Ca' Foscari University of Venice

Master's Degree programme

in Economics and Finance
Second Cycle (D.M. 270/2004)

Final Thesis

Are Systemic Risk Measures Really Useful for Regulators?

An Assessment of the Estimation Risk

Supervisor
Ch. Prof. Fulvio Corsi

Assistant supervisor
Ch. Prof. Andrea Giacomelli

Graduand
Umberto Bertiglia
843700

Academic Year
2016 / 2017

Index

Index of Figures

Index of Tables

Are Systemic Risk Measures Really Useful for Regulators?: An Assessment of the Estimation Risk

ABSTRACT: Systemic risk measures have gained popularity in the recent finance literature and they are widely applied for detecting systemic risk contributions of financial institutions. It has also been suggested that a regulator should base capital requirements upon the contribution to systemic risk of each single institution. We analyse market data based systemic risk measures such as Delta CoVaR, MES and SRISK. We find that the uncertainty about their estimates is high. Using bootstrap techniques, we develop a nonparametric hypothesis test to assess if these measures are statistically different between institutions. We find that it is hard to rank institutions using their CoVaR or MES, while SRISK gives good results. We conclude that confidence intervals, incorporating the uncertainty of the measure should be provided and that systemic risk measures should be applied with caution, especially when they are used for important purposes such as the design of a new regulatory framework, because they can lead to expensive bad decisions.

KEYWORDS: Risk Management – Systemic Risk – CoVaR – MES – SRISK - Bootstrap.

Introduction

Systemic risk has recently emerged as in important topic in the finance literature. Research on this topic has intensified since recent events as the bankruptcy of Lehman Brothers in September 2008 and the Eurozone sovereign debt crisis. These events have highlighted the tremendous externalities that can hit the real economy when the financial system does not function properly, drawing attention to the need for tools that would detect these deficiencies at an early stage; the object is try to control them in order to avoid, or at least, limit these externalities to the real economy. Traditional risk measures have been blamed after the crisis, because they consider only the specific risk of an institution in isolation and not as part of the financial system: due to its inadequacy, the regulatory framework has added fuel to the fire of the crisis and it is not surprising that one of the arguments after the crisis was the design of a new regulatory framework. For these reasons, a new class of risk measures have been developed in order to detect the contribution to systemic risk of each financial institution. These new measures should help regulators and policymakers to detect financial institutions that pose a threat for the financial system and for the economy as a whole and limit the negative effects of externalities that may arise.

Since systemic risk is not an observable quantity, the importance of statistical risk measures is increased in the recent years after the crisis and they will play a key role in measuring and forecasting the risk of financial institutions in the future. There are several proposals from the academic sector for using systemic risk measures for determining capital requirements for financial institutions. Mainly for these reasons, it is important to understand if regulators can rely on systemic risk measures and if they really provide a different signal compared to traditional risk measures. The purpose of our work is to assess how these measures are computed empirically and what is the uncertainty that we have about these estimates. In order to do so we propose a study of the prominent systemic risk models based on market data which involves the application of bootstrap techniques. We assess the estimation risk of each systemic risk measure computing standard errors and confidence intervals using bootstrap techniques. We find that the estimation risk in quite high for all the measures we analyse. Then we exploit the relationship between bootstrap confidence intervals and hypothesis testing in order to test for the statistical difference between the systemic risk of two financial institutions. We find that due to the large uncertainty of the risk measures, very often systemic risk measures of two institutions are not statistically different and thus we have difficulties in identifying which institutions are really systemically important. Only using SRISK as our systemic risk measure of reference, we were able to divide our sample of institutions into groups that represent different levels of risk. A regulator should use this kind of analysis to have a more clear picture of what is the contribution that a particular financial institution brings to the system and then develop a policy that is coherent with these results. For example, if the systemic risk measures of two institutions are not statistically different, then they should be subject to the same capital requirements.

We think that our results would be of considerable interest for regulators and researchers in the field of systemic risk, because they will better understand how reliable are these models. As suggested by Danielsson et al. (2016), the output of these models is used as an input into expensive decisions and thus it is crucial to understand on what basis they are taking a decision and for this reason an assessment of the estimation risk is crucial.

The outline of the rest of the paper is as follows. In section I we start by analysing the various definitions of systemic risk and systemically important financial institutions (SIFIs). Then we review the literature regarding systemic risk measures based on market data from a theoretical point of view. We analyse in detail four systemic risk measures: CoVaR, Co-Risk, MES and SRISK. Then we also consider the role that systemic risk measures can play to design a new regulatory framework. In section II, after presenting our data and methodology, we apply our statistical procedure to Delta CoVaR, MES and SRISK and we present our empirical findings. The conclusions features a discussion about our main findings and a possible extension of future work.

Section I: Literature Review

This section presents a literature review on systemic risk theory developed in the last years. We try to highlight the main findings and the problems that have not already been solved. The first problem we present is how to define systemic risk: there is no universal definition of systemic risk and various alternatives have been proposed by scholars and policy makers. Various definitions of systemic risk have produced different methods to measure it. We will focus on market data based measures, because they are the ones that have received much more attention. We will present the strengths and weaknesses of each measure and the possible implementation in a regulatory framework.

1.1 Defining Systemic Risk

A definition is not just a formalization of a concept, but represents the foundations on which the following work relies on. Differently from other types of risks, the literature on systemic risk is based on multiple definitions which reflect different aspects of the phenomenon: there is no general theoretical framework which means that sometimes the definitions overlap and sometimes they are very different. Several authors deal with the problem of systemic risk definition in the introductory parts of their works: we decide to focus only on the definitions of systemic risk proposed in recent work, omitting the literature developed before the crisis.

1.1.1 Definitions of Systemic Risk

Adrian and Brunnermeier (2016, p. 1705) defines systemic risk as *'the risk that the capacity of the entire financial system is impaired, with potentially adverse consequences for the real economy'*. According to their definition, systemic risk arises because of the spillovers across institutions which can be due to direct contractual links or, indirectly, to price movements and liquidity drains. When these spillovers reach critical levels, it is possible that also the real economy is affected through the credit channel. This phenomenon can arise from a single financial institution that individually generates negative spillovers on the others: for example, a large bank that fails to fulfil its contracts; but it is possible that systemic risk arises as the product of several institutions being systemically important as a herd: for example, this could happen when a group of small institutions hold the same portfolio, at the same time, and a price drop due to external reasons creates a problem for all the institutions.

For Brownlees and Engle (2017, p. 48) systemic risk arises when *'obligations will spread throughout both the financial and real economy, and the natural functions of the financial sector will be curtailed'*. In their work, Brownlees and Engle (2017) apply a definition similar to that of Adrian and Brunnermier (2016), but they focus more on the undercapitalization of large financial institutions, as the main source of systemic risk. They argue that, when the system is undercapitalized, it is not able to fulfil its basic functions. For example, the bankruptcy of a financial institution cannot be absorbed by a bigger competitor, when the system as a whole is undercapitalized. Indeed, an undercapitalized system is unable to provide credit to householders or businesses and this fact will have immediate consequences on the real economy.

The debate on systemic risk involves not only scholars, but also policy makers such as the European Central Bank (ECB). The ECB (2009, p. 134) defines systemic risk as the possibility that *'the instability in the financial industry becomes so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially'*. This definition is very similar to those presented before, because it entails the disruption of the basic functions of the financial system and the link to the real economy.

Also the International Monetary Fund (IMF) (2009, p. 113) recognises the difficulty in defining systemic risk, because *'it is often viewed as a phenomenon that is there when we see it, reflecting a sense of broad-based breakdown in the functioning of the financial system, which is normally realized, ex post, by a large failures of FIs (usually banks)'*. This definition highlights that it is difficult to define *a priori* what causes the financial instability and thus systemic risk; in a different way from the others, their definition does not refer to the consequences that systemic risk has on the real economy. They suggest to investigate the financial institutions and the linkages among them and not the link between the system and a single financial institution. In a way similar to Adrian and Brunnermeier (2016), the interlinkages between institutions are modelled as the dependence in the returns in stressed conditions (when they are in the left tail of the distribution).

Patro et al. (2013, p. 106) define systemic risk as *'the likelihood of experiencing a systemic failure, a broad-based breakdown of financial system that is triggered by a strong systemic event (e.g., a financial institution failure), which severely and* *negatively impacts in the financial markets and the economy in general'*. According to this definition, we can measure the systemic risk with the standard tools of probability and statistics: if the probability of a systemic failure is high, we are in presence of high systemic risk, otherwise the risk is lower. In this definition, the focus is primarily on the financial system as a whole and not on single systemic institutions, as in Adrian and Brunnermeier (2016).

Abdymomunov (2013) links the definition of systemic risk to the systemic financial stress which arises when the market participants change their expectations on future values of assets, losses and, more in general, the economic activity. According to this definition, we can measure the systemic financial risk by looking at the level of stress in the financial markets which is indicated by the level of volatility in asset prices. In a similar way, Billio et al. (2012) suggest that one of the main symptom of systemic risk is the existence of abrupt shifts in the regime, because the economy is characterized by low volatility during economic growth and high volatility during a recession.

It is interesting to note that the way in which systemic risk measures are shaped is a direct consequence of the definition. For example, Adrian and Brunnermier's CoVaR is an attempt to measure the increase in the tail co-movement that can arise due to the spreading of financial distress across institutions. Brownlees and Engle's SRISK instead tries to capture the contribution that a single institution brought to the undercapitalization of the entire system. Therefore, it is not surprizing that we can find different systemic risk definitions and measures that try to capture the different features of each definition. In general, we can observe two types of focus which characterize the recent literature: the former is on the financial system as a whole, while the latter focus is on the single institutions.

We can notice that all the definitions of systemic financial risk have some common features. First of all, they all recognize that an event or shock that alters the normal functioning of the financial system, in order to be considered systemic, must propagate from the financial system itself to the real economy: otherwise we are not dealing with systemic risk. Even a big shock in the financial system does not always trigger a systemic event. For example the failure of an institution is considered a shock for the financial system; but if this institution is acquired by a bigger competitor, it is likely that this shock will not have consequences on the real economy and thus it will not be considered a systemic event. Instead, if the institution must be bailed out by the government with the money of taxpayers, because the market cannot provide a solution, it is more likely that this will have an impact on the real economy. Another common feature of all definitions is that they all require that the basic functions of the financial system are impaired. The functions of the financial system are classified in: clearing and settling payments, pooling resources and subdividing shares, transferring resources across time (and space), managing risk, separating ownership from management, providing information and facilitating price discovery. All of these functions are necessary in a complete financial system and the lack of one of them could trigger a systemic event.

1.1.2 Definition of Systemically Important Financial Institutions

Another important concept related to the one of systemic risk is the definition of systemically important financial institutions (SIFIs). This is very important in particular for what concerns the regulation. According to Tarullo (2009), *'financial institutions are systemically important if the failure of the firm to meet its obligations to creditors and costumers would have significant adverse consequences for the financial system and the broader economy'*. Also in this definition are present two concepts that permit the identification of systemic risk and thus SIFIs; the first one is that the institution is unable to do its work by paying its creditors and providing credit to its consumers and thus permitting the basic functions of a financial system; the second is that this fact creates problems also in the real economy and so it is not only relegated to the financial industry. Therefore, we can claim that this definition is consistent with the definitions of systemic risk proposed by the financial literature. But as Acharya et al. (2012) point out, this definition describe the systemic risk brought by a single financial institution, without referring to the condition of the system. In fact, in normal conditions, if one financial firm has a shortage of capital, this will be absorbed by a bigger competitor in the industry and there will not be direct consequences to the real economy: in this case we are not in presence of systemic risk. Instead, when the system as a whole is undercapitalized and a firm experiences problems, this will probably trigger a systemic event, because the market alone is not able to find a solution to this situation and thus, there is the need of an intervention by the government or the central bank with the relative externalities to the real economy.

For the Financial Stability Board (2010), SIFIs are defined as firms *'whose disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity'*.

This definition might be very similar to the one proposed in Tarullo (2009), but there is an important difference. Here institutions are marked as systemically important also for their interconnectedness with the system, a characteristic that is not present in the previous definition where SIFIs are identified only on the basis of their inability to meet their obligations.

According to the Basel III agreements, the SIFIs must be object of an additional capital requirement proportional to the externalities that they could cause to the real economy. This can be viewed as an additional tax that this type of firms needs to pay to avoid that the burden of a bailout, in case of a crisis, is carried by the householders. In addition, the Basel Committee on Banking Supervision distinguishes between SIFIs and Global Systemically Important Banks (G-SIBs): the former are identified according to size, complexity and interconnectedness, the latter by lack of readily available substitutes for financial infrastructures and their global activity.

1.2 Measuring Systemic Risk

The old quote *'you can't manage what you don't measure'* by Peter Drucker is still appropriate when dealing with systemic risk. For this reason, a large part of the literature is devoted to develop methods to measure this risk. There are mainly two research path in the literature of systemic risk measurement (Borri et al., 2012). The first is called *network analysis* and tries to estimate the links between financial institutions and the possible domino effect, if one firm defaulted. We will not deal with this type of research in this work, but examples of this approach can be found in Markose et al. (2012) and Martinez-Jaramillo et al. (2010). The other strand of the literature is called *micro-evidence approach* and looks at each financial institution through some specific variables in order to evaluate systemic risk brought by each institution. Our work is connected to this second strand of the literature, because we analyse some systemic risk measures such as Delta CoVaR, Co-Risk, MES and SRISK that have been developed with the object of identifying systemic risk in financial institutions. Since the literature on systemic risk is wide and varied, many approaches have been proposed to classify measures of this risk. Bisias et al. (2012) propose different methods to aggregate systemic risk measurement: they classify systemic risk measurement by data requirements, by supervisory scope, by event-/decision-time and by research method. We think that the most appropriate way to classify systemic risk measurement is through data requirements, because it is the key feature that differentiate systemic risk measures. Data can be public or proprietary: public data are available to everyone interested in, while proprietary data are available only to regulators and, in some cases, only to people that work inside the firm. Mainly for this reason we distinguish between systemic risk measurement that requires only public data and those that do not.

1.2.1 Market Data Based Approach

The market data based method is probably the prevalent way used to build systemic risk measures, because it permits to quickly estimate the joint probability distribution of negative outcomes of a group of financial institutions and the market. This is true for several reasons. First of all, most scholars do not have access to proprietary data that only regulators can obtain and thus they focus their research on those measures that can easily be estimated by using public data such as stock prices and balance sheet values. But even for regulators, it could be hard to obtain useful data: players in the financial industry are forced to maintain secret their portfolio exposure in order to be profitable, because the intellectual property in finance cannot be protected by a patent. This fact creates a trade-off between transparency and profitability of the business and an incentive for the institutions to be opaque, unless they are forced by a regulator to disclose some proprietary information. This is one of the reasons why systemic risk measures based on contingent claim analysis of the financial institution's assets relies on very strong assumptions in order to cover the unavailability of this type of data (Lehar, 2005; Gray et al., 2007). If a regulator forced an institution to make public every proprietary information, there would be nothing to regulate anymore. The second important aspect is that market values are for their nature forward-looking and thus they provide an estimate of future values of the firm, taking into account information and factors that are not present in accounting and balance sheet values. They are also hard to manipulate (Lucas, 2014), while there is a long history of balance sheet manipulations. Well known drawbacks of market prices are their excessive volatility and their procyclicality, that could alter significantly the risk measures exactly when we need them the most. Another main problem is that measures based on market data are not applicable to institutions that are not listed (Acharya et al., 2014), but since we are considering only big financial institutions, this problem is negligible.

Here we adopt the classification of systemic risk measures based on market data proposed by Danielsson et al. (2016). Generally, there are two types of systemic risk that can be assessed: the risk of an institution given the risk of the system and the risk of the system given the risk of the institution. This two measures are based on the risk of the institution and the system viewed in isolation which are usually VaR and ES.

Let us define R_i the daily returns of the institution i and R_S the daily returns of the financial system: the joint density function of the returns of an institution i and the system is $f(R_i, R_S)$, where $f(R_i)$ is the marginal density of the institution i, while $f(R_i \mid R_S)$ and $f(R_S \mid R_i)$ are the conditional distributions. By applying the Bayes' theorem and considering $f(R_S)$ a normalizing constant, we can define the risk of the institution i given the system as

$$f(R_i \mid R_S) \propto f(R_S \mid R_i) f(R_i) . \tag{1.1}$$

On the other hand, the risk of the system conditional on a single institution i is defined as

$$f(R_S \mid R_i) \propto f(R_i \mid R_S) f(R_S) . \tag{1.2}$$

Let us consider VaR as our risk measure; we define Q as a negative quantile with the related probability p (95% or 99%) where $P(R \leq Q) = p$. VaR is simply $-Q$, while ES is defined as

$$ES = E(R \mid R \leq Q) . \tag{1.3}$$

Adrian and Brunnermeier's CoVaR of institution i is obtained from (1.1) and it is defined as

$$CoVaR_i = P(R_S \leq Q_S \mid R_i \leq Q_i) = p . \tag{1.4}$$

Instead marginal expected shortfall (MES) proposed by Acharya et al. (2017) is obtained from (1.2) and it is defined as

$$MES_i = E(R_i \mid R_S \leq Q_S) \tag{1.5}$$

We can define also MVaR and CoES, even if these measures are not very much used in practice:

$$MVaR_i = P(R_i \le Q_i \mid R_S \le Q_s) = p\,, \tag{1.6}$$

$$CoES_i = E(R_S \mid R_i \le Q_i)\,. \tag{1.7}$$

Marginal risk measure	**Conditional on system**	**Conditional on institution**
VaR	MVaR	CoVaR
ES	MES	CoES

Table 1: classification of systemic risk measures based on the marginal risk measure and the conditional event.

Table 1 summarises the classification of systemic risk measures. CoVaR and MES are in some sense specular measures, because while MES considers the returns of an institution when the financial system is in its left tail, CoVaR does the opposite by looking at the returns of the system when a particular institution is in its left tail. Other popular risk measures fall under this general classification. The $Co-Risk_{i,j}$ measure developed in IMF (2009) is the risk of institution i conditional on institution j and not on the system, as it is CoVaR: it is a measure of dependence of an institution on another institution. SRISK proposed by Brownlees and Engle (2017) is a nonlinear transformation of MES that takes into account also the leverage, and market capitalization of the firm. In the next paragraphs we will present each systemic risk measure in detail from a theoretical point of view. We will also analyse the single strengths and weaknesses of each measure and the comparative advantages that one measure can have on the other measures.

CoVaR

The idea behind the work of Adrian and Brunnermeier (2016) is that measures that consider the risk of an institution in isolation such as VaR and ES are not able to capture the systemic component of risk. Their intuition is to capture systemic risk by measuring the tail dependence between a single financial institution and the financial system. More formally, the hypothesis is that the stock returns of a systemically important financial institution shows asymptotically left-tail dependence with the returns of a diversified portfolio that represents the financial system, thus left-tail dependence is a proof of systemic risk (Kupiec and Guntay, 2016). A formal definition of left-tail dependence is provided by Kupiec and Guntay (2016). Let (R_1, R_2) be a bivariate random variable representing the returns on 1 and 2. The left tail dependence between 1 and 2 is defined as:

$$L(p) = P(r_1 < F_1^{-1}(p) \mid r_2 < F_2^{-1}(p)) \tag{1.8}$$

where p is a quantile and $F_j^{-1}(\cdot)$ is the inverse of the cumulative distribution function. The asymptotic left-tail dependence is instead defined as

$$\lim_{p \to 0} L(p)\,. \tag{1.9}$$

When the limit is zero, we are in presence of asymptotically independence in the left tail, while the bigger the value of the limit, the higher the degree of asymptotic dependence. Kupicec and Guntay (2016) find also that, if we assume normally distributed stock returns, the asymptotic tail dependence is equal to zero, prohibiting the presence of systemic risk: mainly for this reason, in empirical applications, it is better to assume that stock returns are not normally distributed.

Before defining CoVaR, we recall the definition of VaR. VaR of institution i is defined as the quantile q for which

$$P(R_i \le VaR_i^q) = q \tag{1.10}$$

Adrian and Brunnermeier (2016) defines $CoVaR^q_{S|C(X_i)}$ as the VaR of the financial system conditional on the institution i being in financial distress (or in its left-tail of the distribution). $CoVaR^q_{S|C(X_i)}$ is the q quantile of this conditional distribution for which

$$P(R_S \mid C(R_i) \le CoVaR^q_{S|C(R_i)}) = q\,. \tag{1.11}$$

Delta CoVaR measures the contribution that a single institution brings to the financial distress of the system and it is defined as

$$\Delta CoVaR^q_{S|i} = CoVaR^q_{S|R_i = VaR_i^q} - CoVaR^q_{S|R_i = VaR_i^{50}}\,. \tag{1.12}$$

It is simply the change in CoVaR of institution i when it shifts from its median normal condition (the 50% quantile) to its left-tail (popular values are 5% and 1% quantile).

In a first draft of their work, Adrian and Brunnermeier (2011) use the condition of institution i being exactly at its VaR level ($R_i = VaR_i^q$). This assumption simplifies too much the conditioning event, because it does not consider returns beyond the VaR level which can heavily influence the financial distress of institution i. For this reason, Girardi and Ergun (2013) propose a revised definition of Adrian and Brunnermeier's CoVaR, where the institution is at most at its VaR level ($R_i \leq VaR_i^q$), allowing for more severe losses. Mainik and Schaanning (2014) find that the revised definition is able to model the dependence between R_S and R_i better: they show that, when we use Gaussian or Student-t distributions for the returns, CoVaR is an increasing and continuous function of dependence only in the revised definition (though normality does not allow for asymptotically tail dependence as shown in Kupiec and Guntay (2016)).

CoVaR and Delta CoVaR have important properties. First, if we split an institution into n sub-institutions, the CoVaR of the single institution is equal to the CoVaR of the n sub-institutions. This means that if we condition VaR on a single big institution or on n small institutions that taken together have the same risk of the big one, the result should be the same. This property permits to CoVaR to identify both individual systemic institutions and small institutions systemic as a herd. Acharya et al. (2012) show an undesirable property of CoVaR under Gaussianity assumption: two firms with the same correlation with the system, but different volatilities are considered equally risky by CoVaR. In this case Delta CoVaR does not depend on the volatility of institution i:

$$\Delta CoVaR_i^q = \rho_{i,S} \sigma_S \Phi^{-1}(q) \tag{1.13}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal CDF. This property can create problems when we attempt to rank systemically important institutions by using Delta CoVaR.

