

# Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures

Denisa Banulescu\*, Christophe Hurlin†, Jérémy Leymarie‡, Olivier Scaillet§

May 18, 2016

## Abstract

This paper proposes two backtesting tests to assess the validity of the systemic risk measure forecasts. This new tool meets the need of financial regulators of evaluating the quality of systemic risk measures generally used to identify the financial institutions contributing the most to the total risk of the financial system (SIFIs). The tests are based on the concept of cumulative violations and it is built up in analogy with the recent backtesting procedure proposed for ES (Expected Shortfall). First, we introduce two backtests that apply for the case of the MES (Marginal Expected Shortfall) forecasts. The backtesting methodology is then generalised to MES-based systemic risk measures (SES, SRISK) and to the  $\Delta\text{CoVaR}$ . Second, we study the asymptotic properties of the tests in presence of estimation risk and we investigate their finite sample performances via Monte Carlo simulations. Finally, we use our backtests to assess the validity of the MES, SRISK and  $\Delta\text{CoVaR}$  forecasts on a panel of EU financial institutions.

## 1 Introduction

Many systemic risk measures have been proposed in the academic literature over the past years (see Benoit et al. 2016, for a survey), the most well-known being the Marginal Expected Shortfall (MES) and the Systemic Expected Shortfall (SES) of Acharya et al. (2010), the Systemic Risk Measure (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2015), and the Delta Conditional Value-at-Risk ( $\Delta\text{CoVaR}$ ) of Adrian and Brunnermeier (2016). These measures are designed to summarize the systemic risk contribution of each financial institution into a single figure, in order to identify the so-called systemically important financial institutions (SIFIs), *i.e.* the firms whose failure might trigger a crisis in the whole financial system. The identification of the SIFIs is crucial for the systemic risk regulation, whatever the regulation tools considered (higher capital requirements, specific regulation, systemic risk tax, etc.). As a consequence, regulators and other end-users of these measures thus need guidance on how to select the ones most adapted to their objectives. In this context, the validation is a key requirement for any systemic risk measure to become an industry standard. Some attempts of validation procedures have been proposed in the

---

\*University of Orléans (LEO, UMRS CNRS 7332).

†University of Orléans (LEO, UMRS CNRS 7332).

‡University of Orléans (LEO, UMRS CNRS 7332).

§University of Geneva and Swiss Finance Institute.

litterature. Following the coherent risk approach of Artzner et al. (1999), Chen, Iyengar, and Moallemi (2013) define an axiomatic framework for systemic risk measures. Brownlees and Engle (2015) show that banks with higher SRISK before the financial crisis were more likely to be bailed out by the government and to receive capital injections from the Federal Reserve. Engle, Jondeau, and Rockinger (2015) compare the empirical ranking obtained with a given measure with the one computed by the FSB-BCBS, which is based on confidential bank supervisory data. However, to the best of our knowledge, no backtesting procedure have been proposed yet for the systemic risk measures.

In this paper we propose a general framework for backtesting the MES and the related systemic risk measures. Introduced by Acharya et al. (2010), the MES of a financial firm is defined as its short-run expected equity loss conditional on the market taking a loss greater than its Value-at-Risk (VaR). The MES is simple to compute and therefore easy for regulators to consider. Furthermore, the MES constitutes one of the key elements (with the leverage and the market value) of two popular systemic risk measures, i.e. the SRISK and the SES. The latter is equal to the expected amount a bank is undercapitalized in a future systemic event in which the overall financial system is undercapitalized (Acharya et al. 2010). Thus, the SES increases in the bank's expected losses during a crisis. The SRISK corresponds to the expected capital shortfall of a given financial institution, conditional on a crisis affecting the whole financial system (Brownlees and Engle 2015). This index can be used to rank the financial institutions according to their systemic risk, since the firms with the highest SRISK are the largest contributors to the undercapitalization of the financial system in times of distress. What it is important to note is that these two systemic risk measures are theoretically related to the the current leverage, the current equity value of the financial institution and its MES. As a consequence, testing for the validity of the SRISK or SES forecasts, implies to test for the validity of MES forecasts.

As defined by Jorion (2007), backtesting is a formal statistical framework that consists in verifying if actual losses are in line with projected losses. This involves a systemic comparison of the history of model-generated risk measure forecasts with actual returns. Since the true value of the risk measure is unobservable, this comparison generally relies on violations. When one comes to backtest the VaR, a violation is said to occur when the *ex-post* portfolio return is lower than the VaR forecast. The trick of our paper consists in defining an appropriate concept of violation for the MES forecast. Then, it is possible to adopt the same standard backtesting tests (Kupiec, 1995, Christoffersen, 1998, etc.) as those that are currently used for the VaR forecasts (see Christoffersen, 2009 for a survey on backtesting). Our approach is the following. First, we introduce a concept a concept of Conditional-VaR (CoVaR), inspired from the systemic risk measure proposed by Adrian and Brunnermeier (2016). The  $(\beta, \alpha)$ -CoVaR is defined as the  $\beta$ -quantile of the truncated distribution of the firm's returns given that the market takes a loss greater than its  $\alpha$ -VaR. We express the MES as an integrale of the CoVaRs for all the coverage rate  $\beta$  between 0 and 1. To the best of our knowledge, this is the first time that a relationship is established between the CoVaR and the MES, and so on with the SES and SRISK. Second, we define a concept of *joint*

violation of the  $(\beta, \alpha)$ -CoVaR of the firm's returns and  $\alpha$ -VaR of the market returns. Finally, we extend the concept of cumulative violation recently proposed by Du and Escanciano (2015) for the Expected Shortfall (ES) backtests, to a bivariate case. We define a cumulative *joint* violation process defined as the integrale of the joint violation processes for all the coverage rate  $\beta$  between 0 and 1. This process can be viewed as a "violation" counterpart of the definition that we propose for the MES. Furthermore, we show that this process has the same properties than the cumulative violation process introduced by Du and Escanciano for the ES. In particular, this process is a martingale difference sequence (*mds*). Exploiting this *mds* property, we propose two backtests for the MES : an Unconditional Coverage (UC) test and an Independence (IND) test (Christoffersen, 1998). The UC test refers to the fact that the violations frequency should be in line with the theoretical probability to observe a violation. Failure of UC means that the MES forecast does not measure the risk accurately. Besides UC, MES model should satisfy the independence property. The independence property refers to clustering of violations. One advantage of our approach is that it allows to backtest either conditional (with respect to the past information set) MES (Brownless and Engle 2015) or unconditional MES (Acharya et al. 2010) forecasts.

We consider the same test statistics as those used by Du and Escanciano (2015) for the backtesting of the ES, since these statistics are similar to those generally used by the regulator or the risk manager in the context of the VaR backtesting (Kupiec, 1995). We derive the asymptotic distribution of the two test statistics, while taking into account the estimation risk (Escanciano and Olmo 2010, 2011, Gouriéroux and Zakoian, 2013). Indeed, the MES forecasts are generally issued from a parametric model for which the parameters have to be estimated. Then, the use of standard backtesting procedures to assess the MES model in an out-of-sample basis can be misleading, because these procedures do not consider the impact of the estimation risk. That is why we propose a robust version of our test statistics to the estimation risk. Our Monte Carlo simulations show that these robust test statistics have good finite sample properties for realistic sample sizes.

Finally, from the previous results we derive backtesting tests for the  $\Delta$ CoVaR and the MES-based systemic risk measures (SES, SRISK). The  $\Delta$ CoVaR, which is defined as the difference between the conditional VaR (CoVaR) of the financial system conditional on an institution being in distress and the CoVaR conditional on the median state of the institution. The backtesting for the  $\Delta$ CoVaR is based on a vector of two joint violations defined for the two conditioning events. The intuition of the test is then similar to the backtests proposed for multi-level VaR (Francq and Zakoian, 2015), i.e. the VaR defined for a finite set of coverage rates (Hurlin and Tokpavi 2006, Pérignon and Smith 2008, Leccadito, Boffelli and Urga 2014). Since the SRISK and the SES can be written as a deterministic function of the MES, we adapt the two MES backtesting tests (UC and INDF) and their robust versions for these systemic risk measures.

The paper is organized as follows. In Section 1, we define the MES and we express it as a simple integral of conditional Value-at-Risk (CoVaR). Section 2 introduces the concept of cumulative joint violation process and its properties, which will be used for the backtesting test of the MES. The

backtesting tests for the MES are presented in Section 3 and their finite sample properties are illustrated by Monte Carlo simulations in Section 4. Section 5 extends our backtests to the main systemic risk measures that depends on the MES, i.e. the SES and the SRISK of Acharya et al. (2012) and Brownlees and Engle (2015), and to the  $\Delta$ CoVaR of Adrian and Brunnermeier (2016). An empirical application is proposed in Section 6. Finally, the last section concludes.

## 2 Marginal Expected Shortfall

Let  $Y_t = (Y_{1t}, Y_{2t})'$  denote the vector of stock returns of two assets at time  $t$ . In the specific context of systemic risk,  $Y_{1t}$  generally corresponds to the stock return of a financial institution, whereas  $Y_{2t}$  corresponds to the market return.<sup>1</sup> Denote by  $\Omega_{t-1}$  the information set available at time  $t-1$ , with  $(Y_{t-1}, Y_{t-2}, \dots) \subseteq \Omega_{t-1}$  and  $F(\cdot; \Omega_{t-1})$  the joint cumulative distribution function (*cdf*) of  $Y_t$  given  $\Omega_{t-1}$  such that  $F(y_t; \Omega_{t-1}) \equiv \Pr(Y_{1t} < y_1, Y_{2t} < y_2 | \Omega_{t-1})$  for any  $y = (y_1, y_2)' \in \mathbb{R}^2$ . Assume that  $F(\cdot; \Omega_{t-1})$  is continuous.

Following Acharya et al. (2010), we define the MES of a financial firm as its short-run expected equity loss conditional on the market taking a loss greater than its Value-at-Risk (VaR). This definition of the MES was extended to a  $\Omega_{t-1}$ -conditional version by Brownlees and Engle (2015). Formally, the  $\alpha$ -level MES of the financial institution at time  $t$  given  $\Omega_{t-1}$  is defined as

$$MES_{1t}(\alpha) = \mathbb{E}(Y_{1t} | Y_{2t} \leq VaR_{2t}(\alpha); \Omega_{t-1}), \quad (1)$$

where  $VaR_{2t}(\alpha)$  denotes the  $\alpha$ -level VaR of  $Y_{2t}$ , such that  $\Pr(Y_{2t} \leq VaR_{2t}(\alpha) | \Omega_{t-1}) = \alpha$  with  $\alpha \in [0, 1]$ . Notice that if the market return  $Y_{2t}$  is defined as the value-weighted average of the firm's returns (for all the firms that belong to the financial system), then the MES of one firm corresponds to the derivative of the market's Expected Shortfall (ES) with respect to the firm's market share (Scaillet, 2004), hence the term "Marginal". Hence, the MES can be interpreted as a measure of the participation of one financial institution to the overall systemic risk.

