University of Padua

Department of mathematics

Master Degree in Data Science

MASTER THESIS

# A STUDY ON THE DEPENDENCE AMONG SPANISH BANKS VIA QUANTILE REGRESSION FORESTS AND NETWORK ANALYSIS

**Supervisor:** Prof. BRUNO SCARPA

Department of Statistics

University of Padua

**Co-supervisor:** Prof. CATALINA BOLANCE LOSILLA

Department of Econometrics

University of Barcelona

**Student:** FRANCESCA DE ROBERTIS

Student N° 1206031

# Abstract

The present work is focused on the estimation of CoVaR and ΔCoVaR, two measures that are used to quantify the effects of a financial institution or a financial system in distress on the VaR of another financial institution. Firstly, the studied case consists of five Spanish banks, namely: BBVA, Bankia, Bankinter, CaixaBank, Santander and, in the second part, the MSCI Europe index is used to analyze the financial system effects. The method adopted is the Quantile Regression Forests and the results are compared with the already well-known Quantile Regression. The last part of the thesis studies the relations among the banks and the system as a network. Considering additional European banks, a community detection analysis is performed.

# Contents

# Chapter 1

# Introduction

In the late 2008 the financial crisis has radically changed the entire world economy for more than a decade. This event has highlighted the importance of the concept of systemic risk for all financial institutions.

The concept of systemic risk is not uniquely identified by the literature, it generally refers to the risk that financial instability becomes so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially.

Research advances in the study of financial instability, in particular when combined with market experience and current policy analysis, have been of great help in further developing macro-prudential supervision and in supporting the development of indices that are currently being adopted by financial institutions to deal with future systemic risks.

The Value-at-Risk (VaR) is a measure of the risk of loss for investments. It estimates how much a set of investments might lose - with a certain probability -, given normal market conditions, in a set time period such as a day. The VaR is typically used by firms and regulators in the financial industry to gauge the amount of assets needed to cover possible losses and this is called the Solvency Capital Requirement (SCR). The newer Basel accords (Basel II and Basel III) acknowledge the important role of the VaR as a basis of risk measurement and regulatory capital calculation. Particularly, for regulatory purposes, a risk measure approach needs to have the ability to adequately capture all the risks faced from an institution, which encompass market risk, credit risk and operational risk. The VaR is commonly adapted as an index able to summarize the downside risk of any institution arising out of financial market variability.

In this thesis we will consider the Spanish economy by taking into account the bank sector. An European Financial index is chosen as a benchmark to study the market behaviour while analysing the relationships between five Spanish banks: BBVA, Bankinter, CaixaBank, Bankia and Santander. The aim of this study is trying to understand how the distress in one particular institution increases the VaR of another institution. Although VaR is one of the most widely risk measure for quantifying a financial institution's maximum loss for a given confidence level and time horizon, this measure is centered on the individual risk of an institution and fails to consider potential spillover effects on other institutions and its connection to the over-all systemic risk.

In order to take into account the contribution of individual banks to the risk of other institutions and to the financial system various systemic risk measures that have been developed, we focus our attention on the one proposed by Adrian and Brunnermeier in 2011 [5]. They proposed a conditional VaR called CoVaR, where the prefix 'Co' is added to emphasize the systemic nature of the estimated metric, since this index is able to capture possible risk spillovers between financial institutions.

Different ways of computing CoVaR have been proposed: Quantile Regression (QR), GARCH method, Bayesian, Copula approaches and so on. The present work wants to develop an alternative method for the CoVaR index computation, based on Quantile Regression Forests (QRF).

QRF is an effective method able to study the entire distribution of a random variable, differently from the traditional Random Forests methods that estimate only the expectation of the studied variables conditioned to the regressors. We measure the systemic impact of financial distress from one Spanish bank on other banks and on the European financial system using weekly data for the period from February 2014 to April 2020.

Finally, the last part of the work will be focused on the analysis of the five banks and the financial index as a whole. We will construct a network to understand whether the systemic effect from bank $A$ to bank $B$ imply similar systemic effects when considering the effect from bank $B$ to bank $A$. Based on this methodology, a study on the relations among a larger number of European financial institutions is performed, supported also by community detection techniques.

Besides the CoVaR, other systemic risk measures exist. In [8] the authors develop a conditional autoregressive VaR (CAViaR) model that uses Quantile Regression to capture the tail behaviour of returns. Authors of [9] develop an index for systemic financial distress given by the price of credit default swaps (CDS). Using CDS data, in [10] a banking stability index is constructed with the aim to assess interbank dependence for tail events. [11] introduces systemic expected shortfall and marginal expected shortfall as indicators to quantify downside risk and the contributions of financial institutions to risk. Another work [12] proposes a measure of aggregate systemic risk called CATFIN that can predict declines in aggregate bank lending activity 6 months in advance.

Recent papers extend the $\Delta$CoVaR method and apply it to additional financial sectors [13,14]. In this thesis we will be focused on CoVaR and $\Delta$CoVaR estimation, so it is relevant to highlight that multiple methods have been proposed in the literature. CoVaR is estimated using multivariate GARCH by [15]. The authors of [16] and [17] use copulas methods. Bayesian inference for CoVaR estimation is proposed by [18], while [19] make distributional assumptions about shocks and employ maximum likelihood estimators. The method proposed in this thesis differs from these ones and adopt for the first time a Quantile Regression Forests model.

The structure of the rest of this thesis is as follow:
in Chapter II we describe the systemic risk measures, how they differ from each other, analysing pros and cons of each one, we also introduce the used dataset. In Chapter III we present an introduction to the technical instruments adopted, discussing the regression problem and the techniques adopted to deal with it. In Chapter IV there are the results of the CoVaR estimation and the presentation of the results. In Chapter V we analyze the results obtained by including systemic state variables in the estimation process, that let the CoVaR to be a time-variant variable.

Chapter VI regards the analysis performed including the system into the study. In Chapter VII we discuss the relations among the system and the banks quantified in terms of $\Delta$CoVaR adopting a network analysis and community detection methods. Finally, in Chapter VIII concluding remarks are presented. Notice that the drafting of the present work has been started under the supervision of professor Catalina Bolancé Losilla from University of Barcelona. Then, the work has continued under the supervision of professor Bruno Scarpa from University of Padua.

# Chapter 2

# Theoretical framework

## 2.1 What is the systemic risk?

During the years economists have given different definitions of systemic risk by focusing on different aspects.

According to De Band and Hartmann in [20] it is connected to the disruption of a financial system, having major negative consequences for the real economy, caused by breakdowns in all or parts of the system, at the same time.

Kaufamn and Scott in [21] underline the importance of three concepts: the incidence of a shock having large and simultaneous effects on the entire parts of the system, rather than effecting one or a few institutions; the second and the third concepts are related to the shock transmission across the different institutions. They associated to the systemic risk the idea of *chain reactions* and *domino effects* in the system.

In addition to that, in [22], Rochet and Tirol refer systemic risk to "the propagation of an agent's economic distress to other agents linked to that agent through financial transactions".

## 2.2 Measuring systemic risk

### 2.2.1 Value-at-Risk

One of the most common risk measures used by financial institutions is the Value-at-Risk (VaR). It measures the potential loss in the value of a risky asset or portfolio over a defined period for a given confidence interval.

The $q\%$-VaR$^i$ is the maximum loss of institution $i$ at the $(1-q)\%$ confidence level. For example, if the VaR on an asset is €100 million at a one-week, 95% confidence level, there is a only a 5% chance that the value of the asset will drop more than €100 million over any given week. Regarding the confidence level, it usually varies between 95% to 99%, in relation to specific requests. The time horizon varies between 1-10 days, depending on the type of activity: the more liquid is the asset, the shorter will be the projection horizon. In other words, the VaR can be considered a technique that shows, in terms of money, the risk that a person that owns a capital can take. Obviously, we have to take into account that the wider and complex the portfolio is, the more difficult the VaR computation will be: in this case one has to consider possible relations among them.

The innovative element in the VaR is that it can be seen, in statistical terms, as the standard deviation of the $P\&L$ (profit and loss)[1]: the larger is the volatility (standard deviation), the higher is the risk and the potential monetary losses. In Fig. 2.1 it is shown an example of VaR. It can be computed in different ways, the most common are:

- Variance-Covariance approach;
- Historical Simulation;
- Monte-Carlo simulation;
- Cornish-Fisher VaR.

[1] $P\&L$ (profit and loss) is a financial statement that summarizes the revenues, costs, and expenses incurred during a specified period, usually a fiscal quarter or year.

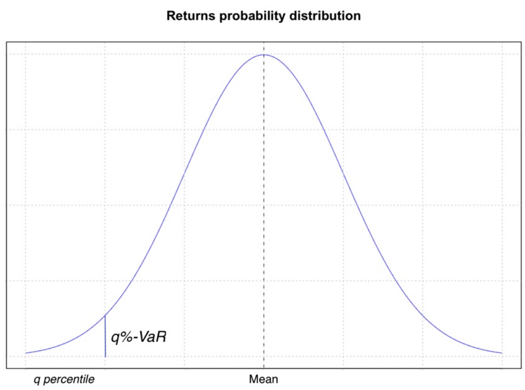


**Figure 2.1:** Example of $q$%-VaR.

The Variance-Covariance method is the only parametric approach among the just mentioned methods. It is based on a specific statistical reference model. This implies the assumption that the distribution of the profits or the losses is normal and so, computing in an easier way the mean and the standard deviation. The other three methods are not parametric and they do not ask for a particular distribution of the data. In particular, the Historical Simulation studies the behaviour of time series that should be long enough. On the other hand, Monte Carlo simulation consists in generating a sufficiently large number of scenarios that can be used to obtain a distribution of possible losses. When the returns are normally distributed, Cornish-Fisher VaR collapses to traditional definition of VaR.

### 2.2.2 Expected Shortfall

Expected Shortfall, also called average Value-at-Risk (AVaR), is a risk measure sensitive to the shape of the tail of the distribution of returns on a portfolio. Given a certain confidence level, this measure represents the expected loss when it is greater than the value of the VaR calculated with that confidence level. It is the average of the portfolio returns that are lower than the value of VaR at that confidence level. Mathematically, it is defined as:

$$ES_\alpha = -\frac{1}{\alpha}\int_0^\alpha VaR_q(X)d_q. \tag{2.1}$$

ES is often characterised as measuring losses "beyond VaR" in the sense that it evaluates the tail of large losses and calculates the average of the losses greater than VaR.

### 2.2.3 Conditional Value-at-Risk

Adrian and Brunnermeier in [5] proposed CoVaR to measure the contribution of financial institutions to systemic risk or the strength of risk spillover of a single financial institution to other financial institutions.

The VaR measure, as explained in the previous section, refers to the risk of a financial institution in isolation. However, a single institution's risk measure does not necessarily reflect its connection to overall systemic risk. For these reasons, the name is born adding the prefix "Co.", that stay for *conditional* to the Value-at-Risk. An institution $i$'s CoVaR relative to the system is defined as the VaR of the whole financial sector conditional on institution $i$ being in a particular state.

The formal definition of CoVaR given in [5] is: $CoVaR_q^{j|C(X^i)}$ is defined as the VaR of institution $j$ (or the financial system) conditional to some event $C(X^i)$ of institution $i$. That is, $CoVaR_q^{j|C(X^i)}$ is implicitly defined by the $q$%-quantile of the conditional probability distribution:

$$Pr(X_j|C(X^i) \leq CoVaR_q^{j|C(X^i)}) = q\%. \tag{2.2}$$

In other words, let $X^i$ and $X^j$ be the returns of two financial institutions $i$ and $j$. The CoVaR of bank $j$ at a confidence level $1-q$ can be defined as the $q$%-quantile of the conditional probability distribution:

$$Pr(X_j \leq CoVaR_q^{j|i}|X_i \leq VaR_q^i) = q\%. \tag{2.3}$$

Another risk measure is computed ΔCoVaR, the difference between the CoVaR conditional on the distress of an institution and the CoVaR conditional on the median state of that institution:

$$\Delta CoVaR_q^{j|i} = CoVaR_q^{j|X^i=VaR_q^i} - CoVaR_q^{j|X^i=VaR_{50}^i}. \quad (2.4)$$

Both the CoVaR and the ΔCoVaR satisfies different properties.

- After splitting one large individually systemic institutions into $n$ smaller clones, the CoVaR of the large institution is the same as the CoVaRs of the $n$ clones.
- If among a large number of small financial institutions that are exposed to the same factors, one of these falls into distress, this will not necessarily cause a systemic crisis. However, if the distress is due to a common factor, then the other institutions will also be in distress. The set of institutions is *systemic as a herd*.
- Each institution's ΔCoVaR is endogenous and depends on other institutions' risk-taking.

There exists other type of ΔCoVaR:

- *Exposure-$\Delta CoVaR$*: $\Delta CoVaR^{j|system}$, which we label Exposure-ΔCoVaR reports institution $j$'s increase in Value-at-Risk in the event of a financial crisis. In other words, it is a measure of an individual institution's exposure to system-wide distress and is similar to the stress tests performed by individual institutions and regulators.
- *Network-$\Delta CoVaR$*: If we compute $CoVaR^{j|i}$ between $j$ and $i$ both individual financial institutions, rather than a set of institutions, we talk of Network - $\Delta CoVaR$. In this case, we study the relationships among the whole network of financial institutions.
- *Forward-$\Delta CoVaR$*: This is an index that is more focused on predict *future*, rather than *contemporaneous* ΔCoVaR. It is based on the dependence of ΔCoVaR on lagged characteristics and on specific institution's variables such as size, leverage and maturity mismatch.

As we have derived the CoVaR from the VaR, there exists the counterpart of the ES called *Conditional Expected Shortfall* (CoES): the expected loss in terms of Expected Shortfall conditional to a VaR event. $CoES_q^{j|i}$ may be defined as the expected loss for institution $j$ conditional on its losses exceeding $CoVaR$. $\Delta CoES_q^{j|i}$ analogously is $CoES_q^{j|i} - CoES_{50}^{j|i}$.

## 2.3 Data

We examine the systemic impact of financial distress for five banks, namely BBVA, Bankinter, CaixaBank, Bankia, Santander on the other. We use weekly data for the period from 1st of February 2014 to 15th of April 2020. The closing price of each bank is considered: it is adjusted for splits but not for dividends. Notice that in the first part of 2020, from February to April (and on), the COVID-19 pandemic is spread into the world and obviously it affected the financial institutions. In Fig. 2.2, the stock price time-series of the five banks are shown.

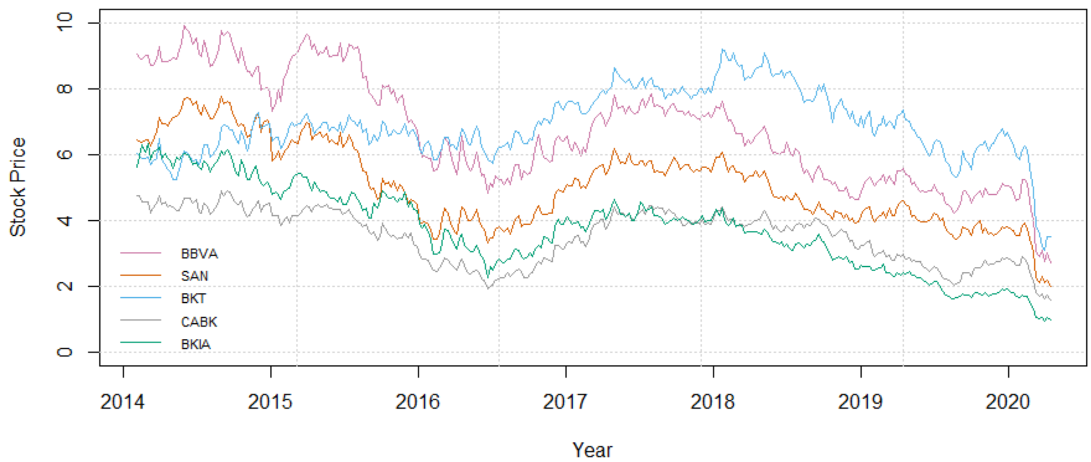


**Figure 2.2:** Stock prices of the five banks considered.

Most financial time series follow random walks, which means that the best estimate of tomorrow's value is today's value. In a time series of asset prices $P_t$, we are not generally interested in the actual prices but in their relative changes. The closing price of each insurance company is converted to the form of logarithmic yield. One period log return from date $t-1$ to date $t$ is:

$$R_t = \log\left[\frac{P_t}{P_{t-1}}\right] \tag{2.5}$$

where $R_t$ is the stock returns for the day $t$, $P_t$ and $P_{t-1}$ are the closing price for the day $t$ and $t-1$ respectively.

Moreover, prices usually have a *unit root*, while returns can be assumed to be stationary. One easy way to visualize the difference among the two is to calculate the *autocorrelation fuction* (ACF) of each series. In Fig. 2.3 are shown the autocorrelation functions of the original BBVA time series and the one of the returns. As we can see from the first plot, the correlation value is always positive and it decreases as the lag increases.

This means that the values of the series are correlated to the lagged values: a trend component is strongly present in the series. On the other hand, the ACF of the returns shows correlation values that vary a lot, oscillating between positive and negative values. Apart from lag 7, all the values are within the band, i.e. the dashed lines. This means that the series is not correlated with the lagged values. The returns of the five banks are represented in Fig. 2.4.