The more popular way to estimate CoVaR is by quantile regression. In this case, CoVaR is the predicted value for the q quantile from the quantile regression of the system returns on the returns of institution i. A general convention is to use the negative returns, thus the losses have positive sign.

$$CoVaR_i^q = VaR^q_{S|R_i = VaR_i^q} = \hat{\alpha}_i^q + \hat{\beta}_i^q VaR_i^q \, . \tag{1.14}$$

The Delta CoVaR estimate is given by

$$\Delta CoVaR_i^q = CoVaR_i^q - CoVaR_{S|VaR_i^{50}}^q = \hat{\beta}_i^q (VaR_i^q - VaR_i^{50}) \tag{1.15}$$

This is the standard procedure if we want to estimates Delta CoVaR in a cross-sectional framework. Instead, if we want to estimate Delta CoVaR in a time series framework we need to add to the model some state variables to explain the evolution of the model over time. The procedure is as follows: first, we estimate the two quantile regressions

$$R_i^t = \alpha_i^t + \gamma_i^q \mathbf{M}_{t-1} + \varepsilon_i^{q,t} \tag{1.16}$$

$$R_{S|i}^t = \alpha_{S|i}^q + \gamma_{S|i}^q \mathbf{M}_{t-1} + \beta_{S|i}^q R_i^t + \varepsilon_{S|i}^{q,t} \tag{1.17}$$

where t denotes time and $\mathbf{M}_{t-1}$ is a set of lagged state variables.

Then the predicted values from the regressions are used in order to compute

$$VaR_i^{q,t} = \hat{\alpha}_i^q + \gamma_i^q \mathbf{M}_{t-1} \tag{1.18}$$

$$CoVaR_i^{q,t} = \hat{\alpha}_{S|i}^q + \hat{\gamma}_{S|i}^q \mathbf{M}_{t-1} + \hat{\beta}_{S|i}^q VaR_i^{q,t} \tag{1.19}$$

Finally, time-varying Delta CoVaR is

$$\Delta CoVaR_i^{q,t} = CoVaR_i^{q,t} - CoVaR_i^{50,t} = \hat{\beta}_{S|i}^q (VaR_i^{q,t} - VaR_i^{50,t}) \tag{1.20}$$

According to Adrian and Brunnermeier (2016), state variables must have the following characteristics; they must be capable of capturing the time variation in the conditional moments of assets returns and they must also be liquid and tractable.

Alternative ways of estimating CoVaR have been proposed. Girardi and Ergun (2013) propose a multivariate GARCH estimation of CoVaR that consists of three steps. In the first step we compute VaR for each institution using an univariate GARCH(1,1) model. In the second step, a GARCH-DCC model is estimated for the returns of each institution i and the financial system. In the third step, the Delta CoVaR is computed numerically. Both Adrian and Brunnermeier (2016) and Girardi and Ergun (2013) find that the two estimation methodologies (quantile regression and GARCH estimation) are consistent and provide more or less the same results.

There are also other alternatives methods to estimate CoVaR. Bernardi et al. (2013) propose a Bayesian approach using Markov Chain Monte Carlo, while Oh and Patton (2013) apply the copulas to the CDS spreads. It is also possible to make some distributional assumption and use maximum likelihood in order to obtain an estimate of

CoVaR: the preferred distribution to model the dependence between system and institutions is the Student-t distribution (Cao, 2014).

Since CoVaR is probably the most popular measure of systemic risk and it can be easily adopted for other uses, various extension have been proposed to the basic definition which are less important, because they are quite equivalent to other more studied systemic risk measures. Dollar Delta CoVaR is simply the Delta CoVaR multiplied by the size of the institution which is measured as the market capitalization:

$$\$\Delta CoVaR_{S|i}^{q} = \$MC_i * \Delta CoVaR_{S|i}^{q} . \tag{1.21}$$

This measure scaled by size allows for comparison of institutions with different market capitalizations (MC) or it can be used also to compare the risk of each division in a single institution, measuring each division using its size with respect to the single institution. Exposure CoVaR instead is simply the opposite of CoVaR; while CoVaR computes the VaR of the system, when a single financial institution is in distress, Exposure CoVaR computes the VaR of the single institution, when the system is in distress. By inverting the condition, we obtain a different measure and thus it is important to highlight that one institution can have a low CoVaR and an high Exposure CoVaR, but also the opposite is possible. This definition is very similar to the idea behind MES; in fact MES and Exposure CoVaR measure the same thing. Network CoVaR measure the VaR of an institution conditioning on another institution being in distress: it is a measure of the tail-dependence between two financial institutions. This measure is basically equivalent to the Co-Risk measure developed in IMF (2009); both can be estimated using quantile regression, but Co-Risk is estimated using the CDS spread for each financial institution. Conditional Expected Shortfall (CoES) is a logical extension of CoVaR, because it is subadditive and considers also losses beyond the VaR threshold. Another interesting extension is Multi CoVaR; Cao (2014) computes the VaR of the system conditional on a set of financial institutions being in financial distress. The intuition is that is more likely that several institutions, rather than only one as it is in the definition of CoVaR, experience financial distress at the same time.

Though very much criticized, CoVaR has laid the foundations for the work on systemic risk measurement based on the tail-dependence between institutions and system. In spite of the several limitations of CoVaR, it has gained a lot of attention and its effectiveness has been tested is several empirical works.

Co-Risk

The approach of CoVaR to measure systemic risk was first developed to capture the tail dependence between a single institution and the financial system. But the same approach can be used to capture the tail dependence of a particular institution to another institution. In IMF (2009), they develop a systemic risk measure which is aimed at assessing the dependence in the default risk of two institutions. As in Adrian and Brunnermeier's Delta CoVaR, they use quantile regression to capture the nonlinearity in the dependence between two market variables. Here, instead of market returns, CDS spreads are used.

The estimating procedure consists in first estimating the following model using quantile regression:

$$CDS_i = \alpha_q + \sum_{i}^{K} \beta_{q,t} R_i + \beta_{q,t} CDS_j \tag{1.22}$$

where CDS_i and CDS_j are the CDS spreads of institution i and j respectively and R_i is a set of state variables that control for common risks factors such as the business cycle and the market volatility. Usually, the 95% quantile is considered for the regression. Once we have estimated the model, the Co-Risk measure is defined as

$$Co-Risk_{i,j} = 100*(\frac{\hat{\alpha}_{95} + \sum_{i}^{K} \hat{\beta}_{95,i} R_i + \hat{\beta}_{95,i} CDS_{95,j}}{CDS_{95,i}} - 1)\,. \tag{1.23}$$

The Co-Risk approach is very similar to the CoVaR one, because both use quantile regression to capture the tail dependence between two market variables. Here the CDS spreads are used instead of the stock returns. This means that we are considering a different type of information, coming from a different market. The stock returns consider information coming from the stock market, while CDS spreads consider information coming from the bond market or more in general fixed-income market and so it is possible that the two approaches are not equivalent.

Co-Risk can be interpreted as the increase in the credit risk of an institution, conditional on the event that another institution is experiencing financial distress (because its CDS spread is above the 95% percentile). It represents also a measure of the link that there is between two financial institutions: if institution i is characterized by an high Co-Risk measure conditioning on institution j, this means that a default of institution j is likely to trigger also the default of institution i. Therefore, Co-Risk can be used to identify SIFIs as those institutions which default will cause a cascade of other defaults of other institutions.

In the final remarks, the IMF (2009) suggests that the degree of interconnectedness of an institution with respect to the other institutions, measured by its Co-Risk, can be used to impose capital surcharges to those institutions whose failure will trigger the failure of other institutions. This approach will create an incentive for the SIFIs to cut their link to other financial institutions by, for example, reducing the counterparty exposure or by holding different portfolios. If the financial firms are loosely linked to each other, they are less systemic as a herd and so they make the system less vulnerable.

MES

Another popular systemic risk measure is the Marginal Expected Shortfall (MES) proposed by Acharya et al. (2017). We start by reviewing the definition of Expected Shortfall (ES). At the $1-q$ confidence level, ES is the expected value of the loss for institution i, when it exceeds its VaR level. Using negative returns R_i, we obtain:

$$ES_i^q = E(R_i \mid R_i \geq VaR_i^q) \tag{1.24}$$

which can be simply calculated by averaging the historical returns that are beyond the VaR threshold.

As a risk measure, ES is superior to VaR for mainly two reasons. First, VaR does not consider losses beyond the q quantile, which can be quite problematic in cases where there could be large losses with small probabilities which is very common when we deal with systemic risk. For example, let us consider a trade that will give a gain of 100 in 99,99% of the cases and a loss of 100,000,000 in 0.01% of the cases. If one uses the usual 95% and 99% VaR to evaluate the risk of this trade he or she will turn out with a value equal to 100, ignoring the possible losses beyond VaR threshold. Therefore, using VaR creates an incentive for the bank to get involved in very asymmetric bets, because it can profit more, without exceeding the parameters of risk imposed by the regulators. A practical example could be a financial institution writing deep out-of-the-money call options which bring a small gain most of the time but if something goes wrong they can theoretically produce an infinite loss (Hull, 2014). The paradox of VaR is that instead of being useful to mitigate risk, it creates new risks, because of the incentive we described. This scenario is avoidable by simply using ES. The second important aspect is that ES is a coherent risk measure while VaR is not, because it lacks the subadditivity property: the VaR of the sum of two portfolios could be higher than the sum of their individual VaRs, while this is not possible if we use ES. In spite of this critical problems, VaR represents the main framework for regulation, but there are plans to move to ES with Basel III in the next years.

The idea behind MES is that the risk of the financial system S as a whole can be decomposed in the sum of the risks brought by each single institution i:

$$ES_S^q = \sum_i W_i * E(R_i \mid R_S \geq VaR_S^q) \tag{1.25}$$

where the variable W_i represents the weight of the institution i in the system and can be approximated by the market capitalization of firm i over the capitalization of the system. At the $1-q$ confidence level, MES for institution i is defined as the first derivative of the ES of the system with respect to W_i:

$$MES_i^q = \frac{\delta ES_S^q}{\delta W_i} = E(R_i \mid R_S \geq VaR_S^q)\,. \tag{1.26}$$

In other words, MES is the estimated loss of institution i, when the system is in its left tail (or right if we consider negative returns). MES can be meaningful even if the firm we are considering is not included in the portfolio: in this case MES should be interpreted as a measure of sensitivity of the firm to extreme bad events in the financial system (Idier et al., 2014).

Despite the fact that MES is universally accepted as a systemic risk measure, the way in which is commonly calculated could be a problem. We define a systemic event as an event that happens once or twice in a decade, a very extreme tail event. Instead, when we compute MES, we usually employ the 5% worse daily returns of the system which, by definition, are a one-in-a-month event (0.05*20=1). This means that we are using not so extreme observations to compute a measure that should capture very extreme and rare events. For this reason, Acharya et al. (2017) propose to use extreme value theory to link the moderately bad days employed in MES calculation to the very extreme tail that represents the true crisis. They specify the returns on security j for the bank i as

$$r_j^i = \eta_j^i - \delta_{i,j}\varepsilon_j^i - \beta_{i,j}\varepsilon_m \tag{1.27}$$

where η_j^i follows a Gaussian distribution and explains the variations in normal days while ε_j^i and ε_m follow independent normalized power law distributions characterized by tail index ζ and they explain the events in the extreme tail of the distribution. Gabaix (2009) demonstrates the following properties:

$$VaR_{i,j}^q \approx (\delta_{i,j}^\zeta + \beta_{i,j}^\zeta)^{1/\zeta} * q^{-1/\zeta}, \tag{1.28}$$

$$ES_{i,j}^q \approx \frac{\zeta}{\zeta - 1} VaR_{i,j}^q. \tag{1.29}$$

Then, the 5% tail event corresponds to the critical values $\bar{\varepsilon}_m^{\%}$ and $\bar{\varepsilon}_m^S$ of the systemic shock ε_m. The severity of the shock is defined as $k = \frac{\bar{\varepsilon}_m^S}{\bar{\varepsilon}_m^{\%}}$. Given these premises, the authors define the Systemic Expected Shortfall (SES) as

$$\frac{SES^i}{w_0^i} = (\frac{za^i}{w_0^i} - 1) + kMES_i^{0.05} + \Delta^i \tag{1.30}$$

where w_0^i is the equity (market capitalization) at time zero, a^i is the value of the assets, z is a percentage that ranges between 8% and 12% depending on the assets held by the institution and Δ^i is defined as

$$\Delta^i = \frac{E(\phi^i \mid W_1 < zA) - k * E(\phi^i \mid I_{0.05})}{w_0^i} - \frac{(k-1)(f^i - f^b)}{w_0^i}. \tag{1.31}$$

The SES is basically a transformation of MES that takes into account:

- the degree of undercapitalization of the firm $za^i / w_0^i - 1$;
- the MES of the firm amplified by a the factor k which takes into account the performance during a real crisis;
- the adjustment term Δ^i which measures the excess cost of the financial distress.

But, in practice, they estimate SES using the following formula:

$$\frac{SES^i}{w_0^i} = \frac{za^i}{w_0^i} - 1 - MES_i^{0.05} \tag{1.32}$$

which is just a simplification of the original theoretical framework because it considers $k=1$ and $\Delta^i = 0$.

Many scholars extended the literature by developing measures based on MES. Banulescu and Dumitrescu (2015) propose an absolute systemic risk measure called Component Expected Shortfall (CES). The CES is simply the portion of the ES of the system due to the *i*-th firm

$$CES_i = W_i \frac{\delta ES_S^q}{\delta W_i} \tag{1.33}$$

which is the product between the MES of firm i and its weight in the system. A larger CES implies a larger contribution to the systemic risk of the system. Of course, the sum of all the CESs is equal to the ES of the system. The size of the firm is undoubtedly an important component of systemic risk, because it is linked to the 'too-big-to-fail' paradigm. In its basic form, MES does not take into account it, while CES does. Another important extension of MES is the SRISK measure, proposed by Brownlees and Engle (2017) which is simply a nonlinear transformation of MES and we present it in the next section.

SRISK

What differentiates SRISK from the other systemic risk measures based on market data is that it incorporates information both from the markets (stock prices) and from balance sheets. Since SRISK takes into account more information, it should have an advantage compared to the other approaches in measuring systemic risk.

Moreover, SRISK is aimed at measuring the possible capital shortfall for a financial firm in presence of a prolonged market decline. In this definition we have a slightly change in the conditional event. When we define CoVaR and MES, we use as conditional events the 1% or 5% worst results from the historical distribution; instead in this case the conditioning event is a prolonged market decline. Brownlees and Engle (2017) suggest a market decline below a value C, in a time span h. For example a drop of 40% in S&P 500 in a period of 6 months (126 trading days). These are very different conditional events: a loss in the 5% quantile happens one in a month, while a 40% drop in the stock market is a one-in-a-decade event. For these reasons, CoVaR and MES are criticized, because they measure systemic risk using not so extreme observations which could not reflect a systemic event. SRISK tries to fill this gap by using the Long Run Marginal Expected Shortfall (LRMES).

First of all, we define a capital shortfall (CS) as the difference between the reserve of capital a firm needs to hold due to regulation and its equity:

$$CS_i = kA_i - MC_i = k(D_i + MC_i) - MC_i \tag{1.34}$$

where MC_i is the market value of equity, D_i is the book value of debt, A_i is the quasi value of assets and k is the prudential capital fraction. SRISK proposed by Brownlees and Engle (2017) is the expected capital shortfall conditional on a systemic event:

$$SRISK_i = E(CS_i \mid R_S > C) = kE(D_i \mid R_S > C) - (1-k)E(MC_i \mid R_S > C) \tag{1.35}$$

where R_S are the negative log-returns of the system. In order to compute the expected value we need to assume that the book value of debt is constant throughout the period of decline of the market. In other words, a financial institution cannot modify or renegotiate its debt structure during the crisis. This might be a very strong assumption but it is necessary in order to going forward with the calculations. Said that, we obtain

$$SRISK_i = kD_i - (1-k)MC_i(1 - LRMES_i)\,. \tag{1.36}$$

SRISK is simply a function of LRMES (which will be analysed later), capital requirement ratio k and leverage of the institution. The capital requirement ratio is an outside factor which can be fixed by the regulators. The authors choose to use a capital ratio based on the one maintained by well managed large financial institutions in normal times. The leverage of a financial institution is approximated using the quasi-leverage ratio which is $(D_i + MC_i)/MC_i$. Assessing the leverage of a financial institution during a crisis can be an hard task, even if we use more detailed information. Let us consider this real example: according to a bankruptcy report, *'Lehman employed off-balance-sheet devices, known within Lehman as Repo 105 and Repo 108 transactions, to remove securities inventory from its balance sheet, usually for a period of seven to ten days, and to create materially misleading picture of the firm's financial condition in late 2007 and*

2008' (Valukas, 2010, p. 732). In this case, if one trusted balance sheet information, it would obtain a biased picture of the leverage of a financial institution.

If we aggregate the SRISKs of all firms composing the financial system

$$SRISK_S = \sum_{i=1}^{N} (SRISK_i)^+ \tag{1.37}$$

where we consider as 0 the contribution of firms that have a negative SRISK, we obtain a measure of the total amount of money required to bail out the system, if a systemic event happened. The SRISK of the single institution i over the total SRISK of the system represent the percentage contribution of the firm to the total risk:

$$SRISK\%_i = \frac{SRISK_i}{SRISK_S}. \tag{1.38}$$

It is also interesting to decompose SRISK in order to gain insight on what causes changes in the systemic risk measure. Directly from the formula of SRISK, we can obtain its total differential as

$$\Delta SRISK_i = k * dD_i - (1-k)(1-LRMES_i) * dMC_i + (1+k) * MC_i * dLRMES_i. \tag{1.39}$$

Therefore, the fluctuations in SRISK can be attributed to changes in three factors:

- $\Delta D_i = k * dD_i$: the first factor that influences SRISK is the debt that the institution is taking. If the burden of debt increases, so will do SRISK. Since debt is updated only once a quarter, changes in SRISK will show up later;
- $\Delta MC_i = -(1-k)(1-LRMES_i) * dMC_i$: the second factor that affects SRISK is the market capitalization. When the market capitalization is increasing, SRISK will decline, because the firm is better capitalized and thus the leverage will be lower. Instead when the market capitalization is falling, the leverage will increase, making the firm more risky. SRISK is very sensitive to changes in the equity value, because it can be updated every day since it is market data;
- $\Delta RISK_i = (1-k) * MC_i * dLRMES_i$: the third factor is the risk of the firm measured by its LRMES. Of course, an increase in the LRMES will increase SRISK. The LRMES can increase because of an increase in the correlation between the firm and the market or because of an increase in the volatility of the firm.

The more challenging issue in computing SRISK is estimating the LRMES. There are various approaches to estimate the LRMES which can yield different results. The more simple is the one applied by the V-Lab of the New York University in order to build a ranking of global financial institutions. In their application they compute LRMES directly using the formula

$$LRMES = 1 - \exp(\log(1-d) * \beta) \tag{1.40}$$

where β is the dynamic conditional beta (Engle, 2016) of the institution and d is the drop in the market in a six-month period which is assumed to be 40%. Similarly, Acharya et al. (2012) compute LRMES as

$$LRMES = 1 - \exp(-18 * MES)\,. \tag{1.41}$$

In this case, the LRMES is a nonlinear transformation of the daily MES of an institution. The last approach proposed by Brownlees and Engle (2017) computes predictions of LRMES using a GARCH-DCC model (Engle, 2002). In particular, this approach requires to specify the equations for the evolution of the time varying volatilities using a GJR-GARCH model (Glosten et al., 1993) and the correlations using a the DCC correlation model. First of all, without specifying any particular distribution, we define the returns of institution i and the system S as

$$\begin{pmatrix} r_{i,t} \\ r_{S,t} \end{pmatrix} | F_{t\text{-}1} \sim D(\mathbf{0}, \begin{bmatrix} \sigma_{i,t}^2 & \rho_{i,t}\sigma_{i,t}\sigma_{S,t} \\ \rho_{i,t}\sigma_{i,t}\sigma_{S,t} & \sigma_{S,t}^2 \end{bmatrix}) \tag{1.42}$$

The GJR-GARCH model equations for the volatility of the institution i and the system S are

$$\begin{aligned} \sigma_{i,t}^2 &= \varpi_{V_i} + \alpha_{V_i} r_{i,t-1}^2 + \gamma_{V_i} r_{i,t-1}^2 I_{i,t-1}^- + \beta_{V_i}\sigma_{i,t-1}^2 \\ \sigma_{S,t}^2 &= \varpi_{V_S} + \alpha_{V_S} r_{S,t-1}^2 + \gamma_{V_S} r_{S,t-1}^2 I_{S,t-1}^- + \beta_{V_S}\sigma_{S,t-1}^2 \end{aligned} \tag{1.43}$$

where $I_{i,t}^- = 1$ if $\{r_{i,t} < 0\}$ and $I_{S,t}^- = 1$ if $\{r_{S,t} < 0\}$. The advantage of the GJR-GARCH model with respect to a basic GARCH(1,1) model is that it takes into account the leverage effect. Given the adjusted returns, $\varepsilon_{i,t} = r_{i,t} / \sigma_{i,t}$ and $\varepsilon_{S,t} = r_{S,t} / \sigma_{S,t}$, the correlation between institution and system is modelled by the following DCC component:

$$Corr\begin{pmatrix}\varepsilon_{i,t}\\ \varepsilon_{S,t}\end{pmatrix}=R_t=\begin{bmatrix}1 & \rho_{i,t}\\ \rho_{S,t} & 1\end{bmatrix}=diag(Q_{i,t})^{-1/2}Q_{i,t}diag(Q_{i,t})^{-1/2} \tag{1.44}$$

where $Q_{i,t}$ is called the pseudo-correlation matrix and it is specified as

$$Q_{i,t}=(1-\alpha_{C_i}-\beta_{C_i})S_i+\alpha_{C_i}\begin{bmatrix}\varepsilon_{i,t-1}\\ \varepsilon_{S,t-1}\end{bmatrix}\begin{bmatrix}\varepsilon_{i,t-1}\\ \varepsilon_{S,t-1}\end{bmatrix}'+\beta_{C_i}Q_{i,t-1} \tag{1.45}$$

where S_i is the unconditional correlation matrix of the firm and system adjusted returns. The estimation can be performed using a two-step QML estimation procedure (Engle, 2009). There are different ways to estimate the LRMES using the GARCH-DCC model. The more sophisticated is to simulate N future paths for the firm and market returns of length h, given the available information. Then the *dynamic* LRMES is the Monte Carlo average of the cumulated returns of each path of the institution returns, when the market is down by 40%. If instead we assume a static bivariate normal framework, the *static* LRMES can by approximated by

$$LRMES_{i,t}=-\sqrt{h}\beta_i E(r_{S,t+1}\mid r_{S,t+1}<c) \tag{1.46}$$

where $\beta_i=\rho_i\dfrac{\sigma i}{\sigma_S}$, $E(r_{S,t+1}\mid r_{S,t+1}<c)=-\sigma_m\dfrac{\phi(c/\sigma_S)}{\Phi(c/\sigma_S)}$ and $c=\log(1+C)/h$. $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the pdf and the CDF of the standard normal distribution.