As usual,  $VaR_{2t}(\alpha)$  is defined as the  $\alpha$ -th percentile of the marginal distribution of  $Y_{2t}$ , denoted  $F_{Y_2}(\cdot, \Omega_{t-1})$ , with  $VaR_{2t}(\alpha) = F_{Y_2}^{-1}(\cdot, \Omega_{t-1})$ .<sup>2</sup> Similarly, the MES can be expressed as a function of the quantiles of the conditional distribution of  $Y_{1t}$  given  $Y_{2t} \leq VaR_{2t}(\alpha)$ . For that, we introduce a concept of Conditional-VaR (CoVaR) inspired from the systemic risk measure proposed by Adrian and Brunnermeier (2016). For any coverage level  $\beta \in [0, 1]$ , the CoVaR for the firm 1 at time  $t$  is the quantity  $CoVaR_{1t}(\beta, \alpha)$  such that

$$\Pr(Y_{1t} \leq CoVaR_{1t}(\beta, \alpha) | Y_{2t} \leq VaR_{2t}(\alpha); \Omega_{t-1}) = \beta. \quad (2)$$

There are two main differences between  $CoVaR_{1t}(\beta, \alpha)$  and the CoVaR introduced by Adrian and Brunnermeier (2014). First, the conditioning event is based on an inequality, i.e.  $Y_{2t} \leq VaR_{2t}(\alpha)$  as in Girardi and Ergun (2013), rather than on the equality  $Y_{2t} = VaR_{2t}(\alpha)$ . Second we introduce

<sup>1</sup>Our results can be easily extended to the general case with  $m \geq 2$  assets.

<sup>2</sup>For simplicity in the notations, we do not use the usual convention that defines the VaR as the opposite of the  $\alpha$ -quantile of the returns distribution. Similarly, we define the MES as the conditional expectation.

a distinction between the coverage level  $\beta$  of the CoVaR and the coverage level  $\alpha$  of the VaR, which is used to define the conditioning event. Then, the  $(\beta, \alpha)$ -level CoVaR can also be defined as

$$CoVaR_{1t}(\beta, \alpha) = F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}^{-1}(\beta; \Omega_{t-1}), \quad (3)$$

where  $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(\cdot; \Omega_{t-1})$  is the *cdf* of the conditional distribution of  $Y_{1t}$  given  $Y_{2t} \leq VaR_{2t}(\alpha)$  and  $\Omega_{t-1}$ . Definition of conditional probability and a change in variables yields a useful representation of the MES in terms of CoVaR.

$$MES_{1t}(\alpha) = \int_0^1 CoVaR_{1t}(\beta, \alpha) d\beta. \quad (4)$$

Equation (4) gives a simple relationship between two risk measures, i.e. the MES and the CoVaR, that are broadly used in the systemic risk literature. Notice that this expression can be used to determine either the  $\Omega_{t-1}$ -conditional  $MES_{1t}(\alpha)$ , as in Brownlees and Engle (2015) or the unconditional  $MES_1(\alpha)$ , as in Acharya et al. (2010). Furthermore, this definition of the MES is valid for any bivariate distribution and for any dynamic model (DCC, CCC, etc.).

For some particular distributions, the conditional *cdf*  $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(\cdot; \Omega_{t-1})$  that defined the CoVaR has a closed form expression. For instance, Arnold et al. (1993) calculate the marginal of a bivariate normal distribution with double truncation over one variable. Horrace (2005) formalized analytical results on the truncated multivariate normal distribution, where the truncation is one-sided and at an arbitrary point. Ho et al. (2012) study of the truncated multivariate  $t$  distribution. However, whatever the distribution considered, it is possible to express the *cdf* of the truncated distribution of  $Y_{1t}$  given  $Y_{2t} \leq VaR_{2t}(\alpha)$  as a simple function of the *cdf* of the joint distribution of  $Y_t$ , with

$$F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(y_1; \Omega_{t-1}) = \frac{1}{\alpha} F(\tilde{y}; \Omega_{t-1}), \quad (5)$$

where the vector  $\tilde{y}$  is defined as  $\tilde{y} = (y_1, VaR_{2t}(\alpha))'$ .

In general, the MES forecasts are issued from a parametric model specified by the researcher, the risk manager or the regulation authority. For instance, Brownlees and Engle (2015) or Acharya, Engle, and Richardson (2012) consider a bivariate DCC model to compute the MES and the SRISK. In practice, the *cdf*  $F(y; \Omega_{t-1}, \theta_0)$  of the joint distribution of  $Y_t$ , the *cdf*  $F_{Y_2}(\cdot; \Omega_{t-1}, \theta_0)$  of the marginal distribution of  $Y_{2t}$  and the *cdf*  $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha; \theta_0)}(y; \Omega_{t-1}, \theta_0)$  of the truncated distribution of  $Y_{1t}$  given  $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$  depend on  $\theta_0$ , an unknown parameter in  $\Theta \in \mathbb{R}^p$ . This parameter has to be estimated before producing MES forecasts.

### 3 Cumulative Joint Violation Process

As defined by Jorion (2007), backtesting is a formal statistical framework that consists in verifying if actual losses are in line with projected losses. This involves a systemic comparison of the history of model-generated risk measure forecasts with actual returns. This comparison generally relies on tests over violations. When one comes to backtest the VaR, a violation is said to occur when the *ex-post* portfolio return is lower than the VaR forecast. In order to backtest the CoVaR and the

MES, we define a *joint* violation of the  $(\beta, \alpha)$ -CoVaR of  $Y_{1t}$  and the  $\alpha$ -VaR of  $Y_{2t}$  at time  $t$ . This violation process is represented by the following binary variable

$$h_t(\alpha, \beta, \theta_0) = \mathbf{1}((Y_{1t} \leq CoVaR_{1t}(\beta, \alpha, \theta_0)) \cap (Y_{2t} \leq VaR_{2t}(\alpha, \theta_0))), \quad (6)$$

where  $\mathbf{1}(\cdot)$  denotes the indicator function. The violation takes the value one if the losses of the firm exceeds its CoVaR and the losses of the market exceeds its VaR, and it is zero otherwise.

The VaR backtesting procedures (Kupiec, 1995; Christoffersen, 1998; Berkowitz et al., 2011, among others) are generally based on the *mds* property of the violation process (see Christoffersen 2009 or Hurlin and Pérignon 2012, for a survey). Here, we adopt the same approach for backtest the CoVaR, and so the MES. Notice that the Bayes' theorem implies that  $\Pr(h_t(\alpha, \beta, \theta_0) = 1 | \Omega_{t-1}) = \alpha\beta$ . Then, equation (2) implies that the violations are Bernoulli variables with mean  $\alpha\beta$ , and that the centered violation  $\{h_t(\alpha, \beta, \theta_0) - \alpha\beta\}_{t=1}^\infty$  is a *mds* for risk levels  $(\alpha, \beta) \in [0, 1]^2$ .

$$\mathbb{E}(h_t(\alpha, \beta, \theta_0) - \alpha\beta | \Omega_{t-1}) = 0. \quad (7)$$

In order to test for the validity of the MES, we consider a cumulative joint violation process which can be viewed as a kind of violations "counterpart" of the MES definition in equation (4). This cumulative *joint* violation process is defined as the integral of the violations  $h_t(\alpha, \beta, \theta_0)$  for all the risk levels  $\beta$  between 0 and 1, with

$$H_t(\alpha, \theta_0) = \int_0^1 h_t(\alpha, \beta, \theta_0) d\beta. \quad (8)$$

This cumulative *joint* violation process is similar to the cumulative violation process introduced by Du and Escanciano (2015) to backtest the ES, even if the two definitions are slightly different. The mean and variance of  $H_t(\alpha, \theta_0)$  are equal to  $\alpha/2$  and  $\alpha(1/3 - \alpha/4)$ , respectively (see appendix A). Furthermore, the Fubini's theorem implies that the *mds* property of the sequence  $\{h_t(\alpha, \beta, \theta_0) - \alpha\beta\}_{t=1}^\infty, \forall (\alpha, \beta) \in [0, 1]^2$  is preserved by integration. As a consequence, the sequence  $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^\infty$  is also a *mds* for any  $\alpha \in [0, 1]$ , that is

$$\mathbb{E}(H_t(\alpha, \theta_0) - \alpha/2 | \Omega_{t-1}) = 0. \quad (9)$$

The backtesting tests that we propose for the MES are based on this *mds* property.

Finally, it is possible to rewrite  $H_t(\alpha, \theta_0)$  in a more convenient way, through the Probability Integral Transformation (PIT). Notice that the cumulative joint violation process  $H_t(\alpha, \theta_0)$  depends on the distribution of  $Y_t$  as follows

$$\begin{aligned} H_t(\alpha, \theta_0) &= \mathbf{1}(Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)) \times \int_0^1 \mathbf{1}(Y_{1t} \leq CoVaR_{1t}(\beta, \alpha, \theta_0)) d\beta \\ &= \mathbf{1}(F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0) \leq \alpha) \times \int_0^1 \mathbf{1}(F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0) \leq \beta) d\beta. \end{aligned} \quad (10)$$

Let us introduce two terms that can be interpreted as "generalized" errors, namely  $u_{2t}(\theta_0) \equiv u_{2t} = F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0)$  and  $u_{12t}(\theta_0) \equiv u_{12t} = F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0)$ . Then, the cumulative

joint violation process becomes

$$H_t(\alpha, \theta_0) = \mathbf{1}(u_{2t} \leq \alpha) \int_0^1 \mathbf{1}(u_{12t} \leq \beta) d\beta = \mathbf{1}(u_{2t} \leq \alpha) \int_{u_{12t}}^1 1 d\beta. \quad (11)$$

Thus, the process  $H_t(\alpha, \theta_0)$  can be expressed as a simple function of the transformed i.i.d. variables  $u_{2t}$  and  $u_{12t}$ , defined over  $[0, 1]$ , such as

$$H_t(\alpha, \theta_0) = (1 - u_{12t}(\theta_0))\mathbf{1}(u_{2t}(\theta_0) \leq \alpha). \quad (12)$$

The PIT implies that the variable  $u_{2t}$  has a uniform  $U_{[0,1]}$  distribution. The generalized error term  $u_{12t}$  has also a  $U_{[0,1]}$  distribution as soon as the transformation  $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta_0)}(\cdot; \Omega_{t-1}, \theta_0)$  is applied to observations  $Y_{1t}$  for which  $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$ .

## 4 Backtesting MES

Exploiting the *mds* property of the cumulative joint violation process, we propose two backtests for the MES. These tests are similar to those generally used by the regulator or the risk manager for VaR backtesting (Christoffersen, 1998). The unconditional coverage (hereafter UC) test corresponds to the null hypothesis

$$H_{0,UC} : \mathbb{E}(H_t(\alpha, \theta_0)) = \alpha/2. \quad (13)$$

Since  $\mathbb{E}(H_t(\alpha, \theta_0)) = \int_0^1 \mathbb{E}(h_t(\alpha; \beta, \theta_0)) d\beta$ , the null  $H_{0,UC}$  can also be written as

$$H_{0,UC} : \Pr(h_t(\alpha, \beta, \theta_0) = 1) = \alpha\beta, \quad \forall \beta \in [0, 1].$$

The null hypothesis UC means that for any coverage level  $\beta$ , the *joint* probability to observe an ex-post return  $y_{1t}$  exceeding its  $(\beta\alpha)$ -CoVaR and an ex-post return  $y_{2t}$  exceeding its  $\alpha$ -VaR must be equal to  $\alpha\beta$ .

