instead of Others.

out-of-sample, which includes the 2007 – 2009 financial crisis, is used to forecast and validate systemic risk measures. It is interesting to analyze the results obtained in different out-of-sample sub-periods comparing, for example, the period before the financial crisis, the period during the crisis, and the period immediately afterwards. During the forecasting period the models are separately re-estimated every week (i.e. every 5 observations) using all data available until as of that date and the forecasts are computed one-step-ahead. We exclude the year before the 2007 – 2009 financial crisis from the in-sample period to permit the models to fit the data without any preliminary information about the crisis. Moreover, the re-estimation done every-week allows the models to fit the data considering the new information set, and the parameters are not fixed to the last estimation at the end of in-sample.

4.2 *In-sample estimation*

We compute the daily CoVaR and MES measures at $\alpha = 5\%$ confidence level assuming Gaussian distribution for all the financial institutions in the panel employing the GARCH-type models, illustrated in Appendix A, over the in-sample period (from 03/01/2000 to 31/05/2006). In particular, we first estimate the volatilities for each financial institution by using the univariate GARCH-type models, then we estimate the correlations of each institution-system pair by using the bivariate DCC model.

Table 2 gives a summary of the DCC-ACGARCH parameters for each sub-industry group over the in-sample. In particular, the 10%, 50% and 90% quantiles values of the obtained parameters estimates are reported to have an idea about the distribution of the parameters across the financial groups under investigation. We notice that the dynamics of the financial institutions in the panel do not have a strong degree of heterogeneity and the parameters estimates are very similar across groups. Columns 4, 5 and 8 in table 2 (i.e. α, β, γ) are dedicated to the parameters estimates of the short-run component of the ACGARCH model, while columns 3, 6 and 7 (i.e. ω, ρ, φ) display the parameters estimates of the long-run component. On the contrary, the last three columns (i.e. const, a, b) show the parameters estimates of the DCC model. The ACGARCH parameters estimates are in line with those that can be found in the related literature. In particular, as expected, the transitory shocks are much less persistent than the permanent shocks ($1 > \rho > \alpha + \beta$), and the impact of transitory shocks is much greater than permanent shocks ($\alpha > \varphi$). The β estimates are on average higher for Broker-Dealers, showing higher persistence. Overall, the parameters estimates do not fluctuate much. Focusing on DCC, parameters are in line with the typical set of estimates (i.e. a is very smaller than b and their sum is close to the unit value) and are similar across groups. Then, we evaluate the volatility estimates obtained by different models using the robust loss functions proposed by Patton (2011), and considering the financial institution squared returns as volatility proxy.

Once estimated volatilities, we compute the CoVaR measure and we backtest CoVaR for verifying the unconditional coverage property employing the UC test by Kupiec (1995) and Christoffersen (1998), according to Girardi and Ergun (2013). We obtain that the UC null hypothesis is rejected at 5% significance level for 88 financial institutions by DCC-GARCH, 87 financial institutions by DCC-GJR, 90 financial institutions by the DCC-CGARCH, 90 financial institutions by the DCC-FIGARCH, 81 financial institutions by the DCC-ACGARCH, over 91 financial institutions in total. In particular, the models satisfy the unconditional coverage property providing adequate CoVaR measures. However, DCC-GARCH-type models tend to capture the same number of VaR and CoVaR violations, hence the UC test provides similar results from different models. We find, in fact, that 34 financial institutions over 91 have the same test statistic value for both DCC-GARCH and DCC-GJR model. This evidence confirms our idea that backtests do not help in discriminating among different competing models, and it is necessary to employ appropriate loss functions in addition to backtests in order to detect the best model.

In addition, we conduct the dynamic quantile (DQ) test by Engle and Manganelli (2014) that aims at verifying whether the model is misspecified. Some financial institutions do not have enough past information to carry out the DQ test, therefore 13 institutions are excluded. The null hypothesis is rejected at 5% significance level for 73 financial institutions by the DCC-GARCH, 70 financial institutions by the DCC-GJR, 74 financial institutions by the DCC-CGARCH, 71 financial institutions by the DCC-FIGARCH model, 64 financial institutions by the DCC-ACGARCH model, over 78 financial institutions in total. Therefore, the DCC-GARCH-type models are not misspecified. Unfortunately, we obtain similar test statistic values for many firms, as well as for the UC test. In particular, 19 financial institutions have the same number of VaR and CoVaR failures, and consequently the same DQ test statistic values. Hence, it is not possible to determine the best model using the DQ results and it is necessary to employ appropriate and more fine-grained loss functions to achieve this goal.

4.3 Out-of-sample forecast accuracy

The out-of-sample period spans from June 2006 to December 2012, including the 2007 – 2009 financial crisis, for a total of 1658 observations and is used to forecast and validate the systemic risk measures. We carry out the comparisons among the volatility predictions obtained by different forecasting methods. In particular, we compare the performances of the DCC-GARCH-type models, outlined in Appendix A, and the benchmark models ones. For comparative purposes and based on their widespread usage in the related literature, we use the rolling volatility (RollVol.) as benchmark for the volatility, the rolling quantile regression model (QR) as benchmark for the CoVaR measure, and the rolling linear regression (LR) model as benchmark for the MES measure. The rolling volatility is computed as the standard deviation of its return time series moving a window of 500 observations forward

one by one for each financial institution. The same rolling window of 500 observations is used for both the rolling quantile regression and the linear regression models. Hence, we evaluate the volatility forecasts with the aim of identifying the best prediction model for volatility, then we compute the models accuracy in forecasting the systemic risk measures, CoVaR and MES. In particular, the comparison of models performances is carried out evaluating the accuracy of the volatility, CoVaR and MES measures using the robust loss functions indicated by Patton (2011) for the volatility, the TTL for the CoVaR and the TMSE loss function for the MES measure.