SRISK is a step forward compared to the other market data based risk measures, because it incorporates a measure of the leverage of the firm and the required capital ratio. But this advantage does come at a cost: we have to assume that the liability structure of the firm during the crisis is unchanged. This means that during the stressed period, the firm cannot renegotiate its debt and its value is constant. But even if we believe that this is what happens in reality, we will never be sure about the true value of the liabilities of the institution. The book value of debt could not be very informative about the liability structure of the firm: balance sheets have been manipulated in the past and will be in the future, thus there is no way of avoiding this problem, unless we are inside the firm and we know what is going on. Another issue is related to the concept of causality. SRISK, but also all the other systemic risk measures we have described, are simply an estimation of one endogenous variable conditioned on another one, without any implication of causality. Acharya et al. (2012) pose the following question: *'are firms weak because of the crisis, or does the crisis happen because the firms are weak?*. Although they argue that both statements are true, because there cannot be a financial crisis without weak firms, this issue is not addressed by this type of measures, because they are simply a test for weakness of individual institutions as a function of market weakness, without any implication of causality.

As we said before, SRISK is the capital an institution will need to raise during a financial crisis in order to reach its capital target ratio. For this reason, SRISK can be thought as an alternative approach to the regulatory stress tests performed by the Fed and the ECB (Acharya et al., 2014): they both try to achieve the same object by using different paths. But there are significant differences between the two paths. While stress tests consider the unconditional failure of a firm, the SRISK methodology is focused on the failure of an institution, when the system as a whole is already undercapitalized. Also the scenario used in the two approaches is different; in the V-Lab approach the only factor in the scenario is a drop of the stock market by 40% in a six-month period; instead central banks use a multitude of macroeconomic and financial factors to build their scenarios. For example, in 2012 the Fed used 25 factors, while the ECB employed more than 70. Also the type and amount of data required is different: V-Lab stress test uses only stock market prices, historical market capitalization and leverage computed using balance sheet information; central banks instead can count on confidential information provided directly by the financial institutions such as capital composition, portfolios of securities and loans, exposures to counterparties and historical profits and losses and the related forecasts. Finally, V-Lab stress test can be performed on a daily basis at a cost approximately equal to zero, while central banks need months in order to obtain the results. Despite the common purpose, Acharya et al. (2014) find that V-Lab methodology and stress tests give very different results. First, the V-Lab method predicts always bigger losses than the ones predicted by the stress tests. The magnitude of the difference cannot be justified by any difference in the methodologies: for example in the European case, the V-Lab net loss is ten times bigger than the one of the stress test. The rankings of the two competing approaches show not significant correlation at the 5% level. And the situation is again critical in Europe, where two institutions, Credit Agricole and Dexia, which are the riskiest using the V-Lab stress test are considered the safest for the stress test of the European Banking Authority. Given these results, a few remarks are necessary. First, it is not possible that both the V-Lab and the stress test ranking are right, because of the conflicting results; it is possible that both are wrong, but this is quite unlikely. Said that, the procedure based on SRISK maybe could be imprecise but it is conceptually consistent, because it has solid theoretical foundations. On the other hand, the stress tests are very complex and their results may be subject to how 'hard' the stress test is set. For example, it is well known that due to political reasons the stress test in Europe considers a mild stress scenario and in fact only a few institutions failed the test. Both methods have their advantages and weaknesses and, as advocated by Lucas (2014), traditional stress tests and new measures based on market data should be viewed more as complements than substitutes, at least until a consensus about the way to base the new regulation is reached.

1.2.2 Other Systemic Risk Measures

For the sake of completeness, in this paragraph we will briefly present other systemic risk measures besides the ones of the market data based approach developed in the previous paragraph. There are alternative ways in which systemic risk measures can be built. For example, systemic risk measures can be based on credit risk techniques, market-implied losses, connectedness and macroeconomic conditions. Zhou (2010), using a multivariate framework and extreme value theory (EVT), proposes two systemic risk indexes: a systemic impact index which captures the risk that an institution represents for the system; and a vulnerability index which indicates the risk that the system imposes on a single financial institution. This approach is comparable to the conditioning approach of CoVaR and MES. Segoviano and Goodhart (2009) focus on the contribution of a single bank to the distress of the financial system looking at probability of distress (PoDs) in a multivariate framework. Huang et al. (2009) measure systemic risk as the price of insurance against financial distress which is based also in this case on probability of default and correlations between financial institutions. Allen et al. (2012) introduce CATFIN that is a measure of systemic risk obtained by averaging VaRs and ESs estimated using different parametric and non-parametric methods. They find that this measure is able to predict macroeconomic downturns. Alessi and Detken (2011) build an early warning indicator for asset bubbles based on real and financial data. They find that measures of liquidity, in particular private credit gap at a global level is the best indicator and it could be used by regulators to intervene promptly to avoid financial imbalances. In a similar way, Borio et Drehmann (2009) study credit and asset prices to develop leading indicators for banking crisis. Tarashev et al. (2010) apply game theory to measure systemic risk attributions: in this case systemic risk is measured using Shapley value methodology. This approach is extended in Drehmann and Tarashev (2013). Jobst and Gray (2013) apply the contingent claim analysis to measure systemic risk. Suh (2012) extends existing correlated default models for measuring systemic risk. Bluhm and Krahnen (2014) introduce system wide Value at Risk (SVaR) as a new risk management tool useful to design macro-prudential capital surcharges. Finally, Billio et al. (2012) apply principal component analysis and Granger-causality tests and propose econometric measures of systemic risk aimed at capturing the interconnectedness among the returns of different categories of financial institutions (banks, brokers, insurance companies and hedge funds).

1.3 Systemic Risk Measures and Regulation

This paragraph deals with what should be the final destination of systemic risk measures: regulation or, in other words, preserving the financial stability. The current financial regulation such as the Basel capital requirements is aimed to limit the risk that an institution can take. This is very important, because financial firms have the incentive to take risks which consequences are borne by everyone, since they do not internalize the external cost that a loss would have on the real economy. But in order to do that, the regulation employs risk measures that consider the risk of the institution in isolation and that are not able to capture the important linkage between the single institution and the financial system. For this reason, Acharya et al. (2017) suggest that the financial regulation should focus on limiting the systemic risk and not the risk of each single institution viewed in isolation. They recommend that the new regulation should be based on the new systemic risk measures. Also Adrian and Brunnermeier (2016) suggest that the *forward*-Delta CoVaR should be implemented in a macro-prudential policy. And Acharya et al. (2014) propose to use SRISK instead of the macro-prudential stress tests employed by central banks. Benoit et al. (2016) identify the capital surcharge for SIFIs as the main application of systemic risk measures.

In the current regulation, we can specify three types of institutions group: G-SIBs, SIFIs and D-SIBs. G-SIBs stands for Global Systemically Important Banks, SIFIs stands for Systemically Important Financial Institutions and D-SIBs for Domestic Systemically Important Banks. The Basel Committee identifies and ranks G-SIBs using a scoring methodology. G-SIBs are required to keep an additional equity capital depending on the category in which they are associated to. However, a bank that is not part of the G-SIBs, can be classified as a D-SIB by the national regulator. This means that it can be subject to additional capital requirements in addition to the ones imposed by the Basel Committee. In this context, the systemic risk measures can be useful in identifying SIFIs and impose on them the fair capital surcharge that preserves the global financial stability.

Before going forward, it is useful to present a unified theoretical framework proposed by Benoit et al. (2016) that highlights the additional information that systemic risk measures are able to provide in comparison to the basic risk measures used today. This can be useful in order to understand if the additional information can lead to a better regulation. First of all, we assume a bivariate GARCH process for the firm and system returns $r_{i,t}$ and $r_{S,t}$ at time t. The related conditional variance-covariance matrix is defined as

$$H_t = \begin{pmatrix} \sigma_{S,t}^2 & \rho_{i,t}\sigma_{i,t}\sigma_{S,t} \\ \rho_{i,t}\sigma_{i,t}\sigma_{S,t} & \sigma_{i,t}^2 \end{pmatrix} \tag{1.47}$$

where $\sigma_{i,t}$ and $\sigma_{S,t}$ are the conditional standard deviations for the firm and system returns, while $\rho_{i,t}$ is the conditional correlation between the returns of the firm and the system.

Benoit et al. (2013) show that Delta CoVaR is proportional to the tail risk of the institution measured by its VaR:

$$\Delta CoVaR_{i,t}^q = \gamma_{i,t}(VaR_{i,t}^q - VaR_{i,t}^{0.5}) \tag{1.48}$$

where $\gamma_{i,t} = \rho_{i,t}\sigma_{S,t} / \sigma_{i,t}$ which is called the proportionality coefficient. This coefficient is specific for each institution and this fact explains why Adrian and Brunnermeier (2016) find a lot of difference between Delta CoVaR and VaR in the cross-section. Nevertheless, if we keep the coefficient constant over time, Delta CoVaR becomes proportional to VaR and thus forecasting its systemic risk contribution is equivalent to forecasting its tail risk.

The same it is also true for MES. Benoit et al. (2013) show that MES can be written as

$$MES^{q}_{i,t} = \beta_{i,t} ES^{q}_{S,t} \tag{1.49}$$

where $\beta_{i,t} = \rho_{i,t}\sigma_{i,t} / \sigma_{m,t}$ is the time-varying beta of the institution. Also in this case our systemic risk measure is proportional to a more general measure of market risk. This means that a ranking of the financial institutions based on MES will give the same result of a ranking based on their betas. This means that the systematic importance of an institution increases with its beta. But in this case, forecasting the systemic risk of an institution is not equivalent to forecasting its systematic risk, because the ES of the system varies over time.

Since SRISK is a function of MES, also this systemic risk measure can be expressed as a function of its beta, in addition to its leverage and capital ratio. SRISK can be written as

$$SRISK_{i,t} = k(L_{i,t} - 1) + (1-k)\exp(18 * \beta_{i,t} * ES^{q}_{S,t}) * MC_{i,t} \tag{1.50}$$

where k is the required capital ratio, $L_{i,t}$ is the quasi-leverage ratio and $MC_{i,t}$ is the market capitalization. In this case the rankings based on SRISK will be not equivalent to those based on the beta, because there are also other factors that are firm specific such as the leverage ratio and the market capitalization. The SRISK is the only measure of systemic risk that takes into account both the too-big-to-fail paradigm (through the market capitalization) and the too-interconnected-to-fail one (through the MES).

This general theoretical framework shows that there is a strong link between the new systemic risk measures and the traditional measures of risk such as beta and correlation coefficient. From the point of view of the regulator, this link must be taken into consideration. If for example MES is implemented in a regulation, we have to be aware that it will give results very similar to those of beta. Or that Delta CoVaR will give the same results as VaR if we keep the proportionality coefficient constant over time.

Despite the great work done to develop systemic risk measures based on market data, very small effort has been put in researches that link these measures to practical tools of regulatory intervention. If one institution is identified to be more systemic, there is still no direct response in order to mitigate this possible problematic position of the institution. A basic regulatory response could be that systemic financial institutions must internalise the externalities that they can create on the real economy when they take too much risk: this is possible by developing a tax system based on the systemic risk contribution of each institution. Acharya at al. (2017) propose a model where the taxation on banks depends on their level of systemic risk. Therefore, a systemic bank that is more likely to be bailed out by the government with the money of the taxpayers will pay higher taxes in order to contribute more to a possible bail out. This could be one of the possible implementations of systemic risk measurements in a regulatory framework.

However, there are mainly two problems that must be addressed in order to make these measures really useful for regulation policies: procyclicality and volatility of the measure. Procyclicality is due to the well-known 'volatility paradox' developed by Adrian and Boyarchenko (2012) for which periods of low volatility tends to be associated to an increase in leverage which leads to an increase in systemic risk and thus a possible crisis. In other words, this means that systemic risk increases in periods where the volatility is low and then explodes during the crisis. Therefore any financial regulation that relies on contemporaneous market data will be biased, because it will be too loose when the volatility is low and too tight when the volatility is high (Adrian and Brunnermeier, 2016).

The volatility paradox and the pro-cyclicality of these measures must be taken into account in a system as the one proposed by Acharya et al. (2017), because otherwise the financial institutions would pay lower taxes during normal periods (before a crisis) and then they would pay higher taxes when the crisis explodes and the additional tax will only make the situation worse. This is true also for SRISK that it will ask for higher capital requirements during the crisis, when the bank is already in a stressed situation. The capital requirement would be also much more volatile during periods of crisis and thus SRISK will ask for more frequent adjustment of the capital. Lucas (2014) proposes to incorporate a smoothing rule in order to reduce the volatility of these measures. This will for sure create a loss on the information provided by the statistical measure, but it will make easy to implement a policy based on these measures.

In order to resolve the problem of procyclicality, Adrian and Brunnermeier (2016) regress their risk measure on lagged explanatory variables and they call it *forward*-Delta CoVaR. First of all, they regress Dollar Delta CoVaR on institution characteristics and macroeconomic variables:

$$\$\Delta CoVaR_{i,t}^{q} = a + \mathbf{c}\mathbf{M}_{\mathbf{t\text{-}h}} + \mathbf{b}\mathbf{X}_{\mathbf{t\text{-}h}} + \eta_{i,t} \tag{1.51}$$

where $\mathbf{M}_{\mathbf{t\text{-}h}}$ is the vector containing the macroeconomic variables and $\mathbf{X}_{\mathbf{t\text{-}h}}$ is the vector of the firm characteristics, lagged of $h = 1,4,8$ quarters. Then the *forward*-Delta CoVaR is obtained as the predicted value of the regression using lagged variables:

$$Fwd\Delta CoVaR_{i,t}^{q} = \hat{a} + \hat{\mathbf{c}}\mathbf{M}_{\mathbf{t\text{-}h}} + \hat{\mathbf{b}}\mathbf{X}_{\mathbf{t\text{-}h}}. \tag{1.52}$$

The lagged explanatory variables are firm characteristic and macroeconomic variables that should have an economic meaning and have the ability to predict the future values of Delta CoVaR. Adrian and Brunnermeier (2016) use leverage, maturity mismatch, market capitalization and the number of consecutive quarters in the top decile of the market-to-book ratio as firm variables; then they use the change in the three-month yield, the change in the slope of the yield curve, the TED spread, the change in credit spread, the market return, the real estate sector return and the equity volatility as macroeconomic variables.

This procedure allows to avoid the procyclicality of the systemic risk measures, but at a high cost: we are no more computing a measure directly, but we are trying to predict this risk measure using other variables that can have in some sense a predicting ability of our risk measure using a regression. The results of the regression shows that the adjusted R^2 is roughly 25%, thus the explanatory variables are able to explain only one fourth of the variability of Delta CoVaR. Increasing the number of variables is possible but then we may obtain a better result just because of overfitting the data.

In this paragraph we have discussed the possible implementations of systemic risk measures in a regulatory framework. Despite the huge work done in developing and testing those types of measures, very few researches discuss how these measures can be possibly used by regulators or other policy makers such as central banks to develop macro and micro prevention policies. Probably the main implementation of these risk measures is the determination of capital for systemically important financial institutions.

Therefore the future research should focus on two paths. The former is try to improve the existing systemic risk measures. A good way to do this could be by using market and balance sheet data together (as done with SRISK): in this way we exploit more information and our measure will be more precise. Of course this does not mean that we achieve a perfect risk measure, because it will be always based on some more or less realistic assumptions. The assumptions are an unavoidable component of every model, we have only to check that they are not too much unrealistic. For example, if we create a risk measure which assumes a Gaussian distribution for the returns, this is very unrealistic given all the proofs that we have against this hypothesis. If instead the risk measure assumes that the returns are modelled using a GARCH-DCC model, even if also in this case we have assumptions, they are much less restrictive and can better represent the stylized fact of financial returns. The second research path that should be developed is the link between these systemic risk measures and possible policies. This path is still unexplored and it is worth develop it to understand if the implementation of these measures of risk can make a step forward in the current financial regulation.

Section II: How Accurate Are Systemic Risk Measures?

In this section we will present and develop our application of systemic risk measures and we will try to interpret the main findings and compare them to the current literature. First of all, we will present the data we used and the methodology we are following. Then, we move to the analysis of the results for each systemic risk measure, starting from Delta CoVaR, then moving to MES and finally to SRISK.

2.1 The Data

In this paragraph we present the data we use for our application. We decide to focus on two main geographical areas: Europe and the United States. Since they represent advanced financial systems, compared to other parts of the world, they are characterized by more easily accessible data and the data themselves are more reliable. Nevertheless, this procedure can be applied to every national or transnational financial system in the world.

Since we are studying the market data based approach to systemic risk measures, we use mainly the historical adjusted prices of each financial institution downloaded from Yahoo! Finance. Usually, a financial institution is listed in more than one stock exchange in the world, so it trades simultaneously in different places. In this case, we choose the stock exchange that is located in the country where there is the headquarter of the financial institution. See Appendix A for a list of the financial institutions considered.

As a proxy for the financial system we use the iShares MSCI Europe Financials (EUFN) for Europe and the iShares Dow Jones US Financial (ETF) for the United States. Other applications in the literature use the CRSP value-weight or equally-weight index for the financial system, but this is not really relevant for the results. For example, in the computations of SRISK, the V-Lab does not use any financial system index but uses directly the S&P 500 index, as a proxy of the market.

As standard in the literature, we compute the daily log-returns from the daily adjusted prices as

$$100*(\log(P_t)-\log(P_{t-1})). \tag{2.1}$$

Usually in risk management, it is more convenient to compute the negative log-returns, since we are more interested in the possible losses. In this case negative returns (losses) will have a positive sign, while positive returns (gains) will have negative sign. Using this reference, we compute

$$-100*(\log(P_t)-\log(P_{t-1})). \tag{2.2}$$

Therefore, each risk measure that represents a loss (e.g. VaR, ES, CoVaR, MES and SRISK) comes with positive sign.

In order to compute SRISK, we need also the book value of debt and the market capitalization that can be found also in this case on Yahoo! Finance or Google Finance. In the application regarding the United States, we do not have problems of different currencies, because all the values are in US Dollars (USD). Instead in the European case, we have to take into consideration the fact that the firms that are listed in London have balance sheets and market capitalizations in British Pound (GBP) and not in Euro (EUR). For this reason, we decide to convert everything in Euro using the mean between the bid and the ask mean values of the exchange GBP/EUR in 2016 as the value of the conversion.

Another problem that can arise when using prices that come from different stock exchanges in different countries is that of asynchronous prices. This happens because stock exchanges in different countries have different hours in which they trade. This means that prices related to the same date are not taking into consideration the same amount of information. For example, let us assume that a stock A trades at exchange 1 which closes at 4 p.m., while stock B trades at exchange 2 which closes at 5 p.m.. If an important news arrives at the market at 4:30 p.m., it will be discounted in the price of stock B, but this will not happen for stock A, because the exchange is already closed: the price will incorporate this news only in the following day. This may be a problem when we analyse a global portfolio which has stocks that trade in Tokyo, London and New York, because they are overlapped only for a small portion of the day. In our application, we are considering only Europe and United States, separately. Therefore, we think that this problem is totally negligible for the United States and can be neglected for Europe, since the exchange hours are overlapped for the vast majority of the day.

2.2 Methodology

A lot of effort has been put on the development and application of systemic risk measures and they are becoming quite popular. A lot of problems have been addressed and resolved, but we think there is more than one reason to be concerned about the precision of these risk measures; we are going to assess the precision of these systemic risk measures and the order of magnitude of the estimation error. When we are dealing with some kind of risk, we are always dealing also with an estimation error, because the risk is not directly observable and it must be estimated using some statistical or econometric methods (Danielsson et al., 2016). The magnitude of the estimation error depends on a lot of factors: the number of observations available in our sample, the presence of outliers in the sample and, of course, the method we are using to estimate our quantity of interest. We should be aware of the fact that the estimate we obtain from our model is just a random variable, because if we observed another sample, we would obtain another estimate. Therefore, the value we observe is just selected at random from many others possible values and thus it can be considered a random variable (Ruppert, 2011). There are popular statistical techniques such as the *bootstrap* to deal with this problem.

The problem of the uncertainty of estimation is central in risk management, because usually risk measures are at the base of financial decision making. Examples of important decisions taken by a financial institution could be the control of the risks taken by one of its divisions. Moreover, in the field of systemic risk measurement, the decisions that a regulator could take based on the estimates of systemic risk are, for example, the capital requirements for systemically important financial institutions. In order to avoid costly mistakes, it is important to evaluate and interpret these measures with caution, by looking at the degree of uncertainty about our estimate.

We can present two examples to make this point more clear. Let us assume that VaR calculated for a particular portfolio held by a financial institution is 10. Let us also assume that the maximum VaR that the portfolio manager is allowed to take is 11. Based only on this information, the portfolio manager should not rebalance its portfolio, because he or she is not exceeding the maximum risk allowed. Now, instead of the point estimate of 10, we present a confidence interval whose upper and lower bounds are 7 and 13, respectively. This different information may require an intervention in the portfolio allocation, because it is quite likely that the maximum limit of risk is surpassed. Another example could be a regulator that has to determine which financial institutions are systemically important. Let us assume that institution A is characterized by a systemic risk measure equal to 3, while institution B is characterized by a systemic risk measure equal to 5. One may conclude that institution B is more systemically risky than institution A. Let us assume that instead of the point estimate we compute a confidence interval: the confidence interval for institution A is (1, 5) and for institution B is (3, 7). These confidence intervals are quite overlapped, so it is likely that the two measures that we thought were different, are statistically the same number.

In the financial literature there are examples of this kind of research regarding VaR and ES. For example, Jorion (1996) investigates the estimation risk of VaR and finds that the estimation risk is large and cannot be neglected, especially when we are considering a small confidence level and a restricted sample. Christoffersen and Goncalves (2005) use bootstrap methods to compute confidence intervals for VaR and ES, using different estimation models. They find that confidence intervals are generally quite large and that estimates of ES are characterized by a smaller degree of accuracy, if compared to those of VaR. These results are useful because, since systemic risk measures are based on VaR and ES, they can give us an idea of what kind of estimation risk we can expect from systemic risk measures. More recently, Danielsson and Zhou (2016) find that, even if ES is superior to VaR from a theoretical point of view (VaR is not subadditive), its estimation is more uncertain with respect to VaR. In particular, a determinant role is played by the sample size: estimation windows of 250 or 500 observations are usually not sufficient to deliver accurate risk estimates and thus should be avoided. Chen (2008) finds that more sophisticated methods to compute ES do not lead to more accurate estimates of this risk measure. Since these kind of analysis are still restricted for what concerns systemic risk measures based on market data, we try to fill this gap by analysing the estimation risk embedded in the current systemic risk measures.

This work tries to be a comprehensive and comparative study of measures of financial systemic risk. In order to do that, we apply our methodologies to the same set of institutions in Europe and in the US. We try to discuss also the advantages and disadvantages of each approach by highlighting the main differences and the main similarities. This could enhance the literature regarding systemic risk and suggest future studies.

2.3 Results

In this paragraph we will present our research on systemic risk measures and the relative results. We will analyse in detail each systemic risk measure starting form Delta CoVaR, then moving to MES and finally we analyse SRISK.

