Journal of Financial Econometrics, 2024, 1–29
https://doi.org/10.1093/jjfinec/nbae005
Article

Measuring and Testing Systemic Risk from the Cross-Section of Stock Returns†

Jesús Gil Jaime[1] and Jose Olmo [1,2,*]

[1]Universidad de Zaragoza, Zaragoza, Spain
[2]University of Southampton, Southampton, UK

*Address correspondence to Jose Olmo, Deaprtment of Economic Analysis, Universidad de Zaragoza, Gran Vía 2, 50005, Zaragoza (Spain), or e-mail: joseolmo@unizar.es

Abstract

This study proposes a novel measure of systemic risk that is obtained by aggregating downside risk information from the cross section of assets. In contrast to existing studies, we expand the analysis of systemic risk to many assets and focus on marginal measures of tail risk that are aggregated using a Fisher-type test to detect the risk of systemic events. The presence of downside risk for each asset of the cross section is examined through a bootstrap test of first-order stochastic dominance between the underlying tail distribution and the tail distribution of the residuals of a multivariate DCC-GARCH model. The application of these methods to the cross section of the FTSE-100 stock returns provides overwhelming evidence on the presence of financial instability during the period 2006–2009. Interestingly, we also find compelling evidence of systemic risk during the 2012–2015 period coinciding with the European debt crisis and after the outbreak of the coronavirus disease 2019 pandemic.

Keywords: downside risk, model misspecification, sequential limit theory, stochastic dominance test, systemic risk, Value-at-Risk

JEL classifications: C22, C23, C53, G01, G20

Systemic risk is identified as the main risk to financial stability. By definition, this threat involves many institutions simultaneously and typically affects the system as a whole. Among others, Brunnermeier et al. (2009) emphasize the distinction between microprudential regulation and macroprudential regulation. The former is focused on prudential controls at the firm level, whereas the latter considers the system as a whole. Although the impact of systemic events is a macroprudential concern, particular metrics of threats to financial stability may be applicable at either a microprudential or a macroprudential level. Alexander (2010) provides a useful perspective on this issue and enumerates four distinct policy applications of systemic risk measures: (i) By identifying individual institutions posing threats to financial stability, systemic risk measures help increase supervisory standards; (ii) by identifying specific structural aspects of the financial system that are particularly vulnerable, systemic risk measures help policymakers identify where regulations need to be improved; (iii) by identifying potential negative shocks to the financial system, systemic risk measures may help guide policy on how best to address those threats; and (iv) by alerting that the potential for financial instability is rising in the form of early warning signals, systemic risk measures can be used to inform policymakers about the need of tightening macroprudential policies.

† Jose Olmo acknowledges financial support from Ministerio de Ciencia e Innovacion/Agencia Estatal de Investigación (Grant PID2019-104326GB-I00), and from Fundación Agencia Aragonesa para la Investigación y el Desarrollo.

Received: August 09, 2023. **Revised:** March 16, 2024. **Editorial decision:** March 18, 2024. **Accepted:** March 22, 2024

The literature quantifying systemic risk has focused on (i) and (iv). In this article, we consider (iii) and (iv). To do this, we propose contemporaneous measures of systemic risk based on existing evidence of statistically significant downside risk in the cross section of stock returns. These measures are useful for gauging the fragility of the financial system and also serve as monitoring indicators of financial distress. In this group, we include the liquidity measures of Khandani and Lo (2011) and Hu, Pan, and Wang (2013); the Mahalanobis distance metric of Kritzman and Li (2010); and the absorption ratio of Kritzman et al. (2010). Prominent models for modeling systemic risk in a predictive setting may also be deployed as contemporaneous monitoring tools; see, for example, the metrics introduced in Chan-Lau et al. (2009), Adrian and Brunnermeier (2016), Brownlees and Engle (2016), and Acharya et al. (2017), among others.

Most of these measures of financial instability and systemic risk focus on the joint distribution of negative outcomes—tail events—of a collection of systemically important financial institutions. Thus, Adrian and Brunnermeier (2016) propose to measure systemic risk via an increase in the conditional Value-at-Risk (CoVaR) of the financial system. This measure, denominated ΔCoVaR, is able to quantify the risk contribution of individually systemically important institutions, which are so interconnected and large that they can cause risk spillover effects on others. The Co-Risk of Chan-Lau et al. (2009) measures the tail codependence between credit default swaps of various financial institutions. Allen, Bali, and Tang (2012) propose a systemic risk index called CATFIN, which associates systemic risk to the VaR of the financial system. Huang, Zhou, and Zhu (2009) measure systemic risk as the marginal contribution of a financial firm to the distress insurance premium of the financial sector. More recently, Brownlees and Engle (2016) introduce the SRISK that measures the expected capital shortfall of a firm conditional on a severe market decline. This metric is a function of the size of the firm, its degree of leverage, and its expected equity loss conditional on a market decline, which is denominated as Long Run Marginal Expected Shortfall (LRMES). Similarly, Acharya et al. (2017) introduce the marginal expected shortfall (MES) and systemic expected shortfall (SES). The first measure is defined as the expected decrease of an institution's net equity return conditional on a market decline. The SES extends the MES and measures the amount an institution's equity would drop below its target level (defined as the prudential capital fraction k of assets) in case of a future crisis when aggregate capital is less than k times aggregate assets. Similarly, the CoES measures the conditional expected shortfall (ES).

The main objective of these measures is to identify those firms that contribute most to systemic risk. These measures provide an informal ranking of firms that are at risk of triggering a systemic event. The literature does not provide, in general, a formal framework to statistically test if the financial system is under distress. Important exceptions considering pointwise tests are Kupiec and Guntay (2016) that construct test statistics for CoVaR and MES that can be used to detect systemic risk at the institutional level. Hurlin et al. (2017) propose a bootstrap-based test to compare statistically conditional risk measures at a single point in time and that includes systemic risk measures as particular cases. These methods also incorporate the presence of estimation uncertainty when comparing risk measures across firms and estimation methods. A related literature extending backtesting procedures to systemic risk has recently emerged. Banulescu-Radu et al. (2021) implement a backtesting procedure to assess the empirical validity of the MES, SRISK, and ΔCoVaR issued from a bivariate GARCH model and a dynamic conditional correlation (DCC) structure given by a GARCH-DCC model. More recently, Fissler and Hoga (2023) introduce the novel concept of multi-objective elicitability and propose bivariate scores equipped with the lexicographic order. Based on this concept, these authors are able to implement backtesting procedures for systemic risk measures such as CoVaR, CoES, and MES, and propose Diebold–Mariano type tests to compare systemic risk forecasts.

Most of the above measures and tests to measure systemic risk proxy the evolution of the financial system by a market return, and use bivariate models to capture the dependence between the systemic stocks and the market under extreme events. This simplification is very useful to obtain reliable econometric measures of systemic risk but may potentially overlook important information from the cross section of stock returns. These considerations motivate us to introduce an empirical methodology to measure and test for the presence of financial instability and systemic risk with information from the cross section of stock returns. We define systemic risk as evidence of simultaneous downside risk for a large proportion of firms in the financial system. In our setting, downside risk takes place if the probability of a tail event characterized by a predicted risk measure (e.g., VaR at $\alpha\%$ or CoVaR at $\beta\%$) is greater than the corresponding coverage probability (e.g., $\alpha\%$ or $\beta\%$). In contrast to the existing literature on financial risk management, we do not focus on a specific tail quantile but consider, instead, a continuum of tail quantiles to measure downside risk. This definition is made operational through a test of first-order stochastic dominance (FSD) comparing the empirical downside probability of the financial return against the downside probability of a suitable risk model. This approach provides a test that assesses the downside risk of the financial institution uniformly over the tail domain. The proposed approach can be interpreted as an extension of conventional backtesting procedures and specification tests for tail risk measures, see Kupiec (1995), Christoffersen (1998), Engle and Manganelli (2004), and Escanciano and Olmo (2010) for single VaR levels, and extensions to multilevel VaR models in which the VaR is defined for a finite set of coverage rates (see Pérignon and Smith (2008), Colletaz, Hurlin, and Pérignon (2013); Leccadito, Boffelli and Urga (2014); and Wied, Weiß, and Ziggel (2016), among others). Recent backtesting procedures have also focused on assessing ES measures such as Du and Escanciano (2017), Kratz, Lok, and McNeil (2018), and Couperier and Leymarie (2020), among others. The latter techniques are, however, applied to fixed tail quantiles, whereas our proposed approach is based on uniform tests of downside risk defined over the tail of the returns distribution.

Our approach is also different from these methods because it is based on contemporaneous information. Whereas the above methods are mainly concerned with testing the validity of out-of-sample predictions of tail risk measures, such as VaR and ES, our setting based on FSD tests is based on testing the ability of suitable risk models to model the probability of tail events uniformly over the tail domain. The use of contemporaneous information implies that the downside probabilities of the proposed risk model and the underlying data-generating process are estimated using the same information. Rejection of the FSD null hypothesis implies that the downside probability of the underlying process driving financial returns is larger than the downside probability of the proposed model for, at least, some values of the tail domain. This is evidence in our setting of downside risk. There are two interpretations of this outcome: (i) the proposed family of distributions underestimates the underlying tail risk. This interpretation is similar to the above backtesting procedures and implies that the proposed risk model is not appropriate; (ii) the underlying tail risk is larger than the best risk model prediction using in-sample information. The latter interpretation assumes that the fitted family of risk models is appropriate throughout the evaluation period but there are periods in which the underlying tail risk cannot be described by such family of distributions. Under this interpretation, the occurrence of downside risk is due to an abnormal market condition with potential to produce a tail event (very negative return) and not to a misspecified risk model. We consider a GJR-GARCH-DCC model as originally proposed by Brownlees and Engle (2016) for modeling the SRISK and also implemented by Banulescu-Radu et al. (2021) for systemic risk backtesting purposes. The main difficulty in a cross-sectional setting with many stocks is the estimation of the conditional volatility and, in particular, the estimation of the unconditional covariance matrix

used as input in the DCC model, see Engle, Ledoit, and Wolf (2019). To overcome this issue, in the empirical application, we apply simple linear shrinkage methods (Ledoit and Wolf 2004) to estimate consistently this large-dimensional covariance matrix.

The outcomes of the individual downside risk tests applied to the residuals of a multivariate GARCH-DCC model are aggregated to construct two indicators of systemic risk. One indicator measures the proportion of firms that report simultaneous downside risk. This indicator is obtained by adding up the binary outcomes (rejection or not) of the individual FSD tests over the cross-section of stock returns. The second indicator aggregates the value of the test statistic rather than the test outcome over the cross section of assets. By construction, the latter indicator is more informative than the test based on violations of the FSD condition because it does not censor the information contained in the FSD tests. Large values of both indicators signal the potential of systemic risk triggered by the presence of simultaneous downside risk in the system.

The article also introduces statistical tests based on these indicators to determine if the proposed risk models are suitable for capturing tail risk for the cross section of stock returns. These tests can be interpreted as tests of financial instability or systemic risk. More specifically, for the first indicator, systemic risk takes place when the expected number of firms exhibiting downside risk is larger than some significance level. This test is constructed from a Binomial distribution modeling the proportion of firms signaling distress in the tails and can be interpreted as a Fisher-type test for panel data. For the second indicator, a large value of the corresponding test statistic indicates that a large share of firms are under tail distress even if the assets do not formally reject the FSD hypothesis individually. Both tests converge, asymptotically, to a Normal distribution after applying sequential limit theory developed in Phillips and Moon (1999) under the assumption that the number of firms in the cross section (N) is small compared to the estimation period (T), such that $N/T \to 0$ as both $N, T \to \infty$.

The concept of FSD to monitor downside risk is appealing. However, hypothesis tests based on this property and, in general, composite tests given by multiple inequality restrictions are over-conservative (see Davidson and Duclos (2000) and Barrett and Donald (2003) for early tests of this condition) yielding underestimates of the nominal size of the tests and low power. Linton, Song, and Whang (2010) and Delgado and Escanciano (2013) partially solve this problem by proposing bootstrap and nonparametric type tests, respectively, that are consistent over the boundary of the null hypothesis. We apply similar resampling procedures based on bootstrap methods to account for the presence of estimation of the parametric risk models used by the modeler to capture tail risk. The use of bootstrap allows us to obtain a valid finite-sample distribution of the FSD test under the null hypothesis. The test is, however, potentially undersized due to the composite inequality constraints. To study the importance of this effect in finite samples, we also explore in the Monte Carlo section the performance of a Kolmogorov–Smirnov type test that assesses the goodness of fit of the proposed risk model in the tails, and the implications for systemic risk of using this test for the cross section of assets.

The proposed methodology to detect systemic risk is applied to daily data for a cross section of 68 stock returns containing the constituents of the FTSE-100 index that have remained in the main UK stock index during the evaluation period 2000–2022. The dynamics of financial returns are modeled using the fat tailed GJR-GARCH-DCC model that assumes a multivariate Student-t distribution with five degrees of freedom. The location-scale model parameters are estimated over rolling windows. We construct the proposed indicators of systemic risk introduced above at $\tau = 0.10, 0.05$ coverage probabilities, and find empirical evidence of episodes of financial distress in the system during the global financial crisis covering the period 2006–2009, the sovereign debt crisis between 2012 and 2015 and after March 2020 due to the outbreak of the coronavirus disease (COVID-

19) pandemic. These empirical findings suggest that during these periods the dynamics of financial returns are too extreme to be driven by the above multivariate time series process.

The article is structured as follows. Section 1 introduces the multivariate setting necessary for measuring systemic risk. The section also discusses standard measures of marginal downside risk such as the VaR and ES. Section 2 introduces the tests of marginal downside risk based on FSD applied to individual firms and derives the corresponding asymptotic and bootstrap distributions. Section 3 introduces the tests of systemic risk based on aggregate cross-sectional measures and derives the corresponding asymptotic distributions. Section 4 presents a comprehensive Monte Carlo simulation to study the finite-sample properties of the tests. Section 5 illustrates the performance of the tests to measure systemic risk with daily data on the constituents of the FTSE-100 index over the period 2000–2022. A final section concludes. Tables and figures are collected at the end of the document.