The accuracy of the CoVaR forecasts, evaluated using the TTL function, is presented in table 1, which reports the TTL averages divided by financial industry groups. The bold values indicate the best prediction models. The TTL results summarize the models performances in forecasting CoVaR, identifying the DCC-FIGARCH model as the model that predicts the most accurate CoVaR measure (see table 1). In addition to the loss functions, it is important to introduce techniques that are able to discriminate among different competing models in terms of significance. We therefore consider the Diebold-Mariano test to check the statistical significance between the best model and the benchmark one. This means that we compare each DCC-GARCH-type model with the rolling quantile regression (QR) for each financial institution individually comparing the whole TTL series². We observe that all the DCC-GARCH-type models are statistically different at 1% significance level from the quantile regression approach for the entire sample. Furthermore, we run the Diebold-Mariano test between the new proposed model, DCC-ACGARCH, and the best model for each financial institution individually, finding that the difference between the TTL of DCC-ACGARCH and the TTL value of the best model is statistically significant at 10% level for 65 firms over 91.

Similarly, the MES predictions accuracy is evaluated using the TMSE loss function. Table 3 shows the related results grouped by financial sectors, where bold values correspond to the lowest values indicating the best prediction models. From table 3, the TMSE results do not indicate a unique best prediction model for all the sub-industry groups. Overall, the DCC-GARCH-type models outperform the rolling linear regression (LR) and the best prediction model for MES is, on average, the DCC-GARCH. The Diebold-Mariano test between the TMSE values of the best model and LR (see table 3) shows that the linear regression model is statistically worse than the best one at 10% significance level for 40 financial institutions over 91. On the contrary, 29 firms show that the TMSE values are statistically significant at the 10% level for the new proposed model, DCC-ACGARCH, over the best prediction model. It is important to notice that the novel ACGARCH model, introduced in Appendix A.1, is an extension of the GARCH, GJR and CGARCH models and all these models are the best ones in some sub-industry sector (see table 3). Therefore, it is necessary to estimate and forecast MES by the DCC-ACGARCH model and then

² We run the Diebold-Mariano test only once for the entire sample (composed by 91 financial institutions) to avoid multiple testing problems. The multiple testing problem results from the increase in type I error that occurs when statistical tests are used repeatedly.

to check the significance of the parameters in order to reach the most appropriate nested model.

Since the TMSE does not identify a unique best prediction model for MES, we investigate other versions of TMSE. The proposed TMSE is, in fact, divided by the standard deviation of the financial system estimated by GARCH for all competing models used in the empirical analysis in order to avoid the influences of volatile periods. Using the standard deviation estimated by the other GARCH-type models as standardization, we obtain similar results to those reported in table 3. In particular, the DCC-GARCH model provides overall the most accurate MES forecasts for all the cases. However, if we standardize the TMSE by the standard deviation estimated by each considered model³, we change the standardization according to the considered model and we obtain different results. In this way, we are able to avoid that volatile periods captured by that particular model influence the related loss function, instead of avoiding that volatile periods captured by a particular model (in our case GARCH) influence all the loss functions. These results identify the DCC-FIGARCH as the best forecasting model for the MES measure for the entire sample and each sub-industry group. This finding is in line with the TTL results, reported in table 1, which indicate DCC-FIGARCH as the best prediction model for systemic risk, confirming the importance of capturing the long-range dependence.

4.4 Out-of-sample systemic risk

In order to check more precisely the ranking of groups' riskiness, after a graphical analysis, we compute the average of each measure. Table 4 contains the ranking of the CoVaR averages over the out-of-sample. From table 4, we notice that Broker-Dealers is always the most risky group, followed by Others, Insurance and Depositories groups, and this groups' riskiness ranking is identified by all the considered models. Moreover, the model that provides the overall lowest average of the CoVaR measure is the DCC-FIGARCH that corresponds to the best prediction model for this measure (see table 4).

Table 5 contains the ranking of the MES averages over the out-of-sample (2006 – 2012). All the models identify the same ranking composed by Broker-Dealers, Depositories, Others, and Insurance respectively. The ranking obtained for the MES measure, reported in table 5, is different from that obtained for CoVaR, reported in table 4. Finally, the model that provides the overall lowest average of the MES measure in table 5 is the DCC-CGARCH.

³ Standardizing the TMSE loss function by the standard deviation estimated by the particular considered model means that the TMSE obtained by DCC-CGARCH is divided by the system standard deviation estimated by the CGARCH, the TMSE obtained by DCC-FIGARCH is divided by the system standard deviation estimated by the FIGARCH, and so on.

4.5 Sub-samples comparison

The out-of-sample period is divided into three different sub-samples in order to compare the systemic risk measures obtained in different historical periods: the pre-crisis sample (251 daily observations, from June 2006 to May 2007), the crisis sample (462 daily observations, from June 2007 to March 2009), and the post-crisis sample (945 daily observations, from April 2009 to December 2012).

