# Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures

Denisa Banulescu-Radu, Christophe Hurlin, Jérémy Leymarie, Olivier Scaillet *

April 3, 2020

**Abstract**

This paper proposes an original approach for backtesting systemic risk measures. This backtesting approach makes it possible to assess the systemic risk measure forecasts used to identify the financial institutions that contribute the most to the overall risk in the financial system. Our procedure is based on simple tests similar to those generally used to backtest the standard market risk measures such as value-at-risk or expected shortfall. We introduce a concept of violation associated with the marginal expected shortfall (MES), and we define unconditional coverage and independence tests for these violations. We can generalize these tests to any MES-based systemic risk measures such as SES, SRISK, or $\Delta$CoVaR. We study their asymptotic properties in the presence of estimation risk and investigate their finite sample performance via Monte Carlo simulations. An empirical application to a panel of U.S. financial institutions is conducted to assess the validity of MES, SRISK, and $\Delta$CoVaR forecasts issued from a GARCH-DCC model. Our results show that this model provides valid forecasts for MES and SRISK when considering a medium-term horizon. Finally, we propose an early warning system indicator for future systemic crises deduced from these backtests. Our indicator quantifies how much is the measurement error issued by a systemic risk forecast at a given point in time which can serve for the early detection of global market reversals.

*Keywords*: Backtesting; Banking regulation; Hypothesis testing; Risk management; Systemic risk.

*JEL classification: C12, C52, C58, G21, G28.*

---

*Banulescu-Radu and Hurlin are at the University of Orléans (LEO, FRE CNRS 2014), Leymarie is at the University of Vienna, and Scaillet is at the University of Geneva and Swiss Finance Institute. We would like to thank the Finance department editor and the two referees for constructive criticism and suggestions for an improved version. We would also like to thank for their valuable comments Charles-Olivier Amédée-Manesme, Katarzyna Bień-Barkowska, Daniel Brunner, Ophélie Couperier, Jessica Fouilloux, Christian Gouriéroux, Alain Hecq, Rachidi Kotchoni, Sébastien Laurent, Nour Meddahi, Phu Nguyen-Van, Andrew Patton, Christophe Pérignon, Joey Soudant, and Jean-Michel Zakoïan. We also thank the seminar participants at Audencia Business School, CREST, Maastricht University, Toulouse Business School, University of Rennes, University of Orléans as well as the participants of the 2016 ACPR Chair "Regulation and Systemic Risk" - Dauphine House of Finance Day Conference, 2016 French Finance Association Conference, 2016 International Conference of the Financial Engineering and Banking Society, 2016 Annual Meeting of the French Economic Association, 2016 European Meeting of the Econometric Society, 2016 Econometric Research in Finance Workshop, 2016 MIFN Conference, 15th Conference "Développements Récents de l'Econométrie Appliquée à la Finance" (Université Paris Nanterre), 2016 French Econometrics Conference, 10th Conference on Computational and Financial Econometrics, 2017 Vienna-Copenhagen Conference on Financial Econometrics, 22nd Spring Meeting of Young Economists, 25th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, 14th Augustin Cournot Doctoral Days, 2017 INFER Conference, 2017 Annual Conference of the Multinational Finance Society, 7th PhD Student Conference in International Macroeconomics and Financial Econometrics, 2018 RiskLab-BoF-ESRB Conference on Systemic Risk Analytics, 2018 Annual Conference of the International Association for Applied Econometrics, and 2019 conference Europlace Institute of Finance and Banque de France Foundation. Finally, we thank the Chair ACPR/Risk Foundation: Regulation and Systemic Risk, ANR programs MultiRisk (ANR-16-CE26-0015-01) and CaliBank (ANR-19-CE26-0002-02), and the Banque de France Foundation for supporting our research. All data and Matlab codes are available at https://zenodo.org/record/3732345.

# 1 Introduction

On October 26, 2014, the European Banking Authority (EBA) published the results of the 2014 EU-wide stress test, which involved 123 banks from 22 countries, covering broadly 70% of total assets in the EU banking sector (EBA, 2014). The main conclusion of this exercise is that only 24 participating banks fall below the defined capital threshold, leading to a maximum aggregate capital shortfall of €24.6bn. None of these banks is French, and Christian Noyer, governor of the Banque de France at that time, immediately saluted the success of the "*French banks which are in the best positions in the Eurozone*" (Noyer, October 26, 2014). Interviewed by the Financial Times a day after these announcements, Viral Acharya (Financial Times, 2014) stated that "*French banks are the weakest in Europe*". He declared that in event of a crisis, French financial institutions would face an aggregate capital shortfall of almost $400bn, i.e., approximately 15 times greater than the shortfall identified by the EBA for the weakest European banks.[1]

This debate illustrates the difficulty of measuring systemic risk and the need for validation tools of such measures.[2] Our aim in this paper is to address this gap by introducing a general framework for assessing the validity of systemic risk measures. This framework is specifically designed for systemic risk measures that are expressed as functions of the expected equity loss conditional on a financial crisis.

The ultimate goal of a systemic risk measure is to better identify the vulnerabilities of the financial system (see De Bandt and Hartmann, 2002; Benoit et al., 2017, for a survey). We can identify two main families of systemic risk measures. The first set aggregates low-frequency regulatory data. A typical example is the systemic risk score currently implemented by the Basel Committee on Banking Supervision (BCBS) and the Financial Stability Board (FSB) to identify so-called systemically important financial institutions (SIFIs), i.e., firms whose failure might trigger a crisis affecting the entire financial system.[3] Another example is the bank capital shortfall computed from the regulatory banking stress tests previously mentioned. Until 2014, the performance of these measures could not be assessed empirically because the necessary data were not in the public domain. Recently, Philippon et al. (2017) proposed the first evaluation of the quality of the EU banking stress

[1] Considering only the subset of French banks tested in the ECB stress tests, he estimates a capital shortfall of $189bn which represents approximately 9% of the French GDP.

[2] For similar discussions see also Acharya et al. (2014), Tavolaro and Visnovsky (2014), Acharya et al. (2016a,b), and more recently Pierret and Steffen (2018).

[3] The systemic risk score aggregates information on five broad categories of systemic importance: size, interconnectedness, substitutability, complexity, and cross-jurisdictional activity. To avoid favoring any particular facet of systemic risk, the BCBS computes an equally weighted average score of these categories. For further details, see Basel Committee on Banking Supervision (2014).

tests, and Benoit et al. (2019) identified some pitfalls in the systemic risk scoring. The second set of systemic risk measures relies on high-frequency market data such as stocks or asset returns, option prices, or CDS spreads. These measures have the advantage of being easily implemented with public data and standard econometric models. Many global risk measures have been proposed in the academic literature.[4] The most prominent examples are the marginal expected shortfall (MES) and the systemic expected shortfall (SES) of Acharya et al. (2017), the systemic risk measure (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2017), and the delta conditional value-at-risk ($\Delta$CoVaR) of Adrian and Brunnermeier (2016).[5] 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). It is the key determinant, together with leverage and firm market value, of the SRISK and the SES. These two systemic risk measures correspond to the expected amount by which a bank is undercapitalized in a future systemic event in which the overall financial system is undercapitalized. These global measures are designed to summarize the systemic risk contribution of a given financial institution into a single figure and can be used to rank financial institutions according to their systemic importance. However, to the best of our knowledge, no backtesting procedures have been proposed to evaluate their ex post validity. This issue is of crucial importance because validation is a key requirement for any systemic risk measure to become an industry standard.

We propose here a general framework for backtesting the MES and by extension the SES and SRISK. As defined by Jorion (2007), backtesting is a formal statistical framework that consists of verifying whether actual losses are in line with projected losses. This involves a systematic comparison of the historical model-generated risk measure forecasts with actual losses. Since the true value of the risk measure is unobservable, this comparison generally relies on violations. For instance, in the case of VaR, a violation is said to occur when the ex post portfolio return is lower than the VaR forecast. The goal here consists of defining an appropriate concept of violation for the MES and then adapting the backtesting tests currently used by regulators and industry for the market risk measures such as VaR or expected shortfall (ES). We proceed as follows. First, we introduce a concept of conditional VaR, inspired by the systemic risk measure proposed by Adrian and Brunnermeier (2016). A $(\beta, \alpha)$-CoVaR is defined as the $\beta$-quantile of the truncated distribution

[4] Bisias et al. (2012) surveyed 31 quantitative measures, whereas Giglio et al. (2016) proposed systemic risk indexes computed from 19 alternative systemic risk measures.

[5] These articles are among the most influential in the academic literature on systemic risk. For instance, the two articles by Acharya et al. (2012) and Brownlees and Engle (2017) that defined the MES and the SRISK have been cited 1,000 times since their publication (source: Google Scholar). For online computation of some of these systemic risk measures, see the Stern-NYU V-Lab initiative website.

of the firm returns given that the financial system takes a loss greater than its $\alpha$-VaR. We express the MES as an integral of the CoVaRs for all coverage rates $\beta$ 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 extended to the SES and SRISK. Second, we define a concept of joint violation of the $(\beta, \alpha)$-CoVaR of the firm returns and $\alpha$-VaR of the market returns. This extends the concept of cumulative violation recently proposed by Du and Escanciano (2017) for ES backtests to a bivariate case. We define a cumulative joint violation process defined as the integral of the joint violation processes for all coverage rates $\beta$ between 0 and 1. We show that if the risk model used to forecast the MES is well specified, this cumulative violation process is a martingale difference sequence (mds). Exploiting this mds property, we propose two backtests for the MES: an unconditional coverage (UC, hereafter) test and an independence (IND, hereafter) test as those generally considered for the VaR (Christoffersen, 1998). The UC test refers to the fact that the violation frequency should be in line with the theoretical probability of observing a violation. Failure of UC means that the MES forecasts underestimate or overestimate the ex post systemic risk. In addition to UC, the violations should satisfy the independence property, which implies that past violations should not be informative about current or future violations if the dynamics of the risk model to forecast MES is well-specified.

Our backtesting procedure has many advantages. It allows backtesting either conditional (with respect to the past information set) MES or unconditional MES. Furthermore, we derive the asymptotic distribution of our test statistics while taking into account the estimation risk (Escanciano and Olmo, 2010). Indeed, the MES forecasts are generally issued from a parametric risk model where parameters have to be estimated. Then, the use of standard backtesting procedures to assess the MES forecasts in an out-of-sample framework can be misleading because these procedures do not consider the impact of estimation risk. For this reason, we propose a robust version of our test statistics. Monte Carlo simulations show that these robust statistics have good finite sample properties for realistic sample sizes.

Another advantage of our backtesting procedure is that it can be applied to any MES-based systemic risk measure, the typical examples being SES and SRISK. As we can express these measures as a linear deterministic function of the (long-run) MES, we show that testing their validity is equivalent to testing the validity of the MES forecast itself.[6] We can further extend our framework

[6]This result is due to the assumptions made on the other constituents of SES and SRISK, i.e., a constant level of liabilities and a constant initial market value of the firm (see Brownlees and Engle, 2017; Acharya et al., 2017).

to test the validity of $\Delta$CoVaR forecasts. To do so, we propose a simple approach based on a vector of two joint violations associated with two conditional VaRs of an institution: one for the situation in which the financial system is in distress and a second with respect to the median state of the financial system. The intuition is similar here to the backtests proposed for a multi-level VaR, i.e., VaR defined for a finite set of coverage rates (see Pérignon and Smith, 2008; Colletaz et al., 2013; Leccadito et al., 2014; Wied et al., 2016, among others). These multi-level VaR backtests have been recently adapted by Kratz et al. (2018) and Couperier and Leymarie (2018) to assess ES forecasts.

We apply our backtesting tests to assess the empirical validity of the MES, SRISK, and $\Delta$CoVaR issued from a GARCH-DCC model for a panel of large U.S. financial institutions over the period from January 3, 2000, to December 30, 2016. First, we observe that the one-day-ahead forecasts of these systemic risk measures are generally misspecified in periods of financial instability. The UC tests reject the validity of the risk forecasts for most of the financial institutions in our panel during the 2008-2009 global financial crisis. Second, when considering a longer forecasting horizon, say one month, the UC hypothesis is no longer rejected for MES or SRISK. The results suggest that for longer horizons, the GARCH-DCC model may provide more reliable forecasts of systemic risk. Finally, we apply the UC test to the $\Delta$CoVAR daily forecasts. Our findings are fully in line with those observed for the short-term MES forecasts. The UC hypothesis is rejected for most of the financial institutions during the global 2008-2009 financial crisis and to a lesser extent during the 2011-2012 European debt crisis. In addition, our tests show that the stressed CoVaR is generally more affected by model misspecification than the median CoVaR.

Some attempts of validation procedures for systemic risk measures have been proposed in the literature. Following the coherent risk approach of Artzner et al. (1999), Chen et al. (2013) define an axiomatic framework for systemic risk measures.[7] However, most of the validation techniques are empirical. They generally consist of testing whether firms with high systemic risk scores are more likely to become insolvent (Wu, 2018) or to suffer the highest financial losses (Idier et al., 2014) in a financial crisis, for example. The latter conclude that some standard balance-sheet ratios are better able to predict large equity losses conditional on a true crisis than is the MES. Brownlees and Engle (2017) 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 et al.

[7] A systemic risk measure must satisfy the main conditions that define any coherent risk measure, namely, the monotonicity, positive homogeneity, and outcome convexity axioms. However, it must also satisfy an additional preference consistency axiom. This axiom states that the risk measure has to reflect the preference of the regulator on the cross-sectional profile of losses across firms and the distribution of the aggregate outcomes across states.

(2015) compare the ranking of European financial institutions obtained with the SRISK to the list of SIFIs produced by the FSB. Recently, Jokivuolle et al. (2018) use statistical tests and find that the systemic risk categories defined by the FSB are different from those computed by $\Delta CoVaR$, MES, and SRISK. Brownlees et al. (2020) propose an historical assessment of SRISK and ΔCoVaR based on two dimensions. The first one, called SIFI ranking challenge, consists of investigating whether ranking financial institutions by SRISK and ΔCoVaR permits the identification of institutions with notable deposit declines around panic events. The second one, called the financial crisis prediction challenge, investigates whether these systemic risk measures are significant predictors of system-wide deposit declines during panic events. Based on an original historical dataset for the New York banking system between 1866 and 1933, their results show that both measures well identify SIFIs, especially in periods of distress. However, SRISK and ΔCoVaR exhibit poor performance as early warning signals of distress in the financial system as a whole.

Our validation approach is different because it relies on backtesting tests in the same spirit as those used for market or credit risk measures (Jorion, 2007). We propose test statistics that are similar to those generally used by regulators or risk managers to backtest the VaR (Kupiec, 1995; Christoffersen, 1998) or the ES (Du and Escanciano, 2017; Kratz et al., 2018). Even if the validation approaches are different, they show similarities in the main. As in Brownlees et al. (2020), albeit with a different empirical approach, we propose an early warning system (EWS) indicator intended to identify the periods of forecast breakdown that depict significant changes in market conditions. Our indicator is defined as the difference between the original MES forecast issued from the risk model and an adjusted MES forecast. The latter represents the forecast that satisfies the null hypothesis of unconditional coverage. The adjustment of risk forecasts issued from a misspecified or mis-estimated risk model has already been considered in the risk management literature. For instance, Gouriéroux and Zakoïan (2013) propose a method to adjust VaR forecasts affected by estimation risk. Boucher et al. (2014) adjust imperfect VaR forecasts by outcomes from backtesting frameworks, considering desirable qualities of VaR models such as the frequency, independence and magnitude of violations. In recent contributions, Couperier and Leymarie (2018) and Lazar and Zhang (2019) apply a similar approach to adjust imperfect ES forecasts. Similarly, we propose to adjust imperfect MES-based forecasts exploiting the statistical properties of the cumulative joint violation process. Formally, it consists of determining an adjusted coverage level for the MES such that the null hypothesis of the UC test is not rejected. Our empirical results show that the adjusted MES diverges from the unadjusted MES at the beginning of the 2007-2009 financial crisis and to a lesser extent during the European debt crisis. Such a divergence should alert regulators to the severity of the financial crisis because the risk models applied by banks before the crisis are no longer compatible with current market conditions.

The remainder of the paper is organized as follows. In Section 2, we define the MES and introduce the cumulative joint violation process. We present the backtesting tests for MES in Section 3 and illustrate their finite sample properties via Monte Carlo simulations. Section 4 extends our backtests to any MES-based systemic risk measure, especially to SES, SRISK and ΔCoVaR. An empirical application is proposed in Section 5. Section 6 illustrates how to use these backtests as early warning systems to detect financial crisis episodes. Finally, we conclude the paper in Section 7.

# 2 MES and cumulative joint violation process

In this section, we define the MES and introduce a concept of cumulative joint violation dedicated to its assessment. We consider the following notations. 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, $Y_{1t}$ corresponds to the stock return of a financial institution, whereas $Y_{2t}$ corresponds to the market return. Denote by $\Omega_{t-1}$ the information set available at time $t-1$, with $(Y_{t-1}, Y_{t-2}, ...) \subseteq \Omega_{t-1}$ and $F(.;\Omega_{t-1})$ being 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 $F(.;\Omega_{t-1})$ is continuous.

## 2.1 Marginal expected shortfall

Following Acharya et al. (2012) and Brownlees and Engle (2017), 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 VaR. Formally, the $\alpha$-level MES of the financial institution at time $t$ given $\Omega_{t-1}$ is defined as

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

where $VaR_{2t}(\alpha)$ is the $\alpha$-level VaR of the marginal distribution of $Y_{2t}$, denoted $F_{Y_2}(.;\Omega_{t-1})$, with $VaR_{2t}(\alpha) = F_{Y_2}^{-1}(\alpha;\Omega_{t-1})$, and $\alpha \in [0,1]$.[8] If we define the market return $Y_{2t}$ as the value-weighted average of firm returns (for all the firms that belong to the financial system), then the MES corresponds to the derivative of the market ES with respect to the firm market share (Scaillet, 2004; Acharya et al., 2017), hence the term "marginal". Thus, MES measures how the financial institution contributes to the overall risk of the financial system.

[8] For ease of notation, we do not use the usual convention that defines the VaR as the opposite of the $\alpha$-quantile of the return distribution.

The MES is a conditional expectation, and as such, it 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 purpose, we introduce the concept of Conditional-VaR (CoVaR) inspired by the systemic risk measure proposed by Adrian and Brunnermeier (2016). For any coverage level $\beta \in [0,1]$, the CoVaR for firm 1 at time $t$ is the quantity $CoVaR_{1t}(\beta,\alpha)$ such that

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

There are two main differences between $CoVaR_{1t}(\beta,\alpha)$ and the CoVaR introduced by Adrian and Brunnermeier (2016). First, the conditioning event is based on an inequality, i.e., $Y_{2t} \leq VaR_{2t}(\alpha)$ as in Girardi and Ergün (2013), rather than on the equality $Y_{2t} = VaR_{2t}(\alpha)$. Second, we introduce 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. The $(\beta,\alpha)$-level CoVaR can also be defined as $CoVaR_{1t}(\beta,\alpha) = F^{-1}_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(\beta;\Omega_{t-1})$ where $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(.;\Omega_{t-1})$ is the cdf of the conditional distribution of $Y_{1t}$ given $Y_{2t} \leq VaR_{2t}(\alpha)$ and $\Omega_{t-1}$. Definition of a conditional probability and a change of variables yield a useful representation of MES in terms of CoVaR.

$$MES_{1t}(\alpha) = \int_0^1 CoVaR_{1t}(\beta,\alpha)\mathrm{d}\beta. \tag{3}$$

Equation (3) is essential for our backtesting approach. It provides a simple relationship between two risk measures, i.e., the MES and the CoVaR. To the best of our knowledge, this is the first attempt to establish a link between these systemic measures that are both broadly used in the systemic risk literature (see Acharya et al., 2017, for a comparison). Notice that this definition of the MES is valid for any bivariate distribution.

For some particular distributions, the conditional cdf $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(.;\Omega_{t-1})$ that defines 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) formalizes analytical results on the truncated multivariate normal distribution, where the truncation is one-sided and at an arbitrary point. Ho et al. (2012) focus on the truncated multivariate $t$-distribution. Regardless of the distribution, it is also 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 $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(y_1;\Omega_{t-1}) = \alpha^{-1} F\left(\widetilde{y};\Omega_{t-1}\right)$, with $\widetilde{y} = (y_1, VaR_{2t}(\alpha))'$.

In general, the MES forecasts are issued from a parametric model specified by the researcher, the risk manager, or the regulatory authority. For instance, Brownlees and Engle (2017) and Acharya et al. (2012) consider a bivariate dynamic conditional correlation (DCC) model to compute the MES and the SRISK. In practice, the cdf $F(.;\Omega_{t-1},\theta_0)$ of the joint distribution of $Y_t$, the cdf $F_{Y_2}(.;\Omega_{t-1},\theta_0)$ of the marginal distribution of $Y_{2t}$ and the cdf $F_{Y_1|Y_2\leq VaR_{2t}(\alpha;\theta_0)}(.;\Omega_{t-1},\theta_0)$ of the truncated distribution of $Y_{1t}$ given $Y_{2t}\leq VaR_{2t}(\alpha,\theta_0)$ depend on $\theta_0\in\Theta\subseteq\mathbb{R}^p$, a vector of unknown parameters. Therefore, we need to estimate these parameters to compute the MES forecasts. In the sequel, we name this parametric model the "risk model" in reference to the internal risk model used by banks to produce their VaR or ES forecasts.

## 2.2 Cumulative joint violation process

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 $t$. We represent this violation process by the following binary variable

$$h_t(\alpha,\beta,\theta_0)=\mathbb{1}\left((Y_{1t}\leq CoVaR_{1t}(\beta,\alpha,\theta_0))\cap(Y_{2t}\leq VaR_{2t}(\alpha,\theta_0))\right), \quad (4)$$

where $\mathbb{1}(.)$ denotes an indicator function. The violation (4) takes value one if the loss of the firm exceeds its CoVaR and the loss of the market exceeds its VaR, zero otherwise.

The VaR backtesting tests (Kupiec, 1995; Christoffersen, 1998; Berkowitz et al., 2011, among others) generally exploit the mds property of the violation process (see Christoffersen, 2010, for a survey). Here, we adopt a similar approach to backtest the CoVaR and the MES. Notice that Bayes theorem implies that $\Pr(h_t(\alpha,\beta,\theta_0)=1|\,\Omega_{t-1})=\alpha\beta$. Then, it follows from Equation (2) that the violations are Bernoulli distributed 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 $(\alpha,\beta)\in[0,1]^2$. Formally, if the $(\beta,\alpha)$-CoVaR of $Y_{1t}$ and the $\alpha$-VaR of $Y_{2t}$ are correctly specified, we have $\mathbb{E}(h_t(\alpha,\beta,\theta_0)-\alpha\beta|\,\Omega_{t-1})=0$.

To test the validity of the MES, we consider a cumulative joint violation process that we can view as a violation "counterpart" of the MES definition in Equation (3). We define this cumulative joint violation process as the integral of the violations $h_t(\alpha,\beta,\theta_0)$ for all the risk levels $\beta$ between 0 and 1, with

$$H_t(\alpha,\theta_0)=\int_0^1 h_t(\alpha,\beta,\theta_0)\mathrm{d}\beta.$$

We can regard this cumulative joint violation process as an extension to the bivariate case of the cumulative violation process recently introduced by Du and Escanciano (2017) to backtest the ES. The random variable $H_t(\alpha,\theta_0)$ has an $\Omega_{t-1}$-conditional distribution defined as the product of a Bernoulli($\alpha$) random variable times a continuous uniform distribution over $[0,1]$. The mean and variance of $H_t(\alpha,\theta_0)$ are equal to $\alpha/2$ and $\alpha(1/3-\alpha/4)$, respectively (see Appendix B for details). Furthermore, the Fubini 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 $\alpha\in[0,1]$, such that we have $\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)-\alpha/2\right|\Omega_{t-1}\right)=0$.