1 Measuring Systemic Risk

To measure systemic risk, we propose a multivariate location-scale process defined as

$$Y_t = \mu_t(W_{t-1};\theta_1) + [\Sigma_t(W_{t-1};\theta_2)]^{1/2}\varepsilon_t, \tag{1}$$

with $Y_t = (y_{1,t},\ldots,y_{N,t})'$; $\mu_t(W_{t-1};\theta_1) = E[Y_t \mid W_{t-1}]$ is a $N\times1$ vector with the location processes and $\Sigma_t(W_{t-1};\theta_2)$ is the $N\times N$ conditional covariance matrix. The agents' information set is given by W_{t-1} which may contain past values of Y_t and other variables of relevance for predicting the multivariate distribution of financial returns. The multivariate error term $\varepsilon_t = (\varepsilon_{1,t},\ldots,\varepsilon_{N,t})'$ is assumed to be *iid* and satisfies that $E[\varepsilon_t \mid W_{t-1}] = 0$.

We extend the existing literature based on bivariate models and consider a GARCH-DCC process for modeling the cross section of N assets, where N is potentially large.

1.1 GARCH-DCC Modeling

The DCC model was proposed in Engle (2002) to capture the time-varying correlation between financial returns. Let $Y_t = (y_{1t},\ldots,y_{Nt})'$ be an $N\times1$ vector of financial returns. The dynamics of the components are assumed to follow an AR(1)-GJR-GARCH(1,1) model such that

$$\begin{aligned} y_{i,t} &= \alpha_{i,0} + \alpha_{i,\mu} y_{i,t-1} + \xi_{i,t}, \\ \xi_{i,t} &= \sigma_{ii,t}\,\varepsilon_{i,t}, \end{aligned} \tag{2}$$

where $\xi_{i,t}$ is the error term and $\varepsilon_{i,t}$ is an innovation process with $E[\varepsilon_{i,t} \mid W_{t-1}] = 0$ and $E[\varepsilon_{i,t}^2 \mid W_{t-1}] = 1$; $\alpha_{i,0}$ and $\alpha_{i,\mu}$ are the parameters of the autoregressive process with $|\alpha_{i,\mu}|<1$ to ensure stationarity of the process $y_{i,t}$ for $i = 1,\ldots,N$. The DCC model is estimated in two steps. In the first step, univariate GARCH type models are fitted to each time series of returns and estimates of their conditional variances are obtained. In the second step, the standardized residuals $\varepsilon_{i,t} = \xi_{i,t}/\sigma_{ii,t}$ are used to estimate the time-varying correlation matrix.

The conditional variance process can be expressed as $\Sigma_t = D_t R_t D_t$, with $R_t = [\rho_{ij,t}]$ the conditional correlation matrix, and D_t a diagonal matrix with time-varying standard deviations on the diagonal. Thus,

$$\begin{aligned} D_t &= diag(\sigma_{11,t},\ldots,\sigma_{NN,t}), \\ R_t &= diag\left(\sigma_{11,t}^{-1/2},\ldots,\sigma_{N\,N,t}^{-1/2}\right)\Sigma_t diag\left(\sigma_{11,t}^{-1/2},\ldots,\sigma_{N\,N,t}^{-1/2}\right). \end{aligned} \tag{3}$$

To capture stylized facts such as the leverage effect, we model the idiosyncratic conditional variances as univariate GJR-GARCH models:

$$\sigma^2_{ii,t} = \omega_i + (\alpha_i + \gamma_i I\{\xi_{i,t-1}<0\})\xi^2_{i,t-1} + \beta_i \sigma^2_{ii,t-1}, i = 1, \ldots, N. \quad (4)$$

The matrix $\Sigma_t = [\sigma_{ij,t}]$ in Equation (3) is symmetric positive definite and is specified as

$$\Sigma_t = (1-\psi_1-\psi_2)\bar{\bar{\Sigma}} + \psi_1 \varepsilon_{t-1}\varepsilon^T_{t-1} + \psi_2 \Sigma_{t-1}, \quad (5)$$

where $\bar{\bar{\Sigma}} = E[\varepsilon_t \varepsilon_t^T]$ is the unconditional covariance matrix of the standardized residuals ε_t obtained from the first step estimation; ψ_1 and ψ_2 are non-negative scalars satisfying $0<\psi_1+\psi_2<1$. The correlation estimator is given by $\rho_{ij,t} = \frac{\sigma_{ij,t}}{\sigma_{ii,t}\sigma_{jj,t}}$.

The GARCH-DCC parameters are estimated using maximum likelihood methods under the assumption of joint normality of the error terms. The matrix $\bar{\bar{\Sigma}}$ is usually estimated using the sample covariance matrix, see Engle (2002). However, in large dimensions, this nonparametric estimator of $\bar{\bar{\Sigma}}$ may be very imprecise and ill-conditioned, see Engle et al. (2019). Ledoit and Wolf (2003, 2004), in a sequel of papers, introduce several shrinkage methods to improve the estimation of the covariance matrix in large-dimensional settings. In the empirical application below, we apply the simple linear shrinkage method introduced in Ledoit and Wolf (2004), see also Engle et al. (2019) for DCC models, and propose the following estimator of $\bar{\bar{\Sigma}}$:

$$\hat{\bar{\bar{\Sigma}}} = \lambda I_N + (1-\lambda)\hat{\bar{\bar{\Sigma}}}, \quad (6)$$

where I_N is the identity matrix of dimension N and $0<\lambda<1$ is the shrinkage parameter obtained from minimizing the quadratic loss function $E[||\hat{\bar{\bar{\Sigma}}} - \bar{\bar{\Sigma}}||^2]$.

Another interesting approach to model the dynamics of the cross section of stock returns and, in particular, the time-varying conditional covariance matrix is the use of unspanned stochastic volatility models, recently introduced by Creal and Wu (2015, 2017). In these models, the cross-sectional dimension is reduced by considering latent factors that are estimated using Bayesian methods and a novel MCMC algorithm. Whereas these authors implement this approach to model the cross section of bond yields and their volatilities, the method can be naturally extended to model the cross section of stock returns. This is, nevertheless, beyond the scope of this study.

1.2 Defining Systemic Risk in a Multivariate Setting

This section discusses two definitions of systemic risk obtained from aggregating marginal measures of downside risk widely used in the literature.

1.2.1 VaR-based measures of downside risk

Let $q_\tau(W_{t-1}) = (q_{1,\tau_1}(W_{t-1}), \ldots, q_{N,\tau_N}(W_{t-1}))'$ be a multivariate quantile process that is defined by the following condition:

$$P\{Y_t \leq q_\tau(W_{t-1}) \mid W_{t-1}\} = \tau, \quad (7)$$

where $P\{\cdot \mid W_{t-1}\}$ denotes the cumulative distribution function of the vector Y_t conditional on the information set W_{t-1}; $\tau \in (0, 1)$ denotes the coverage probability that is usually identified with a probability in the tail. In contrast to the univariate setting, the above condition is not sufficient to identify the quantile process, see White, Kim, and Manganelli (2015). This is overcome in our setting by considering the multivariate location-scale model (1)

and, in particular, taking advantage of the *iid* assumption on the vector ε_t. Thus, the above expression is equivalent to $P\{\varepsilon_t \leq [\Sigma_t(W_{t-1};\theta_2)]^{-1/2}(q_\tau(W_{t-1}) - \mu_t(W_{t-1};\theta_1)) \mid W_{t-1}\}$ and such that

$$P\{\varepsilon_t \leq \Sigma_t^{-1/2}(q_\tau - \mu_t) \mid W_{t-1}\} = P\{\varepsilon_{1,t} \leq \tilde{q}_{1,\tau_1} \mid W_{t-1}\} \cdots P\{\varepsilon_{N,t} \leq \tilde{q}_{N,\tau_N} \mid W_{t-1}\}, \tag{8}$$

with $\tilde{q}_{i,\tau_i}$ the $i-th$ element of the vector $\Sigma_t^{-1/2}(q_\tau - \mu_t)$, where, for notational convenience, we have removed the dependence of the different functions on the conditioning set and parameters. These functions are the marginal quantile processes corresponding to each covariate and satisfy, by construction, the condition $P\{\varepsilon_{i,t} \leq \tilde{q}_{i,\tau_i} \mid W_{t-1}\} = \tau_i$ for $i = 1, \ldots, N$ such that condition (8) implies that $\tau_1 \cdots \tau_N = \tau$. To identify these functions, we take advantage of the homogeneity of the marginal distributions of the components of the vector ε_t in the DCC-GARCH specification such that $\tau_i = \tau^{1/N} \equiv \tau^*$. Thus, under the continuity of the marginal distributions of the error term, the above equality characterizes the marginal quantile process that is defined as $\tilde{q}_{i,\tau^*} = F_{\varepsilon_i}^{-1}(\tau^*)$, where $F_{\varepsilon_i}^{-1}(\tau^*)$ is the inverse of the cumulative distribution function of the covariate $\varepsilon_{i,t}$ evaluated at probability τ^*. Therefore, the multivariate quantile process is defined as

$$q_\tau(W_{t-1}) = \mu_t(W_{t-1};\theta_1) + [\Sigma_t(W_{t-1};\theta_2)]^{1/2} F_\varepsilon^{-1}(\tau^*), \tag{9}$$

with $F_\varepsilon^{-1}(\tau^*) = (F_{\varepsilon_1}^{-1}(\tau^*), \ldots, F_{\varepsilon_N}^{-1}(\tau^*))'$. Using this identification strategy, the marginal quantile processes $q_{i,\tau^*}(W_{t-1}) = \mu_{i,t}(W_{t-1};\theta_1) + \tilde{\Sigma}_{it}(W_{t-1};\theta_2) F_\varepsilon^{-1}(\tau^*)$ can be interpreted as the true conditional quantile process at a coverage probability τ^* associated to the covariate y_{it} for $i = 1, \ldots, N$; $\tilde{\Sigma}_{it}(W_{t-1};\theta_2)$ denotes the $i^{th}-$row of the matrix $[\Sigma_t(W_{t-1};\theta_2)]^{1/2}$.

The conditional multivariate quantile process $q_\tau(W_{t-1})$ is usually not known and needs to be approximated by a risk model imposed by the modeller. Consider the following (semi)parametric risk model $m_\tau(W_{t-1};\theta) = (m_{1,\tau^*}(W_{t-1};\theta), \ldots, m_{N,\tau^*}(W_{t-1};\theta))'$ for the dynamics of the conditional quantile process at $\tau \in (0,1)$. The relevant condition that we propose in this article to assess the presence of systemic risk is

$$P\{Y_t \leq m_\tau(W_{t-1};\theta) \mid W_{t-1}\} > \tau, \tag{10}$$

at a given time t.[1] Importantly, this condition can be interpreted as evidence of misspecification of the multivariate quantile process $m_\tau(W_{t-1};\theta)$. Similar conditions are tested in a univariate setting employing backtesting methods, see Kupiec (1995), Christoffersen (1998), as pioneering examples in the literature and, more formally, with specification tests for the quantile risk measure as in Engle and Manganelli (2004). Therefore, it is important to note that the main difference with this interpretation of the above condition is that we identify the presence of systemic risk with the simultaneous occurrence of downside risk events that lead to the violation of condition (10) at a given point in time. Therefore, we expect the multivariate risk model $m_\tau(W_{t-1};\theta)$ to be correctly specified in general, but at a given point in time, if this risk measure fails to report the appropriate coverage probability for a sufficiently large number of firms then we claim that there is evidence of systemic risk. This is explained in more detail below.

[1] Similar measures of systemic risk are proposed in the literature by considering the conditional version of Equation (10) in bivariate settings $Y_t = (y_{1t}, y_{2t})'$ involving the returns of a potentially systemic firm and the market portfolio. For example, the ΔCoVaR measure of Adrian and Brunnermeier (2016) measures systemic risk by an increase in the quantile measure $m_{1,\beta}(W_{t-1};\theta)$ of the market portfolio return (y_{1t}) conditional on two different realizations of the firm's return given by $y_{2t} = m_{2,\alpha}(W_{t-1};\theta)$ and $y_{2t} = m_{2,0.5}(W_{t-1};\theta)$, with $0 < \alpha, \beta < 0.5$ two potentially different coverage probabilities.