2.3.1 Results for CoVaR

One of the main findings of the work by Adrian and Brunnermeier (2016) is that Delta CoVaR does not provide the same information of VaR, because there is no correlation in the cross-section between Delta CoVaR and VaR of an institution. Figure 1 in Adrian and Brunnermeier (2016), shows the point estimates of VaR and Delta CoVaR for the main 15 financial institutions in the US divided in commercial banks, investments banks, insurance companies and government sponsored enterprises, just before the start of the crisis. Looking carefully at this picture, one might note that investment banks are grouped in the North-East part of the plane, because they are characterized by high VaR and high CoVaR. Instead, commercial banks are grouped in the central part of the plane, because they are characterized by an intermediate VaR and a small Delta CoVaR. In the end, insurance companies are grouped in the North-Est part of the plane, because they have small VaR but high Delta CoVaR. One may conclude that based on Delta CoVaR, insurance companies and investment banks are systemically more risky than commercial banks. This is also intuitive, since the business model of an investment bank is inherently riskier than the one of a commercial bank.

As we can see in Figure 1, this is confirmed by the lack of manifest correlation in the scatter plot. Indeed, the correlation is just -0.0168 and it is not significantly different form zero, thus we are pretty sure that there is no correlation between Delta CoVaR and VaR as risk measures: they provide different information.

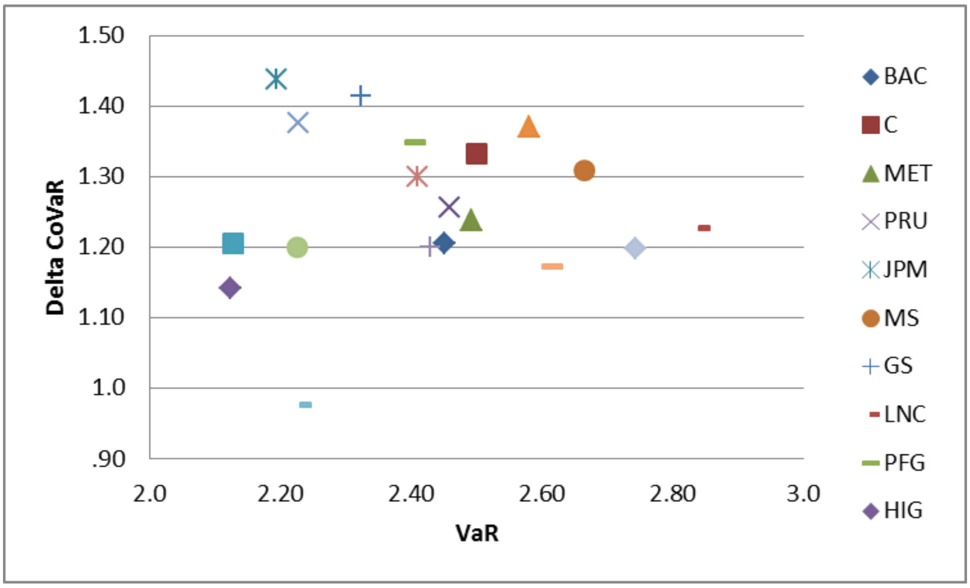

Figure 1: scatterplot of Delta CoVaR and VaR for American institutions.

But what it is interesting to note is that in the European case the relationship between VaR and Delta CoVaR is much more evident and in fact the correlation is equal to 0.5266 which is also statistically different from zero. This finding is in contrast to what we expect and what we obtain in the US case but, since we are considering a restricted number of institutions, it could be possible that these results are due to the small number of observations.

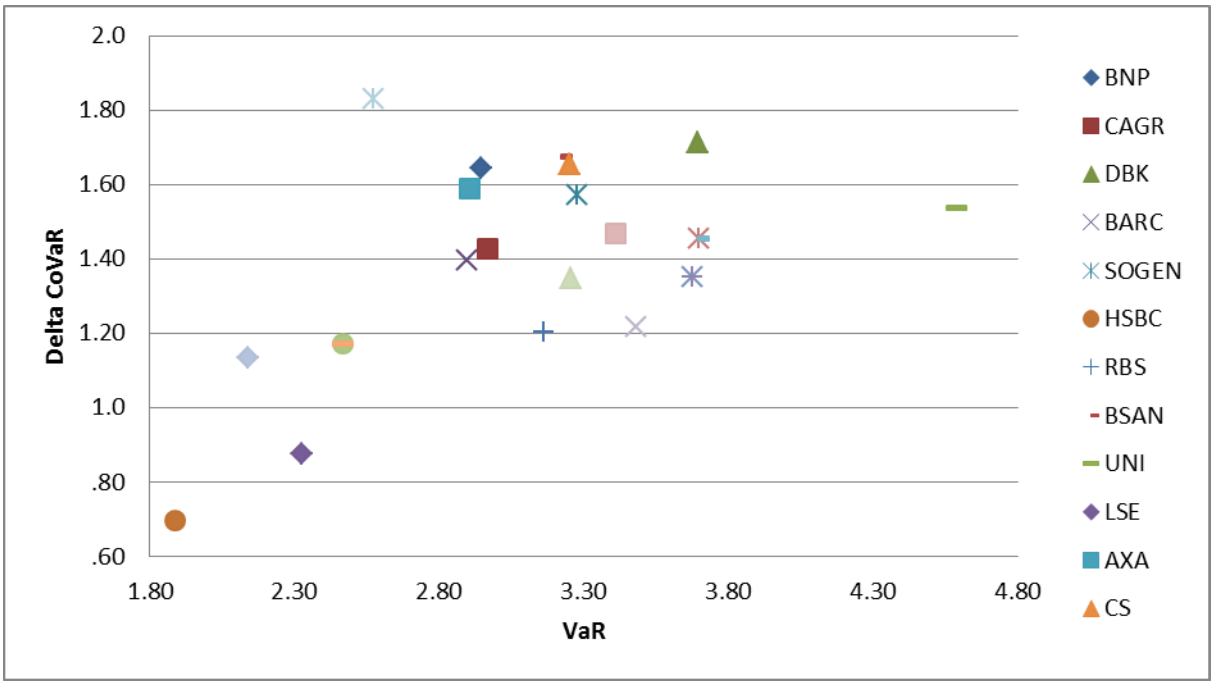

Figure 2: scatterplot of Delta CoVaR and VaR for European institutions.

If we divide the plane of the figure into four quarters, an institution that is located in the North-East part of the plan has a high Delta CoVaR and an high VaR which means that it is a good candidate for being a systemically important financial institution. Instead, firms that are located in the South-West quarter are characterized by small Delta CoVaR and small VaR, thus they can be considered less systemically risky. The main problem of this representation is that, even if it is very intuitive and easily readable, it considers only the point estimate of each risk measure, without any reference to the possible uncertainty of the numbers. This means that we can possibly be mistaken about two numbers that we think they are different but are the facto the same number, because they are not statistically different.

In statistics, uncertainty is usually expressed through standard errors and confidence intervals. But for complex models, as are the ones applied in financial risk management, it can be cumbersome or even impossible to derive analytical formulas for standard errors and confidence intervals using probability theory (Ruppert, 2011). To overcome this problem, a computer simulation technique called *bootstrap* or *resampling* is usually applied. This technique was popularized by the statistician Bradley Efron who coined the phrase *'pulling oneself up by one's bootstrap'* (see Efron and Tibshirani (1993) for a general introduction). The bootstrap technique consists in sampling with replacement a given number of resamples from our original sample which in this case will play the role of the population. There are two types of bootstrap techniques: the nonparametric or model-free bootstrap and the parametric or model-based bootstrap. In the non-parametric bootstrap, the resamples are drawn directly with replacement from the original sample. The important thing is that we have to sample with replacement, otherwise we will obtain only a copy of our original sample. Instead in the nonparametric bootstrap method, the resamples are drown from a density function. The parametric and the nonparametric methods are different, because they make different assumptions in order to draw the resamples. In the nonparametric bootstrap, we do not make any assumption about the distribution of the original sample but we have to assume that the observation are i.i.d. (independent and identically distributed). Instead, in the parametric bootstrap we have to make an assumption about the model or the distribution of the original data and then draw our resamples according to the specified model.

The bootstrap is a convenient and simple method to estimate the uncertainty of our systemic risk measures. First of all, we try the basic version of the bootstrap methods: the i.i.d. bootstrap. The only assumption that this method requires is that the sample is i.i.d. and no other assumptions about the distribution of the data are required. Since we are dealing with financial time series, we know that they are not properly i.i.d. because, even if they are not serially correlated, their variance is time-dependent and thus, they cannot be considered properly an i.i.d. time series. But as a first approximation of our work, we will consider these time series as i.i.d., even if this means that we lose the time dependence of the variance and so the effect of volatility clustering when we resample the data.

Another important fact that we have to consider is the correlation between the financial system and the single institution. In order to capture this fact, we cannot consider the two series as separate univariate time series. We have to consider them jointly, as a multivariate time series. Therefore, we must resample the pairs composed by the returns of the index and the returns of the financial institution. If $\mathbf{Y_1,...,Y_n}$ is our sample where $\mathbf{Y_i = [R_S \quad R_{FI}]}$, we resample directly $\mathbf{Y_i}$ and not the single components $\mathbf{R_S}$ and $\mathbf{R_{FI}}$ separately. In this way we maintain the covariance structure of the time series in the resamples (Ruppert, 2011).

We consider the daily negative log-returns for European and US financial institutions and the relative financial index. Our sample of the US firms ranges form 2013-03-08 to 2017-02-24 for 1000 observations (approximately 4 years of data). The sample for the European firms ranges form 2013-03-01 to 2017-02-23, also in this case for 1000 observations. We draw, with replacement, 5000 samples of length 1000 from the original sample. For each sample we compute the risk measures of interest (in this case VaR and Delta CoVaR).

Let us consider θ our quantity of interest (in our case the systemic risk measure). $\hat{\theta}$ will be the estimate from the original sample and $\hat{\theta}_1^*,...,\hat{\theta}_B^*$ the estimates for each resample. The bootstrap estimate $\bar{\hat{\theta}}^*$ is simply the average of $\hat{\theta}_1^*,...,\hat{\theta}_B^*$. We can also compute the bootstrap bias as

$$BIAS_{boot}(\hat{\theta}) = \bar{\hat{\theta}}^* - \hat{\theta} \tag{2.3}$$

and the bootstrap standard error is computed as

$$s_{boot}(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2} \ . \tag{2.4}$$

In Figure 3 we present the 95% confidence region obtained by bootstrapping the values of VaR and Delta CoVaR for four of the principal financial institutions in the US. The confidence regions are the equivalent of the confidence intervals, but in a two dimension framework, and they are computed using a procedure similar to the one used to compute Gaussian confidence intervals. This kind of analysis is suggested by Danielsson et al. (2016). The figure shows that the regions are quite overlapped, meaning that we are not certain about the statistical significance of the point estimate. As in the case of confidence intervals, in the case of confidence regions, if two regions are overlapped, it means that statistically they represent the same number. This aspect is important because usually, when we compute a systemic risk measure, we focus only on the point estimate and we just relay on that. In this case, relying on the point estimate only could be misleading, because we can think that one institutions is more systemically riskier than another, just because its point estimate is higher. Instead, if the confidence regions are quite overlapped, even if the point estimates are different, the two numbers are statistically the same number.

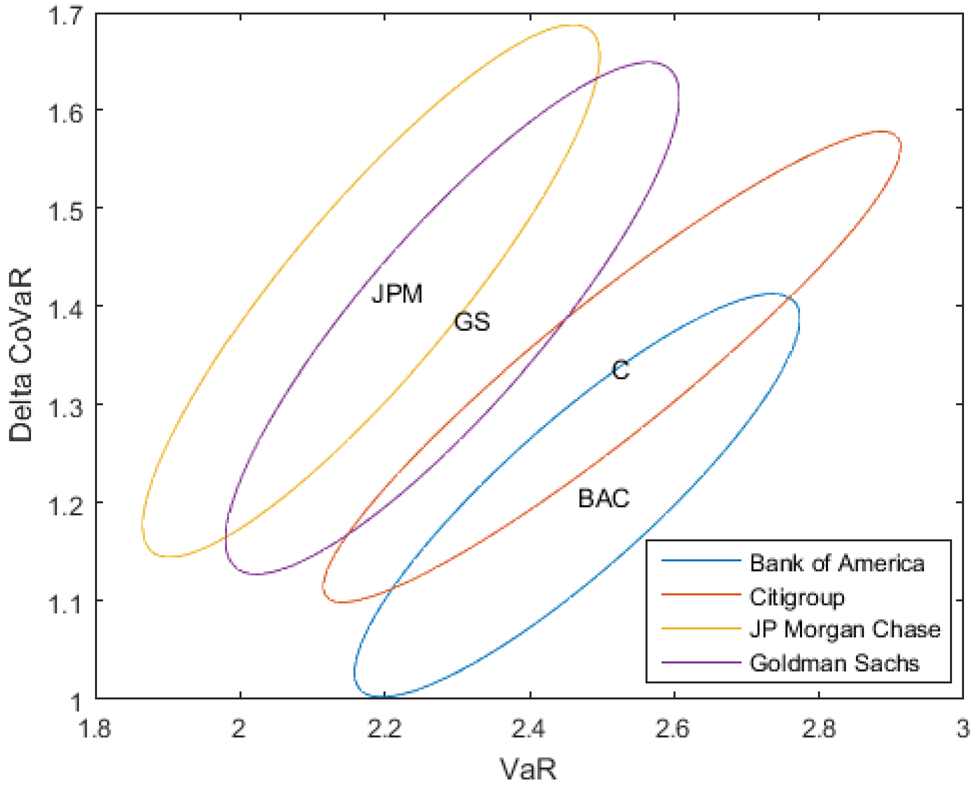

Figure 3: 95% confidence regions for American institutions.

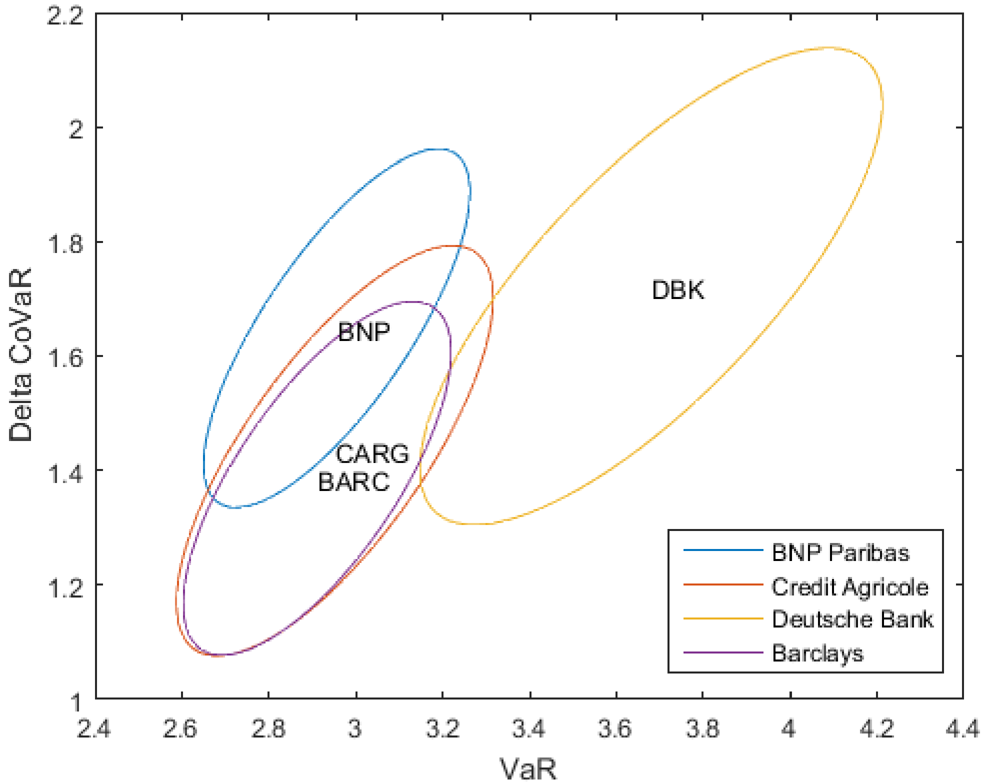

Figure 4: 95% confidence regions for European institutions.

We obtain a similar situation also in the European case in Figure 4. The only difference is that in this case one institution is partially detached from the others, because of its higher VaR. As we can see clearly form Figure 4, Deutsche Bank (DBK) is characterized by an higher VaR and this is why the institution is shifted to the right hand side of the chart. The other three institutions are more or less overlapped. Also in this case, we can notice that there is an high level of uncertainty in the statistical measure, thus it can be hard to distinguish if an institution is more risky than another one. We can notice also that the size of the area of the region is different from one institution to another. A larger area includes more values and thus we are more uncertain about the dimension of the measure. For example, in this case, the area of Deutsche Bank is clearly bigger than the one of the other financial institutions. This is probably because the time series of the returns of Deutsche Bank is more volatile with respect to the other series, leading to a more uncertain estimate.

In Table 2 and Table 3, we present the bootstrap mean, bias, standard error and confidence intervals for the institutions in the US and European case, respectively. There are many ways to construct confidence intervals using the bootstrap techniques. We choose the simpler and more general one that is the percentile method. The basic percentile confidence interval is simply

$$(q_L, q_U) \tag{2.5}$$

where q_L and q_U are the $\alpha/2$ and $(1-\alpha)/2$ quantiles of the empirical distribution of $\hat{\theta}_1^*, ..., \hat{\theta}_B^*$.

	Bootstrap Mean	**Bootstrap Bias**	**Bootstrap Standard Error**	**95% Percentile Confidence Interval**
Bank of America	1.2078	0.0027	0.0839	(1.0466, 1.3838)
Citigroup	1.3389	0.0074	0.0979	(1.1818, 1.5557)
JP Morgan Chase	1.4164	-0.0210	0.1108	(1.2136, 1.6300)

Goldman Sachs	1.3887	-0.0256	0.1065	(1.1652, 1.5714)

Table 2: bootstrap statistics for the American institutions.

	Bootstrap Mean	**Bootstrap Bias**	**Bootstrap Standard Error**	**95% Percentile Confidence Interval**
BNP Paribas	1.6468	0.0039	0.1281	(1.3549, 1.8950)
Credit Agricole	1.4349	0.0068	0.1464	(1.1639, 1.7446)
Deutsche Bank	1.7213	0.0074	0.1702	(1.4153, 2.0681)
Barclays	1.3866	-0.0086	0.1262	(1.1406, 1.6284)

Table 3: bootstrap statistics for the European institutions.

What we can notice by looking at these two tables is that the point estimates of the financial institutions in Europe are slightly higher than the estimates for the US institutions. This fact might signal that the top financial institutions in Europe brought more systemic risk than the top financial institutions in the US. The bias is very small and usually positive for both the samples, but it is not too much relevant in the analysis of the estimation risk. Instead, the standard error is quite large, if compared to the dimension of our point estimate. The standard errors and the confidence intervals show clearly that there is quite a lot of uncertainty in the estimation of Delta CoVaR. This means that if we used Delta CoVaR as our tool to make decisions related to risk management, we would be very uncertain about the decision to take or it could be that Delta CoVaR is not useful at all for our decision process.

One might think that these results are due to the fact that we use an i.i.d. bootstrap which is not always appropriate in the context of time series analysis. A time series represents one realization from a stochastic process and so it is a sample of dimension 1 which cannot be directly resampled (Ruppert, 2011). There are two ways to bootstrap time series: one considering nonparametric bootstrap and the other considering parametric bootstrap. The nonparametric time series bootstrap is called *block bootstrap* and consists in dividing the time series into blocks of consecutive observations and then resample the block with replacement and then paste them together in order to build the resampled time series. In this way the block will preserve the dependence in the data and the only assumption we need is that the blocks are i.i.d., so they can be drawn independently with replacement. The few assumptions required by the block bootstrap are offset by the fact that it is hard to select the block length. A large block length can preserve more autocorrelation in the data, but it reduces the number of possible resamples. Therefore, there is a trade-off between length of the resample and number of possible resamples. There is an algorithm that has been developed in order to choose the optimal block length, given the data (see Lahiri (2003) for further references). Instead, the parametric time series bootstrap relies on the simulation from a time series model. The first step is to specify a model with the relative estimated parameters. In this bootstrap technique, we do not resample directly the data, but we resample only the residuals which, if the model is correctly specified, are i.i.d. and so we can draw them with replacement. Usually we can draw the residuals directly or from a distribution. Once the residuals are resampled, they are attached to the model to simulate a possible path of the time series. The parametric bootstrap requires that the model is correctly specified, because it represents the true data generating process from which the data are drawn. The parametric procedure adds an additional approximation, because it requires the estimation of the parameters of the model. In the end, we cannot claim that one method is superior with respect to the other. We have to evaluate the assumptions made by each method and then choose the one which seems more appropriate for our purpose.

In order to understand if our results are robust to the bootstrap technique applied, we try a block bootstrap to our data. As we have said before, the only assumption that we have to made in the block bootstrap is the length of the block in order to have i.i.d. blocks. The peculiarity of financial returns is that, while they are not serially correlated, the squared returns are. This means that the data are uncorrelated, but they are not independent. Unfortunately, the method proposed by Politis and White (2004) to select the optimal block length takes into account only the autocovariance of the data and not the autocovariance of the squared data. For this reason, the choice of the block length for financial data can be a problem. In fact, if we apply the algorithm to our data we find that the optimal block length is more or less 1, because usually there is no autocorrelation in the data (a sign of efficiency in the market). But if we apply the algorithm to the squared returns, the optimal block length is between 5 and 8, because there is autocorrelation in the squared residuals, due to volatility clustering. For this reason, we think that a good choice for the block length is 8, because it can take into account the volatility clustering and it preserves the possibility of a high number of resamples.

We apply the block bootstrap to our samples and we find that the results are similar to those of the i.i.d. bootstrap. Figure 5 and Figure 6 show that the main results are valid also when we apply the more sophisticated block bootstrap. In Europe and in the US, the situation does not change significantly with respect to the method applied.