# 3 Backtesting MES

Exploiting the mds property of the cumulative joint violation process enables the implementation of various types of backtests for the MES (see Nieto and Ruiz, 2016, for a survey). Here, we propose two tests based on the so-called unconditional coverage hypothesis and independence hypothesis (Christoffersen, 1998). The unconditional coverage test relies on the following null hypothesis:

$$\mathrm{H}_{0,UC}:\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\right)=\alpha/2.$$

Since we have $\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\right)=\int_0^1\mathbb{E}\left(h_t(\alpha;\beta,\theta_0)\right)\mathrm{d}\beta$, the null $\mathrm{H}_{0,UC}$ is equivalent to $H_{0,UC}$ : $\Pr\left(h_t(\alpha,\beta,\theta_0)=1\right)=\alpha\beta$ for any $\beta\in[0,1]$. The null hypothesis UC means that for any risk level $\beta$, the joint probability of observing 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 $\alpha\beta$. Note that if $\mathrm{H}_{0,UC}$ is violated, there are two scenarios. The first corresponds to $\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\right)>\alpha/2$, and indicates that the number of joint exceedances is higher than expected. This scenario is associated with an underestimation of the MES. In the opposite case, $\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\right)<\alpha/2$ indicates that not enough exceedances are experienced on average and reveals an overestimation of the MES.

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

$$\mathrm{H}_{0,IND}:\rho_1=....=\rho_K=0,$$

with $\rho_k=corr\left(H_t\left(\alpha,\theta_0\right),H_{t-k}\left(\alpha,\theta_0\right)\right)$. A rejection of $\mathrm{H}_{0,IND}$ highlights a misspecification in the MES dynamics, and reveals the use of an incorrect risk model.

Implementing the tests requires estimating the parameters $\theta_0\in\Theta$ of the model used by the risk manager to forecast the MES. For simplicity, we adopt a fixed forecasting estimation scheme. We use an in-sample period from $t=1$ to $t=T$ to estimate $\theta_0$. Denote by $\Omega_T$ the information set available at the end of the in-sample period, with $\{Y_1, ..., Y_T\} \subseteq \Omega_T$ and $\widehat{\theta}_T$ a consistent estimator of $\theta_0$. Based on the ex post returns observed from $t = T+1$ to $t = T+n$, we compute the backtesting test statistics from the out-of-sample forecasts of the cumulative violation process given by

$$H_t(\alpha, \widehat{\theta}_T) = \left(1 - u_{12t}(\widehat{\theta}_T)\right) \mathbb{1}\left(u_{2t}(\widehat{\theta}_T) \leq \alpha\right), \quad \forall t = T+1, ..., T+n,$$

with $u_{2t}(\widehat{\theta}_T) \equiv F_{Y_2}(Y_{2t}; \Omega_{t-1}, \widehat{\theta}_T)$ and $u_{12t}(\widehat{\theta}_T) \equiv F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \widehat{\theta}_T)}(Y_{1t}; \Omega_{t-1}, \widehat{\theta}_T)$. See Appendix B for further details.

### 3.1 Unconditional coverage test

By analogy with the backtest proposed by Du and Escanciano (2017) for ES and the well-known VaR backtest proposed by Kupiec (1995), we consider a standard $t$-test for the null hypothesis of unconditional coverage $\mathrm{H}_{0,UC}$ for the MES. We define the test statistic, denoted $UC_{MES}$, as

$$UC_{MES} = \frac{\sqrt{n}\left(\bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2\right)}{\sqrt{\alpha\,(1/3 - \alpha/4)}}, \tag{5}$$

with $\bar{H}(\alpha, \widehat{\theta}_T)$ the out-of-sample mean of $H_t(\alpha, \widehat{\theta}_T)$, namely $\bar{H}(\alpha, \widehat{\theta}_T) = \frac{1}{n}\sum_{t=T+1}^{T+n} H_t(\alpha, \widehat{\theta}_T)$. To provide intuition for 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 $\theta_0$ instead of its estimator $\widehat{\theta}_T$. Under the null hypothesis, the sequence $\{H_t\,(\alpha, \theta_0) - \alpha/2\}_{t=T+1}^{T+n}$ is a mds with 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 $UC_{MES} \equiv UC_{MES}(\alpha, \widehat{\theta}_T)$ when $T \to \infty$ and $n \to \infty$, whereas $\lambda = n/T \to 0$, i.e., when there is no estimation risk.

However, in the general case $T \to \infty$, $n \to \infty$ and $n/T \to \lambda < \infty$, there is an estimation risk as soon as $\lambda \neq 0$. Escanciano and Olmo (2010) show that the use of standard backtesting procedures can be misleading if one does not account for the uncertainty associated with parameter estimation. Similarly, Bontemps and Meddahi (2005) develops a test for normality that is robust to parameter uncertainty and show that standard tests are misleading due to parameter uncertainty. In the presence of estimation error, 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 $T \to \infty$, $n \to \infty$ and $n/T \to \lambda$ with $0 < \lambda < \infty$.

**Theorem 1** *Under Assumptions A1-A4 in the Appendix, we have:*

$$UC_{MES} \stackrel{d}{\to} \mathcal{N}\left(0, \sigma_\lambda^2\right),$$

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

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

*with* $R_{MES} = \mathbb{E}_0\left(\partial H_t\left(\alpha, \theta_0\right)/\partial\theta\right)$ *and* $\mathbb{V}_{as}(\widehat{\theta}_T) = \Sigma_0/T$.

The proof of Theorem 1 is reported in Appendix C. The vector $R_{MES}$ quantifies the parameter estimation effect on the test statistic $UC_{MES}$ due to the difference between the estimate $\widehat{\theta}_T$ and the true value of the parameter $\theta_0$. We can characterize the impact of the estimation risk on the $UC_{MES}$ test statistic as follows

$$\begin{aligned} UC_{MES} &= \underbrace{\frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t\left(\alpha,\theta_0\right) - \alpha/2\right)}_{UC_{MES}(\alpha,\theta_0)} \\ &\quad \underbrace{+\frac{\sqrt{\lambda}}{\sigma_H}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)'}{\partial\theta}\right|\Omega_{t-1}\right)\sqrt{T}(\widehat{\theta}_T - \theta_0)}_{\text{Estimation risk}} + o_p\left(1\right). \end{aligned}$$

This formula is similar to that obtained by Du and Escanciano (2017) to backtest ES, but its constituents are different because the testing procedure now relies on a bivariate distribution. Whatever the dynamic model considered for the returns, we can deduce the vector $R_{MES}$ from the cdf of the joint distribution $Y_t$ given $\Omega_{t-1}$ with

$$R_{MES} = -\frac{1}{\alpha}\mathbb{E}_0\left(\frac{\partial F(\widetilde{y}_t;\Omega_{t-1},\theta_0)}{\partial\theta}\mathbb{1}(u_{2t}\left(\theta_0\right) \leq \alpha)\right) + \mathbb{E}_0\left((1 - u_{12t}\left(\theta_0\right))\frac{\partial\mathbb{1}(u_{2t}\left(\theta_0\right) \leq \alpha)}{\partial\theta}\right),$$

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

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

In the bivariate case, the estimation error affects the cdf of the truncated distribution of $Y_{1t}$ given $Y_{2t} \leq VaR_{2t}\left(\alpha,\theta_0\right)$ not only through its effect on the joint distribution but also through the truncation parameter, i.e., the VaR of the market return. There is no analytical expression for the derivative $\partial F(\widetilde{y}_t;\Omega_{t-1},\theta_0)/\partial\theta$ except for some particular bivariate distributions (see Appendix D for the case of a bivariate normal distribution). In the general case, we have to evaluate that expression via numerical differentiation.

**Corollary 2** *When there is no estimation risk, i.e., when $\lambda = 0$, under Assumptions A1-A4 in the Appendix, we have $UC_{MES} \xrightarrow{d} \mathcal{N}(0,1)$.*

When the estimation period $T$ is much larger than the evaluation period $n$, the unconditional coverage test is simplified because it does not require evaluations of $R_{MES}$ and $\Sigma_0$. Given these results, it is possible to define a test statistic $UC^C_{MES}$ that explicitly takes into account the estimation risk while having a standard limit distribution for any $\lambda$ with $0 \leq \lambda < \infty$ when $T$ and $n$ tend to infinity. The feasible robust UC backtest statistic is

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

where $\widehat{\mathbb{V}}_{as}(\widehat{\theta}_T) = \widehat{\Sigma}_0/T$ is a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_T$ and $\widehat{R}_{MES}$ is a consistent estimator of $R_{MES}$. In Appendix E, we propose an estimator for the vector $R_{MES}$ that we can easily implement.

## 3.2 Independence test

To test the independence hypothesis $\mathrm{H}_{0,IND}: \rho_1 = .... = \rho_m = 0$, we use a Portmanteau Box-Pierce test applied to the sequence of cumulative joint violation forecasts. We define the Box-Pierce test statistic as follows

$$IND_{MES} = n\sum_{j=1}^{m}\widehat{\rho}^2_{nj},$$

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

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

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

**Theorem 3** *Under Assumptions A1-A4 in the Appendix, we have:*

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

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

$$\Delta_{ij} = \delta_{ij} + \lambda R_i' \Sigma_0 R_j,$$

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

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

The proof of Theorem 3 is reported in Appendix F. The test statistic $IND_{MES}$ has an asymptotic distribution, which is a weighted sum of chi-squared variables. The weights depend on the asymptotic variance-covariance matrix of the estimator $\widehat{\theta}_T$, 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 in the Appendix, we have* $IND_{MES} \xrightarrow{d} \chi^2(m)$*.*

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

$$IND_{MES}^{C} = n \widehat{\rho}_n^{(m)\prime} \widehat{\Delta}^{-1} \widehat{\rho}_n^{(m)},$$

where $\widehat{\Delta}$ is a consistent estimator of $\Delta$, such that $\widehat{\Delta}_{ij} = \delta_{ij} + n \widehat{R}_i' \widehat{\mathbb{V}}_{as}(\widehat{\theta}_T) \widehat{R}_j$, where $\widehat{\mathbb{V}}_{as}(\widehat{\theta}_T)$ is a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_T$ and $\widehat{R}_j$ is a consistent estimator of $R_j$. In Appendix E, we provide a simple method to estimate $R_j$. When $T$ and $n$ tend to infinity, the robust statistic $IND_{MES}^{C}$ converges to a chi-squared distribution with $m$ degrees of freedom regardless of the relative value of $n$ and $T$.

### 3.3 Monte Carlo simulations

This section assesses the finite sample properties of our two backtest statistics computed with and without taking into account the estimation risk. We consider two definitions of the MES to study the properties of our tests in various settings. First, we consider the particular case of a marginal (time-invariant) MES, as in Acharya et al. (2017). This configuration allows us to easily control for the degree of misspecification of firm risk and financial interdependencies and thus to assess the power of our tests in relevant cases. Second, we consider a conditional (time-varying) MES based on a multivariate GARCH-DCC model, as in Brownlees and Engle (2017). In the sequel, we briefly present the two data generating processes (DGP) and the results of our Monte Carlo simulations.

**Backtesting marginal MES.** To evaluate the empirical size and power of our tests in the case of a marginal MES, we consider a bivariate normal distribution for the daily demeaned returns $Y_t = (Y_{1t}, Y_{2t})'$, such that $Y_t = \Sigma^{1/2} z_t$, with $Y_{1t}$ being the firm return and $Y_{2t}$ being the market return and where $z_t$ denotes an i.i.d. Gaussian vector error process with $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I_2$. Under the null, the variance-covariance matrix $\Sigma$ is defined as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $\sigma_1$ and $\sigma_2$ denote the volatility of firm and market returns, respectively, and $\rho$ is the correlation between the two returns. We calibrate the parameters $\theta_0 = \left(\sigma_1^2, \sigma_2^2, \rho\right)'$ using the daily log-returns of Bank of America and the CRSP market value-weighted index over the period from January 1, 2012, to December 30, 2016.[9] Notice that, under the normality assumption, the MES of the bank does not depend on the market volatility.[10]

For each Monte Carlo replication $b$, we simulate the series $\{y_{1t}^{(b)}, y_{2t}^{(b)}\}_{t=1}^{T+n}$, and we estimate the variance-covariance parameters $\theta_0$ using the first $T$ simulated observations $\{y_{1t}^{(b)}, y_{2t}^{(b)}\}_{t=1}^{T}$. Then, using the estimated parameters $\widehat{\theta}_T^{(b)}$, we compute the cumulative violation process $H_t(\alpha, \widehat{\theta}_T^{(b)})$ for the out-of-sample periods $t = T+1, \ldots, T+n$. For a finite sample size $T$, the violations are affected by the estimation risk due to the difference between the estimate $\widehat{\theta}_T^{(b)}$ and the true parameter value $\theta_0$. Finally, using the series $\{H_t(\alpha, \widehat{\theta}_T^{(b)})\}_{t=T+1}^{T+n}$, we compute the test statistics $UC_{MES}$ and $IND_{MES}$, without estimation risk correction, and the corresponding robust test statistics $UC_{MES}^C$ and $IND_{MES}^C$. The simulation study is based on $10,000$ replications. In addition, we consider various in-sample sizes ($T = 250$, $500$, and $2,500$) and out-of-sample sizes ($n = 250$ and $500$) to illustrate the impact of $T$ and $n$ on the estimation error and the small-sample properties of the

[9] The estimated values of unconditional variances and correlation, $\sigma_1^2$, $\sigma_2^2$, and $\rho$, are equal to 3.506, 0.722, 0.663, respectively.

[10] Under these assumptions, the MES for the bank return has a closed-form expression (Brownlees and Engle, 2017) given by $MES_1(\alpha) = -\rho\sigma_1\phi\left(\Phi^{-1}(\alpha)\right)/\alpha$, where $\phi(.)$ and $\Phi(.)$ denote the pdf and the cdf of a standard normal distribution. We can also express the MES as the product of the beta of the firm by the ES of the market return (Benoit et al., 2017), such as $MES_1(\alpha) = \beta_1 ES_2(\alpha)$, with $\beta_1 = \rho\sigma_1/\sigma_2$, and $ES_2(\alpha) = \mathbb{E}(Y_{2t}|Y_{2t} \leq VaR_2(\alpha))$.

backtests. Finally, we set the coverage rate $\alpha$ for the MES to 5%, and we compute the empirical size as the rejection frequency at a nominal level of 5%.

Because the MES depends on risk levels (volatilities) and dependencies (correlation), we propose various experiments designed to assess the capabilities of our tests to detect misspecification on these two quantities. We consider three misspecified models given by $Y_t = \widetilde{\Sigma}^{1/2} z_t$ under the alternative hypothesis:

$$\widetilde{\Sigma} = \begin{pmatrix} \widetilde{\sigma}_1^2 & \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 \\ \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 & \widetilde{\sigma}_2^2 \end{pmatrix},$$

- $\mathrm{H}_1(A_1)$: Undervalued variance of firm returns, $\widetilde{\sigma}_1^2 = (1-\tau)\sigma_1^2$, with $0 < \tau < 1$.
- $\mathrm{H}_1(A_2)$: Undervalued variance of market returns, $\widetilde{\sigma}_2^2 = (1-\tau)\sigma_2^2$, with $0 < \tau < 1$.
- $\mathrm{H}_1(A_3)$: Undervalued correlation between firm and market returns, $\widetilde{\rho} = (1-\tau)\rho$, with $0 < \tau < 1$.

Under these alternatives, the variance-covariance matrix is misspecified and underestimates the variance of return ($A_1$ and $A_2$) or the correlation ($A_3$) by a rate $\tau$, equal to either $25\%, 50\%$, or 75%. Such undervaluations of risk or dependence induce an underestimated MES and, in fine, an undervaluation of the systemic risk. In the case of alternatives, we use the first $T$ simulated observations to estimate the variance-covariance parameters $\theta_0$ under the constraint $\mathrm{H}_1$.[11] Denote the vector of estimated and constrained parameters obtained in the $b^{th}$ replication by $\widehat{\theta}_T^{(b)}$. We compute the test statistics from the misspecified cumulative violation process $H_t(\alpha, \widehat{\theta}_T^{(b)})$ for $t = T+1, ..., T+n$, and we compute the empirical power as the rejection frequency at a 5% nominal level. To take into account the potential size distortions for small $T$, our reported powers are all size-corrected.

Table 1 displays the empirical sizes and powers for the $UC_{MES}$, $UC_{MES}^C$, $IND_{MES}$, and $IND_{MES}^C$ tests. Four main results are noteworthy. First, the empirical sizes of the UC and IND tests, with or without correction for estimation risk, converge to the nominal level when both $T$ and $n$ increase. However, for small in-sample sizes $T$, the $UC_{MES}$ test exhibits severe size distortions

[11] For each alternative, the reduction rate $\tau$ is applied to the true value of the parameter instead of its estimated counterpart to avoid variations of the constrained parameters across replications. To ensure positive semi-definiteness of the estimate of $\widetilde{\Sigma}$ for any $\tau$, the other parameters are estimated by using the constrained maximum likelihood. This framework makes it possible to illustrate the power of our tests while taking into account the estimation risk. Notice that the data generating process considered under $\mathrm{H}_1$ $(A_2)$ leads to a misspecified MES: even if MES does not depend on the market volatility $\sigma_2$, the other estimated parameters (firm return volatility and correlation) do not converge to their true value.

due to estimation errors. For instance, for $T = 250$ and $n = 500$, the $UC_{MES}$ test is substantially oversized as its empirical size (11.99%) is twice the nominal size. These size distortions increase with the out-of-sample size $n$, as the test statistic $UC_{MES}$ diverges when $T$ is fixed and small, and $n$ tends to infinity. Second, for small $T$ samples, we recommend the use of robust test statistics that properly control for estimation risk. Empirical sizes of the robust test statistic $UC^C_{MES}$ are close to the nominal size of 5% for all reported samples. For instance, for $T = 250$ and $n = 500$, its empirical size is 5.53%. However, the robust $IND^C_{MES}$ backtest exhibits a slower rate of convergence and displays slight size distortions in small samples.[12] Third, our tests demonstrate good capacity for detecting various misspecifications in the variance-covariance matrix used to compute the MES. The empirical power of UC increases with the misspecification rate $\tau$ and the out-of-sample size $n$. The simulations for very large $n$ (not reported) confirm that our UC tests are consistent, as the rejection frequencies tend to 1. Interestingly, we obtain the highest empirical powers for the market volatility misspecification ($A_2$), indicating that a poor assessment of market distress may have severe consequences on the MES. Finally, for all alternatives, the empirical power of the IND test is still very low whatever $n$ and $T$. This result is consistent with the theory, as the underestimation of volatilities and correlation of returns does not have any consequences on the autocorrelations of the cumulative violation process $H_t$. As a consequence, the IND backtest is non-sensitive to the alternatives $A_1$, $A_2$, and $A_3$.

**Backtesting conditional MES.** For the conditional case, we consider a dynamic conditional correlation (DCC) model as in Brownlees and Engle (2017). This model is widely used in the context of systemic risk since it is able to well reproduce most of the stylized facts on financial data and to capture the interdependencies observed between firm and market returns. Formally, we assume that the vector process of demeaned returns is now defined as $Y_t = \Sigma_t^{1/2} z_t$, where $\Sigma_t$ denotes the conditional variance-covariance matrix and $z_t$ is as defined previously. Under the null, we define the conditional variance-covariance matrix $\Sigma_t$ as $\Sigma_t = D_t R_t D_t$, where $D_t = diag\{\sigma_{1t}; \sigma_{2t}\}$ is a diagonal matrix that contains conditional volatilities and $R_t$ is the conditional correlation matrix of $Y_t$. We consider a GJR-GARCH specification for the conditional variances of the firm and market returns (Glosten et al., 1993; Rabemananjara and Zakoian, 1993),

$$\sigma_{it}^2 = \alpha_{i0} + \alpha_{i1} Y_{i,t-1}^2 + \alpha_{i2} Y_{it-1}^2 \mathbb{1}\left(Y_{i,t-1} < 0\right) + \alpha_{i3} \sigma_{it-1}^2 \quad \forall i = 1, 2.$$

The DCC specification imposes a time-varying correlation structure on the standardized returns

[12] Using an independence test for backtesting expected shortfall, Du and Escanciano (2017) obtain similar size distortions for small sample sizes.

$\epsilon_{1t} = Y_{1t}/\sigma_{1,t}$ and $\epsilon_{2t} = Y_{2t}/\sigma_{2,t}$ through the pseudo-correlation matrix $Q_t$,

$$Q_t = (1 - a - b)\overline{Q} + a\epsilon_{t-1}\epsilon'_{t-1} + bQ_{t-1},$$

where $\overline{Q}$ is the unconditional correlation matrix and $a$ and $b$ are two non-negative parameters such that $a + b < 1$. The conditional correlation matrix is obtained by rescaling $Q_t$, such as $R_t = (diag\ Q_t)^{-1/2} Q_t (diag\ Q_t)^{-1/2}$. In the sequel, we refer to this conditional specification as the GARCH-DCC model.[13] The Monte Carlo study design is similar to that previously described, except that we now consider $1,000$ replications. We consider four misspecified models given by $Y_t = \widetilde{\Sigma}_t^{1/2} z_t$ under the alternative hypothesis:

- $\mathrm{H}_1(B_1)$: Undervalued conditional variance of firm returns, $\widetilde{\sigma}^2_{1t} = (1-\tau)\sigma^2_{1t}$, with $0 < \tau < 1$.
- $\mathrm{H}_1(B_2)$: Undervalued conditional variance of market returns, $\widetilde{\sigma}^2_{2t} = (1-\tau)\sigma^2_{2t}$, with $0 < \tau < 1$.
- $\mathrm{H}_1(B_3)$: Undervalued conditional correlation, $\widetilde{\rho}_t = (1-\tau)\rho_t$, with $0 < \tau < 1$.
- $\mathrm{H}_1(B_4)$: Misspecification of the dynamics of firm and market returns, with $Y_t = \Sigma^{1/2} z_t$.

Under the alternatives $B_1$-$B_3$, the GARCH-DCC model is misspecified because the conditional variances or the conditional correlation are misleading by a rate $\tau$, set to either $25\%, 50\%$, or $75\%$. In contrast, the alternative $B_4$ induces a misspecification of the dynamics of returns: the cumulative violation process is computed with a variance-covariance matrix that is assumed to be constant over time, whereas the true matrix is time-varying and exhibits a leverage effect. We expect that this setting creates autocorrelated violations, which should be detected by our independence test.

Table 2 provides the empirical size and size-corrected power at a 5% nominal level of the conditional MES backtests. Our results are similar to those obtained for the unconditional setting, and the key takeaways are the following. First, we observe size distortions for the UC backtests when there is no correction for the estimation risk and the estimation sample size $T$ is small. For a fixed size $T$, these distortions increase with the out-of-sample size $n$. We also observe that the impact of estimation risk is amplified compared to the unconditional case due to the increase in the number of estimated parameters (11 in the case of the GARCH-DCC model). As expected, the estimated parameters converge to the true parameter values and the impact of estimation error vanishes asymptotically. Second, our correction for estimation risk is efficient because it leads to

[13] We set the parameters $(\alpha_{10}, \alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{20}, \alpha_{21}, \alpha_{22}, \alpha_{23}, a, b)$ at their maximum likelihood parameter estimates, obtained using the same dataset as for the marginal MES. The corresponding estimated values are $(0.149, 0.065, 0.075, 0.855, 0.057, 0.001, 0.268, 0.786, 0.041, 0.754)$. The unconditional correlation is still set to 0.663.

precise inference, even for small $T$ and $n$. The robust UC test statistic has an empirical size that is generally close to the nominal size, typically about 4% for $T = 250$ and $n = 500$. The robust IND test statistic is slightly over-sized for finite sample sizes. Third, our backtests display good power performance. The UC tests generally detect well the misspecified alternatives $B_1$, $B_2$, and $B_3$, and we verify that there is a general improvement in power as the sample size $n$ increases, suggesting that these tests are consistent for these alternatives. Fourth, the results on the empirical power suggest complementarity between UC and IND tests. As in the unconditional case, the UC test displays good power performance against alternatives $B_1$, $B_2$, and $B_3$, while the empirical power of the IND test is still very low, except for large risk undervaluations. By definition, the IND backtest demonstrates better abilities to detect misspecification errors in model dynamics. Under alternative $B_4$, the IND test has an empirical power that is almost twice that of the UC test. As observed in the VaR backtesting literature (Christoffersen, 1998), the combination of UC and IND tests permits the detection of a larger spectrum of potential misspecifications in the dynamic models.

