# Risk Measure Inference[1]
# This Version: October 25, 2013

Christophe Hurlin[a], Sébastien Laurent[b], Rogier Quaedvlieg[c,*], Stephan Smeekes[d]

[a]*Laboratoire d'Economie d'Orléans (LEO), University of Orléans, France*
[b]*IAE Aix en Provence and Greqam, France*
[c]*Department of Finance, Maastricht University, Netherlands*
[d]*Department of Quantitative Economics, Maastricht University, Netherlands*

**Abstract**

We propose a widely applicable bootstrap based test of the null hypothesis of equality of two firms' Risk Measures (RMs) at a single point in time. The test can be applied to any market-based measure. In an iterative procedure, we can identify a complete grouped ranking of the RMs, with particular application to finding *buckets* of firms of equal systemic risk. An extensive Monte Carlo Simulation shows desirable properties. We provide an application on a sample of 94 U.S. financial institutions using the ΔCoVaR, MES and %SRISK, and conclude only the %SRISK can be estimated with enough precision to allow for a meaningful ranking.

*Keywords:* Grouped Ranking, Risk Measures, Bootstrap, Uncertainty.

## 1. Introduction

Financial risk management is fundamentally based on the comparison of a given risk measure for different assets, portfolios or financial institutions. Many examples can be cited: the comparison of the market risk for two portfolios, measured by their volatilities, their value-at-risk (VaR) or their expected shortfall (ES), the comparison of the systematic risk of two assets measured by their beta, the comparison of the systemic risk of two financial institutions according to a systemic risk measure or a score of systemic importance (Basel Committee on Banking Supervision, 2013) and

[1]This research was financially supported by a grant of the Dutch Organization for Scientific Research (NWO). We would like to thank participants at Cluster Risque Financiers, Orléans, the Liège-Luxembourg-Maastricht PhD Workshop and the 7th International MIFN workshop and specifically Eric Beutner and Sylvain Benoit for fruitful discussions.

*Corresponding author: Department of Finance, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.

*Email addresses:* christophe.hurlin@univ-orleans.fr (Christophe Hurlin), sebastien.laurent@iae-aix.com (Sébastien Laurent), r.quaedvlieg@maastrichtuniversity.nl (Rogier Quaedvlieg), s.smeekes@maastrichtuniversity.nl (Stephan Smeekes)

many others. When one considers an unconditional risk measure, this comparison can be achieved through parametric or non-parametric tests. These tests exist for a variety of standard (unconditional) risk measures; the most famous being tests for equal variances. However, most of the risk measures are expressed conditional on an information set available at time $t$ and the corresponding forecasts are generally issued from a dynamic (semi-)parametric model. For instance, we can cite the conditional VaR and ES forecasts derived from a (M-)GARCH model or the dynamic conditional beta (Engle, 2012). In these cases, the conditional distribution of the estimated risk measure is generally unknown and depends on the actual estimation procedure used for the model parameters.

In this paper, we propose a general testing framework that takes into account parameter uncertainty to statistically test for equality of conditional risk measures for different financial units (assets, portfolios or financial institutions). We propose two types of tests. First, we propose a bootstrap-based test of equality for a general class of conditional risk-measures at a single point in time, or over a longer horizon of forecasts. This test can be applied to a wide class of risk measures, estimated by almost any (semi-)parametric model. For example, it can be used to compare the market risk measured by a conditional VaR, ES or volatility for two portfolios or two banks. It can be used to test the relative level of systemic risk for two banks, to compare the systematic risk of two portfolios or the liquidity risk of two different assets. It can also be applied in order to test the equality of a conditional risk measure (for instance the VaR) issued from two different models (say, GARCH and RiskMetrics) for the same asset or the same portfolio.

Second, we propose a generalized procedure aiming at grouping units that are statistically indistinguishable in terms of their riskiness. The method is inspired by the Model Confidence Set of Hansen et al. (2011). This bucketing procedure can be applied to any type of risk measure but has particular application for systemic risk measures.

The financial crisis has renewed the interest in measuring systemic risk and specifically in the identification of the global systemically important banks (G-SIBs).[2] As they pose a major threat to the financial system, regulators and policy makers from around the world have called for tighter supervision, extra capital requirements, and liquidity buffers for these G-SIBs (Financial Stability Board, 2011). However, the task of measuring the degree of systemic risk and determining how much capital is needed for each financial institution is a difficult one. Even though the Basel Committee proposes a score of systemic risk for the banks, it recognizes the inevitable uncertainty in the measure and decided to classify banks into *buckets*, groups of banks representing roughly equal risk (BCBS, 2013).

Meanwhile, academics have been less cautious. Many measures of systemic risk have been proposed over the past years, the most well-known being the *Marginal Expected Shortfall* (MES) and the *Systemic Expected Shortfall* (SES) of Acharya et al. (2010), the *Systemic Risk Measure* (SRISK) of Acharya et al. (2012) and Brownlees

[2] Also called systemically important financial institutions (SIFIs).

and Engle (2011), and the *Delta Conditional Value-at-Risk* ($\Delta$CoVaR) of Adrian and Brunnermeier (2011). They all try to summarize systemic risk into one single number. This has a lot of appeal as there exists an automatic ordering of the financial institutions according to their systemic risk. Generally, the authors then proceed to produce exactly such a list, implying their measures can be used to obtain an exact ranking in real time (with a daily or weekly frequency) that can be compared to the list of G-SIB published once a year by the BCBS. However, claiming that firm A is more risky than firm B, simply because its systemic risk measure is marginally higher, implies that they are estimated with perfect accuracy. This is certainly not the case. These measures rely on publicly available market data (stock returns, option prices, or CDS spreads) supposed to reflect all information about publicly traded firms and require relatively sophisticated estimation techniques, such that even if the model is correctly specified, there is still a lot of parameter uncertainty. Indeed, there is convincing evidence that the signal produced by the systemic risk measures are not reliable and have large uncertainty (e.g. Danielsson et al., 2011). If this is taken into account it is unlikely that one can discern such an absolute ranking.

We are aware of only one other paper that statistically compares firms in terms of systemic riskiness. Castro and Ferrari (2011) propose a method within the linear quantile regression framework of testing whether or not two firms differ in terms of their $\Delta$CoVaR. Our method is more general; it works for any market-based measure, and allows for conditional, or time-varying, versions of the risk measures.

The paper is structured as follows: Section 2 first gives some definitions and defines the general framework. Section 3 establishes hypotheses and the test. Section 4 discusses the bootstrap implementation. Section 5 gives simulation results, showing the procedure has desirable properties: appropriate size and high power. Finally, Section 6 provides an empirical application on 94 of the largest financial institutions in the United States.

## 2. Methodology

Our aim is to propose a general framework to statistically test for equality of two conditional risk measures obtained for two different assets, portfolios or financial institutions. Consider a financial unit (firm, asset or portfolio) indexed by $i$ and a $\mathcal{F}_{i,t-1}$-conditional risk measure (denoted RM) issued from a dynamic parametric model, where $\mathcal{F}_{i,t-1}$ denotes the information set available at time $t-1$. Formally, we define RM as follows:

$$RM_{i,t} = f_i\left(\theta_i, \omega; X_{i,t-1}\right), \tag{1}$$

where $f_i\left(.\right)$ denotes a functional form that depends on $(i)$ the risk measure itself (for instance, the VaR) and $(ii)$ the parametric model used to produce the corresponding forecast (for instance, a GARCH model). $X_{i,t-1}$ is a set of variables belonging to $\mathcal{F}_{i,t-1}$, $\theta_i$ is the vector of model's parameters and $\omega$ is a vector of parameters specific to the risk measure itself. The latter parameters are generally determined by the user. For instance, in the case of the VaR, it corresponds to the risk level, generally fixed to 1% or 5% by convention.

The notation for $RM_{i,t}$ encompasses a wide class of parametric risk measures associated to asset's and portfolio's profit and losses (P&L) or to a financial firm (in the case of the systemic risk analysis). The RM can for instance be a measure of price variation (conditional volatility), a systematic risk measure (beta), a tail risk measure (VaR, ES), or a systemic risk measure (MES, SRISK, ΔCoVaR). The model can be a univariate or a multivariate GARCH model, a quantile or a linear regression model, etc. Hence, this notation can be viewed as a generalization of that used by Gourieroux and Zakoïan (2013) for parametric VaR models.[3]

RM is indexed by $t$, implying we exclude the test of equality of unconditional risk measures. This is not very restrictive as these tests exist for a variety of standard unconditional risk measures; e.g. a test for equal variances. The index $i$ in $f_i(.)$ means that we allow for different functional forms for financial units $i$ and $j$. The risk measure will obviously be common to the two units, but one may use different models to produce the forecasts. For instance, the procedure allows the comparison of the conditional VaR for Bank of America obtained from a GARCH model, and the conditional VaR of Citibank using an internal model based on RiskMetrics. Similarly, one could also test for equality of two different models applied to the same unit.

As examples of the notation we consider (*i*) a conditional VaR based on a student GARCH model, (*ii*) the ES based on a Gaussian GJR-GARCH, (*iii)* the conditional MES of Acharya et al. (2010) and Brownlees and Engle (2011) and (*iv*) the SRISK of Acharya et al. (2012) and Brownlees and Engle (2011). The last two cases correspond to well-known systemic risk measures based on a GARCH-cDCC and non-parametric tail-estimators.

**Example 1 (VaR-GARCH)** *Consider a demeaned return process $r_{i,t}$ associated to an asset or a portfolio indexed by i. If we assume a t-GARCH(1,1) model for demeaned returns $r_{i,t}$, the corresponding conditional VaR($\alpha$) can be expressed as a linear function of the conditional volatility $\sigma_{i,t|t-1}$ of the returns as follows:*

$$f_i^{VaR}\left(\theta_i, \omega; X_{i,t-1}\right) = -t_\nu^{-1}(\alpha)\sqrt{\frac{\upsilon-2}{\upsilon}}\sigma_{i,t|t-1},$$

*with $\sigma^2_{i,t|t-1} = \gamma_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}$, with $\epsilon_{i,t} = r_{i,t}/\sigma_{i,t}$ and $t_\nu^{-1}(\alpha)$ is the $\alpha-$quantile of the standardised student cdf with $\nu$ degrees of freedom. As such $\theta_i = \{\gamma_i, \alpha_i, \beta_i, \nu\}$, $\omega = \{\alpha\}$ and $X_{i,t-1} = \{\tilde{\boldsymbol{r}}_{\boldsymbol{i,t-1}}\}$, where $\tilde{\boldsymbol{r}}_{\boldsymbol{i,t-1}}$ is the set of return observations for firm i up to time $t-1$.*

[3] In the univariate case, assuming a conditional normal distribution, they define the conditional VaR as $VaR_t\left(\alpha\right) = -g\left[y_t; \theta, \Phi^{-1}\left(\alpha\right)\right]$ where $y_t$ denotes the P&L at time $t$, $\theta$ the vector of model's parameters and $g\left(.\right)$ is a functional form that depends on the model.