The second backtest is based on the independence (hereafter IND) property of the cumulative violation process  $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^\infty$ : the cumulative violations observed at two different dates for the same coverage rate  $\alpha$  must be distributed independently. As Christoffersen (1998), we propose a simple Box-Pierce test (Box and Pierce, 1970) in order to test for the nullity of the first  $K$  autocorrelations of  $H_t(\alpha, \theta_0)$ , denoted  $\rho_k$ . The null of the IND test is then defined as

$$H_{0,IND} : \rho_1 = \dots = \rho_K = 0. \quad (14)$$

with  $\rho_k = \text{corr}(H_t(\alpha, \theta_0) - \alpha/2, H_{t-k}(\alpha, \theta_0) - \alpha/2)$ .

These two tests imply to estimate the parameters  $\theta_0 \in \Theta$  in order to forecast the MES. For simplicity, we consider a fixed forecasting scheme. An in-sample period from  $t = -T + 1$  to  $t = 0$  is used to estimate  $\theta_0$ . Denote by  $\Omega_{-1}$  the information set available at the end of the in-sample period, with  $\{Y_{-T+1}, \dots, Y_0\} \subseteq \Omega_{-1}$  and  $\hat{\theta}_T$  a consistent estimator of  $\theta_0$ . This estimator is used to forecast the MES for all the dates from  $t = 1$  to  $t = n$ . The backtesting tests are then based on the out-of-sample forecasts of the cumulative violation process process given by

$$H_t(\alpha, \hat{\theta}_T) = \left(1 - u_{12t}(\hat{\theta}_T)\right) \mathbf{1}\left(u_{2t}(\hat{\theta}_T) \leq \alpha\right), \quad \forall t = 1, \dots, n. \quad (15)$$

## 4.1 Unconditional Coverage Test

By analogy with the backtest proposed by Du and Escanciano (2015) for ES, and the Kupiec's backtest (1995) for VaR, we propose a standard  $t$ -test for the null hypothesis of unconditional coverage  $H_{0,UC}$  for the MES. This test statistic, denoted  $UC_{MES}$ , is defined as

$$UC_{MES} = \frac{\sqrt{n} \left( \bar{H}(\alpha, \hat{\theta}_T) - \alpha/2 \right)}{\sqrt{\alpha(1/3 - \alpha/4)}}, \quad (16)$$

with  $\bar{H}(\alpha, \hat{\theta}_T)$  the out-of-sample mean of  $H_t(\alpha, \hat{\theta}_T)$

$$\bar{H}(\alpha, \hat{\theta}_T) = \frac{1}{n} \sum_{t=1}^n H_t(\alpha, \hat{\theta}_T). \quad (17)$$

In order to give the intuition of the asymptotic properties of the statistic  $UC_{MES}$ , let us define a similar statistic  $UC_{MES}(\alpha, \theta_0)$  based on the true value of the parameters  $\theta_0$  rather than on its estimator  $\hat{\theta}_T$ . Under the null hypothesis, the sequence  $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^n$  is a  $mds$  with a variance equal to  $\alpha(1/3 - \alpha/4)$ . As a consequence, the Lindeberg-Levy central limit theorem implies that  $UC_{MES}(\alpha, \theta_0)$  has an asymptotic standard normal distribution. A similar result holds for the feasible statistic  $UC_{MES} \equiv UC_{MES}(\alpha, \hat{\theta}_T)$  when  $T \rightarrow \infty$  and  $n \rightarrow \infty$ , whereas  $\lambda = n/T \rightarrow 0$ , i.e. when there is no estimation risk.

However, in the general case  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda < \infty$ , there is an estimation risk as soon as  $\lambda \neq 0$ . In this case, the asymptotic distribution of  $UC_{MES}$  is not standard and depends on the ratio of the in-sample size  $T$  to the out-of-sample size  $n$ . Theorem 1 gives the corresponding asymptotic distribution of  $UC_{MES}$  when  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda$  with  $0 < \lambda < \infty$ .

**Theorem 1** *Under assumptions A1-A4*

$$UC_{MES} \xrightarrow{d} \mathcal{N}(0, \sigma_\lambda^2), \quad (18)$$

where  $\xrightarrow{d}$  denotes the convergence in distribution and where the asymptotic variance  $\sigma_\lambda^2$  is

$$\sigma_\lambda^2 = 1 + \lambda \frac{R'_{MES} \Sigma_0 R_{MES}}{\alpha(1/3 - \alpha/4)}, \quad (19)$$

where  $R_{MES} = \mathbb{E}_0(\partial H_t(\alpha, \theta_0)/\partial \theta)$  and  $\mathbb{V}_{as}(\hat{\theta}_T) = \Sigma_0/T$ .

The proof of Theorem 1 is reported in appendix B. The vector  $R_{MES}$  quantifies the parameter estimation effect on the test statistic  $UC_{MES}$ . Indeed, this estimation risk that comes from the difference between the estimate  $\hat{\theta}_T$  and the true value of the parameter  $\theta_0$ . This difference affects the  $UC_{MES}$  test statistic as follows

$$\begin{aligned} UC_{MES} &= \underbrace{\frac{1}{\sigma_H \sqrt{n}} \sum_{t=1}^n (H_t(\alpha, \theta_0) - \alpha/2)}_{UC_{MES}(\theta_0)} \\ &\quad + \underbrace{\frac{\sqrt{\lambda}}{\sigma_H} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left( \frac{\partial H_t(\alpha, \hat{\theta})}{\partial \theta} \Big| \Omega_{t-1} \right)}_{\text{Estimation risk}} \sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1). \end{aligned} \quad (20)$$



Whatever the dynamic model considered for the returns, the vector  $R_{MES}$  can be simply deduced from the cdf of the joint distribution of  $Y_t$  given  $\Omega_{-1}$ . Indeed, since the derivative of a step function is a Dirac function and a continuous distribution has no point mass, we get

$$R_{MES} = \mathbb{E}_0 \left( \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right) = -\frac{1}{\alpha} \mathbb{E}_0 \left( \frac{\partial F(\tilde{y}_t, \theta_0)}{\partial \theta} \right) \quad (21)$$

where  $\tilde{y}_t = (y_{1t}, VaR_{2t}(\alpha, \theta_0))'$  and

$$\begin{aligned} \frac{\partial F(\tilde{y}_t, \theta_0)}{\partial \theta} &= \underbrace{\int_{-\infty}^{y_{1t}} f(u, VaR_{2t}(\alpha, \theta_0); \theta_0) du}_{\text{Impact on the truncation}} \times \frac{\partial VaR_{2t}(\alpha, \theta_0)}{\partial \theta} \\ &+ \underbrace{\int_{-\infty}^{y_{1t}} \int_{-\infty}^{VaR_{2t}(\alpha, \theta_0)} \frac{\partial f(u, v; \theta_0)}{\partial \theta} dudv}_{\text{impact on the pdf of the joint distribution}} \end{aligned} \quad (22)$$

with the pdf of the conditional distribution of  $Y_{1t}$  given  $Y_{2t} = VaR_{2t}(\alpha)$ . An analytical expression for the derivative  $\partial F(\tilde{y}_t, \theta_0)/\partial \theta$  can be determined for some particular bivariate distributions (see appendix C for the case of the bivariate normal distribution). In the general case, it can be obtained by numerical differentiation.

**Corollary 2** *When there is no estimation risk, i.e. when  $\lambda = 0$ , under assumptions A1-A4,*

$$UC_{MES} \xrightarrow{d} \mathcal{N}(0, 1). \quad (23)$$

When the estimation period  $T$  is much larger than the evaluation period  $n$ , the unconditional coverage test is simplified since it does not require to evaluate  $R_{MES}$  and  $\Sigma_0$ .

Given these results, it is possible to define a robust test statistic, denoted  $UC_{MES}^C$ , that explicitly takes into account the estimation risk and that has standard limit distribution for any  $\lambda$  with  $0 \leq \lambda < \infty$ , when  $T$  and  $n$  tends to infinity. The feasible robust UC backtest statistics is

$$UC_{MES}^C = \frac{\sqrt{n} \left( \bar{H}(\alpha, \hat{\theta}_T) - \alpha/2 \right)}{\sqrt{\alpha(1/3 - \alpha/4) + n \hat{R}'_{MES} \hat{\mathbb{V}}_{as}(\hat{\theta}_T) \hat{R}_{MES}}} \quad (24)$$

with  $\hat{\mathbb{V}}_{as}(\hat{\theta}_T) = \hat{\Sigma}_0/T$  a consistent estimator of the asymptotic variance covariance matrix of  $\hat{\theta}_T$  and  $\hat{R}_{MES}$ , a consistent estimator of  $R_{MES}$  given by

$$\hat{R}_{MES} = -\frac{1}{\alpha n} \sum_{t=1}^n \frac{\partial F(\hat{y}_t, \hat{\theta}_T)}{\partial \theta} \mathbf{1}(y_{2t} \leq VaR_{2t}(\alpha, \hat{\theta}_T)). \quad (25)$$

with  $\hat{y}_t = (y_{1t}, VaR_{2t}(\alpha, \hat{\theta}_T))'$ .

## 4.2 Independence Test

To test the independence hypothesis  $H_{0,IND} : \rho_1 = \dots = \rho_m = 0$ , we use a Portmanteau Box-Pierce test on the sequence of cumulative joint violation forecasts. The Box-Pierce test statistic is defined as follows

$$IND_{MES} = n \sum_{j=1}^m \hat{\rho}_{nj}^2, \quad (26)$$

with  $\hat{\rho}_{nj}$  the sample autocorrelation of order  $j$  of the estimated cumulative joint violation  $H_t(\alpha, \hat{\theta}_T)$  given by

$$\hat{\rho}_{nj} = \frac{\hat{\gamma}_{nj}}{\hat{\gamma}_{n0}} \quad \text{and} \quad \hat{\gamma}_{nj} = \frac{1}{n-j} \sum_{t=1+j}^n \left( H_t(\alpha, \hat{\theta}_T) - \alpha/2 \right) \left( H_{t-j}(\alpha, \hat{\theta}_T) - \alpha/2 \right), \quad (27)$$

where  $\hat{\gamma}_{nj}$  denotes a consistent estimator of the  $j$ -lag autocovariance of  $H_t(\alpha, \hat{\theta}_T)$ . Theorem 3 gives the asymptotic distribution of the statistic  $IND_{MES}$  when  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda < \infty$ .