We carry out the comparison among the forecasting performances of the DCC-GARCH-type models in each sub-sample period and we obtain different best prediction models for each measure. For the CoVaR measure, the TTL values suggest the DCC-FIGARCH model as the best prediction model over the pre-crisis and the crisis samples, while the DCC-GJR provides the most accurate CoVaR forecasts after the crisis to the end of the sample among the considered competing models. Finally, for the MES forecasts, the DCC-GARCH outperforms the other models during the pre-crisis and crisis periods, while the DCC-FIGARCH is the best MES prediction model over the post-crisis sample. It is interesting to observe that for the first two sub-samples the best prediction model is the same, whereas after the crisis it is a different model. Note that since the suggested models are nested into the ACGARCH, the latter is a useful starting point in order to find the best possible model specification. This could be due to the fact that before and during the crisis the volatilities and the systemic risk are much higher than during the post-crisis period, hence two different model specifications are necessary.

Table 6 reports the averages of the CoVaR measure obtained by the different models over three different sub-samples. From table 6 it is interesting to observe that the most risky group over all periods is Broker-Dealers. During the pre-crisis sample Broker-Dealers are followed by Depositories, Insurance and Others respectively (as found in Girardi and Ergun, 2013), and this ranking is uniformly identified by all the models (see table 6). As well as during the post-crisis period where Broker-Dealers are followed by Others, Insurance and Depositories by all the models respectively. On the contrary, during the crisis sample in table 6, each model determines a different ranking. This fact can be explained by the fact that during the financial crisis the contagion among financial institutions is very high, hence it is not possible to achieve a well-defined ranking, as all the CoVaR values are very close each other.

Table 7 presents the comparison of the MES averages over the different sub-samples obtained by all the competing models. From table 7 the most risky group is Broker-Dealers over all the considered periods, while Insurance is the group with the lowest risk. During the pre-crisis sample Broker-Dealers are followed by Others, and Depositories, while over the crisis and the post-crisis periods Depositories and Others are reversed. All the models identify the same ranking based on the groups' riskiness (see table 7).

5 Conclusions

In the aftermath of the crisis, the measurement of systemic risk has become a topic of active research. The availability of so many different systemic risk measures and approaches highlights the need to identify which measure and model predict systemic risk more accurately. Statistical tools to test and compare systemic risk forecasts, however, have not been properly developed and a deep analysis on their accuracy is largely unexplored in the existing literature. In this work we have evaluated CoVaR adequacy by VaR-based backtests, and, as expected, we have pointed out that CoVaR backtests do not discriminate among different competing models used to compute CoVaR. These backtests, obtained by extending those used for backtesting VaR, generally provide the same results across all models and have low power. Therefore, employing appropriate loss functions specifically designed to assess the accuracy of the systemic risk forecasts and the forecasting ability of the considered models is necessary. The need of developing loss functions to assess the systemic risk accuracy leads us to the proposal of two loss functions, the Tail Tick Loss (TTL) and the Tail Mean Square Error (TMSE) loss functions, which are suitable for the CoVaR and MES frameworks, respectively. These loss functions are helpful in discriminating among different competing models in order to determine the models that forecasts more precisely CoVaR and MES measures, respectively. In this way, we are able to evaluate the forecasting performances of CoVaR and MES.

An empirical analysis of the 2007 – 2009 financial crisis, carried out on 91 US financial institutions daily return series on the time interval which spans from January 2000 to December 2012, confirms the need of a more fine-grained loss function with the aim of evaluating the forecasts accuracy. The objective of the comprehensive and exhaustive comparison of different bivariate volatility models for forecasting systemic risk is to evaluate the accuracy and performances of these models in forecasting CoVaR and MES in order to identify the best model which provides more accurately systemic risk forecasts. All DCC-GARCH-type models fit the data satisfactorily with a high statistical significance of the parameters and outperform significantly the benchmark models, highlighting the presence of long-range dependence. We have applied, in fact, not only the econometric standard models, but also long-range dependence models with the aim of taking into account other stylized facts, so far not considered in systemic risk framework. In particular, we have introduced a comprehensive model, called the Asymmetric-Component-GARCH (ACGARCH), able to capture the leverage effect jointly with long-range dependence. The proposed DCC-ACGARCH model provides good in- and out-of-sample forecasts of systemic risk and outperforms significantly benchmarks (i.e. quantile regression and linear regression models) according to the Diebold-Mariano statistical test in estimating and forecasting the CoVaR and MES measures. In particular, ACGARCH improves systemic risk forecasts and its results are comparable with those obtained by other existing models. Moreover, there is empirical evidence that both leverage effect and long-range dependence should be considered in measuring and forecasting systemic risk. CoVaR backtests provide quite similar results across all the models, since the CoVaR failures tend to occur on the same days. The CoVaR backtesting results are

very similar among different competing models, confirming the need to employ appropriate loss functions, in addition to the backtests, with the aim of discriminating among the competing models for achieving more accurate forecasts. The TTL loss function, which is employed to evaluate the models performances in forecasting CoVaR, indicates that the DCC-FIGARCH model predicts more accurately the CoVaR and the long-memory models are preferable to the standard ones according to the Diebold-Mariano statistical test. For the MES measure, we have applied the proposed TMSE loss function to assess the models performances in forecasting MES without finding a unique preferred best model. The DCC-GARCH is, in fact, the overall best prediction model for MES, but not for all the sub-industry groups. It is important to notice that all the models indicated for each group are special cases of the proposed ACGARCH model. Therefore, estimating ACGARCH is useful to identify the most appropriate forecasting model. Investigating other versions of the TMSE, it is interesting to observe that dividing the TMSE by the standard deviation estimated by each considered model, the DCC-FIGARCH is found as the best MES predicting model as well as for the CoVaR. Moreover, our empirical analysis results are consistent with those presented in the existing literature by some authors, for example Adrian and Brunnermeier (2011), Girardi and Ergun (2013).