**Example 2 (ES-AR-GJR-GARCH)** *Consider an AR(1)-GJR-GARCH(1,1) model for demeaned returns $r_{i,t}$. The conditional ES($\alpha$) is then defined as follows:*

$$f_i^{ES}(\theta_i, \omega; X_{i,t-1}) = -\mu_{i,t} + \lambda\left(\Phi^{-1}(1-\alpha)\right)\sigma_{i,t|t-1},$$

*with $\mu_{i,t} = \phi_i r_{i,t-1}$ and $\sigma^2_{i,t|t-1} = \gamma_i + \alpha_{1i}\epsilon^2_{i,t-1} + \alpha_{2i} I_{\{\epsilon_{i,t-1}<0\}}\epsilon^2_{i,t-1} + \beta_i\sigma^2_{i,t-1}$, where $\epsilon_{i,t} = (r_{i,t} - \mu_{i,t})/\sigma_{i,t}$. $\lambda(z) = \phi(z)/\Phi(z)$ corresponds to the Mills ratio where $\phi(.)$ and $\Phi(.)$ denote respectively the pdf and the cdf of the standard normal distribution. Then we get $\theta_i = \{\gamma_i, \alpha_{1i}, \alpha_{2i}, \beta_i\}$, $\omega = \{\alpha\}$ and $X_{i,t-1} = \{\tilde{\boldsymbol{r}}_{\boldsymbol{i,t-1}}\}$.*

**Example 3 (MES)** *The MES is the marginal contribution of an institution i to systemic risk, as measured by the ES of the financial system. The market return is denoted $r_{m,t} = \sum_{i=1}^{N} w_{i,t} r_{i,t}$, with $w_{i,t}$ the value-weight of firm i at time t, and $r_{i,t}$ demeaned firm returns. The conditional MES is defined by the first derivative $-\partial\mathbb{E}_{t-1}(r_{m,t} \mid r_{m,t} < C)/\partial w_{i,t}$, where C is a threshold (generally equal to the market VaR). In order to estimate the MES, following Brownlees and Engle (2011), if the vectorial process $(r_{i,t}\ r_{m,t})'$ follows a GARCH-DCC we get:*

$$\begin{aligned} f_i^{MES}(\theta_i, \omega; X_{i,t-1}) &= \sigma_{i,t|t-1}\rho_{im,t|t-1}\mathbb{E}_{t-1}(\epsilon_{m,t}|\epsilon_{m,t} < C/\sigma_{m,t|t-1}) \\ &\quad + \sigma_{i,t|t-1}\sqrt{1-\rho^2_{im,t|t-1}}\,\mathbb{E}_{t-1}(\epsilon_{i,t}|\epsilon_{m,t} < C/\sigma_{m,t|t-1}), \end{aligned}$$

*where $\sigma^2_{i,t|t-1} = \gamma_i + \alpha_i\epsilon^2_{i,t-1} + \beta_i\sigma^2_{i,t-1}$, $\rho_{im,t|t-1} = Q_{im,t|t-1}/\sqrt{Q_{ii,t|t-1}Q_{mm,t|t-1}}$ with $Q_{ij,t}$ the $(i,j)^{th}$ element of $Q_t$, and $Q_{t|t-1} = (1-\alpha_C-\beta_C)\bar{Q} + \alpha_C\epsilon_{t-1}\epsilon'_{t-1} + \beta_C Q_{t-1}$, with $\epsilon_{i,t} = r_{i,t}/\sigma_{i,t}$. Brownlees and Engle (2011) consider a non-parametric estimator (Scaillet, 2005) for the tail expectations of the standardized returns $\epsilon_t$. Then we have $\theta_i = \{\gamma_i, \gamma_m, \alpha_i, \alpha_m, \beta_i, \beta_m, \bar{Q}, \alpha_C, \beta_C\}$, $\omega = \{C\}$ and $X_{i,t-1} = \{\tilde{\boldsymbol{r}}_{\boldsymbol{i,t-1}}, \tilde{\boldsymbol{r}}_{\boldsymbol{m,t-1}}\}$.*

**Example 4 (SRISK)** *The SRISK is defined as the expected capital shortfall of a given financial institution, conditional on a crisis affecting the whole financial system. The daily conditional SRISK can be defined as follows:*

$$f_i^{SRISK}(\theta_i, \omega; X_{i,t-1}) = \max\left(0\ ;\ k\ D_{i,t-1} - (1-k)\,W_{i,t-1}\left(1 - LRMES_{i,t|t-1}\right)\right),$$

*where $D_{i,t}$ and $W_{i,t}$ denote the book value of total liabilities and the market value of the financial institution respectively. k is a prudential capital ratio and $LRMES_{i,t|t-1} = 1 - exp(-18MES_{i,t|t-1})$ denotes the conditional marginal expected shortfall, extrapolated to a full-blown crisis over half a year (see Brownlees and Engle (2011)). $MES_{i,t|t-1}$ is the estimate of the Marginal Expected Shortfall for firm i at time t as defined in Example 3, which can be estimated through $f_i^{MES}(\theta_i, \omega; X_{i,t-1})$. Then we have $\theta_i = \{\gamma_i, \gamma_m, \alpha_i, \alpha_m, \beta_i, \beta_m, \bar{Q}, \alpha_C, \beta_C\}$, $\omega = \{C, k\}$ and $X_{i,t-1} = \{\tilde{\boldsymbol{r}}_{\boldsymbol{i,t-1}}, \tilde{\boldsymbol{r}}_{\boldsymbol{m,t-1}}, D_{i,t-1}, W_{i,t-1}\}$.*

## 3. Hypotheses of Interest and Test

In this section, we present two types of tests for the conditional risk measures: (i) a comparison test of two risk measures and (ii) a bucketing procedure.

### *3.1. A Comparison Test of Risk Measures*

We wish to test whether two financial units, indexed by $i$ and $j$, present the same level of risk at time $t$ with respect to the conditional risk measure RM. We assume that the considered risk measure satisfies the axiom of monotonicity (Artzner et al., 1997, 1999), *i.e.,* if the unit $i$ is strictly riskier than the unit $j$, then $RM_{i,t} < RM_{j,t}$. To fix the notations, let us define $\mathcal{F}_{t-1} = \mathcal{F}_{i,t-1} \cup \mathcal{F}_{j,t-1}$ the information set available at time $t-1$ for the two units and a $\mathcal{F}_{t-1}$-conditional relative riskiness variable denoted $x_{ij,t}$ with:

$$x_{ij,t} \equiv RM_{i,t} - RM_{j,t}. \tag{2}$$

The null hypothesis is:

$$H_0 : \mathbb{E}\left(x_{ij,t} \mid \mathcal{F}_{t-1}\right) = 0. \tag{3}$$

The alternative hypothesis can be expressed as $H_1 : \mathbb{E}\left(x_{ij,t} \mid \mathcal{F}_{t-1}\right) \neq 0$ meaning that the risk of financial unit $i$ is different from the risk of unit $j$, at time $t$ according to the measure RM. The testing framework is clearly very general as it works for any specification of $f_i()$. It can be applied to any conditional (monotonic) risk measure and is relevant in many practical cases in finance.

The similarity to the literature devoted to the comparison of forecasts (Diebold and Mariano, 1995; Hansen, 2005) is clear. But, although there are many parallels, our hypothesis and test differ in some important ways. In most cases, $i$ and $j$ do not represent competing models, but they represent different financial units (assets, portfolios, financial institutions etc.). Second, we do not compare the forecast to an *ex-post* observation. Finally, and most importantly, we test for equality of the conditional expectation of two single forecasts, i.e. the conditional risk measures at time $t$, not means of series.

In the case where our test is used to compare the forecasts of the same risk measure issued from two alternative models, there are also some similarities with the literature devoted to the volatility forecasting comparison (Hansen and Lunde, 2006; Patton, 2011). But the two approaches differ in some important ways. First, our test can be applied to any conditional risk measure and is not specific to the conditional volatility. Second, our test is not designed to determine the 'best' model: the test investigates whether there is enough information in the data to claim that two models are producing significantly different risk forecasts for a single observation. Third, our approach does not require the use of a proxy variable for the risk measure. Note that for most risk measures (except volatility), such proxy variables are not available yet.

Finally, note that the proposed framework can be extended to test the equality of risk measure forecasts for a horizon $h > 1$, either as single estimate, or as aggregate $\sum_{i=0}^{h-1} RM_{i,t-h} - RM_{j,t-h}$, by considering a smaller information set $\mathcal{F}_{t-h} = \mathcal{F}_{i,t-h} \cup \mathcal{F}_{j,t-h}$.

Testing the hypothesis described in (3) is challenging in many respects. First, the conditional distribution of the estimated risk measure $\widehat{RM}_{i,t} = f_i\left(\widehat{\theta}_i, \omega; X_{i,t-1}\right)$ is generally unknown or may be difficult to obtain depending on the actual estimation procedure used for $\widehat{\theta}_i$.[4] The estimates are typically obtained using (M)GARCH-type models, for many of which their estimates' distribution is unknown. Second, even if the conditional distributions of the both estimators $\widehat{RM}_{i,t}$ and $\widehat{RM}_{j,t}$ are known, the distribution of the difference $\widehat{x}_{ij,t}$ is generally unknown because the joint distribution is rarely known (except in the trivial but unlikely case of independence between $\widehat{\theta}_i$ and $\widehat{\theta}_j$). This is clearly the case when we compare two RM forecasts issued from two alternative models for the same asset. Francq and Zakoïan (2012) offer a very specific case where we have a result for GARCH processes. However, a delta method is needed to go from $\theta$ to the $RM$, and there is no result for the possible dependence between $\theta_i$ and $\theta_j$.

One could circumvent these problems by relying on a central limit theorem and showing asymptotic normality of the test statistic under the null. This could be done by taking time averages of $\hat{x}_{ij,t}$, but this is not appropriate in our case. The values of the measures change a lot over time, and moreover, there is no reason to believe the relative ordering (for the true values of $RM_{i,t}$ and $RM_{j,t}$) is constant over time. We therefore only have the single forecast for $\widehat{x}_{ij,t}$.

These restrictions mean that we cannot use traditional testing methods. Instead, we use the assumed data generating process (DGP) to bootstrap the conditional risk measures and obtain their distribution. We therefore propose the following two-sided test statistic:

$$T(\alpha) \equiv \frac{|\hat{x}_{ij,t}|}{c^*_{ij,t}(\alpha)}, \tag{4}$$

where $c^*_{ij,t}(\alpha)$ is the bootstrap critical value obtained from the absolute null-value shifted bootstrap distribution of $\hat{x}_{ij,t}$. The use of the critical value means that the $\alpha\%$ rejection point for all combinations of financial units $(i, j)$ is scaled to 1. Rejection thus occurs at the $\alpha\%$ level if $T(\alpha) > 1$. Ex-post one may draw conclusions on which is the riskiest based on the sign of $x_{ij,t}$[5].