**Theorem 3** *Under assumptions A1-A4:*

$$IND_{MES} \xrightarrow{d} \sum_{j=1}^m \pi_j Z_j^2, \quad (28)$$

where  $\{\pi_j\}_{j=1}^m$  are the eigenvalues of the matrix  $\Delta$  with the  $ij$ -th element given by

$$\Delta_{ij} = \delta_{ij} + \lambda R_i' \Sigma_0 R_j, \quad (29)$$

$$R_j = \frac{1}{\alpha(1/3 - \alpha/4)} \mathbb{E}_0 \left( \left( H_{t-j}(\alpha, \theta_0) - \alpha/2 \right) \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right), \quad (30)$$

$\delta_{ij}$  is a dummy variable that takes a value 1 if  $i = j$  and 0 otherwise,  $\{Z_j\}_{j=1}^m$  are independent standard normal variables and  $\nabla_{as}(\hat{\theta}_T) = \Sigma_0/T$ .

The proof of Theorem 3 is reported in appendix D. The test statistic  $IND_{MES}$  has an asymptotic distribution which is a weighted sum of chi-squared variables. The weights depends on the asymptotic variance-covariance matrix of the estimator  $\hat{\theta}_T$ , on the cumulative joint violation process and on its derivative with respect to the model parameter  $\theta$ , as for the UC test. However, this limit distribution becomes standard when  $\lambda = 0$ , i.e. when there is no estimation risk.

**Corollary 4** *When there is no estimation risk, i.e. when  $\lambda = 0$ , under assumptions A1-A4,*

$$IND_{MES} \xrightarrow{d} \chi^2(m). \quad (31)$$

From the previous results, we can deduce a robust test statistics for the indendence hypothesis which has standard distribution for any  $\lambda$  with  $0 \leq \lambda < \infty$ , when  $T$  and  $n$  tends to infinity. Denote  $\hat{\rho}_n^{(m)}$  the vector  $(\hat{\rho}_{n1}, \dots, \hat{\rho}_{nm})'$ . The feasible robust IND backtest statistics is defined as

$$IND_{MES}^C = n \hat{\rho}_n^{(m)'} \hat{\Delta}^{-1} \hat{\rho}_n^{(m)} \quad (32)$$

where  $\hat{\Delta}$  is a consistent estimator for  $\Delta$ , such that

$$\hat{\Delta}_{ij} = \delta_{ij} + n \hat{R}_i' \hat{\nabla}_{as}(\hat{\theta}_T) \hat{R}_j, \quad (33)$$

$$\hat{R}_j = \frac{1}{\alpha(1/3 - \alpha/4)} \frac{1}{n-j} \sum_{t=j+1}^n \left( H_{t-j}(\alpha, \hat{\theta}_T) - \alpha/2 \right) \frac{\partial H_t(\alpha, \hat{\theta}_T)}{\partial \theta}, \quad (34)$$

and  $\hat{\nabla}_{as}(\hat{\theta}_T)$  is a consistent estimator of the asymptotoc variance covariance matrix of  $\hat{\theta}_T$ . When  $T$  and  $n$  tends to infinity, the robust statistic  $IND_{MES}^C$  converges to a chi-squared distribution with  $m$  degrees of freedom whatever the relative value of  $n$  and  $T$ .

## 5 Monte Carlo Simulation Study

This section assesses the finite sample properties of the test statistics, computed with and without taking into account the estimation risk. The first part describes the data generating process (DGP) used to build up the realistic setup, along with the algorithm required to construct the cumulative violation series. The second part is devoted to the analysis of the empirical size and power of the test.

### 5.1 Monte Carlo Design

In line with the current literature, we consider two definitions of MES to examine the finite sample properties of our test. First, we present MES as a conditional systemic risk measure (see Brownlees and Engle (2015)). Second, we deal with the particular case of a time-invariant MES (defined as in Acharya et al. (2010)). The DGP, as well as the size and power exercises, are drawn up in line with these aspects.

#### Conditional MES

For the dynamic case, the DGP consists in a dynamic conditional correlation (DCC) model computed on demeaned return processes  $Y_t = (Y_{1t} \ Y_{2t})'$ :

$$Y_t = \Sigma_t^{1/2} z_t, \quad (35)$$

where the 2x2 matrix  $\Sigma_t$  is the time-varying conditional covariance matrix of  $Y_t$  and  $z_t$  is an *i.i.d.* Gaussian vector error process such that  $\mathbb{E}[z_t] = 0$  and  $\mathbb{E}[z_t z_t'] = I_{(2)}$ . The *conditional* covariance matrix is defined as follows:

$$\Sigma_t = D_t R_t D_t, \quad (36)$$

where  $D_t = \text{diag}\{\sqrt{\sigma_{1,t}^2}, \sqrt{\sigma_{2,t}^2}\}$  contains conditional standard deviations of the  $Y_t$  system and  $R_t = \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix}$  is the *conditional* correlation matrix of  $Y_t$ . The conditional variances are most often modeled using the standard univariate GARCH(1,1) framework:

$$\sigma_{i,t}^2 = \alpha_{i,0} + \alpha_{i,1} Y_{i,t-1}^2 + \alpha_{i,2} \sigma_{i,t-1}^2, \quad \forall i = 1, 2$$

Furthermore, the time varying correlation matrix,  $R_t$ , that can be obtained as follows:  $\mathbb{E}_{t-1}(\varepsilon_t \varepsilon_t') = D_t^{-1} \Sigma_t D_t^{-1} = R_t$ , since  $\varepsilon_t = D_t^{-1} Y_t$ . We consider:

$$Q_t = (1 - a - b) \bar{R} + a \varepsilon_{t-1} \varepsilon_{t-1}' + b Q_{t-1},$$

where  $a$  and  $b$  are non-negative scalar parameters such that  $a + b < 1$ ,  $\bar{R}$  is the unconditional correlation matrix of the standardized errors  $\varepsilon_t$ , and  $Q_0$  is positive definite. The correlation matrix is subsequently obtained by rescaling  $Q_t$ , such as:  $R_t = (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2}$ .

For a more realistic scenario we use the parameter estimates obtained from the GARCH(1,1)-DCC(1,1) model on real log-returns series (*i.e.*, JP Morgan Chase Co. and S&P500 index from January 1st, 2005 to October 9th, 2015).<sup>3</sup>

Based on this dynamic specification, we compute the conditional *MES* as follows :

$$\begin{aligned} MES_t &= E_{t-1}[Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)] \\ &= \int_{\mathbb{R}} x f_{Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)}(x; \theta_0^c | \Omega_{t-1}) dx \end{aligned} \quad (37)$$

where  $f_{Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)}(x; \theta_0^c | \Omega_{t-1})$  is the pdf of  $Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)$ ,  $\theta_0^c$  is the set of true parameters, and  $\Omega_{t-1}$  is the known information set at time  $t - 1$ , so that  $MES_t$  depends on past information. The vector of parameters is estimated at each iteration using the following two-step procedure: (*i*) first, we estimate the univariate GARCH parameters  $(\alpha_{1,0}, \alpha_{1,1}, \alpha_{1,2})$ ,  $(\alpha_{2,0}, \alpha_{2,1}, \alpha_{2,2})$  by CMLE; (*ii*) second, given the estimated parameters from step one, we adjust the DCC parameters  $(a, b)$  based on a bivariate Normal distribution.

### Unconditional MES

A natural extension of the conditional specification is represented by the time invariant representation, given by:

$$Y_t = \Sigma^{1/2} z_t \quad (38)$$

where the 2x2 matrix  $\Sigma$  is the non time-varying covariance matrix of  $Y_t$  and  $z_t$  is the same error term than in dynamic case. The unconditional covariance matrix is defined as follows :

$$\Sigma = DRD \quad (39)$$

where  $D_t = \text{diag}\{\sqrt{\sigma_1^2}, \sqrt{\sigma_2^2}\}$  contains unconditional standard deviations of the  $Y_t$  system and  $R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  is the unconditional correlation matrix of  $Y_t$ .

In line with the previous specification, the parameters of this specific static case correspond with the unconditional counterpart of the GARCH(1,1)-DCC(1,1) model parameters.<sup>4</sup>

The unconditional *MES* associated with this static specification is computed as follows:

$$\begin{aligned} MES &= E[Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)] \\ &= \int_{\mathbb{R}} x f_{Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)}(x; \theta_0^u) dx \end{aligned} \quad (40)$$

where  $f_{Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)}(x; \theta_0^u)$  is the pdf of  $Y_{1t}|Y_{2t} \leq VaR_{2t}(\alpha)$ , and  $\theta_0^u$  is the vector of parameters aggregating all the parameters of this unconditional specification. At each replication, we estimate  $\theta_0^u$  by MLE using the proprieties of the bivariate Normal distribution.

The Monte Carlo simulation exercise is based on 10,000 replications, and we consider in-sample sizes  $T = 250, 2500$  and out-of-sample sizes  $N = 250, 500$ . The coverage levels are set to  $\alpha = 0.01, 0.05$ .

<sup>3</sup>GARCH(1,1) parameters for the conditional variances of the first and second asset,  $(\alpha_{1,0}, \alpha_{1,1}, \alpha_{1,2})$ ,  $(\alpha_{2,0}, \alpha_{2,1}, \alpha_{2,2})$ , are set to (0.02893, 0.09696, 0.90053) and (0.02100, 0.10346, 0.87903). DCC(1,1) parameters for the conditional correlation between the two series,  $(a, b)$ , are set to (0.03640, 0.91189), and  $\bar{R} = \begin{bmatrix} 1 & 0.74826 \\ 0.74826 & 1 \end{bmatrix}$ .

<sup>4</sup>Thus,  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\rho$  match the unconditional variances and the unconditional correlation of our dynamic specification. We set hence  $(\sigma_1^2, \sigma_2^2, \rho)$  to (11.50177, 1.19961, 0.74826).

### 5.1.1 Size and Power

The empirical size and power of the test are assessed within the two configurations previously presented. Note that when studying the power, two types of misspecification are considered under the alternative: (*i*) on the volatility dynamics; (*ii*) on the correlation dynamics. For instance, in the  $H_1(A)$  and  $H_1(B)$  frameworks, the volatility of  $Y_{1,t}$  (*i.e.*, the firm) and  $Y_{2,t}$  (*i.e.*, the market), respectively, is undervalued. Such underestimation of the riskiness on the firm/market losses might wrongly reduce the estimated MES. In the  $H_1(C)$  structure, the misspecification occurs in the dependence between the two series. We focus hence on a scenario in which the correlation between the firm and the market is undervalued. This can imply a strong undervaluation of MES when the market is in times of distress.

Consider first the conditional MES case. The DGP structure of  $Y_t$  under the null and the alternative is given by:

$H_0$  : GARCH(1,1)-DCC(1,1) model:

$$\begin{aligned} Y_t &= \Sigma_t^{1/2} z_t \\ \Sigma_t &= D_t R_t D_t \\ z_t &\sim i.i.d. N(0, I) \end{aligned}$$

$H_1(A)$  : GARCH(1,1)-DCC(1,1) model with undervalued variance of the firm ( $Y_{1,t}$ ):

$$\begin{aligned} Y_t &= \Sigma_t^{1/2} z_t \\ \sigma_{1,t}^2 &= \alpha_{1,0}^{H_1(A)} + \alpha_{1,1} Y_{1,t-1}^2 + \alpha_{1,2} \sigma_{1,t-1}^2 \end{aligned}$$

with  $\alpha_{1,0}^{H_1(A)} = 25\% \times \alpha_{1,0}, 50\% \times \alpha_{1,0}, 75\% \times \alpha_{1,0}$ , successively.