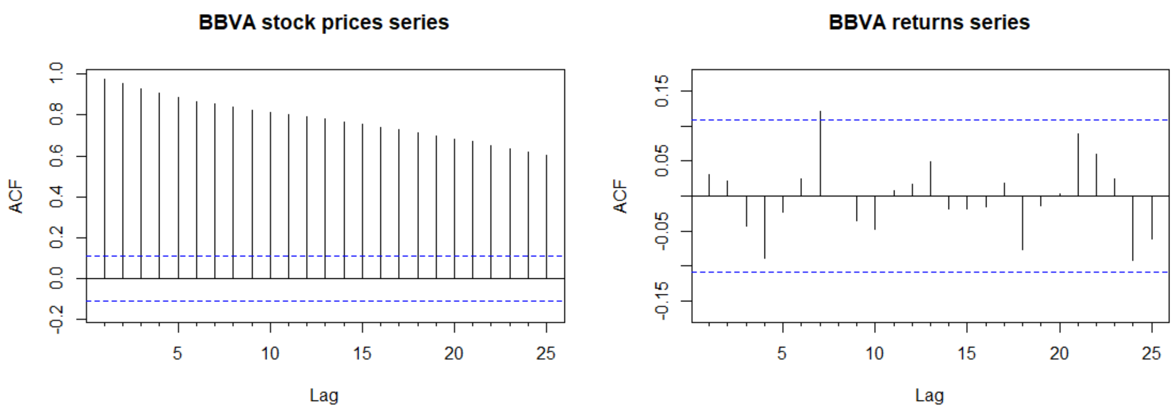


**Figure 2.3:** ACF of BBVA stock prices series and BBVA returns series.

| | BBVA | BANKINTER | BANKIA | CAIXA | SANTANDER |
|---|---|---|---|---|---|
| Mean Value | -0.0034 | -0.0017 | -0.0052 | -0.0031 | -0.0034 |
| Max | 0.1133 | 0.1246 | 0.1346 | 0.1567 | 0.1266 |
| Min | -0.2819 | -0.2447 | -0.2288 | -0.1860 | -0.2965 |
| Std. Dev. | 0.0423 | 0.0377 | 0.0467 | 0.0435 | 0.0446 |
| Skewness | -0.9360 | -0.8809 | -0.4674 | -0.1801 | -1.0393 |
| Kurtosis | 5.7178 | 5.4310 | 2.4304 | 1.1456 | 5.9256 |
| Jacque-Bera | 518.6225 | 436.2074 | 92.1191 | 21.1505 | 556.3621 |

**Table 2.1:** Descriptive statistics of the returns series for the five considered banks.

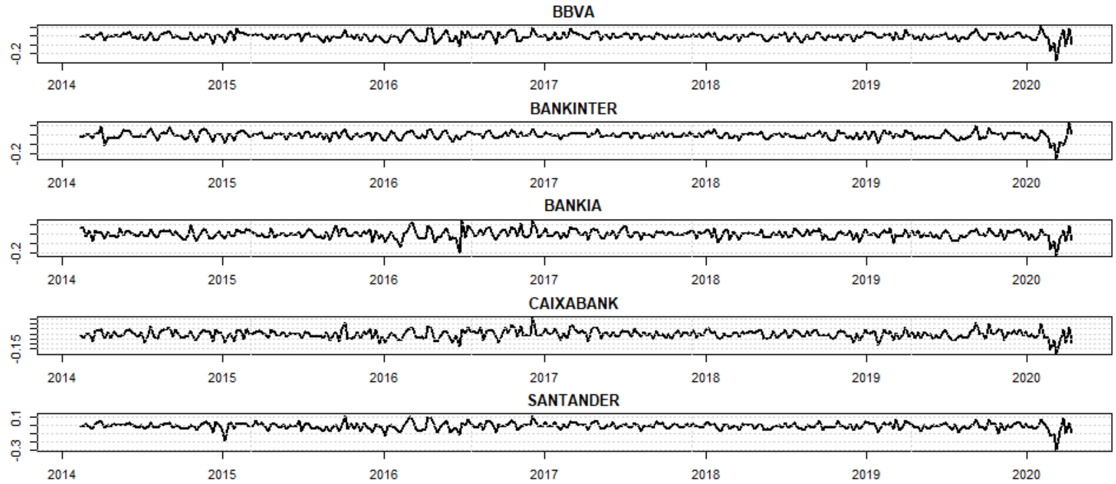


**Figure 2.4:** Returns of the five banks series.

Before fitting the model, the stationary test of the return time series is carried out. The Augmented Dickey-Fuller method in unit root test is used to test the stationarity of the return series. Since the alternative hypothesis is that the series is stationary, obtaining a small p-value means rejecting the null hypothesis. In all the cases we got a p-value smaller than 0.01 and we can conclude that the series are stationary.

### 2.3.1 Modified Cornish-Fisher Value-at-Risk

Some descriptive statistics are computed to analyse in depth the distributions of the returns and, in particular, to test the normality. The results are shown in Table 2.1. It can be seen that the skeweness of the series is not zero and also, the kurtosis, in

the majority of the cases, is greater than three, taking form of "peak kurtosis and tail thickness" [26], common in the cases of financial series. The Jarque-Bera statistics are far from 0 and the p-values are very small: this let us reject the hypothesis of normality of the returns. Since the distribution of the returns is not normal, computing the Value-at-Risk is not direct.
Zangari (1996) [6] and Favre and Galeano (2002) [7] provide a modified VaR calculation that takes the higher moments of non-normal distributions into account through the use of a Cornish Fisher expansion. This measure is now widely cited and used in the literature and is usually referred as *"Modified VaR"* or *"Modified Cornish-Fisher VaR"*.

The Cornish-Fisher expansion provides a simple relation between the skewness and kurtosis parameters and the Value-at-Risk facilitating the optimizations of this value. It is based on the computation of this value:

$$z_{cf} = z_c + \frac{(z_c^2 - 1)S}{6} + \frac{(z_c^3 - 3z_c)K}{24} - \frac{(2z_c^3 - 5z_c)S^2}{36}, \tag{2.6}$$

where $S$ in the skewness of the returns $R$, $K$ is the excess of kurtosis of $R$ and $z_c$ is the critical value at level $1 - \alpha$. So, the Modified Value-at-Risk is:

$$VaR = -\bar{R} - \sigma * z_{cf}, \tag{2.7}$$

where $\bar{R}$ is the mean and $\sigma$ is the standard deviation of returns. Notice that when the returns are normally distributed, $S$ and $K$ are 0 and Cornish-Fisher VaR collapses to traditional mean-VaR.

In this work, we have computed the Modified VaR using rolling windows of width 12 weeks (a quarter of year). The distributions of the Cornish-Fisher VaR for the five banks are shown in Fig. 2.5.

We notice that the risk seems to be stationary along all the series until 2019, except for some slightly decrease in the first period of 2015, especially for the Santander bank and in the middle of 2016 when Bankia is the most affected one.

At the beginning of 2020, we notice a very strong decrease concurrently with the spread of the COVID-19 in the world.

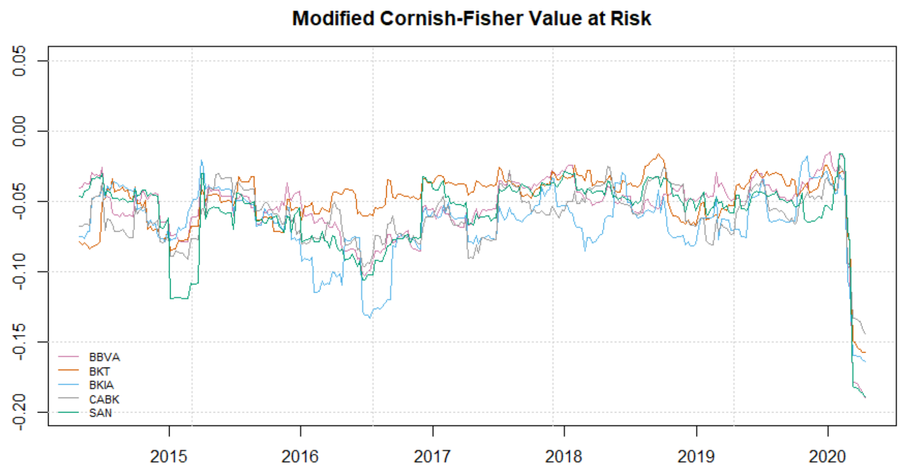


**Figure 2.5:** Modified Cornish-Fisher Value-at-Risk of the five considered banks.

# Chapter 3

# Technical instruments

## 3.1 The regression problem

In the machine learning literature and in the statistical learning theory, we tend to refer to problems with a quantitative response as regression problems, while those involving a qualitative response are often referred to as classification problems. Regression methods found their origins in the linear regression, when at the beginning of the nineteenth century, Legendre and Gauss published papers on the method of least squares, which implemented the earliest form of linear regression. Linear regression is used for predicting quantitative values and the approach was first successfully applied to problems in astronomy. It is well known that linear regression is still a useful and widely used tool, although more advanced models have been proposed in the literature.

In this thesis we will consider some of the most recent development of this technique and how they can be applied to estimate financial system risks.

## 3.2 Quantile Regression

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The objective of standard univariate regression is the estimation of the conditional expectation $\hat{y} = \mathbb{E}(Y|X = x)$, where $Y$ is a random variable indicating the response given a realization $x$ of the independent random variable $X$. The value $\hat{y}$ is obtained through the estimation of a vector of unknown parameters $\beta$ that can be computed in multiple way, e.g. using Ordinary Least Squares estimation or Maximum Likelihood.

In the literature multiples methods have been proposed to modify this model with the purpose to improve it. Those are generally referred to different forms of regression, adopting different procedures of estimation. For example, estimating the conditional expectation across a broader collection of non-linear models yields to the so called non-parametric regression. Another approach can be retrieved by noticing that only the mean of the response variable is estimated using standard regression, neglecting all the other aspect that could guarantee a better quantification of $Y$.

Motivated by finding a solution to cope with this problem, Koenker and Bassett [1] proposed the Quantile Regression method (QR) in January 1978. In this framework, quantiles are adopted to describe the distribution of the dependent variable, in this way a Quantile Regression models the relationship between $X$ and the conditional quantiles of $Y$ rather than just the conditional mean of $Y$, yielding to a more comprehensive picture of the effect of the independent variable on the dependent variable.

The conditional distribution function is given by the probability that, for $X = x$, $Y$ is smaller than $y \in \mathbb{R}$,

$$F(y|X = x) = P(Y \leq y|X = x). \tag{3.1}$$

For a continuous distribution function, the $\alpha$-quantile $Q_\alpha(x)$ is then defined such that the probability of $Y$ being smaller than $Q_\alpha(x)$ is, for a given $X = x$, exactly equal to $\alpha$, i.e.

$$P(Y \leq Q_\alpha(x)|X = x) = \alpha. \tag{3.2}$$

In general we choose $Q_\alpha(x)$ as:

$$Q_\alpha(x) = \inf \{y : F(y|X = x) \geq \alpha\}. \tag{3.3}$$

As described in Equation 3.1, we start by considering a random variable $Y$ having distribution function $F$, from which a set of ordered observations has been collected along the time interval horizon $t \in 1, ..., T$.
We also denote by $x_t$ a sequence of $K$-vectors of a multivariate regression problem. The $\alpha$-th quantile is defined as any of the solution of the minimization problem:

$$\min_{\beta \in \mathbb{R}} \left[ \sum_{t \in \{t: y_t \geq x_t \beta\}} \alpha \left| y_t - x_t \beta \right| + \sum_{t \in \{t: y_t < x_t \beta\}} (1 - \alpha) \left| y_t - x_t \beta \right| \right]. \tag{3.4}$$

If the mean and the median coincide and $\alpha = 1/2$, this is the case of standard linear regression, in fact the QR can be considered a natural generalization of the linear model. This non-differentiable function is minimized via numerical methods rather than maximum likelihood, such as using the simplex method which is guaranteed to yield a solution in a finite number of iterations.

QR is more robust to non-normal errors, outliers and is invariant to monotonic transformations, such as the logarithm, of the dependent variable.

## 3.3 Random Forests

Random Forests is a supervised machine learning method whose main element is the decision tree. It is made by a series of yes/no questions about data leading to a class or to continuous values in the case of regression. In the CART algorithm, a decision tree is built by determining the questions (called splits of nodes $m$) that, when answered, leads to the greatest reduction of a chosen metric, such as the Gini Impurity:

$$G = \sum_{c=1}^{K} \hat{p}_{mc}(1 - \hat{p}_{mc}), \tag{3.5}$$

where $K$ are the classes and $\hat{p}_{mc}$ is the fraction of examples of class $c$ among all examples in node $m$. In Fig. 3.1, there is an example of representation of decision tree. Each root node represents a single input variable $X$ and a split point on that variable. The leaves of the tree contain the output variable $Y$ divided into the obtained classes. The Random Forests is a model made up of many decision trees

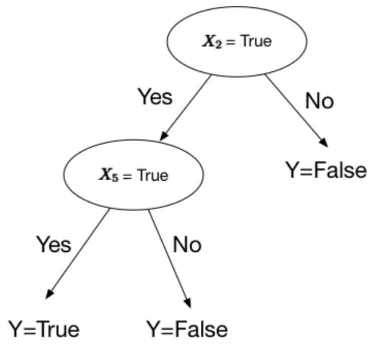


**Figure 3.1:** Example of Decision Tree.

whose prediction by committee is more accurate than that of any individual tree. This model uses two key concepts:

- *Random sampling of training data points when building trees*: During the training phase, each tree learns from a random sample of the data points. The samples are drawn with replacement, known as *bootstrapping*. At test time, predictions are made by averaging the predictions of each decision tree.
- *Random subsets of features considered when splitting nodes*: Only a subset of all the features are considered for spitting each node.

Recalling the notation in Breiman [2], we call $\theta$ the random parameter vector that determines how a tree is grow and the corresponding tree is denoted by $T(\theta)$. Considering $\Gamma$ the space in which $X$ lives, every leaf $l$ of a tree corresponds to a rectangular subspace of $\Gamma$. For every $x$ in $\Gamma$ there is one and only leaf $l$ such that $x \in R_l$ (the leaf obtained when dropping $x$ down the tree). It is called $l(x, \theta)$ for tree $T(\theta)$. The prediction of a single tree $T(\theta)$ for a new data point $X = x$ is obtained by averaging over the observed values in leaf $l(x, \theta)$.

Let the weight vector $w_i(x, \theta)$ be given by a positive constant if observation $X_i$ is part of leaf $l(x, \theta)$ and 0 if it is not. The weights sum to one, and thus:

$$w_i(x, \theta) = \frac{1_{\left\{X_i \in R_{l(x,\theta)}\right\}}}{\#\left\{j : X_j \in R_{l(x,\theta)}\right\}}. \tag{3.6}$$

The prediction of a *single tree*, given covariate $X = x$, is the weighted average of the original observations $Y_i$

$$\hat{\mu} = \sum_{i=1}^{n} w_i(x, \theta) Y_i. \tag{3.7}$$

Considering the average prediction of $k$ single trees, each constructed with an i.i.d vector $\theta_t$, with $t = 1, ..., k$, we obtain $w_i(x)$:

$$w_i(x) = k^{-1} \sum_{i=1}^{k} w_i(x, \theta_t). \tag{3.8}$$

This led to the Random Forests prediction of the conditional mean $\mathbb{E}(Y|X = x)$:

$$\hat{\mu} = \sum_{i=1}^{n} w_i(x) Y_i. \tag{3.9}$$

## 3.4 Quantile Regression Forests

In section 3.3 we have discussed the Random Forests method that is used to estimate the conditional mean $\mathbb{E}(Y|X = x)$. In 2006, Meinshausen [3] proposed a method that is the very close to the Quantile Regression (Section 3.2) but in the field of Random Forests: the Quantile Regression Forests. We want to obtain an approximation of the full conditional distribution and, in particular, estimate quantiles of variable $Y$ given $X = x$, $Q_\alpha(x)$. According to [3], the model gives a non-parametric way of estimating conditional quantiles for high-dimensional predictor variables and also, the algorithm is competitive in terms of predictive power. In order to estimate the quantiles of the variable $Y$, the method starts from the estimation of the conditional distribution function of $Y$ given $X = x$ that is:

$$F(y|X = x) = P(Y \leq y|X = x) = E(1_{\{Y \leq y\}}|X = x). \tag{3.10}$$

In analogy with Random Forests, we define an approximation of $\mathbb{E}(1_{\{Y \leq y\}}|X = x)$ , the weighted mean over the observations of $1_{\{Y \leq y\}}$:

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) 1_{\{Y_i \leq y\}}. \tag{3.11}$$

The weights $w_i(x)$ are the same of the Random Forests, Equation 3.8.

Meinhausen summarizes the algorithm for computing $\hat{F}(y|X = x)$ in the following procedures:

1. Grow $k$ trees $T(\theta_t)$ , $t = 1, ..., k$, as in Random Forests. However, for every leaf of every tree, take note of all observations in this leaf, not just their average.