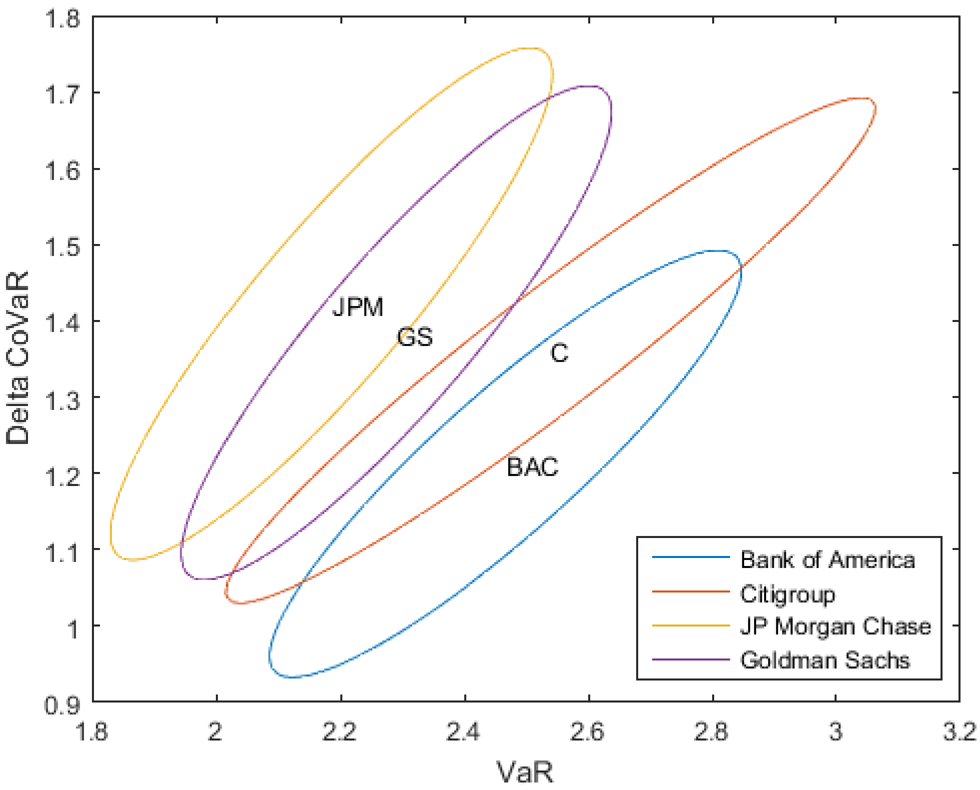

Figure 5: 95% confidence regions for American institutions using block bootstrap.

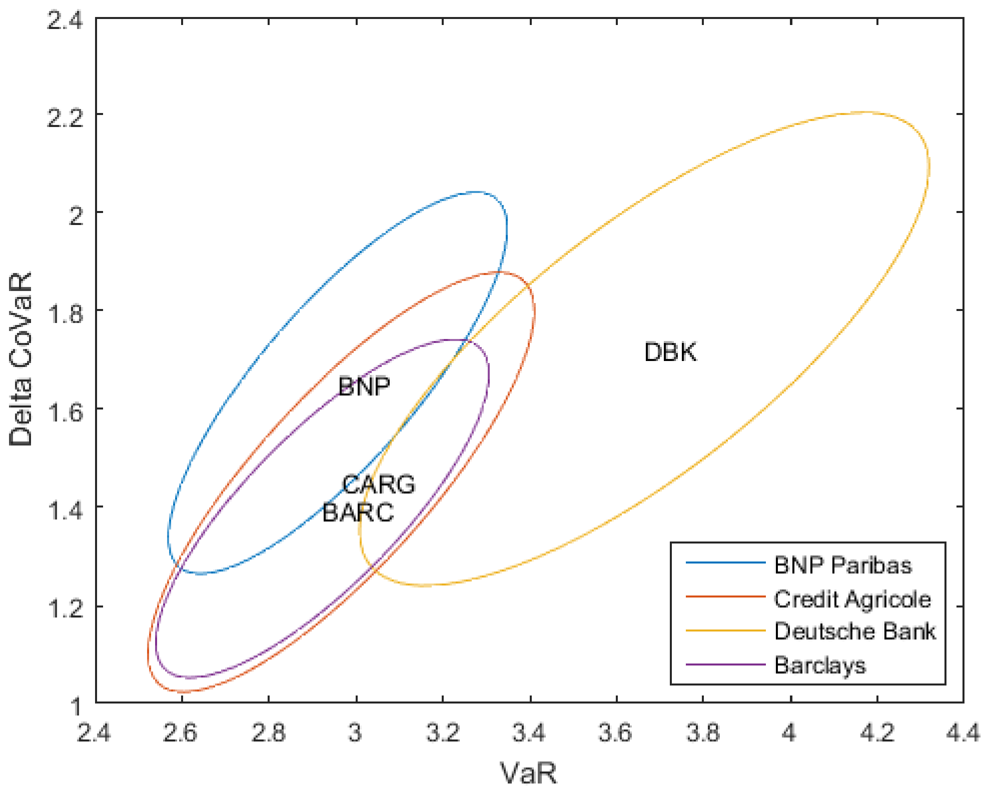

Figure 5: 95% confidence regions for European institutions using block bootstrap

Table 4 and Table 5 present the bootstrap means, biases, standard errors and percentile confidence intervals for the US and European case.

	Bootstrap Mean	Bootstrap Bias	Bootstrap Standard Error	Percentile Confidence Interval
Bank of America	1.2130	0.0079	0.1143	(0.9909, 1.4492)
Citigroup	1.3614	0.0299	0.1352	(1.1379, 1.6743)
JP Morgan Chase	1.4220	-0.0154	0.1372	(1.1711, 1.6900)
Goldman Sachs	1.3841	-0.0302	0.1323	(1.1203, 1.6097)

Table 4: block bootstrap statistics for the American institutions.

	Bootstrap Mean	Bootstrap Bias	Bootstrap Standard Errors	Percentile Confidence Interval
BNP Paribas	1.6537	0.0090	0.1588	(1.3092, 1.9567)
Credit Agricole	1.4513	0.0223	0.1747	(1.1490, 1.8321)
Deutsche Bank	1.7230	0.0091	0.1971	(1.3852, 2.1412)
Barclays	1.3976	0.0024	0.1408	(1.1275, 1.6618)

Table 5: block bootstrap statistics for the European institutions.

We can notice that the estimates of the block bootstrap are slightly higher than in the i.i.d. case, both in the European and in the US case. Also the biases and the standard errors are smaller in the i.i.d. case. This leads to narrower confidence intervals as well. It is possible that the percentile confidence intervals based on the i.i.d. bootstrap are artificially narrow and that they might not represent the true uncertainty about our estimates. This happens also when we use Gaussian confidence intervals, but the distribution in reality has heavy tails and so we underestimate the probability in the tails. For these reasons, we believe that by using block bootstrap, we obtain a more realistic estimate of confidence intervals and standard errors. Despite this problem, the main results that we find using the i.i.d. bootstrap are true also if we use the block bootstrap.

Given these results, it could be useful for a policy maker or a regulator to assess the statistical significance and the interpretation of these findings. In particular it could be useful to know if the systemic risk contribution of a given institution is significant and weather the systemic risk contribution of a financial institution is statistically different from the contribution of another financial firm. Indeed, this second aspect could be useful in order to develop a ranking of SIFIs. Statistical tests to assess the results provided by systemic risk measures have not yet been developed, but they are very important in order to draw conclusions and use them for policymaking.

We were able to find only a few researches that try to fill this gap. For example, Castro and Ferrari (2014) develope some statistical tests using directly the inference in the quantile regression framework. Since CoVaR is computed using quantile regression, they use the inference on this procedure to understand if the contribution to systemic risk of one institution is statistically significant and if one institution is more systemically risky than another one. The power of their tests shows optimality only when the sample size is very large (5000 observations). The tests have only moderate performance for the number of observations that it is usually available for financial applications. Kupiec and Guntay (2016) develop a statistical test for Delta CoVaR, MES and SRISK using Gaussian stock returns: in this way it is possible to derive analytical formulas for the tests. Also in this case they show that their tests have weak power, especially when they test for CoVaR.

We try to contribute to this restricted strand of the literature by using our bootstrap confidence intervals to build a hypothesis testing procedure. In fact we can exploit the one-to-one relationship that exists between confidence intervals and hypothesis tests. This one-to-one relationship makes possible the construction of an hypothesis test using any bootstrap method to obtain confidence intervals (Chernick and LaBudde, 2011). In books of statistics dedicated on bootstrap techniques, there are chapters on this topic (see for example Chernick (2007) and Chernick and LaBudde (2011)). Let us assume that we have built a $(100-\alpha)\%$ bootstrap confidence interval. We define the null hypothesis H_0 as $\theta = \theta_0$, where θ is a parameter of interest and θ_0 is a predetermined value. We obtain a bootstrap test by rejecting H_0, if and only if θ_0 lies outside the confidence interval. The significance level of the test is related to the confidence level of the interval and so the more accurate the bootstrap confidence interval is, the more powerful would be the hypothesis test. The confidence level of the test is the same level α of the confidence interval. Given a confidence level α, we want a confidence level to be as narrow as possible, in order to gain power in the hypothesis testing.

An important application for a regulator would be to test if a financial institution is systemically important or not. We can do it by specifying the null and the alternative hypothesis. For example, when we consider Delta CoVaR, MES and SRISK as systemic risk measures (SRMs), we can specify that the null hypothesis is that one of these measures is equal to zero for institution i and the alternative hypothesis is that it is not equal to zero:

$$\begin{aligned} H_0 &: SRM_i = 0 \\ H_1 &: SRM_i \neq 0 \end{aligned} \qquad (2.6)$$

In this way we can identify which institutions are systemically important with a transparent and quick procedure. In this case we reject the null hypothesis if and only if the value zero falls outside the confidence interval.

In the same way, we can test if the contribution to systemic risk of a financial institution i is different from the contribution of institution j. In this case, the null hypothesis of the test is that the difference between the two SRMs is equal to zero and the alternative hypothesis is that the difference is not equal to zero:

$$\begin{aligned} H_0 &: SRM_i = SRM_j \rightarrow SRM_i - SRM_j = 0 \\ H_1 &: SRM_i \neq SRM_j \rightarrow SRM_i - SRM_j \neq 0 \end{aligned} \qquad (2.7)$$

In this case we reject the null hypothesis if and only if the value of the difference between the two measures is not included in our confidence interval. If we substitute the symbol $\neq$ with the symbol $>$, we can test if the SRM of institution i is statistically greater than the SRM of institution j.

The use of bootstrap confidence intervals for hypothesis testing is very attractive, because it does not require any assumption about the distribution of the data or the statistic of the test: for this reason they are called nonparametric tests. What really matters is the precision of the bootstrap confidence intervals and how they are constructed. For example, the percentile confidence intervals we used in the previous application could be a good choice, because they only require that the distributions of interest are symmetric. In addition, we can even improve their precision by considering the biased corrected (BC) bootstrap method or the adjusted percentile method (BCa). In this way, we can obtain a narrower confidence interval, without decreasing the confidence level and improving the power of the test.

We apply our statistical tests to the data of European and American financial institutions. First of all, we will use out test for identifying financial institutions that are systemically important, because they have a Delta CoVaR that is statistically different from zero. We take $\alpha = 0.05$ for our test. The null and the alternative hypothesis of our test are:

$$\begin{aligned} &H_0 : \Delta CoVaR_i = 0 \\ &H_1 : \Delta CoVaR_i \neq 0 \end{aligned}. \tag{2.8}$$

For this test we consider 20 US financial institutions. A list of these institutions can be found in Appendix A. We consider the percentile confidence interval computed using the block bootstrap procedure applied before. The percentile confidence intervals are not the 'state-of-the-art' in terms of bootstrap confidence intervals (BC_a intervals are probably a little bit more precise) but they have an important feature that could be useful when we will compute SRISK: in fact percentile confidence intervals are invariant to monotonic transformations. On this premise, at the 5% significance level, we reject the null hypothesis that Delta CoVaR is equal to zero for all the 20 institutions. This means that the contribution of these 20 financial firms is statistically different from zero and thus they all contribute to the systemic risk of the financial system.

Then we repeat the same test for 20 European financial institutions. A list of the European financial institutions can be found in the Appendix A. At the 5% significance level, we reject the null hypothesis for all the 20 institutions. Therefore, also in this case these 20 European institutions contribute all to the financial systemic risk of the economy.

Since we have understood that the contributions to systemic risk of many financial institutions are significant, now we are interested in comparing our results with other similar results present in the literature. Castro and Ferrari (2014) study if the Delta CoVaR of 26 European banks is statistically different from zero using weekly returns from 1999 to 2012. Their test indicates that all banks considered have a significant systemic risk contribution. This result is in line with our findings. They do not study the American situation, so we cannot compare it. Kupiec and Guntay (2016) study the 50 largest US financial institutions between 2006 and 2007 and they find that only one financial institution (MetLife) has a Delta CoVaR that is statistically significant. This result is completely in contrast with our findings: in our case all the financial institutions we tested have a significant Delta CoVaR. This is quite surprising, because we expected similar results, even if the tests developed are different and relies on different assumptions. One possible explanation is the power of the test. As demonstrated by the authors using a Monte Carlo simulation, their test has a weak power, especially for what concerns Delta CoVaR. This means that the null hypothesis is not rejected many more times we would expect and for this reason the power of the test is unsatisfactory. Another reason could be the sample period. Kupiec and Guntay (2016) use data from 2006 to 2007, thus before the crisis. We instead use a much more recent time span from 2013 to 2017. These results confirm that this field of research is just at an early stage: there is the need for additional research that will address these problems and find possible solutions.

The next step now consists in understanding if the contribution of systemic risk of one institution is statistically different from the one of another institution. This is very important, especially for a regulator, because it can identify a ranking of the institutions that are systemically more risky and then develop a policy or a regulatory framework with the aim of limiting the risk taken by financial institutions and protecting the real economy from externalities that could arise from the financial system. In this case we can simply test if the difference between the Delta CoVaR of institution i and the Delta CoVaR of institution j is statistically different from zero:

$$\begin{aligned} &H_0 : \Delta CoVaR_i = \Delta CoVaR_j \rightarrow \Delta CoVaR_i - \Delta CoVaR_j = 0 \\ &H_1 : \Delta CoVaR_i \neq \Delta CoVaR_j \rightarrow \Delta CoVaR_i - \Delta CoVaR_j \neq 0 \end{aligned} . \tag{2.9}$$

If the bootstrap confidence interval of the difference between the two risk measures contains the value 0, we accept the null hypothesis: the confidence interval represents the acceptance area of our test.

Firstly, we test for the US institutions. The result shows that the institutions can hardly be discriminated based on their Delta CoVaR. At the 5% confidence level, the risk measures of different institutions for the vast majority of the cases are not statistically different and thus their contribution to the systemic risk of the financial system is statistically the same. These results for the US case are confirmed by other researches. Danielsson et al. (2016) use the same data and procedure of Adrian and Brunnermeier (2016) and conclude that it is hard to distinguish between financial institutions by only using their Delta CoVaR. Table 6 presents the result of this test for the US case, using as alternative hypothesis $\Delta CoVaR_i > \Delta CoVaR_j$. For each institution in our sample we present the number of institutions that are dominated by the institution in question.

Financial Institution	**Number of Dominated Financial Institutions**
Bank of America	1
Citigroup	2
MetLife	0
Prudential	1
JP Morgan Chase	2
Morgan Stanley	1
Goldman Sachs	2
Lincoln National	1
Principal Financial Group	4
Wells Fargo	1
Hartford Financial Services	1
Capital One	1
State Street	1
Bank of New York Mellon	1
Ameriprise Financial	1
Suntrust Banks	1
Key Corp	1
Cit Group	0
Regions Financial	1
Zions Bancorporation	1

Table 6: the column indicates the number of other American financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its Delta CoVaR is statistically higher than the Delta CoVaR of 5 institutions in the sample.

Now we analyse the European case. The result of the tests shows that also for the European case Delta CoVaR is most of the time not statistically different between institutions. Table 7 presents the result of the tests. These results are also confirmed by the findings of Castro and Ferrari (2014). Appling their test, they find that for only 38 bank pairs out of 352 the test indicates that Delta CoVaR is statistically different between institutions.

Financial Institution	**Number of Dominated Financial Institutions**
BNP Paribas	4
Credit Agricole	2
Deutsche Bank	2
Barclays	2
Societe Generale	2
HSBC	0
Royal Bank of Scotland	1
Banco Santander	4
Unicredit	2
London Stock Exchange	0

Table 7: the column indicates the number of other European financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its Delta CoVaR is statistically higher than the Delta CoVaR of 5 institutions in the sample.

These results which, as we have seen, are confirmed also by other similar findings in the literature cast some shadows about the ability of Delta CoVaR in identifying systemically important financial institutions. For a regulator, the information provided by Delta CoVaR would be of limited benefit. One should conclude that all the banks bring more or less the same risk and so they should be subject to the same regulation.

We think there are mainly two reasons that can explain why we obtain these results. The former is the information used to compute Delta CoVaR and the latter is the estimation error. We think that Delta CoVaR, considered in isolation, as a systemic risk measure is inadequate to measure the contribution to systemic risk of a financial institution. The problem is that Delta CoVaR takes into account only the information provided by the market through the stock prices: this information is important but not sufficient. We need also other types of information. In order to deal with this problem, instead of Delta CoVaR, we can consider Dollar Delta CoVaR as our systemic risk measure. As we have seen in the theoretical part, Dollar Delta CoVaR is simply the Delta CoVaR of an institution multiplied by its market capitalization. In this way we take into account also the size of the institution which is considered to be important, because of the well-known problem of ‘too-big-to-fail’ institutions. But, in our opinion, also this result will fail to give adequate results because, since the Delta CoVaR of institutions is very similar, Dollar Delta CoVaR will pick as systemically important the institutions that have large market capitalization. In this way institutions such as JP Morgan in the US and HSBC in Europe will be treated as more systemically risky only because they have large market capitalizations; but the market does not consider them as a the more risky financial institutions in the system.

For what concerns the estimation error, we have demonstrated that this is large and it cannot be avoided, at least using the statistical method that has been applied in the literature until today. We think that there are mainly two reasons that can explain the large estimation error. Firstly, the computations required for Delta CoVaR are based on VaR which is well-known that it has a large estimation error as well (see Christoffersen and Goncalves (2005) and more recently Danielsson and Zhou (2016)). This means that the large estimation error provided by VaR is transferred to the estimation of Delta CoVaR, making difficult to infer some precise conclusions about the estimates. Secondly, VaR and Delta CoVaR are difficult to estimate, because they are measures that consider the tail of a distribution. As we know, we have a few observations in the tails of a distribution and thus this fact makes the estimation much more imprecise in these extreme regions. This fact is an inherent problem in the whole discipline of financial risk management and cannot be avoided. But we must be aware that this problem exists and thus the results should be assessed with caution.

Until now, we have considered Delta CoVaR only in a cross-sectional framework; in other words, we have computed Delta CoVaR for different institutions at the same time instant t. Now instead, we move to analyse the time series framework of Delta CoVaR by computing Delta CoVaR for an institution i in a time interval. We want to understand how the value of Delta CoVaR evolves through time, its volatility and the differences between two institutions. The most important thing for a regulator is understanding how much Delta CoVaR is volatile as a measure of systemic risk. If Delta CoVaR is too much volatile, it will be difficult to use it to set capital requirements for financial institutions. Capital requirements should not change too much rapidly, because otherwise it will be impossible for the financial institution to meet these requirements.

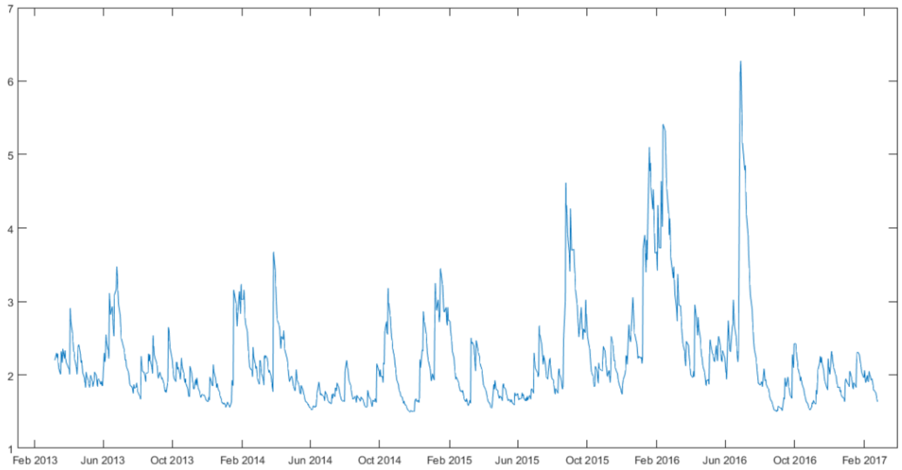

Figure 6: time-varying Delta CoVaR for Bank of America.

Figure 6 shows the time series of Delta CoVaR for one US institution from 2013 to 2017. We can notice that, even if our sample period is not characterized by financial crisis or other relevant market stress, the estimate of Delta CoVaR is very volatile, because it ranges from 1.5 to 6. The high volatility of this systemic risk measure indicates that the systemic risk of one financial institution changes very rapidly, at least if we measure it with quantile regression Delta CoVaR. This is not a good thing from a regulatory point of view, because a regulator wants a measure that is as smooth as possible (Lucas, 2014). This is another proof of the fact that Delta CoVaR has some problems and it is not readily applicable as a systemic risk measure: it needs some modifications in order to deal with the problems we have presented so far.

Another interesting thing is looking at how the Delta CoVaR of two institutions varies across time. In this case we are interested in the correlation between the two time series. If the correlation is high, it will be difficult to distinguish between the risk of the two financial institutions and so it will be difficult to apply a different regulation. As we can see from Figure 8 which shows the time series of Delta CoVaR for two different financial institutions, they are highly correlated, because the series are quite overlapped over time. This means that the risk of the two institutions is pretty much the same over time and so they should be subject to the same level of regulation. But this means also that Delta CoVaR can hardly identify a difference in the systemic risk of a financial institution, if compared to another financial institution.

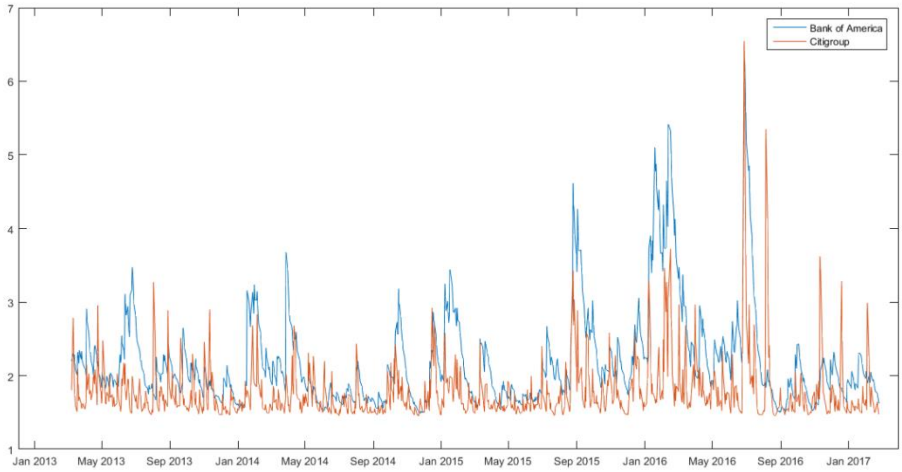

Figure 7: comparison of time-varying Delta CoVaR for two American institutions.