We work under the assumption that the bootstrap is asymptotically valid for the risk measures considered, in the sense that the bootstrap correctly reproduces the asymptotic distribution of the risk measure estimator. While this validity cannot be established in general due to the absence of results about the asymptotic distributions of the estimators of the parameters $\theta_i$, we expect the bootstrap to be valid for "well-behaved" estimators. In particular, if the estimators are $\sqrt{T}$-consistent and asymptotically normal, as they are for the class of risk measures considered by Francq and

[4] We assume that the parameters $\omega$ specific to the risk measure are predetermined.

[5] The absolute value is not necessary for the single test, and a general version $T(\alpha) = \hat{x}_{ij,t}/c^*_{ij,t}(\alpha)$ may be considered, taking into account that the left and right tail of the distribution may differ. The absolute version will however facilitate the procedure in the multiple comparison setting outlined in the next section.

Zakoïan (2012), we conjecture the bootstrap to be asymptotically valid. Our simulation study in Section 5 supports this conjecture.

### *3.2. Bucketing Procedure*

An increasingly important aspect of risk management is the measurement of systemic risk (see Bisias et al., 2012, for a survey). The general idea of this literature is that the financial institutions contributing the most to the risk of the financial system (the G-SIBs or SIFIs) will be subject to more intense supervision and will have to maintain more regulatory capital. In practice, there are two ways to measure the contribution of a given firm to the overall risk of the system. A first approach, called the Supervisory Approach, relies on firm-specific information on size, leverage, liquidity, interconnectedness, complexity, and substitutability. In this specific context, the latest draft released by the Basel Committee on Banking Supervision (BCBS, 2013) recommends to classify the financial institutions according to a score of systemic importance. Based on their systemic risk score, banks are allocated into five buckets that correspond to specific higher loss absorbency requirement.[6]

A second approach relies on publicly available market data, such as stock returns, option prices, or CDS spreads, as they are believed to reflect all information about publicly traded firms. Four prominent examples of such measures are the *Marginal Expected Shortfall* (MES) and the *Systemic Expected Shortfall* (SES) of Acharya et al. (2010), the *Systemic Risk Measure* (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2011), and the *Delta Conditional Value-at-Risk* ($\Delta$CoVaR) of Adrian and Brunnermeier (2011). These measures are based on sophisticated econometric models and estimation methods designed to measure the dependence between firm and market returns. Even if these models are correctly specified, there is still a lot of parameter uncertainty. For this reason, the market-based approach requires a procedure that allows to classify banks into buckets, groups of banks representing roughly equal risk, as it is done in the BCBS's approach.

For these reasons, we propose an iterative bucketing procedure that can be used to obtain a grouped ranking, or buckets, of financial units. The objective is to get a complete ranking by means of a procedure inspired by the Model Confidence Set of Hansen et al. (2011). Our procedure produces buckets of equally risky units, in the sense that we cannot statistically distinguish the units within one bucket in terms of their riskiness.[7] This testing procedure can be applied to any type of monotonic risk measure (market risk, liquidity risk, etc.), but it has particular application in the context of the systemic risk.

[6]There are four populated equally sized buckets (1 to 4), with an additional empty bucket (5) with a higher loss absorbency requirement of 3.5% of risk-weighted assets to provide an incentive against banks further increasing their systemic importance. Any bank with a score lower than the cutoff level of the first bucket is classified as non-systemically important and has no additional capital charge.

[7]Thereby we acknowledge that it is near impossible to estimate with such precision to be able to determine an absolute ranking of riskiness. While we may observe large differences in the very top of the ranking, the level of risk in the middle of the field may be very small with indistinguishable differences.

We now describe the testing procedure. Consider the set of all financial units $\mathcal{N}^0$. We start with the identification of the set of most risky units, defined at time $t$ as

$$\mathcal{N}_t^{(1)} \equiv \{i \in \mathcal{N}^0 : \mathbb{E}\left(x_{ij,t} \mid \mathcal{F}_{t-1}\right) \geq 0 \text{ for all } j \in \mathcal{N}^0\}. \tag{5}$$

The goal is to find the set $\mathcal{N}_t^{(1)}$. This is achieved through a sequence of significance tests where objects in $\mathcal{N}^0$ are removed from the set under consideration if they are found to be less risky. The null we want to test is therefore

$$H_{0,\mathcal{N}} : \mathbb{E}\left(x_{ij,t} \mid \mathcal{F}_{t-1}\right) = 0 \text{ for all } i, j \in \mathcal{N}, \tag{6}$$

with $\mathcal{N} \subseteq \mathcal{N}^0$, the subset containing the not yet eliminated firms. The null hypothesis states that all financial units in the final set after the elimination procedure should be equally risky. For any set $\mathcal{N}$ this can be tested using an equivalence test and an elimination rule. If the equivalence test is rejected, we use the elimination rule to remove the most significantly different unit, reducing the size of $\mathcal{N}$, and re-apply the equivalence test. Our set of most risky units is the subset of $\mathcal{N}^0$ that contains $\mathcal{N}_t^{(1)}$ with a certain probability which can be controlled. This procedure identifies the most risky set only. To obtain the full ranking, we apply the procedure on the set $\mathcal{N}^0 \setminus \hat{\mathcal{N}}_t^{(1)}$ to obtain a second bucket, $\hat{\mathcal{N}}_t^{(2)}$. This is repeated until all units have been allocated to a bucket.

In order to carry out the procedure we need an equivalence test and an elimination rule. In case of equivalence we have that $\mathbb{E}\left(x_{ij,t} \mid \mathcal{F}_{t-1}\right) = 0$ for all $i, j \in \mathcal{N}$. We propose the following test statistic

$$T^{max}(\alpha) \equiv \max_{i,j \in \mathcal{N}} \frac{|\hat{x}_{ij,t}|}{c^*_{ij,t}(\alpha)}. \tag{7}$$

Here the need for standardization becomes evident, as we want to identify the firm which is most likely to be different from the rest. If there is a significant difference, an elimination rule follows naturally. We eliminate the unit $\arg\max_{j\in\mathcal{N}} \sup_{i\in\mathcal{N}} \hat{x}_{ij,t}/c^*_{ij,t}(\alpha)$, or to put it simply, the most significantly rejected financial unit. By taking the absolute value of $\hat{x}_{ij,t}$ we do not have to concern ourselves with the comparison of positive or negative critical values. Our method is equivalent to having the non-absolute version of the test statistic, and selecting the maximum absolute $T(\alpha)$ in the elimination rule.

Given a rejection point of 1, and independence of the different statistics, this test will have size $1 - (1 - \alpha)^{N(N-1)/2}$. Independence is however unlikely, which is another reason to use the bootstrap approach. To control the size at $\alpha$, we estimate the critical value of $T^{max}$, denoted $d^*_t(\alpha)$, in the same bootstrap procedure. This way we obtain our estimated buckets, which we denote $\mathcal{N}_{1-\alpha}^{(k)}$, $k = 1, 2, ...$, where each bucket contains the true set of most risky firms in $\mathcal{N}$ with probability $1 - \alpha$.

*3.3. Procedure Implications*

Of course, there are many different ways to obtain buckets of equally risky financial units, and even rank them. However, the implications of our procedure are ideally suited to ranking systemic firms.

First, the approach is one directional, which means we only control the size, the Type I error of the null of equal risk, in one direction as well. By using a top-down approach, we are controlling the probability of falsely removing a firm from the set of firms under consideration. A false rejection leads to a firm being assigned to a less risky cluster in the next iteration. Under-estimating the risk is in our opinion much more hazardous than the reverse, and therefore this is controlled. Moreover, if a false rejection occurs, the procedure has a self-correcting mechanism that minimizes the effect on the relative ranking of the remaining firms. The firm that is falsely rejected, is by definition the most risky firm in the set of firms not yet assigned to a bucket. As such, with enough power, it will be assigned to a bucket on its own. Even though it is estimated as less risky than it ought to be, it is still deemed more risky than all the remaining, less risky firms.

Second there is the Type II error; failing to eliminate a firm, and assigning it to a too risky bucket. In practice, what might happen is that a firm with a low point estimate but a high standard error may be assigned to a riskier bucket then a firm with a higher point estimate, but a low standard error. In some sense, these firms are *loose cannons*. Their series have characteristics that make it difficult to estimate their true risk with accuracy. Again, due to the top-down approach, the resulting ranking will be prudent; in case of large uncertainty, a firm is always put in the most risky bucket.

Finally, we want to emphasize the number of buckets is not specified ex-ante. This is the main difference with the approach proposed by the BCBS. It can be anywhere between one and the number of firms under consideration, depending on the precision of the estimates. It therefore automatically strikes a balance between compression and accuracy of the ranking.

## 4. Bootstrap Implementation

This section describes how to obtain $c^*_{ij,t}$ and $d^*_t$ for all $i, j$ and a general time of interest $T$, for the broad class of RMs. Here, we assume a general DGP, $r_t = g(\theta, \epsilon_t|\mathcal{F}_{t-1})$, with $r_t$, $\epsilon_t$ vectors of dimension $N$, and $\theta$ the set of model parameters. Here we assume $\epsilon_t$ to be iid with means zero and covariance matrix equal to the identity matrix. For instance, for the returns in Example 1, $r_t = g(\theta, \epsilon_t|\mathcal{F}_{t-1}) = \sigma_{t|t-1}\epsilon_t$, where $\sigma_{t|t-1}$ follows a GARCH with parameters $\theta$. We define the inverse, $\epsilon_t = g^{-1}(\theta, r_t|\mathcal{F}_{t-1})$, which retrieves the residuals from the observed process. Hence, for the same example, $\epsilon_t = g^{-1}(\theta, r_t|\mathcal{F}_{t-1}) = r_t/\sigma_{t|t-1}$.

We employ the general concept of the methodology suggested for GARCH forecasts by Pascual et al. (2006). In general, their approach is as follows. First estimate $\theta_i$ on the original series $r$, i.e. $r_t$ for $t = 1, ..., T-1$. Generate bootstrap series, $r^*$, using $\hat{\theta}_i$, and innovations drawn with replacement from the empirical distribution of the centred residuals. Estimate the same model on the bootstrap series, to obtain $\hat{\theta}_i^*$. The bootstrap RM forecast, $f_{i,t}^*$ is computed based on the true series $r$ and bootstrap parameter estimates $\hat{\theta}^*$.