2. For a given $X = x$, drop $x$ down all trees. Compute the weight $w_i(x, \theta_t)$ of observation $i \in \{1, ..., n\}$ for every tree as in 3.6. Compute weight $w_i(x)$ for every observation $i \in \{1, ..., n\}$ as an average over $w_i(x, \theta_t)$, $t = 1, ..., k$, as in 3.8.

3. Compute the estimate of the distribution function as in 3.11 for all $y \in$ R, using the weights from Step 2.

An estimate of $Q_\alpha(x)$ is obtained by plugging $\hat{F}(y|X = x)$ in Equation 3.3.

The main advantage of this method with respect to the classical Quantile Regression is that it is able to better estimate the extreme quantiles i.e. $\alpha$ very close to 0 or 1. This aspect is useful in our case since the metrics we are going to study are generally evaluated at a level of $\alpha = 0.01$ or $\alpha = 0.05$.

# Chapter 4

# CoVaR estimation

## 4.1 Why common approaches do not work

As seen in Chapter 2, the real financial data are usually not normally distributed, but take the form of "peak kurtosis and tail thickness" [26]. The classical linear regression method is based on the mean value estimation, so it is not useful in the estimation of this kind of models. Based on the evaluations of the quantiles of a variable, all the Quantile Regression models effectively compensates the drawbacks of the traditional linear regression and is widely used in the measurement of financial risk.

In this chapter we will use these methods to estimate the CoVaR between two generic financial institutions $i$ and $j$. In particular, starting from the simple Quantile Regression, already implemented by Adrian and Brunnemier in [5], we will extend it to the Quantile Regression Forests method and then comparing the results of these estimation models.

## 4.2 Quantile Regression

Define $X^j$ and $X^i$ as the returns of two financial institutions, $j$ and $i$ respectively. Using the Quantile Regression method we obtain the following equation:

$$\hat{X}_q^{j|i} = \hat{\alpha}_q^i + \hat{\beta}_q^i X^i, \tag{4.1}$$

where $\hat{X}_q^{j|i}$ denotes the predicted value for a $q\%$-quantile of the returns of institution $j$ conditional on returns $X^i$ of institution $i$. The coefficient $\hat{\beta}_q^i$ estimates the change in a specified quantile $q$ of $X^j$ produced by a one unit change in $X^i$.

From the definition of Value-at-Risk, the predicted value gives the VaR of the financial institution $j$ conditional on the returns $X^i$.
Using the of $VaR_q^i$ instead of $X^i$, we obtain $CoVaR_q^{j|i}$:

$$CoVaR_q^{j|i} = VaR_q^{j|X^i=VaR_q^i} = \hat{\alpha}_q^i + \hat{\beta}_q^i VaR_q^i. \tag{4.2}$$

In summary, we have computed $VaR_q^{j|X^i} = F_{X^j}^{-1}(q|X^i)$ and then, setting $X^i = VaR_q^i$, we have obtained $CoVaR_q^{j|i} = F_{X^j}^{-1}(q|VaR_q^i)$.

### 4.2.1 Analysis of the parameters estimates

Fig. 4.1 describes how the regression coefficients change with quantiles, regressing BBVA's returns on Bankinter's returns. The same reasoning can be done for the other regressions. In the above plot, the dashed line shows the behaviour of the intercept for all the studied quantiles. While, the plot below regards the $\beta$ coefficient. The grey band around the dashed line represents the 90% confidence interval.

## 4.3 Quantile Regression Forests

Starting from the reasoning in Section 4.2, we have estimated the CoVaR using the Quantile Regression Forests method.

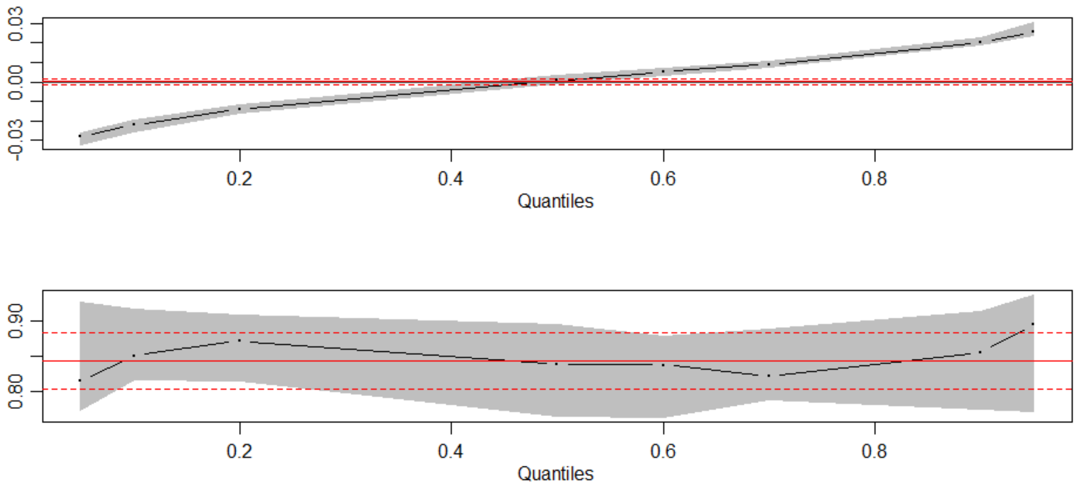


**Figure 4.1:** $\alpha$ (first plot) and $\beta$ (second plot) parameters estimates according to different level of quantiles in the Quantile Regression of BBVA given Bankinter.

As proposed in [3], as a first step QRF estimates the conditional distribution function of $Y = X^j$ given $X = X^i$, where $X^j$ and $X^i$ are defined in Section 4.2. In particular, it is approximated by a weighted mean over the observations of $1_{\{Y_i \leq y\}}$:

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) 1_{\{Y_i \leq y\}}. \tag{4.3}$$

Once the conditional distribution function is estimated, it is plugged in the quantile estimation Equation 3.3 and so, QRF solves the equation:

$$Q_\alpha(x) = \hat{F}^{-1}(\alpha) = \inf \left\{ y : \sum_{i=1}^{n} w_i(x) 1_{\{Y_i \leq y\}} \geq \alpha \right\}. \tag{4.4}$$

Basically, we have to compute the functions: $VaR_q^{j|X^i} = F_{X^j}^{-1}(q|X^i)$ and $CoVaR_q^{j|i} = F_{X^j}^{-1}(q|VaR_q^i)$.

An important problem that can arise in this procedure, is that we cannot use directly the returns values $X^i$ and $X^j$ if they are not in the same range of $VaR_q^i$. This is due to the fact that the weights obtained in the first QRF have to be comparable with the values of $VaR_q^i$ in the computation of $CoVaR_q^{j|i}$ of the second QRF. In order to do this, we have decided to standardize both the returns and the VaR values. In this way, the data have the same mean and variance and they are comparable.

### 4.3.1 Setting of the parameters in Quantile Regression Forests

Regarding the Quantile Regression Forests, some parameters have been set to fit the models. One of the most important is the number of trees used to train the forests. Having about 300 observations, considering a large number of trees could have been useless. In fact, as the reader can see in Fig. 4.2, in which we plot the error against the number of trees, it stabilizes around 50 trees. Another parameter

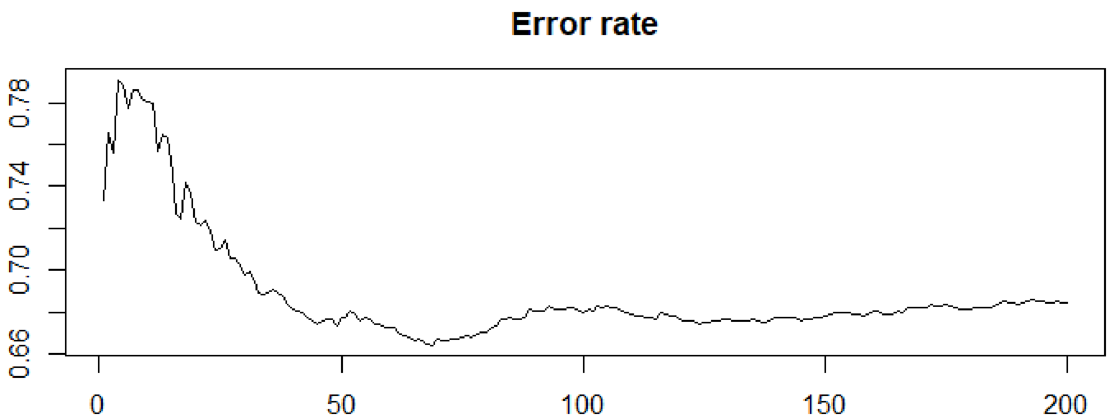


**Figure 4.2:** Error Rate evaluation for the number of considered trees.

we had to set is the vector of quantiles on which we have to calibrate the forests. We have used $q = (0.05, 0.5, 0.95)$ in order to catch all the aspect of the distribution. The estimation of the CoVaR is done using $q = 0.05$. The number of variables tried for each split is set as the default value in Random Forests for regression, namely $\frac{p}{3}$, where $p$ is the total number of variables. In Fig. 4.3, 4.4, 4.5, there is a comparison among the Value-at-Risk, the CoVaR estimated using the Quantile Regression and the one estimated using the Quantile Regression Forests, for each of the five banks.

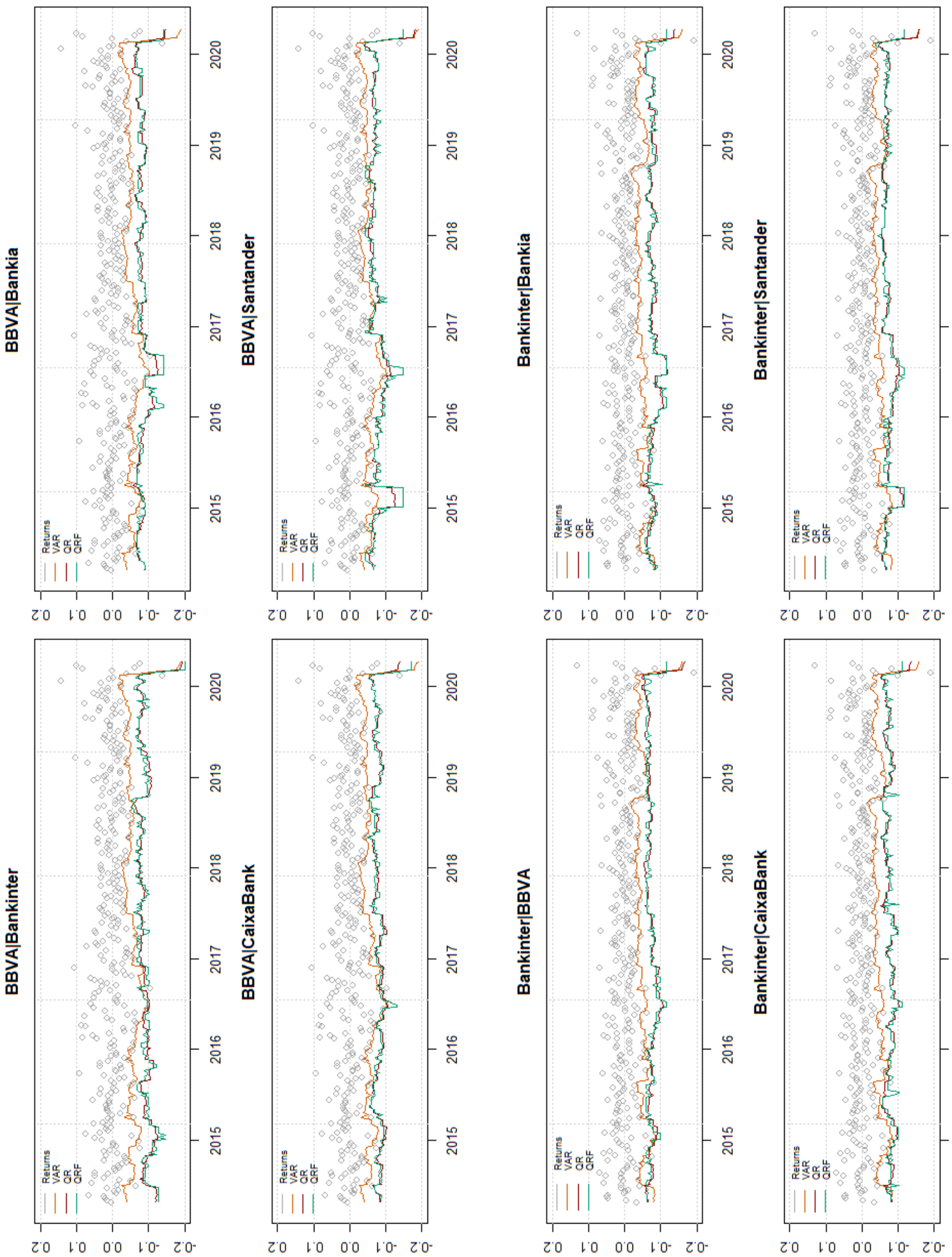


**Figure 4.3:** Effect of each bank at its 5% VaR on BBVA and Bankinter 5% VaR in terms of CoVaR.

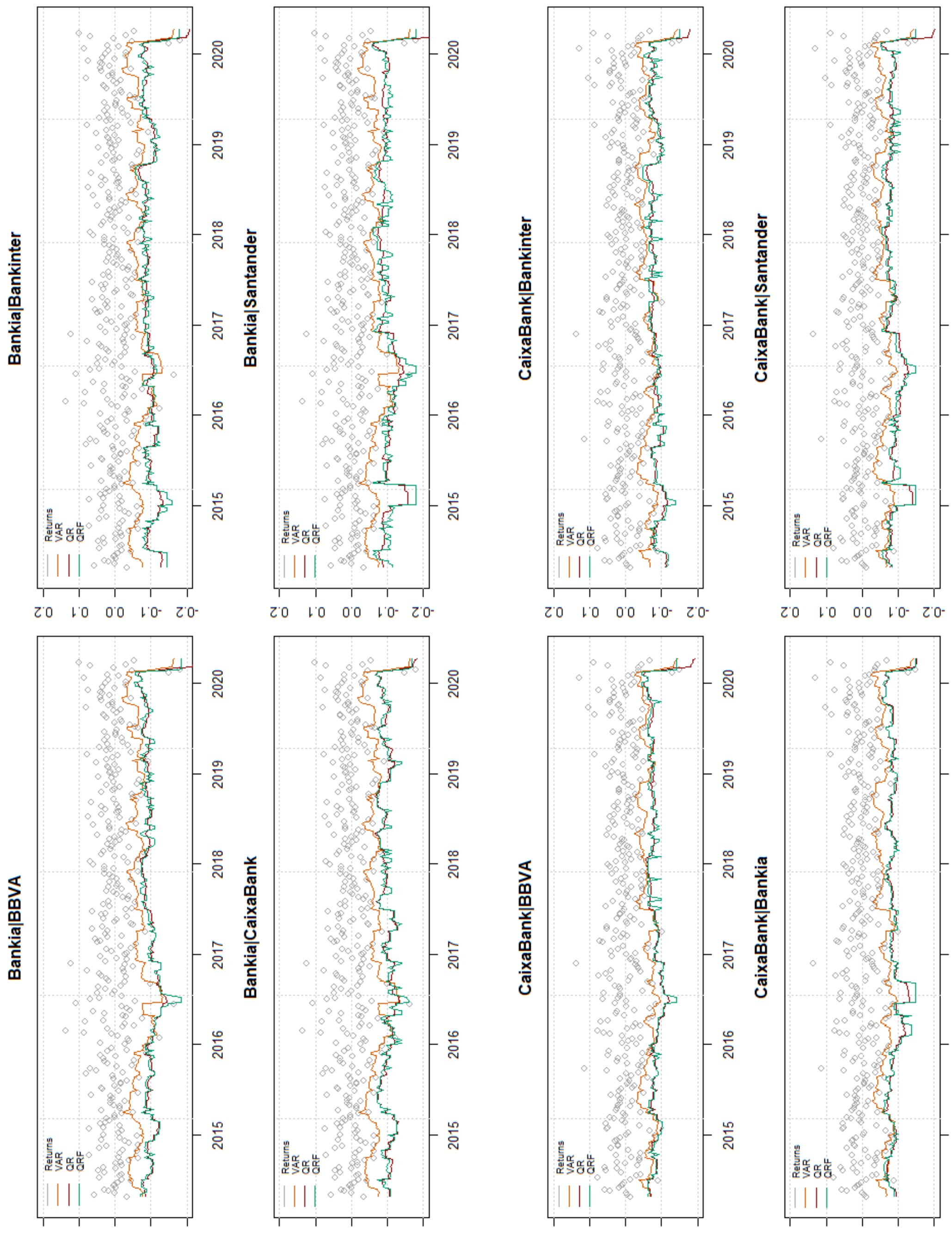


**Figure 4.4:** Effect of each bank at its 5% VaR on Bankia and CaixaBank 5% VaR in terms of CoVaR.

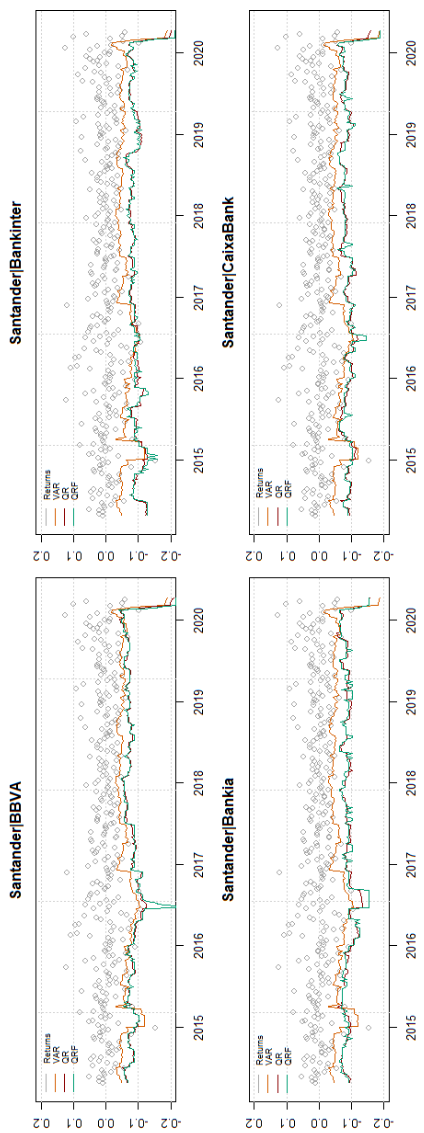


**Figure 4.5:** Effect of each bank at its 5% VaR on Santander 5% VaR in terms of CoVaR.

As we have already seen in Fig. 2.5, the Value-at-Risk series computed using different time windows are stationary in the interval $[-0.10, -0.05]$, except for the last part of the series in which there is a significant decrease, reaching also $-0.20$. Beyond this, there are other points in which we notice a less pronounced decrease, for example at the beginning of 2015 or in the middle of 2016.