TABLES 1-2 ABOUT HERE

# 4 Backtesting other systemic risk measures

A great advantage of our backtesting approach lies in its simple extension to any systemic risk measures defined as a function of the MES or the CoVaR. In the sequel, we focus our analysis on the SES of Acharya et al. (2017), the SRISK of Acharya et al. (2012) and Brownlees and Engle (2017), and the $\Delta$CoVaR of Adrian and Brunnermeier (2016), as they are the most prominent in the literature.[14]

## 4.1 Backtesting SES and SRISK

Brownlees and Engle (2017) define the SRISK as the expected capital shortfall of a financial institution, conditional on a crisis affecting the entire financial system. The capital shortfall, denoted $CS_{1t}$, is defined as the capital reserves that the firm needs to hold to conform with regulation and/or prudential management minus firm equity. Formally, we define the capital shortfall of the firm indexed by 1 on day $t$ as $CS_{1t} = k\left(L_{1t} + W_{1t}\right) - W_{1t}$, where $L_{1t}$ is the book value of debt, $W_{1t}$ is the market value of firm equity, and $k$ is a prudential ratio. We define a systemic event as a

[14] Our backtests apply to many other measures, such as the $\Delta$CoVaR proposed by Girardi and Ergün (2013), the component expected shortfall (CES) of Banulescu and Dumitrescu (2015), or the delta conditional expected shortfall ($\Delta$CoES) of Ferreiro (2018), among others. However, they cannot be applied to the class of network systemic risk measures, such as those proposed by Billio et al. (2012) or Hué et al. (2019).

market decline below a given threshold over a time horizon $h$. Denote by $\widetilde{Y}_{t+h} = (\widetilde{Y}_{1t+h}, \widetilde{Y}_{2t+h})'$, or equally by $Y_{t+1:t+h} = (Y_{1t+1:t+h}, Y_{2t+1:t+h})'$, the vector of multi-period arithmetic firm and market returns between $t+1$ and $t+h$ and by $\widetilde{V}aR_{2t+h} \equiv VaR_{2t+1:t+h}(\alpha)$ the corresponding $\alpha$-VaR of $\widetilde{Y}_{2t+h}$ used as the market threshold. We define the SRISK of the firm on day $t$ for an horizon $h$ as

$$\begin{aligned} SRISK_{1t}(h) &= \mathbb{E}_t(CS_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)) \\ &= k\mathbb{E}_t(L_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)) - (1-k)\,\mathbb{E}_t(W_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)), \end{aligned}$$

where $\mathbb{E}_t(.)$ denotes the conditional expectation with respect to $\Omega_t$. To compute this expectation, Brownlees and Engle assume that the debt is constant during a systemic event, i.e., $\mathbb{E}_t(L_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)) = L_{1t}$. Furthermore, they introduce the concept of long-run marginal expected shortfall (LRMES) to compute the expected market value of the firm. The LRMES is simply a MES defined in terms of cumulative returns after $h$ periods:[15]

$$MES_{1t}(\alpha; h) = \mathbb{E}_t(\widetilde{Y}_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)).$$

Thus, we get $\mathbb{E}_t(W_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)) = W_{it}(1 + \mathbb{E}_t(\widetilde{Y}_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{V}aR_{2t+h}(\alpha)))$, and finally, the SRISK is defined as[16]

$$SRISK_{1t}(h) = k\ L_{1t} - (1-k)\,W_{1t}(1 + MES_{1t}(\alpha; h)). \tag{6}$$

A similar systemic risk measure, the SES, is proposed by Acharya et al. (2017). The SES represents the amount that a bank equity drops below its target level (defined as a fraction $k$ of assets) in the event of a systemic crisis when the aggregate capital is less than $k$ times the aggregate assets. Acharya et al. (2017) provide theoretical justification for how SES relates to MES and show that:

$$SES_{1t}(h) = (k\ LV_{1t} - 1 - \Pi\ MES_{1t}(\alpha; h) + \Delta)\ W_{1t},$$

where $\Pi$ and $\Delta$ are two constant terms with $\Pi > 0$ and $LV_{1t} = (L_{1t} + W_{1t})/W_{1t}$ denotes the quasi-leverage ratio.

These two formulas show how the SRISK and the SES extend the LRMES to take into account both the liabilities and the size of the financial institution. Assuming that the level of debt cannot be negotiated in the case of a systemic event implies that the level of debt is known at time $t$. Thus,

[15] For $h = 1$, the LRMES boils down to the MES given by Equation (1).

[16] Equation (6) is similar to the definition reported by Brownlees and Engle (2017) in Equation (1) (page 52) of their paper, except that we do not adopt the same sign convention for the MES. Here, we have defined the MES as a negative quantity according to our Equation (1).

the need for inference for these risk measures comes only from the expected value of the firm at horizon $h$, which can break down into an initial value known at time $t$ and the long-run MES. More generally, we define a MES-based systemic risk measure as a risk measure for which the need for inference comes only from its MES component.

**Definition 5** *A MES-based systemic risk measure $RM_{1t}(h)$ for firm 1 at time $t$ and for an horizon $h \geq 1$ is defined as a deterministic function of the (long-run) MES with*

$$RM_{1t}(h) = g_t(MES_{1t}(\alpha; h), X_t),$$

*where $g_t(.)$ is a monotonic (in MES) function and $X_t$ is a set of variables that belong to $\Omega_t$.*

By definition, testing the validity of any MES-based systemic risk measure is equivalent to testing the validity of the LRMES itself. The intuition is as follows. We can rewrite any MES-based risk measure as a function of the CoVaR (defined in terms of cumulative returns), with

$$RM_{1t}(h) = \int_0^1 g_t(Co\widetilde{V}aR_{1t+h}(\beta, \alpha, \theta_0), X_t) \mathrm{d}\beta.$$

This expression is similar to that obtained for the MES in Equation (3). Furthermore, the violation process $h_t(\alpha, \beta, \theta_0, X_t)$ associated with the quantity $g_t(Co\widetilde{V}aR_{1t+h}(\beta, \alpha, \theta_0), X_t)$ is equivalent to the violation process $h_t(\alpha, \beta, \theta_0)$ associated with the CoVaR, $Co\widetilde{V}aR_{1t+h}(\beta, \alpha, \theta_0)$: both binary processes will take value 1 on the same dates. Thus, the cumulative joint violation process used to backtest the SRISK, the SES, or any MES-based measure is equivalent to the cumulative joint violation process used to backtest the MES itself (see Appendix G). As a consequence, the UC and IND tests are identical to those reported in Section 3.

To implement the aforementioned test statistics, we need to compute the cumulative violation process $H_t(\alpha, \theta_0)$ associated with the joint distribution of the multi-period returns $(\widetilde{Y}_{1t+h}, \widetilde{Y}_{2t+h})$ after $h$ periods. For $h > 1$, it is generally not available in closed form for the class of standard dynamic models (e.g., GARCH-type models) typically used for daily returns. The problem here is similar for the estimation of the LRMES itself (see Brownlees and Engle, 2017, for more details). However, it is straightforward to implement a simulation-based procedure to obtain an estimate of the cdf of the joint distribution for any horizon $h$. In the empirical application, we simulate a large number of paths of returns for the periods $t+1$ to $t+h$, conditional on the information available at time $t$, and we compute the corresponding cumulative returns. Then, we apply a kernel estimator to the simulated cumulative returns to estimate the cdf of their joint distribution.

### 4.2 Backtesting ΔCoVaR

Our testing procedure can be extended to assess the validity of the ΔCoVaR (Adrian and Brunnermeier, 2016). In the sequel, we define this indicator as the difference between the conditional VaR (CoVaR) of an institution conditional on the financial system being in distress and the CoVaR conditional on the financial system being in its median state.[17] Adrian and Brunnermeier define the stress for the financial system as a situation in which the market return $Y_{2t}$ is equal to its $VaR_{2t}(\alpha)$ and consider a quantile regression model for the estimation of the CoVaR. A more general approach consists of defining the financial stress as a situation in which $Y_{2t} \leq VaR_{2t}(\alpha)$, as in Girardi and Ergün (2013). In both cases, we can estimate the CoVaR with M-GARCH-type models and some usual results about truncated distributions. Similarly, we can represent the normal or median state of the system by a situation in which $VaR_{2t}(\beta_{\inf}) \leq Y_{2t} \leq VaR_{2t}(\beta_{\sup})$, with $\alpha < \beta_{\inf} < \beta_{\sup}$, for instance $\beta_{\inf} = 25\%$ and $\beta_{\sup} = 75\%$. Formally, if we denote by $Y_{2t}$ the return of a portfolio of financial institutions (financial system), then the ΔCoVaR of the financial firm is given by[18]

$$\Delta\text{CoVaR}_t(\alpha) = CoVaR_{1t}(\alpha, \alpha, \theta_0) - CoVaR_{1t}(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0), \tag{7}$$

where $CoVaR_{1t}(\alpha, \alpha, \theta_0)$ (hereafter, stressed CoVaR) verifies $\Pr(Y_{1t} \leq CoVaR_{1t}(\alpha, \alpha, \theta_0) | Y_{2t} \leq VaR_{2t}(\alpha, \theta_0); \Omega_{t-1}) = \alpha$ and $CoVaR_{1t}(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)$ (hereafter, median CoVaR) verifies $\Pr(Y_{1t} \leq CoVaR_{1t}(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0) | VaR_{2t}(\beta_{\inf}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\sup}, \theta_0); \Omega_{t-1}) = \alpha$. To backtest the stressed CoVaR and the median CoVaR, we define two violations through the following binary processes: $h_t(\alpha, \alpha, \theta_0) = \mathbb{1}((Y_{1t} \leq CoVaR_{1t}(\alpha, \alpha, \theta_0)) \cap (Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)))$, and $h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0) = \mathbb{1}((Y_{1t} \leq CoVaR_{1t}(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)) \cap (VaR_{2t}(\beta_{\inf}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\sup}, \theta_0)))$. The logic of the test is then similar to that used for backtesting the MES. If the risk model is correctly specified, the two violation processes satisfy the mds property with $\mathbb{E}\left( h_t(\alpha, \alpha, \theta_0) - \alpha^2 \middle| \Omega_{t-1}\right) = 0$ and $\mathbb{E}( h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0) - \alpha(\beta_{\sup} - \beta_{\inf}) | \Omega_{t-1}) = 0$. By exploiting the mds property, we can propose various types of backtests for the ΔCoVaR or for each of its constituents. As for the MES, we consider an unconditional coverage test with the following joint null hypothesis:

$$\text{H}_{0,UC} : \mathbb{E}\left(\mathbf{h}_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)\right) = \mu,$$

[17] Adrian and Brunnermeier provide various definitions of CoVaR depending on the direction of the conditioning. To maintain consistency with the conditioning event used in the MES definition, we consider hereinafter the indicator referred to as *Exposure*-ΔCoVaR, which is a measure of an individual institution's exposure to system-wide distress.

[18] Adrian and Brunnermeier deal with the special case in which $\beta_{\inf} = \beta_{\sup} = 0.5$, implying that the $CoVaR_{1t}(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)$ is null. For ease of presentation, we only present the ΔCoVaR at horizon $h = 1$.

where $\mathbf{h}_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0) = (h_t(\alpha, \alpha, \theta_0), h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0))'$ denotes the vector of violations, and where $\mu = \left(\alpha^2, \alpha\left(\beta_{\sup} - \beta_{\inf}\right)\right)'$. Here, the intuition is similar to the backtests proposed for the multi-level VaR, i.e., the VaR defined for a finite set of coverage rates (see Francq and Zakoïan, 2016, for estimation issues). We can cite the tests proposed by Hurlin and Tokpavi (2006), Pérignon and Smith (2008), Colletaz et al. (2013), Leccadito et al. (2014), and Wied et al. (2016), among others. These multi-level VaR backtesting procedures have been recently adapted to test the validity of ES forecasts by Kratz et al. (2018) and Couperier and Leymarie (2018). Given this joint null hypothesis, we consider a Wald test statistic with

$$UC_{\Delta\text{CoVaR}} = n\left(\overline{\mathbf{h}}(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T) - \mu\right)' \gamma^{-1} \left(\overline{\mathbf{h}}(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T) - \mu\right),$$

where $\overline{\mathbf{h}}(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T)$ denotes the out-of-sample mean of $\mathbf{h}_t(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T)$ and $\gamma$ is the conditional variance-covariance matrix of $\mathbf{h}_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)$ such that

$$\gamma = \begin{pmatrix} \alpha^2(1-\alpha^2) & -\alpha^3(\beta_{\sup} - \beta_{\inf}) \\ -\alpha^3(\beta_{\sup} - \beta_{\inf}) & \alpha(\beta_{\sup} - \beta_{\inf})(1 - \alpha(\beta_{\sup} - \beta_{\inf})) \end{pmatrix}.$$

Under the null hypothesis, the sequence $\{\mathbf{h}_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0) - \mu\}_{t=T+1}^{T+n}$ is a mds with variance equal to $\gamma$. Thus, the Lindeberg-Levy central limit theorem implies that $UC_{\Delta\text{CoVaR}}$ converges to a chi-squared distribution with two degrees of freedom when we evaluate the test statistic in $\theta_0$ instead of $\widehat{\theta}_T$. The asymptotic distribution remains valid for the feasible statistic $UC_{\Delta\text{CoVaR}} \equiv UC_{\Delta\text{CoVaR}}(\widehat{\theta}_T)$ as soon as $T \to \infty$ and $n \to \infty$, with $\lambda = n/T \to 0$, i.e., without estimation risk. The proof is reported in Appendix H. In the general case $n/T \to \lambda < \infty$, the test statistic $UC_{\Delta\text{CoVaR}}$ is no longer chi-squared distributed. However, by considering the same reasoning as in Section 3, we can derive a robust test statistic $UC^C_{\Delta\text{CoVaR}}$ that has an asymptotic chi-squared distribution regardless of the values $n$ and $T$. Monte Carlo simulations show that the robust test statistic provides satisfactory size performance regardless of the sample size (see Table 3).

Finally, if the $\Delta$CoVaR risk model fails the UC test, it is worth examining whether this rejection is due to the the CoVaR of the distress state of the financial system (stressed CoVaR) and/or of the median state of the financial system (median CoVaR). To identify the origin of the error, we propose two UC sub-tests defined as $\mathrm{H}_{0,UC}^{\text{distress}} : \mathbb{E}\left(h_t(\alpha, \alpha, \theta_0)\right) = \alpha^2$ and $\mathrm{H}_{0,UC}^{\text{median}} : \mathbb{E}\left(h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)\right) = \alpha\left(\beta_{\sup} - \beta_{\inf}\right)$. The testing procedure is then similar to that previously described.

# 5 Empirical application

In the section, we apply our UC and IND backtests to a panel of U.S. financial institutions to assess the validity of MES, SRISK, and $\Delta$CoVaR. In the first part, we describe the data and the empirical setup, with a special focus on the multiple testing problem. In the second part, we apply our UC and IND backtests and present our empirical results for short-term (one day) and medium-term (one month) systemic risk forecasts.

## 5.1 Data description and empirical setup

Our study focuses on the same panel of large U.S. financial institutions as that considered by Brownlees and Engle (2017). The dataset contains all U.S. financial firms with a market capitalization greater than 5 billion dollars as of the end of June 2007 and covers the period from January 3, 2000, to December 30, 2016. The panel is unbalanced because some companies were not traded continuously during the sample period (for instance, Lehman Brothers after its bankruptcy on September 15, 2008). The corresponding list of tickers and company names is reported in Appendix I. For each firm in the panel, we compute the daily logarithmic returns and market capitalization from CRSP data, and we consider the daily CRSP market value-weighted index return as the market return. For the SRISK, we collect the quarterly book value of total liabilities from Compustat.

To model the systemic indicators, we use a GARCH-DCC model as in Brownlees and Engle (2017), Engle et al. (2015), Idier et al. (2014), and Acharya et al. (2012), among others. We estimate the parameters by maximum likelihood over the $T$ in-sample daily observations assuming a normal joint distribution. We consider two types of estimation schemes: ($i$) a rolling-window scheme and ($ii$) a recursive scheme.[19] In the rolling-window scheme, we estimate the parameters using the most recent $T$ daily observations up to the end of each month. Here, we consider a 2-year rolling window with $T = 500$. In the recursive scheme, we estimate the parameters using all available information from January 3, 2000, up to the end of each month. As a result, we observe a gradual increase in $T$ as we move forward in the exercise. For each company, we compute MES, SRISK, and $\Delta$CoVaR at the end of each month from January 2005 to December 2016. As in Brownlees and Engle (2017), we only consider positive estimates for the SRISK because it represents a capital shortfall, and the negative values are set to zero. Finally, we compute the systemic risk forecasts at coverage level $\alpha$ equal to 5%, and the prudential capital ratio $k$ is set to 8%.

[19] The Stern-NYU V-Lab website uses the recursive estimation scheme to compute the SRISK for a large set of U.S. and European financial firms.

Our empirical assessment requires testing several hypotheses simultaneously. Each month, we apply (a maximum of) 95 backtests representing the maximum number of firms in our panel. This framework causes a multiple testing problem.[20] To that end, we use a controlling method based on the Family Wise Error rate (FWE). It allows control for the probability of erroneously rejecting at least one of the null hypotheses (UC or IND) among the institutions. In the sequel, we consider the Bonferroni procedure, which provides a strong control of the FWE (Romano et al., 2008). The method is applied as follows. Denote by $\gamma$ the significance level used for a single backtest and by $M$ the number of backtests to be applied simultaneously, with typically $M \leq 95$ in our case. For each individual backtest, i.e., each firm $s = 1, \ldots, M$, we compute an individual $p$-value $p_s$, and we reject the null hypothesis at level $\gamma$ if $p_s < \gamma/M$ for all backtests.

### 5.2 Empirical results for short-term MES and SRISK forecasts

In this section, we apply our backtests to assess the validity of MES and SRISK by considering a forecasting time horizon of one day, i.e., $h = 1$. We evaluate the empirical validity of these indicators for Bank of America (BAC), JP Morgan (JPM), American International Group (AIG), and Lehman Brothers (LEH). Figure 1 displays the case of the UC hypothesis and Figure 2 presents the case of the IND hypothesis. The rejection of the null hypothesis at a 5% significance level is represented by shaded areas. Finally, the backtests are computed with $n = 250$ out-of-sample observations.[21]

For each of these firms, Figures 1 and 2 display the one-day-ahead forecasts of the MES (blue line) and SRISK (red line) issued from a recursive estimation window. The figures reveal that both MES and SRISK increase during periods of financial instability, particularly during the subprimes crisis (2008-2009) and the European debt crisis (2011-2012). At first glance, these measures capture well the impact of the financial crisis on the capital shortfalls of these four institutions. However, this statement is less clear when one considers the empirical validity of these indicators. Figure 1 highlights a significant number of rejections of the UC hypothesis during the 2008-2009 financial crisis, and as a result, one can question the validity of the unconditional coverage property during this period. The GARCH-DCC model produces short-term MES forecasts that are associated with cumulative violations that are not observed with the right out-of-sample frequency. Furthermore,

[20] We only control for the multiple testing problem in the cross-sectional dimension. We have decided to not control for the multiple testing problem in the time series dimension, as we do not aim to provide a "global test" over the whole period.

[21] For ease of presentation, we only report the backtests that are robust to estimation risk. Overall, the unadjusted backtests display more rejections, especially for the rolling estimation scheme, as the ratio $\lambda = n/T$ increases. The corresponding results are available upon request.

we note for Lehman Brothers severe rejections of the unconditional coverage hypothesis just before its bankruptcy, indicating a sharp change in its market conditions. For the rest of the sample, the unconditional coverage hypothesis is generally not rejected, and we can conclude that MES and SRISK forecasts disclose the true level of systemic risk.

Figure 2 looks at the empirical validity of the IND hypothesis. The Box-Pierce test statistic is computed for a lag order $m = 5$. Strikingly, the cluster of rejections of the UC hypothesis observed during the global financial crisis is no longer experienced for the IND hypothesis. Instead, we observe for AIG and BAC a cluster of rejections of the independence hypothesis during the European debt crisis. The two statistics take alternative approaches to identify measurement errors that allow for the separate detection of the two recent episodes of financial market instability. Furthermore, we also identify for JPM (and to a lesser extent for BAC and AIG) several rejections of the IND hypothesis in 2016-2017, which should warn regulators of the difficulties encountered by MES and SRISK in capturing the correct dynamics of the firms and market returns in recent years. In Appendix J, we provide various robustness check exercises by considering (i) a rolling window estimation scheme and (2) a larger out-of-sample size, with $n = 500$. We observe the same overall findings.

FIGURES 1-2 ABOUT HERE

To complete this study, we compute the UC and IND backtests for the full list of tickers and report the rejection frequencies of the tests. As discussed above, we apply both tests for up to 95 firms simultaneously, and we implement a controlling method to address the multiple testing problem. Figure 3 displays the rejection rates of the UC test (top panel) and of the IND test (bottom panel) adjusted via the Bonferroni controlling method. We still consider a recursive estimation scheme and $n = 250$ out-of-sample observations. The UC hypothesis is rejected for 20% to 40% of the firms in the panel during the 2008-2009 subprime crisis. The GARCH-DCC model fails to capture the average level of systemic risk in times of crisis for a large set of firms. Our results are consistent with Danielsson et al. (2016), who show that daily risk forecasts are subject to significant model risk during periods of financial distress, which are unfortunately when they are most needed. We observe less pronounced rejections of the IND hypothesis, and the 2008-2009 financial crisis is no longer identified. Thus, the non-autocorrelation property of the violations is more easily satisfied than the unconditional coverage in global market distress situations. However, we observe up to 20% of rejections during the European debt crisis. Finally, we also highlight a cluster of rejections of the independence assumption at the end of the sample period, i.e., 2015-2017.

Finally, we display in Figure 4 the results obtained for $n = 500$, i.e., twice the previous out-of-sample size. The UC and IND backtests identify the same periods of invalid risk forecasts, but the rejections are more frequent and more important due to the large size $n$ that increases the power of the tests. In this situation, we observe up to 80% of rejections of the UC hypothesis during the global financial crisis, providing evidence of the failure of the GARCH-DCC model to deliver valid systemic indicators for most of the financial institutions during this period. We obtain similar results when considering a rolling estimation scheme (see Appendix J).

FIGURES 3-4 ABOUT HERE

### 5.3 Empirical results for medium-term forecasts

Due to the difficulty for banks of immediately adjusting their capital structure, it is generally preferable to forecast systemic risk at a longer horizon than a single day. Here, we follow Brownlees and Engle (2017) and consider a forecasting horizon of one month, i.e., $h = 22$ days, for LRMES and SRISK. Appendix K provides a detailed description of the method used to compute both measures over a time horizon $h > 1$. However, the use of a longer forecasting horizon implies drastically increasing the size of the out-of-sample dataset used to assess the unconditional coverage property. Indeed, as we consider cumulative firm and market returns $\widetilde{Y}_{t+h} \equiv Y_{t+1:t+h}$, the computation of only one observation of the cumulative violation process $H_t(\alpha, \widehat{\theta}_T)$ requires $h$ observations of daily returns. Because the UC test statistic defined in Equation (5) depends on the out-of-sample mean of the violations $H_t(\alpha, \widehat{\theta}_T)$, its computation requires $n \times h$ daily observations. For instance, one year of out-of-sample observations (250 daily observations) and a forecasting horizon of 22 days allows us to compute only 11 observations of $H_t(\alpha, \widehat{\theta}_T)$. To obtain 100 observations of $H_t(\alpha, \widehat{\theta}_T)$, we need about 9 years of daily returns. As the UC test statistic converges with $n$, this is clearly problematic.