The original series are used for forecasting such that the state of the variable at the time of the forecast is taken into account for the forecast. The bootstrapping is therefore only used to measure the parameter uncertainty. Forecasts over longer horizons can be achieved by simulating a return path based on draws from the empirical distribution. For our setting this amounts to the following algorithm

*Bootstrap Algorithm.*

1. Estimate the models to obtain $\hat{\theta}$. Use the parameter estimates to forecast $\hat{x}_{ij,t}$.
2. Compute the residuals $\hat{\epsilon}_t = g^{-1}(\hat{\theta}, r_t|\mathcal{F}_{t-1})$ for all $t = 1, ..., T-1$.
3. Draw $\tau_1, ..., \tau_{T-1}$ i.i.d. from the uniform$(1, T-1)$ distribution and construct the bootstrap errors from the centred residuals $\epsilon_t^{*b} = \hat{\epsilon}_{\tau_t}$, $\forall\ t = 1, ..., T-1$.
4. Construct the bootstrap return series. That is, $r_t^{*b} = g(\hat{\theta}, \epsilon_t^{*b}|\mathcal{F}_{t-1})$.
5. Estimate the model on the bootstrapped series to obtain $\hat{\theta}^{*b}$. Compute $RM_{i,T}^{*b}$ using $f_i(\hat{\theta}_i^{*b}; X_i)$, and similarly for $RM_{j,T}^{*b}$ to obtain $x_{ij,T}^{*b}$.
6. Repeat steps 3 to 5 $B$ times, obtaining bootstrap statistics $x_{ij,T}^{*b}$, $b = 1, ..., B$. Calculate the bootstrap critical value $c_{ij,T}^*(\alpha)$ as the $\alpha-$quantile of the ordered null-value shifted series $|x_{ij,T}^{*b} - \hat{x}_{ij,T}|$. Similarly, obtain $d_T^*$ as the $\alpha-$quantile of $T^{max,*b} - T^{max}$.

The time-concordant sampling we propose in Step 3 ensures that the possible tail-dependence between firms is preserved. We want to emphasize that the conditioning variables $X_i$ are not bootstrapped for the bootstrap forecasts. This is how we take in the recent state of the series at the relevant time point.
To illustrate how to obtain the critical values in Step 6 consider the following table.

| b/(i,j) | (1,2) | (1,3) | ... | (N-1,N) | |
|---|---|---|---|---|---|
| 1 | $x_{12}^{*1}$ | $x_{13}^{*1}$ | ... | $x_{N-1,N}^{*1}$ | $max_{(i,j)}T^{*1} = \lvert x_{ij}^{*1}\rvert / c_{ij}^*$ |
| 2 | $x_{12}^{*2}$ | $x_{13}^{*2}$ | ... | $x_{N-1,N}^{*2}$ | $max_{(ij)}T^{*2} = \lvert x_{ij}^{*2}\rvert / c_{ij}^*$ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| B | $x_{12}^{*B}$ | $x_{13}^{*B}$ | ... | $x_{N-1,N}^{*B}$ | $max_{(i,j)}T^{*B} = \lvert x_{ij}^{*B}\rvert / c_{ij}^*$ |
| | $c_{12}^*$ | $c_{13}^*$ | ... | $c_{N-1,N}^*$ | $d^*$ |

Given all the $x_{ij}^*$ the first step consists of obtaining the critical values $c_{ij}^*$, which are used to compute $T^{max,*}$ for every bootstrap sample. This delivers the distribution of $T^{max}$ and $d^*$ is its centred $\alpha$-quantile.

## 5. Simulation Study

We use Monte Carlo Simulations to study the properties of both the single test and the Bucketing procedure. The Monte Carlo Simulation is performed on 1,000 replications. For the bootstraps we generate $B = 999$ samples. We always forecast the conditional risk measure at time $T$ and estimate on the sample 1 to $T - 1$. We apply the single test to a tail-risk measure, the $VaR$, and both the single test and the bucketing procedure to the systemic risk measure $MES$, as defined in Examples 1 and 3 respectively.[8]

### 5.1. VaR

We apply the single test to the VaR estimated at the 5% level. For the simulation we consider the following DGP

$$\begin{aligned} r_{i,t} &= \sigma_{i,t}\epsilon_{i,t} \\ \epsilon_{i,t} &\overset{\text{iid}}{\sim} ST(0, 1, \nu_i) \\ i &= 1, 2, \end{aligned} \tag{8}$$

where $\sigma^2_{i,t}$ follows a GARCH(1,1) with parameters $(\gamma, \alpha_1, \beta_1) = (0.05, 0.10, 0.85)$ for both series. The innovations follow a student distribution with mean zero, variance one and degrees of freedom $\nu_i$. We need to be able to simulate under the null, where $VaR_{1,T} = VaR_{2,T} \iff t^{-1}_{\nu_1}\sqrt{(\nu_1 - 2)/\nu_1}\sigma_{1,T} = t^{-1}_{\nu_2}\sqrt{(\nu_2 - 2)/\nu_2}\sigma_{2,T}$. To impose this equality, we simulate processes, and re-scale the series ex-post such that the VaRs are equal. See Appendix A for more details on the re-scaling. We consider two cases, i.e. both firms have equal or different degrees of freedom. In the former case the volatility at time $t$ is equal for both firms, in the latter case the volatility will be higher for the firm with higher degrees of freedom.

For the equal degrees of freedom case, we set $\nu_1 = \nu_2 = 5$. We set $\sigma_{1,T} = 2$ and define $\sigma_{2,T}$ relative to that. Define $\Delta\sigma = \sigma_{2,T} - \sigma_{1,T}$. We use $\Delta\sigma = \{0.0, 0.5, 1.0\}$ to simulate under the null hypothesis and local alternatives. In the case of different degrees of freedom we set $\nu_1 = 5, \nu_2 = 7$. Again $\sigma_{1,T} = 2$. We scale $\sigma_{2,T}$ such that the $VaR_{2,T}$ has the same value as in the previous case (with equal degrees of freedom), i.e. $\sigma_{2,T} = \frac{t^{-1}_5}{t^{-1}_7}\sqrt{21/25}(\sigma_{1,T} + \Delta\sigma)$. Consequently, the VaRs are equal when $\Delta\sigma = 0$, while the first series has a higher volatility and the second has a higher degree of freedom.

The results are reported in Table 1. We set the nominal size of the test to 5% and consider two different sample sizes, $T = 1,000$ and $T = 2,000$. Results suggest that the test does not suffer from serious size distortions. Indeed, the rejection frequencies reported in column $\Delta = 0$ are very close to the nominal size of 5% in all cases, even for 1,000 observations. As expected, power increases with the distance between the VaRs and reaches almost one for both DGPs when $\Delta = 1$. It seems to be slightly higher in the case where there is a difference in degrees of freedom of the innovation distributions.

[8] For the estimation of the conditional volatility models we use the G@RCH-package for OxMetrics (Laurent, 2013).

Table 1: Rejection frequencies of the single test of equal VaR

| | T=1,000 | | | T=2,000 | | |
|---|---|---|---|---|---|---|
| $\nu_2 \backslash \Delta\sigma$ | 0.0 | 0.5 | 1.0 | 0.0 | 0.5 | 1.0 |
| 5 | 0.045 | 0.669 | 0.966 | 0.049 | 0.838 | 0.994 |
| 7 | 0.052 | 0.721 | 0.976 | 0.050 | 0.864 | 0.998 |

*Note:* The table contains the rejection rates of a single test of equal VaR. Nominal size is 5%.

### *5.2. MES*

For the MES we consider the general DGP as proposed by Brownlees and Engle (2011), i.e.

$$\begin{aligned} r_{m,t} &= \sigma_{m,t}\epsilon_{m,t} \\ r_{i,t} &= \sigma_{i,t}\left(\rho_{i,t}\epsilon_{m,t} + \sqrt{1-\rho_{i,t}^2}\xi_{i,t}\right) \\ (\epsilon_{m,t}, \xi_{i,t}) &\sim F, \end{aligned} \tag{9}$$

where $\sigma_{m,t}$ and $\sigma_{i,t}$ follow GARCH models, while $\rho_{i,t}$ follow a cDCC as described in Example (3). $F$ is a general zero mean, unit variance distribution, with unspecified dependence structures. Under the assumption that the innovations are i.i.d. and all dependence between firms and the market is captured by the correlation, Benoit et al. (2012) show that the MES can be written as:

$$MES_{i,t}(\alpha) = \beta_{i,t} ES_{m,t}(\alpha), \tag{10}$$

where $\beta_{i,t} = cov(r_{i,t}, r_{m,t})/var(r_{m,t}) = \rho_{i,t}\sigma_{i,t}/\sigma_{m,t}$, and $ES_{m,t}$ is as in Example 2.

The Expected Shortfall has a closed form expression for several innovation distributions, such as the Normal and Student $t_\nu$. For instance, in the Gaussian case, letting $\phi(.)$ and $\Phi(.)$ denote the standard normal univariate pdf and cdf respectively, the MES can be written as follows:

$$\begin{aligned} MES_{i,t}(\alpha) &= \beta_{i,t}\sigma_{m,t}\lambda(\Phi^{-1}(\alpha)) \\ &= \rho_{i,t}\sigma_{i,t}\lambda(\Phi^{-1}(\alpha)), \end{aligned} \tag{11}$$

where $\lambda(z) = \phi(z)/\Phi(z)$ denotes the Mills ratio. Given a distribution, the MES therefore solely depends on the volatility of the firm and its correlation with the market. Under the assumption of normality, two firms have equal MES if the product of conditional volatilities and conditional correlation with the market at time $t$ is equal. We use this result to control the relative risk of simulated firms.

To keep our simulations computationally feasible, we make a simplifying assumption of constant correlations. The model reduces to a CCC (Bollerslev, 1990). As there are no time-varying correlations we save time in our simulations by estimating the correlation by its sample counterpart.[9] Compared to Equation (9), we set $\rho_{i,t} = \rho_i$. Of course, this assumption will be relaxed in the application.

Next to the constant correlation, another difference with (9) is that innovations are assumed i.i.d. standard normal. That is, again comparing to Equation (9), $F = N(0, I_2)$, the bivariate normal distribution. GARCH models are estimated by maximum likelihood and the tail expectation using the non-parametric estimator of Scaillet (2005) as in Brownlees and Engle (2011).

GARCH parameters $(\gamma, \alpha_1, \beta_1)'$ are set to $(0.05, 0.10, 0, 85)'$ for both models and to simulate under the null and the alternative we employ the same trick used in Section 5.1 to rescale the volatilities for time $T$.

For the single test we generate two firms and the market. The market has $\sigma_{m,T} = 1$. The first firm has $\sigma_{1,T} = 2$ and $\rho_1 = 0.5$. We vary the volatility and correlation of the second firm. Let us define $\Delta\sigma = \sigma_{2,T} - \sigma_{1,T}$ and $\Delta\rho = \rho_2 - \rho_1$. The distance between the MES of firms 1 and 2 is therefore a function of the parameters $(\Delta\sigma, \Delta\rho)$. For instance, setting $(\Delta\sigma, \Delta\rho) = (0.5, 0.05)$ results in $MES_{1,T} = 2.063$ and $MES_{2,T} = 2.673$. We choose $\Delta\sigma = \{0, 0.5, 1\}$ and $\Delta\rho = \{0.00, 0.05, 0.10\}$.