Regarding the comparison between the two obtained estimates of the CoVaR, we can see that the red line (QR estimate) is more smooth with respect to the green one that shows numerous and small peaks. Moreover, QRF sometimes overestimate the lower part of the distributions. This could be due to the fact that QRF is more influenced by the 2020 values.

Changing point of view, we could compare the effect of each bank on the others or, on the contrary, how much each bank's VaR is modified by the others. If we overlap the CoVaR estimate of each bank with respect to the others, we see that the effects are similar for all of them. All the banks receive almost the same contribution from the others.

Zooming in, we could see that there are some cases in which the CoVaR lines that are more distant from the VaR estimate: for example the effect of Bankia on BBVA seems to be higher than the effect of Santander on BBVA. Here we have done just a graphical analysis: the quantification of these relations will be done in Chapter 7, dealing with networks and using the $\Delta$CoVaR metric.

These are the results obtained studying the relation between the returns of banks at a level $q = 0.05$. In the next chapter there will be a more accurate analysis considering also some macroeconomic lagged variables that could affect the CoVaR computation.

# Chapter 5

# Systemic State Variables for CoVaR estimation

Until now we have studied a methodology for estimating CoVaR that is constant over time. In order to capture the time-variation in the joint distribution of the returns, the next step is to estimate both VaR and CoVaR as a function of state variables that change over time. As suggested in [5], a new vector of lagged state variables $\boldsymbol{M}_{t-1}$ is introduced. These variables influence the mean and volatility of the risk measures.

State variables are added in the model in order to obtain an ever better estimate of the risk measures. It has been studied in [4], that among international variables that have the greatest impact on the Spanish financial system we can mention: US stock market, US interest rates, oil price and two financial stress indices, CFSI and VIX. However, since this work is more focused on the European context, we have decided to choose just the VIX as an element of the US financial market, considering it fundamental on an international scenario. It captures the implied volatility in the stock market.

In addition to it, it has been used the Euribor, a daily reference rate coming from the averaged interest rates at which Eurozone banks offer unsecured funds lending to other banks in the euro money market and the Euro Stoxx 50, a stock index of the Eurozone. In summary the variables that are considered in the present work are:

- VIX
- Euribor 3 months
- Euro Stoxx 50

A smaller number of variables is considered to avoid overfitting.

## 5.1 Estimation

Starting from the description of the CoVaR computation using Quantile Regression in [5], a proof for the CoVaR estimation using Quantile Regression Forests and considering additional time-lagged variables, is given.

### 5.1.1 Theoretical results

It is assumed that the losses $X_t^j$ have the following linear structure:

$$X_{t+1}^j = \phi_0 + \boldsymbol{M}_t\phi_1 + X_{t+1}^i\phi_2 + (\phi_3 + \boldsymbol{M}_t\phi_4)\Delta Z_{t+1}^j, \tag{5.1}$$

where $\boldsymbol{M}_t$ is a vector of state variables. The error term $\Delta Z_{t+1}^j$ is assumed to be *i.i.d.* with zero mean and unit variance.
The conditional expected return is:

$$\mu^j \left[X_{t+1}^j | \boldsymbol{M}_t, X_{t+1}^i\right] = \phi_0 + \boldsymbol{M}_t\phi_1 + X_{t+1}^i\phi_2 \tag{5.2}$$

and the conditional volatility:

$$\sigma^{jj} \left[X_{t+1}^j | \boldsymbol{M}_t, X_{t+1}^i\right] = \phi_3 + \boldsymbol{M}_t\phi_4. \tag{5.3}$$

The coefficients $\phi_0$, $\phi_1$, $\phi_2$ could be estimated via OLS of $X_{t+1}^j$ on $\boldsymbol{M}_t$ and $X_{t+1}^i$. The predicted value of such an OLS regression would be the mean of $X_{t+1}^j$ conditional on $\boldsymbol{M}_t$ and $X_{t+1}^i$.

The Quantile Regression uses conditional quantiles instead of the conditional mean, that one could use to estimate model at Eq. 5.1.

Knowing that $F_X$ denotes the cumulative distribution function of $X$, it follows that:

$$F^{-1}_{X^j_{t+1}}(q|\boldsymbol{M}_t, X^i_{t+1}) = \alpha_q + \boldsymbol{M}_t\gamma_q + X^i_{t+1}\beta_q, \tag{5.4}$$

where: $\alpha_q = \phi_0 + \phi_3 F^{-1}_{\Delta Z^j}(q)$, $\gamma_q = \phi_1 + \phi_4 F^{-1}_{\Delta Z^j}(q)$ and $\beta_q = \phi_2$ .

From the definition of VaR, we obtain:

$$\begin{aligned} VaR^j_{q,t+1} &= \inf\left\{Pr(X^j_{t+1}|\left\{\boldsymbol{M}_t, X^i_{t+1}\right\} \leq VaR^j_{q,t+1}) \geq q\%\right\} & (5.5)\\ &= F^{-1}_{X^j_{t+1}}(q|\boldsymbol{M}_t, X^i_{t+1}), & (5.6)\end{aligned}$$

where:

$$F^{-1}_{X^j_{t+1}}(q|\boldsymbol{M}_t, X^i_{t+1}) \tag{5.7}$$

is called conditional quantile function and it is the $VaR^j_{q,t+1}$ conditional on $\boldsymbol{M}_t$ and $X^i_{t+1}$. In order to obtain the $CoVaR^{j|i}_{q,t}$ from the conditional quantile function, we condition on $VaR^i_{q,t+1}$:

$$\begin{aligned} CoVaR^{j|i}_{q,t+1} &= \inf\left\{Pr(X^j_{t+1}|\left\{\boldsymbol{M}_t, VaR^i_{q,t+1}\right\} \leq VaR^j_{q,t+1}) \geq q\%\right\} & (5.8)\\ &= F^{-1}_{X^j_{t+1}}(q|\boldsymbol{M}_t, VaR^i_{q,t+1}). & (5.9)\end{aligned}$$

To estimate the conditional quantile function, the Quantile Regression method solves Eq. 3.4. However, the main goal of the present work is to estimate the CoVaR using the Quantile Regression Forests method. As proposed in [3] and seen in Chapter 4, the QRF firstly estimates the conditional distribution function of $Y$ given $X$, $\hat{F}(y|X = x)$, and then it is plugged in the quantile equation, Eq. 3.3. Substituting $Y$ and $X$ with the variables of our interest, we have to compute the functions:

$$VaR^j_{q,t+1} = F^{-1}_{X^j_{t+1}}(q|M_t, X^i_{t+1})$$

and then

$$CoVaR^{j|i}_{q,t+1} = F^{-1}_{X^j_{t+1}}(q|M_t, VaR^i_{q,t+1}).$$

### 5.1.2 Practical analysis

As a starting point we have to compute the VaR conditioned only to the vector of state variables $\boldsymbol{M}_{t-1}$, estimating firstly the parameters of our interest in this way:

$$X_t^i = \alpha_q^i + \gamma_q^i \boldsymbol{M}_{t-1} + \epsilon_{q,t}^i, \tag{5.10}$$

and use it to predict the $VaR_{q,t}^i = F_{X_t^i}^{-1}(q|\boldsymbol{M}_{t-1})$:

$$VaR_{q,t}^i = \hat{\alpha}_q^i + \hat{\gamma}_q^i \boldsymbol{M}_{t-1}. \tag{5.11}$$

Then, we compute $F_{X_t^j}^{-1}(q|\boldsymbol{M}_{t-1}, X_t^i)$, that is the $VaR_{q,t}^j$ conditioned to $\boldsymbol{M}_{t-1}$ and $X_t^i$:

$$X_t^{j|i} = \alpha_q^{j|i} + \gamma_q^{j|i} \boldsymbol{M}_{t-1} + \beta_q^{j|i} X_t^i + \epsilon_{q,t}^i. \tag{5.12}$$

Using the estimated parameters from this regression but conditioning to $VaR_{q,t}^i$, computed in Equation 5.11, we get the $CoVaR_{q,t}^{j|i}$:

$$CoVaR_{q,t}^{j|i} = \hat{\alpha}_q^{j|i} + \hat{\gamma}_q^{j|i} \boldsymbol{M}_{t-1} + \hat{\beta}_q^{j|i} VaR_{q,t}^i. \tag{5.13}$$

So, the steps for estimating CoVaR through Quantile Regression Forests are:

- Compute $VaR_{q,t}^i = F_{X_t^i}^{-1}(q|\boldsymbol{M}_{t-1})$ using a QRF in which the response variable is $X_t^i$ and the explanatory variables $\boldsymbol{M}_{t-1}$;
- Another QRF is computed in which $Y = X_t^{j|i}$ and $X = (\boldsymbol{M}_{t-1}, X_t^i)$, to estimate the relation between the returns of the two banks and the vector of macroeconomic variables;
- The model computed at the previous point with its relative weights is used to predict the $CoVaR_{q,t}^{j|i}$, using as explanatory variables $X = (\boldsymbol{M}_{t-1}, VaR_{q,t}^i)$.

We have compared the results obtained using the Quantile Regression with those obtained with the Quantile Regression Forests are shown in Fig. 5.1, 5.2, 5.3.

Different aspects stand out: the CoVaR estimates as a function of lagged variables is more variable and follow better the VaR lines rather than the method of rolling windows used in Chapter 4. We could expect this because the previous way of computing the VaR could be considered as the juxtaposition of a sequence of VaRs computed for different windows without keeping track of any reference at time $t$.

Moreover, we notice that the QRF's CoVaR estimates show very deep peaks in correspondence of decrease of the VaR. This effect could be seen for example when studying the effect of Santander on BBVA: in 2016, we can see a sequence of very pronounced peaks that can also reach $-0.2$. Looking at the study of the Bankia's VaR when Santander is in crisis, we notice that the CoVaR estimates using QR seems like a smoothed version of the QRF's one.

The behaviour of QRF estimates can be due to the fact that we are doing overfitting of the data that could be avoid considering a smaller number of trees or also, because it is more influenced by the explanatory variables.

This part of the work could be integrated using other macroeconomics variables that affect the behavior of the studied banks.

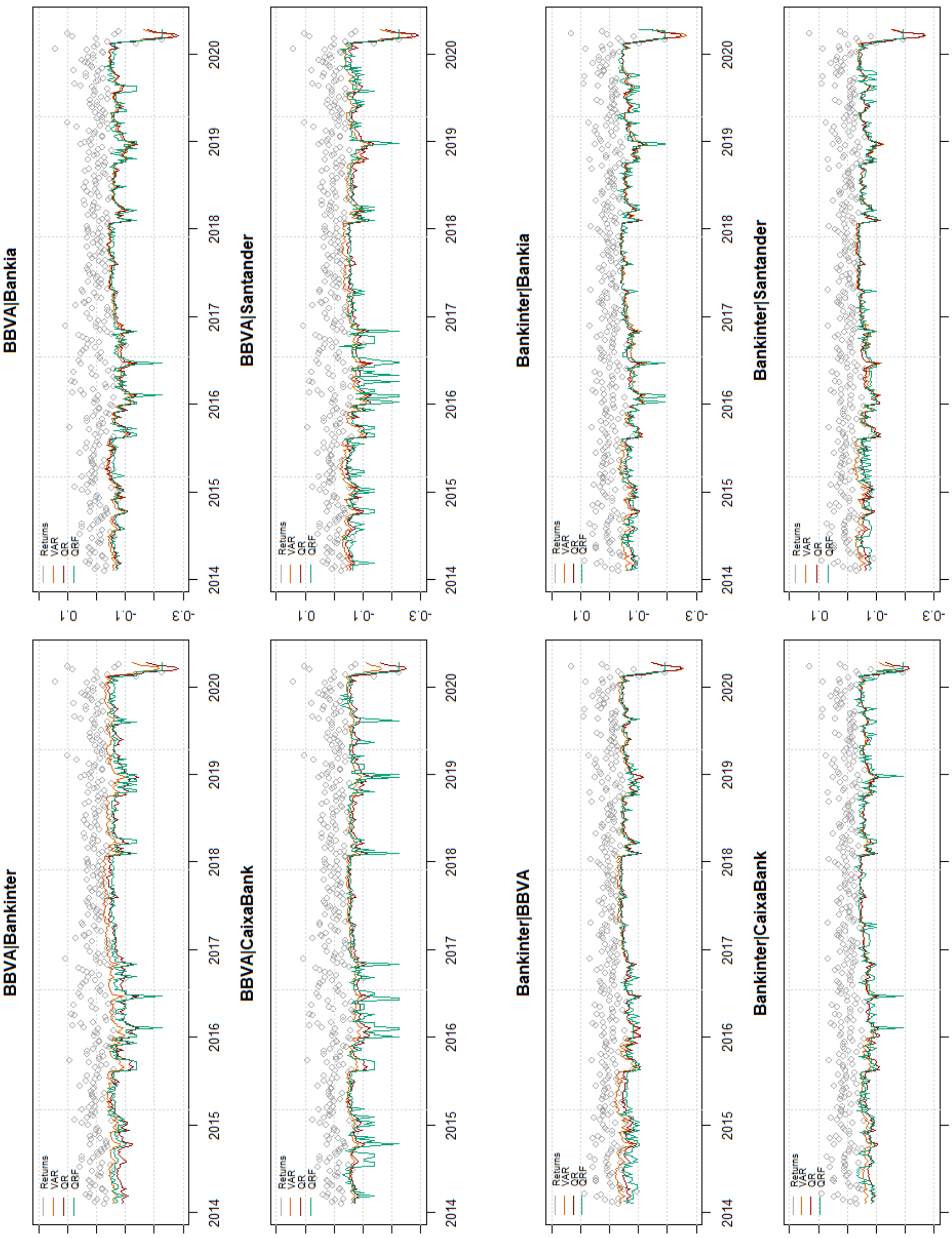


**Figure 5.1:** Effect of each bank at its 5% VaR and additional systemic variables on BBVA and Bankinter 5% VaR in terms of CoVaR.