$H_1(B)$  : GARCH(1,1)-DCC(1,1) model with undervalued variance for the market ( $Y_{2,t}$ ):

$$\begin{aligned} Y_t &= \Sigma_t^{1/2} z_t \\ \sigma_{2,t}^2 &= \alpha_{2,0}^{H_1(B)} + \alpha_{2,1} Y_{2,t-1}^2 + \alpha_{2,2} \sigma_{2,t-1}^2 \end{aligned}$$

with  $\alpha_{2,0}^{H_1(B)} = 25\% \times \alpha_{2,0}, 50\% \times \alpha_{2,0}, 75\% \times \alpha_{2,0}$ , successively.

$H_1(C)$  : GARCH(1,1)-DCC(1,1) model with undervalued correlation :

$$\begin{aligned} Y_t &= \Sigma_t^{1/2} z_t \\ \Sigma_t &= D_t R D_t \\ \bar{R} &= \begin{bmatrix} 1 & \bar{\rho}^{H_1(C)} \\ \bar{\rho}^{H_1(C)} & 1 \end{bmatrix} \end{aligned}$$

with  $\bar{\rho}^{H_1(C)} = 0, 0.3, 0.6$ , successively.

Second, in the unconditional MES case, the DGP structure of  $Y_t$  under the null and the alternative is given by:

$H_0$  : Static model:

$$\begin{aligned} Y_t &= \Sigma^{1/2} z_t, \\ \Sigma &= DRD \\ z_t &\sim i.i.d. N(0, I) \end{aligned}$$

$H_1(A)$  : Static model with undervalued variance of  $Y_{1,t}$ :

$$\begin{aligned} Y_t &= \Sigma^{1/2} z_t \\ \Sigma &= \begin{bmatrix} \sigma_1^{2, H_1(A)} & \rho \sigma_1^{H_1(A)} \sigma_2 \\ \rho \sigma_1^{H_1(A)} \sigma_2 & \sigma_2^2 \end{bmatrix} \end{aligned}$$

with  $\sigma_1^{2, H_1(A)} = 25\% \times \sigma_1^2, 50\% \times \sigma_1^2, 75\% \times \sigma_1^2$ , successively.

$H_1(B)$  : Static model with undervalued variance of  $Y_{2,t}$ :

$$\begin{aligned} Y_t &= \Sigma^{1/2} z_t \\ \Sigma &= \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2^{H_1(B)} \\ \rho \sigma_1 \sigma_2^{H_1(B)} & \sigma_2^{2, H_1(B)} \end{bmatrix} \end{aligned}$$

with  $\sigma_2^{2, H_1(B)} = 25\% \times \sigma_2^2, 50\% \times \sigma_2^2, 75\% \times \sigma_2^2$ , successively.

$H_1(C)$  : Static model with undervalued correlation :

$$\begin{aligned} Y_t &= \Sigma^{1/2} z_t \\ \Sigma &= \begin{bmatrix} \sigma_1^2 & \rho^{H_1(C)} \sigma_1 \sigma_2 \\ \rho^{H_1(C)} \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \end{aligned}$$

with  $\rho^{H_1(C)} = 0, 0.3, 0.6$ , successively.

## 6 Backtesting other Systemic Risk Measures

Other systemic risk measures can be written as a function of the MES or the CoVaR, and therefore, can be backtested according to our methodology. It is especially the case for the SES (Acharya et al. 2010) and the SRISK (Acharya et al. 2012 and Brownlees and Engle 2015) which extend the MES in order to take into account both the liabilities and the size of the financial institution. It is also the case for the  $\Delta$ CoVaR (Adrian and Brunnermeier 2016) which is fundamentally based on the difference of two conditional VaR.

## 6.1 Backtesting MES-based Systemic Risk Measures

The SRISK corresponds to the expected capital shortfall of the financial institution, conditional on a crisis affecting the whole financial system. Brownlees and Engle (2015) define the capital shortfall ( $CS_{it}$ ) as the capital reserves the firm needs to hold for regulation or prudential management minus the firm's equity. Then, we have

$$CS_{1t-1} = k(L_{1t-1} + W_{1t-1}) - W_{1t-1}, \quad (41)$$

with  $L_{1t}$  the book value of debt,  $W_{1t}$  the market value of the firm's equity and  $k$  a prudential ratio. As a consequence, if the financial system crisis is defined as a situation in which  $Y_{2t} < VaR_{2t}(\alpha)$ , the SRISK become

$$SRISK_{1t} = \mathbb{E}_{t-1}(CS_{1t} | Y_{2t} < VaR_{2t}(\alpha)) \quad (42)$$

$$= k\mathbb{E}_{t-1}(L_{1t} | Y_{2t} < VaR_{2t}(\alpha)) - (1-k)\mathbb{E}_{t-1}(W_{1t} | Y_{2t} < VaR_{2t}(\alpha)) \quad (43)$$

Brownlees and Engle assume that the debt is constant, i.e.  $\mathbb{E}_{t-1}(L_{1t} | Y_{2t} < VaR_{2t}(\alpha)) = L_{1t-1}$ . Since  $\mathbb{E}_{t-1}(W_{1t} | Y_{2t} < VaR_{2t}(\alpha)) = W_{1t-1}(1 + \mathbb{E}_{t-1}(Y_{1t} | Y_{2t} < VaR_{2t}(\alpha)))$ , then we get

$$SRISK_{1t} = k L_{1t-1} - (1-k) W_{1t-1} MES_{1t}(\alpha) \quad (44)$$

Similarly, we can define a linear relationship between the SES (Acharya et al. 2010) and the MES. Indeed, the SES corresponds to the amount a bank's equity drops below its target level (defined as a fraction  $k$  of assets) in case of a systemic crisis when aggregate capital is less than  $k$  times aggregate assets. Acharya et al. (2010) show that the SES of bank  $i$  can be expressed as linear function of its MES, with:

$$SES_{it} = (k LV_{it} - 1 + \theta MES_{it}(\alpha) + \Delta) W_{it-1}, \quad (45)$$

where  $\theta$  and  $\Delta$  are constant terms,  $W_{it}$  the market capitalization or market value of equity, and  $L_{it} = (L_{it-1} + W_{it-1})/W_{it-1}$  the leverage.

Within this two examples, a MES-based systemic risk measure for the firm  $i$  at time  $t$ , denoted  $RM_{it}$ , can be defined as a deterministic function of the  $MES_{it}(\alpha)$  given  $\Omega_{t-1}$ , such as.

$$RM_{it} = g_t(MES_{it}(\alpha), X_{i,t-1}), \quad (46)$$

with  $g_t(\cdot)$  an decreasing function (with MES) and  $X_{t-1}$  a set of variables that belong to  $\Omega_{t-1}$ . Define a new joint violation process  $h_t(\alpha, \beta, \theta_0, X_{t-1})$  such that

$$\begin{aligned} h_t(\alpha, \beta, \theta_0, X_{t-1}) &= (1 - \mathbf{1}(g_t(Y_{1t}, X_{t-1}) \leq g_t(CoVaR_{1t}(\beta, \alpha, \theta_0), X_{t-1}))) \\ &\quad \times \mathbf{1}(Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)), \end{aligned} \quad (47)$$

and a corresponding cumulative joint violation process

$$H_t(\alpha, \theta_0, X_{t-1}) = \int_0^1 h_t(\alpha, \beta, \theta_0, X_{t-1}) d\beta. \quad (48)$$

For these systemic risk measures, we consider the following test

$$H_{0,RM} : \mathbb{E}(H_t(\alpha, \theta_0, X_{t-1})) = \frac{\alpha}{2}.$$

The test statistic is then similar to that used for the MES test

$$UC_{RM} = \frac{\sqrt{n} \left( \bar{H}(\alpha, \hat{\theta}_T) - \alpha/2 \right)}{\sqrt{\alpha(1/3 - \alpha/4)}},$$

with

$$\bar{H}(\alpha, \hat{\theta}_T) = \frac{1}{n} \sum_{t=1}^n H_t(\alpha, \hat{\theta}_T, X_{t-1}).$$

When there is no estimation risk, i.e. when  $\lambda = 0$ , under assumptions A1-A4,

$$UC_{RM} \xrightarrow{d} \mathcal{N}(0, 1). \quad (49)$$

When  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda < \infty$ , test statistic  $UC_{RM}$  has a non standard asymptotic distribution. However, by using the same approach as in Section 4, it is possible to derive a robust test statistic that has a standard normal distribution whatever the relative value of  $n$  and  $T$ .

## 6.2 Backtesting $\Delta\text{CoVaR}$

Adrian and Brunnermeier (2016) propose a measure for systemic risk, called the  $\Delta\text{CoVaR}$ , which is defined as the difference between the conditional VaR (CoVaR) of the financial system conditional on an institution being in distress and the CoVaR conditional on the median state of the institution. Contrary to the MES case, denote by  $Y_{1t}$  the return of a portfolio of financial institutions and by  $Y_{2t}$  the stock return of a given institution. Adrian and Brunnermeier define the stress of the institution as a case in which the return  $Y_{2t}$  is equal to  $VaR_{2t}(\alpha)$ . This choice is justified by their estimation methodology based on a (quantile) regression model. A more general approach consists in defining the financial stress as a situation in which  $Y_{2t} \leq VaR_{2t}(\alpha)$  and in using truncated distributions (Girardi and Ergun, 2013). Similarly, the median state of the institution can be represented by an interquartile range,  $VaR_{2t}(\beta_{\text{inf}}) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}})$ , with for instance  $\beta_{\text{inf}} = 25\%$  and  $\beta_{\text{sup}} = 75\%$ . Then, the  $\Delta\text{CoVaR}$  of the financial system corresponds to the quantity

$$\Delta\text{CoVaR}_{1t}(\alpha) = \text{CoVaR}_{1t}(\alpha, \alpha, \theta_0) - \text{CoVaR}_{1t}(\alpha, \tilde{\beta}, \theta_0), \quad (50)$$

with  $\tilde{\beta} = (\beta_{\text{inf}}, \beta_{\text{sup}})'$  and

$$\Pr(Y_{1t} \leq \text{CoVaR}_{1t}(\alpha, \alpha, \theta_0) | Y_{2t} \leq VaR_{2t}(\alpha, \theta_0); \Omega_{t-1}) = \alpha,$$

$$\Pr(Y_{1t} \leq \text{CoVaR}_{1t}(\alpha, \tilde{\beta}, \theta_0) | VaR_{2t}(\beta_{\text{inf}}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}}, \theta_0); \Omega_{t-1}) = \alpha. \quad (51)$$