To address this issue, we propose a backtesting framework with overlapping blocks of data. Figure 5 summarizes the procedure for the case of $h = 22$. Instead of considering the sequence of cumulative returns $\{Y_{1:22}, Y_{23:44}, \dots\}$ (in blue) to compute the violations, we consider the sequence $\{Y_{1:22}, Y_{12:33}, Y_{23:44}, \dots\}$ for which two subsequent cumulative returns share a common component of 11 daily returns (in blue and red). We can view this method as a rolling window procedure applied to the out-of-sample observations that increases the number of observations available for the out-of-sample assessment and the power of our test. Obviously, it also creates dependence across the violations $H_t(\alpha, \widehat{\theta}_T)$.[22] Therefore, the property of independence of the cumulative joint violation process is no longer satisfied, and thus we cannot apply the IND backtest in this case.

Figure 6 reports the rejection frequencies of the UC backtest for SRISK (and LRMES) calculated for the full panel of firms. We estimate the GARCH-DCC parameters either with a recursive estimation (top panel) or with a rolling-window scheme (bottom panel). For each case, we assess the empirical validity of SRISK over a time horizon $h = 22$ with $n = 100$ out-of-sample observations, requiring 1111 daily log returns given the 11-day overlap window.[23] As mentioned above, the significance level is adjusted via the Bonferroni controlling method. We find that the one-month-ahead SRISK forecasts issued by the GARCH-DCC model are generally valid. For both estimation schemes, there are very few rejections of the UC hypothesis. Our findings differ drastically from those associated with daily forecasts, where we found large rejection frequencies.

Such a difference should be interpreted with cautious. One the one hand, it could be the sign that the risk model performs well at longer horizons, while it is less effective in capturing systemic risk over short-term horizons. This interpretation is in line with that of Brownlees and Engle (2017) who show that the predictive ability of SRISK is superior over longer horizons. On the other hand, it may be indicative of the limits in the rejection capabilities of the test for longer horizons. Indeed, the dependence induced by the overlapping scheme may decrease the power of the backtesting tests when considering longer horizons.

Finally, we note that the rolling estimation setup is characterized by more rejections than the recursive estimation setup. This highlights the fact that a widened sample period is more suitable to capture the long-run dynamics.

FIGURES 5-6 ABOUT HERE

### 5.4 Empirical results for ΔCoVaR forecasts

This section is devoted to the empirical evaluation of the ΔCoVaR daily forecasts. We compute the ΔCoVaR with probability levels $\alpha = 0.05$, $\beta_{\text{inf}} = 0.25$, and $\beta_{\text{sup}} = 0.75$. As emphasized by

[22]If the overlapping is limited and generates only weak dependence across violations, the central limit theorem still applies, and thus the UC test statistic remains normally distributed under the null. To correct the long-run variance estimate, we consider an HAC estimator (Newey and West, 1987) for estimating the variance of the violation process. Monte Carlo experiments (not reported) show that this adjusted test is well sized as soon as the overlap does not exceed 11 trading days for $h = 22$.

[23]The joint distribution of the cumulative returns $\widetilde{Y}_{t+h} \equiv (Y_{1t+1:t+h}, Y_{2t+1:t+h})'$ is not available in closed form, and thus the derivative of their cdf is more difficult to compute numerically. For this reason, we do not compute the robust statistics, and alternatively, we set $T$ to 500 to limit the impact of estimation risk.

Equation (7), if the $\Delta$CoVaR does not successfully pass the test, it is interesting to know whether this rejection stems from the CoVaR associated with the distress state (stressed CoVaR) or with the median state (median CoVaR). Thus, we also report the two UC sub-tests associated with each type of CoVaR to complete the analysis and identify which component of the $\Delta$CoVaR is misspecified.

Figure 7 reports the forecasts of $\Delta$CoVaR for AIG, BAC, JPM, and LEH and the rejection regions of the UC backtest in shaded areas. For each firm, we observe that the $\Delta$CoVaR predictions increase in 2008-2009 and in 2012, suggesting that this indicator is sensitive to systemic events and helpful for systemic risk monitoring. However, at the same time, we observe several rejections of the unconditional coverage hypothesis for the $\Delta$CoVaR forecasts. In this respect, our empirical results associated with the $\Delta$CoVaR and with the MES coincide fairly well. Such similarities are not surprising given the relationship between MES and CoVaR described by Equation (3).[24]

To characterize which component of the $\Delta$CoVaR is misspecified, Figures 8 and 9 report the results for separate backtests of the stressed CoVaR and the median CoVaR, respectively. We find that the CoVaR calculated in distress conditions generally increases more markedly during financial crisis than the CoVaR evaluated in median conditions. Furthermore, we observe that the rejection periods of the $\Delta$CoVaR identified in Figure 7 coincide with those of the stressed CoVaR, while the median CoVaR is overall not affected by misspecification (except for AIG). Thus, the GARCH-DCC model has noticeable problems in properly estimating the CoVaR in case of the financial crisis, as the corresponding frequency of out-of-sample violations is not correct.

Finally, the empirical assessment is extended to the full list of tickers. Figure 10 reports the rejection rates of the 95 U.S. firms for the $\Delta$CoVaR (top panel), the stressed CoVaR (middle panel), and the median CoVaR (bottom panel). We implement the Bonferroni controlling method and the corresponding results are based on a recursive estimation scheme and $n = 250$ out-of-sample observations. Our results confirm that the rejections of the risk model are generally observed for the $\Delta$CoVaR and for the stressed CoVaR, while the validity of the median CoVaR is generally not rejected. Furthermore, these rejections mainly occur during the 2008-2009 financial crisis, indicating that the $\Delta$CoVaR and the stressed CoVaR may be severely affected by forecasting errors during crisis periods. We obtain similar results for $n = 500$ with more pronounced rejections (not reported).

FIGURES 7-10 ABOUT HERE

[24]In the same spirit, Benoit et al. (2013) theoretically show that the higher the correlation between the returns of the SIFIs and the market are, the more likely it is that MES and CoVaR will lead to a convergent diagnostic.

# 6 Early warning system

A by-product of our backtesting procedure is to provide an early warning system (EWS) indicator that allows us to depict significant changes in the stability of the financial system. The idea is the following: a given risk model, say a GARCH-DCC, is used to produce a systemic risk measure, say MES or SRISK. If the model is correctly specified, we will observe the cumulative violations with the right frequency, and we will not reject the null hypothesis of unconditional coverage. On the contrary, if the model is misspecified and/or if the market conditions are likely to change, the model produces invalid systemic risk measures. Let us define an *adjusted* systemic risk measure produced by a potentially misspecified model but for which we do not reject the null hypothesis of unconditional coverage. We can use the difference observed between adjusted and unadjusted risk measures as a predictor of financial distress. This provides useful insights for monitoring the financial system on a real-time basis.

In the same spirit as done for VaR or ES by Gouriéroux and Zakoïan (2013), Boucher et al. (2014) or Lazar and Zhang (2019), we propose to adjust imperfect MES-based forecasts by considering the mean property of the cumulative joint violation process. Formally, we define an adjusted coverage level $\widetilde{\alpha}$ for the MES-based forecast such that the null hypothesis of unconditional coverage is valid, i.e., $\mathbb{E}(H_t(\widetilde{\alpha}, \theta_0)) = \alpha/2$. If the risk model is correctly specified then $\widetilde{\alpha} = \alpha$. We can obtain a feasible adjusted coverage level $\widetilde{\alpha}$ as the solution of the program

$$\widetilde{\alpha} = \underset{\alpha \in ]0,1[}{\arg\min} \left( \bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2 \right)^2 ,$$

where $\bar{H}(\alpha, \widehat{\theta}_T) = (1/n) \sum_{t=T+1}^{T+n} H_t(\alpha, \widehat{\theta}_T)$ denotes the mean of the cumulative joint violation process and $\alpha/2$ represents the expected value of the cumulative joint violation process under the null hypothesis of correct unconditional coverage. Given the definition of $\widetilde{\alpha}$, the adjusted forecast $MES_{1t}(\widetilde{\alpha}, \widehat{\theta}_T)$ is valid as we observe the corresponding cumulative joint violations with the right frequency. MES is appealing as an EWS because it reacts very quickly to changes in financial market conditions through its short-run information content. Thereafter, we devise an adjustment for MES, but we can easily generalize this adjustment to any MES-based systemic risk measure.

Considering the difference between the adjusted and unadjusted MES, we can build an indicator that depicts the unexpected changes in market conditions and provides early warning signals of distress in the financial system. Formally, this indicator, denoted $EWS_t\,(\alpha, \widetilde{\alpha})$, is defined as

$$EWS_t\,(\alpha, \widetilde{\alpha}) = MES_{1t}(\widetilde{\alpha}, \widehat{\theta}_T) - MES_{1t}(\alpha, \widehat{\theta}_T).$$

This quantity depicts the magnitude of model misspecification at time $t$ expressed in terms of systemic risk and represents the unexplained systemic risk component not anticipated by market practitioners. In line with our notations, a positive value of $EWS_t\left(\alpha,\widetilde{\alpha}\right)$ induces an unexpected growth in systemic risk, while a negative value indicates an unexpected decline in systemic risk.

Figure 11 reports the indicator $EWS_t\left(\alpha,\widetilde{\alpha}\right)$ for AIG, BAC, JPM, and LEH. The results are reported using a recursive window estimation scheme and $n = 250$ out-of-sample observations. In line with our previous results, we observe a large increase in $EWS_t\left(\alpha,\widetilde{\alpha}\right)$ during the 2007-2009 financial crisis and to a lesser extent during the European debt crisis. During these periods, the GARCH-DCC model is unable to provide valid systemic risk forecasts, and an adjustment of the coverage rate is required to get acceptable measures. Interestingly, we also observe a sharp increase in this indicator for Lehman Brothers at the beginning of 2007, substantially before its bankruptcy.

Finally, we perform the same exercise at an aggregated level, that is, summing our indicator over the 95 U.S. firms in our panel. Our aim is to evaluate the ability of the EWS for the financial system as a whole. Figure 12 displays the aggregated $EWS_t\left(\alpha,\widetilde{\alpha}\right)$. Our results are consistent to those observed at the firm level. There is a large increase of the indicator during the global financial crisis. This increase starts in mid-2007 confirming the ability of our indicator to detect reversals early. The highest value of our aggregated early warning indicator, 3.23, is observed in September 2008 and coincides with the collapse of Lehman Brothers. Similar results are obtained when considering a rolling window estimation scheme, and using $n = 500$ for the out-of-sample period.

In their recent work, Brownlees et al. (2020) evaluate the ability of CoVaR and SRISK to provide early warning signals of distress in the financial system. They conclude that even if CoVaR and SRISK may be helpful for identifying systemic institutions in periods of distress, neither indicator would be efficient for predicting when the next crisis is likely to occur. Our approach here is complementary: by examining the number of violations experienced by a systemic risk indicator, our aim is to detect systemic risk measurement errors caused by changes in global market conditions.

FIGURES 11-12 ABOUT HERE

# 7 Conclusion

This article develops a statistical procedure to backtest systemic risk measures. The tests are built-up in analogy with the recent backtests for ES proposed by Du and Escanciano (2017). Our procedure is consistent with the testing strategies applied by risk managers to assess the validity of market risk measures such as VaR and ES. Its implementation is easy since it only implies evaluating a cdf of a bivariate distribution. Furthermore, it allows performing a separate test for unconditional coverage and independence hypothesis (Christoffersen, 1998). Monte Carlo simulations show that for realistic sample sizes, the tests have good finite sample properties. Also, we pay particular attention to the consequences of estimation risk and derive a robust version of the test statistics.

We apply our backtests to large U.S. financial institutions considering a GARCH-DCC model for modeling systemic risk. We find that the short-term forecasts of systemic risk (one-day-ahead forecasts) do not satisfy the unconditional coverage hypothesis during the 2007-2009 financial crisis. The risk model does not capture the true level of systemic risk on a short-term basis, and it is not possible to accurately conclude which institution is (systemically) riskier than another, or to determine the right level of regulatory capital for the SIFIs. In contrast, we observe for longer forecasting horizons (one-month-ahead forecasts) that the unconditional coverage hypothesis is no longer rejected. Our overall recommendation for users of these tools, first and foremost academics and regulators, is assessing ex ante the validity of the risk forecasts before providing any interpretation of bank exposure to systemic risk.

As a by-product of our backtesting methodology, we develop a procedure that adjusts misleading systemic risk predictions. The technique consists of modifying the coverage level of a MES-based risk measure through the severity of the market distress event, such that the adjusted indicator satisfies the unconditional coverage assumption. This technique has been used for market risk measures such as ES and VaR (Boucher et al., 2014; Lazar and Zhang, 2019, etc.), but to date, it has not been available for systemic risk measures. To illustrate the merit of our method, we introduce an EWS indicator defined as the difference between the misleading forecast and its adjusted counterpart. Our EWS exhibits a sharp increase before the early signs of the crisis and takes its highest value during the historic collapse of Lehman Brothers. As a forward-looking measure, it may complete the toolbox used by academics and regulators to capture the accumulation of systemic risk and improve the allocation efficiency of regulatory capital among banks.

# TABLES

Table 1: Empirical rejection rates for backtesting MES(5%) at 5% nominal level (marginal case)

| | | $T=250$, $n=250$ | | | | $T=250$, $n=500$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.0809 | 0.0499 | 0.0895 | 0.0728 | 0.1199 | 0.0553 | 0.0795 | 0.0626 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.0909 | 0.0906 | 0.0479 | 0.0467 | 0.1037 | 0.1042 | 0.0446 | 0.0484 |
| | $\tau=50\%$ | 0.2010 | 0.1886 | 0.0413 | 0.0408 | 0.2460 | 0.2250 | 0.0482 | 0.0462 |
| | $\tau=75\%$ | 0.3824 | 0.3516 | 0.0386 | 0.0344 | 0.5021 | 0.4436 | 0.0460 | 0.0418 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.2129 | 0.2359 | 0.0364 | 0.0353 | 0.2794 | 0.3333 | 0.0417 | 0.0384 |
| | $\tau=50\%$ | 0.7161 | 0.7174 | 0.0437 | 0.0220 | 0.8862 | 0.8912 | 0.0808 | 0.0346 |
| | $\tau=75\%$ | 0.9877 | 0.9830 | 0.1848 | 0.0165 | 0.9993 | 0.9984 | 0.4400 | 0.0296 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1351 | 0.1291 | 0.0407 | 0.0391 | 0.1533 | 0.1487 | 0.0455 | 0.0440 |
| | $\tau=50\%$ | 0.2676 | 0.2462 | 0.0365 | 0.0337 | 0.3497 | 0.3095 | 0.0460 | 0.0424 |
| | $\tau=75\%$ | 0.4032 | 0.3590 | 0.0368 | 0.0340 | 0.5218 | 0.4409 | 0.0456 | 0.0388 |
| | | $T=500$, $n=250$ | | | | $T=500$, $n=500$ | | | |
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.0648 | 0.0473 | 0.0891 | 0.0774 | 0.0898 | 0.0509 | 0.0755 | 0.0652 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.1014 | 0.1006 | 0.0437 | 0.0450 | 0.1111 | 0.1113 | 0.0437 | 0.0441 |
| | $\tau=50\%$ | 0.2161 | 0.2079 | 0.0391 | 0.0389 | 0.2751 | 0.2617 | 0.0446 | 0.0397 |
| | $\tau=75\%$ | 0.4115 | 0.3904 | 0.0394 | 0.0358 | 0.5572 | 0.5146 | 0.0452 | 0.0393 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.2372 | 0.2526 | 0.0327 | 0.0352 | 0.3481 | 0.3846 | 0.0460 | 0.0464 |
| | $\tau=50\%$ | 0.7524 | 0.7556 | 0.0438 | 0.0319 | 0.9365 | 0.9411 | 0.0826 | 0.0550 |
| | $\tau=75\%$ | 0.9942 | 0.9932 | 0.1774 | 0.0497 | 0.9998 | 0.9998 | 0.4554 | 0.1408 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1379 | 0.1355 | 0.0375 | 0.0383 | 0.1814 | 0.1740 | 0.0393 | 0.0393 |
| | $\tau=50\%$ | 0.2813 | 0.2648 | 0.0347 | 0.0311 | 0.3905 | 0.3637 | 0.0420 | 0.0373 |
| | $\tau=75\%$ | 0.4231 | 0.3963 | 0.0367 | 0.0314 | 0.5765 | 0.5310 | 0.0447 | 0.0399 |
| | | $T=2500$, $n=250$ | | | | $T=2500$, $n=500$ | | | |
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.0540 | 0.0503 | 0.0883 | 0.0851 | 0.0581 | 0.0498 | 0.0773 | 0.0724 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.0971 | 0.0971 | 0.0420 | 0.0402 | 0.1204 | 0.1189 | 0.0446 | 0.0447 |
| | $\tau=50\%$ | 0.2169 | 0.2144 | 0.0422 | 0.0389 | 0.3122 | 0.3079 | 0.0417 | 0.0406 |
| | $\tau=75\%$ | 0.4287 | 0.4239 | 0.0373 | 0.0325 | 0.6130 | 0.6018 | 0.0432 | 0.0402 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.2515 | 0.2544 | 0.0344 | 0.0346 | 0.4008 | 0.4125 | 0.0411 | 0.0443 |
| | $\tau=50\%$ | 0.7961 | 0.7968 | 0.0444 | 0.0414 | 0.9656 | 0.9669 | 0.0796 | 0.0775 |
| | $\tau=75\%$ | 0.9974 | 0.9973 | 0.1666 | 0.1336 | 1.0000 | 1.0000 | 0.4319 | 0.3710 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1421 | 0.1409 | 0.0416 | 0.0402 | 0.1893 | 0.1875 | 0.0417 | 0.0427 |
| | $\tau=50\%$ | 0.2874 | 0.2824 | 0.0354 | 0.0324 | 0.4357 | 0.4251 | 0.0439 | 0.0430 |
| | $\tau=75\%$ | 0.4331 | 0.4281 | 0.0329 | 0.0301 | 0.6487 | 0.6367 | 0.0423 | 0.0401 |

Note: This table displays the Monte Carlo results associated to the marginal MES setting. $UC_{MES}$ and $IND_{MES}$ denote the unconditional coverage test and independence test, respectively. The test statistics robust to estimation risk are superscripted by $C$. For the independence test, we consider a maximum lag order $m=5$. Reported powers are sized-corrected.

Table 2: Empirical rejection rates for backtesting MES(5%) at 5% nominal level (conditional case)

| | | $T=250$, $n=250$ | | | | $T=250$, $n=500$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.1080 | 0.0460 | 0.1000 | 0.0590 | 0.1500 | 0.0400 | 0.1100 | 0.0530 |
| $H_1(B_1)$ | $\tau=25\%$ | 0.0780 | 0.0670 | 0.0400 | 0.0470 | 0.0820 | 0.0800 | 0.0400 | 0.0550 |
| | $\tau=50\%$ | 0.1550 | 0.1540 | 0.0540 | 0.0480 | 0.1970 | 0.1980 | 0.0610 | 0.0630 |
| | $\tau=75\%$ | 0.3090 | 0.2750 | 0.0430 | 0.0440 | 0.4470 | 0.3630 | 0.0480 | 0.0600 |
| $H_1(B_2)$ | $\tau=25\%$ | 0.1520 | 0.1850 | 0.0320 | 0.0250 | 0.2210 | 0.2730 | 0.0270 | 0.0310 |
| | $\tau=50\%$ | 0.6490 | 0.6290 | 0.0310 | 0.0110 | 0.8500 | 0.7920 | 0.0530 | 0.0120 |
| | $\tau=75\%$ | 0.9750 | 0.9100 | 0.1710 | 0.0020 | 0.9970 | 0.9540 | 0.3610 | 0.0040 |
| $H_1(B_3)$ | $\tau=25\%$ | 0.0970 | 0.1050 | 0.0410 | 0.0420 | 0.1180 | 0.1320 | 0.0390 | 0.0470 |
| | $\tau=50\%$ | 0.2240 | 0.2060 | 0.0450 | 0.0490 | 0.3050 | 0.2820 | 0.0600 | 0.0620 |
| | $\tau=75\%$ | 0.3190 | 0.2890 | 0.0450 | 0.0420 | 0.4600 | 0.3800 | 0.0490 | 0.0430 |
| $H_1(B_4)$ | | 0.2140 | 0.3170 | 0.4470 | 0.4690 | 0.2490 | 0.3710 | 0.5920 | 0.6460 |
| | | $T=500$, $n=250$ | | | | $T=500$, $n=500$ | | | |
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.0760 | 0.0430 | 0.0940 | 0.0600 | 0.0920 | 0.0360 | 0.0900 | 0.0610 |
| $H_1(B_1)$ | $\tau=25\%$ | 0.0810 | 0.0760 | 0.0470 | 0.0550 | 0.1050 | 0.1140 | 0.0470 | 0.0490 |
| | $\tau=50\%$ | 0.1800 | 0.1680 | 0.0420 | 0.0450 | 0.2760 | 0.2790 | 0.0430 | 0.0420 |
| | $\tau=75\%$ | 0.3810 | 0.3600 | 0.0530 | 0.0610 | 0.5360 | 0.5050 | 0.0570 | 0.0580 |
| $H_1(B_2)$ | $\tau=25\%$ | 0.2250 | 0.2460 | 0.0260 | 0.0330 | 0.3260 | 0.3940 | 0.0430 | 0.0380 |
| | $\tau=50\%$ | 0.7340 | 0.7160 | 0.0420 | 0.0290 | 0.9310 | 0.9290 | 0.0680 | 0.0290 |
| | $\tau=75\%$ | 0.9900 | 0.9770 | 0.1750 | 0.0140 | 1.0000 | 0.9960 | 0.3640 | 0.0140 |
| $H_1(B_3)$ | $\tau=25\%$ | 0.1090 | 0.1100 | 0.0450 | 0.0540 | 0.1630 | 0.1960 | 0.0420 | 0.0430 |
| | $\tau=50\%$ | 0.2790 | 0.2670 | 0.0450 | 0.0540 | 0.3960 | 0.3880 | 0.0470 | 0.0400 |
| | $\tau=75\%$ | 0.3830 | 0.3460 | 0.0320 | 0.0380 | 0.5790 | 0.5400 | 0.0500 | 0.0520 |
| $H_1(B_4)$ | | 0.2200 | 0.2880 | 0.4640 | 0.5030 | 0.2860 | 0.3850 | 0.5970 | 0.6390 |
| | | $T=2500$, $n=250$ | | | | $T=2500$, $n=500$ | | | |
| | | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ | $UC_{MES}$ | $UC^C_{MES}$ | $IND_{MES}$ | $IND^C_{MES}$ |
| $H_0$ | | 0.0500 | 0.0410 | 0.0890 | 0.0810 | 0.0650 | 0.0380 | 0.0680 | 0.0590 |
| $H_1(B_1)$ | $\tau=25\%$ | 0.0980 | 0.1010 | 0.0540 | 0.0540 | 0.1000 | 0.0980 | 0.0480 | 0.0370 |
| | $\tau=50\%$ | 0.2080 | 0.2080 | 0.0450 | 0.0440 | 0.3050 | 0.2990 | 0.0560 | 0.0510 |
| | $\tau=75\%$ | 0.4120 | 0.4100 | 0.0330 | 0.0330 | 0.6190 | 0.6070 | 0.0590 | 0.0540 |
| $H_1(B_2)$ | $\tau=25\%$ | 0.2700 | 0.2810 | 0.0300 | 0.0280 | 0.4020 | 0.4220 | 0.0570 | 0.0580 |
| | $\tau=50\%$ | 0.8200 | 0.8220 | 0.0420 | 0.0290 | 0.9680 | 0.9690 | 0.0870 | 0.0600 |
| | $\tau=75\%$ | 0.9990 | 0.9980 | 0.1620 | 0.0640 | 1.0000 | 0.9990 | 0.4380 | 0.2270 |
| $H_1(B_3)$ | $\tau=25\%$ | 0.1380 | 0.1350 | 0.0280 | 0.0330 | 0.1940 | 0.2030 | 0.0560 | 0.0560 |
| | $\tau=50\%$ | 0.2940 | 0.2910 | 0.0360 | 0.0340 | 0.4470 | 0.4460 | 0.0500 | 0.0430 |
| | $\tau=75\%$ | 0.4630 | 0.4570 | 0.0380 | 0.0340 | 0.6160 | 0.6020 | 0.0520 | 0.0400 |
| $H_1(B_4)$ | | 0.2770 | 0.2960 | 0.4040 | 0.4200 | 0.2860 | 0.3210 | 0.6580 | 0.6660 |

Note: This table displays the Monte Carlo results associated to the conditional MES setting. $UC_{MES}$ and $IND_{MES}$ denote the unconditional coverage test and independence test, respectively. The test statistics robust to estimation risk are superscripted by $C$. For the independence test, we consider a maximum lag order $m=5$. Reported powers are sized-corrected.