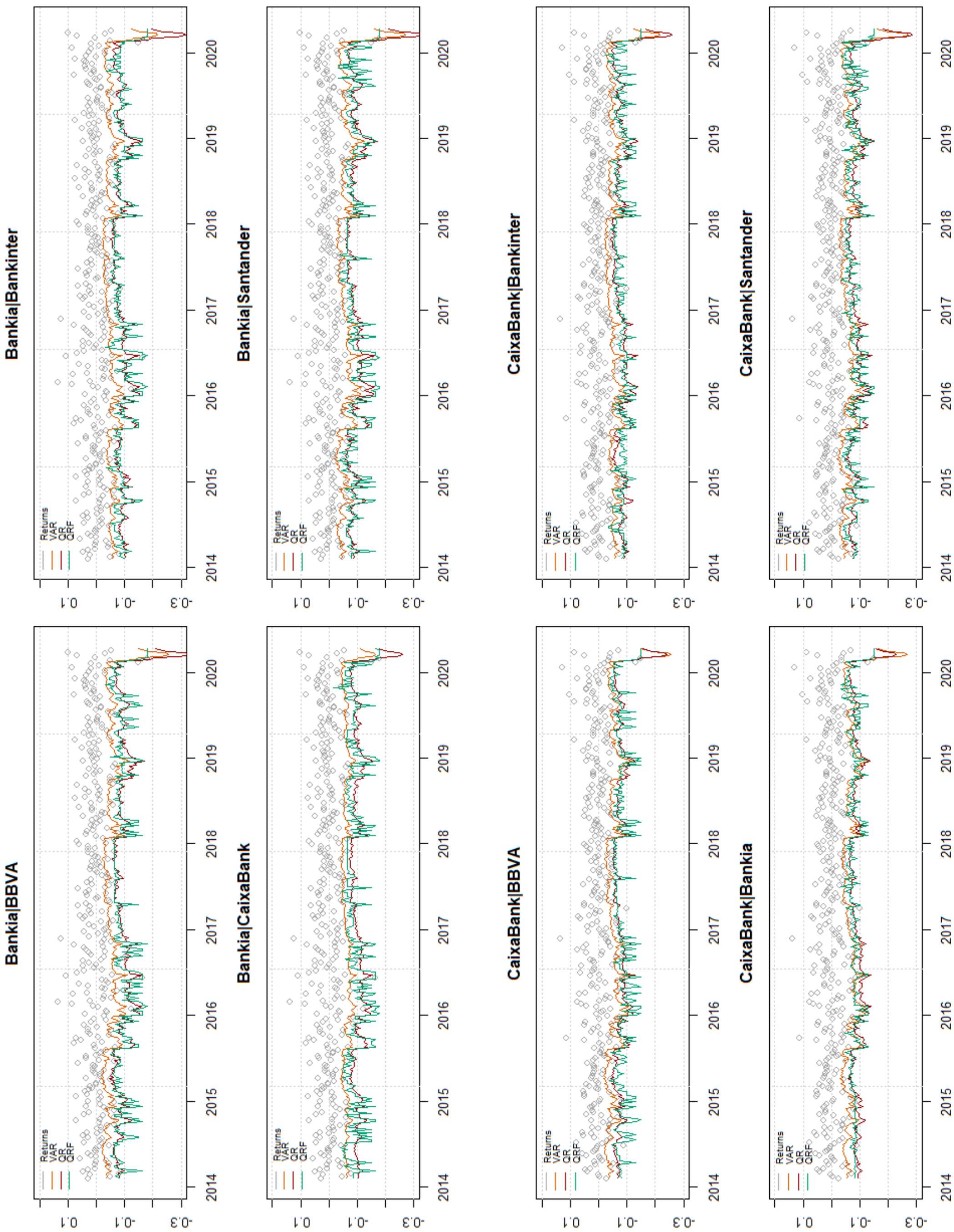


**Figure 5.2:** Effect of each bank at its 5% VaR and additional systemic variables on Bankia and CaixaBank 5% VaR in terms of CoVaR.

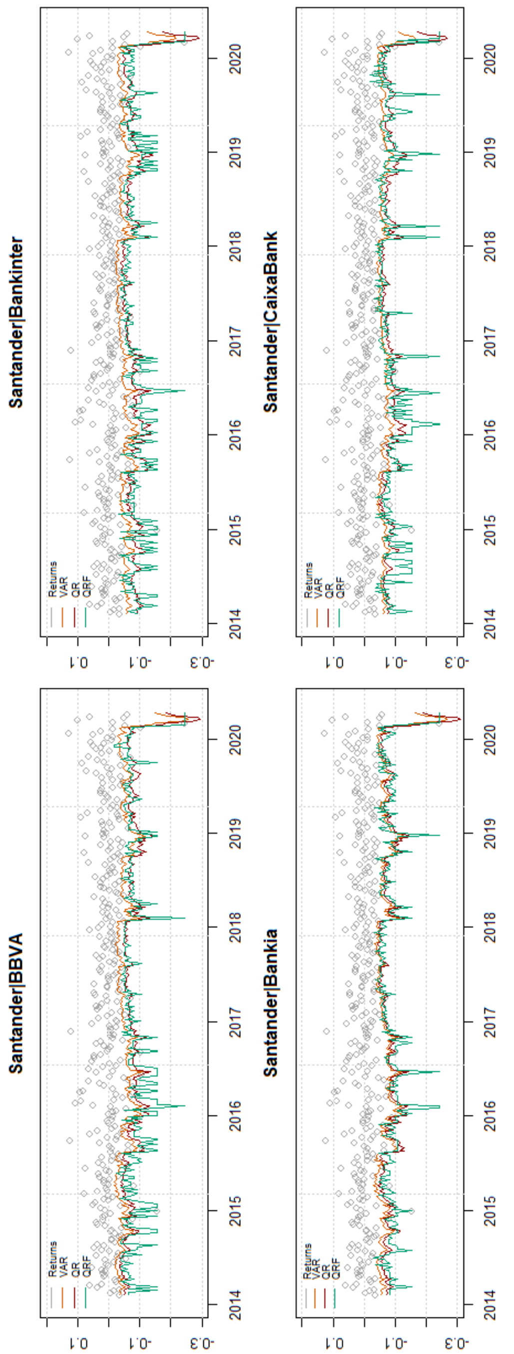


**Figure 5.3:** Effect of each bank at its 5% VaR and additional systemic variables on Santander 5% VaR in terms of CoVaR.

# Chapter 6

# The role of the system in the estimation of CoVaR

In the original paper [5], Adrian and Brunnermeier focus on the CoVaR as a measure to visualize the behaviour of the VaR of the financial institution when a financial system is under distress. This is useful for detecting which institutions are most at risk if a financial crisis occur. After we have studied the relationship among the single financial institutions, we are interested in studying the effect of the system on the studied banks and vice versa. In fact, CoVaR is directional: reversing the conditioning shifts the focus to the question of how the whole system's risk changes given that a particular financial institution is in distress. In the present work, as a *system*, we have considered the Morgan Stanley Capital International (MSCI) financial index for Europe, whose values are shown in Fig. 6.1.

The MSCI Europe Index captures *Large and Mid Cap* [1] representation for 14 Developed Markets (DM) countries in Europe. The index covers approximately 85% of the market capitalization across the European Developed Markets excluding the UK. In Fig. 6.1, we can see that the index shows a similar behaviour with respect to the banks only in the first part, until the decrease in 2016. Then, this series has an increasing trend until the beginning of 2020, while the banks decrease slowly.

Banks and MSCI Index share the fall in 2020 due to the spread of COVID-19,

[1] Large and Middle Size capitalization.

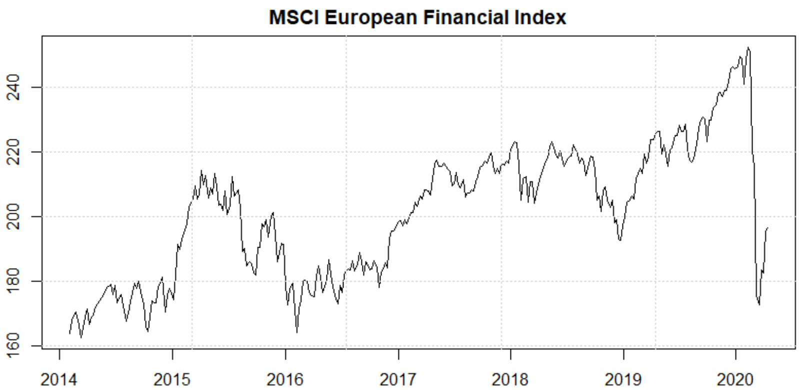


**Figure 6.1:** MSCI European Financial Index time series.

definitely more pronounced in the MSCI Europe Index case. In order to obtain the VaR, as explained in Chapter 2, we have to compute the returns and check the normality of this series. Unfortunately, after having performed the Jarque-Bera test, we cannot consider this series normal and this let us to compute the Cornish - Fisher VaR using different time windows. In Fig. 6.2 there are five plots that show the effect of the system at its 5% VaR level on each bank computed with both the Quantile Regression and with Quantile Regression Forests.

The reasoning behind these computations has been already explained in Chapter 4. In this case we have not considered the effect of additional variables. The first aspect we notice is that, in different cases, the CoVaR line is very close to the VaR, so we can say that the System has not a so strong effect on each bank. In the cases of Bankia and Santander banks, the CoVaR seems to be more distant from the VaR and shows an higher variability.

This means that if there would be a financial crisis in the whole system, Bankia and Santander could be more affected rather than others. Regarding the comparison between the Quantile Regression and Quantile Regression Forests estimations, we confirm the results obtained in Chapter 4 and 5. The Quantile Regression Forests method is more variable and the CoVaR estimates have deeper peaks with respect to the other method.

On the other hand, if we reverse the protagonists of the comparison, measuring the effect of each bank with a VaR at 5 % on the system we have the results in Fig.

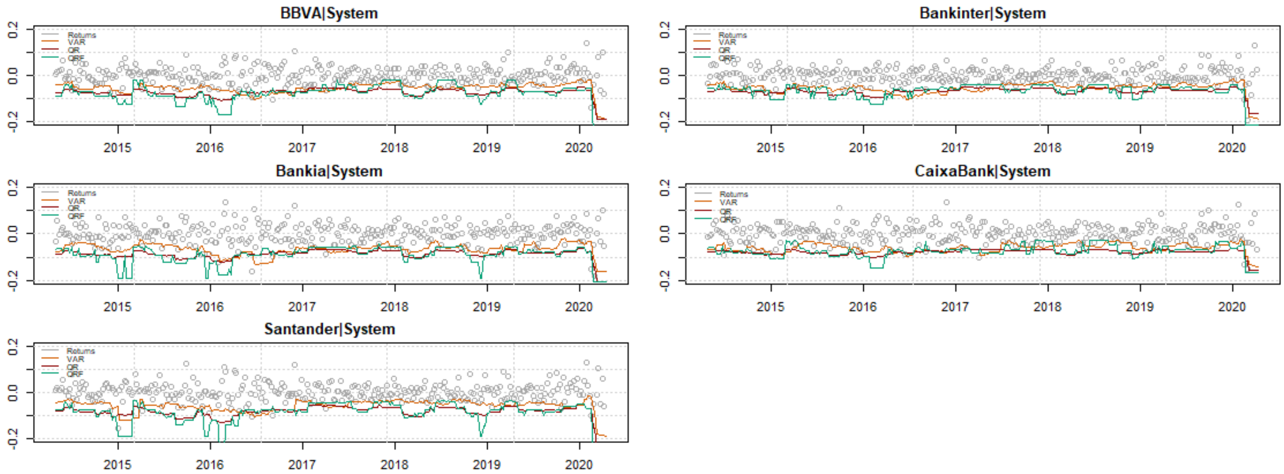


**Figure 6.2:** Effect of the system at its 5% VaR on each bank 5% VaR in terms of CoVaR.

6.3. In this case the situation is almost unchanged. The VaR and CoVaR estimates diverge a bit but, in general, they are similar and we can see a smaller variability. The effect of a bank on the system can be considered negligible and, as it is obvious to think, the contribution of a single bank to the system is not so crucial. In the

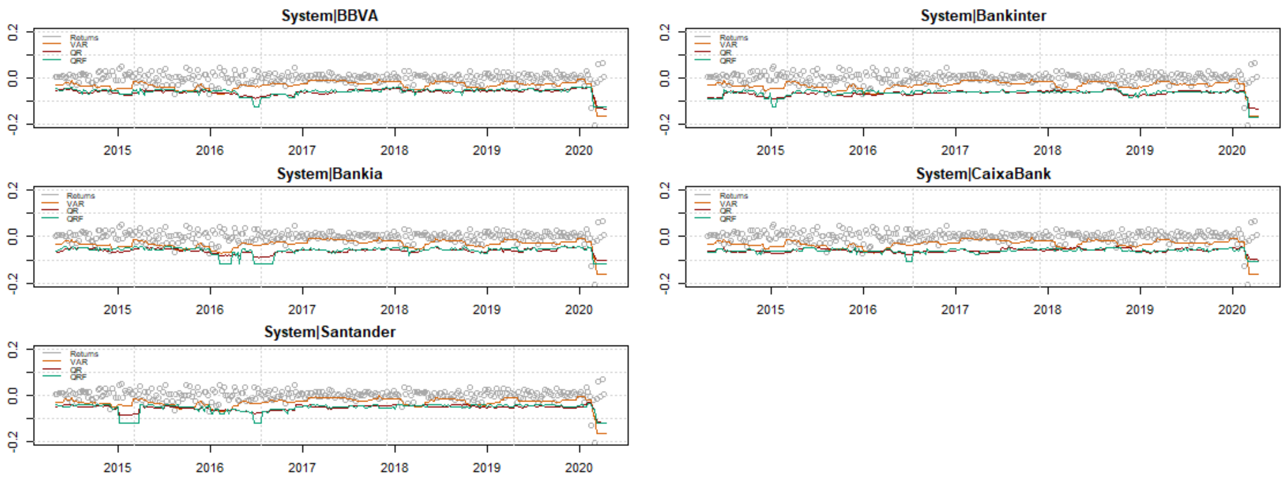


**Figure 6.3:** Effect of each bank at its 5% VaR on the system 5% VaR in terms of CoVaR.

following chapter we are going to analyse in details, which bank would be the most at risk if a financial crisis would occur.

# Chapter 7

# Network analysis

This chapter will focus on the study of the relations among the five banks and the system through the use of network analysis. In particular, contrary to the first part of the thesis, the main metric we are going to analyse is the ΔCoVaR. We recall that the ΔCoVaR is computed in this way:

$$\Delta \mathrm{CoVaR}_q^{j|i} = \mathrm{CoVaR}_q^{j|X^i=VaR_q^i} - \mathrm{CoVaR}_q^{j|X^i=VaR_{50}^i} \tag{7.1}$$

and it denotes the part of $j$ 's systemic risk that can be attributed to $i$. In particular, the measure quantifies how much $i$ adds to the $q$%-VaR of $j$, when $i$ moves from its median state to its $q$%-VaR level.

In this chapter, the VaR has not been computed using rolling windows but it consists of a single value for all the studied period. The 5%-quantile level for the computation is kept. In this way we have a value that summarizes the behaviour of the financial institution and the system in the whole period and let us to compute a unique value for the CoVaR and ΔCoVaR that we are going to study.

Moreover, a more complex analysis has been performed on a larger number of banks that come from different European nations. This is done in order to look for the presence of particular kind of relations among different countries in terms of ΔCoVaR thanks also to community detection techniques.

## 7.1 Relations among banks

In this section, we go deeper into the relations among the banks. This means that for the first part of the analysis, the system's contribution is not considered.

Fig. 7.1 and Fig. 7.2 show two networks that represent graphically the relations among the banks in terms of $\Delta$CoVaR, respectively for the Quantile Regression and Quantile Regression Forest. $\Delta$CoVaR indicates each bank exposure risk, i.e. how sensitive a bank is to another bank going into distress. For example, looking at Fig. 7.1, the effect of Bankinter on BBVA 5%-VaR is $-0.055$, when the Bankinter moves from its median state to its 5%-VaR level.

From the comparison of the networks, it is easy to see that a part of the values are similar, but there are some cases in which they diverge. For example, the effect of Bankia are higher in the Quantile Regression Forests rather than Quantile Regression. In some cases, such as Santander and BBVA, values even double.

We have already pointed out in Chapter 4, that Quantile Regression Forests, in general, gives higher and more pessimistic estimates with respect to the Quantile Regression. This is more visible in the period of abrupt falls. The discrepancy between some values can be also attributed to the fact that we are considering a single value to summarize all the period, so, in some cases, they could be more affected by the lower values of the first part of 2020. Since 2014 there have been different times in which the series have decreased due to different reasons, for example there is a period in the middle of 2016 in which the values are very low: $\Delta$CoVaR can be affected by this aspect.

The analysis of the $\Delta$CoVaR as a network of interconnected banks is a useful tool to understand the relations among them. Thought the study of the network, we are able to discover the complex system that is hidden behind it. In this case, we are not considering only the presence or absence of edges among the different nodes, but they are weighted. Also, we keep into account the direction of the links: the effect of bank $i$ on bank $j$ and vice versa.