For simplicity, we assume the risk model $m_\tau(W_{t-1};\theta)$ proposed by the modeller is obtained from the multivariate location-scale distribution (1). In particular, the conditional mean and covariance processes are assumed to be correctly specified and the only difference is in the choice of the distribution of the error term ε_t. The presence of downside risk for a given asset is due to the inability of the proposed error distribution, denoted as $F_{o,i}(\cdot)$ hereafter, to capture the magnitude of the tail event. Aggregation of these downside risk events across firms provides evidence of systemic risk. More formally, the proposed risk model is given by

$$m_\tau(W_{t-1};\theta) = \mu_t(W_{t-1};\theta_1)+[\Sigma_t(W_{t-1};\theta_2)]^{1/2}F_o^{-1}(\tau^*), \tag{11}$$

with $F_o^{-1}(\tau^*) = (F_{o,1}^{-1}(\tau^*),\ldots,F_{o,N}^{-1}(\tau^*))'$ and $F_{o,i}^{-1}(\tau^*)$ is the inverse of the cumulative distribution function of the error term assumed by the modeller for asset i. Using the above algebra, we obtain $m_{i,\tau^*}(W_{t-1};\theta) = \mu_{i,t}(W_{t-1};\theta_1)+\tilde{\Sigma}_{it}(W_{t-1};\theta_2)F_o^{-1}(\tau^*)$ that can be interpreted as the VaR measure associated to the covariate y_{it} for $i = 1,\ldots,N$. Similarly, $\tilde{m}_{\tau^*}(W_{t-1};\theta) = [\Sigma_t(W_{t-1};\theta_2)]^{-1/2}(m_\tau(W_{t-1};\theta)-\mu_t(W_{t-1};\theta_1))$ such that $\tilde{m}_{i,\tau^*}(W_{t-1};\theta) = F_{o,i}^{-1}(\tau^*)$. In this setting, we say that asset i faces downside risk if

$$\tau^* = P\{\varepsilon_{i,t}\leq\tilde{q}_{i,\tau^*}(W_{t-1};\theta)\mid W_{t-1}\}<P\{\varepsilon_{i,t}\leq\tilde{m}_{i,\tau^*}(W_{t-1};\theta)\mid W_{t-1}\}. \tag{12}$$

This condition is equivalent to $P\{y_{i,t}\leq q_{i,\tau^*}(W_{t-1})\mid W_{t-1}\}<P\{y_{i,t}\leq m_{i,\tau^*}(W_{t-1})\mid W_{t-1}\}$ and implies that the risk measure $m_{i,\tau^*}(W_{t-1};\theta)$ is an under-conservative estimate of the true underlying conditional quantile process $q_{i,\tau^*}(W_{t-1})$. In practice, the latter quantile function is not observable, instead, we use a condition equivalent to Equation (12) to monitor the occurrence of downside risk for asset i at time t:

$$F_{\varepsilon_i}(\tilde{m}_{i,\tau^*}(W_{t-1};\theta))\leq F_{o,i}(\tilde{m}_{i,\tau^*}(W_{t-1};\theta)). \tag{13}$$

This condition guarantees that the proposed risk model $\tilde{m}_{i,\tau^*}$ is able to capture the underlying tail risk at a fixed coverage probability τ^* given that $F_{o,i}(\tilde{m}_{i,\tau^*}(W_{t-1};\theta)) = \tau^*$, by construction. Furthermore, if condition (13) is satisfied by all assets in the cross section, it follows that $P\{Y_t\leq m_\tau(W_{t-1};\theta)\mid W_{t-1}\}\leq\tau$, for $\tau = \tau^{*N}$, and there is no systemic risk. Alternatively, if the condition is violated by a significant number of assets then it is likely that $P\{Y_t\leq m_\tau(W_{t-1};\theta)\mid W_{t-1}\}>\tau$, entailing the possibility of systemic risk. More formally,

$$P\{Y_t\leq m_\tau(W_{t-1};\theta)\mid W_{t-1}\} = \tilde{\tau}_1\cdots\tilde{\tau}_N, \tag{14}$$

with $\tilde{\tau}_i = F_{\varepsilon_i}(\tilde{m}_{i,\tau^*})$, for $i = 1,\ldots,N$. Assessing the presence of systemic risk is equivalent to assessing if $\tilde{\tau}_1\cdots\tilde{\tau}_N>\tau$, with $\tau = \tau^{*N}$, or alternatively, assessing the condition

$$\frac{1}{N}\sum_{i=1}^{N}\ln\tilde{\tau}_i/\tau^*>0. \tag{15}$$

The above results are presented for a fixed coverage probability τ as is usually the case in the risk management literature. This can be easily extended to monitor downside risk uniformly over the tail domain. Thus, the relevant condition replacing Equation (13) is

$$F_{\varepsilon_i}(x)\leq F_{o,i}(x)\ \text{ for all }\ x\in(-\infty,q], \tag{16}$$

where $q = F_o^{-1}(\bar{\tau})$, with $\bar{\tau}$ a coverage probability defining the upper limit of the left tail of the returns distribution.

1.2.2 ES-based measures of downside risk

An alternative tail risk measure that has gained popularity for managing downside risk in recent years (see Artzner et al. (1999); Rockafellar and Uryasev (2002); Topaloglou, Vladimirou, and Zenios (2002); Du and Escanciano (2017); and Patton, Ziegel and Chen (2019), among many others) is the ES.

Systemic risk can be defined applying this risk measure in a multivariate setting. Thus, we say that the financial market represented by the vector of asset returns Y_t does not present evidence of systemic risk if

$$ES(m_{i,\tau^*}(W_{t-1};\theta)) \leq ES_o(m_{i,\tau^*}(W_{t-1};\theta)), \tag{17}$$

for all $i = 1, \ldots, N$, with $ES(m_{i,\tau^*}(W_{t-1};\theta)) = E[-y_{i,t} \,|\, y_{i,t} \leq m_{i,\tau^*}(W_{t-1};\theta), W_{t-1}]$ and $ES_o(m_{i,\tau^*}(W_{t-1};\theta))$ the same expectation but computed under the probability distribution $F_{o,i}$ introduced above. Similarly, let $P_o\{\cdot \,|\, y_{it} \leq m_{i,\tau^*}(W_{t-1};\theta), W_{t-1}\}$ denote the probability computed under $F_{o,i}$. Applying integration by parts, we note that a sufficient condition for the above inequality to be satisfied is

$$P\{y_{it} \leq x \,|\, y_{it} \leq m_{i,\tau^*}(W_{t-1};\theta), W_{t-1}\} \leq P_o\{y_{it} \leq x \,|\, y_{it} \leq m_{i,\tau^*}(W_{t-1};\theta), W_{t-1}\},$$

for $x \in (-\infty, m_{i,\tau^*}(W_{t-1};\theta)]$. Operating with these expressions and using the location-scale specification (1), the above set of conditions can be expressed as

$$F_{\varepsilon_i,\tilde{m}_i}(x) \leqslant F_{o,\tilde{m}_i}(x), \quad \text{for } x \in (-\infty, m_{i,\tau^*}(W_{t-1};\theta)], \tag{18}$$

with $F_{\varepsilon_i,\tilde{m}_i}(x) = P\{\varepsilon_{it} \leq x \,|\, \varepsilon_{it} \leq \tilde{m}_{i,\tau^*}(W_{t-1};\theta), W_{t-1}\}$ and $F_{o,\tilde{m}_i}(x)$ the same probability distribution but evaluated under the probability distribution $F_{o,i}(\cdot)$ corresponding to the risk model $\tilde{m}_{i,\tau^*}(W_{t-1};\theta)$. Systemic risk can be interpreted as the simultaneous violation of condition (18) for a significant number of assets in the cross section as explained below.

To finish the section it is worth noting that the above FSD condition (18) for the conditional tail probabilities also implies, for suitable choices of $\bar{\tau}$ (e.g., $\tilde{m}_{i,\bar{\tau}}(W_{t-1};\theta) = \tilde{q}_{i,\bar{\tau}}(W_{t-1};\theta) = q$), the set of conditions in Equation (16). Therefore, it is sufficient to test the FSD condition (18) to jointly test the validity of conditions (16) and (17) uniformly over the tail domain characterized by the quantile q. Similarly, we can extend the above systemic risk conditions to make them uniform over the tail domain. Thus, condition (10) characterizing systemic risk is replaced by a uniform tail condition given by

$$P\{Y_t \leq x \,|\, Y_t \leq m_{\bar{\tau}}(W_{t-1};\theta),\ W_{t-1}\} > \tau(x),\ \forall x \in (-\infty, q]^N, \tag{19}$$

with $x = (x_1, \ldots, x_N)'$ and $\tau(x) = P\{Y_t \leq x \,|\, Y_t \leq q_{\bar{\tau}}(W_{t-1};\theta),\ W_{t-1}\}$. Similarly, condition (15) for measuring systemic risk is replaced by

$$\frac{1}{N}\sum_{i=1}^{N}\left(\ln\frac{\tilde{\tau}_i(x)}{\bar{\tau}^*} - \ln\frac{\tau^*(x)}{\bar{\tau}^*}\right) > 0,\ \forall x \in (-\infty, q]^N. \tag{20}$$

This condition will be explored further in Section 3 to develop statistical tests of systemic risk.

2 Testing Downside Risk

This section proposes a statistical test for assessing the existence of downside risk as characterized in expression (18). This condition can be interpreted as a test of FSD:

$$\begin{cases} H_{0F} : F_{\varepsilon_i,q}(x) \leq F_{o,q}(x), \text{ for all } x \leq q, \\ H_{AF} : F_{\varepsilon_i,q}(x) > F_{o,q}(x), \text{ for some } x \leq q. \end{cases} \tag{21}$$

To obtain the asymptotic distribution of the above test, we consider the worst-case scenario that is characterized by equality between both tail distributions. This condition provides a conservative test, thus, under the null hypothesis given by the composite hypothesis $F_{\varepsilon_i,q}(x) = F_{o,q}(x)$, for all $x \leq q$, a suitable test statistic is

$$D_{i,T}(q) = \sqrt{T} \sup_{\{x \leq q\}} (\hat{F}_{\varepsilon_i,q}(x) - F_{o,q}(x)), \tag{22}$$

where $\hat{F}_{\varepsilon_i,q}(x) = \hat{F}_{\varepsilon_i}(x)/\hat{F}_{\varepsilon_i}(q)$, with $\hat{F}_{\varepsilon_i}(x) = \frac{1}{T}\sum_{t=1}^{T} 1(\varepsilon_{i,t} \leq x)$ the empirical distribution function and $1(\cdot)$ an indicator function that takes a value of one if the argument is true and zero, otherwise. Under the null hypothesis H_{0F} and assuming that the vector of innovations $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is *iid*, the asymptotic distribution of $D_{i,T}(q)$ is given by the distribution of the supremum of a Brownian bridge $G_{F_{o,q}}$. More formally,

$$D_{i,T}(q) \xrightarrow{d} \sup_{\{x \leq q\}} G_{F_{o,q}}(x), \tag{23}$$

where $\xrightarrow{d}$ denotes convergence in distribution, see Kolmogorov (1933) and van der Vaart (1998). The critical value c_{α_d} of the test (22) is obtained under the null hypothesis from the condition $P\{\sup_{\{x \leq q\}} G_{F_{o,q}} > c_{\alpha_d}\} = \alpha_d$, with $0 < \alpha_d < 1$.

Critical values are obtained under the assumption that the tested distributions are equal. Note, however, that if the composite inequality condition characterizing the null hypothesis (21) holds strictly, these asymptotic critical values are conservative choices of the true critical values, that is, $P\{D_{i,T}(q) > c_{\alpha_d}\} \leq \alpha_d$, see Linton, Maasoumi, and Whang (2005), Linton, Song, and Whang (2010), Delgado and Escanciano (2013), and Gonzalo and Olmo (2014) for a discussion of stochastic dominance tests valid under the boundary of the null hypothesis. This choice of critical values implies, under the null hypothesis, that the risk model $m_{i,\tau}(W_{t-1};\theta)$ may be over-conservative, that is, $q_{i,\tau}(W_{t-1}) \leq m_{i,\tau}(W_{t-1};\theta) < 0$, with the strict inequality condition for some $\tau \leq \bar{\tau}$. This property also implies a loss of power under the alternative hypothesis H_{AF}.

The empirical properties of the test are studied in the simulation section below for different data-generating processes. The simulation section also discusses a goodness of fit test based on a Kolmogorov–Smirnov statistic to assess power losses due to considering the composite inequality condition under the null hypothesis.

2.1 Estimation Effects

In the above setting, the *iid* standardized residuals are obtained from the GARCH-DCC model. Let $\mu_t(W_{t-1};\theta_1)$ be the $N \times 1$ location process and $\Sigma_t(W_{t-1};\theta_2)$ be a $N \times N$ matrix characterizing the scale function in the multivariate location-scale process

$$Y_t = \mu_t(W_{t-1};\theta_1) + [\Sigma_t(W_{t-1};\theta_2)]^{1/2}\varepsilon_t,$$

with $Y_t = (y_{1t}, \ldots, y_{Nt})'$ and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$. In practice, the parameters characterizing the location-scale functions are not known and need to be estimated from the data. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$ be a $\sqrt{T}-$consistent estimator of $\theta = (\theta_1, \theta_2)'$ and assume that

$$\sqrt{T}(\hat{\theta}-\theta) = \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\psi_\theta(Y_s)+o_{P_\theta}(1), \tag{24}$$

with ψ_θ an influence function such that $E[\psi_\theta] = 0$ and $E[||\psi_\theta||^2]<\infty$. This property is satisfied for the two-stage quasi-maximum likelihood estimator of the DCC-GARCH parameters, see Goncalves et al. (2023).

In this scenario, the convergence result (23) is no longer valid. There is an additional term in the asymptotic distribution of the univariate test statistics due to the estimation of the model parameters. Let $\hat{D}_{i,T}(q) = \sqrt{T}\sup_{\{x\leq q\}}(\hat{F}_{e_i,q}(x)-\hat{F}_{o,q}(x))$ be the modified test statistic for the FSD hypothesis, with $\hat{F}_{e_i,q}(x) = \frac{\sum_{t=1}^{T}1(e_{i,t}\leq x, e_{i,t}\leq q)}{\sum_{t=1}^{T}1(e_{i,t}\leq q)}$ the empirical distribution function and $e_t = \hat{\Sigma}_t^{-1/2}(Y_t-\hat{\mu}_t)$ the vector of residuals $e_t = (e_{1,t}, \ldots, e_{N,t})'$ of the estimated location-scale process (1), with $\hat{\mu}_t = \mu_t(W_{t-1};\hat{\theta}_1)$ and $\hat{\Sigma}_t = \Sigma_t(W_{t-1};\hat{\theta}_2)$. Similarly, $\hat{F}_{o,q}(x) = \hat{F}_o(x)/\hat{F}_o(q)$ is the estimated parametric conditional distribution function. Under these conditions, it is well known that

$$\hat{D}_{i,T}(q)\xrightarrow{d}\sup_{\{x\leq q\}}G_{\hat{F}_{o,q}}(x), \tag{25}$$

with $G_{\hat{F}_{o,q}}$ denoting a zero-mean Gaussian process. Applying the results in van der Vaart (1998), one can show that this distribution is defined as the supremum of the sum of a centered Gaussian process $G_{F_{o,q}}$ and the limiting distribution of $\sqrt{T}(\hat{F}_{o,q}(x)-F_{o,q}(x))$. The latter distribution is driven by the asymptotic behavior of the estimator $\hat{\theta}$ and the parametric form of the distribution function $F_o(\cdot)$. Thus, the distribution $G_{\hat{F}_{o,q}}$ cannot be universally tabulated. Fortunately, simulation and resampling methods can be applied to approximate the critical values in finite samples.