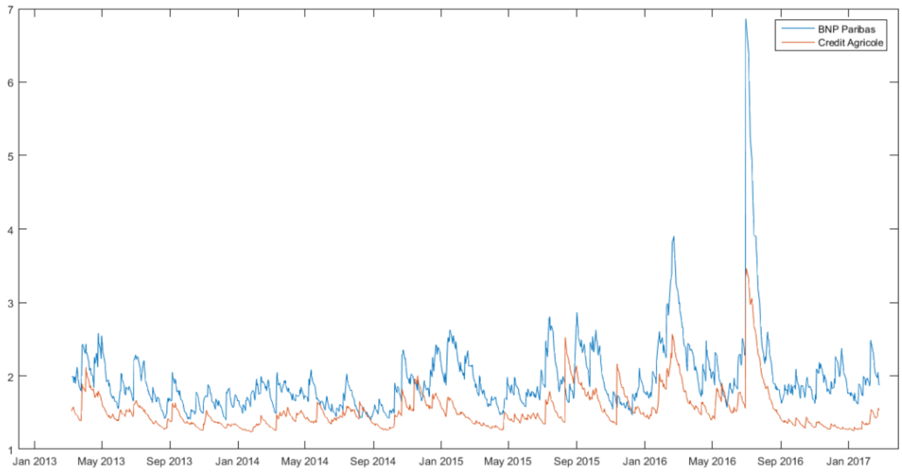

Figure 8: comparison of time-varying Delta CoVaR for two European institutions.

In this paragraph we have analysed Delta CoVaR which is probably the most popular and the most applied systemic risk measure today. We have computed the estimation error related to this measure and we have explained how it affects the inference and the decisions that we can take based on this measure. We can say that the uncertainty of a risk measure such as VaR is transferred also to Delta CoVaR: in fact the computations of Delta CoVaR are based on VaR. In conclusion, it is difficult that a regulator will take important decisions based on this type of measure, because the uncertainty is too high and so the risk of taking the wrong decision is also high.

2.3.2 Results for MES

In the previous paragraph we have explained our methodology and we have analysed Delta CoVaR from an empirical point of view. Now we are going to use the same procedure to analyse another important systemic risk measure: the marginal expected shortfall (MES). MES is another popular systemic risk measure and it is applied in a lot of theoretical and empirical works.

There are contrasting ideas about the usefulness of MES. For example, Acharya et al. (2017) find that MES, measured on the verge of the last financial crisis, was a good predictor of the real equity losses that these institutions experienced during the crisis. On the other hand, Idier et al. (2014), using data from the same period (2007-2009) for US financial institutions, suggest that MES is not able to predict large equity losses of institutions and that standard balance-sheet ratios are more useful than MES. On this premise, the fact that MES can really help regulators in identifying systemically important banks is still an open question.

First of all, as we have done for Delta CoVaR, we assess the estimation error of MES. The methodology and structure of this paragraph are very similar to the previous one, a part from some necessary modifications.

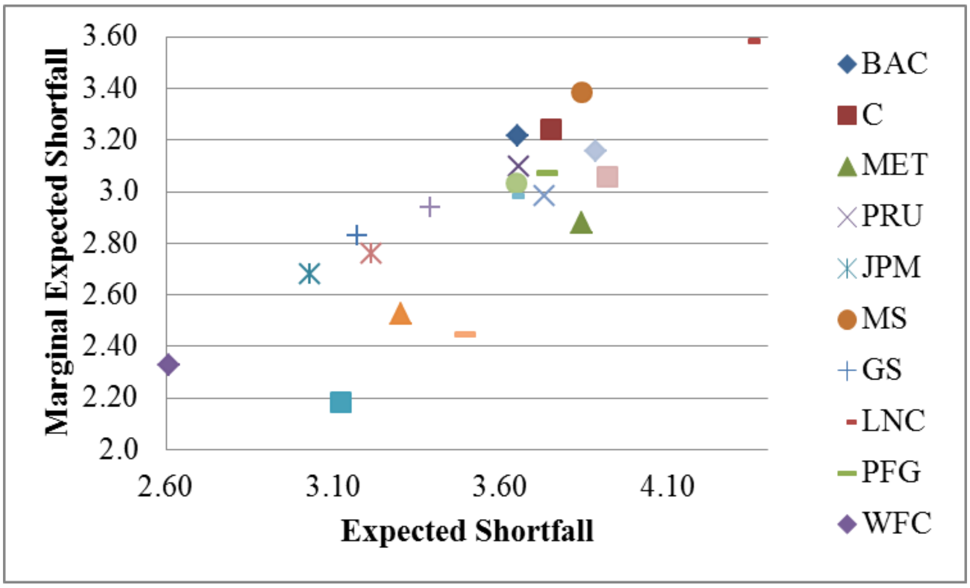

Figure 9: scatterplot of institutions' MES and ES in the American case.

As we have done for Delta CoVaR and VaR, we propose a scatterplot of MES and ES. We choose ES because it is the risk measure from which MES is derived. By looking at Figure 9, we can notice that in the US case there is a clear correlation between MES and ES. In fact the Pearson correlation coefficient is equal to 0.8263 and it is statistically significant. This could mean that, despite what we have observed in the case of Delta CoVaR and VaR, MES and ES provide the same type of information and they are not so different between each other. Indeed, this is confirmed by looking at Figure 10 that presents the scatterplot for the European case. Here we can identify a correlation equal to 0.7352 which is statistically different from zero. Therefore, we are pretty sure that the information provided by MES is very similar to the information provided by ES and this is an undesirable feature of MES, because we want that MES provides us a different signal with respect to the traditional statistical risk measure such as ES.

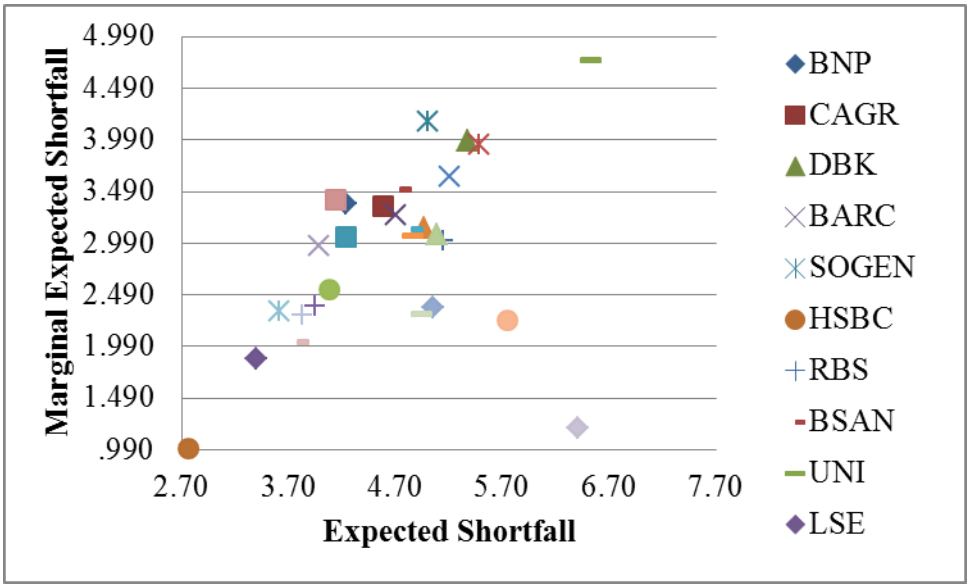

Figure 10: scatterplot of institutions' MES and ES in the European case.

Another interesting point to evaluate is the difference in information provided by two measures: Delta CoVaR and MES. Since both are systemic risk measures, we expect them to provide more or less the same signal. The Spearman's rank correlation coefficient is useful in this case, because it measures the correlation in the rank between two random variables. It ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation). For the US case, the Spearman's rho between Delta CoVaR and MES is equal to 0.0391 and it is not statistically different from zero. Instead for the European case, the Spearman's rho is 0.7953 and it is statistically different from zero. As happened for the case of Delta CoVaR, we obtain conflicting results from our two samples representing Europe and the US, probably because the number of observations is very small. Therefore, based on this information, we are not able to determine whether Delta CoVaR and MES provide the same information or not.

Now we will analyse the estimation error associated to MES. In order to ensure comparability, we apply the same bootstrap techniques that we used for Delta CoVaR. First, we use the i.i.d. bootstrap in the US case. The situation is very similar to the one found for Delta CoVaR, but in this case there is a stronger correlation between MES and ES of different institutions and thus the point estimates and the relative confidence regions are more aligned: this is in line with the findings in the scatterplots. The fact that the regions are quite overlapped means that there is a lot of uncertainty about our estimates and it could be that the estimates are statistically the same number: we will deal with this problem later. We can notice also that the estimation risk for each single institution is different, because the area of each region is different. A bigger area means that the estimate includes more extreme values and thus more uncertainty.

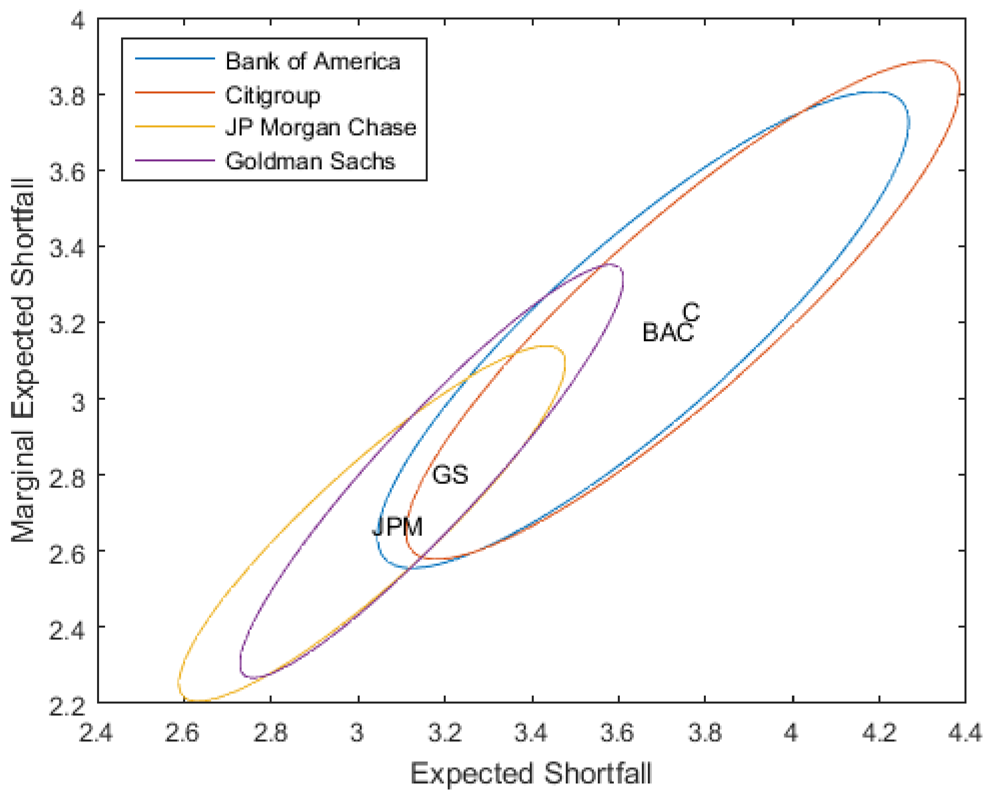

Figure 11: 95% confidence regions using the i.i.d. bootstrap in the American case.

If we look at the European case in Figure 12, we can notice that the situation is not so different from the one we presented in the previous paragraph related to Delta

CoVaR. The only difference is that here the institution that in the previous case was sticking clearly from the others, now it is not so detached. In fact Deutsche Bank is still in this case detached from the other institutions (BNP Paribas, Credit Agricole and Barclays), but not as much as in the previous case.

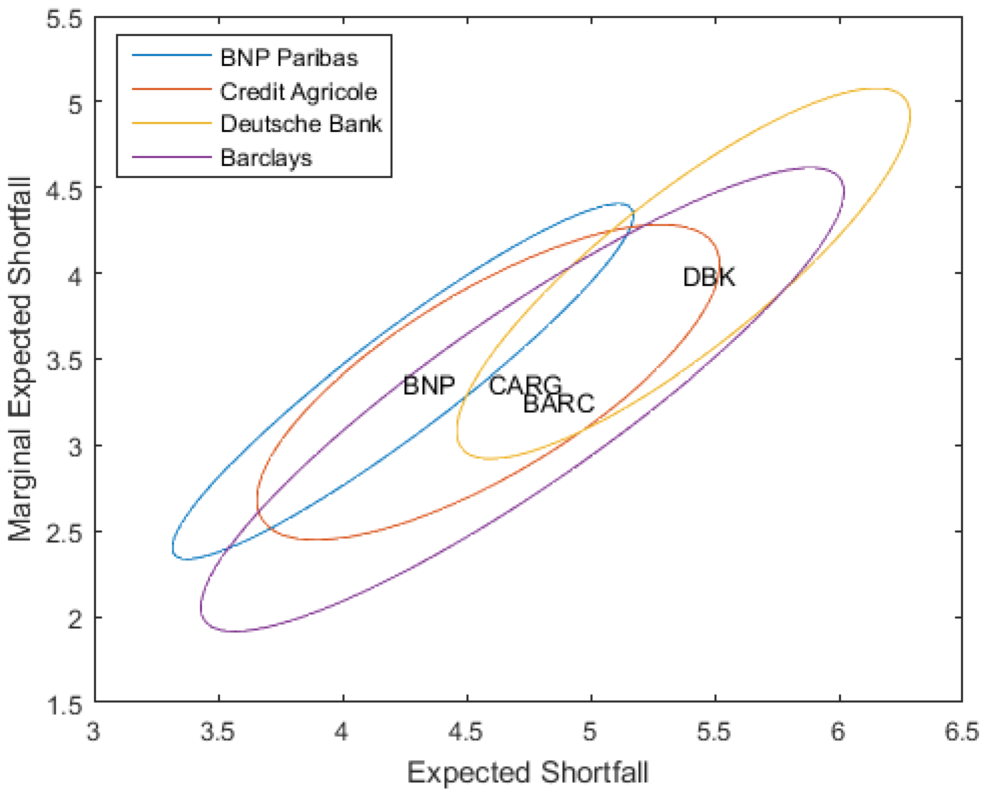

Figure 12: 95% confidence regions using the i.i.d. bootstrap for the European case.

	Bootstrap Mean	Bootstrap Bias	Bootstrap Standard Error	Percentile Confidence Interval
Bank of America	3.1809	-0.0377	0.2553	(2.6818, 3.6843)
Citigroup	3.2327	-0.0109	0.2671	(2.7247, 3.7735)
JP Morgan Chase	2.6720	-0.0082	0.1903	(2.3086, 3.0524)
Goldman	2.8091	-0.0201	0.2213	(2.3903,

Sachs				3.2502)

Table 8: statistics for the i.i.d. bootstrap in the American case.

	Bootstrap Mean	**Bootstrap Bias**	**Bootstrap Standard Error**	**Percentile Confidence Interval**
BNP Paribas	3.3710	0.0028	0.4225	(2.6621, 4.2967)
Credit Agricole	3.3665	0.0137	0.3750	(2.6783, 4.1538)
Deutsche Bank	4.0009	0.0195	0.4411	(3.2010, 4.9191)
Barclays	3.2654	0.0042	0.5516	(2.3351, 4.4694)

Table 9: statistics for the i.i.d. bootstrap in the European case.

It is interesting to assess the estimation error of MES compared to the one of Delta CoVaR. In order to do that, we must use a standardized measure of the variability of the estimate: the coefficient of variation is suitable for our application. The coefficient of variation is obtained by the ratio of the standard error (that measures the variability) over the estimate of the mean. It this way, we obtain a measure that is standardized and can be used to compare estimates of Delta CoVaR and MES. An high coefficient of variation indicates that the estimate is less precise and more volatile. By looking at the results for the i.i.d. bootstrap, we find that the coefficient of variation is higher for the estimates of MES than for the ones of Delta CoVaR. This is in line with what one should expect from the previous literature on VaR and ES. In fact there are strong proofs that the estimation of ES is more uncertain than the estimation of VaR (see Christoffersen and Goncalves (2005) and Danielsson and Zhou (2017)). Therefore, since Delta CoVaR and MES are based respectively on VaR and ES, it is logic to imagine that the estimation of MES will be more uncertain than the estimation of Delta CoVaR. Probably, the reason why ES is estimated with less precision with respect to VaR is that it is more difficult to estimate an expected value than a quantile. This fact can explain why VaR is estimated with more precision than ES and consequently, why Delta CoVaR is more precise than MES.

Then we move from the i.i.d. bootstrap to the block bootstrap, as we have made for Delta CoVaR, in order to take into account the dependence of the time series and make our results more robust. We use also in this case a block length of 8 observations, because it seems adequate to take into account the dependence in the squared residuals due to volatility clustering. The situation does not change too much between i.i.d. bootstrap and block bootstrap, both in the European and American application. For what concerns MES, we can repeat what we have already remarked for Delta CoVaR. Using the block bootstrap procedure, we obtain a more reliable picture of the estimation error of MES, because we take into account the dependence in the variance. In fact, by using the i.i.d. bootstrap the confidence intervals are artificially narrow.

We can also notice that the estimation error is different from sample to sample. For example, the estimation error of the series regarding Bank of America and Citigroup is much more higher than the one of the series regarding JP Morgan Chase and Goldman Sachs. In the European case depicted in Figure 14, we can notice that the confidence region of Barclays is extremely large, compared to the one of the other financial institutions: therefore, we are more uncertain about Barclays than about other institutions. We can conclude by saying that the estimation of MES is at least as imprecise as the estimation of Delta CoVaR. The fact that the estimation of MES is based on ES could be an explanation: in fact we have evidences in the literature about the imprecision of ES estimation.

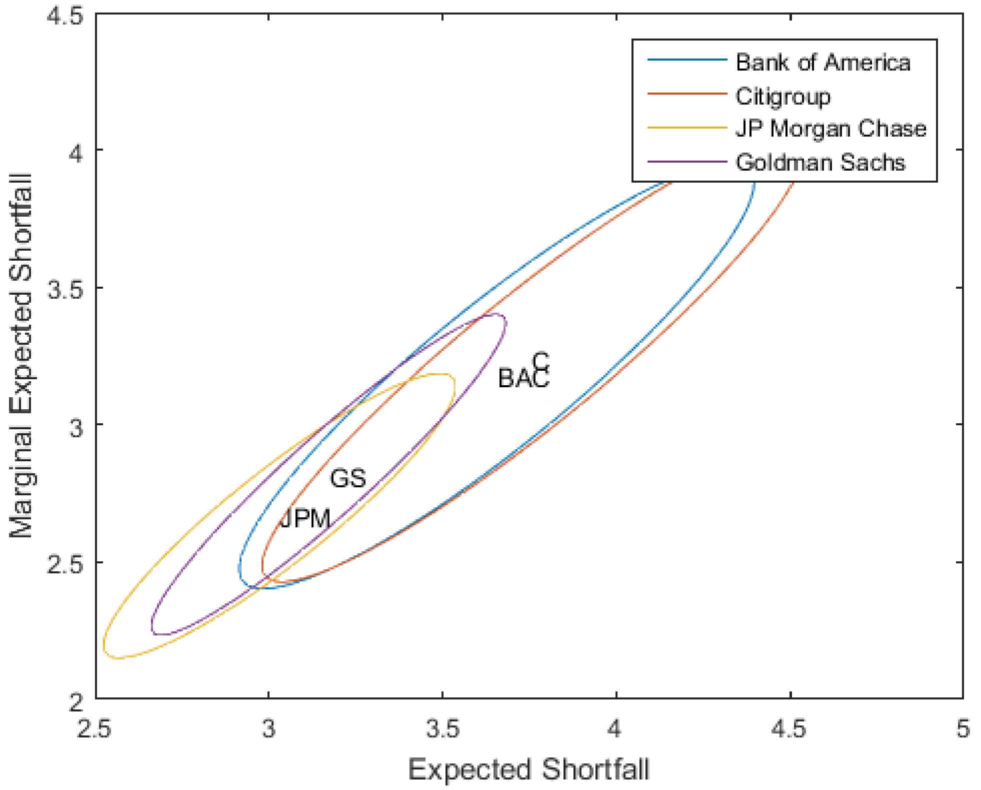

Figure 13: 95% confidence regions for American financial institutions obtained using block bootstrap.

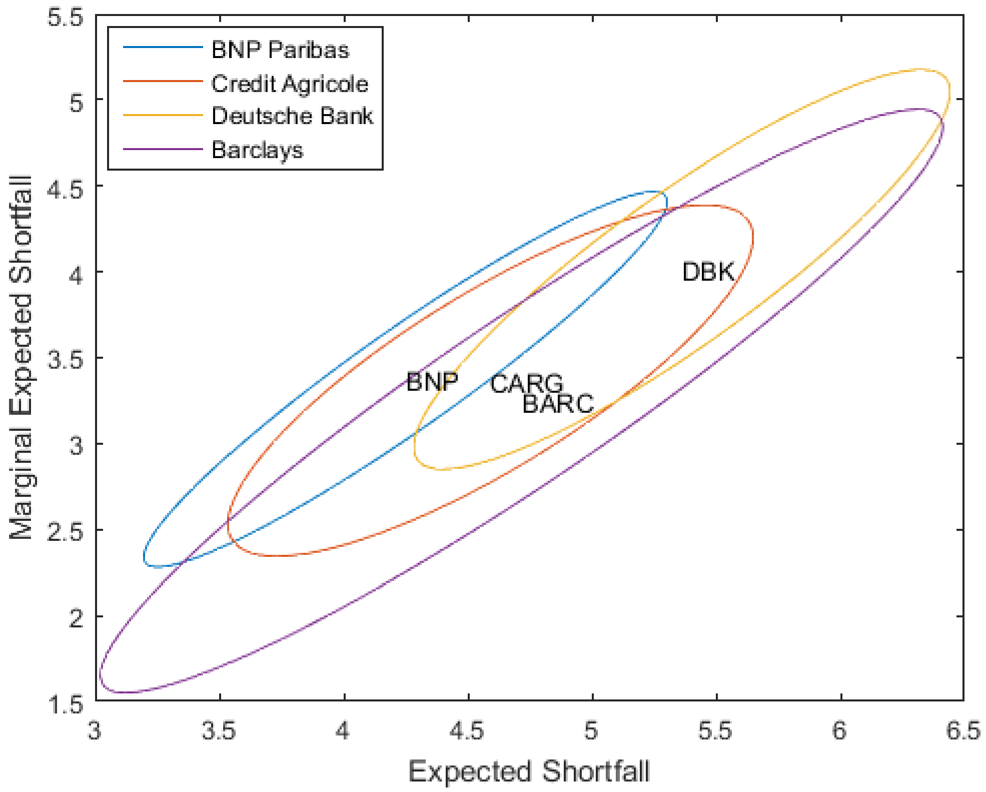

Figure 14: 95% confidence region using block bootstrap in the European case

	Bootstrap Mean	Bootstrap Bias	Bootstrap Standard Error	Percentile Confidence Interval
Bank of America	3.1785	-0.0401	0.3166	(2.5885, 3.8208)
Citigroup	3.2362	-0.0074	0.3299	(2.6140, 3.9118)
JP Morgan Chase	2.6695	-0.0106	0.2116	(2.2778, 3.0978)
Goldman Sachs	2.8178	-0.0115	0.2384	(2.3861, 3.3013)

Table 10: statistics for the block bootstrap in the American case.