Knowing that $n = 5$ is the number of nodes, each graph can be written in terms of sociomatrix: an $n \times n$ square matrix, with an undefined diagonal, where entry

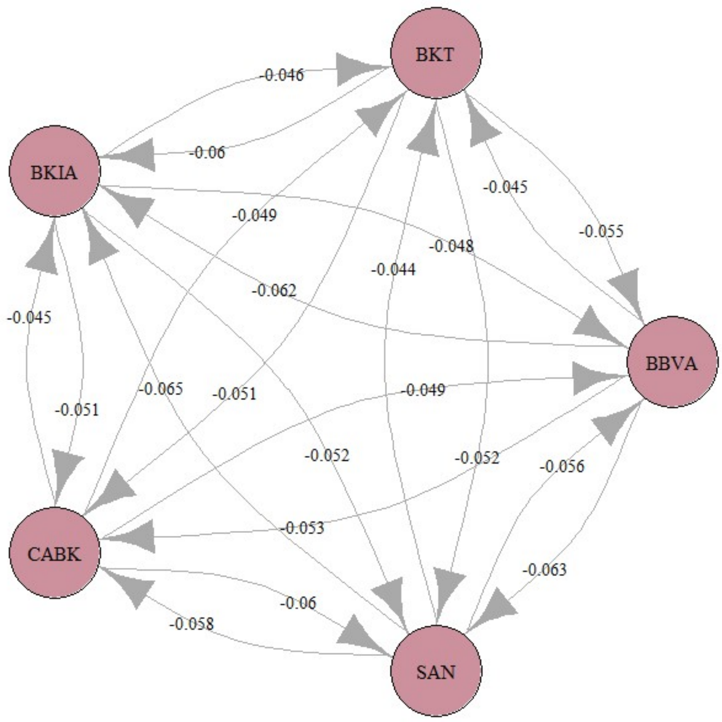


**Figure 7.1:** Network of ΔCoVaR using the Quantile Regression.

$y_{i,j}$ denotes the value of the relationship between nodes $i$ and $j$, from $i$ to $j$. In our case, both sociomatrices represent fully directed relations with $n = 5$ nodes.
The sociomatrices consider the banks in the following order:

- BBVA
- Bankinter
- Bankia
- CaixaBank
- Santander

The following **V** and **Z** matrices are the sociomatrices that represent the ΔCoVaR

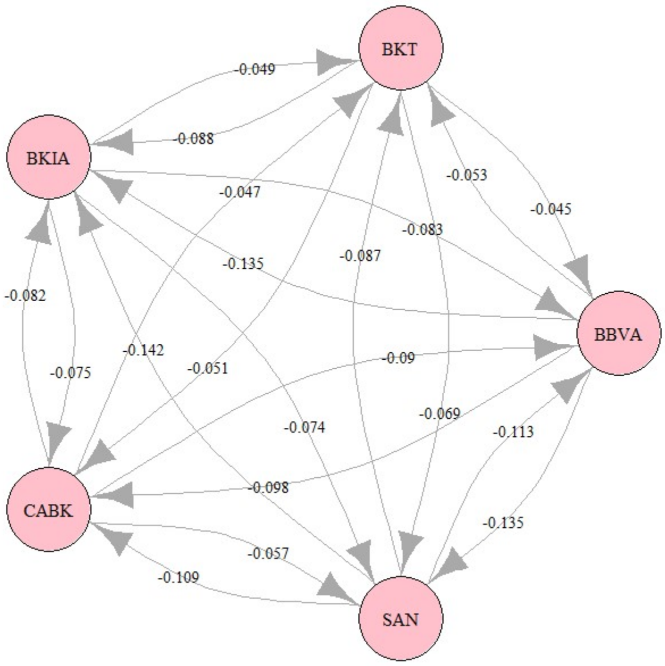


**Figure 7.2:** Network of ΔCoVaR using Quantile Regression Forests.

estimates obtained respectively from the Quantile Regression and the Quantile Regression Forests.

$$
\mathbf{V} = \begin{array}{c} \begin{array}{ccccc} \text{BBVA} & \text{BKT} & \text{BKIA} & \text{CABK} & \text{SAN} \end{array} \\ \begin{bmatrix} - & -0.045 & -0.062 & -0.053 & -0.063 \\ -0.055 & - & -0.060 & -0.051 & -0.052 \\ -0.048 & -0.046 & - & -0.051 & -0.052 \\ -0.049 & -0.049 & -0.045 & - & -0.060 \\ -0.056 & -0.044 & -0.065 & -0.058 & - \end{bmatrix} \end{array} \begin{array}{l} \\ \text{BBVA} \\ \text{BKT} \\ \text{BKIA} \\ \text{CABK} \\ \text{SAN} \end{array} \tag{7.2}
$$

$$
\mathbf{Z} = \begin{array}{c} \begin{array}{ccccc} \text{BBVA} & \text{BKT} & \text{BKIA} & \text{CABK} & \text{SAN} \end{array} \\ \begin{bmatrix} - & -0.053 & -0.135 & -0.098 & -0.135 \\ -0.045 & - & -0.088 & -0.051 & -0.069 \\ -0.083 & -0.049 & - & -0.075 & -0.074 \\ -0.090 & -0.047 & -0.082 & - & -0.057 \\ -0.113 & -0.087 & -0.142 & -0.109 & - \end{bmatrix} \end{array} \begin{array}{l} \\ \text{BBVA} \\ \text{BKT} \\ \text{BKIA} \\ \text{CABK} \\ \text{SAN} \end{array} \tag{7.3}
$$

Since this work is focused on the introduction of the Quantile Regression Forests as a method to estimate the CoVaR and the ΔCoVaR, the following reasoning will be done focusing on **Z** matrix.

The most basic statistic we can study in this matrix is the mean. For a directed relation it is computed in the following way:

$$
\bar{z}_{..} = \frac{1}{n(n-1)} \sum_{i \neq j} z_{i,j} \tag{7.4}
$$

which is the sum of all the relational measurements divided by the number of relational measurements. This is called *Grand mean*. Moreover, it could be possible to compute also the *Row mean* and *Column mean*, respectively:

$$
\bar{z}_{i.} = \frac{1}{n-1} \sum_{j:j \neq i} z_{i,j} \tag{7.5}
$$

$$
\bar{z}_{.j} = \frac{1}{n-1} \sum_{i:i \neq j} z_{i,j} \tag{7.6}
$$

Starting from these metrics, it is possible to retrieve two important metrics:

- Measure of *sociability*: $\hat{a}_i = \bar{z}_{..} - \bar{z}_{i.}$
- Measure of *popularity*: $\hat{b}_j = \bar{z}_{..} - \bar{z}_{.j}$

We can summarize all these values in the Tab.7.1.

Regarding the row means, so the effect that each bank has on another bank, we can notice that BBVA, BKT and SAN differ more from the mean. Also, looking at the

| | **BBVA** | **BKT** | **BKIA** | **CABK** | **SAN** |
|---|---|---|---|---|---|
| Grand mean | -0.0841 | -0.0841 | -0.0841 | -0.0841 | -0.0841 |
| Row mean | -0.1053 | -0.0633 | -0.0703 | -0.0690 | -0.1128 |
| Column mean | -0.0828 | -0.0590 | -0.1118 | -0.0833 | -0.0838 |
| $\hat{a}_i$ | -0.0211 | 0.0209 | 0.0139 | 0.0151 | -0.0287 |
| $\hat{b}_j$ | 0.0013 | 0.0251 | -0.0276 | 0.0008 | 0.0004 |

**Table 7.1:** Descriptive statistics of the sociomatrix $\mathbf{Z}$.

$\hat{a}_i$, we see that BBVA and SAN effects are below the mean, but above in absolute value: they give a higher contribution than the average to the other banks. On the contrary BKT, BANKIA and CABK are above the grand mean (below in absolute value) and they add a smaller part to the VaR of the other banks.

The same reasoning can be done looking at the column means. They measure how much each bank's VaR is modified by other financial institutions in the system. In this case, however, it is possible to notice that $\hat{b}_j$ of BBVA, CABK, SAN is close to 0: the column means are very close to the grand mean.

Keeping in mind that the average of $\hat{a}_j$ and $\hat{b}_j$ is 0, we can see that there is an higher variability among the row effects than the columns one.

We go deeper into the study of the network constructing a statistical model for $\mathbf{Z}$ based on its structure as a sociomatrix. The main intuitive approach is the simple ANOVA decomposition of the sociomatrix. The row variance can be considered as the heterogeneity of the nodes in terms of sociability. Similarly, the column variance is a way to represent the nodal heterogeneity in popularity. In particular, it evaluates the variability of $z_{i,j}$'s around the overall mean. In the ANOVA we evaluate the null hypothesis of no row and column heterogeneity through the F-test. Also, we assume that the errors are *i.i.d* from a mean-zero normal distribution.

In our analysis, we got small p-values which mean that we should reject the null hypothesis. We can affirm that there is significant degree of heterogeneity among the banks both as conditioner and conditioned. This result can be interpreted with the quote: the case present an high level of heterogeneity much more than would be expected if the $\hat{a}_i$ were all zero, or the $\hat{b}_j$ were all zero. A model-based version of the ANOVA decomposition is:

$$z_{i,j} = \mu + a_i + b_j + \epsilon_{i,j} \tag{7.7}$$

In Table 7.1, estimates for the metrics of *sociability* and *popularity* have been already computed.

Going back to the network, we give some definitions. A pair of nodes is referred to as a *dyad*, and a quantity that is measured or observed for multiple dyads is called a *dyadic variable*. The classical ANOVA ignores some fundamentals aspects:

- Since each pair of effect $(a_i, b_i)$ share a node, the relation between these vectors should be studied.
- Each dyad $i, j$ has two outcomes, $z_{i,j}$ and $z_{j,i}$. So, we should analyse the possibility that also $\epsilon_{i,j}$ and $\epsilon_{j,i}$ are correlated.

The response to the first point can be seen in Fig. 7.3.

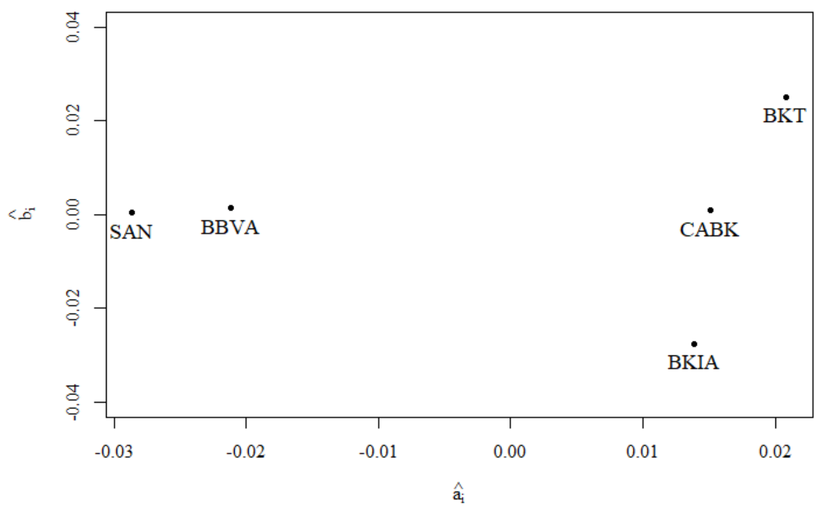


**Figure 7.3:** Scatterplot of sender and receiver of ΔCoVaR effect.

In this plot, we have compared the estimated row and columns effects, $\hat{a}_i$ and $\hat{b}_i$, for each row and column of the matrix and so, for each bank. The effects seems to be not linearly correlated at all, in fact the correlation value is very low: 0.066. *sociable* nodes are not *popular* and vice versa.

The next step consists in computing the correlation among $\epsilon_{i,j} = z_{i,j} - (\hat{\mu} + \hat{a}_i + \hat{b}_i)$ and $\epsilon_{j,i} = z_{j,i} - (\hat{\mu} + \hat{a}_i + \hat{b}_i)$. The correlation value is 0.305, a low value also in this case.

As we have already said, the ANOVA model does not quantify the dyadic correlations. There are other models that provide a more complete description of the sociomatrix like the *Social relations model.* (SRM), introduced by Warner et al. (1979), which estimate the parameters through a Markov Chain Monte Carlo (MCMC) algorithm that provide Bayesian inference for the parameters in the model.

Moreover, there is a family of statistical model for analyzing data about social network called *Exponential Random Graph models* (ERGM): it is based on the assumption that the structure in an observed graph can be explained by any statistics depending on the network and nodal attributes. In this case, we do not think they should be applied because we have a very small network and we have already tested that there is no correlation within dyads. However, some of these methods will be applied in Section 7.3 when dealing with a network made by a larger number of nodes.

## 7.2 Study of banks versus system

In this section we are going to consider the role of the system in the network in relation to the banks. In particular, we are going to study the VaR of the system conditional to a financial institution i when this move from it median state to a situation of distress and vice versa. This will be done through a network and the computation of the ΔCoVaR, as we have done in Section 7.1. Fig. 7.5 shows the two networks that represent graphically the relations between the system and the banks in terms of ΔCoVaR, respectively using Quantile Regression and Quantile Regression Forests. In this case, the sociomatrices will be the same as matrix **V** and **Z** with an additional row and column that is referred to the system.

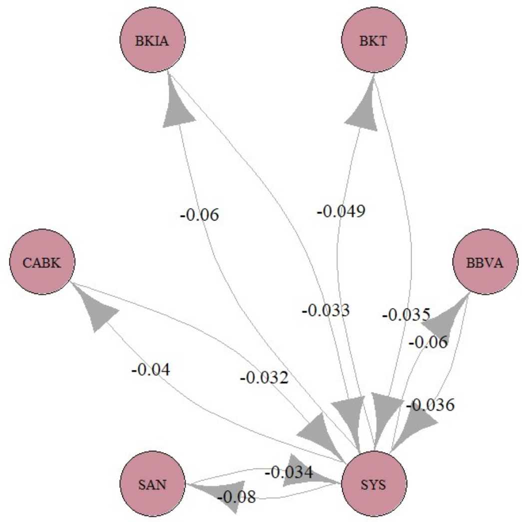


**Figure 7.4:** Network of ΔCoVaR using Quantile Regression.

$$
\mathbf{V_1} = \begin{array}{c}
\begin{array}{cccccc} \text{BBVA} & \text{BKT} & \text{BKIA} & \text{CABK} & \text{SAN} & \text{SYS} \end{array} \\
\begin{bmatrix}
- & -0.045 & -0.062 & -0.053 & -0.063 & -0.036 \\
-0.055 & - & -0.060 & -0.051 & -0.052 & -0.036 \\
-0.048 & -0.046 & - & -0.051 & -0.052 & -0.033 \\
-0.049 & -0.049 & -0.045 & - & -0.060 & -0.032 \\
-0.056 & -0.044 & -0.065 & -0.058 & - & -0.034 \\
-0.060 & -0.049 & -0.060 & -0.040 & -0.080 & -
\end{bmatrix}
\end{array}
\begin{array}{l} \\ \text{BBVA} \\ \text{BKT} \\ \text{BKIA} \\ \text{CABK} \\ \text{SAN} \\ \text{SYS} \end{array}
\tag{7.8}
$$

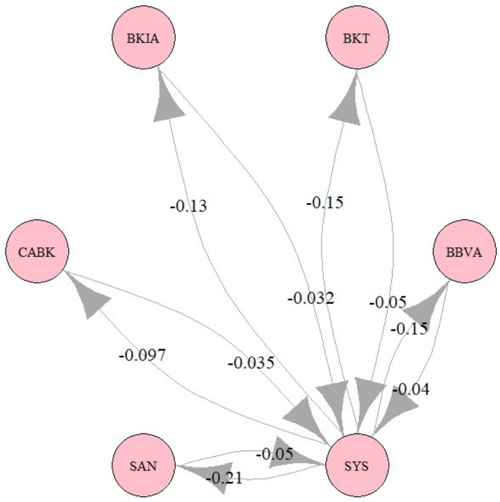


**Figure 7.5:** Network of ΔCoVaR using Quantile Regression Forests.

$$\mathbf{Z_1} = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{cccccc} \text{BBVA} & \text{BKT} & \text{BKIA} & \text{CABK} & \text{SAN} & \text{SYS} \end{array} & \\ \begin{bmatrix} - & -0.053 & -0.135 & -0.098 & -0.135 & -0.040 \\ -0.045 & - & -0.088 & -0.051 & -0.069 & -0.050 \\ -0.083 & -0.049 & - & -0.075 & -0.074 & -0.032 \\ -0.090 & -0.047 & -0.082 & - & -0.057 & -0.035 \\ -0.113 & -0.087 & -0.142 & -0.109 & - & -0.050 \\ -0.150 & -0.150 & -0.130 & -0.097 & -0.210 & - \end{bmatrix} & \begin{array}{l} \text{BBVA} \\ \text{BKT} \\ \text{BKIA} \\ \text{CABK} \\ \text{SAN} \\ \text{SYS} \end{array} \end{array} \tag{7.9}$$

A common aspect between the estimates of the ΔCoVaR using the two methods is that the effect of the system on each bank is always higher than the contrary effects. In lots of cases, it also doubles the bank effects. This aspect is partially visible from the CoVaR plots in Fig. 6.2 and Fig. 6.3: we notice that the CoVaR of the banks computed with respect to the system in distress is more distant from the VaR and the estimates are more variable with respect to the opposite case (the effect of that bank on the system).

Moving on, looking at the last column and, in particular, to the last row of matrix $\mathbf{V}_1$ and $\mathbf{Z}_1$, the difference between the Quantile Regression and the Quantile Regression Forests estimates stands out. The system contributions to the VaR of the banks diverge a lot in the two cases. In QR case we have values in the range of $[-0.080, -0.040]$, while in QRF values vary between $-0.210$ to $-0.097$. On the other hand, however, the gap between the estimated effects of the banks on the system computed using QR and those computed using QRF is smaller than the previous case. Coherently with the results obtained in Fig. 6.2 the system's stronger effect is on Santander bank's VaR both with QR and QRF. The banks effects are similar, they behave in the same way with respect to the system. The next step is computing in mathematical terms how the behaviour of matrix $\mathbf{Z}$ changes adding the row and column's system to it.[1]

Some descriptive statistics and the measures of *sociability* and *popularity* of matrix $\mathbf{Z_1}$ are computed in Table 7.2.

| | **BBVA** | **BKT** | **BKIA** | **CABK** | **SAN** | **SYS** |
|---|---|---|---|---|---|---|
| Grand mean | -0.0875 | -0.0875 | -0.0875 | -0.0875 | -0.0875 | -0.0875 |
| Row mean | -0.0922 | -0.0606 | -0.0626 | -0.0622 | -0.1002 | -0.1474 |
| Column mean | -0.0962 | -0.0772 | -0.1154 | -0.0860 | -0.1090 | -0.0414 |
| $\hat{a}_i$ | -0.0046 | 0.0269 | 0.0249 | 0.0253 | -0.0126 | -0.0598 |
| $\hat{b}_j$ | -0.0086 | 0.0103 | -0.0278 | 0.0015 | -0.0214 | 0.0461 |

**Table 7.2:** Descriptive statistics of the sociomatrix $\mathbf{Z_1}$.

Through the ANOVA test we reject the null hypothesis: there is a significant variability among the row and column means of the matrix. Then, we have plotted the new row and column effects $\hat{a}_i$ and $\hat{b}_j$ in Fig. 7.6.