2.2 Bootstrap Approximation

This section introduces a bootstrap approximation of the finite-sample distribution of the test (21) under the null hypothesis H_{0F}. The p-value of the test given by $P_{H_{0F}}\left\{\hat{D}_{i,T}(q) > \sup_{\{x\leq q\}}G_{\hat{F}_{o,q}}(x)\right\}$ can be approximated in finite samples by $P\{\hat{D}^*_{i,T}(q) > \hat{D}_{i,T}(q)\,|\,\{Y_t\}_{t=1}^T\}$. The bootstrap version of the FSD test is $\hat{D}^\star_{i,T}(q) = \sqrt{T}\sup_{\{x\leq q\}}(\hat{F}_{e^*_i,q}(x)-\hat{F}^\star_{o,q}(x))$ that is obtained from the bootstrap empirical distribution function $\hat{F}_{e^*_i,q}(x) = \frac{\sum_{t=1}^{T}1(e^\star_{iT}\leq x, e^\star_{iT}\leq q)}{\sum_{t=1}^{T}1(e^\star_{iT}\leq q)}$, where $e^\star_{iT}$ denotes the bootstrap residuals of the estimated location-scale process (1) obtained under the null hypothesis H_{0F}. Similarly, $\hat{F}^\star_{o,q}(x) = \hat{F}^\star_o(x)/\hat{F}^\star_o(q)$ is the estimated parametric distribution function.

Although the distribution of $\hat{D}^*_{i,T}(q)$ is not directly observed, it can be approximated to any degree of accuracy by operating conditionally on $\{Y_t\} = \{y_{1,t}, \ldots, y_{N,t}\}_{t=1}^T$. The algorithm to compute the p-value of the test is described below.

Algorithm:

1. Estimate the parameter vector θ from the multivariate DCC-GARCH process applied to $Y_t = (y_{1,t}, \ldots, y_{N,t})'$ in Equation (1) using a $\sqrt{T}-$ consistent estimation procedure. Let $\hat{\mu}_t = \mu_t(W_{t-1}; \hat{\theta}_1)$ and $\hat{\Sigma}_t = \Sigma_t(W_{t-1}; \hat{\theta}_2)$ denote the corresponding estimated conditional mean and covariance processes.
2. Compute the test statistic $\hat{D}_{i,T}(q) = \sqrt{T}\sup_{\{x \leq q\}}(\hat{F}_{e_i,q}(x) - \hat{F}_{o,q}(x))$ for each asset in the cross section, where $\hat{F}_{e_i,q}(\cdot)$ is the empirical distribution function conditional on the tail event and $\hat{F}_{o,q}(x) = \hat{F}_o(x)/\hat{F}_o(q)$, with $x \leq q$, where $q = F_o^{-1}(\bar{\tau})$ is the quantile that defines the tail event.
3. Generate B bootstrap replicas of the vector $\{Y_t\}_{t=1}^T$ under the null hypothesis H_{0F} from the location-scale model (1). More specifically, let $Y_t^{\star(b)}$, with $b = 1, \ldots, B$, denote such process obtained as

$$Y_t^{\star(b)} = \hat{\mu}_t + \hat{\Sigma}_t^{1/2} \varepsilon_t^{\star(b)}, \tag{26}$$

   where $\varepsilon_t^{\star(b)} = (\varepsilon_{1,t}^{\star(b)}, \ldots, \varepsilon_{N,t}^{\star(b)})'$ is a vector of bootstrap error terms drawn from the parametric distribution F_o under H_{0F}.
4. Estimate the bootstrap conditional mean and covariance processes from the bootstrap sequence $\{Y_t^{\star(b)}\}_{t=1}^T$, and denote it as $\hat{\mu}_t^{\star(b)} = \mu_t(W_{t-1}; \hat{\theta}_1^{\star(b)})$ and $\hat{\Sigma}_t^{\star(b)} = \Sigma_t(W_{t-1}; \hat{\theta}_2^{\star(b)})$.
5. Obtain the vector of *iid* standardized bootstrap residuals $e_{0t}^{\star(b)} = [\hat{\Sigma}_t^{\star(b)}]^{-1/2}(Y_t^{\star(b)} - \hat{\mu}_t^{\star(b)})$ under H_{0F}.
6. Compute the bootstrap test statistic $\hat{D}_{i,T}^{\star(b)}(q) = \sqrt{T}\sup_{\{x \leq q\}}(\hat{F}_{e_i^{\star(b)},q}(x) - \hat{F}_{o,q}^{\star}(x))$ for each asset in the cross section, as defined above.

This algorithm yields a random sample of B observations from the distribution of $\sup_{\{x \leq q\}} G_{\hat{F}_{oq}}(x)$. Using the Glivenko–Cantelli theorem and previous assumptions, the empirical p-value conditional on $\{Y_t\}_{t=1}^T$ defined by $\hat{p}_{i,TB}^{\star} = \frac{1}{B}\sum_{b=1}^{B} 1(\hat{D}_{i,T}^{\star(b)}(q) > \hat{D}_{i,T}(q))$ converges, in probability, to the bootstrap distribution $P\{\hat{D}_{i,T}^{\star}(q) > \hat{D}_{i,T}(q) \mid \{Y_t\}_{t=1}^T\}$, as $B \to \infty$. For each asset, the conditional probability converges to the p-value obtained from the asymptotic distribution of the test statistic $\hat{D}_{i,T}(q)$, for $i = 1, \ldots, N$, as $T \to \infty$.

3 Measuring Systemic Risk

In this section, we propose a novel measure to monitor systemic risk that is based on the individual downside risk tests introduced above. For illustrative purposes, the exposition focuses first on the scenario given by $\mu_t(W_{t-1}; \theta_1)$ and $\Sigma_t(W_{t-1}; \theta_2)$ known. A natural strategy for testing for systemic risk is to adapt condition (20) by constructing its empirical counterpart. This condition involves the cross-sectional average of the difference in logs between the conditional tail distributions under the probability laws $F_{o,i}$ and $F_{e_i,q}$, for $i = 1, \ldots, N$. This condition has to be assessed uniformly over the interval $(-\infty, q]^N$. Importantly, the distribution of such statistics is not straightforward and requires of suitable normalization to obtain a tractable asymptotic distribution for devising a statistical test.

A simpler testing strategy is, instead, to aggregate the outcomes (rejection/nonrejection) of the individual downside risk tests with asymptotic distribution in Equation (23) over the cross section of assets. This technique takes advantage of the same information as the test based on condition (20) but is simpler to implement. It is also similar in spirit to various methods used in meta-analysis and, in particular, to Fisher's p-value combination method, see Fisher (1932), which integrates statistical significance from many statistical hypothesis tests to examine a joint null hypothesis that is given by an intersection of the individual null hypotheses. Let $d_i(q) = 1(D_{i,T}(q) > c_{\alpha_d})$ be a binary variable, for $i = 1, \ldots, N$, that takes a value of one if the null hypothesis H_{0F} is rejected and zero, otherwise; N denotes the number of firms and c_{α_d} is the critical value of the test (23) at an α_d significance level. The proposed systemic risk indicator is $S_N^D(q) = \sum_{i=1}^{N} d_i(q)$ that reports the number of institutions for which the null hypothesis H_{0F} is rejected. A large value of this statistic indicates that a significant number of institutions exhibit downside risk.

Under the null hypothesis H_{0F}, the statistic $S_N^D(q)$ is a binomial random variable $Bin(N, \alpha_d)$, with α_d satisfying that $\lim_{T\to\infty} P_{H_{0F}}\{D_{i,T}(q) > c_{\alpha_d}\} \leq \alpha_d$. The corresponding hypothesis test for the presence of systemic risk is

$$\begin{cases} H_{0,R_1} : E[d(q)] \leq \alpha_d, \\ H_{A,R_1} : E[d(q)] > \alpha_d, \end{cases} \tag{27}$$

where the expectation is computed over the cross section of assets. This is a cross- sectional test that is repeated at each point t in the evaluation period. Therefore, rejection of the null hypothesis implies that a significant number of firms are at risk of a downside event in period t. There is an alternative, more traditional, interpretation that suggests that rejection of the null hypothesis provides evidence of misspecification of the risk model for describing the joint tail behavior of the vector of assets. The first interpretation gains relevance if the risk model can be considered a suitable representation of the dynamics of the quantile process in calm periods and only fails during turmoil periods due to an overall increased probability of downside risk.

A suitable test statistic for the hypothesis H_{0,R_1} is $\bar{S}_N^D(q) = \frac{1}{N}\sum_{i=1}^{N} d_i(q)$ that reports the proportion of institutions for which the null hypothesis H_{0F} is rejected. Thus,

Proposition 1. *Let $0 < \alpha_d < 1$ be the significance level of the FSD test (21), and assume that the sequence of test outcomes $\{d_i(q)\}_{i=1}^N$ is iid. Then, under the null hypothesis H_{0,R_1}, the test statistic $\bar{S}_N^D(q)$ satisfies that*

$$\sqrt{N}\left(\frac{\bar{S}_N^D(q) - \alpha_d}{\sqrt{\alpha_d(1-\alpha_d)}}\right) \xrightarrow{d} N(0,1), \tag{28}$$

as $N, T \to \infty$, with $N/T \to 0$.

Proof. The proof of this result follows from applying sequential limit theory, see Phillips and Moon (1999). First, as $T \to \infty$, we obtain the asymptotic convergence of the test statistics $D_{i,T}$ in Equation (23) for $i = 1, \ldots, N$. The FSD test is over-conservative due to the composite inequality constraint characterizing H_{0F}. More formally, under the null hypothesis H_{0F}, we obtain $\lim_{T\to\infty} P\{D_{i,T}(q) > c_{\alpha_d}\} \leq \alpha_d$. In a second step, we note that the sequence $\{d_i(q)\}_{i=1}^N$ is *iid*. This property is inherited from applying the FSD test Equation (18) to the elements of the vector ε_t obtained from the DCC-GARCH model. By construction, this vector is a zero-mean *iid* error term. Therefore, the conditions in the sequential limit theory developed by Phillips and Moon (1999) are

satisfied such that the central limit theory applies to the statistic $\bar{S}_N^D(q)$ for $N \to \infty$, with $N/T \to 0$, and the asymptotic result in Equation (28) is satisfied. □

Exceeding the critical value indicating systemic risk may take time to build up. An alternative, more sensitive to the presence of financial instability, is obtained by aggregating the value of the test statistic $D_{i,T}$ across firms. This systemic risk measure is defined as $V_N^D(q) = \sum_{i=1}^N D_{i,T}(q)$, and the corresponding hypothesis test as

$$\begin{cases} H_{0,R_2} : E[D_T(q)] \leq \eta_G, \\ H_{A,R_2} : E[D_T(q)] > \eta_G, \end{cases} \tag{29}$$

with η_G the expected value of the distribution of the asymptotic process $\sup_{\{x \leq q\}} G_{F_{oq}}(x)$ in Equation (23) obtained under the null hypothesis H_{0F} of no downside risk. The asymptotic distribution of the test $D_T(q)$ is the same for all firms in the cross section if the parameter vector θ is known, implying the same value of η_G across firms. A suitable test statistic for the hypothesis H_{0,R_2} is $\bar{V}_N^D(q) = V_N^D(q)/N$. Intuitively, there is no systemic risk if the average value of the sequence of test statistics $\{D_{i,T}(q)\}_{i=1}^N$ is equal to the mean of the distribution of the asymptotic process characterizing the null hypothesis H_{0F}. In contrast, if the average is significantly larger than η_G then there is evidence of downside risk for some stocks. The following result shows how to choose the critical value that determines statistically the presence of systemic risk under this approach.

Proposition 2. *Let the sequence of test statistics $\{D_{i,T}(q)\}_{i=1}^N$ be iid. Then, under the null hypothesis H_{0,R_2}, the test statistic $\bar{V}_N^D(q)$ satisfies that*

$$\sqrt{N}\left(\frac{\bar{V}_N^D(q) - \eta_G}{\lambda_G}\right) \xrightarrow{d} N(0,1), \tag{30}$$

as $T, N \to \infty$, with $N/T \to 0$; η_G and λ_G are the expected value and standard deviation, respectively, of the distribution of the asymptotic process $\sup_{\{x \leq q\}} G_{F_{oq}}(x)$ defined in Equation (23).

Proof. The proof of this result is obtained by applying sequential limit theory arguments derived in Phillips and Moon (1999). More specifically, for $T \to \infty$, the asymptotic distribution of $D_{i,T}$ is dominated by the distribution of the supremum of a Brownian bridge such that $\lim_{T\to\infty} E[D_{i,T}(q)] \leq \eta_G$ and $\lim_{T\to\infty} V[D_{i,T}(q)] = \lambda_G^2$ for $i = 1, \ldots, N$. In a second step, $N \to \infty$, with $N/T \to 0$, such that the central limit theorem applies to the random sample $\{D_{1,T}(q), \ldots, D_{N,T}(q)\}$ to obtain the asymptotic result in Equation (30). □

3.1 Estimation Effects on Systemic Risk Tests

The asymptotic distribution of the test H_{0F} varies when the parameter vector θ and the functional form of the parametric distribution F_o are estimated. To obtain valid estimates of the corresponding critical values, we apply data-dependent bootstrap procedures. Estimation of the model parameters implies that the critical values of the above tests differ across assets making the method computationally more demanding.

For the first test of systemic risk H_{0,R_1} the test statistic is constructed from $S_N^{D\star}(q) = \sum_{i=1}^N d_i^\star(q)$ with $d_i^\star(q) = 1(\hat{D}_{i,T}(q) > \hat{c}_{i,\alpha_d}^\star)$, where $\hat{c}_{i,\alpha_d}^\star$ is the bootstrap critical value of the

test H_{0F} obtained from a sample $\{y_{it}\}_{t=1}^{T}$ as detailed in Section 2.2. More specifically, $\hat{c}^{\star}_{i,\alpha_d}$ is the $\lceil(1-\alpha_d)B\rceil$ order statistic of the bootstrap sample $\hat{D}^{*(b)}_{i,T}(q)$, where $b=1,\ldots,B$ and B is the number of bootstrap replicas.