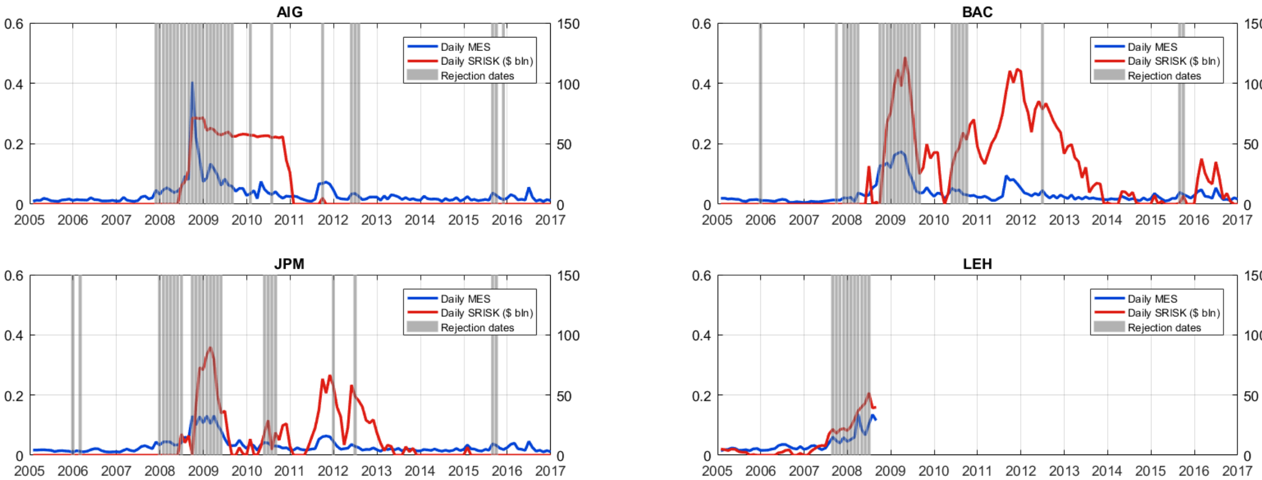
Table 3: Empirical rejection rates for backtesting ΔCoVaR at 5% nominal level (marginal case)

| | | $T=250$, $n=250$ | | $T=250$, $n=500$ | |
|---|---|---|---|---|---|
| | | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ |
| $H_0$ | | 0.0766 | 0.0458 | 0.1051 | 0.0572 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.3100 | 0.3637 | 0.4881 | 0.4966 |
| | $\tau=50\%$ | 0.9002 | 0.9300 | 0.9941 | 0.9928 |
| | $\tau=75\%$ | 1.0000 | 0.9999 | 1.0000 | 1.0000 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.0585 | 0.0599 | 0.0555 | 0.0686 |
| | $\tau=50\%$ | 0.0681 | 0.0748 | 0.1001 | 0.1292 |
| | $\tau=75\%$ | 0.0938 | 0.1492 | 0.2513 | 0.3323 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1027 | 0.1141 | 0.1449 | 0.1799 |
| | $\tau=50\%$ | 0.2669 | 0.3026 | 0.4957 | 0.5299 |
| | $\tau=75\%$ | 0.5644 | 0.5917 | 0.8471 | 0.8253 |
| | | $T=500$, $n=250$ | | $T=500$, $n=500$ | |
| | | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ |
| $H_0$ | | 0.0601 | 0.0396 | 0.0806 | 0.0525 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.3534 | 0.3922 | 0.5325 | 0.5341 |
| | $\tau=50\%$ | 0.9336 | 0.9467 | 0.9974 | 0.9969 |
| | $\tau=75\%$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.0493 | 0.0501 | 0.0541 | 0.0634 |
| | $\tau=50\%$ | 0.0735 | 0.0798 | 0.1180 | 0.1278 |
| | $\tau=75\%$ | 0.1313 | 0.1635 | 0.3234 | 0.3444 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1109 | 0.1210 | 0.1732 | 0.1886 |
| | $\tau=50\%$ | 0.3085 | 0.3453 | 0.5299 | 0.5525 |
| | $\tau=75\%$ | 0.5945 | 0.6273 | 0.8629 | 0.8613 |
| | | $T=2500$, $n=250$ | | $T=2500$, $n=500$ | |
| | | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ | $UC_{\Delta CoVaR}$ | $UC^C_{\Delta CoVaR}$ |
| $H_0$ | | 0.0547 | 0.0468 | 0.0642 | 0.0522 |
| $H_1(A_1)$ | $\tau=25\%$ | 0.3769 | 0.3870 | 0.5723 | 0.5760 |
| | $\tau=50\%$ | 0.9414 | 0.9421 | 0.9974 | 0.9973 |
| | $\tau=75\%$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $H_1(A_2)$ | $\tau=25\%$ | 0.0493 | 0.0529 | 0.0612 | 0.0677 |
| | $\tau=50\%$ | 0.0715 | 0.0799 | 0.1234 | 0.1319 |
| | $\tau=75\%$ | 0.1336 | 0.1512 | 0.3164 | 0.3457 |
| $H_1(A_3)$ | $\tau=25\%$ | 0.1100 | 0.1196 | 0.1864 | 0.1977 |
| | $\tau=50\%$ | 0.3074 | 0.3349 | 0.5794 | 0.5933 |
| | $\tau=75\%$ | 0.6112 | 0.6361 | 0.8943 | 0.9001 |

Note: This table displays the Monte Carlo results for the ΔCoVaR associated to a time-invariant setting as defined in Subsection "Backtesting marginal MES". $UC_{\Delta CoVaR}$ denotes the unconditional coverage test for ΔCoVaR. The probability levels are set to $\alpha = 0.05$, $\beta_{\text{inf}} = 0.25$, and $\beta_{\text{sup}} = 0.75$. The test statistic robust to estimation risk is superscripted by $C$. Reported powers are sized-corrected.

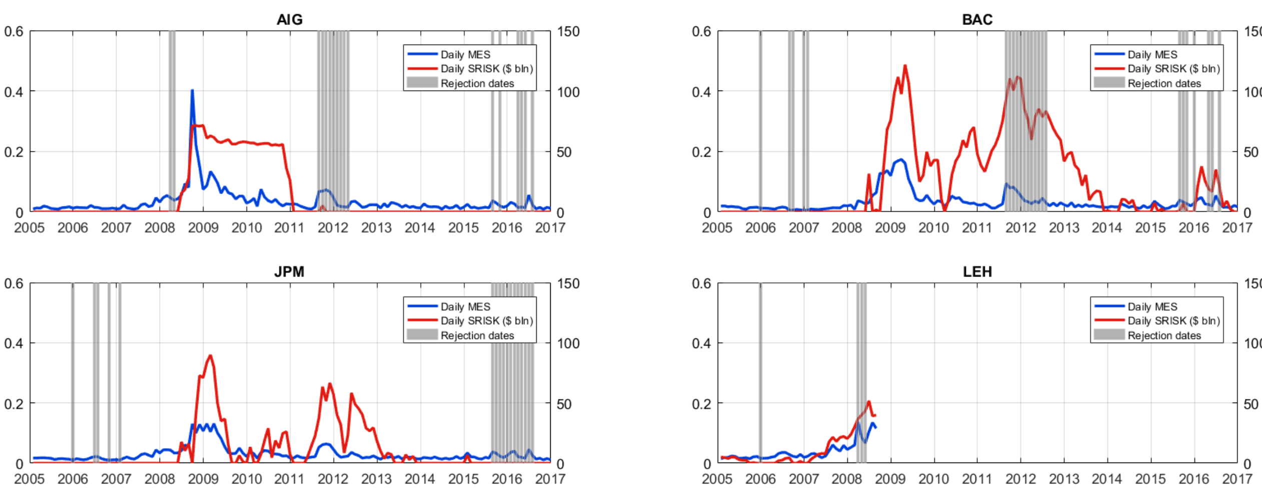
# FIGURES

Figure 1: UC backtests for one-day risk forecast horizon (recursive estimation, $n = 250$)



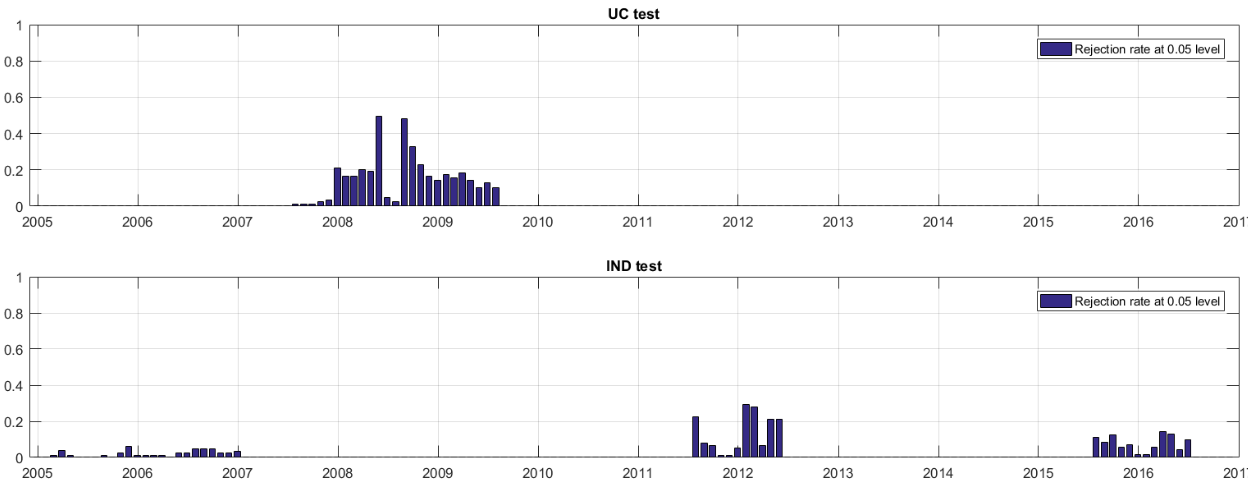

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation. Reported results are robust to estimation risk.

Figure 2: IND backtests for one-day risk forecast horizon (recursive estimation, $n = 250$, $m = 5$)



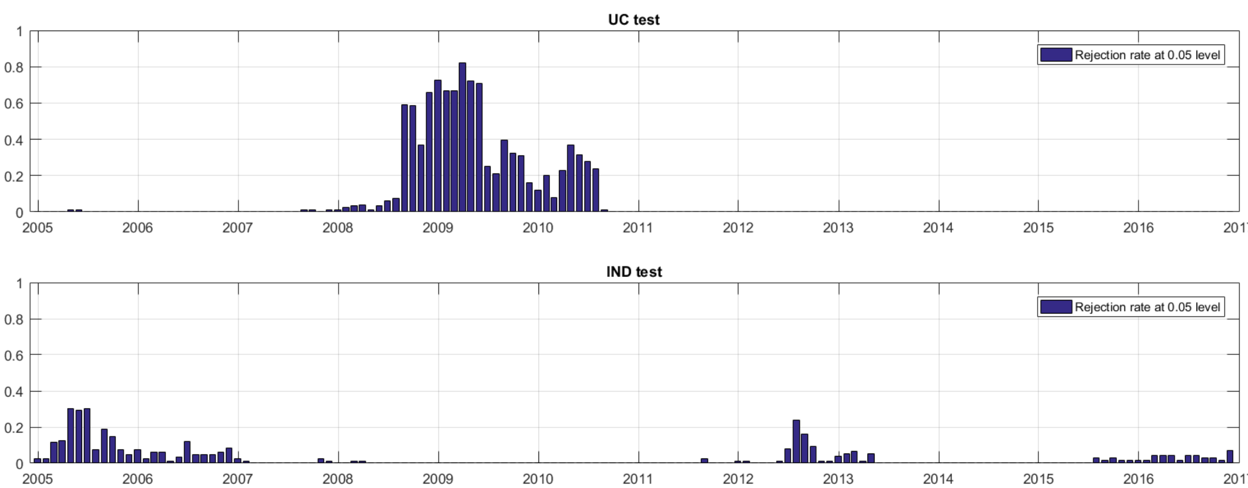

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the IND backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 3: Rejection rates of UC and IND backtests (recursive estimation, $n = 250$, $m = 5$)



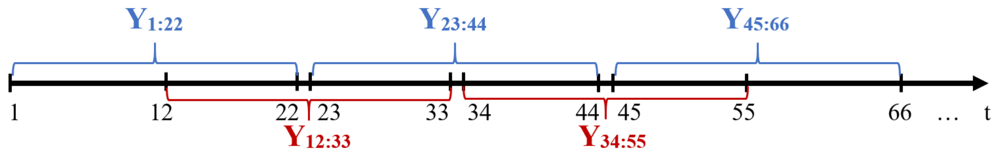

Note: This figure displays the rejection rates of the UC backtest (top panel) and IND backtest (bottom panel) at 5% significance level for the full list of tickers. We use $n = 250$ out-of-sample observations, a recursive scheme for parameter estimation, and a Bonferroni correction for multiple testing. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 4: Rejection rates of UC and IND backtests (recursive estimation, $n = 500$, $m = 5$)



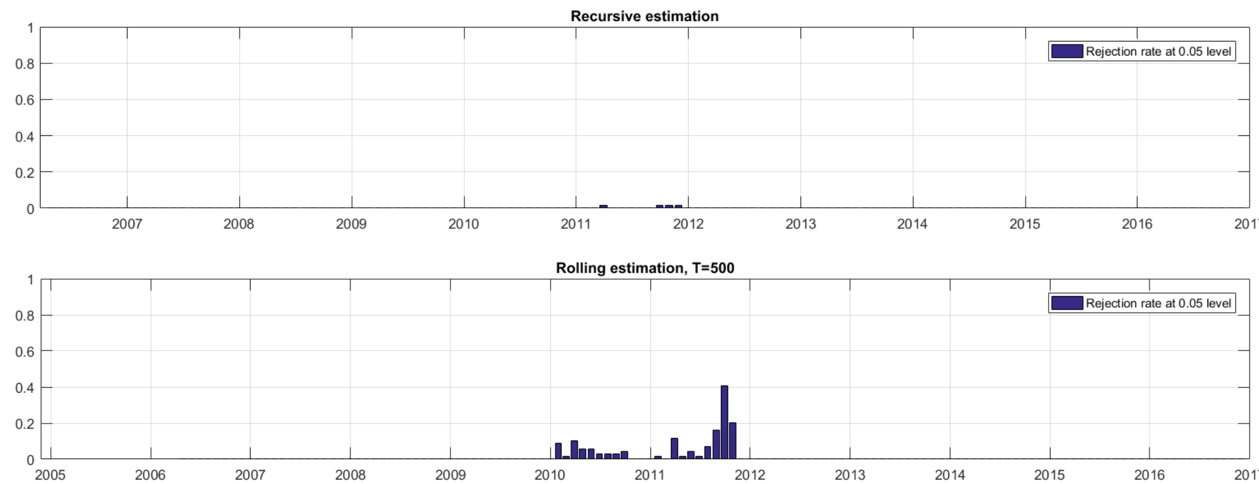

Note: This figure displays the rejection rates of the UC backtest (top panel) and IND backtest (bottom panel) at 5% significance level for the full list of tickers. We use $n = 500$ out-of-sample observations, a recursive scheme for parameter estimation, and a Bonferroni correction for multiple testing. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 5: Overlapping procedure



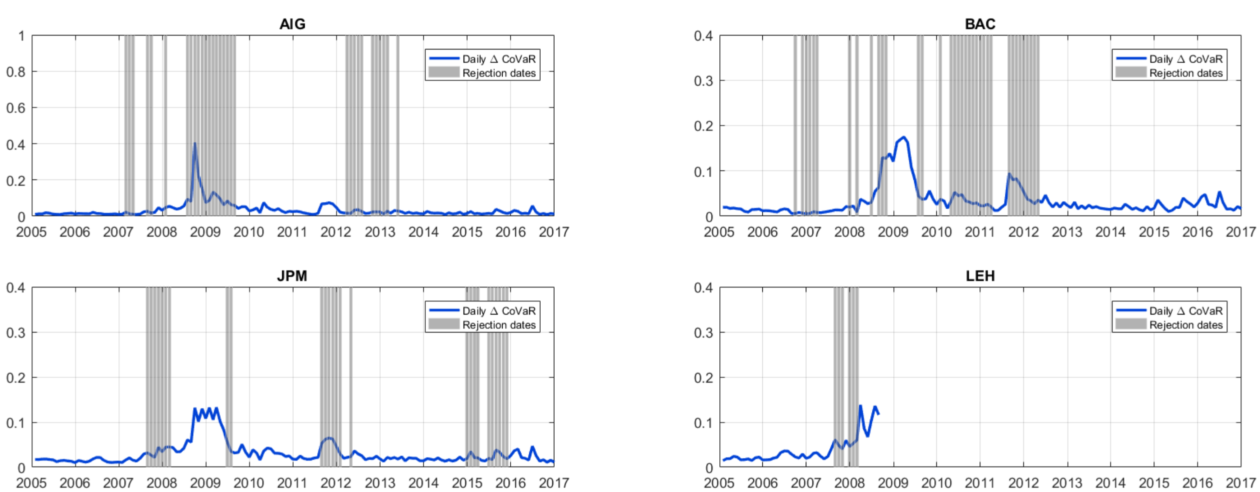

Note: This figure provides an illustration of the overlapping procedure. The variables $\{Y_{1:22}, Y_{12:33}, \ldots\}$ represent the successive out-of-sample returns considered in the test to backtesting $h$-day-ahead systemic risk forecasts with $h = 22$.

Figure 6: Rejection rates of UC backtests ($n = 100$, $h = 22$, overlap of 11 days)



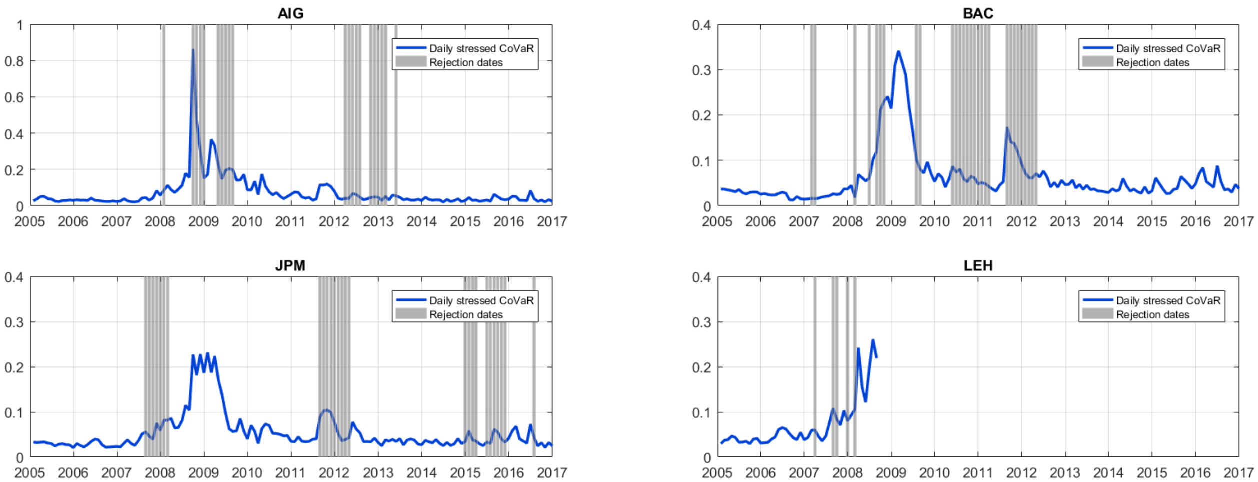

Note: This figure displays the rejection rates of the UC backtest at 5% significance level for the full list of tickers. We use a recursive scheme (top panel) and a rolling scheme with $T = 500$ observations (bottom panel) for parameter estimation. We consider a forecasting horizon $h = 22$, an overlap of 11 trading days, $n = 100$ out-of-sample observations, and a Bonferroni correction for multiple testing.

Figure 7: UC backtests for $\Delta$CoVaR (recursive estimation, $n = 250$)



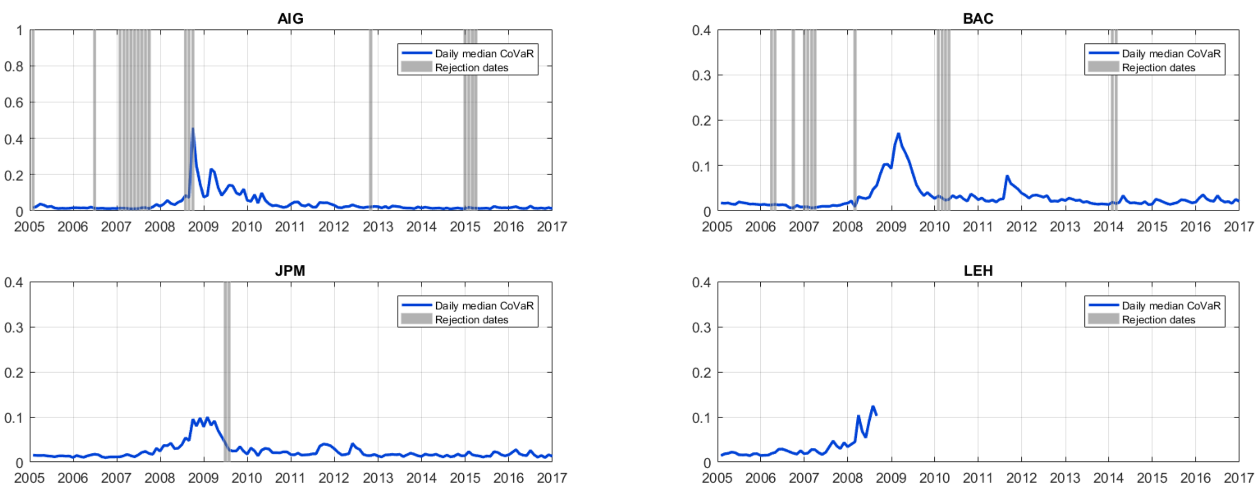

Note: This figure displays the daily $\Delta$CoVaR (blue line) and the rejection dates (shaded area) of the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation. Reported results are robust to estimation risk.