In order to backtest each of the two CoVaRs, we define two violations. These violation processes are represented by the following binary variables

$$h_t(\alpha, \alpha, \theta_0) = \mathbf{1}((Y_{1t} \leq \text{CoVaR}_{1t}(\alpha, \alpha, \theta_0)) \cap (Y_{2t} \leq VaR_{2t}(\alpha, \theta_0))),$$



$$h_t(\alpha, \tilde{\beta}, \theta_0) = \mathbf{1} \left( \left( Y_{1t} \leq \text{CoVaR}_{1t}(\alpha, \tilde{\beta}, \theta_0) \right) \cap \left( \text{VaR}_{2t}(\beta_{\text{inf}}, \theta_0) \leq Y_{2t} \leq \text{VaR}_{2t}(\beta_{\text{sup}}, \theta_0) \right) \right).$$

If the risk model is well specified, the two centered violation processes satisfy

$$\mathbb{E} \left( h_t(\alpha, \alpha, \theta_0) - \alpha^2 \mid \Omega_{t-1} \right) = 0 \quad (52)$$

$$\mathbb{E} \left( h_t(\alpha, \tilde{\beta}, \theta_0) - \alpha (\beta_{\text{sup}} - \beta_{\text{inf}}) \mid \Omega_{t-1} \right) = 0 \quad (53)$$

The intuition of the test is then similar to the backtests proposed for multi-level VaR (see Francq and Zakoian, 2015 for the estimation method), i.e. the VaR defined for a finite set of coverage rates (Hurlin and Tokpavi 2006, Pérignon and Smith 2008, Leccadito, Boffelli and Urga 2014). It consists in considering a *vector* a violations, denoted  $\mathbf{h}_t(\alpha, \tilde{\beta}, \theta_0)$ , with

$$\mathbf{h}_t(\alpha, \tilde{\beta}, \theta_0) = \left( h_t(\alpha, \alpha, \theta_0), h_t(\alpha, \tilde{\beta}, \theta_0) \right)'. \quad (54)$$

Thus, a test for the validity of the two components of the  $\Delta\text{CoVaR}$  can be expressed as

$$H_{0,UC} : \mathbb{E} \left( \mathbf{h}_t(\alpha, \tilde{\beta}, \theta_0) \right) = \mu_{\mathbf{h}} = \left( \alpha^2, \alpha (\beta_{\text{sup}} - \beta_{\text{inf}}) \right)'. \quad (55)$$

The corresponding test statistic is defined as

$$UC_{\Delta\text{CoVaR}} = n \left( \bar{\mathbf{h}}(\alpha, \tilde{\beta}, \hat{\theta}_T) - \mu_{\mathbf{h}} \right)' \hat{\mathbb{V}}(\bar{\mathbf{h}}(\alpha, \tilde{\beta}, \hat{\theta}_T))^{-1} \left( \bar{\mathbf{h}}(\alpha, \tilde{\beta}, \hat{\theta}_T) - \mu_{\mathbf{h}} \right), \quad (56)$$

with  $\bar{\mathbf{h}}(\alpha, \tilde{\beta}, \hat{\theta}_T)$  the out-of-sample mean of  $\mathbf{h}_t(\alpha, \tilde{\beta}, \hat{\theta}_T)$ , and  $\hat{\mathbb{V}}(\bar{\mathbf{h}}(\alpha, \tilde{\beta}, \hat{\theta}_T))$  a consistent estimator of the variance-covariance matrix of  $\bar{\mathbf{h}}(\alpha, \tilde{\beta}, \theta_0)$ . Notice that as soon as  $\alpha < \beta_{\text{inf}}$ , this matrix is diagonal since the covariance between the two violations  $\mathbf{h}_t(\alpha, \alpha, \theta_0)$  and  $\mathbf{h}_t(\alpha, \tilde{\beta}, \theta_0)$  is null.

When there is no estimation risk, i.e. when  $\lambda = 0$ , under assumptions A1-A4,

$$UC_{\Delta\text{CoVaR}} \xrightarrow{d} \chi^2(2). \quad (57)$$

When  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda < \infty$ , test statistic  $UC_{\Delta\text{CoVaR}}$  has an asymptotic distribution which is a weighted sum of chi-squared variables. However, by using the same approach as in Section 4, it is possible to derive a robust test statistic that has a chi-squared distribution with 2 degrees of freedom whatever the relative value of  $n$  and  $T$ .

## 7 Empirical Application

## 8 Conclusion

This paper develops a backtesting procedure for systemic risk measures. These tests are based on the concept of cumulative joint violation process. This original approach has many advantages. First, its implementation is easy since it only implies to evaluate a cdf of a bivariate distribution. Second, it allows for a separate test for unconditional coverage and independence hypothesis (Christoffersen, 1998). Third, Monte-Carlo simulations show that for realistic sample sizes, our tests have good finite sample properties. Finally, we pay a particular attention the consequences of the estimation risk and propose a robust version of our test statistics.

**Table 1.** Empirical rejection rates for backtesting MES(0.05) at 5% significance level

		$UC_{MES}(\widehat{\theta}_T)$	$UC_{MES}^C(\widehat{\theta}_T)$	$IND_{MES}(\widehat{\theta}_T)$	$IND_{MES}^C(\widehat{\theta}_T)$
$T = 250, n = 250$ , Size and Power (size corrected)					
$H_0$	—	0.089	0.047	0.094	0.075
$H_1^A$	$\Delta\sigma_1^2 = 25\%$	0.308	0.351	0.030	0.040
	$\Delta\sigma_1^2 = 50\%$	0.834	0.866	0.051	0.052
$H_1^B$	$\Delta\sigma_2^2 = 25\%$	0.382	0.452	0.035	0.037
	$\Delta\sigma_2^2 = 50\%$	0.984	0.989	0.080	0.053
$H_1^C$	$\rho_{H_1} = 0.6$	0.346	0.383	0.038	0.037
	$\rho_{H_1} = 0.3$	0.714	0.766	0.041	0.045
$T = 2500, n = 2500$ , Size and Power (size corrected)					
$H_0$	—	0.091	0.051	0.059	0.057
$H_1^A$	$\Delta\sigma_1^2 = 25\%$	0.975	0.981	0.085	0.085
	$\Delta\sigma_1^2 = 50\%$	1.000	1.000	0.472	0.458
$H_1^B$	$\Delta\sigma_2^2 = 25\%$	0.999	1.000	0.111	0.110
	$\Delta\sigma_2^2 = 50\%$	1.000	1.000	0.931	0.922
$H_1^C$	$\rho_{H_1} = 0.6$	0.994	0.995	0.095	0.092
	$\rho_{H_1} = 0.3$	1.000	1.000	0.223	0.215

**Table 2.** Empirical rejection rates for backtesting MES(0.05) at 5% significance level

		$UC_{MES}(\widehat{\theta}_T)$	$UC_{MES}^C(\widehat{\theta}_T)$	$IND_{MES}(\widehat{\theta}_T)$	$IND_{MES}^C(\widehat{\theta}_T)$
$T = 2500, n = 250, \text{ Size and Power (size corrected)}$					
$H_0$	—	0.059	0.054	0.099	0.094
$H_1^A$	$\Delta\sigma_1^2 = 25\%$	0.357	0.362	0.039	0.038
	$\Delta\sigma_1^2 = 50\%$	0.891	0.895	0.047	0.048
$H_1^B$	$\Delta\sigma_2^2 = 25\%$	0.467	0.473	0.044	0.045
	$\Delta\sigma_2^2 = 50\%$	0.986	0.988	0.099	0.100
$H_1^C$	$\rho_{H_1} = 0.6$	0.448	0.454	0.033	0.035
	$\rho_{H_1} = 0.3$	0.798	0.802	0.053	0.056
$T = 250, n = 2500, \text{ Size and Power (size corrected)}$					
$H_0$	—	0.312	0.045	0.076	0.056
$H_1^A$	$\Delta\sigma_1^2 = 25\%$	0.641	0.867	0.069	0.071
	$\Delta\sigma_1^2 = 50\%$	0.999	1.000	0.416	0.317
$H_1^B$	$\Delta\sigma_2^2 = 25\%$	0.899	0.997	0.077	0.084
	$\Delta\sigma_2^2 = 50\%$	1.000	1.000	0.910	0.774
$H_1^C$	$\rho_{H_1} = 0.6$	0.773	0.901	0.087	0.075
	$\rho_{H_1} = 0.3$	0.989	0.999	0.251	0.208

## A Moments of the Cumulative Joint Violation Process

Let us define the binary variable  $H_t(\alpha, \theta_0) = (1 - u_{12t}(\theta_0)) \mathbf{1}(u_{2t}(\theta_0) \leq \alpha)$  with  $u_{2t}(\theta_0) \equiv u_{2t} = F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0)$  and  $u_{12t}(\theta_0) \equiv u_{12t} = F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0)$ . The two first conditional moments of the cumulative joint process  $H_t(\alpha, \theta_0)$  are then given by

$$\begin{aligned} \mathbb{E}(H_t(\alpha, \theta_0) | \Omega_{t-1}) &= \Pr(u_{2t}(\theta_0) \leq \alpha | \Omega_{t-1}) \times \mathbb{E}(H_t(\alpha, \theta_0) | u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}) \\ &= \alpha - \alpha \mathbb{E}(u_{12t}(\theta_0) | u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}), \end{aligned}$$

$$\mathbb{E}(H_t^2(\alpha, \theta_0) | \Omega_{t-1}) = \alpha \mathbb{E}(1 - 2u_{12t}(\theta_0) + u_{12t}^2(\theta_0) | u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}).$$

Since the conditional distribution of  $u_{2t}(\theta_0)$  given  $\Omega_{t-1}$  is  $U_{[0,1]}$  with  $\mathbb{E}(u_{2t}(\theta_0) | \Omega_{t-1}) = 1/2$  and  $\mathbb{E}(u_{2t}^2(\theta_0) | \Omega_{t-1}) = 1/3$ , then we get

$$\mathbb{E}(H_t(\alpha, \theta_0) | \Omega_{t-1}) = \frac{\alpha}{2}, \quad \mathbb{V}(H_t(\alpha, \theta_0) | \Omega_{t-1}) = \alpha \left( \frac{1}{3} - \frac{\alpha}{4} \right). \quad (58)$$

## B Proof of Theorem 1

To derive the asymptotic properties of the statistic  $CC_{MES}$ , we introduce the following assumptions.

**A1:** The vectorial process  $Y_t = (Y_{1t}, Y_{2t})'$  is strictly stationary and ergodic.

**A2:** The marginal distribution of  $Y_{2t}$  is given by  $F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0)$  and the truncated distribution of  $Y_{1t}$  given  $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$  is given by  $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0)$ .

**A3:**  $\theta_0 \in \Theta$ , with  $\Theta$  a compact subspace of  $\mathbb{R}^p$ .

**A4:** The estimator  $\hat{\theta}_T$  is consistent for  $\theta_0$  and is asymptotically normally distributed such that:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_0),$$

with  $\Sigma_0$  a positive definite  $p \times p$  matrix. Denote  $\mathbb{V}_{as}(\hat{\theta}_T) = \Sigma_0/T$ .