References

1. Acharya V., Pedersen L., Philippon T., Richardson M. (2010). *Measuring Systemic Risk*. Working Paper, New York University.
2. Adrian T., Brunnermeier M. (2011). *COVAR*. Working Paper, Federal Reserve Bank of New York.
3. Baillie R.T., Bollerslev T., Mikkelsen H.O. (1996). *Fractionally integrated generalized autoregressive conditional heteroskedasticity*. *Journal of econometrics*, 74(1), 3 – 30.
4. Benoit S., Colletaz G., Hurlin C., Prignon C. (2013). *A theoretical and empirical comparison of systemic risk measures*. HEC Paris Research Paper No. FIN-2014 – 1030.
5. Berkowitz J. (2001). *Testing density forecasts, with applications to risk management*. *Journal of Business and Economic Statistics*, 19(4), 465 – 474.
6. Bisias D., Flood M., Lo A., Valavanis S. (2012). *A Survey of Systemic Risk Analytics*. Working Paper of the Office of Financial Research, U.S. Department of Treasury.
7. Bollerslev T. (1986). *Generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics*, 31, 307 – 327.
8. Brownlees C.T., Engle R.F. (2017). *SRISK: A conditional capital shortfall measure of systemic risk*. *The Review of Financial Studies*, 30(1), 48 – 79.
9. Brownlees C.T., Gallo G.M. (2008). *On variable selection for volatility forecasting: The role of focused selection criteria*. *Journal of Financial Econometrics*, 6(4), 513 – 539.
10. Cabrera W., Hurtado J., Morales M., Rojas J.S. (2014). *A Composite Indicator of Systemic Stress (CISS) for Colombia*. Borradores de Economía, (826).
11. Cao Z. (2013). *Multi-CoVaR and Shapley Value: A Systemic Risk Measure*. Manchester University.
12. Cheong C.W., Nor A.H.S.M., Isa Z. (2007). *Asymmetry and long-memory volatility: Some empirical evidence using GARCH*. *Physica A: Statistical Mechanics and its Applications*, 373, 651 – 664.
13. Christoffersen P.F. (1998). *Evaluating interval forecasts*. *International economic review*, 841 – 862.

14. Christoffersen P., Jacobs K., Ornathanalai C., Yintian Wang Y. (2008). *Option valuation with long-run and short-run volatility components*. Journal of Financial Economics, 90, 272 – 297.
15. Engle R.F., Lee G. (1999). *A long-run and short-run component model of stock return volatility*. In: Engle R., White H. (Eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive WJ Granger*. Oxford University Press, 475 – 497.
16. Engle R.F. (2002). *Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models*. Journal of Business and Economic Statistics, 20(3), 339 – 350.
17. Engle R.F., Manganelli S. (2004). *CAViaR: Conditional autoregressive value at risk by regression quantiles*. Journal of Business and Economic Statistics, 22(4), 367 – 381.
18. Engle R.F., Rangel J. (2008). *The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes*. Review of Financial Studies, 21(3), 1187 – 1222.
19. Engle R.F., Jondeau E., Rockinger M. (2015). *Systemic Risk in Europe*. Review of Finance, 19, 145 – 190.
20. Escanciano J.C., Pei P. (2012). *Pitfalls in backtesting historical simulation VaR models*. Journal of Banking and Finance, 36(8), 2233 – 2244.
21. Gallant A.R., Rossi P.E., Tauchen G. (1993). *Nonlinear dynamic structures*. Econometrica: Journal of the Econometric Society, 871 – 907.
22. Giacomini R., Komunjer I. (2005). *Evaluation and combination of conditional quantile forecasts*. Journal of Business and Economic Statistics, 23(4), 416 – 431.
23. Girardi G., Ergun A. (2013). *Systemic Risk Measurement: Multivariate GARCH Estimation of CoVaR*. Journal of Banking and Finance, 37, 3169 – 3180.
24. Glosten L.R., Jagannathan R., Runkle D.E. (1993). *On the relation between the expected value and the volatility of the nominal excess return on stocks*. The journal of finance, 48(5), 1779 – 1801.
25. Guoa H., Neely C. (2008). *Investigating the intertemporal riskreturn relation in international stock markets with the component GARCH model*. Economics Letters, 99, 371 – 374.
26. Huang X., Zhou H., Zhu H. (2009). *A framework for assessing the systemic risk of major financial institutions*. Journal of Banking and Finance, 33, 2036 – 2049.
27. Komunjer I. (2005). *Quasi-maximum likelihood estimation for conditional quantiles*. Journal of Econometrics, 128(1), 137 – 164.
28. Kupiec P.H. (1995). *Techniques for verifying the accuracy of risk measurement models*. The Journal of Derivatives, 3(2).
29. Li D., Ghoshray A., Morley B. (2012). *Measuring the risk premium in uncovered interest parity using the component GARCH-M model*. International Review of Economics and Finance, 24, 167 – 176.
30. Patton A.J. (2011). *Volatility forecast comparison using imperfect volatility proxies*. Journal of Econometrics, 160(1), 246 – 256.
31. Popescu A., Turcu C. (2014). *Sovereign Debt and Systemic Risk: An MES and CES Approach*. Working paper.
32. Pramor M., Tamirisa N. (2006). *Common Volatility Trends in the Central and Eastern European Currencies and the Euro*. IMF Working Paper.
33. Sosvilla-Rivero S., Morales-Zumaquero A. (2012). *Volatility in EMU sovereign bond yields: permanent and transitory components*. Applied Financial Economics, 22(17), 1453 – 1464.
34. Yun J., Moon H. (2014). *Measuring Systemic Risk in the Korean Banking Sector via Dynamic Conditional Correlation Models*. Pacific-Basin Finance Journal, 27, 94 – 114.