Proposition 3. *Let $0<\alpha_d<1$ be the significance level of the FSD test (21), and assume that the sequence of bootstrap test outcomes $\{d^{*}_{i}(q)\}_{i=1}^{N}$ is asymptotically iid for $T\to\infty$. Then, under the null hypothesis H_{0,R_1} and conditional on the vector of realized observations $\{y_{1,t},\ldots,y_{N,t}\}_{t=1}^{T}$, the test statistic $\bar{S}^{D\star}_{N}(q)=S^{D\star}_{N}(q)/N$ satisfies that*

$$\sqrt{N}\left(\frac{\bar{S}^{D\star}_{N}(q)-\alpha_d}{\sqrt{\alpha_d(1-\alpha_d)}}\right)\xrightarrow{d}N(0,1), \tag{31}$$

as $N,T\to\infty$, with $N/T\to 0$.

Proof. As shown in Section 2,2 , the bootstrap critical value is obtained conditional on the available sample and satisfies, under the null hypothesis H_{0F}, that $\lim_{T\to\infty} P\{\hat{D}_{i,T}(q)>\hat{c}^{\star}_{i,\alpha_d}\,|\,\{y_{i,t}\}_{t=1}^{T}\}\leq\alpha_d$, for $i=1,\ldots,N$. Therefore, the indicator function $d^{\star}_{i}(q)$ is such that $\lim_{T\to\infty}E[d^{\star}_{i}(q)\,|\,\{y_{it}\}_{t=1}^{T}]\leq\alpha_d$ for $i=1,\ldots,N$. In a second step, for $N\to\infty$, with $N/T\to 0$, under H_{0,R_1}, and assuming the test outcomes $\{d^{*}_{i}(q)\}_{i=1}^{N}$ are asymptotically *iid* for $T\to\infty$, the central limit theorem can be applied to the sequence to obtain the asymptotic result in Equation (31). □

Similarly, we can derive the asymptotic distribution of the systemic risk test based on the null hypothesis H_{0,R_2} under parameter estimation. The asymptotic distribution characterizing the limiting behavior of the test statistics $D_{i,T}$ is replaced by a bootstrap version. Thus, $\hat{\eta}^{\star}_{iG}$ and $\hat{\lambda}^{\star}_{iG}$ are the expected value and standard deviation of the bootstrap distributions approximating the finite-sample distribution of the test H_{0,R_2} under the null hypothesis.

Proposition 4. *Let the sequence of test statistics $\{\hat{D}_{i,T}(q)\}_{i=1}^{N}$ be asymptotically iid for $T\to\infty$. Then, under the null hypothesis H_{0,R_2} and conditional on the vector of realized observations $\{y_{1t},\ldots,y_{Nt}\}_{t=1}^{T}$, the test statistic $V^{D\star}_{N}(q)=\sum_{i=1}^{N}\hat{D}_{i,T}(q)$, satisfies that*

$$\frac{V^{D\star}_{N}(q)-\sum_{i=1}^{N}\hat{\eta}^{\star}_{iG}}{\sqrt{\sum_{i=1}^{N}\hat{\lambda}^{\star 2}_{iG}}}\xrightarrow{d}N(0,1), \tag{32}$$

as $N,T\to\infty$, with $N/T\to 0$.

Proof. The proof of this result is analogous to the proof of Propositions 2 and 3 and omitted for space considerations. □

4 Monte Carlo Simulation

This section explores the finite-sample properties of the tests for H_{0F}, H_{0,R_1} and H_{0,R_2} proposed above. We study the empirical size and power of the downside risk tests FSD for the univariate case and the extension of these tests to a panel of N firms. Data are simulated from the location-scale process (1). For computational tractability, we consider the

variables to be cross-sectionally independent such that Σ_t is the $N \times N$ identity matrix.[2] Thus, for the simulation exercise, we obtain N independent location-scale processes

$$y_{i,t} = \mu_{i,t}(\theta_1) + \sigma_{ii,t}(\theta_2)\varepsilon_{i,t}, \text{ for } i = 1, \ldots, N, \tag{33}$$

where $\varepsilon_{ii,t} \sim F_{\varepsilon_i}(\cdot)$ is the process of innovations. We consider two models for the data-generating process. First, the processes μ_t and σ_t are constant parameters μ and σ, and the innovations ε_t are assumed to be *iid* and driven from a (*i*) N(0,1) distribution or (*ii*) t_ν, with ν the degrees of freedom of a Student-t distribution. The second model considers the more realistic case of serial dependence in the conditional volatility process. To do this, we simulate an EGARCH(1,1) process $y_{i,t} = \sigma_{ii,t}(\theta_2)\varepsilon_{i,t}$ characterized by the following equation:

$$\sigma^2_{ii,t}(\theta_2) = \beta_0 + \beta_1 \left(\frac{|y_{i,t-1}|}{\sigma_{ii,t-1}} - \sqrt{\frac{2}{\pi}} \right) + \beta_2 \sigma^2_{ii,t-1}(\theta_2) + \gamma \frac{y_{i,t-1}}{\sigma_{ii,t-1}}, \tag{34}$$

and $\varepsilon_{i,t} \sim F_{\varepsilon_i}(\cdot)$ is an *iid* process; $\theta_2 = \{\beta_0, \beta_1, \beta_2, \gamma\}$ denote the model parameters driving the constant, ARCH, GARCH, and leverage effects, respectively, of the conditional volatility process.

As an additional simulation exercise, we explore versions of the above tests that consider the goodness of fit of the proposed risk model $F_o(x)$ in lieu of the FSD hypothesis in Equation (18). These tests replace the inequality condition characterizing FSD by the equality between distribution functions uniformly over the tail domain:

$$\begin{cases} H_{0K} : F_{\varepsilon_i,q}(x) = F_{o,q}(x), \text{ for all } x \leq q, \\ H_{AK} : F_{\varepsilon_i,q}(x) \neq F_{o,q}(x), \text{ for some } x \leq q. \end{cases} \tag{35}$$

A relevant statistic to test the above hypothesis is

$$K_{i,T}(q) = \sqrt{T} || \hat{F}_{\varepsilon_i,q}(x) - F_{o,q}(x) ||_\infty, \tag{36}$$

where $|| \cdot ||_\infty$ is the supremum norm defined over the interval $(-\infty, q]$. Under the null hypothesis H_{0K}, and assuming that the innovations $\varepsilon_{i,t}$ are *iid*, it follows that

$$K_T(q) \xrightarrow{d} || G_{F_{oq}} ||_\infty, \tag{37}$$

see Kolmogorov (1933) and van der Vaart (1998). The critical value c_{α_k} of the test (35), called KS hereafter, is obtained from the condition $P\{||G_{F_{o,q}}||_\infty > c_{\alpha_k}\} = \alpha_k$, with $0 < \alpha_k < 1$ a suitable significance level.

4.1 Simulation Setup

We consider three scenarios for studying the finite-sample properties of the tests. These scenarios are (*i*) model parameters (μ, σ) are known so there are no estimation effects on the asymptotic distribution of the tests; (*ii*) model parameters (μ, σ) are estimated. The latter scenario entails the presence of estimation effects in the relevant asymptotic distributions. In this case, the simulation exercise explores the reliability of the bootstrap approximations; and (*iii*) the conditional volatility process follows an EGARCH model. For (*ii*) and

[2] The empirical application implements the bootstrap method for a fat tailed DCC-GARCH model applied to the 68 stocks of the FTSE-100 index that have remained in the index over the 2000–2022 evaluation period.

Table 1. Empirical size and power for the location-scale DGP $y_t = \mu + \sigma\varepsilon_t$, with $\varepsilon_t \sim F_\varepsilon(\cdot)$; $F_\varepsilon(\cdot) = Z$, with Z a standard Normal distribution, or $F_\varepsilon(\cdot) = t_5$, with t_5 a standardized Student distribution with five degrees of freedom

			FSD test			KS test		
T	F_ε	F_o	θ	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$	θ	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$
$\bar{\tau} = 0.10$								
100	Z	Z	0.064	0.056	0.020	0.068	0.032	0.012
250	Z	Z	0.052	0.052	0.028	0.064	0.020	0.012
500	Z	Z	0.032	0.048	0.016	0.048	0.036	0.012
100	Z	t_5	0.020	0.004	0.012	0.088	0.032	0.024
250	Z	t_5	0.004	0.004	0.004	0.200	0.088	0.084
500	Z	t_5	0.004	0.000	0.000	0.496	0.420	0.472
100	t_5	Z	0.236	0.264	0.200	0.104	0.284	0.184
250	t_5	Z	0.372	0.524	0.432	0.252	0.468	0.292
500	t_5	Z	0.616	0.744	0.584	0.472	0.704	0.568
100	t_5	t_5	0.040	0.048	0.060	0.048	0.108	0.076
250	t_5	t_5	0.064	0.072	0.092	0.060	0.024	0.044
500	t_5	t_5	0.040	0.048	0.056	0.032	0.036	0.040
$\bar{\tau} = 0.05$								
100	Z	Z	0.076	0.060	0.084	0.100	0.068	0.100
250	Z	Z	0.056	0.064	0.072	0.080	0.084	0.104
500	Z	Z	0.084	0.060	0.052	0.052	0.040	0.040
100	Z	t_5	0.080	0.000	0.000	0.384	0.568	0.524
250	Z	t_5	0.048	0.060	0.060	0.652	0.600	0.584
500	Z	t_5	0.020	0.016	0.016	0.748	0.784	0.776
100	t_5	Z	0.192	0.076	0.112	0.084	0.028	0.08
250	t_5	Z	0.332	0.432	0.504	0.228	0.208	0.248
500	t_5	Z	0.728	0.704	0.712	0.576	0.588	0.624
100	t_5	t_5	0.132	0.000	0.000	0.060	0.056	0.036
250	t_5	t_5	0.064	0.028	0.024	0.068	0.056	0.052
500	t_5	t_5	0.064	0.048	0.040	0.068	0.060	0.036

FSD denotes the First-order Stochastic Dominance test H_{0F} and KS denotes the Kolmogorov–Smirnov test H_{0K}. The rejection rates are obtained from 250 simulations. Columns θ report the rejection rates for the simulated test distributions where the model parameters μ and σ are known. Columns $\hat{\theta}_{asy}$ report the tests' rejection rates using simulated critical values from the asymptotic distribution not considering estimation effects. This is corrected in columns $\hat{\theta}_{boo}$ that report the rejection rates using bootstrap critical values obtained from $B = 500$ bootstrap replications; $\bar{\tau}$ is the threshold value that characterizes the domain of the tail of the distribution. T denotes the sample size.

(*iii*) the asymptotic distributions of the FSD tests need to be corrected for the presence of estimation effects when computing suitable critical values

The simulation exercise generates different combinations of the null and alternative hypotheses of the FSD test (H_{0F}) and the KS test (H_{0K}) for scenarios (*i*) and (*ii*). The size of the test H_{0K} is captured in the combinations (Z, Z) and (t_5, t_5), with Z denoting the standard Normal distribution and t_5 a Student-t distribution with five degrees of freedom. The power is reflected by the remaining combinations of F_ε and F_o. For the FSD test, the null hypothesis is given by the inequality $F_\varepsilon(x) \leq F_o(x)$ for $x \leq q = F_o^{-1}(\bar{\tau})$. This condition is naturally satisfied for the combinations (Z, Z) and (t_5, t_5). The null hypothesis for the FSD test also includes the pair (Z, t_5). In contrast, the combination (t_5, Z) is under the alternative hypothesis. As discussed above, the critical value for the FSD test is obtained under the assumption $F_{\varepsilon q}(\cdot) = F_{oq}(\cdot)$ implying an over-conservative test for the combination (Z, t_5) and correct size for the combinations (Z, Z) and (t_5, t_5). The top panel focuses on the tests for the tail coverage probability $\bar{\tau} = 0.10$ and the bottom panel for $\bar{\tau} = 0.05$.