For the bucketing procedure we generate the market and $N$ firms. In order to obtain firms that satisfy the null hypothesis of equal systemic risk, we give all firms within the same bucket identical variance and correlation. The number of simulated buckets is set to $c = 5$. The market again has $\sigma_{m,T} = 1$. All firms $i$ in bucket 1 have $\sigma_{i,T}^{(1)} = 2, \rho_i^{(1)} = 0.5$. All firms $i$ in bucket $k = 2, ..., c$ have $\sigma_{i,T}^{(k)} = 2 + (k-1)\Delta\sigma, \rho_i^{(k)} = 0.5 + (k-1)\Delta\rho$. The difference between two successive buckets in terms of their volatility and correlation is therefore equal to that between the two firms in the single test of equal MES. We also take the same values for $\{\Delta\sigma, \Delta\rho\}$. We simulate $N = 20$ with $c = 5$ buckets of 4 firms each. This means that the difference between the most and least risky cluster is $4\Delta\sigma$ and $4\Delta\rho$.

#### *5.2.1. Single test*

Table 2 reports the rejection frequencies of the null hypothesis of equal MES for $T = 1,000$ and $2,000$ at the 5% significance level. The size of the test corresponds to $\Delta\sigma = \Delta\rho = 0$, and is close to the nominal value. There appear to be no size distortions. The other entries in the table correspond to power. Like in the test of equal VaR presented in Section 5.1, power increases rapidly with the distance between the MES and reaches values close to 100% when $\Delta = 1$ For $T = 1,000$, the power is close to 70% when the difference in correlation is only 0.05 and volatility is 0.5 higher. When $T = 2,000$ the power is over 90% for the same parameters.

[9]We have performed the simulations using cDCC correlations for a few parameter settings with a small number of replications and found very similar results.

Table 2: Rejection frequencies of the single test of equal MES

| | T=1,000 | | | T=2,000 | | |
|---|---|---|---|---|---|---|
| $\Delta\rho\backslash\Delta\sigma$ | 0.0 | 0.5 | 1.0 | 0.0 | 0.5 | 1.0 |
| 0.00 | 0.046 | 0.346 | 0.835 | 0.044 | 0.605 | 0.967 |
| 0.05 | 0.078 | 0.687 | 0.944 | 0.117 | 0.915 | 0.998 |
| 0.10 | 0.285 | 0.895 | 0.984 | 0.530 | 0.988 | 1.000 |

*Note:* The table contains the rejection rates of a single test of equal MES. Nominal size is 5%.

*5.2.2. Bucketing Procedure*

In order to save space, we only report the results for $T = 2,000$ and choose a significance level of $\alpha = 10\%$. This means that we expect each bucket to contain the most risky firms in the remaining set with 90% probability.

It is difficult to evaluate the bucketing procedure in terms of size and power. This is mainly due to the fact that an error in any of the iterations has an impact on the next steps. Indeed, the composition of the second bucket will be affected by the composition of the first one, and so on. Moreover, we may overestimate the number of buckets if, for instance, the first bucket is split up into two separate buckets, such that the third estimated bucket is in fact the second bucket implied by the DGP. Therefore, we do not expect to always have a one-to-one correspondence between the generated ranking and the estimated ranking.

Despite these problems, we can comment on two types of relevant statistics. We can compute typical size and power statistics of the first estimated bucket, as they are not affected by previous iterations. Second, we can talk about relative ranking. If one firm is more risky than another according to the DGP, it has to be allocated to a higher bucket.

When only looking at the first bucket, size and power are meaningful. The test is defined in such a way that the estimated first bucket contains the true first bucket with probability $1-\alpha$. Hence, the size for the first iteration can be defined as the fraction of firms in the true first bucket that are included in the estimated first bucket. Formally, we define the size as the frequency at which $\mathcal{N}^{(1)} \subset \hat{\mathcal{N}}^{(1)}_{90\%}$.

Power cannot be explicitly defined for the procedure. The power should relate to the extent to which we are able to eliminate firms that are less risky. Only four firms belong to the first bucket. That means the remaining sixteen have to be eliminated. The power is one when the sixteen least risky firms are eliminated. We therefore report the fraction of these firms not assigned to the first cluster.

Results are reported in Panel A of Table 3. The frequency at which the most risky firms are contained in the first bucket is very close to $(1-\alpha)$, even in the case when $\Delta\sigma = \Delta\rho = 0$, and all 20 firms are in the first bucket. For the second statistic, the power increases when the distance between buckets becomes greater. Overall, the procedure has quite high power, rapidly increasing to one with reasonable distance to the null.

Panel B of Table 3 concerns relative rankings. Again, we consider two different statistics. For the first one, labelled 'Pairwise relative ranking correct', we check whether the relative ranking is correct. If firm $i$ is more risky than firm $j$, it has to be allocated to an earlier bucket. If they are equally risky they need to be in the same bucket. This measure is in our opinion appropriately severe to the aforementioned problem with one bucket being estimated as two separate buckets. The measure will deteriorate as these two buckets should be one bucket, but both buckets are still ranked higher than the next bucket, keeping the relative ranking intact.

The second statistic we consider, labelled 'Top 10 correctly identified', has great practical relevance. The only firms we really care about are the most risky. We therefore compute the frequency at which the top ten of estimated firms only contains ten of the most risky firms. Note that the number ten has no real significance in our simulation. The top ten should therefore contain only firms out of the top twelve of our DGP.

Table 3 suggests that both statistics go to one; when the firms are sufficiently far from each other relative ranking is close to perfect. The fact that the numbers for the pairwise relative ranking are higher than those for the size, may be surprising at first sight. This is in fact due to the self-correcting mechanism within the procedure. Due to Type I errors, wrongful rejection occurs by definition. The next bucket should contain only the wrongly rejected firm(s) as these are the most risky remaining firms.

The rest of the ranking is unaffected, a single bucket is simply split into two. This is also what happens in practise. To illustrate this, Figure 1 plots the allocation of firms to buckets 1, 2 and 3 for those simulations where the first estimated bucket contained two or less firms, instead of the required four. The firms within the lines, firms 1 to 4, and 5 to 8, are equally risky and should therefore be in a single bucket. The second bucket almost always takes the remaining firms and the third becomes what was the second under the DGP.

## 6. Empirical Application

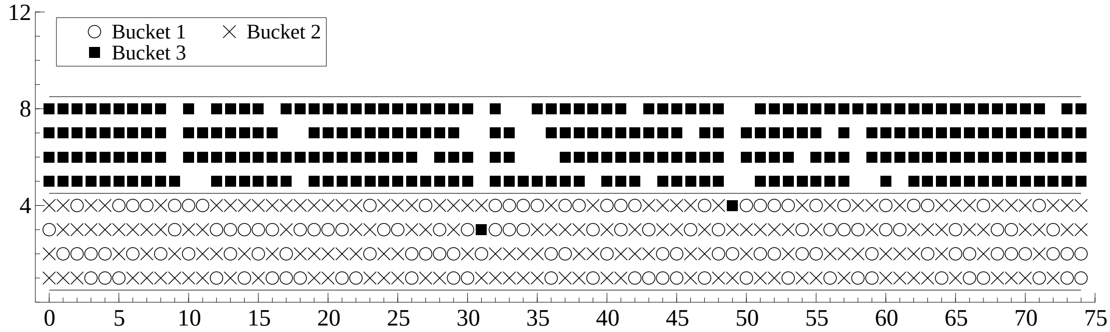
In this empirical application we apply the bucketing procedure to 94 of the largest U.S. financial firms. The dataset we use is identical to the panel studied by Acharya et al. (2010), Brownlees and Engle (2011) and many other papers on similar topics. It contains daily returns and market capitalizations retrieved from CRSP and quarterly book value of equity from Compustat. The data covers the period between January 3, 2000 and December 31, 2012, for a total of 3,269 daily observations. The market return is approximated by the CRSP market value-weighted index return. Market value is determined by CRSP daily closing prices and number of shares outstanding. Quarterly

Table 3: Simulation results on the Bucketing procedure for MES

| $\Delta\rho \backslash \Delta\sigma$ | 0.0 | 0.5 | 1.0 |
|---|---|---|---|
| **Panel A: First bucket** | | | |
| | $\mathcal{N}^{(1)} \subset \hat{\mathcal{N}}^{(1)}_{90\%}$ (size) | | |
| 0.00 | 0.882 | 0.918 | 0.910 |
| 0.05 | 0.911 | 0.907 | 0.905 |
| 0.10 | 0.912 | 0.910 | 0.909 |
| | Firms $i \notin \mathcal{N}^{(1)}$ not in $\hat{\mathcal{N}}^{(1)}_{90\%}$ (power) | | |
| 0.00 | - | 0.764 | 0.988 |
| 0.05 | 0.822 | 0.987 | 1.000 |
| 0.10 | 0.914 | 1.000 | 1.000 |
| **Panel B: All buckets** | | | |
| | Pairwise relative ranking correct | | |
| 0.00 | - | 0.654 | 0.917 |
| 0.05 | 0.836 | 0.939 | 0.949 |
| 0.10 | 0.899 | 0.958 | 0.967 |
| | Top 10 correctly identified | | |
| 0.00 | - | 0.995 | 0.999 |
| 0.05 | 1.000 | 1.000 | 1.000 |
| 0.10 | 1.000 | 1.000 | 1.000 |

*Note:* The table is divided into two parts. Panel A reports results on the first bucket only, showing the frequency at which the true first bucket is in the estimated bucket and the fraction of firms correctly eliminated. Panel B concerns all buckets and shows the fraction of pairwise rankings correct, and the fraction of simulations where the Top 10 was correctly identified.

Figure 1: Bucket allocation when the first estimated bucket is too small.




*Note:* The graph shows the cluster allocation of the different firms when the size of the first bucket two firms or less. Parameters are $\Delta\sigma = 0.5, \Delta\rho = 0.05$.

book values of total liabilities are from Compustat (LTQ). This results in a dataset which contains all U.S. financial firms with a market capitalization greater than 5bln USD as of the end of June 2007. A full list of ticker symbols and firms is given in Appendix C.

The objective of this application is twofold. First, we want to test whether the risk measures are estimated with enough precision to deliver a complete ranking of firms. If no absolute ranking can be distinguished, we want to test whether we can at least identify buckets of firms that are indistinguishable from each other within the bucket but distinguishable from firms belonging to lower ranked buckets. Different measures can be estimated with varying degrees of uncertainty, and may also differ in the ordering of point estimates. As such, different risk measures can lead to different rankings.

Second, given a suitable risk measure, our bucketing procedure can be used to obtain rankings very similar to the ranking of G-SIBs (Financial Stability Board, 2011, 2012), but on a daily basis.