Figure 8: UC backtests for stressed CoVaR (recursive estimation, $n = 250$)



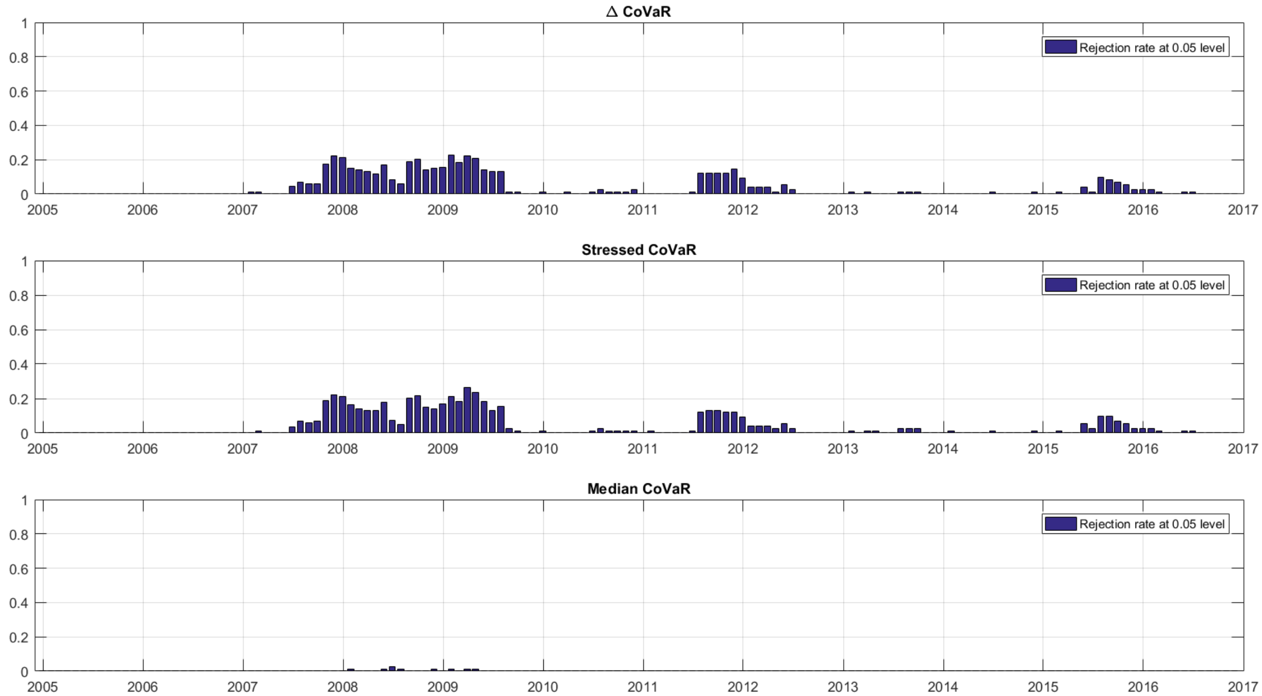

Note: This figure displays the daily stressed CoVaR (blue line) and the rejection dates (shaded area) of the UC sub-test for stressed CoVaR at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation. Reported results are robust to estimation risk.

Figure 9: UC backtests for median CoVaR (recursive estimation, $n = 250$)



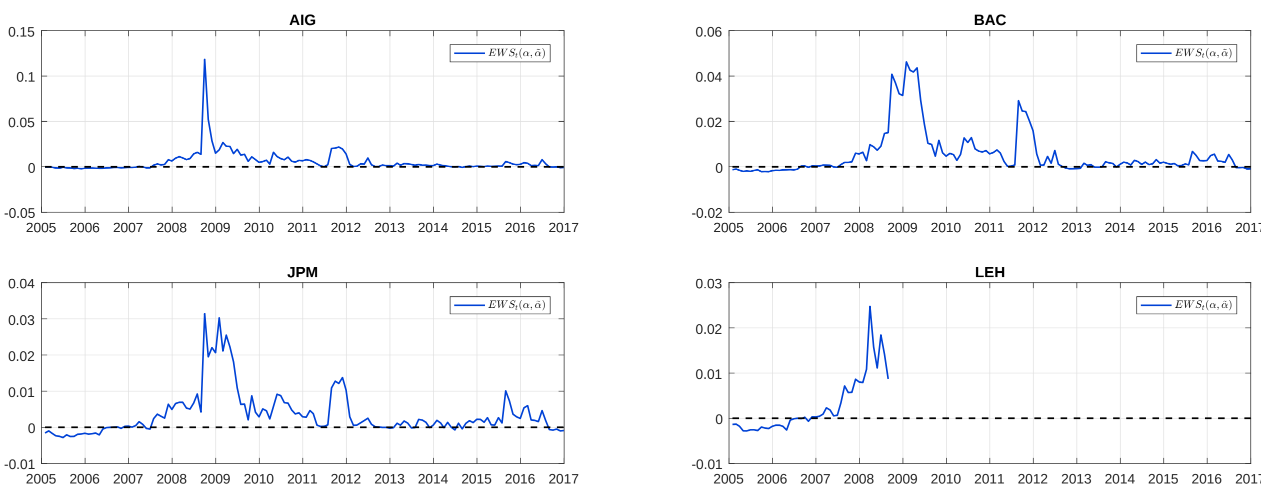

Note: This figure displays the daily median CoVaR (blue line) and the rejection dates (shaded area) of the UC sub-test for median CoVaR at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation. Reported results are robust to estimation risk.

Figure 10: Rejection rates for $\Delta$CoVaR, stressed, and median CoVaRs (recursive estimation, $n = 250$)



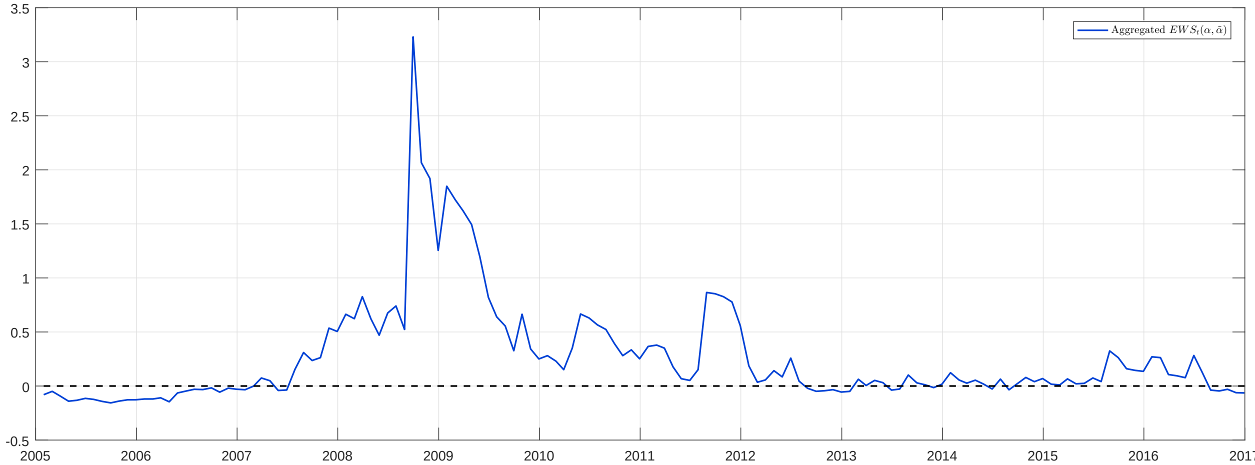

Note: This figure displays the rejection rates of the UC backtest for $\Delta$CoVaR (top panel), the UC sub-test for stressed CoVaR (middle panel), and the UC sub-test for median CoVaR (bottom panel) at 5% significance level for the full list of tickers. We use $n = 250$ out-of-sample observations, a recursive scheme for parameter estimation, and a Bonferroni correction for multiple testing. Reported results are robust to estimation risk.

Figure 11: $EWS_t\left(\alpha, \widetilde{\alpha}\right)$ for a panel of four firms (recursive estimation, $n = 250$)




Note: This figure displays the EWS (blue line) issued by the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation.

Figure 12: Aggregated $EWS_t\left(\alpha, \widetilde{\alpha}\right)$ (recursive estimation, $n = 250$)




Note: This figure displays the aggregated EWS (blue line) issued by the UC backtest at 5% significance level for the full list of tickers. We use $n = 250$ out-of-sample observations and a recursive scheme for parameter estimation.

# 8 Appendix

## Appendix A: Assumptions

To derive the asymptotic properties of the test statistics and the robust test statistics, 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 $\widehat{\theta}_T$ is consistent for $\theta_0$ and is asymptotically normally distributed such that:

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

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

## Appendix B: Cumulative joint violation process

**Proof.** First, let us 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:

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

Let us introduce two terms that we can interpret 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

$$H_t(\alpha, \theta_0) = \mathbb{1}\left(u_{2t} \leq \alpha\right) \int_0^1 \mathbb{1}\left(u_{12t} \leq \beta\right) \mathrm{d}\beta = \mathbb{1}\left(u_{2t} \leq \alpha\right) \int_{u_{12t}}^1 1 \mathrm{d}\beta.$$

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

$$H_t(\alpha, \theta_0) = \left(1 - u_{12t}(\theta_0)\right) \times \mathbb{1}(u_{2t}(\theta_0) \leq \alpha).$$

The PIT implies that the variable $u_{2t}$ has a uniform $U_{[0,1]}$ distribution. The binary variable $\mathbb{1}(u_{2t}(\theta_0) \leq \alpha)$ has a Bernouilli distribution with a success probability equal to $\alpha$. The variable

$u_{12t}$ has also a $U_{[0,1]}$ distribution as soon as the PIT transformation $F_{Y_1|Y_2\leq VaR_{2t}(\alpha,\theta_0)}(.;\Omega_{t-1},\theta_0)$ is applied to observations $Y_{1t}$ for which $Y_{2t}\leq VaR_{2t}(\alpha,\theta_0)$.[25]

$$\mathbb{1}(u_{2t}(\theta_0)\leq\alpha)|\,\Omega_{t-1}\sim\text{Bernouilli}(\alpha),$$

$$(1-u_{12t}(\theta_0))|\,\{\Omega_{t-1},u_{2t}(\theta_0)\leq\alpha\}\sim U_{[0,1]}.$$

The two first conditional moments of the cumulative joint process $H_t(\alpha,\theta_0)$ are then given by

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

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

Since the conditional distribution of $u_{12t}(\theta_0)$ given $\Omega_{t-1}$ is $U_{[0,1]}$ with $\mathbb{E}\left(u_{12t}(\theta_0)|\,u_{2t}(\theta_0)\leq\alpha,\Omega_{t-1}\right)=1/2$ and $\mathbb{E}\left(u_{12t}^2(\theta_0)\big|\,u_{2t}(\theta_0)\leq\alpha,\Omega_{t-1}\right)=1/3$, then we get

$$\mathbb{E}\left(H_t(\alpha,\theta_0)|\,\Omega_{t-1}\right)=\frac{\alpha}{2},$$

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

## Appendix C: Proof of Theorem 1

**Proof.** Denote $H_t(\alpha,\theta)=(1-u_{12t}(\theta))\,1\,(u_{2t}(\theta)\leq\alpha)$ the cumulative violation process, with $u_{2t}(\theta)=F_{Y_2}(Y_{2t};\Omega_{t-1},\theta)$ and $u_{12t}(\theta)=F_{Y_1|Y_2\leq VaR_{2t}(\alpha,\theta)}(Y_{1t};\Omega_{t-1},\theta)$, $\forall t=T+1,...,T+n$ and $\forall\theta\in\Theta$. Under the null hypothesis $H_{0,UC}$, the sequence $\{H_t(\alpha,\theta_0)-\alpha/2\}_{t=T+1}^{T+n}$ 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. We can rewrite the test statistic $UC_{MES}$ as

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

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

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

[25] We thank Andrew Patton for this suggestion.

Rearranging these terms gives

$$\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\middle|\,\Omega_{t-1}\right)\right) = \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t\left(\alpha,\theta_0\right)-\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)\middle|\,\Omega_{t-1}\right)\right) + \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left(H_t(\alpha,\widehat{\theta}_T)-H_t\left(\alpha,\theta_0\right)\right)\middle|\,\Omega_{t-1}\right)+o_p\left(1\right). \tag{8}$$

The mean value theorem implies that

$$H_t(\alpha,\widehat{\theta}_T)=H_t\left(\alpha,\theta_0\right)+\left(\widehat{\theta}_T-\theta_0\right)'\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta},$$

where $\widetilde{\theta}$ is an intermediate point between $\theta_0$ and $\widehat{\theta}_T$. Equation (8) becomes

$$\frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\alpha/2\right) = \frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t\left(\alpha,\theta_0\right)-\alpha/2\right) + \frac{\sqrt{\lambda}}{\sigma_H}\sqrt{T}(\widehat{\theta}_T-\theta_0)'\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta}\middle|\,\Omega_{t-1}\right)+o_p\left(1\right).$$

Assume that $T\to\infty$, $n\to\infty$ and $n/T\to\lambda$ 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 $\sqrt{T}(\widehat{\theta}_T-\theta_0)$ is 0 as $\widehat{\theta}_T$ depends on the in-sample observations and the summand in the first term is for out-of-sample observations. Under Assumption A4, $\widetilde{\theta}\xrightarrow{p}\theta_0$ and since $\partial H_t\left(\alpha,\theta_0\right)/\partial\theta$ is also a mds, we have

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

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

$$\frac{\sqrt{\lambda}}{\sigma_H}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta}\middle|\,\Omega_{t-1}\right)'\sqrt{T}(\widehat{\theta}_T-\theta_0)\xrightarrow{d}\mathcal{N}\left(0,\frac{\lambda}{\sigma_H^2}R'_{MES}\Sigma_0R_{MES}\right),$$

and finally

$$UC_{MES}\xrightarrow{d}\mathcal{N}\left(0,1+\lambda\frac{R'_{MES}\Sigma_0R_{MES}}{\alpha\left(1/3-\alpha/4\right)}\right).$$

## Appendix D: Bivariate normal case

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

$$Y_t=\Sigma_t^{1/2}\varepsilon_t$$

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

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

$$F(y, \Sigma_t) = \Pr\left((Y_1 \leq y_1) \cap (Y_2 \leq y_2)\right) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b.$$

Using the Dwyer formula (1967), we know that

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

and we get

$$\begin{aligned} \frac{\partial F(y, \Sigma_t)}{\partial \Sigma_t} &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \frac{\partial f(a, b, \Sigma_t)}{\partial \Sigma_t} \mathrm{d}a\mathrm{d}b \\ &= -\frac{\Sigma_t^{-1}}{2} F(y, \Sigma_t) + \frac{1}{2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) \, \Sigma_t^{-1} \Delta(a, b) \, \Sigma_t^{-1} \mathrm{d}a\mathrm{d}b, \end{aligned} \tag{9}$$

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

$$\Sigma_t = \begin{pmatrix} \sigma_{1t}^2 & \sigma_{12t} \\ \sigma_{12t} & \sigma_{2t}^2 \end{pmatrix},$$

we can decompose the matrix expression given in Equation (9) as follows

$$\frac{\partial F(y, \Sigma_t)}{\partial \sigma_{1t}^2} = -\frac{\sigma_{2t}^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \left(\sigma_{2t}^2 a^2 - 2\sigma_{12t}\sigma_{2t}^2 ab + \sigma_{12t}^2 b^2\right) f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b,$$

$$\frac{\partial F(y, \Sigma_t)}{\partial \sigma_{2t}^2} = -\frac{\sigma_1^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \left(\sigma_{12t}^2 a^2 - 2\sigma_{12t}\sigma_1^2 ab + \sigma_{1t}^2 b^2\right) f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b,$$

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

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 $\theta \in \Theta$, we have

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

where vec(.) denotes the vectorization operator of a matrix.

## Appendix E: Consistent estimates of $\mathbf{R_{MES}}$, $\mathbf{R_j}$, and $\boldsymbol{\gamma_\lambda}$

We provide consistent estimates of $R_{MES}$, $R_j$, and $\gamma_\lambda$ involved in the computation of the robust test statistics (see Theorem 1 and 3, and Appendix H).

**Estimation of $\mathbf{R_{MES}}$.** As a starting point, we consider the approach that consists in evaluating the convergence of

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}\right|\Omega_{t-1}\right).$$

Deriving the cumulative joint violation process with respect to $\theta$ yields

$$\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}=-\frac{\partial u_{12t}(\widetilde{\theta})}{\partial\theta}\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)+\left(1-u_{12t}\left(\widetilde{\theta}\right)\right)\frac{\partial\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)}{\partial\theta}.$$

Since $F_{Y_2}\left(VaR_{2t}\left(\alpha,\theta\right);\Omega_{t-1},\theta\right)=\alpha$ for any $\theta\in\Theta$, we have

$$-\frac{\partial u_{12t}\left(\theta\right)}{\partial\theta}=-\frac{1}{\alpha}\frac{\partial F(\widetilde{y}_t\left(\theta\right);\Omega_{t-1},\theta)}{\partial\theta},$$

with the vector $\widetilde{y}_t\left(\theta\right)$ defined as $\widetilde{y}_t\left(\theta\right)=\left(y_{1t},VaR_{2t}\left(\alpha,\theta\right)\right)'$. Taking sums and conditional expectations yield

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}\right|\Omega_{t-1}\right)=-\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial F(\widetilde{y}_t(\widetilde{\theta});\Omega_{t-1},\widetilde{\theta})}{\partial\theta}\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)\right|\Omega_{t-1}\right)$$
$$+\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\left(1-u_{12t}(\widetilde{\theta})\right)\frac{\partial\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)}{\partial\theta}\right|\Omega_{t-1}\right).$$

Now assume that $T\rightarrow\infty$, under assumption A4, this quantity converges to

$$\begin{aligned}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\right|\Omega_{t-1}\right)=&-\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)}{\partial\theta}\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)\right|\Omega_{t-1}\right)\\&+\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\left(1-u_{12t}(\theta_0)\right)\frac{\partial\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)}{\partial\theta}\right|\Omega_{t-1}\right).\end{aligned}\quad(10)$$

Assume that $n\rightarrow\infty$, this quantity converges to

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\right|\Omega_{t-1}\right)\xrightarrow{p}R_{MES}=\mathbb{E}_0\left(\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\right).\quad(11)$$

We deduce from Equations (10) and (11) that

$$R_{MES}=-\frac{1}{\alpha}\mathbb{E}_0\left(\frac{\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)}{\partial\theta}\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)\right)+\mathbb{E}_0\left(\left(1-u_{12t}(\theta_0)\right)\frac{\partial\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)}{\partial\theta}\right),\quad(12)$$

Note that explicit formulas for $\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)/\partial\theta$ in the left hand side of Equation (12) are available for some particular bivariate distributions (see Appendix D for the normal distribution). In any case, we can evaluate the derivative numerically. The right hand side of Equation (12) is trickier to implement due to the presence of the indicator function $\mathbb{1}(u_{2t}(\theta_0) \leq \alpha)$ that cannot be continuously differentiated (see Engle and Manganelli, 2004; Lambert et al., 2012, for similar issues). We first approximate the indicator function with a continuously differentiable function. Denote by $\mathbb{1}^{\oplus}(u \leq \tau)$ a continuous approximation function of $\mathbb{1}(u \leq \tau)$, where $u$ is a random variable, and $\tau$ is a probability level,

$$\mathbb{1}^{\oplus}(u \leq \tau) = \int_0^{\tau} \frac{1}{h}\phi\left(\frac{u-v}{h}\right)\mathrm{d}v = \Phi\left(\frac{u}{h}\right) - \Phi\left(\frac{u-\tau}{h}\right).$$

with $\phi(.)$ a Gaussian kernel function, and $h > 0$ a bandwidth parameter. Under a suitable form of the law of large number, we can replace the expectation operator by the sample mean, and a consistent estimator of $R_{MES}$ of Equation (12) is given by

$$\widehat{R}_{MES} = -\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathbb{1}(u_{2t}(\widehat{\theta}_T) \leq \alpha) + \frac{1}{n}\sum_{t=T+1}^{T+n}(1-u_{12t}(\widehat{\theta}_T))\frac{\partial \mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T) \leq \alpha)}{\partial\theta},$$

with

$$\begin{aligned}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta} &= \int_{-\infty}^{y_{1t}} f(u, VaR_{2t}(\alpha,\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)\mathrm{d}u \times \frac{\partial VaR_{2t}(\alpha,\widehat{\theta}_T)}{\partial\theta} \\ &\quad + \int_{-\infty}^{y_{1t}}\int_{-\infty}^{VaR_{2t}(\alpha,\widehat{\theta}_T)} \frac{\partial f(u,v;\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathrm{d}u\mathrm{d}v,\end{aligned}$$

$$\frac{\partial \mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T) \leq \alpha)}{\partial\theta} = \frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)}{h}\right) - \phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\alpha}{h}\right)\right) \times \frac{\partial F_{Y_{2t}}(y_{2t},\widehat{\theta}_T)}{\partial\theta}.$$

This approach requires to choose an appropriate value for the bandwidth parameter $h$ for the kernel smoothing which must ensure $\underset{n\to\infty}{lim}\, h \equiv h(n) = 0$. Throughout the paper, we select a bandwidth parameter $h = n^{-1}$ which verifies the above properties.