A Appendix - Multivariate conditional volatility models

One of the most widely used models to obtain CoVaR and MES forecasts is the DCC-GARCH model. The estimation of this model is carried out in two steps. First

the conditional volatilities are estimated on the basis of a univariate GARCH-type model, then the conditional correlations are estimated by the bivariate Dynamic Conditional Correlation (DCC) model (Engle, 2002). DCC is a parsimonious model used to forecast high dimensional covariance matrices with clear computational advantages⁴. Motivated by the aim of investigating the role of long-range dependence in volatility, not yet been considered in systemic risk framework (to the best of our knowledge), we consider two long-memory volatility models as FIGARCH(1, d , 1) and Component-GARCH(1, 1), both implemented using the DCC framework. In order to evaluate the relative benefits of allowing for long-memory when forecasting systemic risk, we compare the long-memory multivariate models with two short-memory multivariate volatility models. These are the multivariate DCC-GARCH(1, 1) and DCC-GJR(1, 1) models, the most widely used models in systemic risk measurement. Finally, to jointly capture the long-range dependence and the leverage effect we introduce the comprehensive Asymmetric-Component-GARCH (ACGARCH) model combined with DCC. The details of these five models are presented.

A.1 The DCC-ACGARCH(1,1) model

Engle and Lee (1999) propose a version of the CGARCH model (see section A.4) with asymmetric shocks structure to capture the asymmetric volatility pattern, namely leverage effect. The authors use the GJR treatment to allow shocks affect both the volatility components asymmetrically. However, they find that the leverage effect is mainly temporary since the leverage term is statistically significant for the transitory component, but not in the trend component. Moreover, Gallant et al. (1993) find that the leverage effect is heavily damped over time. Hence, we introduce a comprehensive model called Asymmetric-Component-GARCH (ACGARCH) where shocks affect only the volatility transitory component asymmetrically and this specification is the most appropriate to capture important stylized facts of financial data, in particular the leverage effect and the long-range dependence. We have combined the leverage effect, already considered in systemic risk framework (e.g. Girardi and Ergun, 2013; Brownlees and Engle, 2017), and the long-range dependence, not yet considered in systemic risk framework, but applied successfully to other fields by many authors (e.g. Christoffersen et al., 2008; Guoa and Neely, 2008; Li et al., 2012). The ACGARCH(1, 1) specification for the conditional variance is:

$$\begin{aligned} q_t &= \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \\ \sigma_t^2 - q_t &= \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) + \gamma(\varepsilon_{t-1}^2 - q_{t-1})I_{(\varepsilon_{t-1} < 0)} \end{aligned} \quad (\text{A.1})$$

⁴ Alternative models to DCC are reported in the survey of the literature provided by Bauwens et al. (2006).

where γ captures the short-run leverage effect. Similarly to the CGARCH model, q_t is the long-run component and $\sigma_t^2 - q_t$ is the short-run component. The one-step-ahead forecast of the conditional volatility is:

$$\begin{aligned} q_{t+1} &= \omega + \rho q_t + \varphi(\varepsilon_t^2 - \sigma_t^2) \\ \sigma_{t+1}^2 - q_{t+1} &= \alpha(\varepsilon_t^2 - q_t) + \beta(\sigma_t^2 - q_t) + \gamma(\varepsilon_t^2 - q_t)I_{(\varepsilon_t < 0)} \end{aligned} \quad (\text{A.2})$$

To implement the ACGARCH(1,1) model in the multivariate context, we use the DCC framework to model the dynamic processes of the correlations directly, where the covariance matrix can be decomposed into time-varying conditional standard deviations and a correlation matrix:

$$\boldsymbol{\Sigma}_{si,t} = \mathbf{D}_{si,t}^{1/2} \mathbf{C}_{si,t} \mathbf{D}_{si,t}^{1/2} = \begin{bmatrix} \sigma_{s,t} & 0 \\ 0 & \sigma_{i,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{si,t} \\ \rho_{si,t} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{s,t} & 0 \\ 0 & \sigma_{i,t} \end{bmatrix}$$

where $\mathbf{C}_{si,t}$ is the conditional correlation matrix, whose elements are $\rho_{si,t}$, and $\mathbf{D}_{si,t}^{1/2}$ is the diagonal matrix of conditional standard deviations estimated by ACGARCH(1,1). Engle (2002) specifies the conditional correlation matrix as follows:

$$\begin{aligned} \mathbf{C}_{si,t} &= \text{diag}(\mathbf{Q}_{si,t})^{-1/2} \mathbf{Q}_{si,t} \text{diag}(\mathbf{Q}_{si,t})^{-1/2} \\ \mathbf{Q}_{si,t} &= (1 - a - b)\bar{\mathbf{Q}}_{si} + a(\mathbf{z}_{si,t-1} \mathbf{z}'_{si,t-1}) + b \mathbf{Q}_{si,t-1} \end{aligned}$$

where $\mathbf{Q}_{si,t}$ is the pseudo correlation matrix and $\bar{\mathbf{Q}}_{si}$ is the unconditional correlation matrix of the standardized residuals $\mathbf{z}_{si,t} = \mathbf{D}_{si,t}^{-1/2} \boldsymbol{\varepsilon}_{si,t}$. The k -step-ahead prediction of the conditional covariance matrix:

$$\boldsymbol{\Sigma}_{si,t+k} = \mathbf{D}_{si,t+k}^{1/2} \mathbf{C}_{si,t+k} \mathbf{D}_{si,t+k}^{1/2}$$

is solved by forecasting $\mathbf{D}_{si,t+k}^{1/2}$ and $\mathbf{C}_{si,t+k}$ separately. The predictions of the univariate variances, contained in $\mathbf{D}_{si,t+k}^{1/2}$, are obtained by the ACGARCH(1,1) forecast. Since \mathbf{C}_t is a non-linear process, according to Engle and Sheppard (2001) we assume for simplicity that $\mathbb{E}_t[\mathbf{z}_{si,t+i} \boldsymbol{\eta}'_{si,t+i}] \approx \mathbf{Q}_{si,t+i}$ for $i \in \{1, \dots, k\}$. Hence, the k -step-ahead forecasts of the pseudo correlation and conditional correlation matrices are given by:

$$\begin{aligned} \mathbf{Q}_{si,t+k} &= \sum_{i=0}^{k-2} (1 - a - b)\bar{\mathbf{Q}}_{si} (a + b)^i + (a + b)^{k-1} \mathbf{Q}_{t+1} \\ \mathbf{C}_{si,t+k} &= \text{diag}(\mathbf{Q}_{si,t+k})^{-1/2} \mathbf{Q}_{si,t+k} \text{diag}(\mathbf{Q}_{si,t+k})^{-1/2} \end{aligned}$$

A.2 The DCC-GARCH(1, 1) model

The short-memory GARCH models (Bollerslev, 1986) are very popular in the literature for the volatility analysis of the financial returns and their application to study these financial phenomena is almost consolidate. They model in a parsimonious way the conditional heteroskedasticity. The GARCH(1, 1) specification for the univariate conditional variance is:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

The k -step-ahead forecast of the conditional volatility for the GARCH(1, 1) model is:

$$\sigma_{t+k}^2 = \sigma^2 + (\alpha + \beta)^{k-1} (\sigma_{t+1}^2 - \sigma^2)$$

where σ^2 is the unconditional variance. The one-step-ahead forecast is: $\sigma_{t+1}^2 = \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2$. In order to implement the GARCH(1, 1) model in the multivariate context, we use the DCC approach described above, with the same forecast functions for $\mathbf{Q}_{si,t+k}$ and $\mathbf{C}_{si,t+k}$.

A.3 The DCC-GJR(1,1) model

GJR-GARCH (GJR) (Glosten et al., 1993) is an alternative short-memory model that captures the leverage effect present in the financial data, allowing the conditional variance to respond differently to the past negative and positive innovations. Stock market volatility responds to stock price movements asymmetrically: bad news (negative shocks) tends to increase investors expectation about future market volatility more than good news (positive shocks). The GJR(1, 1) specification for the conditional variance is:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{(\varepsilon_{t-1} < 0)} + \beta \sigma_{t-1}^2$$

where γ captures the leverage effect, in particular the different impact of negative shocks on volatility than positive shocks, and $I_{(\varepsilon_{t-1} < 0)}$ denotes the indicator function ($I_{\varepsilon_{t-1} < 0} = 1$ when $\varepsilon_{t-1} < 0$ and 0 otherwise). The k -step-ahead forecast of the conditional volatility for the GJR(1, 1) model, with $k > 2$, is:

$$\sigma_{t+k}^2 = \omega + \sum_{i=0}^{k-1} \left(\alpha + \frac{\gamma}{2} + \beta \right)^i + \left(\alpha + \frac{\gamma}{2} + \beta \right)^{k-1} \sigma_{t+1}^2$$

where the one-step-ahead forecast is $\sigma_{t+1}^2 = \omega + \alpha \varepsilon_t^2 + \gamma \varepsilon_t^2 I_{(\varepsilon_t < 0)} + \beta \sigma_t^2$. As with the GARCH(1, 1) model, in order to implement the GJR (1, 1) model in the multivariate context, we use the DCC approach described above, with the same forecast functions $\mathbf{Q}_{si,t+k}$ and $\mathbf{C}_{si,t+k}$.

A.4 The DCC-CGARCH(1,1) model

An alternative model that captures the long-range dependence in volatility is the Component-GARCH (CGARCH) model (Engle and Lee, 1999), where the conditional volatility is additively decomposed into two components, a long-run or permanent component q_t and a short-run or transitory component $\sigma_t^2 - q_t$. The CGARCH(1, 1) specification for the conditional variance is:

$$\begin{aligned} q_t &= \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \\ \sigma_t^2 - q_t &= \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \end{aligned} \quad (\text{A.3})$$

where q_t determines the unconditional variance and is interpreted as volatility trend, and $\sigma_t^2 - q_t$ is the distance between the conditional variance and its trend. The long-run component captures the long-run impact of an innovation (movements) and is more persistent than the short-run component accounts for the noisier short-run movements or the transitory effect from a variance innovation. According to Li et al. (2012), “separating permanent and transitory risk is important in assessing whether this uncertainty is driven by macroeconomic fundamentals or by market sentiments, which could affect the investment strategies”. In addition, the long-run component is mainly driven by shocks to economic fundamentals, whereas the short-run component is driven by transitory shifts in financial market sentiment or short-term position-taking (Pramor and Tamirisa, 2006; Christoffersen et al., 2008; Sosvilla-Rivero and Morales-Zumaquero, 2012; Li et al., 2012). The k -step-ahead forecast of the conditional volatility for the CGARCH(1, 1) model is:

$$\begin{aligned} q_{t+k} &= \frac{\omega}{1-\rho} + \rho^k \left(q_t - \frac{\omega}{1-\rho} \right) \\ \sigma_{t+k}^2 - q_{t+k} &= (\alpha + \beta)^k (\sigma_t^2 - q_t) \end{aligned}$$

As for the GJR(1, 1) model, to implement CGARCH(1, 1) in the multivariate context, we use the DCC approach described above, with the same forecast functions for $\mathbf{Q}_{si,t+k}$ and $\mathbf{C}_{si,t+k}$.