**Proof.** Denote  $H_t(\alpha, \theta) = (1 - u_{12t}(\theta)) \mathbf{1}(u_{2t}(\theta) \leq \alpha)$  the cumulative violation process, with  $u_{2t}(\theta) = F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta)$  and  $u_{12t}(\theta) = F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \theta)}(Y_{1t}; \Omega_{t-1}, \theta)$ ,  $\forall t = 1, \dots, n$  and  $\forall \theta \in \Theta$ . Under the null hypothesis  $H_{0,UC}$ , the sequence  $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^n$  is a *mds* with  $\sigma_H^2 = \mathbb{V}(H_t(\alpha, \theta_0)) = \alpha(1/3 - \alpha/4)$ . For simplicity, we assume that  $\Omega_{t-1}$  only includes a finite number of lagged values of  $Y_t$ , i.e. there is no information truncation. The test statistic  $UC_{ES}$  can be rewritten as

$$UC_{MES} = \frac{1}{\sigma_H \sqrt{n}} \sum_{t=1}^n \left( H_t(\alpha, \hat{\theta}_T) - \alpha/2 \right). \quad (59)$$

Under assumptions A1-A4, the continuous mapping theorem implies that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left( H_t(\alpha, \hat{\theta}_T) - \mathbb{E} \left( H_t(\alpha, \hat{\theta}_T) \middle| \Omega_{t-1} \right) \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( H_t(\alpha, \theta_0) - \mathbb{E}(H_t(\alpha, \theta_0) | \Omega_{t-1}) \right) + o_p(1).$$

Rearranging these terms gives

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( H_t(\alpha, \hat{\theta}_T) - \mathbb{E}(H_t(\alpha, \theta_0) | \Omega_{t-1}) \right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (H_t(\alpha, \theta_0) - \mathbb{E}(H_t(\alpha, \theta_0) | \Omega_{t-1})) \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{E} \left( \left( H_t(\alpha, \hat{\theta}_T) - H_t(\alpha, \theta_0) \right) \middle| \Omega_{t-1} \right) + o_p(1). \end{aligned} \quad (60)$$

The mean value theorem implies that

$$H_t(\alpha, \hat{\theta}_T) = H_t(\alpha, \theta_0) + \left( \hat{\theta}_T - \theta_0 \right)' \frac{\partial H_t(\alpha, \tilde{\theta})}{\partial \theta}, \quad (61)$$

where  $\tilde{\theta}$  is an intermediate point between  $\theta_0$  and  $\hat{\theta}_T$ . Equation (60) becomes

$$\begin{aligned} \frac{1}{\sigma_H \sqrt{n}} \sum_{t=1}^n \left( H_t(\alpha, \hat{\theta}_T) - \alpha/2 \right) &= \frac{1}{\sigma_H \sqrt{n}} \sum_{t=1}^n (H_t(\alpha, \theta_0) - \alpha/2) \\ &+ \frac{\sqrt{\lambda}}{\sigma_H} \sqrt{T} (\hat{\theta}_T - \theta_0)' \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left( \frac{\partial H_t(\alpha, \tilde{\theta})}{\partial \theta} \middle| \Omega_{t-1} \right) + o_p(1). \end{aligned} \quad (62)$$

Assume that  $T \rightarrow \infty$ ,  $n \rightarrow \infty$  and  $n/T \rightarrow \lambda$  with  $0 \leq \lambda < \infty$ . Under the null hypothesis  $H_{0,UC}$ , the first term on the right hand converges in distribution to a standard normal distribution. The covariance between the first term and  $\sqrt{T}(\hat{\theta}_T - \theta_0)$  is 0 as  $\hat{\theta}_T$  depends on the in-sample observations and the summand in the first term is for out-of-sample observations. Under assumption A4,  $\tilde{\theta} \xrightarrow{p} \theta_0$  and since  $\partial H_t(\alpha, \theta_0) / \partial \theta$  is also a *mds*, we have

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E} \left( \frac{\partial H_t(\alpha, \tilde{\theta})}{\partial \theta} \middle| \Omega_{t-1} \right) \xrightarrow{p} R_{MES} = \mathbb{E}_0 \left( \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right), \quad (63)$$

where  $\mathbb{E}_0(\cdot)$  denotes the expectation with respect to the true distribution of  $H_t(\alpha, \theta_0)$ . So, we get

$$\frac{\sqrt{\lambda}}{\sigma_H} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left( \frac{\partial H_t(\alpha, \tilde{\theta})}{\partial \theta} \middle| \Omega_{t-1} \right)' \sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\lambda}{\sigma_H^2} R'_{MES} \Sigma_0 R_{MES} \right). \quad (64)$$

and finally

$$UC_{MES} \xrightarrow{d} \mathcal{N} \left( 0, 1 + \lambda \frac{R'_{MES} \Sigma_0 R_{MES}}{\alpha (1/3 - \alpha/4)} \right). \quad (65)$$

■

## C Computation of $R_{MES}$ in the Bivariate Normal Case

Let us assume that  $Y_t = (Y_{1t}, Y_{2t})'$  such that

$$Y_t = \Sigma_t^{1/2} \varepsilon_t \quad (66)$$

where  $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$  are i.i.d.  $\mathcal{N}(0, I_2)$ , where  $I_2$  denotes the  $2 \times 2$  identity matrix and  $\Sigma_t = \Sigma_t(\theta_0)$  is the conditional variance covariance matrix of  $Y_t$  given  $\Omega_{t-1}$ . Denote by  $f(y, \Sigma_t) \equiv f(y_1, y_2, \Sigma_t)$  the pdf and by  $F(y, \Sigma_t) \equiv F(y_1, y_2, \Sigma_t)$  the cdf of the joint distribution of  $Y_t$  such that

$$f(y, \Sigma_t) = \frac{1}{2\pi} |\Sigma_t|^{-\frac{1}{2}} \exp \left( -\frac{y' \Sigma_t^{-1} y}{2} \right) \quad (67)$$

$$F(y, \Sigma_t) = \Pr((Y_1 \leq y_1) \cap (Y_2 \leq y_2)) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) da db \quad (68)$$

Using the Dwyer's formula (1967), we know that

$$\frac{\partial f(y, \Sigma_t)}{\partial \Sigma_t} = \frac{\partial \ln f(y, \Sigma_t)}{\partial \Sigma_t} \times f(y, \Sigma_t) = -\frac{f(y, \Sigma_t)}{2} (\Sigma_t^{-1} - \Sigma_t^{-1} y y' \Sigma_t^{-1}) \quad (69)$$

and we get

$$\begin{aligned} \frac{\partial F(y, \Sigma_t)}{\partial \Sigma_t} &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \frac{\partial f(a, b, \Sigma_t)}{\partial \Sigma_t} da db \\ &= -\frac{\Sigma_t^{-1}}{2} F(y, \Sigma_t) + \frac{1}{2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) \Sigma_t^{-1} \Delta(a, b) \Sigma_t^{-1} da db \end{aligned} \quad (70)$$

with  $\Delta(a, b)$  a  $2 \times 2$  matrix equal to  $(a, b)' \times (a, b)$ . If the conditional variance covariance matrix is defined as

$$\Sigma_t = \begin{pmatrix} \sigma_{1t}^2 & \sigma_{12t} \\ \sigma_{12t} & \sigma_{2t}^2 \end{pmatrix} \quad (71)$$

the matrix expression given in equation (70) can be decomposed as follows

$$\begin{aligned} \frac{\partial F(y, \Sigma_t)}{\partial \sigma_{1t}^2} &= -\frac{\sigma_{2t}^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} (\sigma_{2t}^2 a^2 - 2\sigma_{12t}\sigma_{2t}^2 ab + \sigma_{12t}^2 b^2) f(a, b, \Sigma_t) da db \\ \frac{\partial F(y, \Sigma_t)}{\partial \sigma_{2t}^2} &= -\frac{\sigma_{1t}^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} (\sigma_{12t}^2 a^2 - 2\sigma_{12t}\sigma_{1t}^2 ab + \sigma_{1t}^2 b^2) f(a, b, \Sigma_t) da db \\ \frac{\partial F(y, \Sigma_t)}{\partial \sigma_{12t}} &= -\frac{\sigma_{12t}}{2\Delta_t} F(y, \Sigma_t) \\ &\quad + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} (-\sigma_{2t}^2 \sigma_{12t} a^2 + (\sigma_{12t}^2 + \sigma_{1t}^2 \sigma_{2t}^2) ab - \sigma_{1t}^2 \sigma_{12t} b^2) f(a, b, \Sigma_t) da db \end{aligned}$$

with  $\Delta_t = \sigma_{1t}^2 \sigma_{2t}^2 - \sigma_{12t}^2$ . If  $\theta$  denotes the parameters vector of the conditional variance-covariance model (CCC, DCC, etc.), then for any vector  $\theta \in \Theta$ , we have

$$\frac{\partial F(y, \Sigma_t)}{\partial \theta} = \left( \text{vec} \left( \frac{\partial F(y, \Sigma_t)}{\partial \Sigma_t} \right) \right)' \text{vec} \left( \frac{\partial \Sigma_t}{\partial \theta} \right) \quad (72)$$

where  $\text{vec}(\cdot)$  denotes the vectorization operator of a matrix. A consistent estimate of  $R_{MES}$  is given by

$$\widehat{R}_{MES} = -\frac{1}{\alpha n} \sum_{t=1}^n \left( \text{vec} \left( \frac{\partial F(\tilde{y}_t, \Sigma_t)}{\partial \Sigma_t} \Big|_{\widehat{\theta}_T} \right) \right)' \frac{\partial \text{vec}(\Sigma_t)}{\partial \theta} \Big|_{\widehat{\theta}_T} \quad (73)$$

## D Proof of Theorem 3

**Proof.** Define the lag- $j$  autocovariance and autocorrelation of the cumulative joint violation  $H_t(\alpha, \theta_0)$  for  $j \geq 0$  by

$$\rho_j = \frac{\gamma_j}{\gamma_0} \quad \text{and} \quad \gamma_j = \text{Cov}(H_t(\alpha, \theta_0), H_{t-j}(\alpha, \theta_0)). \quad (74)$$

We drop the dependence of  $\gamma_j$  and  $\rho_j$  on  $\alpha$  and  $\theta_0$  for simplicity of notation. The sample counterparts of  $\gamma_j$  and  $\rho_j$  based on the sample  $\{H_t(\alpha, \theta_0)\}_{t=1}^n$  are

$$\rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}} \quad \text{and} \quad \gamma_{nj} = \frac{1}{n-j} \sum_{t=1+j}^n (H_t(\alpha, \theta_0) - \alpha/2) (H_{t-j}(\alpha, \theta_0) - \alpha/2), \quad (75)$$

respectively. Similarly, define  $\hat{\rho}_{nj}$  and  $\hat{\gamma}_{nj}$ , the sample counterparts of  $\rho_j$  and  $\gamma_j$  based on the sample  $\{H_t(\alpha, \hat{\theta}_T)\}_{t=1}^n$  with

$$\hat{\rho}_{nj} = \frac{\hat{\gamma}_{nj}}{\hat{\gamma}_{n0}} \quad \text{and} \quad \hat{\gamma}_{nj} = \frac{1}{n-j} \sum_{t=1+j}^n \left( H_t(\alpha, \hat{\theta}_T) - \alpha/2 \right) \left( H_{t-j}(\alpha, \hat{\theta}_T) - \alpha/2 \right). \quad (76)$$