**Estimation of $\mathbf{R_j}$.** The sketch of the proof is similar to that of the quantity $R_{MES}$. We use a continuously differentiable function in place of the indicator function, and we compute a consistent estimator of $R_j$ by the law of large number. The quantity $R_j$ is given by

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

Given the above equation, we need a consistent estimator of $\partial H_t(\alpha,\theta_0)/\partial\theta$ to estimate $R_j$. As it has been done for $R_{MES}$, we substitute the indicator function by the kernel approximation, and estimate $\partial H_t(\alpha,\theta_0)/\partial\theta$ as follows

$$\frac{\partial H_t^{\oplus}(\alpha,\widehat{\theta}_T)}{\partial\theta}=-\frac{1}{\alpha}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha)+(1-u_{12t}(\widehat{\theta}_T))\frac{\partial\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)}{\partial\theta}.$$

Finally, by the law of large number, we get

$$\widehat{R}_j=\frac{1}{\alpha\,(1/3-\alpha/4)}\frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\left((H_{t-j}(\alpha,\theta_0)-\alpha/2)\,\frac{\partial H_t^{\oplus}(\alpha,\widehat{\theta}_T)}{\partial\theta}\right).$$

**Estimation of $\boldsymbol{\gamma_\lambda}$.** Let us rewrite $h_t(\alpha,\alpha,\widehat{\theta}_T)$ and $h_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)$ in terms of their generalized errors,

$$h_t(\alpha,\alpha,\widehat{\theta}_T)=\mathbb{1}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha),$$

$$h_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)=\mathbb{1}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times(\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\sup})-\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\inf})),$$

with

$$u_{12t}(\alpha,\widehat{\theta}_T)=F_{Y_1|Y_2\leq VaR_{2t}(\alpha,\widehat{\theta}_T)}(Y_{1t};\Omega_{t-1},\widehat{\theta}_T),$$

$$u_{12t}(\beta,\widehat{\theta}_T)=F_{Y_1|VaR_{2t}(\beta_{\inf},\widehat{\theta}_T)\leq Y_2\leq VaR_{2t}(\beta_{\sup},\widehat{\theta}_T)}(Y_{1t};\Omega_{t-1},\widehat{\theta}_T).$$

As a key point, we need estimates of $R_C^{\alpha}$ and $R_C^{\beta}$ (see Appendix H for a definition). As for $R_{MES}$ and $R_j$, the quantities $R_C^{\alpha}$ and $R_C^{\beta}$ depend on indicator functions, and we propose to approximate them with kernel smoother. A consistent estimator of $R_C^{\alpha}$ is given by

$$\widehat{R}_C^{\alpha}=\frac{1}{n}\sum_{t=T+1}^{T+n}\left(\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha)+\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\right),$$

with $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)=\frac{1}{h}\left(\phi\left(\frac{u_{12t}(\alpha,\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{12t}(\alpha,\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{12t}(\alpha,\widehat{\theta}_T)$, and $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)=\frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{2t}(\widehat{\theta}_T)$. Similarly, a consistent estimator of $R_C^{\beta}$ is given by

$$\begin{aligned}\widehat{R}_C^{\beta}=\frac{1}{n}\sum_{t=T+1}^{T+n}\Bigg[&\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times\Big(\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\sup})-\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\inf})\Big)\\&+\mathbb{1}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times\left(\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\sup})-\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\inf})\right)\Bigg],\end{aligned}$$

with $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)=\frac{1}{h}\left(\phi\left(\frac{u_{12t}(\beta,\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{12t}(\beta,\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{12t}(\beta,\widehat{\theta}_T)$, and $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\sup})-\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\inf})=\frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\beta_{\inf}}{h}\right)-\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\beta_{\sup}}{h}\right)\right)\frac{\partial}{\partial\theta}u_{2t}(\widehat{\theta}_T)$. Given the results in Appendix H, we get

$$\widehat{\gamma}_\lambda=\begin{pmatrix}\lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\alpha}} & \lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}}\\ \lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}} & \lambda\widehat{R_C^{\beta}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}}\end{pmatrix}.$$

## Appendix F: 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

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

For ease of notations, we drop the dependence of $\gamma_j$ and $\rho_j$ on $\alpha$ and $\theta_0$. The sample counterparts of $\gamma_j$ and $\rho_j$ based on the sample $\{H_t(\alpha, \theta_0)\}_{t=T+1}^{T+n}$ are

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

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

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

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

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \mathbb{E}\left(\widehat{\rho}_{nj} \middle| \Omega_{t-1}\right)\right) = \sqrt{n-j}\left(\rho_{nj} - \mathbb{E}\left(\rho_{nj} \middle| \Omega_{t-1}\right)\right) + o_p(1).$$

Rearranging these terms gives

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \rho_{nj}\right) = \sqrt{n-j}\,\mathbb{E}\left(\widehat{\rho}_{nj} - \rho_{nj} \middle| \Omega_{t-1}\right) + o_p(1). \tag{13}$$

The mean value theorem implies that

$$\widehat{\rho}_{nj} = \rho_{nj} + \left(\widehat{\theta}_T - \theta_0\right)' \frac{\partial \widetilde{\rho}_{nj}}{\partial \theta},$$

with $\widetilde{\rho}_{nj}$ the lag-$j$ autocorrelation of the process $H_t(\alpha, \widetilde{\theta})$, where $\widetilde{\theta}$ is an intermediate point between $\theta_0$ and $\widehat{\theta}_T$. Define $\lambda = (n-j)/T$, Equation (13) becomes

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \rho_j\right) = \sqrt{n-j}\left(\rho_{nj} - \rho_j\right) + \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T - \theta_0)'\mathbb{E}\left(\frac{\partial \widetilde{\rho}_{nj}}{\partial \theta} \middle| \Omega_{t-1}\right) + o_p(1).$$

Under Assumption A4, when $T \to \infty$ we have $\widetilde{\theta} \xrightarrow{p} \theta_0$. Then, we get for $j \neq 0$

$$\mathbb{E}\left(\frac{\partial \widetilde{\rho}_{nj}}{\partial \theta} \middle| \Omega_{t-1}\right) \xrightarrow{p} \mathbb{E}\left(\frac{1}{\gamma_{n0}} \frac{\partial \gamma_{nj}}{\partial \theta} \middle| \Omega_{t-1}\right) - \mathbb{E}\left(\frac{\rho_{nj}}{\gamma_{n0}} \frac{\partial \gamma_{n0}}{\partial \theta} \middle| \Omega_{t-1}\right).$$

When $n \to \infty$, $\gamma_{n0} \xrightarrow{p} \gamma_0$ and $\rho_{nj} \xrightarrow{p} \rho_j$. Since $\mathbb{E}((H_t(\alpha,\theta_0) - \alpha/2)\partial H_{t-j}(\alpha,\theta_0)/\partial\theta | \Omega_{t-1}) = \partial H_{t-j}(\alpha,\theta_0)/\partial\theta \, \mathbb{E}( (H_t(\alpha,\theta_0) - \alpha/2) | \Omega_{t-1}) = 0$ for $j > 0$, we get

$$\mathbb{E}\left(\frac{\partial \widetilde{\rho}_{nj}}{\partial \theta}\bigg| \Omega_{t-1}\right) \xrightarrow{p} R_j - 2\rho_j R_0,$$

with

$$R_j = \frac{1}{\gamma_0}\mathbb{E}_0\left(\frac{\partial \gamma_{nj}}{\partial \theta}\right) = \frac{1}{\gamma_0}\mathbb{E}_0\left((H_{t-j}(\alpha,\theta_0) - \alpha/2)\,\frac{\partial H_t(\alpha,\theta_0)}{\partial\theta}\right),$$

and $\gamma_0 = \alpha\,(1/3 - \alpha/4)$. Under the null $\rho_j = 0$ for $j = 1, ..., m$. Therefore

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

Notice that $\sqrt{n}\,(\rho_{n1}, ..., \rho_{nm})' \xrightarrow{d} \mathcal{N}(0, I_m)$ 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 $(\widehat{\rho}_{n1}, ..., \widehat{\rho}_{nm})'$. Under Assumptions A1-A4, we have

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

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

$$\Delta_{ij} = \delta_{ij} + \lambda R_i'\Sigma_0 R_j,$$

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

$\forall\,(i,j) \in \{1, ..., m\}^2$, where $\delta_{ij}$ is a dummy variable that takes a value 1 if $i = j$ and 0 otherwise. We 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

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

Finally

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

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

## Appendix G: Backtesting MES-based risk measure

**Proof.** Consider a MES-based risk measure expressed as

$$RM_{1t}(h) = \int_0^1 g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0), X_t)\mathrm{d}\beta,$$

with

$$\Pr\left(\widetilde{Y}_{1t+h} \leq Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0)|\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0);\Omega_t\right) = \beta.$$

In order to backtest the CoVaR and the MES, we define a joint violation of the $(\beta,\alpha)$-CoVaR of $\widetilde{Y}_{1t+h}$ and the $\alpha$-VaR of $\widetilde{Y}_{2t+h}$. This violation process is represented by the following binary variable

$$h_t(\alpha,\beta,\theta_0) = \mathbb{1}(\widetilde{Y}_{1t+h} \leq Co\widetilde{V}aR_{1t}(\beta,\alpha,\theta_0)) \times \mathbb{1}(\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0)).$$

Denote by $h_t(\alpha,\beta,\theta_0,X_t)$ the violation process associated to $g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0),X_t)$ and used to backtest the MES-based risk measure, for instance the SRISK. If the function $g_t(.)$ is monotonic decreasing with the MES (as it is the case for the SRISK given our sign convention for the MES), we have

$$h_t(\alpha,\beta,\theta_0,X_t) = \mathbb{1}(g_t(\widetilde{Y}_{1t+h},X_t) \geq g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0),X_t)) \times \mathbb{1}(\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0)).$$

The violation processes $h_t(\alpha,\beta,\theta_0)$ and $h_t(\alpha,\beta,\theta_0,X_t)$ are identical, in the sense that they take a value 1 at the same date. As a consequence, the cumulative joint violation processes used for backtesting the MES and a MES-based risk measure are identical.

$$H_t(\alpha,\theta_0,X_t) = \int_0^1 h_t(\alpha,\beta,\theta_0,X_t)\mathrm{d}\beta = \int_0^1 h_t(\alpha,\beta,\theta_0)\mathrm{d}\beta = H_t(\alpha,\theta_0).$$

The unconditional coverage test for a MES-based risk measure, say SRISK, corresponds to the null hypothesis

$$\mathrm{H}_{0,UC}^{\mathrm{SRISK}} : \mathbb{E}(H_t(\alpha,\theta_0,X_t)) = \alpha/2.$$

The corresponding test statistic $UC_{\mathrm{SRISK}}$ is given by

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

with $\bar{H}(\alpha,\widehat{\theta}_T,X)$ the out-of-sample mean of $H_t(\alpha,\widehat{\theta}_T,X_t)$. As $H_t(\alpha,\widehat{\theta}_T,X_t) = H_t(\alpha,\widehat{\theta}_T)$, $\forall t$, this statistic is equivalent to that used for backtesting the MES

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

## Appendix H: Backtesting ΔCoVaR

**Proof.** The first step of the proof consists in evaluating the two first conditional moments of $\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)$. The Bayes theorem implies that

$$\mathbb{E}(h_t(\alpha,\alpha,\theta_0)|\Omega_{t-1}) = \alpha^2,$$

$$\mathbb{E}\left(h_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)|\Omega_{t-1}\right)=\alpha(\beta_{\text{sup}}-\beta_{\text{inf}}),$$

leading to

$$\mathbb{E}\left(\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)|\Omega_{t-1}\right)\equiv\mu=\begin{pmatrix}\alpha^2\\ \alpha(\beta_{\text{sup}}-\beta_{\text{inf}})\end{pmatrix}. \tag{14}$$

According to Equation (14), $h_t\left(\alpha,\alpha,\theta_0\right)$ and $h_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)$ are Bernouilli distributed with conditional success probabilities $\alpha^2$ and $\alpha(\beta_{\text{sup}}-\beta_{\text{inf}})$, and conditional variances given by

$$\mathbb{V}\left(h_t\left(\alpha,\alpha,\theta_0\right)|\Omega_{t-1}\right)=\alpha^2\left(1-\alpha^2\right),$$

$$\mathbb{V}\left(h_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)|\Omega_{t-1}\right)=\alpha\left(\left(\beta_{\text{sup}}-\beta_{\text{inf}}\right)\left(1-\alpha\left(\beta_{\text{sup}}-\beta_{\text{inf}}\right)\right)\right).$$

In order to compute the statistic $UC_{\Delta CoVaR}$, we have to determine the general expression of the conditional variance-covariance matrix of $\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)$, and in particular, we need the conditional covariance between the two violations,

$$\begin{aligned}\text{Cov}\left(h_t(\alpha,\alpha,\theta_0),h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)|\Omega_{t-1}\right)&=\mathbb{E}\left(h_t\left(\alpha,\alpha,\theta_0\right)h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)|\Omega_{t-1}\right)\\&\quad-\mathbb{E}\left(h_t(\alpha,\alpha,\theta_0)|\Omega_{t-1}\right)\mathbb{E}\left(h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)|\Omega_{t-1}\right).\end{aligned} \tag{15}$$

Since $\alpha<\beta_{\text{inf}}$, $Y_{2t}\leq VaR_{2t}\left(\alpha,\theta_0\right)$ and $VaR_{2t}\left(\beta_{\text{inf}},\theta_0\right)\leq Y_{2t}\leq VaR_{2t}\left(\beta_{\text{sup}},\theta_0\right)$ are incompatible events, we get $h_t(\alpha,\alpha,\theta_0)\times h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)=0$ implying that the first expectation in Equation (15) is 0. We thus have

$$\text{Cov}\left(h_t(\alpha,\alpha,\theta_0),h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)|\Omega_{t-1}\right)=-\alpha^3(\beta_{\text{sup}}-\beta_{\text{inf}}),$$

and the conditional variance-covariance matrix of $\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)$ is given by

$$\mathbb{V}\left(\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)|\Omega_{t-1}\right)\equiv\gamma=\begin{pmatrix}\alpha^2(1-\alpha^2) & -\alpha^3(\beta_{\text{sup}}-\beta_{\text{inf}})\\ -\alpha^3(\beta_{\text{sup}}-\beta_{\text{inf}}) & \alpha(\beta_{\text{sup}}-\beta_{\text{inf}})(1-\alpha(\beta_{\text{sup}}-\beta_{\text{inf}}))\end{pmatrix}.$$

With the two first conditional moments of $\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)$ in hand, we now turn to the identification of its asymptotic distribution. Equation (14) implies that the sequence $\{\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)-\mu\}_{t=1}^{\infty}$ is a mds for any $(\alpha,\beta_{\text{inf}},\beta_{\text{sup}})\in[0,1]^3$ with

$$\mathbb{E}\left(\mathbf{h}_t\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)-\mu|\Omega_{t-1}\right)=0.$$

As a consequence, the Lindeberg-Levy central limit theorem implies that

$$\sqrt{n}\left(\bar{\mathbf{h}}\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0\right)-\mu\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\gamma\right), \tag{16}$$

with $\bar{\mathbf{h}}\left(\alpha,\beta_{\inf},\beta_{\sup},\theta_0\right) = 1/n\sum_{t=T+1}^{T+n}\mathbf{h}_t\left(\alpha,\beta_{\inf},\beta_{\sup},\theta_0\right)$. From Equation (16), we get immediately that

$$UC_{\Delta CoVaR} = n\left(\bar{\mathbf{h}}\left(\alpha,\beta_{\inf},\beta_{\sup},\theta_0\right)-\mu\right)'\gamma^{-1}(\bar{\mathbf{h}}\left(\alpha,\beta_{\inf},\beta_{\sup},\theta_0\right)-\mu) \xrightarrow{d} \chi(2).$$

We now turn to the case of the feasible test statistic $UC_{\Delta CoVaR} \equiv UC_{\Delta CoVaR}(\widehat{\theta}_T)$ and its asymptotic distribution. The sketch of the proof is similar to that of Theorem 1. First, rewrite

$$\sqrt{n}(\bar{\mathbf{h}}(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mu) = \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1})).$$

The continuous mapping theorem implies that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T))|\Omega_{t-1}) =&\\ \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1}))+o_p(1).&\end{aligned}$$

Rearranging these terms gives

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1})) =&\\ \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1}))&\\ +\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1})+o_p(1).&\end{aligned} \tag{17}$$

The mean value theorem implies that

$$\begin{aligned}\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T) &\equiv \begin{pmatrix} h_t(\alpha,\alpha,\widehat{\theta}_T)\\ h_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)\end{pmatrix}\\ &= \begin{pmatrix} h_t(\alpha,\alpha,\theta_0)\\ h_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)\end{pmatrix}+\begin{pmatrix}\left(\widehat{\theta}_T-\theta_0\right)'\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}\\ \left(\widehat{\theta}_T-\theta_0\right)'\frac{\partial h_t(\alpha,\beta_{\inf},\beta_{\sup},\tilde{\theta})}{\partial\theta}\end{pmatrix},\end{aligned}$$

where $\tilde{\theta}$ is an intermediate point between $\theta_0$ and $\widehat{\theta}_T$. Equation (17) becomes

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\widehat{\theta}_T)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)|\Omega_{t-1})) = \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\inf},\beta_{\sup},\theta_0)-\mu)&\\ +\begin{pmatrix}\sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)'\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\\ \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)'\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\beta_{\inf},\beta_{\sup},\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\end{pmatrix}+o_p(1).&\end{aligned} \tag{18}$$

Assume that $T \to \infty$, $n \to \infty$ and $n/T \to \lambda$ with $0 \leq \lambda < \infty$. Under the null hypothesis $\mathrm{H}_{0,UC}$, the first term of Equation (18) converges in distribution to a bivariate normal distribution with mean $\mathbf{0}$ and variance $\gamma$. The covariance between the first term and $\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right)$is 0 as $\widehat{\theta}_T$ depends on the in-sample observations and the summand in the first term is for out-of-sample observations. Under Assumption A4, $\tilde{\theta} \to \theta_0$, and since the two derivatives $\partial h_t(\alpha, \alpha, \theta_0)/\partial\theta$ and $\partial h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)/\partial\theta$ are also mds, we have

$$\frac{1}{n}\sum_{t=T+1}^{T+n} \mathbb{E}\left(\frac{\partial h_t(\alpha, \alpha, \tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \xrightarrow{p} R_C^{\alpha} = \mathbb{E}_0\left(\frac{\partial h_t(\alpha, \alpha, \theta_0)}{\partial\theta}\right),$$

$$\frac{1}{n}\sum_{t=T+1}^{T+n} \mathbb{E}\left(\frac{\partial h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \xrightarrow{p} R_C^{\beta} = \mathbb{E}_0\left(\frac{\partial h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)}{\partial\theta}\right).$$

where $\mathbb{E}_0(.)$ denotes the expectation with respect to the true distribution of $h_t(\alpha, \alpha, \theta_0)$ and $h_t(\alpha, \beta_{\inf}, \beta_{\sup}, \theta_0)$. As a consequence, we have

$$\begin{pmatrix} \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n} \mathbb{E}\left(\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \\ \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n} \mathbb{E}\left(\frac{\partial h_t(\alpha,\beta_{\inf},\beta_{\sup},\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \gamma_\lambda\right),$$

where $\gamma_\lambda = \begin{pmatrix} \lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\alpha} & \lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\beta} \\ \lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\beta} & \lambda R_C^{\beta\prime}\Sigma_0 R_C^{\beta} \end{pmatrix}$, and we can deduce that

$$\sqrt{n}\left(\bar{\mathbf{h}}\left(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T\right) - \mu\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \gamma + \gamma_\lambda\right).$$

The feasible test statistic $UC^C_{\Delta CoVaR}$ that takes into account the presence of estimation risk is then given by

$$UC^C_{\Delta CoVaR} = n\left(\bar{\mathbf{h}}\left(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T\right) - \mu\right)' (\gamma + \widehat{\gamma}_\lambda)^{-1} (\bar{\mathbf{h}}\left(\alpha, \beta_{\inf}, \beta_{\sup}, \widehat{\theta}_T\right) - \mu),$$

where $\widehat{\gamma}_\lambda$ denotes a consistent estimator of $\gamma_\lambda$ (see Appendix E). The robust statistic $UC^C_{\Delta CoVaR}$ converges to a chi-square distribution with two degrees of freedom whatever the value of $\lambda$, and is thus free of estimation risk.

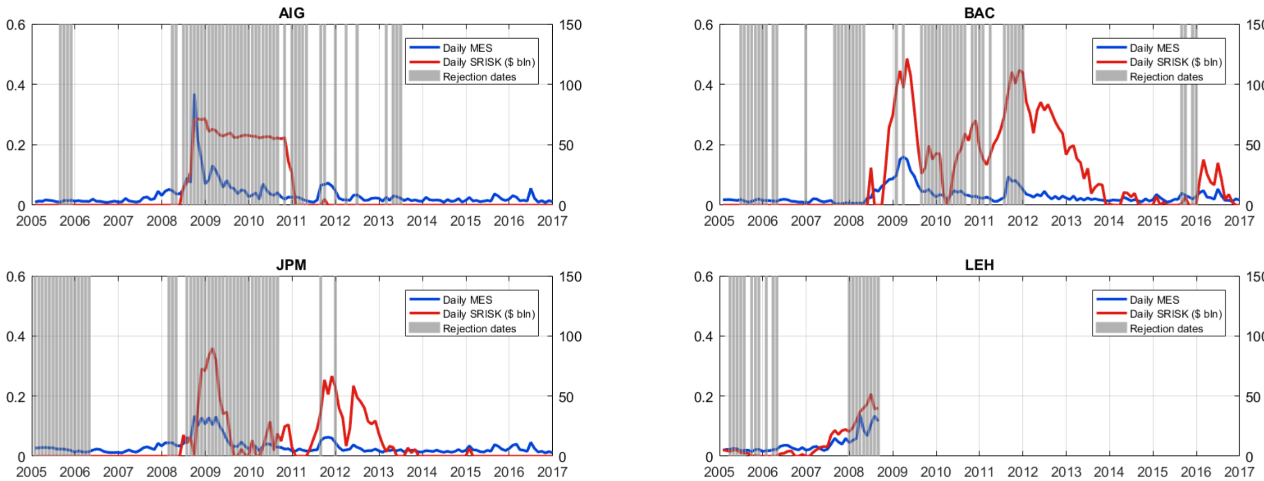
## Appendix I: List of tickers

Tickers and company names

| | | | |
|---|---|---|---|
| AET | Aetna | HBAN | Huntington Bancshares |
| AFL | Aflac | ICE | Intercontinental Exchange |
| ALL | Allstate Corp | JNS | Janus Capital |
| ABK | Ambac Financial Group | JPM | JP Morgan Chase |
| ACAS | American Capital | KEY | Keycorp |
| AXP | American Express | LM | Legg Mason |
| AIG | American International Group | LEH | Lehman Brothers |
| AMP | Ameriprise financial | LNC | Lincoln National |
| AMTD | TD Ameritrade | L | Loews |
| ANTM | Anthem | MI | Marshall & Ilsley |
| AON | Aon Corp | MMC | Marsh & McLennan |
| AIZ | Assurant | MA | Mastercard |
| MTB | M & T Bank Corp | MBI | MBIA |
| BAC | Bank of America | MER | Merrill Lynch |
| BK | Bank of New York Mellon | MET | Metlife |
| BBT | BB&T | MS | Morgan Stanley |
| BSC | Bear Stearns | NCC | National City Corp |
| WRB | W.R. Berkley Corp | NYCB | New York Community Bancorp |
| BRK | Berkshire Hathaway | NTRS | Northern Trust |
| BLK | Blackrock | NMX | Nymex Holdings |
| COF | Capital One Financial | NYX | NYSE Euronext |
| BOT | CBOT Holdings | PBCT | Peoples United Financial |
| CBG | CBRE Group | PNC | PNC Financial Services |
| CB | Chubb Corp | PFG | Principal Financial Group |
| CI | CIGNA Corp | PGR | Progressive |
| CINF | Cincinnati Financial Corp | PRU | Prudential Financial |
| CIT | CIT Group | RF | Regions Financial |
| C | Citigroup | SAF | Safeco |
| CME | CME Group | SCHW | Schwab Charles |
| CNA | CNA Financial Corp | SEIC | SEI Investments Company |
| CMA | Comerica | SLM | SLM Corp |
| CBH | Commerce Bancorp | SOV | Sovereign Bancorp |
| CBSS | Compass Bancshares | STT | State Street |
| CFC | Countrywide Financial | STI | Suntrust Banks |
| CVH | Coventry Health Care | SNV | Synovus Financial |
| AGE | A.G. Edwards | TMK | Torchmark |
| ETFC | E-Trade Financial | TRV | Travelers |
| FNM | Fannie Mae | TROW | T. Rowe Price |
| FNF | Fidelity National Financial | UB | Unionbancal Corp |
| FITB | Fifth Third Bancorp | UNH | Unitedhealth Group |
| BEN | Franklin Resources | UNM | Unum Group |
| FRE | Freddie Mac | USB | US Bancorp |
| GNW | Genworth Financial | WB | Wachovia |
| GS | Goldman Sachs | WM | Washington Mutual |
| HIG | Hartford Financial Group | WFC | Wells Fargo & Co |
| HNT | Health Net | WU | Western Union |
| HCBK | Hudson City Bancorp | ZION | Zion |
| HUM | Humana | | |

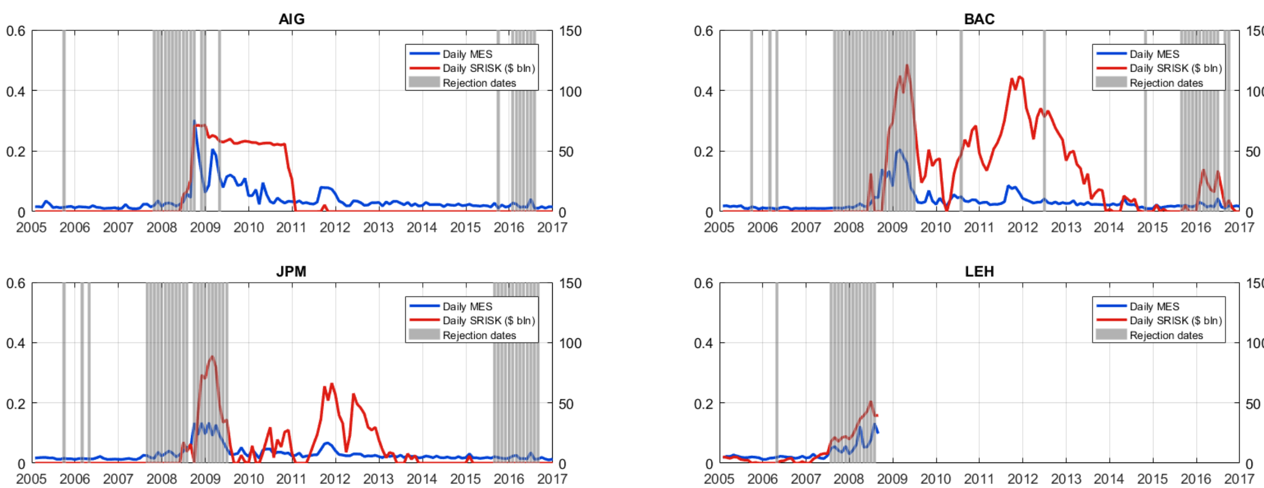
## Appendix J: Robustness checks for backtesting short-term risk measures

Figure 13: UC backtests for one-day risk forecast horizon (recursive estimation, $n = 500$)



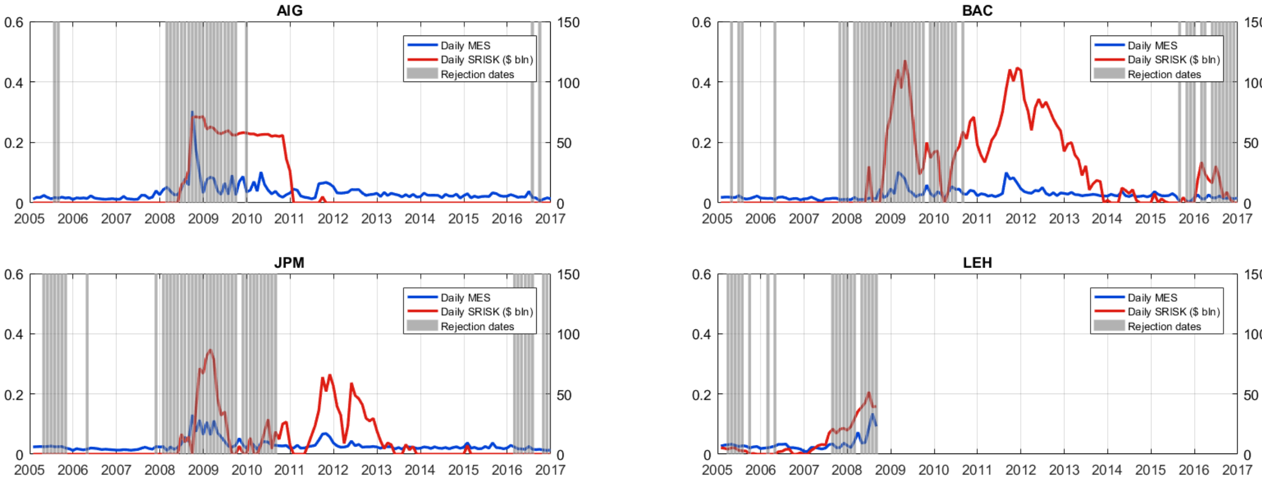

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 500$ out-of-sample observations and a recursive scheme for parameter estimation. Reported results are robust to estimation risk.