Table 1 reports the empirical size and power for each type of test for different simulation scenarios. There are three columns θ, $\hat{\theta}_{boo}$ and $\hat{\theta}_{asy}$. The first column reports the empirical

Table 2. Empirical size and power for the location-scale DGP $y_t = \sigma_t(\theta_2)\varepsilon_t$, with $\sigma_t(\theta_2)$ an EGARCH process as in Equation (34) with $\beta_0 = 0.05$, $\beta_1 = 0.1$; $\beta_2 = 0.85$ and $\gamma = 0.1$

			FSD test		KS test	
T	F_ε	F_o	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$
$\bar{\tau} = 0.10$						
100	Z	Z	0.064	0.076	0.052	0.076
250	Z	Z	0.040	0.044	0.052	0.056
500	Z	Z	0.072	0.06	0.072	0.072
100	Z	t_5	0.016	0.016	0.116	0.656
250	Z	t_5	0.004	0.000	0.244	0.908
500	Z	t_5	0.000	0.000	0.572	0.996
100	t_5	Z	0.200	0.264	0.192	0.244
250	t_5	Z	0.320	0.320	0.260	0.228
500	t_5	Z	0.656	0.652	0.520	0.500
100	t_5	t_5	0.040	0.080	0.044	0.300
250	t_5	t_5	0.016	0.016	0.064	0.380
500	t_5	t_5	0.040	0.008	0.068	0.536
$\bar{\tau} = 0.05$						
100	Z	Z	0.048	0.052	0.048	0.056
250	Z	Z	0.072	0.072	0.076	0.088
500	Z	Z	0.064	0.076	0.068	0.076
100	Z	t_5	0.004	0.072	0.192	0.640
250	Z	t_5	0.024	0.024	0.288	0.732
500	Z	t_5	0.008	0.016	0.536	0.892
100	t_5	Z	0.228	0.348	0.188	0.276
250	t_5	Z	0.364	0.420	0.296	0.268
500	t_5	Z	0.680	0.648	0.576	0.588
100	t_5	t_5	0.016	0.144	0.040	0.344
250	t_5	t_5	0.052	0.108	0.016	0.336
500	t_5	t_5	0.036	0.040	0.072	0.324

The innovations to the volatility process are $\varepsilon_t \sim F_\varepsilon(\cdot)$, with $F_\varepsilon(\cdot) = Z$, where Z is a standard Normal distribution, or $F_\varepsilon(\cdot) = t_5$, with t_5 a standardized Student-t distribution with five degrees of freedom. FSD denotes the First-order Stochastic Dominance test H_{0F} and KS denotes the Kolmogorov–Smirnov test H_{0K}. The rejection rates are obtained from 250 simulations. Columns $\hat{\theta}_{asy}$ report the tests' rejection rates using simulated critical values from the asymptotic distribution not considering estimation effects. This is corrected in columns $\hat{\theta}_{boo}$ that report the rejection rates using bootstrap critical values obtained from $B = 500$ bootstrap replications; $\bar{\tau}$ is the threshold value that characterizes the domain of the tail of the distribution. T denotes the sample size.

size and power for scenario (i). The second column $\hat{\theta}_{boo}$ reports the results when the parameter vector θ is estimated and the critical values of the test are obtained using bootstrap methods. Column $\hat{\theta}_{asy}$ reports the results when the parameter vector θ is estimated but one erroneously uses the asymptotic critical values as in column θ. In the bootstrap case, each generated sample Y_t entails a different bootstrap critical value that is conditional on the available sample. The empirical size is close to the nominal size across simulation experiments and coverage probabilities. Size improves as the sample size grows. For the FSD test, the empirical rejection rates are close to zero for the pair (Z, t_5), as expected. In contrast, the rejection rates are higher than 0.20 and increase with the sample size for the pair (t_5, Z). Similar results are obtained for the coverage probability $\bar{\tau} = 0.05$. In this case, the empirical power is slightly higher than for $\bar{\tau} = 0.10$. The results for the KS test are also according to theory and similar to the FSD tests. The only significant difference is that the pair (Z, t_5) is now under the alternative hypothesis yielding rejection probabilities around 0.5 for $T = 500$. There are no significant differences in the finite-sample properties of the tests between the bootstrap method $(\hat{\theta}_{boo})$ and the asymptotic one $(\hat{\theta}_{asy})$.

Table 3. Empirical size and power for the location-scale DGP $y_t = \mu + \sigma\varepsilon_t$, with $\varepsilon_t \sim F_\varepsilon(\cdot)$; $F_\varepsilon(\cdot) = N(0,1)$ under H_{0K} and H_{0F} and $F_\varepsilon(\cdot) = t_5$ under H_{AK} and H_{AF}

		FSD test				KS test			
		H_{0,R_1}		H_{0,R_2}		H_{0,S_1}		H_{0,S_2}	
N/T	% H_A	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	$\hat{\theta}_{asy}$
$\bar{\tau} = 0.10$									
30/100	0%	0.072	0.012	0.048	0.004	0.044	0.000	0.024	0.000
	25%	0.432	0.252	0.512	0.332	0.340	0.016	0.400	0.004
	75%	0.96	0.700	0.996	0.992	0.920	0.340	0.984	0.716
	100%	0.996	0.936	1.000	1.000	0.968	0.632	0.996	0.912
60/250	0%	0.076	0.004	0.048	0.012	0.092	0.000	0.044	0.000
	25%	0.972	0.784	0.976	0.92	0.936	0.344	0.948	0.068
	75%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100/500	0%	0.056	0.000	0.052	0.004	0.08	0.000	0.056	0.000
	25%	1.000	1.000	1.000	1.000	1.000	0.992	1.000	0.852
	75%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\bar{\tau} = 0.05$									
30/100	0%	0.04	0.148	0.000	0.112	0.056	0.112	0.000	0.136
	25%	0.092	0.508	0.000	0.52	0.048	0.244	0.000	0.244
	75%	0.264	0.836	0.004	0.968	0.116	0.268	0.000	0.316
	100%	0.364	0.756	0.000	1.000	0.116	0.312	0.000	0.344
60/250	0%	0.088	0.108	0.056	0.108	0.096	0.204	0.048	0.088
	25%	0.764	0.944	0.896	0.984	0.552	0.688	0.648	0.684
	75%	1.000	1.000	1.000	1.000	0.992	0.992	1.000	1.000
	100%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100/500	0%	0.068	0.044	0.032	0.028	0.072	0.044	0.056	0.020
	25%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	75%	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.996
	100%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

T denotes the time series dimension and N the number of variables. Column % H_A denotes the percentage of variables generated under the alternative hypothesis given by t_5. Column $\hat{\theta}_{boo}$ reports the rejection rates using bootstrap critical values obtained from $B = 500$ replications and Column $\hat{\theta}_{asy}$ reports the rejection rates using simulated critical values from the asymptotic distribution assuming θ known. The rejection rates are obtained from 250 simulations. $\bar{\tau}$ is the threshold value that characterizes the domain of the tail of the distribution.

Table 2 reports the empirical size and power for the EGARCH simulation exercise. The simulation setting is the same as the preceding exercise. Estimation effects are, however, more important in this setting because of the heavily parameterized structure of the conditional volatility model. The conditional mean process is zero and the conditional variance is driven by the parameters $\beta_0 = 0.05$, $\beta_1 = 0.1$, $\beta_2 = 0.85$ and $\gamma = 0.1$. The empirical results are similar to the simple location-scale model in Table 1. Importantly, the test based on asymptotic critical values and estimated parameters ($\hat{\theta}_{asy}$) reports very biased estimates of the test size for the pair (t_5, t_5) for both coverage probabilities.

The empirical properties of the tests of systemic risk are studied in Table 3 for different combinations of the null and alternative hypotheses for the above tests. The column $\%H_A$ reports the null and alternative hypotheses considered for the simulation exercise. The data-generating process under the null hypothesis is (Z, Z) and under the alternative is (t_5, Z). Thus, the first case is given by 0% of variables generated from the Student-t distribution with five degrees of freedom. This scenario corresponds to a full cross section of asset returns driven by a Normal distribution. The other three scenarios are given by 25, 75%, and 100% of asset returns generated by a t_5 distribution, respectively, in a setting

Table 4. Empirical size and power for the location-scale DGP $y_t = \sigma_t(\theta_2)\varepsilon_t$, with $\sigma_t(\theta_2)$ an EGARCH process as in Equation (34) with $\beta_0 = 0.05$, $\beta_1 = 0.1$; $\beta_2 = 0.85$ and $\gamma = 0.1$

		FSD test						KS test					
		H_{0,SR_1}			H_{0,SR_2}			H_{0,SR_1}			H_{0,SR_2}		
N/T	% H_A	θ	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	θ	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	θ	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$	θ	$\hat{\theta}_{asy}$	$\hat{\theta}_{boo}$
$\bar{\tau} = 0.10$													
66/200	0%	0.140	0.216	0.050	0.068	0.420	0.015	0.116	0.116	0.045	0.128	0.456	0.030
	25%	0.720	0.720	0.390	0.788	0.836	0.130	0.408	0.456	0.290	0.556	0.828	0.110
	75%	0.980	0.976	0.920	1.000	1.000	0.845	0.948	0.916	0.865	0.996	1.000	0.635
	100%	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.988	1.000	1.000	1.000	1.000
100/500	0%	0.192	0.136	0.090	0.084	0.228	0.050	0.116	0.048	0.090	0.052	0.172	0.010
	25%	0.980	0.956	0.966	1.000	0.984	0.445	0.924	0.836	0.875	0.976	0.916	0.190
	75%	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	0.995
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100/1000	0%	0.076	0.032	0.066	0.052	0.036	0.056	0.072	0.024	0.068	0.108	0.060	0.057
	25%	1.000	1.000	1.000	1.000	1.000	1.000	0.984	0.984	1.000	1.000	0.996	1.000
	75%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100%	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	0.996	1.000
$\bar{\tau} = 0.05$													
66/200	0%	0.236	0.304	0.065	0.148	0.228	0.035	0.424	0.340	0.065	0.392	0.488	0.050
	25%	0.800	0.816	0.420	0.816	0.816	0.325	0.648	0.656	0.290	0.716	0.824	0.225
	75%	1.000	0.992	0.965	0.988	0.992	0.885	0.984	0.972	0.850	0.988	0.992	0.790
	100%	1.000	1.000	1.000	0.992	1.000	1.000	0.992	0.988	1.000	0.992	1.000	1.000
100/500	0%	0.068	0.068	0.110	0.104	0.120	0.010	0.084	0.068	0.140	0.068	0.160	0.025
	25%	0.980	0.968	0.975	0.988	0.964	0.885	0.964	0.944	0.920	0.988	0.956	0.615
	75%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100/1000	0%	0.016	0.012	0.046	0.032	0.012	0.042	0.028	0.012	0.044	0.028	0.020	0.046
	25%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	1.000
	75%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100%	1.000	1.000	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000	0.992	1.000

The innovations to the volatility process are $\varepsilon_t \sim F_\varepsilon(\cdot)$, with $F_\varepsilon(\cdot) = N(0,1)$ or a standardized Student-t_5 distribution. T denotes the time series dimension and N the number of variables. % H_A denotes the percentage of variables generated under the alternative hypothesis given by t_5. Column $\hat{\theta}_{boo}$ reports the rejection rates using bootstrap critical values obtained from $B = 500$ replications and Column $\hat{\theta}_{asy}$ reports the rejection rates using simulated critical values from the asymptotic distribution assuming θ known. The rejection rates are obtained from 250 simulations. $\bar{\tau}$ is the threshold value that characterizes the domain of the tail of the distribution.

where the benchmark distribution is Gaussian. There are two columns for each test; each column reports the test statistic using the bootstrap correction ($\hat{\theta}_{boo}$) and the asymptotic critical values ($\hat{\theta}_{asy}$). The results show strong performance of both tests to reject the null hypothesis of Gaussianity in the tails. Empirical power increases when the proportion of firms with Student-t distributed tails increases, and with the sample size. These results are robust to the choice of critical values. Importantly, though, the rejection rates reflecting the empirical size slightly underestimate the nominal size for the coverage probability $\bar{\tau} = 0.10$ and overestimate it for $\bar{\tau} = 0.05$. Size distortions are less significant as the sample size increases.

Table 4 repeats the simulation exercise for the tests of systemic risk for the EGARCH(1,1) process. The results are very similar to those obtained in Table 3. Both tests report strong power when the proportion of variables driven by a Student-t distribution is as small as 25%. The results improve as the proportion increases to 100%. The power of the test also increases with the sample size N and T, with $N/T \to 0$. We observe significant biases in the empirical size of the tests that use the asymptotic critical values. These deviations of the nominal size are corrected for larger sample sizes and when the bootstrap tests are used instead.

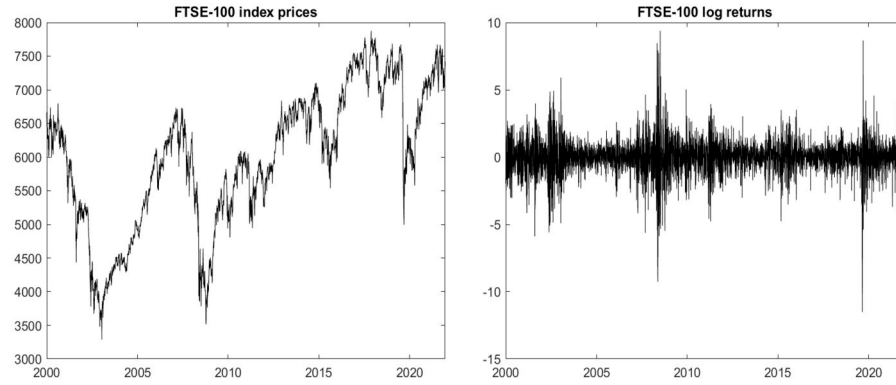


Figure 1. Left panel reports the price dynamics of the FTSE100 index over the period January 4, 2000 to July 22, 2022 (5755 observations). Right panel reports the corresponding log returns over the same period.

5 Empirical Application

The procedures developed above are applied in this section to assess the presence of financial instability to a vector of $N = 68$ stock returns $Y_t = (y_{1t}, \ldots, y_{Nt})'$ containing the constituents of the FTSE-100 index that have remained in the UK financial index during the whole evaluation period from January 2000 to July 2022. The remaining companies have entered and exited the index during the period under consideration. The left panel in Figure 1 presents the evolution of the overall index over the evaluation period and the right panel the dynamics of log returns. There are periods of significant market turbulence and financial distress between 2006 and 2009. Other periods exhibiting strong price fluctuations are 2012–2015 that correspond to the sovereign bond market crisis, the period 2016–2017 corresponding to Brexit, and the period 2020–2021 driven by the outbreak of the COVID-19 pandemic.

Systemic risk is usually interpreted as evidence of market distress spreading over the cross section of firms in the financial system. To monitor the occurrence of this phenomenon, we fit a fat-tailed DCC-GARCH model with Student-t innovations introduced in Equations (3)–(5) to the vector $Y_t = (y_{1t}, \ldots, y_{Nt})'$, with $N = 68$. To compute the probability of tail events we consider a benchmark model given by a multivariate Student-t distribution with five degrees of freedom, which seems a conservative measure of risk in calm periods. The parameters of the DCC model are estimated using the shrinkage estimator proposed in Ledoit and Wolf (2004) although (unreported) results based on the sample covariance matrix yield similar findings. The dynamics of the correlation parameters ψ_1 and ψ_2 obtained from the estimation of the DCC-GARCH model for the vector of 68 stocks are reported in Figure 2. Interestingly, the contribution of past dynamics, given by $\psi_1+\psi_2$, to predict the present correlation is not larger than 0.7 in most periods. This gives more importance to the contribution of the unconditional covariance matrix $\bar{\Sigma}$ in Equation (5), or its correction $\bar{\bar{\Sigma}}$ in Equation (6), for predicting the conditional covariance process.[3]

The aim of this empirical exercise is to assess statistically if this multivariate model is able to capture the underlying tail risk across assets or if there is statistical evidence of

[3] Unreported simulations show that for a smaller number of assets in the cross section ($N = 5, 10, 20$) the estimation of the DCC parameters is $\Psi_1 \in (0.05, 0.10)$ and $\Psi_2 \in (0.85, 0.90)$.