A.5 The DCC-FIGARCH(1,d,1) model

Let $\mathbf{r}_{si,t} = (r_t^s, r_t^i)'$ be the vector denoting the system-institution return pair at time $t = 1, \dots, T$, whose joint dynamics are given by the following bivariate model:

$$\mathbf{r}_{si,t} = \boldsymbol{\varepsilon}_{si,t}, \quad \boldsymbol{\varepsilon}_{si,t} = \boldsymbol{\Sigma}_{si,t}^{1/2} \mathbf{z}_{si,t} \quad (\text{A.4})$$

where $\boldsymbol{\varepsilon}_{si,t} = (\varepsilon_{s,t}, \varepsilon_{i,t})' \sim iid(\mathbf{0}, \boldsymbol{\Sigma}_{si,t})$, $\mathbf{z}_{si,t} = (z_{s,t}, z_{i,t})' \sim iid(\mathbf{0}, \mathbf{I}_2)$, \mathbf{I}_2 is the two-by-two identity matrix, $z_{s,t}$ and $z_{i,t}$ are assumed to be independent, and, for empirical

simplicity, the time-varying covariance matrix $\Sigma_{si,t}$ is obtained by the Cholesky decomposition:

$$\Sigma_{si,t} = \begin{bmatrix} \sigma_{s,t}^2 & \sigma_{s,t}\sigma_{i,t}\rho_{si,t} \\ \sigma_{s,t}\sigma_{i,t}\rho_{si,t} & \sigma_{i,t}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{s,t} & 0 \\ \sigma_{i,t}\rho_{si,t} & \sigma_{i,t}\sqrt{1-\rho_{si,t}^2} \end{bmatrix}$$

The univariate variances $\sigma_{s,t}^2$ and $\sigma_{i,t}^2$ can be separately modeled by the Fractionally Integrated GARCH (FIGARCH) model proposed by Baillie et al. (1996). This model better captures the long-run dynamic dependencies in the conditional variance than standard GARCH. The short-run dynamics, in fact, are modeled by GARCH parameters, while shocks to the conditional variance will die out at a slow hyperbolic rate of decay for the influence of lagged squared innovations determined by a fractional differencing parameter d . The FIGARCH(p, d, q) specification for the conditional variance is:

$$\sigma_t^2 = \frac{\omega}{1-\beta(L)} + \left(1 - \frac{\phi(L)(1-L)^d}{1-\beta(L)}\right) \varepsilon_t^2 \equiv \frac{\omega}{1-\beta(L)} + \lambda(L)\varepsilon_t^2 \quad (\text{A.5})$$

where L denotes the lag or backshift operator, $\lambda(L) \equiv \lambda_1 L + \lambda_2 L^2 + \dots$, $0 < d < 1$ and all the roots of $\phi(L)$ and $[1-\beta(L)]$ lie outside the unit circle. For $d = 0$, the FIGARCH(p, d, q) model reduces to a GARCH(p, q) model with an exponential decay of the shocks, while for $d = 1$, it is a IGARCH(p, q) model with an infinite persistence. By a direct extension of the corresponding proof for the IGARCH case, Baillie et al. (1996) show that the FIGARCH class of processes is strictly stationary and ergodic for $0 = d \leq 1$. The specification of FIGARCH(1, d , 1) is obtained by replacing $\beta(L) = \beta L$ and $\phi(L) = 1 - \phi L$. The k -step-ahead forecast of the FIGARCH(1, d , 1) conditional volatility is:

$$\sigma_{t+k+k-1}^2 = \frac{\omega}{1-\beta} + \left(1 - \frac{(1-\phi L)(1-L)^d}{1-\beta L}\right) \varepsilon_{t+k-1}^2$$

In order to implement the FIGARCH(1, d , 1) model in the multivariate context, we again use the DCC approach described above, with the same forecast functions for $\mathbf{Q}_{si,t+k}$ and $\mathbf{C}_{si,t+k}$.

Table 1 Out-of-sample performances of DCC-GARCH-type models in forecasting CoVaR measured by TTL loss function on 91 US financial institutions, 2006 – 2012.

	QR	DCC GARCH	DCC GJR	DCC CGARCH	DCC FIGARCH	DCC ACGARCH
Overall	1.046	0.228	0.212	0.227	0.204	0.240
Dep.	0.912	0.214	0.203	0.213	0.194	0.223
Ins.	1.156	0.238	0.220	0.239	0.211	0.251
Bro.	1.205	0.244	0.228	0.242	0.223	0.269
Oth.	0.976	0.220	0.204	0.219	0.197	0.232

Table 2 Selected quantiles (10%, 50% and 90%) of the parameter estimates of the DCC-ACGARCH models on 91 US financial institutions, 2000 – 2006.