	Bootstrap Mean	Bootstrap Bias	Bootstrap Standard Error	Percentile Confidence Interval
BNP Paribas	3.3744	0.0061	0.4466	(2.6033, 4.3578)
Credit Agricole	3.3670	0.0143	0.4180	(2.5945, 4.2314)
Deutsche Bank	4.0135	0.0321	0.4751	(3.1500, 5.0231)
Barclays	3.2481	-0.0130	0.6934	(2.2002, 4.8597)

Table 11: statistics for the block bootstrap in the European case.

We now turn to the analysis of results of hypothesis testing for MES. As in the case of Delta CoVaR, we are interested in understanding if the contribution to systemic risk of each financial institution is statistically different from zero. In order to do that, we apply the nonparametric statistical test based on the block bootstrap confidence intervals which has the following null and alternative hypothesis:

$$\begin{aligned} H_0 &: MES_i = 0 \\ H_1 &: MES_i \neq 0 \end{aligned} . \qquad (2.10)$$

At the 5% significance level, we reject the null hypothesis for all the 20 financial institutions considered in the US case. The same is also true for our sample of 10 European financial institutions. Therefore, in our samples, all the firms contribute significantly to systemic risk. Our results are partially supported by Kupiec and Guntay (2016) which find that MES of 27 US firms (out of 50) is statistically significant, at the 5% level. We did not find any comparable result for the European case.

Since we have understood that all the institutions contributes positively to the systemic risk of the system, now we want to understand if these contributions are statistically different between each other and which institutions are the most important in terms of contribution. In order to do that, we use the following nonparametric statistical test based on block bootstrap confidence intervals:

$$\begin{aligned} H_0 &: MES_i = MES_j \rightarrow MES_i - MES_j = 0 \\ H_1 &: MES_i \neq MES_j \rightarrow MES_i - MES_j \neq 0 \end{aligned} \tag{2.11}$$

The results of the test are similar to those that we have found in the previous section for Delta CoVaR. At the 5% significance level, we find that most of the time MES between two institutions is not significantly different from zero. Table 12 reports the number of institutions that are dominated by the financial institution, according to the significant difference of their systemic risk measures, using the alternative hypothesis $MES_i > MES_j$. For the American case, we can notice that for example the systemic risk of Bank of America is higher than the risk of only 8 institutions out of 19. This means that it is hard to distinguish between financial institutions using only as a reference their MES.

Financial Institution	**Number of Dominated Financial Institutions**
Bank of America	8
Citigroup	7
MetLife	1
Prudential	5
JP Morgan Chase	2

Morgan Stanley	8
Goldman Sachs	2
Lincoln National	16
Principal Financial Group	6
Wells Fargo	0
Hartford Financial Services	0
Capital One	0
State Street	3
Bank of New York Mellon	2
Ameriprise Financial	4
Suntrust Banks	4
Key Corp	3
Cit Group	0
Regions Financial	6
Zions Bancorporation	3

Table 12: the column indicates the number of other American financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its MES is statistically higher than the MES of 5 institutions in the sample.

These results are confirmed also for the European sample. Table 13 shows the results for the European financial institutions. Most of the time, the difference of the MES of two financial institutions is not statistically different from zero. Apart from two financial institutions (Societe Generale and Unicredit) which dominate 7 institutions, the others are difficult to discern. For example, BNP Paribas dominates only two financial institutions out of 10.

Financial Institution	**Number of Dominated Financial Institutions**
BNP Paribas	2
Credit Agricole	2
Deutsche Bank	4
Barclays	2
Societe Generale	7
HSBC	0
Royal Bank of Scotland	2

Banco Santander	2
Unicredit	7
London Stock Exchange	1

Table 13: the column indicates the number of other European financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its MES is statistically higher than the MES of 5 institutions in the sample.

These findings highlight an important weakness of MES. From the point of view of a regulator, MES can hardly be used to rank or identify which institution is more systemically important. It is difficult to establish a policy framework based on MES, because MES is not an adequate tool to measure the contribution to systemic risk of one financial institution. Even a bucketing approach seems difficult to apply. In a bucketing approach, we divide the financial institutions in buckets (or groups) that are characterized by an homogenous risk factor: therefore, institutions in the same bucket have more or less the same risk. These procedure cannot be applied using MES as our systemic risk measure, because all institutions would be classified in the same bucket. We will try again this approach later with SRISK.

Now we look at the time series variation of MES. So far, we have analysed the ability of MES in identifying systemically important financial institutions. Now we are interested in analysing the time variation of this systemic risk measure for a single institution.

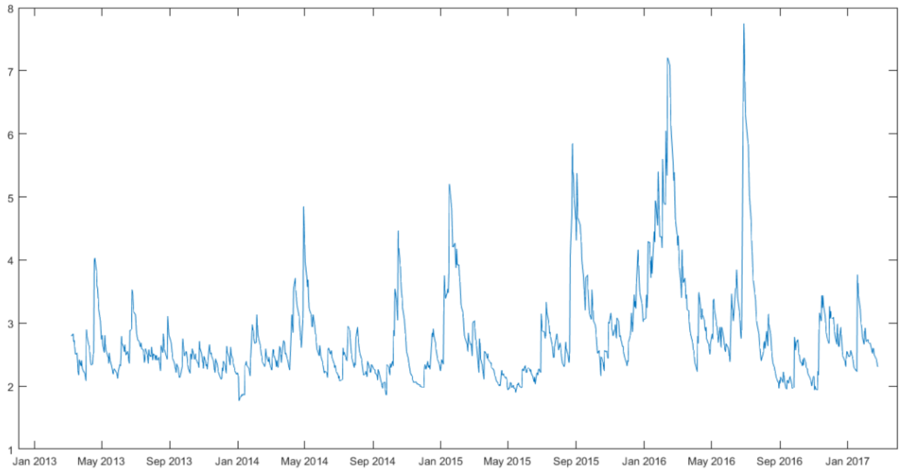

Figure 15: time-varying estimate of MES for Bank of America.

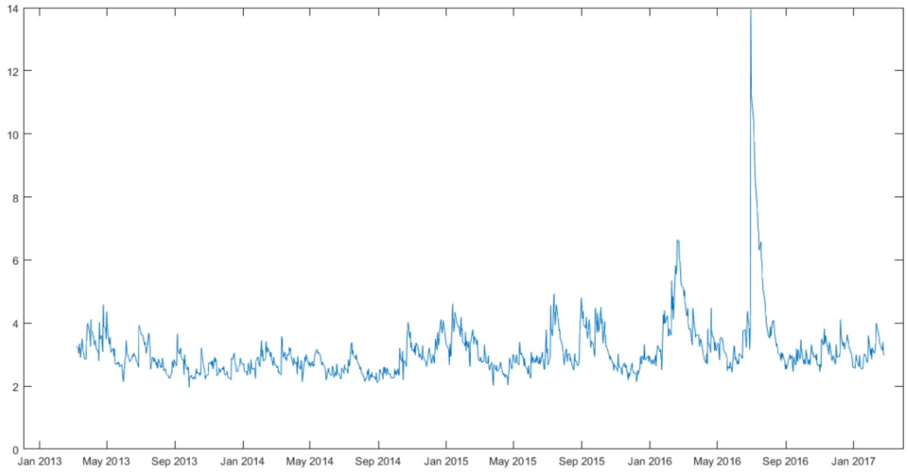

Figure 16: time-varying estimate of MES for BNP Paribas.

Figure 15 and Figure 16 show the time-varying estimate of MES for two different financial institutions. We can see that the time series of MES is highly volatile, as it is the one of Delta CoVaR. In our sample, MES of Bank of America is more volatile than the time series of BNP Paribas. This fact shows that MES is maybe too much volatile to be used as a regulatory tool to make decisions about the capital requirements of financial institutions. The capital requirements or the rankings of systemically important financial institutions based on MES would change too much rapidly and thus it will be impossible to establish a regulatory framework.

Let us now analyse the MES of two institutions over time in order to understand how they vary over time and if the difference between the time series of two financial institutions is relevant or not. If two institutions can hardly be distinguished by their MES over time, this creates a relevant problem in the application of MES to a regulatory framework.

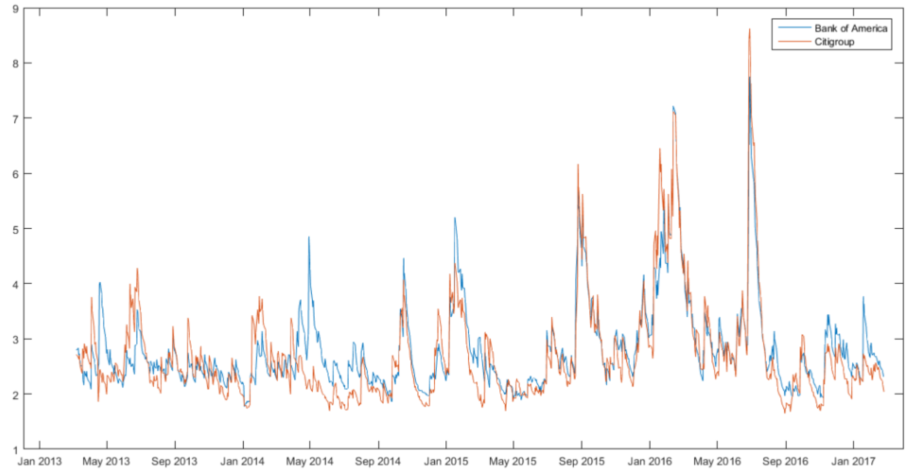

Figure 17: comparison of time-varying MES for two American institutions.

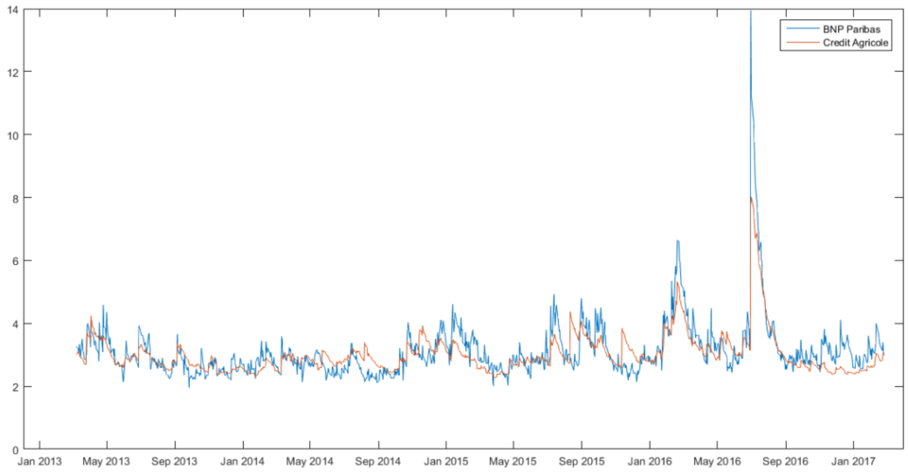

Figure 18: comparison of time-varying MES for two European institutions.

As we can see from Figure 17 and Figure 18 which represent respectively the American and the European case, it is hard to distinguish two financial institutions by their MES over time. The two time series are very much overlapped and this means that the correlation between the risk measure of the two institutions in very high.

Finally, we analyse the difference between Delta CoVaR and MES in a time series framework, in order to understand if these two systemic risk measures provide the same information or not.

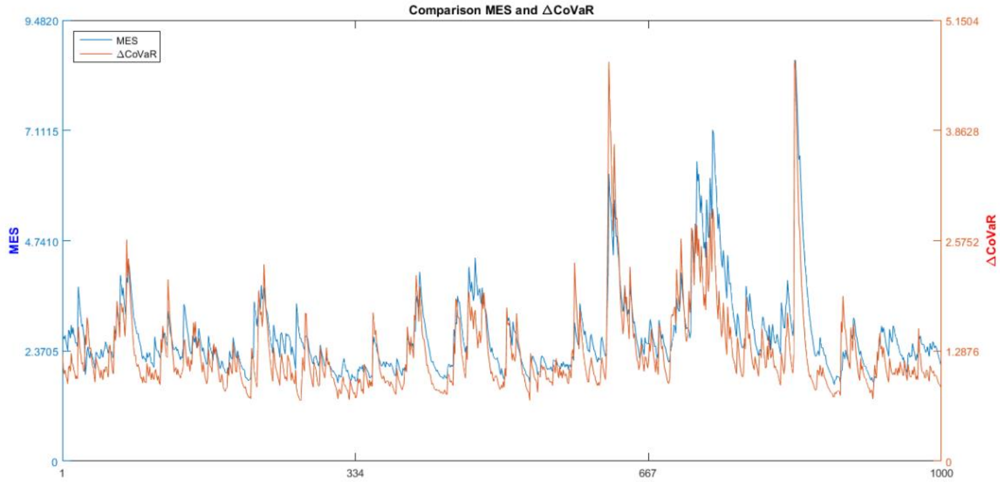

Figure 19: comparison of time-varying Delta CoVaR and MES for Citigroup.

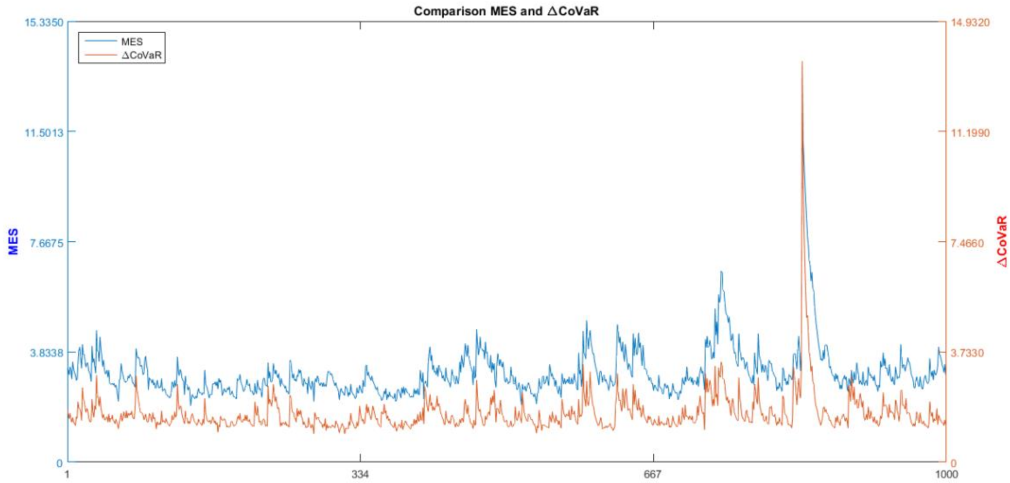

Figure 20: comparison of time-varying Delta CoVaR and MES for BNP Paribas.

Figure 19 and 20 show the time series of Delta CoVaR and MES for two financial institutions, respectively in the American and in the European case. We can see that in both cases the two systemic risk measures are highly correlated and thus we can conclude that they provide the same signal and therefore the same information for a regulator: a combination of both systemic risk measures would not be more informative than a single systemic risk measure taken in isolation. This is in contrast to what we have found for the cross-section between MES and Delta CoVaR, where we found no correlation between MES and Delta CoVaR.

In conclusion, despite the important theoretical differences between MES and Delta CoVaR, they seem to be equivalent, when we apply them to real data. On the top of this, MES seems to suffer from the same problems of Delta CoVaR: it can hardly identify and rank different financial institutions and it is also highly volatile in a time series framework. In addition, MES, as Delta CoVaR, takes into account only the so-called 'too-interconnected-to-fail' paradigm and not the 'too-big-to-fail' paradigm, missing important information. In the next paragraph we will analyse SRISK which is nothing else that a function of MES that tries to address and solve the problems of the other systemic risk measures.

2.3.3 Results for SRISK

The last systemic risk measure we are considering in our work is SRISK of Brownlees and Engle (2017). SRISK is in some kind different from the other systemic risk measures we have analysed so far, because it does not consider only the information provided by the stock returns, as Delta CoVaR and MES do. In fact, in order to compute SRISK we need also information about the market capitalization and the book value of the total liabilities of the institutions. From a theoretical point of view, we can say that SRISK is the only systemic risk measure that takes into account the 'too-interconnected-to-fail' paradigm (through the stock returns) and the 'too-big-to-fail' paradigm (through the market capitalization and the total liabilities value). This fact should positively affect the estimation of SRISK, because we have more information for each financial institution and this additional information is not as volatile as the stock returns. On the other hand, SRISK is nothing more than a function of MES and thus we can expect SRISK to suffer the same estimation problems of MES.

As we have seen in the theoretical part, the critical problem for the computations of SRISK is how to estimate the LRMES. There are mainly two methods: without simulation or with simulation. We will start by considering first the method without simulation. Since in the two previous paragraphs we have understood that between i.i.d. and block bootstrap there is not too much difference, we will present the results only for the block bootstrap.

LRMES without simulation is simply a nonlinear transformation of MES

$$LRMES = 1 - \exp(-18 * MES) . \tag{2.12}$$

Therefore we can take directly the results of the previous paragraph and use them in order to compute LRMES. We can use directly the percentile confidence intervals, because they are invariant to a monotone transformation: LRMES is a monotone transformation of MES because it involves only an exponential transformation. Once we have our estimate of LRMES, we use the other data and we compute SRISK.

Now we test for the significance of SRISK using the nonparametric test based on the block bootstrap percentile intervals. The null and alternative hypothesis are respectively:

$$\begin{aligned} &H_0 : SRISK_i = 0 \\ &H_1 : SRISK_i \neq 0 \end{aligned}. \tag{2.13}$$

At the 5% significance level, we find that 14 out of 20 American institutions have an SRISK statistically different from zero, thus they can be considered systemically important. SRISK produces different results with respect to the other systemic risk measures we analysed. When we tested Delta CoVaR and MES, all the institutions had a measure statistically different from zero. Instead in the case of SRISK, we can distinguish between institutions that are systemically relevant and those that are not. It is interesting to note that among the institutions that have a SRISK statistically equal to zero, there is also a big firm as JP Morgan Chase; instead all the other major players have a SRISK that is statistically significant. At the 5% significance level in the European sample, we find that all the institutions have a SRISK statistically different from zero. This fact can be interpreted as a sign that today there is much more systemic risk in Europe than in the US.

We now want to understand if the difference in SRISK between two financial institutions is statistically significant. We use the usual statistical test which null and alternative hypothesis are respectively:

$$\begin{aligned} &H_0 : SRISK_i = SRISK_j \rightarrow SRISK_i - SRISK_j = 0 \\ &H_1 : SRISK_i \neq SRISK_j \rightarrow SRISK_i - SRISK_j \neq 0 \end{aligned}. \tag{2.14}$$

In this case we reject the null hypothesis for a much larger number of institution pairs. Table 14 and Table 15 present the results for, respectively the American case and the

European case, at the 5% significance level, using the alternative hypothesis $SRISK_i > SRISK_j$.

Financial Institution	Number of Dominated Financial Institutions
Bank of America	15
Citigroup	17
MetLife	18
Prudential	16
JP Morgan Chase	1
Morgan Stanley	14
Goldman Sachs	9
Lincoln National	10
Principal Financial Group	8
Wells Fargo	0
Hartford Financial Services	9
Capital One	1
State Street	10
Bank of New York Mellon	5
Ameriprise Financial	4
Suntrust Banks	5
Key Corp	1
Cit Group	1
Regions Financial	5
Zions Bancorporation	1

Table 14: the column indicates the number of other American financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its SRISK is statistically higher than the SRISK of 5 institutions in the sample.

Financial Institution	Number of Dominated Financial Institutions
BNP Paribas	8
Credit Agricole	6
Deutsche Bank	8

Barclays	6
Societe Generale	5
HSBC	0
Royal Bank of Scotland	3
Banco Santander	1
Unicredit	4
London Stock Exchange	0

Table 15: the column indicates the number of other European financial institutions dominated by the institution in question. For example, if the institution in question dominates 5 institutions, this means that its SRISK is statistically higher than the SRISK of 5 institutions in the sample.

The results show that SRISK is able to detect more differences between institutions, from a systemic point of view. For example the SRISK of Bank of America is higher than the SRISK of other 15 institutions out of 20: with MES the number was only 8. In the European case, the SRISK of BNP Paribas is higher than the SRISK of 8 other institutions out of 10: with MES the number was only 2. Therefore, it is clear that SRISK is in some way better than Delta CoVaR and MES, because its measure is more heterogeneous between institutions and it is simpler to distinguish between them. But there is still considerable uncertainty about SRISK and thus also in this case it is not possible to build a ranking from position 1 to position N, because the risk of some institutions is equivalent and thus they should share the same position.

In alternative, we can try to divide the financial institutions into groups or buckets that represent statistically the same risk. We can try to divide our sample of financial institutions into buckets: institutions in different buckets have a risk measure that is statistically different from institutions in other groups.

Group 1	Bank of America
	Prudential
	Morgan Stanley
	Citigroup
	MetLife
	Lincoln

	National
Group 2	Goldman Sachs
	Principal Financial Group
	Hartford Financial Services
	State Street
Group 3	Capital One
	Key Corp
	Regions Financial

Table 16: American institutions divided in groups based on their SRISK and the statistical significance of this measure.

Table 16 considers the 14 US financial institutions that have a SRISK that is statistically different from zero. We can approximately divide them into 3 groups. The risk of each institution in a group is statistically different from the risk of institutions in other groups. This could be a valid alternative to rankings of positions, because it takes into account also the estimation risk of the risk measure and so we do not rank an institution in a higher ranking position just because its point estimate is higher, while it is not statistically different from other estimates. By looking at the group classification, we can notice that in addition to the big banks that we expect to find in the Group 1, we find also two insurance companies which usually are not believed as systemic as the banks. This classification could be useful for a regulator in order to monitor the systemic risk of the system and the contribution of each single financial institution.

Surprisingly, in the European case we were able to achieve a better classification of the firms, in comparison to the American one. We can divide the 10 institutions into 6 groups. Three groups contain only one institution so they are not actually groups.

Group 1	BNP Paribas
	Deutsche Bank
Group 2	Credit Agricole
	Barclays
Group 3	Societe Generale
Group 4	Unicredit
Group 5	Royal Bank of Scotland
Group 6	HSBC
	Banco Santander
	London Stock Exchange

Table 17: European institutions divided in groups based on their SRISK and the statistical significance of this measure.

What we can notice is that in the European case there are two institutions that lead the ranking: BNP Paribas and Deutsche Bank. They are followed by two other financial institutions: Credit Agricole and Barclays. Then we have three firms that are statistically different between each other. Finally, we have a group of three institutions that are less risky. What we can notice is that the London Stock Exchange is as systemic as a big bank as HSBC, at least in our sample. This is quite surprising, because we are not expecting that a stock exchange is relevant in the contribution of systemic risk, since it does not take too much risk in its business model and its leverage is not so relevant. Indeed, our results are confirmed by Engle et al. (2015). They find that the institutions that mainly contribute to the systemic risk of Europe are all banks, namely Deutsche Bank, Credit Agricole, Barclays, Royal Bank of Scotland and BNP Paribas. They do not include a measure of the uncertainty of their estimates, but their ranking is very similar to ours.