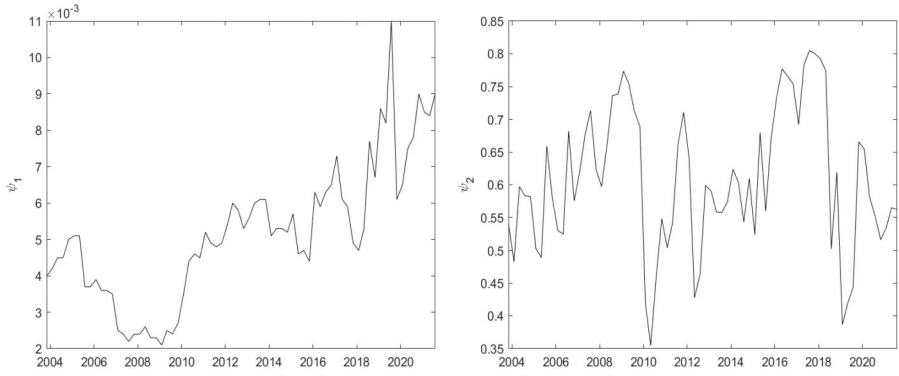


Figure 2. Dynamics of DCC parameters ψ_1 and ψ_2 for DCC-GARCH model (3)–(5) estimated by quasi-maximum likelihood for the vector of 68 stocks remaining in the FTSE-100 index over the period January 4, 2000 to July 22, 2022 (5754 × 68 observations). The estimates are obtained from rolling windows of 66 observations with an in-sample estimation period of $T = 1000$ observations.

systemic risk during specific periods. To do this, we apply the FSD and KS tests to each element of the vector of residuals of the fitted DCC-GARCH model. The second step consists of collecting the outcomes of the individual downside risk tests and the values of the test statistics to implement the systemic risk tests H_{0,R_i}, for $i = 1, 2$. For completeness, we also compute the tests of systemic risk based on aggregating information from the marginal KS tests. The asymptotic theory is not reported but is analogous to the tests H_{0,R_i}. These tests of systemic risk based on the KS statistics are denoted as H_{0,S_i}, for $i = 1, 2$.

The empirical exercise is repeated over rolling windows of 1000 observations moving in time steps of 66 observations corresponding to three months of data, that is, the first window covers the period January 3, 2000 to October 30, 2003; the second window covers the period April 1, 2000 to January 31, 2004 and so on such that the evaluation exercise contains 5754 observations split into 72 rolling windows. The dates in the x axes of Figures 3 and 5 correspond to the terminal dates of each rolling window.

We focus first on testing for systemic risk at moderate tail quantiles given by a coverage probability of $\bar{\tau} = 0.10$. Figure 3 reports the dynamics of the FSD tests $\bar{S}_N^D(q)$ for the null hypothesis H_{0,R_1} and $\bar{V}_N^D(q)$ for the hypothesis H_{0,R_2}, respectively. The results show similar dynamics across systemic risk measures with regards to the presence of periods of financial instability. There is a steady increase in the value of the test statistics from 2006 to 2008, which starts to sharply decline afterward. We should note that the benchmark risk model is given by a fat-tailed DCC-GARCH model with Student-t innovations with five degrees of freedom. The period between 2009 and mid-2012 witnessed a strong drop in the value of the systemic risk statistics. However, from 2013 (2010–2013 evaluation period), these values increase quite sharply and remain high for the rest of the evaluation period. The period 2019–2022 reports low values of the statistics that signal a period of financial stability. The outbreak of the COVID-19 pandemic does not seem to have a strong effect at a $\bar{\tau} = 0.10$ coverage tail probability. Figure 4 reports the evolution of the analogous systemic risk tests based on the marginal KS statistics. In this case, the tests are based on testing the goodness of fit of a multivariate Student-t distribution with five degrees of freedom for the 68 stocks under investigation. As expected, the results are more informative than in the previous case. This is because the tests based on the FSD hypothesis are very conservative whereas the tests based on the goodness of fit use the correct critical values. The dynamics are similar to the patterns observed for the FSD tests; however, the results are more

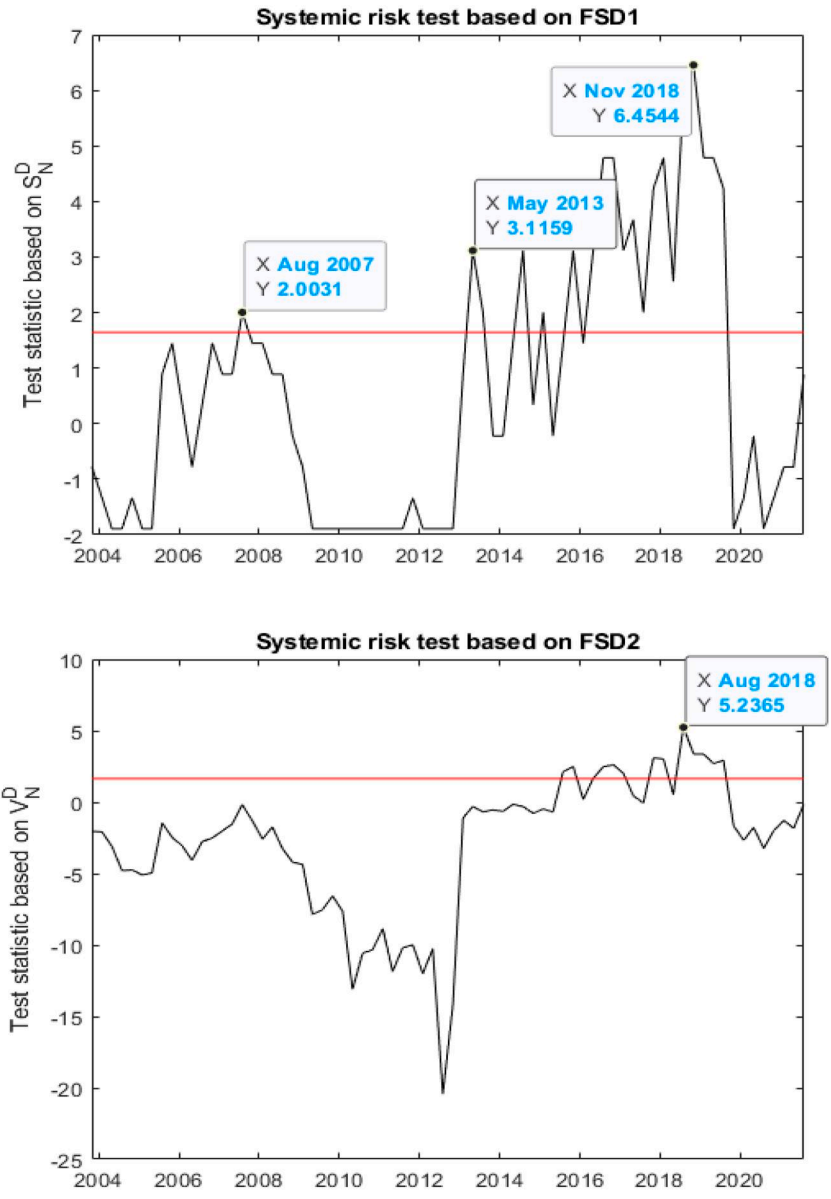
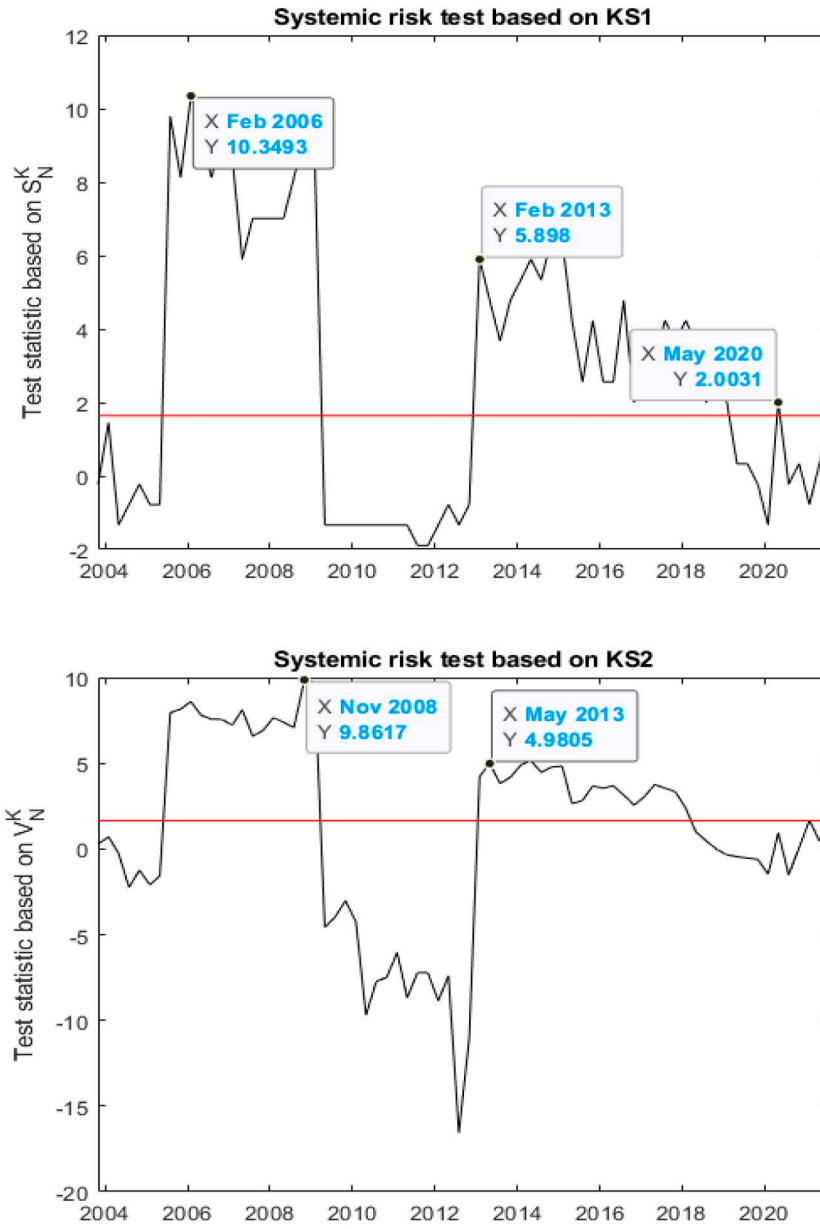


Figure 3. Evolution of systemic risk statistics for coverage probability $\bar{\tau} = 0.10$ and FSD-based test statistics $\bar{S}_N^D(q)$ and $\bar{V}_N^D(q)$. The cross section of stocks is comprised by $N = 68$ assets of the FTSE-100 index computed over 72 rolling windows of $T = 1000$ in-sample observations. The red flat line denotes the Gaussian critical value of the systemic risk tests at an $\alpha = 0.05$ significance level. The evaluation period is October 2003 to July 2022.

statistically significant than in the top two panels. There is strong evidence of systemic risk during the 2006–2009 interval and after 2012. The indicators drop after 2018 but rise again to alert of the presence of financial instability at the start of 2020.



Figures 5 and 6 report the same exercise to detect financial instability and systemic risk for the coverage probability $\bar{\tau} = 0.05$. In this case, the results are more prominent in showing the presence of systemic risk in the extreme left tail during the periods highlighted

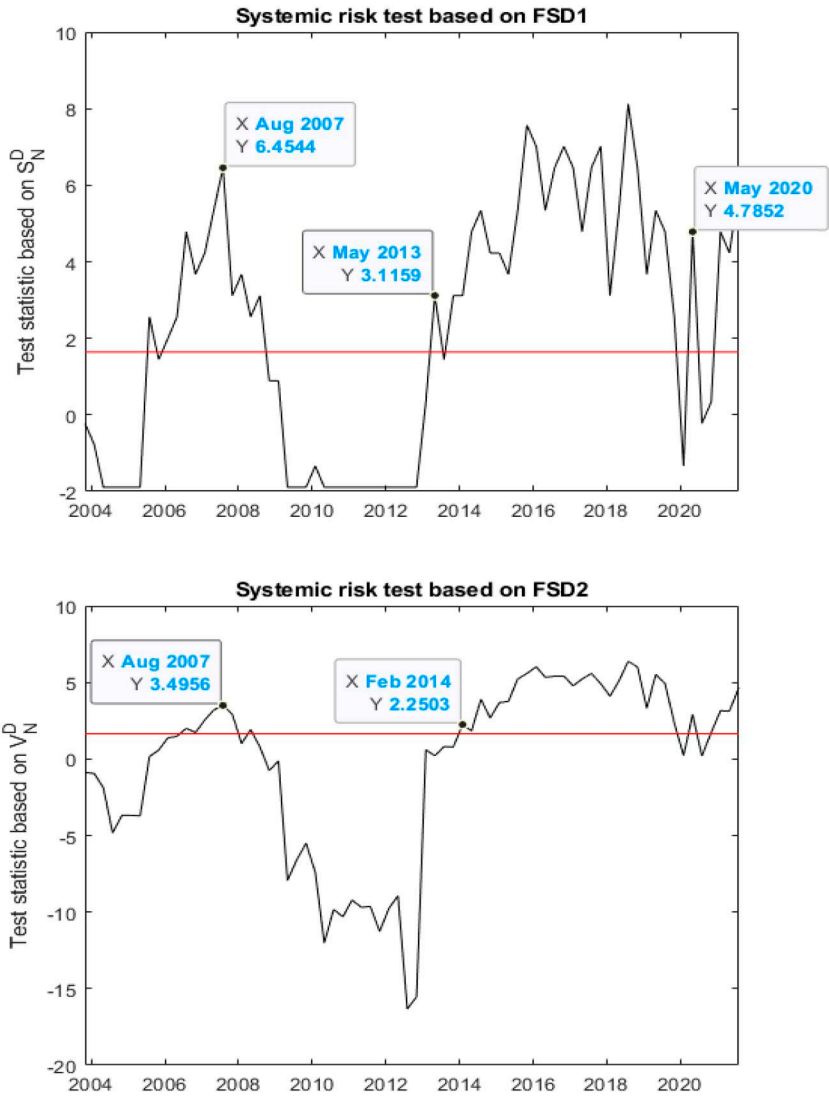


Figure 5. Evolution of systemic risk statistics for coverage probability $\bar{\tau} = 0.05$ and FSD-based test statistics $\bar{S}_N^D(q)$ and $\bar{V}_N^D(q)$. The cross section of stocks comprised $N = 68$ assets of the FTSE-100 index computed over 72 rolling windows of $T = 1000$ in-sample observations. The red flat line denotes the Gaussian critical value of the systemic risk tests at an $\alpha = 0.05$ significance level. The evaluation period is October 2003 to July 2022.

above. The probability of joint extreme losses for $\bar{\tau} = 0.05$ is well above the predictions of the benchmark model, in particular during the 2007–2009 financial crisis and the interval 2013–2018. There is a short period of financial stability until May 2020 that finishes with the outbreak of the COVID-19 pandemic. All of the statistics report similar results and highlight the severity of the 2007–2009 financial crisis but also of several turmoil periods in the UK stock market during the interval 2013–2018.