We consider three measures, the MES, the %SRISK and the ΔCoVaR. The %SRISK is the percentage version of the SRISK, defined in Example 4, and is defined as

$$\%SRISK_{i,t} = SRISK_{i,t} / \sum_{j=1}^{N} SRISK_{j,t}.$$

The ΔCoVaR (Adrian and Brunnermeier, 2011) has yet to be defined. It is a measure based on the CoVaR, i.e. the Value-at-Risk of market returns, conditional on some event $\mathbb{C}(r_{i,t})$ observed for firm $i$:

$$\Pr\left(r_{m,t} \leq CoVaR_{i,t}^{m|\mathbb{C}(r_{i,t})} \mid \mathbb{C}(r_{i,t})\right) = \alpha. \tag{12}$$

The ΔCoVaR is defined as the difference between the VaR of the financial system conditional on the distress of a particular financial institution $i$ and the VaR of the system conditional on the median state of that same firm $i$. Although the framework it was developed in is completely different from the one we consider, the ΔCoVaR can also be represented in our assumed DGP of Equation (9) (Benoit et al., 2012). Adrian and Brunnermeier (2011) suggest using $r_{i,t} = VaR_i(\alpha)$ as the distress event $\mathbb{C}(r_{i,t})$. We obtain a time-varying measure by considering $r_{i,t} = VaR_{i,t}(\alpha)$. We then obtain

$$\Delta CoVaR_{i,t}(\alpha) = CoVaR_{i,t}^{m|r_{i,t}=VaR_{i,t}(\alpha)} - CoVaR_{i,t}^{m|r_{i,t}=Median(r_{i,t})}. \tag{13}$$

The authors propose to estimate the ΔCoVaR using quantile regressions.[10]

$$r_{m,t} = \mu_\alpha^i + \gamma_\alpha^i r_{i,t}. \tag{14}$$

The estimated CoVaR for the $i^{th}$ firm is then given by $CoVaR^{m|VaR_{i,t}(\alpha)} = \hat{\mu}_\alpha^i + \hat{\gamma}_\alpha^i VaR_{i,t}(\alpha)$, where $\hat{\mu}_\alpha^i$ and $\hat{\gamma}_\alpha^i$ are the estimated parameters of the quantile regression. Our definition differs from Adrian and Brunnermeier (2011) in that $\mathbb{C}(r_{i,t})$ is a function of the conditional and not the unconditional VaR. This leads to

$$\Delta CoVaR_{i,t}^{m|VaR_{i,t}(\alpha)} = \hat{\gamma}_\alpha^i \left(\widehat{VaR}_{i,t}(\alpha) - \widehat{VaR}_{i,t}(0.5)\right). \tag{15}$$

### *6.1. Bucketing different measures*

We report the bucket allocation for the MES, %SRISK and ΔCoVaR on eight pre-determined dates coinciding with those in Brownlees and Engle (2011). A firm is included in the ranking at a certain date if the firm still exists and furthermore there are at least 1,000 observations up until the date. We set $\alpha = 0.2$. This implies that every bucket contains the true set of most risky remaining firm with 80% probability. In practice a higher value will increase average bucket size and decrease the total number of buckets, and vice versa. Hansen et al. (2011) and Boudt et al. (2013) choose $\alpha = 0.25$.

In our tables, the firms are first ranked in terms of their bucket, and within buckets we order the firms in descending value of their RM estimate, even though there is no statistical evidence that their risk is statistically different. We then report the ten highest ranked firms, as is done in amongst others Brownlees and Engle (2011), Benoit et al. (2012) and **?**. This is an arbitrary cut-off point, and as such we also provide the size of the top 10. For instance, the first nine firms might be allocated to three buckets while the fourth one might contain more than one firm, in which case the top 10 is a

[10]Quantile regressions are estimated using RQ 1.0 for OxMetrics by Daniel Morillo and Roger Koenker.

Table 4: MES Rankings

| 30-03-2007 | | | 29-06-2007 | | | 31-12-2007 | | | 29-02-2008 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 5/83 | | | #Buckets/Firms: 5/83 | | | #Buckets/Firms: 5/81 | | | #Buckets/Firms: 6/82 | | |
| Size Top 10: 29 | | | Size Top 10: 53 | | | Size Top 10: 29 | | | Size Top 10: 36 | | |
| LEH | 1 | 3.54 | AMTD | 1 | 3.01 | ETFC | 1 | 9.41 | ABK | 1 | 8.27 |
| BSC | 1 | 3.47 | BSC | 1 | 2.80 | MBI | 1 | 9.29 | MBI | 1 | 7.64 |
| MS | 1 | 3.45 | MER | 1 | 2.63 | ABK | 1 | 7.53 | LEH | 1 | 5.28 |
| AMTD | 1 | 3.43 | MBI | 1 | 2.62 | FRE | 1 | 6.96 | CIT | 1 | 5.28 |
| ETFC | 1 | 3.26 | SCHW | 1 | 2.60 | CFC | 1 | 6.04 | WM | 1 | 5.15 |
| AGE | 1 | 3.22 | GS | 1 | 2.54 | WM | 1 | 5.82 | MER | 1 | 5.11 |
| JNS | 1 | 3.16 | LEH | 1 | 2.43 | CIT | 1 | 5.38 | FRE | 1 | 4.82 |
| GS | 1 | 3.16 | ETFC | 1 | 2.35 | FNM | 1 | 5.28 | CNA | 1 | 4.78 |
| BEN | 1 | 3.15 | TROW | 1 | 2.24 | MS | 1 | 5.15 | MI | 1 | 4.61 |
| MER | 1 | 2.82 | MS | 1 | 2.20 | LEH | 1 | 4.90 | BSC | 1 | 4.61 |

| 30-06-2008 | | | 29-08-2008 | | | 30-01-2009 | | | 30-06-2010 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 6/82 | | | #Buckets/Firms: 8/81 | | | #Buckets/Firms: 6/73 | | | #Buckets/Firms: 5/75 | | |
| Size Top 10: 49 | | | Size Top 10: 18 | | | Size Top 10: 26 | | | Size Top 10: 39 | | |
| LEH | 1 | 10.29 | FRE | 1 | 13.59 | STT | 1 | 22.19 | ABK | 1 | 7.62 |
| MBI | 1 | 9.78 | FNM | 1 | 13.40 | C | 1 | 20.88 | CBG | 1 | 6.97 |
| CIT | 1 | 8.11 | ABK | 1 | 12.94 | HBAN | 1 | 20.78 | MI | 1 | 6.79 |
| PFG | 1 | 6.56 | LEH | 1 | 12.42 | FITB | 1 | 19.82 | JNS | 1 | 6.75 |
| ABK | 2 | 7.81 | MBI | 1 | 9.64 | PNC | 1 | 19.82 | ETFC | 1 | 6.68 |
| FITB | 2 | 7.73 | AIG | 2 | 8.66 | AFL | 1 | 19.50 | ACAS | 1 | 6.62 |
| WM | 2 | 7.46 | MER | 2 | 8.63 | LNC | 1 | 19.03 | LNC | 1 | 6.57 |
| MER | 2 | 6.25 | RF | 2 | 7.68 | BAC | 1 | 18.49 | PFG | 1 | 6.32 |
| C | 2 | 5.82 | BAC | 2 | 7.23 | HIG | 1 | 17.41 | MBI | 1 | 6.23 |
| FRE | 2 | 5.71 | WM | 2 | 6.97 | PFG | 1 | 17.10 | AMP | 1 | 6.19 |

meaningless concept. Additionally, we report the total number of firms included in the ranking for that day as well as the total number of buckets they are allocated to.
The MES is estimated using $C = VaR_{m,t}(0.05)$ as the conditioning event. We first check for possible dynamics in the mean by for each series individually minimizing the Schwarz Information Criteria of the ARMA(m,n)-GJR-GARCH(1,1) model over $m, n = 0, ..., 3$. We tested for the presence of serial correlation in the residuals and their squares and failed to reject the null for all series. As such the i.i.d. bootstrap described in Section 4 will suffice. Table 4 reports the bucket allocation for the MES.

Results suggest that there is too much estimation uncertainty to be able to truly distinguish firms using the daily MES forecasts. We obtain anywhere between only five and eight buckets. On six out of eight days we cannot distinguish any differences between the ten highest estimates. However, two buckets are identified within the top ten in June and September 2008. Regardless, we conclude that the MES is not estimated with sufficient precision to dissociate firms with respect to this systemic risk measure. Indeed, for most dates, the first thirty firms belong to the same bucket. Consequently, ranking firms on the basis of point forecasts of MES seems hazardous. Table 5 shows the top ten of the bucket allocation for %SRISK. For the %SRISK we have to choose a value for the capital ratio $k$. Following Brownlees and Engle (2011) we

Table 5: %SRISK Rankings

| 30-03-2007 | | | 29-06-2007 | | | 31-12-2007 | | | 29-02-2008 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 7/14 | | | #Buckets/Firms: 7/13 | | | #Buckets/Firms: 19/36 | | | #Buckets/Firms: 17/37 | | |
| Size Top 10: 11 | | | Size Top 10: 10 | | | Size Top 10: 11 | | | Size Top 10: 13 | | |
| MS | 1 | 21.16 | MS | 1 | 17.92 | C | 1 | 16.59 | C | 1 | 15.13 |
| FRE | 2 | 14.61 | FRE | 1 | 16.35 | MER | 2 | 10.15 | MER | 2 | 8.86 |
| FNM | 2 | 13.37 | MER | 2 | 15.52 | MS | 3 | 9.16 | MS | 3 | 8.30 |
| LEH | 3 | 10.51 | BSC | 3 | 11.14 | FRE | 4 | 8.41 | FRE | 3 | 7.80 |
| GS | 3 | 10.38 | LEH | 3 | 10.90 | FNM | 5 | 7.65 | FNM | 3 | 7.74 |
| MER | 3 | 10.26 | FNM | 3 | 9.42 | GS | 5 | 7.60 | JPM | 3 | 7.43 |
| BSC | 4 | 9.10 | GS | 3 | 9.21 | LEH | 6 | 6.11 | GS | 4 | 7.36 |
| MET | 5 | 3.43 | MET | 4 | 3.34 | JPM | 6 | 6.01 | LEH | 5 | 6.09 |
| HIG | 5 | 1.87 | PRU | 4 | 2.63 | BAC | 6 | 3.42 | BAC | 5 | 4.49 |
| PRU | 5 | 1.78 | HIG | 4 | 2.12 | BSC | 7 | 3.96 | BSC | 6 | 3.77 |

| 30-06-2008 | | | 29-08-2008 | | | 30-01-2009 | | | 30-06-2010 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 19/39 | | | #Buckets/Firms: 16/36 | | | #Buckets/Firms: 32/53 | | | #Buckets/Firms: 18/37 | | |
| Size Top 10: 10 | | | Size Top 10: 10 | | | Size Top 10: 10 | | | Size Top 10: 10 | | |
| C | 1 | 15.17 | C | 1 | 13.43 | JPM | 1 | 15.28 | C | 1 | 16.38 |
| BAC | 2 | 9.06 | JPM | 2 | 9.75 | C | 2 | 14.22 | BAC | 1 | 16.12 |
| JPM | 3 | 8.06 | BAC | 2 | 9.56 | BAC | 3 | 12.92 | JPM | 2 | 13.85 |
| MER | 3 | 7.82 | MER | 3 | 7.41 | WFC | 4 | 9.27 | AIG | 3 | 8.56 |
| MS | 4 | 7.31 | FRE | 3 | 7.39 | AIG | 5 | 6.27 | MS | 4 | 7.06 |
| FRE | 5 | 6.53 | AIG | 3 | 7.04 | GS | 6 | 6.08 | WFC | 5 | 4.83 |
| FNM | 5 | 6.29 | FNM | 4 | 7.17 | MS | 7 | 4.59 | MET | 5 | 4.64 |
| GS | 6 | 5.66 | MS | 5 | 6.75 | MET | 7 | 3.61 | GS | 5 | 4.39 |
| AIG | 6 | 5.58 | GS | 6 | 6.05 | PRU | 8 | 3.37 | PRU | 5 | 4.23 |
| LEH | 7 | 5.22 | LEH | 7 | 5.14 | HIG | 9 | 2.19 | HIG | 6 | 3.05 |

set it to $k = 0.08$.[11] The total number of firms now denotes those firms with non-zero SRISK, although all firms are included in the bucketing.