We can notice that the System has a completely different behaviour with respect to the banks. It is on the extreme left on the plot, far from the banks, meaning that its row and column mean differ from the grand mean. Also, its row and column

[1] As in the previous case, we focus on Quantile Regression Forests estimates but the same reasoning could be repeated on matrix $\mathbf{V}$.

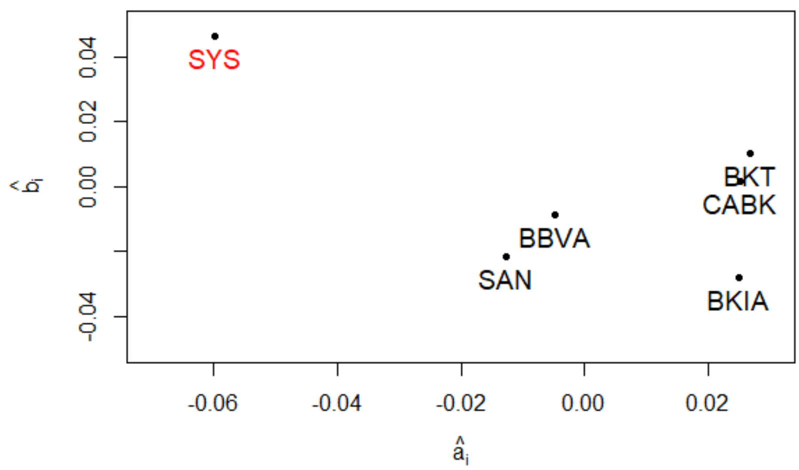


**Figure 7.6:** Scatterplot of sender and receiver of ΔCoVaR effect.

effects differ from those of the banks. In fact, as shown in Table 7.2, the row mean of the last row of the matrix (the one referred to the effect of the system) is $-0.1474$, the higher in absolute value. On the other hand, the column mean is $-0.0414$, the smaller among the others in absolute value.

Regarding the correlation between the row and column effects $\hat{a}_i$ and $\hat{b}_j$ is $-0.62$. We have a negative correlation and an higher value with respect to the case of matrix $\mathbf{Z}$. However, we cannot say that there is a linear relation among the points. The system point can be considered like a leverage, it is an isolated point that does not follow the behaviour of the other banks. As we have done in Section 7.1, we have computed also the correlation between $\epsilon_{i,j}$ and $\epsilon_{j,i}$ that, in this case, is $-0.1$. Since the correlation coefficient varies between 1, in case of strong positive linearly relationship and $-1$, in case of strong negative linearly relationship, a value near 0, as our $-0.1$, cannot be considered significant.

## 7.3 Analysis of the European situation

Once we have compared different methods to compute the CoVaR and ΔCoVaR measures, it would be interesting to replicate the study on the relations among banks through the use of network methods on a larger number of financial institutions. We have considered a subset of European banks, adding to the already mentioned Spanish ones, financial institutions from Italy, France and Germany. In particular, the banks considered in this section are:

- **Spain**: Banco Bilbao Vizcaya Argentaria (BBVA), Bankinter (BKT), Bankia (BKIA), CaixaBank (CABK), Santander(SAN), Sabadell (SAB), Liberbank (LIB);
- **Italy**: Intesa San Paolo (ISP), Unicredit (UNI), UBI banca (UBI), BPER banca (BPE), Banca Popolare di Milano (BPM);
- **Germany**: Commerzbank (CMZ), Deutsche bank (DEU);
- **France**: Credite Agricole (CA), BNP Paribas (BNP), Natixis (NAX).

Our goal is studying the effect of the banks on each other in terms of ΔCoVaR in order to discover clusters of banks that behave in the same way. For example, the nationality could be a significant aspect to take into account. After having computed the ΔCoVaR for each couple of banks, the resulting data are shown in matrix **N**, that can be found in the Appendix. In total we have 17 banks, so we are analyzing a $17 \times 17$ asymmetric matrix with undefined diagonal that brings to a directed and weighted network.

Fig. 7.7 represents the network coming from sociomatrix **N**. The 17 nodes are coloured in different ways according to the country from which the bank belongs. Regarding the edges, only the most significant ones are represented. In particular, the weights are multiplied by -20, in order to obtain positive values, and a threshold of 2 is applied. In this way, the values nearer to 0 are not considered. Finally, the width of the edges is proportional to the the new obtained values. From this plot, different aspects can be highlighted: a large number of edges that start from BPM

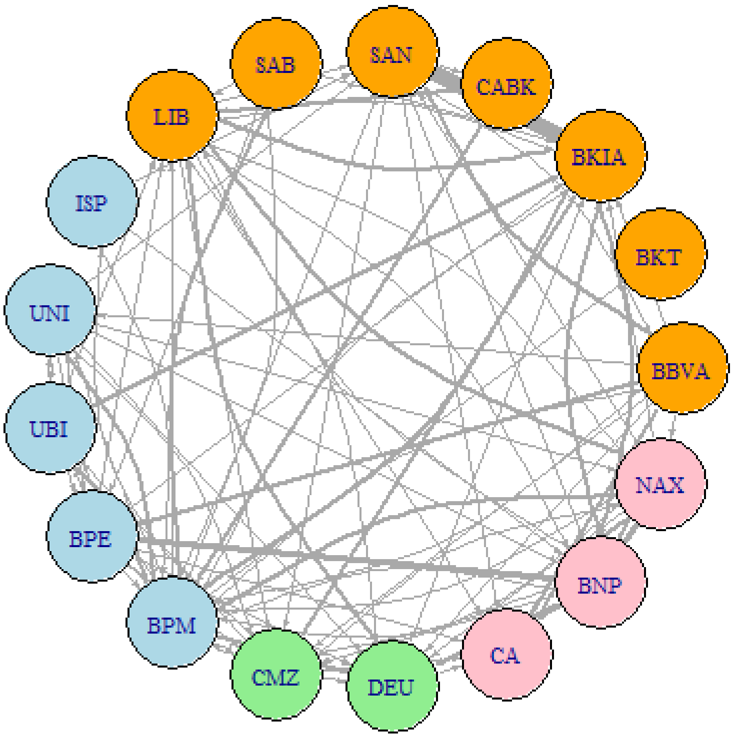


**Figure 7.7:** Network of ΔCoVaR using Quantile Regression Forests for the 17 banks.

are thicker than the others, this suggests that this bank have a greater impact on the VaR of the others. Also, only in some cases there are strong relations among banks from the same country, for example the Spanish or French group of banks.

Since the graph is fully connected, the ratio of the number of edges and the number of possible edges, called density, is 1. Due to the characteristics of the network, it is more useful focusing our attention on the weights of the edges. We have decided to study the distribution of the values of each row of the adjacency matrix but the same reasoning could be done considering the columns values. In this way we are able to study the behaviour of each banks in terms of $\Delta$CoVaR. The row values show the effect of each bank on the VaR of another bank when it moves from its median state to its $5\% - VaR$. The row values denote the part of the systemic risk of one bank that can be attributed to the effect of another bank.

The results are shown in Fig. 7.8 and Fig. 7.9, that represent the distributions grouped by nation.

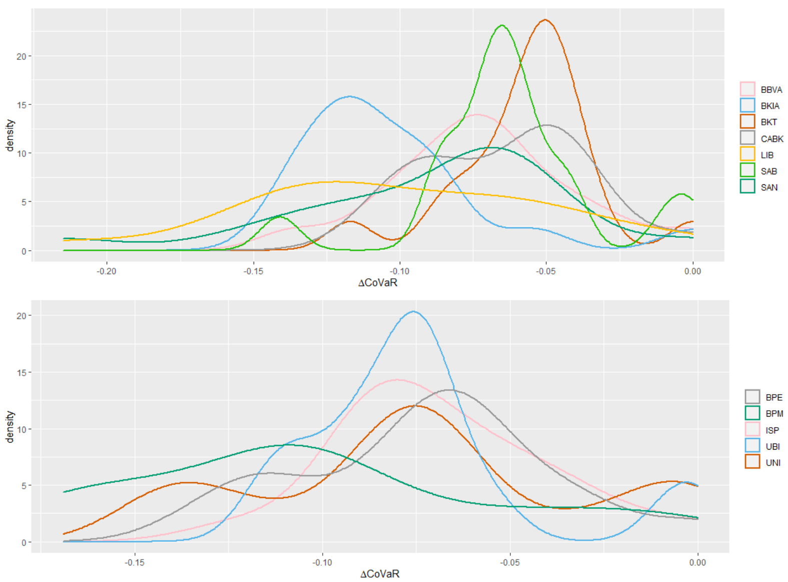


**Figure 7.8:** Distributions of row effects of Spanish (above) and Italian (below) banks

In the Spanish case, we have very similar distributions, mostly negatively skewed, while, in the Italian case, we can see that BPM has a very thick left tail so it shows a prevalence of very low values. In the comparison of the French banks, NAX shows

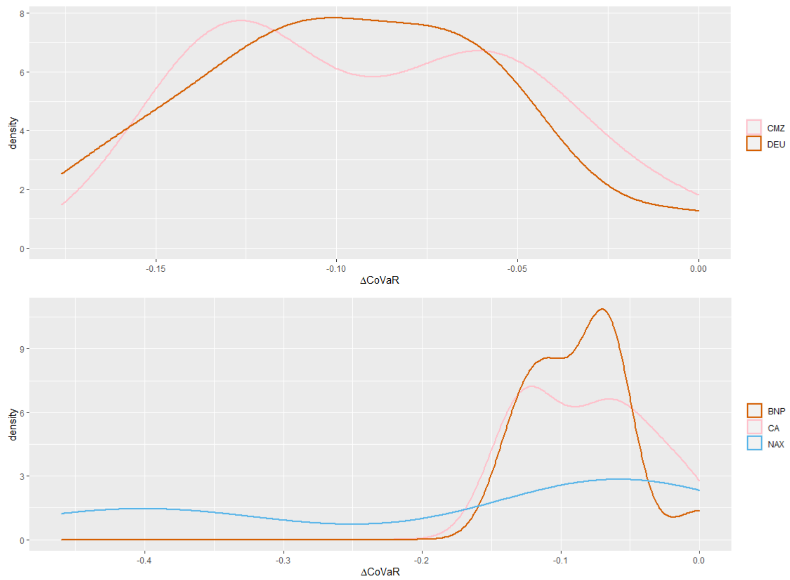


**Figure 7.9:** Distributions of row effects of German (above) and French (below) banks

a flatter distribution with respect to the other two. Finally, there is not a significant difference among CMZ and DEU distributions.

Now, we quantitatively analyse the difference among the row and column means. This is done through the ANOVA decomposition, as it has been done in Section 7.1 and 7.2: the model suggests the presence of a significant across-row and across-column heterogeneity.

Fig.7.10 represents a scatterplot that graphically compares the $\hat{a}_i$'s and the $\hat{b}_j$'s. The banks that show a greater influence on the VaR of the others are: BPM, BBVA, UNI. While, the banks that show a higher value in terms of column effects are: NAX, BKIA and LIB.

In this plot there is not a cluster of banks that stands out from the others, also the nationality component seems not to play an important role in the analysis. The next section is focused on the application of community detection techniques in order to find clusters of banks that could be grouped together.

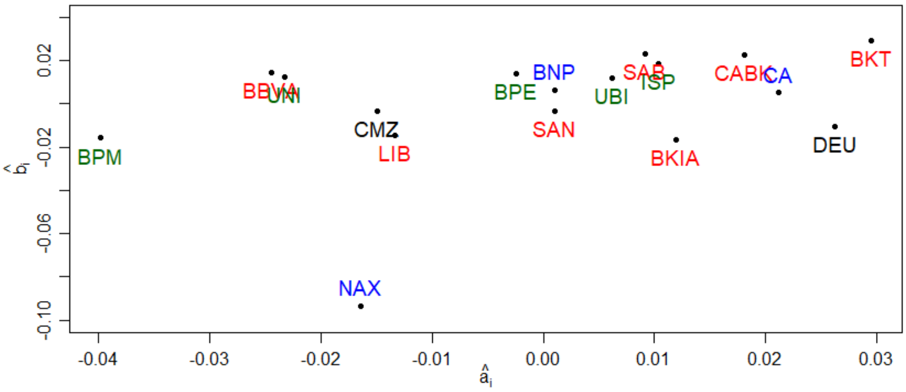


**Figure 7.10:** Scatterplot of sender and receiver of ΔCoVaR effect for the 17 banks.

### 7.3.1 Community detection

Having a network composed by a larger number of nodes with respect to the first part of our analysis (sociomatrices $\mathbf{V}$,$\mathbf{Z}$ and $\mathbf{V}_1$,$\mathbf{Z}_1$), it could be interesting to discover if there are some groups of banks that behave in the same way. In order to do this, different methods could be applied. In our case, I have considered the *Louvain* method based on the *modularity* measure and method based on maps of random walks.

### 7.3.2 Louvain method

The Louvain method is used to extract communities from large networks. Created by Bondel at al. [23] from the University of Louvain, it is based on the optimization of the modularity defined as the value that measures the density of links inside communities compared to links between communities:

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \tag{7.10}$$

where:

- $A_{ij}$ represents the edge weight between nodes $i$ and $j$;

- $k_i, k_j$ are the sum of the weights of the edges attached to nodes $i$ and $j$;
- $m$ is the sum of all of the edge weights in the graph;
- $c_i$ and $c_j$ are the communities of the nodes;
- $\delta$ is the *Kronecker delta function*: $\delta(x, y) = 1$ if $x = y$ or 0 otherwise.

Even if Duguè and Perez [24] have developed an algorithm for directed graph based on direct modularity, we have decided to keep the original version of the Louvain algorithm transforming our network into an undirected network. This is done summing the two weights and creating a single edge for each pair of vertices. In this way, we lose the information of direction but we are able to summarize the relations among each couple of banks and apply the classical concept of modularity. Algorithm 1 explains the steps in the Louvain methods that brings to the definition of communities.

**Algorithm 1** LOUVAIN ALGORITHM

1: Initialize the algorithm putting every node into a different community.
2: For each node $i$ is computed the gain in *modularity* obtained moving $i$ from its community to node $j$'s community, with $i$ and $j$ not connected.
3: Node $i$ is allocated in the community with the largest positive increase in modularity, otherwise it stays in its community.
4: The process is repeated until the modularity does not increase anymore.
5: The union of the new communities creates a new network and the process starts again from 2.

As shown in Fig.7.11 the algorithm suggests the presence of two clusters. The first is made by Natixis, Credite Agricole, BBVA, Commerzbank and Unicredit while the second by the remaining banks. We can see that the banks do not belongs to the same country and also, looking at Fig. 7.10 Credite Agricole bank is far from the other banks in its cluster. Now, we move to the application of another clustering method to our data to see if the results are coherent with the just obtained ones.

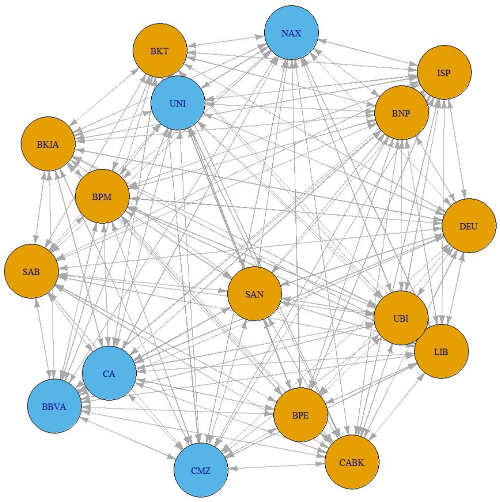


**Figure 7.11:** Cluster of banks resulting from the application of the Louvain method.

### 7.3.3 The Map method

Another common method used for community detection is based on maps of random walks. The concept at the base of this method is measuring the quality of a cluster by estimating the length of a code for paths in the graph through the probability of random walks.

Clusters have a connected flow inside, meaning that if you randomly follow the direction of edges, you tend to stay inside the same cluster. So, this method shows a duality structure: on one hand finding community structure in networks, on the other minimizing the description length of a random walker's movements on a network.

Each partition of the network is associated with an information cost for describing the movements of the random walker. Our goal is to find a module partition $\mathbf{M}$ of $n$ nodes into $m$ modules in order to minimize the expected description length of a random walk and so, to capture the communities of the network. According to [25], the average description length of a single step is :

$$L(\mathbf{M}) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p_{\circlearrowright}^{i} H(P^i). \tag{7.11}$$

This equation comprises two terms: the first one is the entropy of the movement between modules, the second is the entropy of movements within modules (where exiting the module also is considered a movement). Each is weighted by the frequency with which it occurs in the particular partitioning. The entropy of a variable $\mathbf{X}$ of $l$ possibile outcomes, that occur with frequency $p_i$ is defined as:

$$H(\mathbf{X}) = -\sum_{i=1}^{l} p_i \log p_i.$$

$q_{\curvearrowright}$ is the probability that the random walk switches modules on any step. $H(Q)$ is the entropy of the module while $H(P^i)$ us the entropy of the within-module movements. The weight $p^i_{\circlearrowright}$ is the fraction of within-module movements that occur in module $i$, plus the probability of exiting module $i$ such that $\sum_{i=1}^{m} p^i_{\circlearrowright} = 1 + q_{\curvearrowright}$.