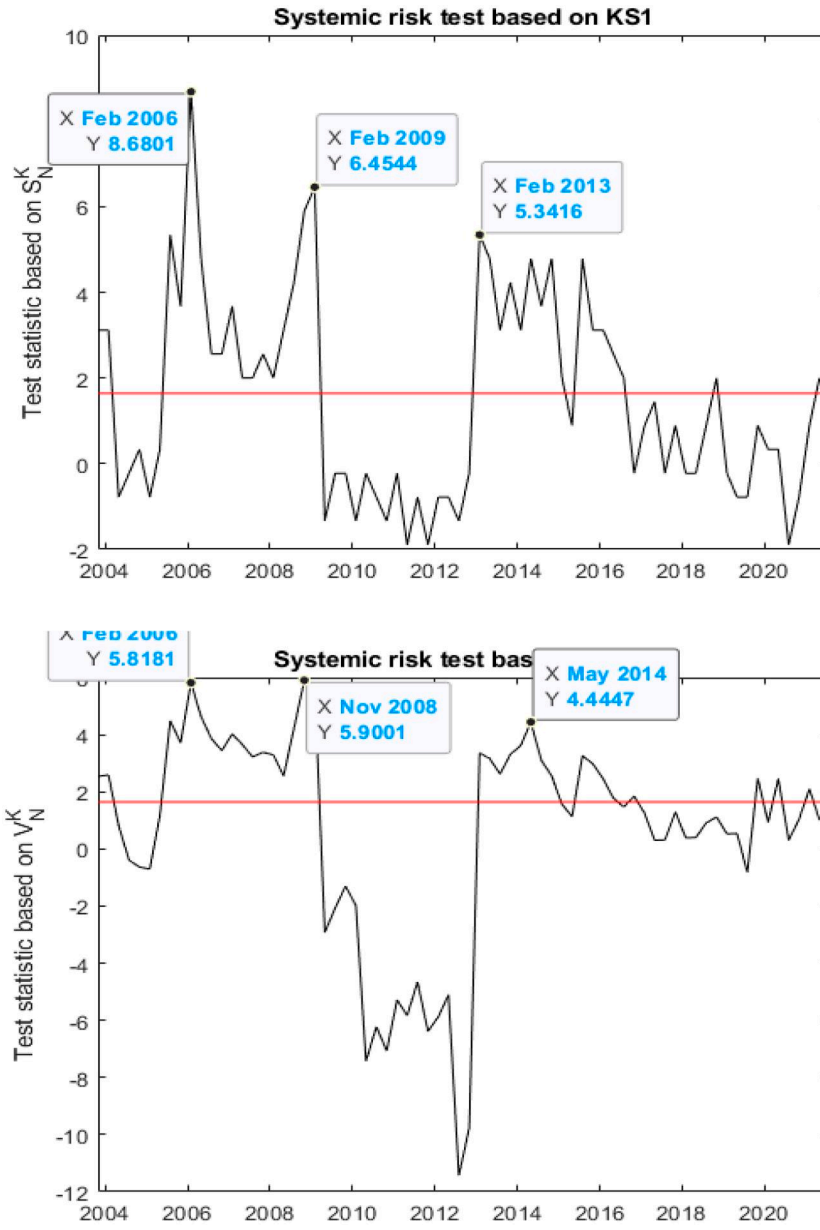


Figure 6. Evolution of systemic risk statistics for coverage probability $\bar{\tau} = 0.05$ and KS-based test statistics introduced in Equation (36). The cross section of stocks comprised $N = 68$ assets of the FTSE-100 index computed over 72 rolling windows of $T = 1000$ in-sample observations. The red flat line denotes the Gaussian critical value of the systemic risk tests at an $\alpha = 0.05$ significance level. The evaluation period is October 2003 to July 2022.

6 Conclusion

The purpose of this study is to measure the probability of systemic risk using information from the cross section of stock returns. Financial instability is defined as the occurrence of simultaneous downside risk for a large proportion of firms in the financial system. This definition is made operational through a bootstrap test of FSD that corrects for estimation of the risk model parameters. Rejection of the null hypothesis implies that the downside probability of the underlying process driving financial returns is larger than the downside probability of the proposed model for, at least, some values of the tail domain. Using a procedure similar in spirit to statistical meta-analysis, the outcomes of these individual tests are aggregated to construct tests of financial instability or systemic risk.

These methods are illustrated by fitting a fat-tailed DCC-GARCH model with Student-t innovations applied to daily data for a cross section of 68 stock returns containing the constituents of the FTSE-100 index that have remained in the index during the evaluation period 2000–2022. The results obtained from the analysis evaluated at different tail coverage probabilities provide overwhelming evidence of financial instability in the UK stock market during the period 2006–2009, 2012–2018, and 2020–2021 due to the occurrence of different crises. The sensitivity of the different proposed tests to detect systemic risk is more pronounced for tests based on goodness of fit hypotheses than on FSD and also depends on the coverage probability characterizing the tail events.

Funding

There is no funding specific to this article.

Conflicts of interest

The authors declare no conflict of personal, financial or other interest.

References

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson. 2017. Measuring Systemic Risk. *Review of Financial Studies* 30: 2–47.

Adrian, T., and M. K. Brunnermeier. 2016. CoVaR. *American Economic Review* 106: 1705–1741.

Alexander, L. 2010. "Opening Remarks," working paper, Measuring Systemic Risk: A Conference Sponsored by the Milton Friedman Institute, the Chicago Fed, and the New York Fed.

Allen, L., T. G. Bali, and Y. Tang. 2012. Does Systemic Risk in the Financial Sector Predict Future Economic Downturns? *Review of Financial Studies* 25: 3000–3036.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–228.

Banulescu-Radu, D., C. Hurlin, J. Leymarie, and O. Scaillet. 2021. Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures. *Management Science* 67: 5730–5754.

Barrett, G., and S. Donald. 2003. Consistent Tests for Stochastic Dominance. *Econometrica* 71: 71–104.

Brownlees, C., and R. F. Engle. 2016. Srisk: A Conditional Capital Shortfall Measure of Systemic Risk. *Review of Financial Studies* 30: 48–79.

Brunnermeier, M. K., A. Crockett, C. A. Goodhart, A. D. Persaud, and H. S. Shin. 2009. *The Fundamental Principles of Financial Regulation.* Geneva, Switzerland: International Center for Monetary and Banking Studies.

Chan-Lau, J. A, M. Espinosa, K. Giesecke, and J. A. Sole. 2009. "Assessing the Systemic Implications of Financial Linkages." *Global Financial Stabiliy Report.* Apr 09:73–110, IMF, Washington, DC.

Christoffersen, P. 1998. Evaluating Interval Forecasts. *International Economic Review* 39: 841–862.

Colletaz, G., C. Hurlin, and C. Pérignon. 2013. The Risk Map: A New Tool for Validating Risk Models. *Journal of Banking & Finance* 37: 3843–3854.

Couperier, O., and J. Leymarie. 2020. "Backtesting Expected Shortfall via Multi-Quantile Regression." *Working Paper Halshs-01909375v4f*, ENSAE Paris, France.
Creal, D. D., and J. C. Wu. 2015. Estimation of Affine Term Structure Models with Spanned or Unspanned Stochastic Volatility. *Journal of Econometrics* 185: 60–81.
Creal, D. D., and J. C. Wu. 2017. Monetary Policy Uncertainty and Economic Fluctuations. *International Economic Review* 58: 1317–1354.
Davidson, R., and J. Y. Duclos. 2000. Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica* 68: 1435–1464.
Delgado, M. A, and J. C. Escanciano. 2013. Conditional Stochastic Dominance Testing. *Journal of Business & Economic Statistics* 31: 16–28.
Du, Z., and J. C. Escanciano. 2017. Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science* 63: 940–958.
Engle, R. F. 2002. Dynamic Conditional Correlation – a Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–350.
Engle, R. F., and S. Manganelli. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics* 22: 367–381.
Engle, R. F., O. Ledoit, and M. Wolf. 2019. Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics* 37: 363–375.
Escanciano, J. C., and J. Olmo. 2010. Backtesting Parametric Value-at-Risk with Estimation Risk. *Journal of Business & Economic Statistics* 28: 36–51.
Fissler, T., and Y. Hoga. 2023. Backtesting Systemic Risk Forecasts Using Multi-Objective Elicitability. *Journal of Business & Economic Statistics* 1. Available at https://doi.org/10.1080/07350015.2023.2200514
Goncalves, S., U. Hounyo, A. J. Patton, and K. Sheppard. 2023. Bootstrapping Two-Stage Quasi-Maximum Likelihood Estimators of Time Series Models. *Journal of Business & Economic Statistics* 41: 683–694.
Gonzalo, J., and J. Olmo. 2014. Conditional Stochastic Dominance Tests in Dynamic Settings. *International Economic Review* 55: 819–838.
Huang, X., H. Zhou, and H. Zhu. 2009. Assessing the Systemic Risk of a Heterogeneous Portfolio of Banks during the Recent Financial Crisis. *Bank of International Settlements Working Paper No 296*: 1–44.
Hu, X., J. Pan, and J. Wang. 2013. Noise as Information for Illiquidity. *The Journal of Finance* 68: 2341–2382.
Hurlin, C., S. Laurent, R. Quaedvlieg, and S. Smeekes. 2017. Risk Measure Inference. *Journal of Business & Economic Statistics* 35: 499–512.
Fisher, R. A. 1932. *Statistical Methods for Research Workers*. 4th edn. Edinburgh: Oliver and Boyd.
Khandani, A. E., and A. W. Lo. 2011. What Happened to the Quants in August 2007: Evidence from Factors and Transactions Data. *Journal of Financial Markets* 14: 1–46.
Kolmogorov, A. N. 1933. *Foundations of Probability Theory*. Berlin. Germany: Julius Springer.
Kratz, M., Y. Lok, and A. McNeil. 2018. Multinomial VaR Backtests: A Simple Implicit Approach to Backtesting Expected Shortfall. *Journal of Banking & Finance* 88: 393–407.
Kritzman, M., and Y. Li. 2010. Skulls, Financial Turbulence, and Risk Management. *Financial Analysts Journal* 66: 30–41.
Kritzman, M, Y. Li, S. Page, and R. Rigobon. 2010. "Principal components as a measure of systemic risk." Working Paper Series Financial Economics, Revere Str, 272–28.
Kupiec, P. 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives* 3: 73–84.
Kupiec, P., and L. Guntay. 2016. Testing for Systemic Risk Using Stock Returns. *Journal of Financial Services Research* 49: 203–227.
Leccadito, A., S. Boffelli, and G. Urga. 2014. Evaluating the Accuracy of Value-at-Risk Forecasts: New Multilevel Tests. *International Journal of Forecasting* 30: 206–216.
Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10: 603–621.
Ledoit, O., and M. Wolf. 2004. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88: 365–411.

Linton, O., E. Maasoumi, and Y. J. Whang. 2005. Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies* 72: 735–765.
Linton, O., K. Song, and Y. J. Whang. 2010. An Improved Bootstrap Test of Stochastic Dominance. *Journal of Econometrics* 154: 186–202.
Nelson, D. B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59: 347–370.
Patton, A. J., J. F. Ziegel, and R. Chen. 2019. Dynamics Semiparametric Models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics* 211: 388–413.
Pérignon, C, and D. R. Smith. 2008. A New Approach to Comparing VaR Estimation Methods. *The Journal of Derivatives* 16: 54–66.
Phillips, P. C. B., and H. R. Moon. 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67: 1057–1111.
Rockafellar, R. T., and S. Uryasev. 2002. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking & Finance* 26: 1443–1471.
Topaloglou, N., H. Vladimirou, and S. A. Zenios. 2002. Cvar Models with Selective Hedging for International Asset Allocation. *Journal of Banking & Finance* 26: 1535–1561.
van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press.
White, H., T. H. Kim, and S. Manganelli. 2015. VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles. *Journal of Econometrics* 187: 169–188.
Wied, D., G. Weiß, and D. Ziggel. 2016. Evaluating Value-at-Risk Forecasts: A New Set of Multivariate Backtests. *Journal of Banking & Finance* 72: 121–132.

Journal of Financial Econometrics, 2024, 00, 1–29
https://doi.org/10.1093/jjfinec/nbae005
Article