The transformations from MES to SRISK induce extra variability and distance between firms, without adding additional uncertainty. As such, they turn out to be much easier to distinguish. The top ten now often consists of more buckets than the full sample for the MES. Unlike MES, results suggest that %SRISK can be estimated with a sufficient amount of precision to obtain a useful ranking.

Note also that our ranking produces some examples of the *loose cannons*. For instance, in August 2008, FNM has a higher point estimate than AIG, but is allocated to a lower bucket. This is due to FNM being estimated more precisely than AIG. The greater uncertainty in the estimate of AIG leads to a higher upper confidence bound than FNM's. As such AIG was put in a high risk bucket.

Finally, Table 6 shows the bucket allocation for the ΔCoVaR using $r_{i,t} = VaR_{i,t}(0.05)$ as the conditioning event. The combination of our assumed DGP and the estimation procedure leads to such large amounts of estimation uncertainty that we simply cannot

[11] Brownlees and Engle (2011) show limited sensitivity of the SRISK point estimate ordering to this variable.

Table 6: ΔCoVaR Rankings

| 30-03-2007 | | | 29-06-2007 | | | 31-12-2007 | | | 29-02-2008 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 1/83 | | | #Buckets/Firms: 2/83 | | | #Buckets/Firms: 2/81 | | | #Buckets/Firms: 1/82 | | |
| Size Top 10: 83 | | | Size Top 10: 75 | | | Size Top 10: 74 | | | Size Top 10: 82 | | |
| LEH | 1 | 1.09 | MBI | 1 | 0.94 | MBI | 1 | 3.17 | AIG | 1 | 1.77 |
| AGE | 1 | 1.05 | GS | 1 | 0.85 | ABK | 1 | 2.38 | CNA | 1 | 1.71 |
| MS | 1 | 0.98 | C | 1 | 0.81 | ETFC | 1 | 2.29 | MI | 1 | 1.67 |
| BSC | 1 | 0.95 | LEH | 1 | 0.80 | SLM | 1 | 2.12 | MER | 1 | 1.67 |
| BEN | 1 | 0.90 | MER | 1 | 0.79 | NCC | 1 | 1.74 | C | 1 | 1.57 |
| MER | 1 | 0.90 | JPM | 1 | 0.76 | WM | 1 | 1.73 | RF | 1 | 1.57 |
| GS | 1 | 0.89 | BSC | 1 | 0.75 | C | 1 | 1.69 | LEH | 1 | 1.56 |
| LM | 1 | 0.84 | SCHW | 1 | 0.73 | FRE | 1 | 1.62 | JPM | 1 | 1.50 |
| C | 1 | 0.83 | EV | 1 | 0.71 | FITB | 1 | 1.58 | EV | 1 | 1.47 |
| BBT | 1 | 0.82 | HRB | 1 | 0.69 | BBT | 1 | 1.55 | HBAN | 1 | 1.40 |

| 30-06-2008 | | | 29-08-2008 | | | 30-01-2009 | | | 30-06-2010 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Buckets/Firms: 2/82 | | | #Buckets/Firms: 3/81 | | | #Buckets/Firms: 2/73 | | | #Buckets/Firms: 1/75 | | |
| Size Top 10: 56 | | | Size Top 10: 31 | | | Size Top 10: 63 | | | Size Top 10: 75 | | |
| FITB | 1 | 3.48 | AIG | 1 | 2.97 | AFL | 1 | 9.05 | MTB | 1 | 1.93 |
| HBAN | 1 | 2.87 | LEH | 1 | 2.49 | PNC | 1 | 8.27 | BEN | 1 | 1.88 |
| LEH | 1 | 2.68 | MI | 1 | 2.49 | STT | 1 | 6.89 | TROW | 1 | 1.79 |
| KEY | 1 | 2.37 | FRE | 1 | 2.43 | FITB | 1 | 6.41 | EV | 1 | 1.67 |
| RF | 1 | 2.33 | MER | 1 | 2.31 | BAC | 1 | 5.97 | MI | 1 | 1.66 |
| C | 1 | 2.28 | RF | 1 | 2.30 | ACAS | 1 | 5.54 | AFL | 1 | 1.61 |
| STI | 1 | 2.03 | KEY | 1 | 2.05 | ALL | 1 | 5.49 | AXP | 1 | 1.57 |
| BBT | 1 | 2.02 | FNM | 1 | 2.01 | WFC | 1 | 5.40 | CINF | 1 | 1.51 |
| AIG | 1 | 2.01 | SNV | 1 | 1.97 | STI | 1 | 5.26 | GS | 1 | 1.49 |
| MI | 1 | 1.95 | C | 1 | 1.94 | C | 1 | 5.14 | SCHW | 1 | 1.47 |

distinguish any firms in the top ten at all. Despite large variations in point estimates of ΔCoVaR, only one bucket is identified for most dates, and the maximum number of buckets is three. This is despite large variation in the point estimates. For instance, the highest point forecast of ΔCoVaR is 9.05 for AFL, but its bootstrap standard deviation is close to 4. In an unreported simulation we find that even if the true DGP is exactly the one assumed here, the standard deviation of the ΔCoVaR is still on average over 40% of its value. This is fully due to the quantile regression estimates, which have a very wide distribution. As a result, the procedure simply fails to make enough rejections to make the ranking useful.

To conclude, results suggest that ranking firms with respect to MES and ΔCoVaR is hazardous and that %SRISK is the only systemic risk measure (among the three considered here) that delivers a meaningful ranking. Indeed, using our bucketing procedure on %SRISK, we are able to identify several buckets of firms and to obtain a meaningful ranking of buckets containing equally risky firms in each bucket. As such we use it to obtain a ranking on dates close to those when the G-SIB rankings were published.

Table 7 contains the top 20 of firms according to %SRISK on the BIS Ranking dates, beginning of November 2011 and 2012. The ranking is however difficult to compare

Table 7: SRISK BIS-Dates

| 2011-11-01 | | | 2012-11-01 | | |
|---|---|---|---|---|---|
| TICK | Bucket | %SRISK | TICK | Bucket | %SRISK |
| BAC | 1 | 17.05 | BAC | 1 | 19.48 |
| JPM | 1 | 15.44 | C | 1 | 17.00 |
| C | 1 | 14.35 | JPM | 1 | 14.26 |
| MS | 1 | 6.43 | MET | 2 | 9.00 |
| GS | 2 | 6.25 | MS | 3 | 7.81 |
| MET | 2 | 6.08 | GS | 3 | 7.58 |
| WFC | 2 | 5.76 | PRU | 3 | 7.22 |
| PRU | 2 | 4.63 | HIG | 4 | 3.63 |
| HIG | 2 | 3.02 | LNC | 5 | 2.52 |
| AIG | 3 | 2.43 | SLM | 6 | 1.85 |
| BK | 3 | 1.60 | PFG | 6 | 1.54 |
| LNC | 4 | 1.63 | AIG | 6 | 1.31 |
| SLM | 4 | 1.58 | BK | 6 | 1.02 |
| STI | 4 | 1.11 | GNW | 7 | 1.39 |
| PFG | 4 | 1.06 | RF | 8 | 0.85 |
| PNC | 4 | 0.85 | AMP | 8 | 0.72 |
| STT | 4 | 0.85 | STI | 8 | 0.65 |
| COF | 4 | 0.71 | ETFC | 8 | 0.58 |
| BBT | 4 | 0.70 | STT | 8 | 0.15 |
| BLK | 4 | 0.62 | BBT | 8 | 0.10 |

to the BIS ranking. They have an international sample, and we have non-banking institutions. To facilitate comparison, Table 8 contains only the U.S. firms in the G-SIB list and our ranking omitting the Insurance and Other companies (see Appendix C). We show all firms in the BIS ranking and the first four buckets identified with the bucketing procedure on %SRISK. The ranking is very similar. BAC is ranked one bucket riskier, and BK one bucket less risky. We were not able to distinguish STT from many firms which are not present in the G-SIB list. The one real outlier is WFC, which is in the third BIS Bucket, but has zero estimated SRISK, and as such does not appear in our ranking. Overall, our ranking is remarkably close to the BIS ranking.

## 7. Conclusion

This paper introduces a test of equality of single point forecasts for a general class of risk measures, as well as an iterative procedure to produce a grouped ranking.

Table 8: BIS ranking published November 2012 and our buckets on November 1 2012

| Bucket | BIS | Allocation 2012-11-01 |
|---|---|---|
| 1 | C, JPM | C, JPM, BAC |
| 2 | BAC, BK, GS, MS | GS, MS |
| 3 | WFC, STT | BK |
| 4 | | RF, STI, ETFC, STT, BBT |

*Note:* The table reports the BIS ranking excluding non-US firms, and the first four clusters of our buckets excluding "Insurance" and "Other" companies.

Simulation results on VaR and MES forecasts suggest that the test has good properties in finite samples, both in terms of size and power.

We applied the bucketing procedure on real forecasts for three popular systemic risk measures. For the MES and ΔCoVaR, we find that for most of the eight dates considered in the application, the first thirty firms belong to the same bucket of riskiest firms. Consequently, ranking firms on the basis of point forecasts of MES and ΔCoVaR seems hazardous. However, when applied on %SRISK, our bucketing procedure is able to identify a meaningful ranking of buckets containing equally risky firms in each bucket. Interestingly, this ranking is very close to the G-SIB ranking.

## 8. References


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

Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2010. Measuring systemic risk. Working paper.

Adrian, T., Brunnermeier, M. K., 2011. CoVaR. Working paper.