	ACGARCH						DCC			
	ω	α	β	ρ	φ	γ	const	a	b	
Dep.	$q_{0.1}$	0.001	0.000	0.563	0.995	0.000	0.000	0.412	0.006	0.960
	$q_{0.5}$	0.004	0.052	0.830	0.999	0.024	0.055	0.611	0.014	0.982
	$q_{0.9}$	0.013	0.143	0.888	1.000	0.051	0.145	0.666	0.033	0.988
Ins.	$q_{0.1}$	0.000	0.002	0.261	0.996	0.000	0.000	0.306	0.005	0.601
	$q_{0.5}$	0.002	0.073	0.798	0.999	0.008	0.034	0.455	0.014	0.981
	$q_{0.9}$	0.017	0.169	0.907	1.000	0.041	0.153	0.587	0.083	0.993
Bro.	$q_{0.1}$	0.002	0.000	0.695	0.997	0.014	0.034	0.573	0.010	0.960
	$q_{0.5}$	0.006	0.000	0.903	0.997	0.017	0.058	0.681	0.014	0.982
	$q_{0.9}$	0.014	0.028	0.943	1.000	0.038	0.115	0.725	0.032	0.986
Oth.	$q_{0.1}$	0.000	0.000	0.046	0.996	0.000	0.000	0.289	0.000	0.471
	$q_{0.5}$	0.008	0.033	0.847	1.000	0.010	0.066	0.476	0.010	0.957
	$q_{0.9}$	0.040	0.233	0.932	1.000	0.035	0.123	0.665	0.040	0.989

Table 3 Out-of-sample performances of DCC-GARCH-type models in forecasting MES measured by TMSE loss function on 91 US financial institutions, 2006 – 2012.

	LR	DCC GARCH	DCC GJR	DCC CGARCH	DCC FIGARCH	DCC ACGARCH
Overall	26.324	21.654	21.899	21.719	21.970	22.004
Dep.	15.732	12.744	13.026	12.778	12.996	12.843
Ins.	9.053	6.581	6.675	6.538	6.966	6.528
Bro.	157.39	134.02	133.40	134.30	134.61	136.62
Oth.	11.535	9.184	10.026	9.387	9.351	9.565

Table 4 Ranking of the CoVaR averages over the out-of-sample 2006 – 2012.

	DCC GARCH	DCC GJR	DCC CGARCH	DCC FIGARCH	DCC ACGARCH
Overall	-3.174	-3.203	-3.125	-3.393	-3.169
1st: Bro.	-3.420	-3.449	-3.353	-3.600	-3.392
2nd: Oth.	-3.193	-3.219	-3.139	-3.453	-3.204
3rd: Ins.	-3.150	-3.184	-3.105	-3.461	-3.159
4th: Dep.	-3.121	-3.145	-3.075	-3.343	-3.093

Table 5 Ranking of the MES averages over the out-of-sample 2006 – 2012.

	DCC GARCH	DCC GJR	DCC CGARCH	DCC FIGARCH	DCC ACGARCH
Overall	-3.091	-3.081	-3.155	-3.060	-3.131
1st: Bro.	-4.104	-4.082	-4.221	-4.124	-4.166
2nd: Dep.	-3.349	-3.349	-3.406	-3.200	-3.413
3rd: Oth.	-3.173	-3.187	-3.251	-3.093	-3.209
4th: Ins.	-2.645	-2.615	-2.696	-2.712	-2.663

Table 6 Comparison of CoVaR averages over different sub-samples.

	DCC	DCC	DCC	DCC	DCC
Pre-crisis	GARCH	GJR	CGARCH	FIGARCH	ACGARCH
Dep.	-2.175	-2.171	-2.109	-2.147	-2.095
Ins.	-1.907	-1.930	-1.862	-1.872	-1.859
Bro.	-2.224	-2.189	-2.136	-2.188	-2.128
Oth.	-1.881	-1.878	-1.834	-1.896	-1.823
Crisis					
Dep.	-4.166	-4.157	-4.259	-4.250	-4.008
Ins.	-4.163	-4.165	-4.262	-4.269	-4.070
Bro.	-4.333	-4.345	-4.443	-4.408	-4.157
Oth.	-4.156	-4.151	-4.243	-4.308	-4.054
Post-crisis					
Dep.	-2.861	-2.910	-2.753	-3.217	-2.910
Ins.	-2.985	-3.037	-2.869	-3.488	-3.059
Bro.	-3.291	-3.345	-3.143	-3.581	-3.353
Oth.	-3.071	-3.119	-2.945	-3.448	-3.155

Table 7 Comparison of MES averages over different sub-samples.

	DCC	DCC	DCC	DCC	DCC
Pre-crisis	GARCH	GJR	CGARCH	FIGARCH	ACGARCH
Dep.	-2.209	-2.155	-2.251	-2.250	-2.249
Ins.	-2.023	-1.942	-2.051	-2.227	-2.013
Bro.	-4.048	-4.084	-4.157	-4.156	-4.132
Oth.	-2.628	-2.599	-2.675	-2.698	-2.599
Crisis					
Dep.	-3.800	-3.740	-3.736	-3.665	-3.878
Ins.	-2.586	-2.530	-2.586	-2.563	-2.619
Bro.	-4.701	-4.655	-4.747	-4.586	-4.925
Oth.	-3.423	-3.363	-3.404	-3.257	-3.533
Post-crisis					
Dep.	-3.432	-3.476	-3.505	-3.225	-3.495
Ins.	-2.838	-2.836	-2.922	-2.914	-2.857
Bro.	-3.827	-3.801	-3.981	-3.890	-3.803
Oth.	-3.196	-3.257	-3.328	-3.117	-3.213