The results we find on SRISK highlight also a weakness of the too-big-to-fail paradigm. If we look at two of the biggest financial institutions in the world, namely JP Morgan Chase and HSBC, we can see that they are not as systemic as one could though; even if they are the largest institutions measured by market capitalization, their risk is lower than other smaller financial institutions. This means that the size of an institution measured by its market capitalization is only an ambiguous measure of its systemic risk.

We also consider the simulation-based method used in Brownlees and Engle (2017) in order to compute LRMES. The steps of the simulation procedure are as follows. First we estimate a GARCH-DCC model for the returns of the system and the financial institution. The univariate volatilities are modelled using a GJR-GARCH specifications which take into account asymmetric effects of the negative returns on the level of volatility. Once we have estimated our model, we can use it in order to simulate paths of the returns for the system and for the financial institution in question. In order to compute LRMES, we need only the paths where the system fall by 40% or more in a time period of 6 months (126 observations): LRMES is simply the average of the cumulated returns of the financial institution, when the system is down by 40% or more. We use data for 2013 to 2017 in order to estimate our model: we take as an example Bank of America. We then obtain 100,000 simulated paths and compute the relative quantity of interest. The problem of this type of simulation is that the probability that the system falls by 40% or more is very small and thus in order to compute LRMES, we use only a very restricted part of the simulated paths. In fact, out of the 100,000 paths simulated, only in 126 cases the system falls by 40% or more: this corresponds to 0.126% of the cases which indicates that is a very rare event. Our simulation procedure produces a very large standard error. There are two reasons that can explain our result: the first one is a problem in the simulation procedure which could not be efficient, and the second one is that the uncertainty of the model is high and so we cannot really know the value of LRMES with precision. Unfortunately, we cannot compare our standard error with the one of the original work of Brownlees and Engle (2017), but one of the authors said that the standard error of their simulation was quite large, but not as large as ours. Probably, if we applied some variance reduction techniques, we would obtain a smaller standard error. But still, we believe that there is considerable uncertainty also about SRISK and the methods to compute LRMES. Indeed, if we look at the rankings published on V-Lab by the New York University, where SRISK is computed with or without simulation, we can notice that they have considerable differences. For example on May 27, 2017 the SRISK of Bank of America computed without simulation is \$86,461 million with an LRMES of 66.21%; instead the same measure computed with simulation is \$60,236 million with a LRMES of 53.88%. And we have additional uncertainty if we consider that we can change the capital requirement ratio which in this case is fixed at 8%. Coleman et al. (2017) show that SRISK is very sensible to the chosen capital requirement ratio.

SRISK should be also considered as a valid alternative to the stress tests provided by the Federal Reserve System or by the European Central Bank (Acharya et al., 2014). But they also point out that the results between their stress tests based on SRISK and the ones provided by central banks have surprisingly conflicting results. This is also true for our work. At the time of writing, the most recent documentation for Europe is the 2016 European Union bank stress test conducted by the European Banking Authority. According to the results of the test, no bank will have failed in the stressed conditions of the test. Instead, our results show clearly that there are two financial institutions (namely BNP Paribas and Deutsche Bank) whose risk is not negligible and should be considered carefully by regulators, because of the amount of money they would require in case of a bail-out. This is also confirmed by the rankings provided by the V-Lab website. Our findings highlight the pointlessness of these kind of tests, because they do not reflect the consensus of the market. They are only made to impart confidence as explained in Elliot (2016). On the other hand, stress tests established by the Federal Reserve for US financial institutions are more reliable. The Fed tested also American subsidiaries of European institutions. In June 2016, nearly all US banks passed the Fed’s stress test: only Morgan Stanley did not receive an unconditional passing grade. Indeed, Deutsche Bank and Banco Santander (their American division) failed the test (Corkery and Popper, 2016). According to our results, Morgan Stanley was included in Group 1 of our ranking, thus we should expect this situation. More interesting is instead the situation of the other two banks: while, according to the American regulators, they would be a big concern if a new financial crisis exploded, they are not considered a problem according to the European regulators. This fact highlights the unreliability of European stress test and why the results of the analysis of SRISK could be useful in future times.

In this paragraph, we have shown that SRISK provides a better estimate of systemic risk compared to other measures such as Delta CoVaR and MES, even if SRISK is strictly related to MES. We think that this is due to the fact that SRISK exploits a larger array of information that it is not limited only to market stock returns. However, more information is not always better information, even in this case. Relying on balance sheet data could be misleading as there are a lot of cases of balance sheet manipulation, even in the recent financial crisis as we have seen. Indeed, the assumption regarding the debt structure during the crisis necessary in order to estimate SRISK could be a serious pitfall. SRISK has for sure addressed and resolved problems that are present in other systemic risk measures but there are still problems also regarding SRISK. For example this measure, as the others analysed in this work, suffers from the problem of procyclicality. This is a serious problem especially when we compute SRISK, because we use data where the volatility is low to estimate a model. And then we use this model to simulate future paths of the returns in order to simulate a crisis and then compute the measure in question. But this could lead to results that are smaller than what we would expect if a crisis actually happened. In conclusion, from our point of view, SRISK is still superior to the other systemic risk measures.

Conclusions

In this work we analyse recently developed methods to measure the contribution of a financial institution to systemic risk which are based on market data. This is an important topic, because a regulator should base capital requirements for financial institutions upon its contribution to systemic risk (Crockett, 2000). We show that despite of the elegant theoretical foundations of these risk measures, from an empirical point of view, they have serious limits. In particular, we show that the estimation risk is high and it is not negligible. We think that this factor should be taken into account when we use these measures to analyse systemically important financial institutions (SIFIs).

There are many approaches to analyse the systemic risk of financial institutions using market data. For sure, the approach that has received much more attention is the one of Adrian and Brunnermeier (2016). In fact their work is the one with the highest number of citations and a lot of scholars tested this systemic risk measure with different samples. Despite its popularity, we show that this systemic risk measure, when estimated with quantile regression, suffers from an high estimation risk. In order to do that we apply bootstrap techniques to compute standard errors and confidence intervals. This uncertainty about the statistical measure translates into uncertainty about the decisions that a regulator can take based on this measure. And this is a very undesirable feature of Delta CoVaR.

But this is not only a problem of Delta CoVaR. In fact we show that also MES suffers from the same problem. Despite this fact, this problem has been in some way resolved by using a more sophisticated measure based on MES which is SRISK. We can interpret SRISK as a function of MES which, a part from the information provided by the returns, takes into account also other types of information such as the leverage of an institution, computed using balance sheet values. We think that, by adding relevant information to the computations of systemic risk measures, we can obtain a more precise estimate of what would be the contribution that a financial institution brings to the risk of the system. From this point of view, SRISK represents a step forward with respect to the other systemic risk measures such as Delta CoVaR and MES. We also point out that sometimes the information provided by the balance sheet could be misleading and thus we can end up with a leverage that is not the true leverage of the financial institution, underestimating their risk.

A simple conclusion of our analysis is that, based on these measures, it is not possible to build a ranking with single positions for each financial institution and that the estimation risk of these measures should be considered: for this reason we have developed hypothesis tests based on confidence intervals to assess if two measures are statistically different. The best we can do is to divide our sample of institutions into groups; each group represents a class of risk. The rules to construct these groups are two: if two institutions have systemic risk measures that are not statistically different, they should be put in the same group; instead if two institutions have systemic risk measures that are statistically different, they should be put in different groups. We can imagine that institutions that are part of the group with the highest risk will be subject to a special attention and maybe to a special regulation by the policymakers. This approach is very attractive, because it is free and it is very rapid to implement; it could be competitive with stress tests (Acharya et al., 2014). Stress tests are very expensive and they require several moths in order to review the value of each asset in the balance sheet of a bank. In addition, they can also be biased by political decisions as in the European case. Indeed, they apply only to banks, while the approach we present can be applied to all kind of financial institutions, not only banks.

Our results for SRISK highlight that in the US there is a group of six financial institutions that represents the main risk. It is interesting to note that apart from banks, also insurance companies are systemically important. This is in contrast with the popular wisdom for which we have to worry only about banks. In addition, a big bank as JP Morgan Chase is found not systemically relevant. In Europe the situation is slightly different. We have two banks that represent the main risk according to our analysis. We think that they should be subject to particular attention by the regulators, but according to the results of stress tests this is not the case.

As a second step, we investigate the reasons of the large uncertainty about these systemic risk measures. We think there are mainly two reasons that can explain this uncertainty. First, systemic risk measures are strictly linked to the left tail of a distribution of the returns. Estimating the tail of a distribution is a demanding task, because usually we do not have a lot of observations in the tail of the distribution and so we cannot precisely infer how the returns behave in their tail. Secondly, since VaR and ES have been used as the basis for constructing Delta CoVaR and MES, we can think that the estimation risk embedded in VaR and ES is in some way shared with the more sophisticated Delta CoVaR and MES.

Given our results, we think that there are important implications related to how statistical risk measures for systemic risk should be used in decision-making by policymakers or regulators. As suggested by Danielsson and Zhou (2017), we think that the more useful approach is to use these statistical risk measures as one indication among others of the underlying systemic risk. Indeed, one may use different models simultaneously to have a more clear picture of the real systemic risk. These systemic risk measures are not the ‘truth’, but at the same time they are not completely useless: they provide a noisy signal of the risk we want to measure. For this reason, even if our results are in a sense pessimistic, we believe that systemic risk measures should be computed and reported, but a detail analysis of the statistical accuracy should be reported too (e.g. presenting confidence intervals instead of point estimates). We think this is the best way to take important decisions based on these measures.

Further research should be focused on elaborating new models that take advantage of a more wide array of information. In fact we believe that the information provided by stock prices alone is not sufficient in order to have a clear picture of the contribution to systemic risk of a financial institution. More sophisticated systemic risk measures should use also balance sheet information or information coming from the prices of other financial assets (e.g. bond prices, option prices and CDS spreads). Then, after we have reliable systemic risk measures, the focus should be on developing a regulatory framework based on these measures. The main task of this new regulatory framework is to impose capital requirements that are proportional to the systemic risk of the financial institution. In this way each institution pays for the externalities that might impose to the real economy.

References

Abdymomunov, A., 2013. Regime-switching Measure of Systemic Financial Stress. *Annals of Finance,* 9(3), pp. 455-470.

Acharya, V., Engle, R. and Pierret, D., 2014. Testing Macroprudential Stress Tests: The Risk of Regualtory Risk Weights. *Journal of Monetary Economics,* Volume 65, pp. 36-53.

Acharya, V., Engle, R. and Richardson, M., 2012. Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks. *American Economic Review: Papers & Proceedings*, 102(3), pp. 59-64.

Acharya, V., Pedersen, L., Philippon, T. and Richardson, M., 2017. Measuring Systemic Risk. *The Review of Financial Studies,* 30(1), pp. 2-45.

Adrian, T. and Boyarchenko, N., 2012. *Intermediary Leverage Cycles and Financial Stability.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2133385 [Accessed 03 June 2017].

Adrian, T. and Brunnermeier, M., 2011. *CoVaR.* [Online] Available at: http://econpapers.repec.org/paper/nbrnberwo/17454.htm [Accessed 03 June 2017].

Adrian, T. and Brunnermeier, M., 2016. CoVaR. *American Economic Review,* 106(7), pp. 1705-1741.

Alessi, L. and Detken, C., 2011. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity. *European Journal of Political Economy,* 27(3), pp. 520-533.

Allen, L., Bali, T. G. and Tang, Y., 2012. Does Systemic Risk in the Financial Sector Predict Future Economic Downturns?. *The Review of Financial Studies,* 25(10), pp. 3000-3036.

Banulescu, G.-D. and Dumitrescu, E.-I., 2015. Which are the SIFIs? A Component Expected Shortfall Approach to Systemic Risk. *Journal of Banking & Finance,* Volume 50, pp. 575-588.

Benoit, S., Colletaz, G., Hurlin, C. and Perignon, C., 2013. *A Theoretical and Empirical Comparison of Systemic Risk Measures.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1973950 [Accessed 03 June 2017].

Benoit, S., Colliard, J.-E., Hurlin, C. and Pérignon, C., 2016. Where the Risks Lie: A Survey on Systemic. *Review of Finance,* pp. 1-44.

Bernardi, M., Gayraud, G. and Petrella, L., 2013. *Bayesian Inference for CoVaR.* [Online] Available at: https://arxiv.org/pdf/1306.2834.pdf [Accessed 03 June 2017].

Billio, M., Getmansky, M., Lo, A. and Pellizon, L., 2012. Econometric Measures of Connectedness and Systemic Risk in the Financial and Insurance Sector. *Journal of Financial Economics,* 104(3), pp. 535-559.

Bisias, D., Flood, M., Lo, A. W. and Valavanis, S., 2012. A Survey of Systemic Risk Analytics. *The Annual Review of Financial Economics,* Volume 4, pp. 255-296.

Bluhm, M. and Krahnen, J. P., 2014. Systemic Risk in an Interconnected Banking System with Endogenous Asset Markets. *Journal of Financial Stability,* Volume 13, pp. 75-94.

Borio, C. and Drehmann, M., 2012. *Assessing the Risk of Banking Crisis - Revised.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1513316 [Accessed 03 June 2017].

Borri, N., Caccavaio, M., Di Giorgio, G. and Sorrentino, A. M., 2012. *Systemic Risk in the European Banking Sector.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2112919 [Accessed 03 June 2017].

Brownlees, C. and Engle, R., 2017. SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *The Review of Financial Studies,* 30(1), pp. 48-79.

Cao, Z., 2014. *Multi-CoVaR and Shapley Values: A Systemic Risk Measure.* [Online] Available at: http://risk2015.institutlouisbachelier.org/uploads/440754460_Cao_systemic_risk_measure_Risk_forum.pdf
[Accessed 03 June 2017].

Castro, C. and Ferrari, S., 2014. Measuring and Testing for the Systemically Important Financial Institutions. *Journal of Empirical Finance,* Volume 25, pp. 1-14.

Chen, S. X., 2008. Nonparametric Estimation of Expected Shortfall. *Journal of Financial Econometrics,* 6(1), pp. 87-107.

Chernick, M. R., 2007. *Bootstrap Methods: A Guide for Practitioners and Researchers.* Second Edition. Hoboken, NJ: John Wiley & Sons.

Chernick, M. R. and LaBudde, R. A., 2011. *An Introduction to Bootstrap Methods with Applicatioins to R.* Hoboken, NJ: John Wiley & Sons.

Christoffersen, P. F. and Goncalves, S., 2005. Estimation Risk in Financial Risk Management. *Journal of Risk,* pp. 1-29.

Coleman, T. F., LaPlante, A. and Rubstov, A., 2017. *Analysis of the SRISK Measure and Its Application to the Canadian Banking and Insurance Industries.* [Online] Available at: http://globalriskinstitute.org/wp-content/uploads/2017/05/Adjusted-SRISK-FINAL.pdf
[Accessed 29 May 2017].

Corkery, M. and Popper, N., 2016. *Nearly All U.S. Banks Pass Fed's Stress Test.* [Online]
Available at: https://www.nytimes.com/2016/06/30/business/dealbook/nearly-all-us-banks-pass-feds-stress-test.html?_r=1
[Accessed 05 June 2017].

Crocket, A., 2000. *Marrying the Micro- and Macro-Prudential Dimensions of Financial Stability.* [Online]

Available at: https://www.bis.org/review/rr000921b.pdf [Accessed 03 June 2017].

Danielsson, J., James, K. R., Valenzuela, M. and Zer, I., 2016. Model Risk of Risk Models. *Journal of Financial Stability,* Volume 23, pp. 79-91.

Danielsson, J. and Zhou, C., 2017. *Why Risk Is So Hard to Measure.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2597563 [Accessed 03 June 2017].

Daniel, T., 2009. *Regulatory Restructuring, Testimony before the US Committee on Banking, Housing and Urban Affairs* (23 July 2009).

Drehmann, M. and Tarashev, N., 2013. Measuring the Systemic Importance of Interconnected Banks. *Journal of Financial Intermediation,* Volume 22, pp. 586-607.

ECB, December 2009. *Financial Stability Review,* s.l.: European Central Bank.

Efron, B. and Tibshirani, R., 1993. *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Elliot, L., 2016. *The EBA's stress tests reveal their own lack of credibility.* [Online] Available at: https://www.theguardian.com/business/economics-blog/2016/aug/01/eba-stress-tests-reveal-their-own-lack-credibility#img-1
[Accessed 05 June 2017].

Engle, R., 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistic,* 20(3), pp. 339-350.

Engle, R., 2016. Dynamic Conditional Beta. *Journal of Financial Econometrics,* 14(4), pp. 643-667.

Engle, R., Jondeau, E. and Rockinger, M., 2015. Systemic Risk In Europe. *Review of Finance,* Volume 19, pp. 145-190.

Financial Stability Board, 2010. *Reducing the Moral Hazard Posed by Systemically Important Financial Institutions.* [Online]

Available at: http://www.fsb.org/wp-content/uploads/r_101111a.pdf [Accessed 29 May 2017].

Gabaix, X., 2009. Power Laws in Economics and Finance. *Annual Review of Economics,* 1(1), pp. 225-294.

Girardi, G. and Ergun, T. A., 2013. Systemic Risk Measurement: Multivariate GARCH Estimation of CoVaR. *Journal of Banking & Finance,* 37(8), pp. 3169-3180.

Glosten, L. R., Jagannathan, R. and Runkle, D., 1993. Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance,* Volume 48, p. 1779–1802.

Gray, D., Merton, R. C. and Bodie, Z., 2007. *New Framework for Measuring and Managing Macrofinancial Risk and Financial Stability*. [Online] Available at: http://www.nber.org/papers/w13607.pdf [Accessed 03 June 2017].

Huang, X., Zhou, H. and Zhu, H., 2009. A Framework for Assessig the Systemic Risk of Major Financilal Institutions. *Journal of Banking & Finance,* Volume 33, pp. 2036-2049.

Hull, J. C., 2014. *Risk Management and Financial Institutions.* Fourth Edition, Hoboken, NJ: John Wiley & Sons, Inc..

Idier, J., Lamé, G. and Mésonnier, J.-S., 2014. How Useful is the Marginal Expected Shortfall for the Measurement of Systemic Exposure? A Practical Assessment. *Journal of Banking & Finance,* Volume 47, pp. 134-146.

IMF, April 2009. *Global Financial Stability Report,* Washington DC: International Monetary Found.

Jobst, A. and Gray, D., 2013. *Systemic Contingent Claims Analysis – Estimating Market-Implied Systemic Risk.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2233756 [Accessed 03 June 2017].

Jorion, P., 1996. Risk^2: Measuring the Risk in Value-At-Risk. *Financial Analysts Journal*, Volume 52, pp. 47-56.

Kupiec, P. and Guntay, L., 2016. Testing for Systemic Risk Using Stock Returns. *Journal of Financial Services Research,* Volume 49, pp. 203-227.

Lahiri, S. N., 2003. *Resampling Methods for Dependent Data.* New York: Springer.

Lehar, A., 2005. Measuring Systemic Risk: A Risk Management Approach. *Journal of Banking & Finance,* 29(10), pp. 2577-2603.

Lucas, D., 2014. Comment on: 'Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights' by Viral Acharya, Robert Engle and Diane Pierret. *Journal of Monetary Economics,* Volume 65, pp. 54-56.

Mainik, G. and Schaanning, E., 2014. On Dependence Consistency of CoVaR and Some Other Systemic Risk Measures. *Statistics & Risk Modeling with Applications in Finance and Insurance,* 31(1), pp. 49-77.

Markose, S., Giansante, S. and Shaghaghi, A. R., 2012. 'Too Interconnected to Fail' Financial Network of US CDS Market: Topological Fragility and Systemic Risk. *Journal of Economic Behavior & Organization,* 83(3), pp. 627-646.

Martinez-Jaramillo, S., Perez, P. O., Embriz, F. A. and Gallo Dey, F. L., 2010. Systemic Risk, Financial Contagion and Financial Fragility. *Journal of Economic Dynamics and Control*, 34(11), pp. 2358-2374.

Oh, D. H. and Patton, A. J., 2016. Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads. *Journal of Business & Economic Statistics,* pp. 1-47.

Patro, D., Qi, M. and Sun, X., 2013. A Simple Indicator of Systemic Risk. *Journal of Financial Stability,* 9(1), pp. 105-116.

Politis, D. N. and White, A., 2004. Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews,* 23(1), pp. 53-70.

Ruppert, D., 2011. *Statistics and Data Analysis for Financial Engineering.* New York: Springer Texts in Statistics.

Segoviano, M. and Goodhart, C., 2009. *Banking Stability Measures.* [Online] Available at:

https://papers.ssrn.com/sol3/Delivery.cfm/wp094.pdf?abstractid=1356460&mirid=1
[Accessed 03 June 2017].

Suh, S., 2012. Measuring Systemic Risk: A Factor-Agumented Correlated Default Approach. *Journal of Financial Intermediation*, Volume 21, pp. 341-358.

Tarashev, N. A., Borio, C. and Tsatsaronis, K., 2013. *Attributing Systemic Risk to Individual Institutions.* [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1631761
[Accessed 03 June 2017].

Tarullo, D., 2009. *Regulatory Restructuting: Testimony before the Committee on Banking, Housing and Urban Affairs at US Senate.* [Online] Available at: https://www.federalreserve.gov/newsevents/testimony/tarullo20090723a.htm
[Accessed 29 May 2017].

Valukas, A. R., 2010. *Lehman Brothers Holdings Inc. Chapter 11 Proceedings Examiner's Report.* [Online] Available at: https://web.stanford.edu/~jbulow/Lehmandocs/menu.html
[Accessed 03 June 2017].

Zhou, C., 2010. Are Banks Too Big to Fail? Measuring Systemic Importance of Financial Institutions. *International Journal of Central Banking*, Volume 6, pp. 205-250.

Appendix A

List of US Financial Institutions

Financial Institution	Symbol
Bank of America	BAC
Citigroup	C
MetLife	MET
Prudential	PRU
JP Morgan Chase	JPM
Morgan Stanley	MS
Goldman Sachs	GS
Lincoln National	LNC
Principal Financial Group	PFG
Wells Fargo	WFC
Hartford Financial Services	HIG
Capital One	COF
State Street	STT
Bank of New York Mellon	BK
Ameriprise Financial	AMP
Suntrust Banks	STI
Key Corp	KEY
Cit Group	CIT
Regions Financial	RF
Zions Bancorporation	ZION

List of European Financial Institutions

Financial Institution	Number of Dominated Financial Institutions
BNP Paribas	BNP
Credit Agricole	CAGR

Deutsche Bank	DBK
Barclays	BARC
Societe Generale	SOGEN
HSBC	HSBC
Royal Bank of Scotland	RBS
Banco Santander	BSAN
Unicredit	UNI
London Stock Exchange	LSE