Aielli, G. P., 2013. Dynamic conditional correlation: on properties and estimation. Journal of Business & Economic Statistics 31 (3), 282–299.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1997. Thinking coherently. Risk 10, 68–71.

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

Basel Committee on Banking Supervision, July 2013. Global systemically im- portant banks: updated assessment methodology and the higher loss absorbency requirement. Bank for international settlements.

Benoit, S., Colletaz, G., Hurlin, C., 2012. A theoretical and empirical comparison of systemic risk measures. Working paper.

Bisias, D., Flood, M., Lo, A., Valavanis, S., 2012. A survey of systemic risk analytics. US Department of Treasury, Office of Financial Research (0001).

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. The Review of Economics and Statistics, 498–505.

Boudt, K., Danielsson, J., Laurent, S., 2013. Robust forecasting of dynamic conditional correlation garch models. International Journal of Forecasting 29 (2), 244–257.

Brownlees, C., Engle, R., 2011. Volatility, correlation and tails for systemic risk measurement. Working paper.

Castro, C., Ferrari, S., 2011. Measuring and testing for the systemically important financial institutions. Working paper.

Danielsson, J., James, K. R., Valenzuela, M., Zer, I., 2011. Model risk of systemic risk models. Working paper.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Doornik, J., 2013. Object-Oriented Matrix Programming Using Ox. Timberlake Consultants Press.

Engle, R., 2012. Dynamic conditional beta. Working paper.

Financial Stability Board, 2011. Policy measures to address systemically important financial institutions. Fsb publication, Financial Stability Board.

Financial Stability Board, 2012. Update of group of global systemically important banks (g-sibs). Fsb publication, Financial Stability Board.

Francq, C., Zakoïan, J.-M., 2012. Risk-parameter estimation in volatility models. Working paper.

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

Gourieroux, C., Zakoïan, J.-M., 2013. Estimation adjusted VaR. Econometric Theory 29 (4), 735–770.

Hansen, P. R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23 (4), 365–380.

Hansen, P. R., Lunde, A., 2006. Consistent ranking of volatility models. Journal of Econometrics 131 (1), 97–121.

Hansen, P. R., Lunde, A., Nason, J., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Laurent, S., 2013. G@RCH 7.0: Estimating and forecasting arch models. London: Timberlake Consultants Ltd.

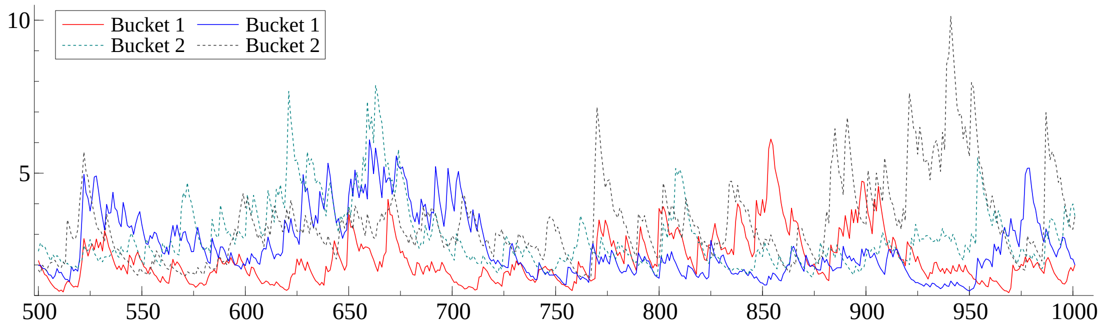
Pascual, L., Romo, J., Ruiz, E., 2006. Bootstrap prediction for returns and volatilities in garch models. Computational Statistics & Data Analysis 50 (9), 2293–2312.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160 (1), 246–256.

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

Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. Insurance and Risk Management Journal 74, 382–406.

Figure A.2: Simulated and scaled variance processes




## Appendix A. Rescaling GARCH processes

We rescale the GARCH process, such that the sum of the conditional variances from $T+1$ to $T+H$ exactly equals $K$. We start with a process whose sum of variances is $\sum_{i=1}^{H} \sigma^2_{T+i}$. We divide the entire series by $\sqrt{\sum_{i=1}^{H} \sigma^2_{T+i}/K}$. For instance, for $H = 2$ we have

$$y_{T+1} = \epsilon_{T+1}\sigma_{T+1}$$
$$y_{T+2} = \epsilon_{T+2}\sigma_{T+2}$$

with sum of conditional variances $\sigma^2_{T+1} + \sigma^2_{T+2}$. The transformed series

$$\tilde{y}_{T+1} = \epsilon_{T+1}\sqrt{K}\sigma_{T+1}/\sqrt{\sigma^2_{T+1} + \sigma^2_{T+2}}$$
$$\tilde{y}_{T+2} = \epsilon_{T+2}\sqrt{K}\sigma_{T+2}/\sqrt{\sigma^2_{T+1} + \sigma^2_{T+2}}$$

which has sum of variances $K(\sigma^2_{T+1} + \sigma^2_{T+2})/(\sigma^2_{T+1} + \sigma^2_{T+2}) = K$.
As an example, Figure A.2 depicts the variance processes of four simulated series belonging to two different buckets. The variance at time 1000 is equal for the series belonging to the same bucket.

## Appendix B. Estimation of the MES

*Volatilities.* The volatilities are estimated using the GJR-GARCH specification (see Glosten et al. (1993); Rabemananjara and Zakoïan (1993)), where the conditional variance is updated as

$$\sigma^2_{i,t} = \gamma + \alpha_{1i}\epsilon^2_{i,t-1} + \alpha_{2i}\epsilon^2_{i,t-1}I_{\epsilon_{i,t}<0} + \beta_i\sigma^2_{i,t-1}. \tag{B.1}$$

*Correlation.* The time-varying conditional correlation is modelled using the cDCC model of Aielli (2013). In this specification, the conditional covariance matrix $\Sigma_t$ is decomposed as

$$\Sigma_t = D_t R_t D_t, \tag{B.2}$$

where $R_t$ is the conditional correlation matrix and $D_t$ a diagonal matrix of conditional volatilities $\sigma_{i,t}$. The cDCC framework introduces a positive definite pseudo-correlation matrix $Q_t$, defined and updated as

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \tag{B.3}$$

$$Q_t = (1 - \alpha_C - \beta_C)S + \alpha_C \epsilon^*_{t-1} \epsilon^{*'}_{t-1} + \beta_C Q_{t-1}, \tag{B.4}$$

where $\epsilon^*_t = diag(Q_t)^{1/2} \epsilon_t$ and $S$ is the sample covariance matrix of $\epsilon^*_t$.

*Tail expectations.* Brownlees and Engle (2011) use a non-parametric kernel estimator of Scaillet (2005) to obtain the tail expectations.

$$\hat{\mathbb{E}}(\epsilon_{mt} \mid \epsilon_{mt} < \kappa) = \frac{\sum_{t=1}^{T} \epsilon_{mt} \Phi\left(\frac{\kappa - \epsilon_{mt}}{h}\right)}{\sum_{t=1}^{T} \Phi\left(\frac{\kappa - \epsilon_{mt}}{h}\right)}, \tag{B.5}$$

and

$$\hat{\mathbb{E}}(\xi_{it} \mid \epsilon_{mt} < \kappa) = \frac{\sum_{t=1}^{T} \xi_{it} \Phi\left(\frac{\kappa - \epsilon_{mt}}{h}\right)}{\sum_{i=1}^{T} \Phi\left(\frac{\kappa - \epsilon_{mt}}{h}\right)}, \tag{B.6}$$

where $\kappa = C/\sigma_{mt}$, and $\Phi(u)$ is the standard normal cdf. Following Scaillet (2005), we set the bandwidth $h = T^{-1/5}$.

## Appendix C. Company Tickers

| Depositories(29) | | Insurance (32) | |
|---|---|---|---|
| BAC | Bank of America | ABK | Ambac Financial Group |
| BBT | BB&T | AET | Aetna |
| BK | Bank of New York Mellon | AFL | Aflac |
| C | Citigroup | AIG | American International Group |
| CBH | Commerce Bancorp | AIZ | Assurant |
| CMA | Comerica inc | ALL | Allstate Corp |
| HBAN | Huntington Bancshares | AOC | Aon Corp |
| HCBK | Hudson City Bancorp | WRB | W.R. Berkley Corp |
| JPM | JP Morgan Chase | BRK | Berkshire Hathaway |
| KEY | Keycorp | CB | Chubb Corp |
| MI | Marshall & Ilsley | CFC | Countrywide Financial |
| MTB | M&T Bank Corp | CI | CIGNA Corp |
| NCC | National City Corp | CINF | Cincinnati Financial Corp |
| NTRS | Northern Trust | CNA | CNA Financial corp |
| NYB | New York Community Bancorp | CVH | Coventry Health Care |
| PBCT | Peoples United Financial | FNF | Fidelity National Financial |
| PNC | PNC Financial Services | GNW | Genworth Financial |
| RF | Regions Financial | HIG | Hartford Financial Group |
| SNV | Synovus Financial | HNT | Health Net |
| SOV | Sovereign Bancorp | HUM | Humana |
| STI | Suntrust Banks | LNC | Lincoln National |
| STT | State Street | MBI | MBIA |
| UB | Unionbancal Corp | MET | Metlife |
| USB | US Bancorp | MMC | Marsh & McLennan |
| WB | Wachovia | PFG | Principal Financial Group |
| WFC | Wells Fargo & Co | PGR | Progressive |
| WM | Washington Mutual | PRU | Prudential Financial |
| WU | Western Union | SAF | Safeco |
| ZION | Zion | TMK | Torchmark |
| | | TRV | Travelers |
| | | UNH | Unitedhealth Group |
| | | UNM | Unum Group |

| Broker-Dealers (10) | | Others (23) | |
|---|---|---|---|
| AGE | A.G. Edwards | ACAS | American Capital |
| BSC | Bear Stearns | AMP | Ameriprise Financial |
| ETFC | E-Trade Financial | AMTD | TD Ameritrade |
| GS | Goldman Sachs | AXP | American Express |
| LEH | Lehman Brothers | BEN | Franklin Resources |
| MER | Merill Lynch | BLK | Blackrock |
| MS | Morgan Stanle | BOT | CBOT Holdings |
| NMX | Nymex Holdings | CBG | C.B. Richard Ellis Group |
| SCHW | Schwab Charles | CBSS | Compass Bancshares |
| TROW | T.Rowe Price | CIT | CIT Group |
| | | CME | CME Group |
| | | COF | Capital One Financial |
| | | EV | Eaton Vance |
| | | FITB | Fifth Third bancorp |
| | | FNM | Fannie Mae |
| | | FRE | Freddie Mac |
| | | HRB | H&R Block |
| | | ICE | Intercontinental Exchange |
| | | JNS | Janus Capital |
| | | LM | Legg Mason |
| | | NYX | NYSE Euronext |
| | | SEIC | SEI Investments Company |
| | | SLM | SLM Corp |