The sketch of the proof is similar to that used for Theorem 1. Under assumptions A1-A4, the continuous mapping theorem implies

$$\sqrt{n-j} (\hat{\gamma}_{nj} - \mathbb{E}(\hat{\gamma}_{nj} | \Omega_{t-1})) = \sqrt{n-j} (\gamma_{nj} - \mathbb{E}(\gamma_{nj} | \Omega_{t-1})) + o_p(1). \quad (77)$$

Rearranging these terms gives

$$\sqrt{n-j} (\hat{\gamma}_{nj} - \gamma_{nj}) = \sqrt{n-j} \mathbb{E}(\hat{\gamma}_{nj} - \gamma_{nj} | \Omega_{t-1}) + o_p(1). \quad (78)$$

The mean value theorem implies that

$$\hat{\gamma}_{nj} = \gamma_{nj} + (\hat{\theta}_T - \theta_0)' \frac{\partial \tilde{\gamma}_{nj}}{\partial \theta}, \quad (79)$$

whith  $\tilde{\gamma}_{nj}$  the lag- $j$  autocovariance of the process  $H_t(\alpha, \tilde{\theta})$ , where  $\tilde{\theta}$  is an intermediate point between  $\theta_0$  and  $\hat{\theta}_T$ . Define  $\lambda = (n-j)/T$ , equation (78) becomes

$$\sqrt{n-j} (\hat{\gamma}_{nj} - \gamma_{nj}) = \sqrt{\lambda} \sqrt{T} (\hat{\theta}_T - \theta_0)' \mathbb{E} \left( \frac{\partial \tilde{\gamma}_{nj}}{\partial \theta} \Big| \Omega_{t-1} \right) + o_p(1). \quad (80)$$

Under assumption A4, when  $T \rightarrow \infty$ , we have  $\tilde{\theta} \xrightarrow{p} \theta_0$  and

$$\mathbb{E} \left( \frac{\partial \tilde{\gamma}_{nj}}{\partial \theta} \Big| \Omega_{t-1} \right) \xrightarrow{p} \mathbb{E} \left( \frac{\partial \gamma_{nj}}{\partial \theta} \Big| \Omega_{t-1} \right). \quad (81)$$

Notice that  $\mathbb{E}((H_t(\alpha, \theta_0) - \alpha/2) \partial H_{t-j}(\alpha, \theta_0) / \partial \theta | \Omega_{t-1}) = \partial H_{t-j}(\alpha, \theta_0) / \partial \theta \mathbb{E}((H_t(\alpha, \theta_0) - \alpha/2) | \Omega_{t-1}) = 0$  for  $j > 0$ . Then, when  $T \rightarrow \infty$  and  $n \rightarrow \infty$ , we get

$$\mathbb{E} \left( \frac{\partial \tilde{\gamma}_{nj}}{\partial \theta} \Big| \Omega_{t-1} \right) \xrightarrow{p} \gamma_0 R_j \quad (82)$$

with

$$R_j = \frac{1}{\gamma_0} \mathbb{E} \left( \frac{\partial \gamma_{nj}}{\partial \theta} \right) = \frac{1}{\gamma_0} \mathbb{E} \left( (H_{t-j}(\alpha, \theta_0) - \alpha/2) \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right), \quad (83)$$

and  $\gamma_0 = \alpha(1/3 - \alpha/4)$ . Therefore

$$\sqrt{n} \hat{\gamma}_{nj} = \sqrt{n} \gamma_{nj} + \sqrt{\lambda} \sqrt{T} \gamma_0 R_j' (\hat{\theta}_T - \theta_0) + o_p(1). \quad (84)$$

By applying the delta method, we finally derive the impact of the estimation risk on the sample autocorrelation. Note that when  $n \rightarrow \infty$ ,  $\rho_{nj}$  converges to 0, since under the null  $\rho_j = 0$  for

$j = 1, \dots, m$ . Then, the estimation risk in the autocorrelation  $\widehat{\rho}_{nj}$  only comes from the estimation risk in the autocovariance  $\widehat{\gamma}_{nj}$ .

$$\sqrt{n}\widehat{\rho}_{nj} = \sqrt{n}\frac{\widehat{\gamma}_{nj}}{\widehat{\gamma}_{n0}} = \sqrt{n}\frac{\widehat{\gamma}_{nj}}{\gamma_0} + o_p(1) \quad (85)$$

Hence, we proved that

$$\sqrt{n}\widehat{\rho}_{nj} = \sqrt{n}\rho_{nj} + \sqrt{\lambda}\sqrt{T}R'_j(\widehat{\theta}_T - \theta_0) + o_p(1). \quad (86)$$

Notice that  $\sqrt{n}(\rho_{n1}, \dots, \rho_{nm})' \xrightarrow{d} \mathcal{N}(0, I_m)$  and the covariance between the first term and  $\sqrt{T}(\widehat{\theta}_T - \theta_0)$  is 0 as  $\widehat{\theta}_T$  depends on the in-sample observations and the correlation  $\rho_{nj}$  depends on the out-of-sample observations. Denote  $\widehat{\rho}_n^{(m)}$  the vector  $(\widehat{\rho}_{n1}, \dots, \widehat{\rho}_{nm})'$ . Under assumptions A1-A4, we have

$$\sqrt{n}\widehat{\rho}_n^{(m)} \xrightarrow{d} \mathcal{N}(0, \Delta), \quad (87)$$

with the  $ij$ -th element of  $\Delta$  given by

$$\Delta_{ij} = \delta_{ij} + \lambda R'_i \Sigma_0 R_j, \quad (88)$$

$$R_j = \frac{1}{\alpha(1/3 - \alpha/4)} \mathbb{E} \left( (H_{t-j}(\alpha, \theta_0) - \alpha/2) \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right),$$

$\forall (i, j) \in \{1, \dots, m\}^2$ , where  $\delta_{ij}$  is a dummy variable that takes a value 1 if  $i = j$  and 0 otherwise. One can write  $\Delta = Q\Lambda Q'$ , where  $Q$  is an orthogonal matrix, and  $\Lambda$  is a diagonal matrix with elements  $\{\pi_j\}_{j=1}^m$ . So, we have

$$Q' \sqrt{n}\widehat{\rho}_n^{(m)} \xrightarrow{d} \mathcal{N}(0, \Lambda). \quad (89)$$

Finally

$$IND_{MES} = n \sum_{j=1}^m \widehat{\rho}_{nj}^2 \xrightarrow{d} \sum_{j=1}^m \pi_j Z_j^2, \quad (90)$$

where  $\{\pi_j\}_{j=1}^m$  are the eigenvalues of the matrix  $\Delta$  and  $\{Z_j\}_{j=1}^m$  are independent standard normal variables. ■

## References

- [1] Acharya, V. V., R. Engle, and M. Richardson (2012): "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks," *The American Economic Review*, 102(3), 59-64.
- [2] Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2010): "Measuring systemic risk", Working paper.
- [3] Adrian, T., and M. Brunnermeier (2016): "CoVaR", forthcoming in *American Economic Review*.
- [4] Arnold, B. C., R.J. Beaver, R.A Groeneveld, and W.Q. Meeker (1993): "The nontruncated marginal of a truncated bivariate normal distribution", *Psychometrika*, 58 (3), 471-488.



- [5] Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): "Coherent measures of risk", *Mathematical Finance*, 9(3), 203-228.
- [6] Benoit S., J.E. Colliard, C. Hurlin and C. Pérignon (2016): "Where the Risks Lie: A Survey on Systemic Risk", forthcoming in *Review of Finance*.
- [7] Berkowitz, J., P. Christoffersen and D. Pelletier (2011): "Evaluating value-at-risk models with desk-level data", *Management Science* 57 (12), 2213-2227.
- [8] Box, G. E. P. and D.A. Pierce (1970): "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, 65, 1509–1526.
- [9] Brownlees, T. C., and R. F. Engle (2015): "SRISK: A conditional capital shortfall index for systemic risk measurement", Working paper.
- [10] Chen, C., G. Iyengar, and C. C. Moallemi (2013): "An Axiomatic Approach to Systemic Risk", *Management Science*, 59(6), 1373-1388.
- [11] Christoffersen, P. F. (1998): "Evaluating interval forecasts", *International Economic Review*, 39, 841-862.
- [12] Christoffersen, P. F. (2009), "Backtesting". In *Encyclopedia of Quantitative Finance*, R. Cont (ed). John Wiley and Sons.
- [13] Du, Z. and J.C. Escanciano (2015): "Backtesting expected shortfall: Accounting for tail risk", forthcoming in *Management Science*.
- [14] Dwyer, P.S. (1967), Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318), 607-625.
- [15] Engle, R., E. Jondeau, and M. Rockinger (2015): "Systemic Risk in Europe", *Review of Finance*, 19(1), 145-190.
- [16] Francq, C. and J.M., Zakoian (2015): "Looking for Efficient QML Estimation of Conditional Value-at-Risk at Multiple Risk Levels," MPRA Working Paper 67195.
- [17] Girardi, G. and A. T. Ergun (2013): "Systemic risk measurement: Multivariate GARCH estimation of CoVaR", *Journal of Banking & Finance*, 37(8), 3169-3180.
- [18] Guntay, L. and P. Kupiec (2015): "Testing for Systemic Risk using Stock Returns". Working paper
- [19] Ho, H. J., T.I. Lin, H.Y. Chen, and W.L. Wang (2012): "Some results on the truncated multivariate t distribution". *Journal of Statistical Planning and Inference*, 142(1), 25-40.
- [20] Hurlin C. and C. Pérignon (2012): "Margin Backtesting". *Review of Futures Market*, 20, 179-194.

- [21] Hurlin C. and S. Tokpavi (2006): "Bactesting Value-at-Risk Accuracy: A Simple New Test". *Journal of Risk*, 9(2), 19-37.
- [22] Horrace, W. C. (2005): "Some results on the multivariate truncated normal distribution", *Journal of Multivariate Analysis*, 94(1), 209-221.
- [23] Idier, J. G. Lamé, J.S. Mésonnier (2014): "How useful is the Marginal Expected Shortfall for the Measurement of Systemic Exposure? A Practical Assessment", *Journal of Banking and Finance*, 47, 134-146.
- [24] Jorion, P. (2007): *Value-at-Risk*, Third edition, McGraw-Hill.
- [25] Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models". *Journal of Derivatives*, 3, 73-84.
- [26] Leccadito A., S. Boffelli, and G. Urga (2014): "Evaluating the Accuracy of Value-at-Risk Forecasts: New Multilevel Tests", *International Journal of Forecasting*, 30(2), 206-216.
- [27] Pérignon, C. and D. R. Smith (2008): "A New Approach to Comparing VaR Estimation Methods". *Journal of Derivatives*, 16, 54-66.
- [28] Scaillet, O. (2004): "Nonparametric estimation and sensitivity analysis of expected shortfall," *Mathematical Finance*, 14(1), 115-129.