Figure 14: UC backtests for one-day risk forecast horizon (rolling estimation, $T = 500$, $n = 250$)



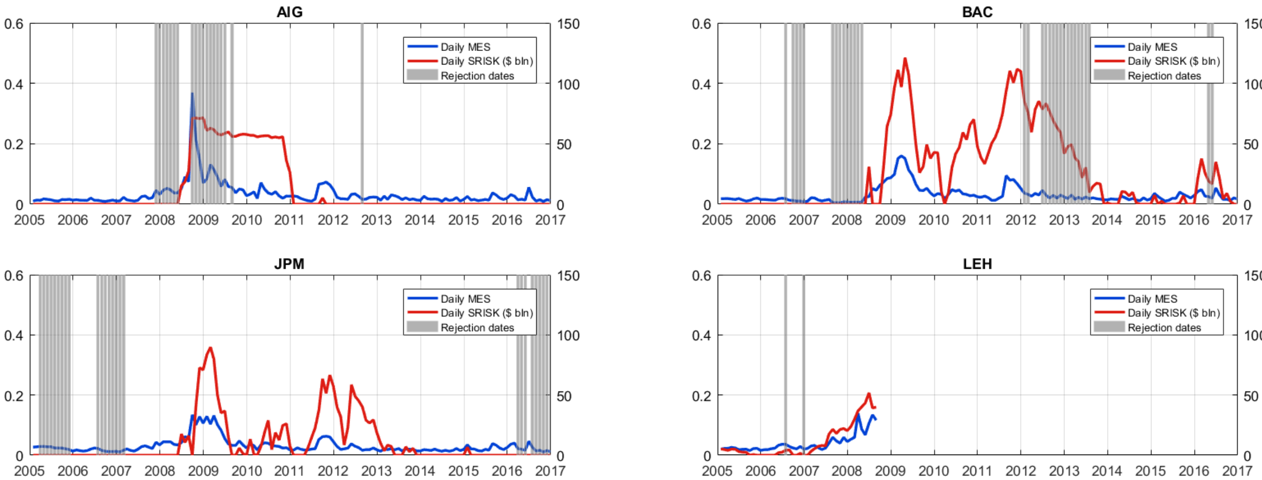

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a rolling scheme with $T = 500$ observations for parameter estimation. Reported results are robust to estimation risk.

Figure 15: UC backtests for one-day risk forecast horizon (rolling estimation, $T = 500$, $n = 500$)



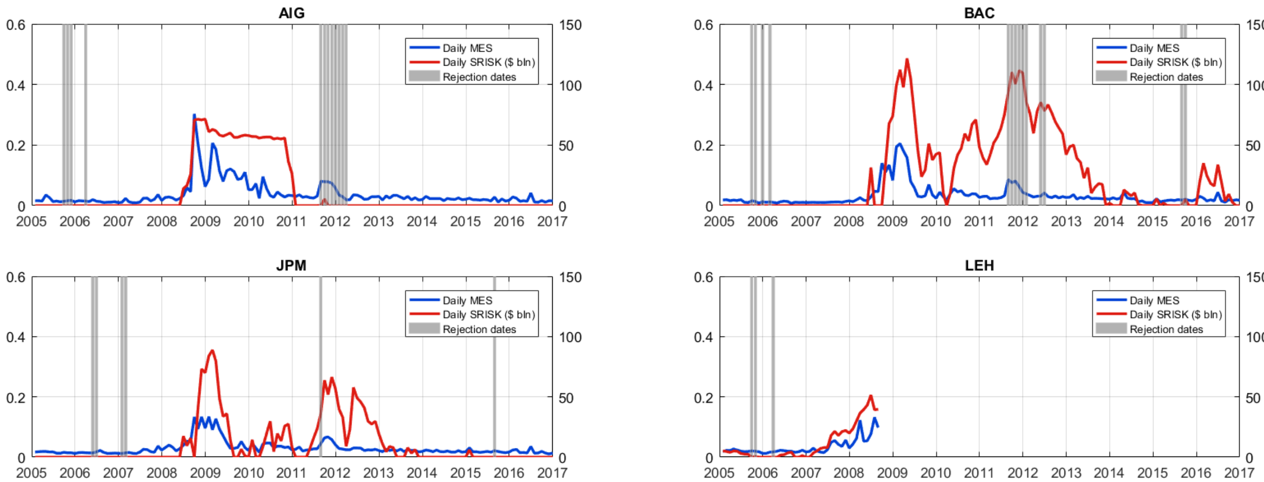

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the UC backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 500$ out-of-sample observations and a rolling scheme with $T = 500$ observations for parameter estimation. Reported results are robust to estimation risk.

Figure 16: IND backtests for one-day risk forecast horizon (recursive estimation, $n = 500$, $m = 5$)



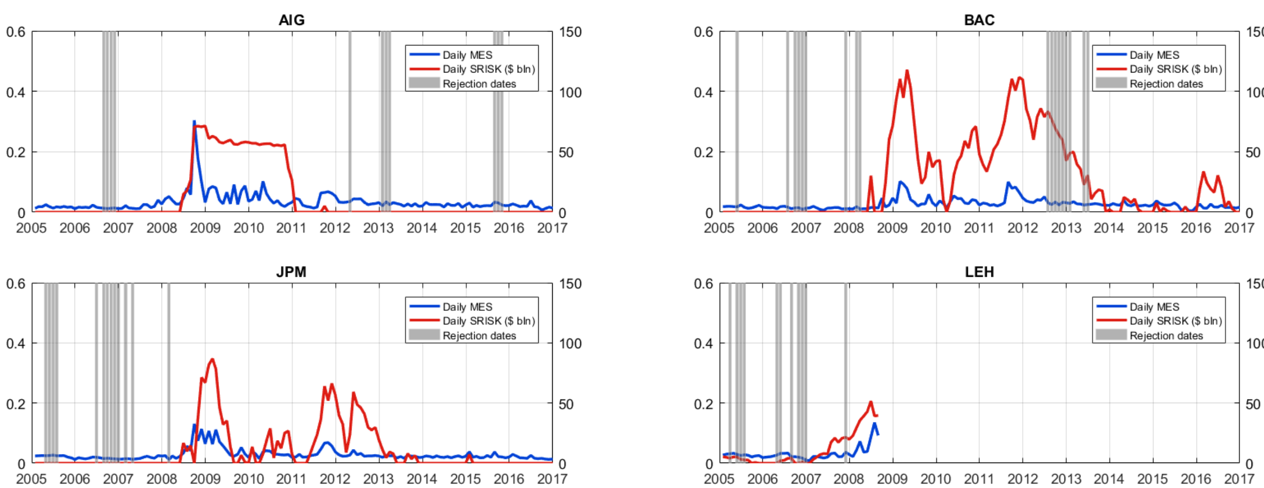

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the IND backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 500$ out-of-sample observations and a recursive scheme for parameter estimation. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 17: IND backtests for one-day risk forecast horizon (rolling estimation, $T = 500$, $n = 250$, $m = 5$)



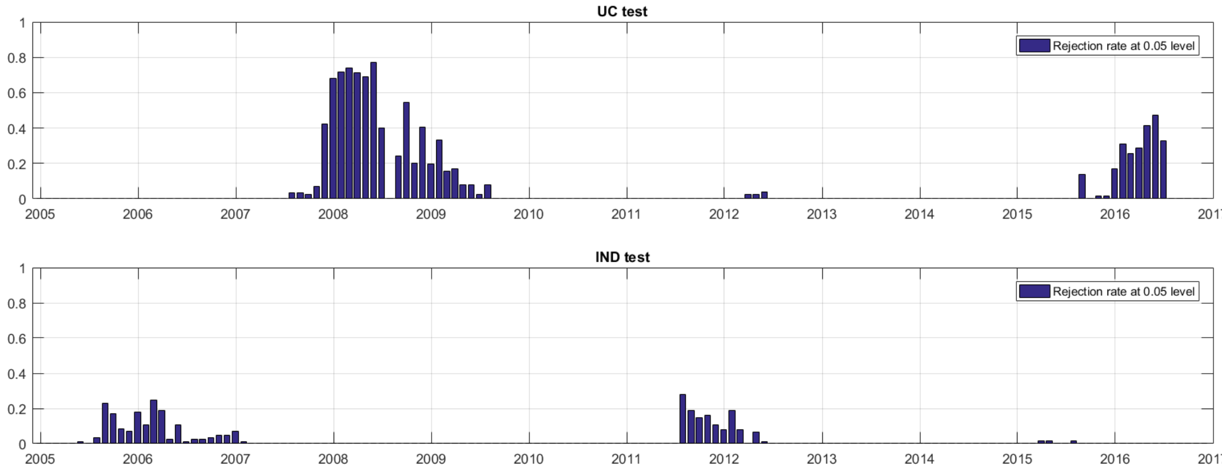

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the IND backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 250$ out-of-sample observations and a rolling scheme with $T = 500$ observations for parameter estimation. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 18: IND backtests for one-day risk forecast horizon (rolling estimation, $T = 500$, $n = 500$, $m = 5$)



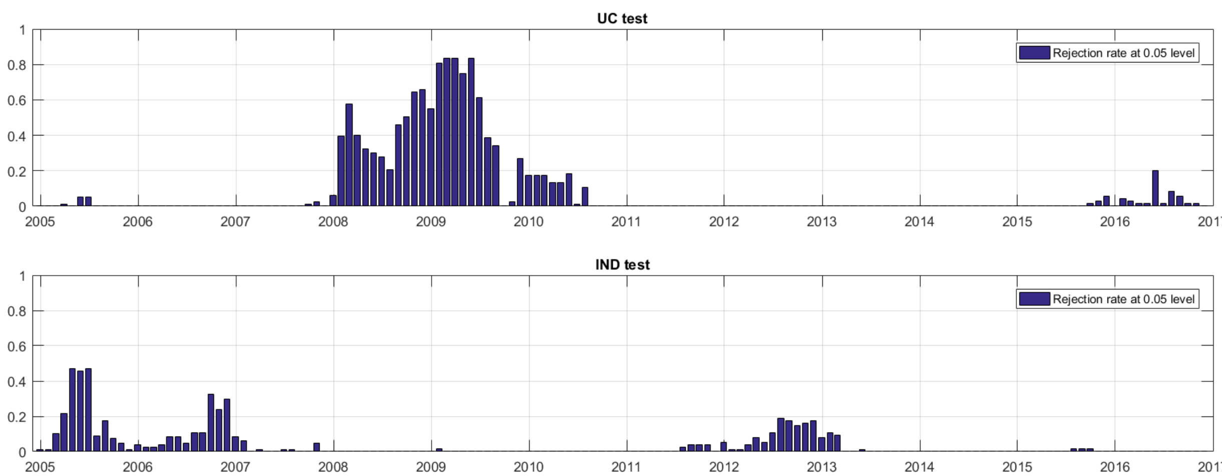

Note: This figure displays the daily MES (blue line), the daily SRISK (red line) and the rejection dates (shaded area) of the IND backtest at 5% significance level for AIG, BAC, JPM, and LEH. We use $n = 500$ out-of-sample observations and a rolling scheme with $T = 500$ observations for parameter estimation. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 19: Rejection rates of UC and IND backtests (rolling estimation, $T = 500$, $n = 250$, $m = 5$)




Note: This figure displays the rejection rates of the UC backtest (top panel) and IND backtest (bottom panel) at 5% significance level for the full list of tickers. We use $n = 250$ out-of-sample observations, a rolling scheme with $T = 500$ observations for parameter estimation, and a Bonferroni correction for multiple testing. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

Figure 20: Rejection rates of UC and IND backtests (rolling estimation, $T = 500$, $n = 500$, $m = 5$)




Note: This figure displays the rejection rates of the UC backtest (top panel) and IND backtest (bottom panel) at 5% significance level for the full list of tickers. We use $n = 500$ out-of-sample observations, a rolling scheme with $T = 500$ observations for parameter estimation, and a Bonferroni correction for multiple testing. We consider a maximum lag order $m = 5$. Reported results are robust to estimation risk.

## Appendix K: Computation of LRMES

In this section, we describe the methodology used to compute the MES, hereafter Long-Run MES (LRMES), and SRISK over an horizon $h > 1$. The computation approach differ from that used for the daily predictions, since the capital shortfall is calculated from the multi-period arithmetic returns instead of the arithmetic daily returns. The technique is thus closely related to that proposed by Brownlees and Engle (2017). Because we do not have closed form distribution for the multi-period arithmetic returns, we simulate daily log returns for which the distribution is known, and then aggregate these pseudo daily realizations to obtain its multi-period counterpart. The algorithm is as follows.

1. Randomly draw $S \times h$ pairs $\{z^s_{1T+t}, z^s_{2T+t}\}_{t=T+1}^{T+h}$ for $s = 1, \ldots, S$ of daily firm and market returns standardized innovations from the normal distribution. The number of pair is set to $S = 500000$ in the empirical application.

2. Build $S$ paths of the couple $\{Y^s_{1T+t}, Y^s_{2T+t}\}_{t=T+1}^{T+h}$ given the information known at time $T$ using the simulated standardized innovations.

3. Apply the formula $exp\left(\sum_{t=T+1}^{T+h} Y^s_{1T+t}\right) - 1$, and $exp\left(\sum_{t=T+1}^{T+h} Y^s_{2T+t}\right) - 1$, for $s = 1, \ldots, S$ to obtain the pseudo multi-period arithmetic returns $\{R^s_{1T+1:T+h}, R^s_{2T+1:T+h}\}_{s=1}^{S}$.

4. Compute the LRMES as the Monte Carlo average of the multi-period arithmetic returns as follows

$$LRMES_{1T+1:T+h}(\alpha, \widehat{\theta_T}) = -\frac{\sum_{s=1}^{S} R^s_{1T+1:T+h} \mathbb{1}\left(R^s_{2T+1:T+h} \leq VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)\right)}{\sum_{s=1}^{S} \mathbb{1}\left(R^s_{2T+1:T+h} \leq VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)\right)},$$

   where $VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)$ is the multi-period market decline, and is defined as

$$VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right) = \text{percentile}\left(\{R^s_{2T+1:T+h}\}_{s=1}^{S}, 100\alpha\right).$$

5. Build the $h$ step ahead SRISK forecasts as follows

$$SRISK_{1T+1:T+h}(\alpha, \widehat{\theta_T}) = kL_{1T} - (1-k)W_{1T}\left(1 + LRMES_{1T+1:T+h}(\alpha, \widehat{\theta_T})\right),$$

   where $k$ is the value of the prudential capital ratio, $L_{1T}$ is the book value of debt of the firm at time $T$, and $W_{1T}$ is the market value of the firm at time $T$. As for the daily case, we set $\alpha$ to 5%.

Overall, our simulated SRISK forecasts (not reported) are similar to those reported by Brownlees and Engle (2017) even if we consider a conditional multivariate normal distribution, and not a semi-parametric GARH-DCC model estimated by QML.

# References


Acharya, V., Engle, R., and Pierret, D. (2014). Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics*, 65:36–53.

Acharya, V., Engle, R., and Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64.

Acharya, V., Pedersen, L., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies*, 30(1):2–47.

Acharya, V., Pierret, D., and Steffen, S. (2016a). Capital shortfalls of european banks since the start of the banking union. New York: Stern School of Business, New York University.

Acharya, V., Pierret, D., and Steffen, S. (2016b). High time to tell european banks: No dividends. New York: Stern School of Business, New York University.

Adrian, T. and Brunnermeier, M. (2016). Covar. *American Economic Review*, 106(7):1705–1741.

Arnold, B., Beaver, R., Groeneveld, R., and Meeker, W. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58(3):471–488.

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

Banulescu, G.-D. and Dumitrescu, E.-I. (2015). Which are the sifis? a component expected shortfall approach to systemic risk. *Journal of Banking & Finance*, 50:575–588.

Basel Committee on Banking Supervision (2014). The g-sib assessment methodology – score calculation. Report, November 6.

Benoit, S., Colletaz, G., Hurlin, C., and Pérignon, C. (2013). A theoretical and empirical comparison of systemic risk measures. SSRN working paper.

Benoit, S., Colliard, J.-E., Hurlin, C., and Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. *Review of Finance*, 21(1):109–152.

Benoit, S., Hurlin, C., and Pérignon, C. (2019). Pitfalls in systemic-risk scoring. *Forthcoming in Journal of Financial Intermediation.*

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Billio, M., Getmansky, M., Lo, A., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

Bisias, D., Flood, M., Lo, A., and Valavanis, S. (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics*, 4(1):255–296.

Bontemps, C. and Meddahi, N. (2005). Measuring systemic risk. *Journal of Econometrics*, 124(1):149–186.

Boucher, C., Daníelsson, J., Kouontchou, P., and Maillet, B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, 44:72–92.

Box, G. and Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.

Brownlees, C., Chabot, B., Ghysels, E., and Kurz, C. (2020). Back to the future: backtesting systemic risk measures during historical bank runs and the great depression. *Journal of Banking & Finance*, 113.

Brownlees, C. and Engle, R. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies*, 30(1):48–79.

Chen, C., Iyengar, G., and Moallemi, C. (2013). An axiomatic approach to systemic risk. *Management Science*, 59(6):1373–1388.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Christoffersen, P. (2010). Backtesting. *Encyclopedia of Quantitative Finance*, R. Cont (ed). John Wiley and Sons.

Colletaz, G., Hurlin, C., and Pérignon, C. (2013). The risk map: A new tool for validating risk models. *Journal of Banking & Finance*, 37(10):3843–3854.

Couperier, O. and Leymarie, J. (2018). Backtesting expected shortfall via multi-quantile regression. Working paper.

Danielsson, J., James, K. R., Valenzuela, M., and Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, 23:79 – 91.

De Bandt, O. and Hartmann, P. (2002). Systemic risk: A survey, in financial crisis, contagion and the lender of last resort: A book of readings. ed. by C. Goodhart, and G. Illing. Oxford University Press.

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, 63(4):901–1269.

Dwyer, P. (1967). Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625.

Engle, R., Jondeau, E., and Rockinger, M. (2015). Systemic risk in europe. *Review of Finance*, 19(1):145–190.

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

Escanciano, J. C. and Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, 28(1):36–51.

European Banking Authority (2014). Results of 2014 eu-wide stress test. Report, October 26.

Ferreiro, J. (2018). A component delta conditional expected shortfall measure in the financial systemic risk framework. Working paper.

Financial Times (2014). Alternative stress tests find french banks are weakest in europe. October 27.

Francq, C. and Zakoïan, J.-M. (2016). Looking for efficient qml estimation of conditional vars at multiple risk levels. *Annals of Economics and Statistics*, (123-124):9–28.

Giglio, S., Kelly, B., and Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471.

Girardi, G. and Ergün, T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Gouriéroux, C. and Zakoïan, J.-M. (2013). Estimation-adjusted var. *Econometric Theory*, 29(4):735–770.

Ho, H., Lin, T.-I., Chen, H.-Y., and Wang, W.-L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, 142(1):25–40.

Horrace, W. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis*, 94(1):209–221.

Hué, S., Lucotte, Y., and Tokpavi, S. (2019). Measuring network systemic risk contributions: A leave-one-out approach. *Journal of Economic Dynamics and Control*, 100:86–114.

Hurlin, C. and Tokpavi, S. (2006). Backesting value-at-risk accuracy: A simple new test. *Journal of Risk*, 9(2):19–37.

Idier, J., Lamé, G., and Mésonnier, J.-S. (2014). How useful is the marginal expected shortfall for the measurement of systemic exposure? a practical assessment. *Journal of Banking & Finance*, 47:134–146.

Jokivuolle, E., Tunaru, R., and Vioto, D. (2018). Testing the systemic risk differences in banks. Bank of Finland Discussion Paper.

Jorion, P. (2007). Value-at-risk. McGraw-Hill, Third edition.

Kratz, M., Lok, Y., and McNeil, A. (2018). Multinomial var backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88:393–407.

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

Lambert, P., Laurent, S., and Veredas, D. (2012). Testing conditional asymmetry: A residual-based approach. *Journal of Economic Dynamics and Control*, 36(8):1229–1247.

Lazar, E. and Zhang, N. (2019). Model risk of expected shortfall. *Journal of Banking & Finance*, 105:74–93.

Leccadito, A., Boffelli, S., and Urga, G. (2014). Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. *International Journal of Forecasting*, 30(2):206–216.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Nieto, M. and Ruiz, E. (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting*, 32(2):475–501.

Noyer, C. (2014). Présentation des résultats de l'evaluation complète des bilans bancaires. ACPR Banque de France, October 26.

Pérignon, C. and Smith, D. (2008). A new approach to comparing var estimation methods. *Journal of Derivatives*, 16(2):54–66.

Philippon, T., Pessarossi, P., and Camara, B. (2017). Backtesting european stress tests. NBER working paper.

Pierret, D. and Steffen, S. (2018). Capital shortfalls at european banks after the 2018 stress test. *AllNews*, November 7.

Rabemananjara, R. and Zakoian, J.-M. (1993). Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, 8(1):31–49.

Romano, J., Shaikh, A., and Wolf, M. (2008). Formalized data snooping based on generalized error rates. *Econometric Theory*, 24(2):404–447.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14(1):115–129.

Tavolaro, S. and Visnovsky, F. (2014). What is the information content of the srisk measure as a supervisory tool? ACPR Banque de France, Economics and Financial Debates.

Wied, D., Weiß, G., and Ziggel, D. (2016). Evaluating value-at-risk forecasts: A new set of multivariate backtests. *Journal of Banking & Finance*, 72:121–132.

Wu, D. (2018). Systemic risk and bank failure. SSRN working paper, http://dx.doi.org/10.2139/ssrn.2492883.