Generally, this method is used for weighted but undirected links. However, [25] introduces a slightly modified version for directed graphs. The map is the same but a "teleportation probability" $\tau$ is introduced in the random walk that is the probability with which the process jump to a random node in the network.

In our case, this method is not able to find clusters of banks differently from the Louvain method. In the following section we will use the results obtained from the Louvain method as an explanatory variable in the SRRM models, in order to understand if it could help to explain better the relations into the network.

### 7.3.4 Models

Considering the lack of interpretation of some aspects of the classical ANOVA model, we move to the computation of the correlation between the row and column effects and then, the dyadic correlation evaluated in terms of residuals correlation. Unfortunately, in both cases, we have low values of correlation, 0.35 and 0.33 respectively. Going deeper into this aspect, there is a particular kind of ANOVA decomposition that describes variability of the sociomatrix also accounting for the within-dyad variability. It is called *Social Relations Model* (SRM) and takes the following form:

$$y_{i,j} = \mu + a_i + b_j + \epsilon_{i,j}$$

$$\{(a_1, b_1), ..., (a_n, b_n)\} \sim N(0, \Sigma_{ab}) \ i.i.d.$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i}) : i \neq j\} \sim N(0, \Sigma_e) \ i.i.d.$$

where:

$$\Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}, \Sigma_\epsilon = \sigma_\epsilon^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

In order to fit this model we run an iterative Markov Chain Monte Carlo (MCMC) that provides Bayesian inference for the parameters in the model. Fig.7.12 shows the results of the estimation of this model on our data.

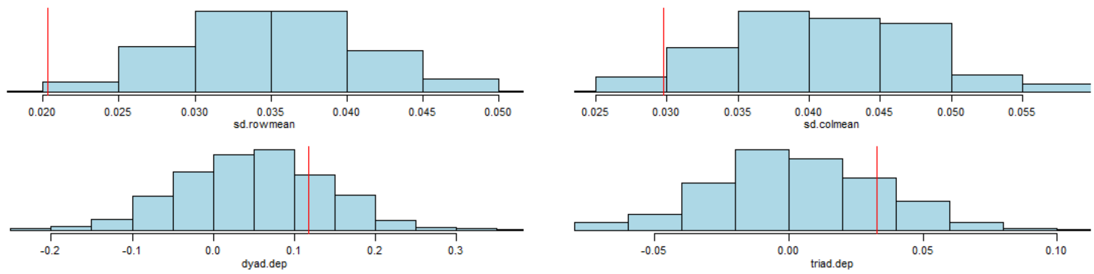


**Figure 7.12:** Estimation of the parameters of the SRM model.

The two plots of Fig.7.12 in the above panel show the posterior predictive goodness of fit summaries for the standard deviation of the row and column means. While, in the plot below there are the estimates of the empirical within-dyad correlation and a normalized measure of triadic dependence. Knowing that the red lines are the observed values, we try to avoid large discrepancies with the histogram that would mean a model lack of fit.

In our case, the model does not represent well the data with respect to the first two statistics, it represents better the within-dyad and triadic statistics. This can be due to the fact that the model is too simple and it is not able to explain the relations into the sociomatrix.

To address this problem, it is possible to add both dyadic and nodal variables obtaining a new model called *Social Relations Regression Model* (SRRM), which combines a linear regression model with the SRM model as follows:

$$y_{i,j} = \beta_d^T x_{d,i,j} + \beta_r^T x_{r,i} + \beta_c^T x_{c,j} + a_i + b_j + \epsilon_{i,j},$$

where $x_{d,i,j}$ is a vector of characteristics of dyad $\{i, j\}$, $x_{r,i}$ is a vector of characteristics of node $i$ as a sender and $x_{c,j}$ is a vector of characteristics of node $j$ as a receiver. In our case we assume that $x_{r,i} = x_{c,j}$ and they are referred as nodal covariates. The nodal variables used in the model are:

- Total asset, net income and number of employees, in order to consider aspects related to the dimension of the bank;
- Nationality and population of the country the bank belongs to;
- The clusters obtained from the Louvain method.

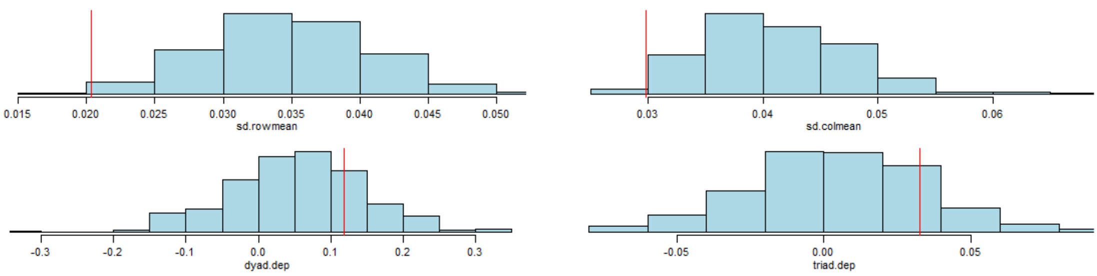


**Figure 7.13:** Estimation of the parameters of the SRRM model.

In Fig.7.13 there are the new results. We can see that the estimations are not improved and the model is still not able to evaluate the first two statistics. This means that the considered nodal variables do not add any contributions to the analysis. In fact, going deeper into the fitted model we can see that these variables are not significant for our model. This result is coherent with the plot in Fig. 7.10: the effect of a bank on another one does not depend neither on the dimension of the bank nor on the country the bank belongs to.

A last attempt is done considering a model, called AME, that combines an *additive main effects*, *multiplicative interaction* model (AMMI) and the SRRM model. In this case we have a model of the following type:

$$y_{i,j} = \beta_d^T x_{d,i,j} + \beta_r^T x_{r,i} + \beta_c^T x_{c,j} + a_i + b_j + u_i^T v_j + \epsilon_{i,j}.$$

$u_i$ is a vector of latent, unobserved factors of characteristics that describe node $i$'s behavior as a sender and similarly, $v_i$ describes node $j$'s behavior as a receiver. This model has been introduced to study whether there could be other nodal attributes, not already considered, whose multiplicative interactions might help in the description of the data and, in particular, to measure the triadic dependence among the banks. We have considered different ranks but the estimations do not show a significant improvement.

We can conclude that the obtained models partially explain the relations into the sociomatrix $\mathbf{N}$. The standard deviation of the row and column means are not well represented by any fitted model. The nodal variables we have taken into account are not useful for our analysis that could be improved considering variables that regards other aspects of the financial institutions or through the introduction of elements that could explain better the dyadic relations.

# Chapter 8

# Conclusion

The present work compares two ways of estimating the CoVaR and $\Delta$CoVaR: the well-known Quantile Regression versus the more recent Quantile Regression Forests. In the first part, the studied case consists of five banks, namely: BBVA, Bankinter, Bankia, CaixaBank, Santander. Then, the MSCI Europe index is used to analyze the financial system effects and lastly, a wider study is carried out through the introduction of 17 financial institutions coming from different European countries (Spain, Italy, France, Germany). Dealing with financial series, they are analyzed in terms of returns. All the series do not follow a normal distribution so, a modified version of VaR is computed, the Cornish-Fisher VaR.

We have proved how to estimate the CoVaR using the QRF, starting from the QR estimation. In fact, the key aspect in the computation of the CoVaR, is that the inverse distribution function is also the quantile function. In this way, conditioning the VaR of a financial institution to the VaR of another financial institution, we are able to compute the CoVaR and successively the $\Delta$CoVaR.

From our analysis, we have highlighted the higher variability of the the CoVaR estimates obtained by the QRF with respect to the CoVaR estimated through the QR. This aspect is accentuated by the introduction of additional lagged variables in the computation of the metrics that could affect the mean and the volatility of the risk measures. VIX, Euribor 3 months and Euro Stoxx 50 have been chosen in our analysis as systemic variables. An aspect that stands out is the estimation of the

QRF of the CoVaR in correspondence to low levels of the VaR: it shows very deep peaks suggesting us that it is high influenced from some even lower values, such as those characterizing the 2020. The same conclusion can be drawn when dealing with the estimation of the CoVaR of the MSCI Europe index.

Comparing graphically the effects of each financial institutions and the system on each other, it is difficult to say which of them gives the higher contribution to the other. In order to do this, the computation of the CoVaR has been replaced by the $\Delta$CoVaR. It quantifies how much $i$ adds to the $q$%-VaR of $j$, when $i$ moves from its median state to its $q$%-VaR level. This reasoning has been done using the values obtained from the QRF. We can conclude that among the banks, BBVA and SAN have a stronger effect on the others. On the base of the row and column effects, we can split the banks into two groups: SAN and BBVA are far from the group composed by BKT, CABK and BKIA. The situation changes with the introduction of the system. It has a very strong effect on the banks but, on the contrary, the banks effect is negligible, as we could have expected.

The last part of the present work uses the just explained tools to study the behaviour of banks that came from different European countries and how they interact with each other. We can affirm that the nationality component and several variables that are related to the dimension of the banks do not play an important role in the analysis of this network of banks. From our reasoning, the BPM, an Italian bank, is the one that have the greatest effect on the others, while NAX, a French bank, is the more influenced by the others.

We can conclude that the QRF is a good and innovative method to estimate the CoVaR. As the QR, instead of computing only the mean, it takes into account the quantiles, i.e. the whole distribution, a very helpful aspect when dealing with financial time series. Particular attention has to be put in tuning the parameters in the estimation procedure. For example, the number of trees used to train the forest should be adequate to the number of observations, otherwise there would be a useless use of resources, leading to overfitting the data. However, some drawbacks are present in the use of the QRF method for CoVaR estimation. The main disadvantage regards the use of the weights of a previously estimated model into another model, the values of the considered variables must fall in the same range in both the models. In addition, from our analysis we have seen that the QRF overestimate, in absolute value, the lower part of the distributions, suggesting it is influenced a lot by the 2020 crisis. We can conclude that there is not a preferable model between QR and QRF for the estimation of the CoVaR. All the latter aspects should be kept into account when choosing between QR and QRF, leaving to the researcher the choice of which method could be considered the best, according to the phenomenon under investigation. In future, this work could be integrated with additional aspects: other systemic variables could be added in the estimation of the CoVaR and most of all, it would be interesting studying and comparing the financial situation of the banks and the system in the periods during and after the COVID-19 pandemic.

# Bibliography

[1] R. Koenker, G. Bassett; Regression Quantiles; Econometrica,Vol. 46, No. 1 (January, 1978). 16

[2] Leo Breiman; Random Forests; Machine Learning, 45, 5–32, 2001, Statistics Department, University of California, Berkeley, CA 94720. 18

[3] Meinshausen;Quantile Regression Forests; Journal of Machine Learning Research 7 (2006) 983–999. 19, 23, 31

[4] Maria-Caridad Sevillano, Francisco Jareno; The impact of relevant international factors on the returns of IBEX 35 companies, 2000-2016; Applied Econometrics and International Development. 29

[5] T.Adrian and M.K.Brunnermeier: CoVaR;American Economic Review 2016, 106(7): 1705–1741. 2, 8, 21, 29, 30, 37

[6] Zangari, Peter. A VaR Methodology for Portfolios that include Options. 1996. RiskMetrics Monitor, First Quarter, 4-12. 13

[7] Laurent Favre and Jose-Antonio Galeano. Mean-Modified Value-at-Risk Optimization with Hedge Funds. Journal of Alternative Investment, Fall 2002, v 5. 13

[8] Robert F Engle and Simone Manganelli (2004) CAViaR, Journal of Business and Economic Statistics, 22:4, 367-381. 3

[9] Huang, Xin and Zhou, Hao and Zhu, Haibin, A Framework for Assessing the Systemic Risk of Major Financial Institutions (April 2009). BIS Working Paper No. 281. 3

[10] Segoviano Basurto, Miguel and Goodhart, Charles A.E., Banking Stability Measures (Janurary 2009). IMF Working Papers, 2009. 3

[11] Acharya, Viral V. and Pedersen, Lasse Heje and Philippon, Thomas and Richardson, Matthew P., Measuring Systemic Risk (May 2010). AFA 2011 Denver Meetings Paper. 3

[12] Allen, Linda and Bali, Turan G. and Tang, Yi, Does Systemic Risk in the Financial Sector Predict Future Economic Downturns? (July 7, 2012). 3

[13] Nikolaus Hautsch, Julia Schaumburg, Melanie Schienle, Financial Network Systemic Risk Contributions, Review of Finance, Volume 19, Issue 2, March 2015, Pages 685–738. 3

[14] Adams, Zeno and Füss, Roland and Gropp, Reint, Spillover Effects among Financial Institutions: A State-Dependent Sensitivity Value-at-Risk (SDSVaR) Approach (June 7, 2015). Journal of Financial and Quantitative Analysis (JFQA), Vol. 49, No. 3, 2014. 3

[15] Girardi, Giulio and Ergun, A. Tolga, Systemic Risk Measurement: Multivariate GARCH Estimation of CoVaR (November 5, 2012). 3

[16] Georg Mainik and Eric Schaanning, On dependence consistency of CoVaR and some other systemic risk measures, 2012, arXiv. 3

[17] Dong Hwan Oh and Andrew J. Patton (2018) Time-Varying Systemic Risk: Evidence From a Dynamic Copula Model of CDS Spreads, Journal of Business and Economic Statistics, 36:2, 181-195. 3

[18] Mauro Bernardi and Ghislaine Gayraud and Lea Petrella, Bayesian inference for CoVaR, 2013, arXiv. 3

[19] Mauro Bernardi and Antonello Maruotti and Lea Petrella, Multivariate Markov-Switching models and tail risk interdependence, 2013, arXiv. 3

[20] De Bandt, O., Hartmann, P., (1998), What Is Systemic Risk Today?, Risk Measurement and Systemic Risk. p.38-83 5

[21] Kaufman, G.G., Scott, K.E. (2003), What is Systemic Risk, and Do Bank Regulators Retard or Contribute to It?, The Independent Review, v.VII, nr. 3, pp. 371-391 5

[22] J.C. Rochet and J. Tirole, Why there are so many banking crises? The Politics and Policy of Bank Regulation - 2008 5

[23] Blondel, Vincent D; Guillaume, Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne. "Fast unfolding of communities in large networks". Journal of Statistical Mechanics: Theory and Experiment. 2008 57

[24] Dugué, Nicolas; Perez, A. Directed Louvain: maximizing modularity in directed networks, 2015 58

[25] Martin Rosvall, Carl T. Bergstrom: Maps of random walks on complex networks reveal community structure, 2007 59, 60

[26] Xiaohua Duan. Systematic Risk Measurement Based on CoVaR Model. International journal of Circuits, systems and signal processing. 2019. 13, 21

BIBLIOGRAPHY

# Appendix A

**N** =

| | BBVA | BKT | BKIA | CABK | SAN | SAB | LIB | ISP | UNI | UBI | BPE | BPM | CMZ | DEU | CA | BNP | NAX |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BBVA | – | −0.053 | −0.135 | −0.098 | −0.135 | −0.008 | −0.058 | −0.090 | −0.132 | −0.077 | −0.109 | −0.109 | −0.133 | −0.116 | −0.139 | −0.065 | −0.420 |
| BKT | −0.045 | – | −0.088 | −0.051 | −0.069 | −0.058 | −0.056 | −0.036 | −0.093 | −0.070 | −0.067 | −0.038 | −0.056 | −0.059 | −0.045 | −0.061 | −0.067 |
| BKIA | −0.083 | −0.049 | – | −0.075 | −0.074 | −0.087 | −0.140 | −0.063 | −0.074 | −0.071 | −0.089 | −0.113 | −0.029 | −0.096 | −0.049 | −0.105 | −0.061 |
| CABK | −0.090 | −0.047 | −0.082 | – | −0.057 | −0.072 | −0.153 | −0.049 | −0.068 | −0.072 | −0.054 | −0.106 | −0.054 | −0.064 | −0.058 | −0.063 | −0.063 |
| SAN | −0.113 | −0.087 | −0.142 | −0.109 | – | −0.071 | −0.058 | −0.092 | −0.055 | −0.054 | −0.073 | −0.083 | −0.132 | −0.059 | −0.124 | −0.142 | −0.049 |
| SAB | −0.069 | −0.061 | −0.131 | −0.046 | −0.089 | – | −0.137 | −0.063 | −0.070 | −0.077 | −0.113 | −0.168 | −0.053 | −0.067 | −0.015 | −0.072 | −0.074 |
| LIB | −0.099 | −0.072 | −0.118 | −0.086 | −0.105 | −0.068 | – | −0.064 | −0.006 | −0.065 | −0.054 | −0.055 | −0.128 | −0.176 | −0.109 | −0.124 | −0.359 |
| ISP | −0.062 | −0.044 | −0.054 | −0.033 | −0.115 | −0.043 | −0.087 | – | −0.080 | −0.089 | −0.074 | −0.169 | −0.069 | −0.096 | −0.086 | −0.129 | −0.054 |
| UNI | −0.040 | −0.052 | −0.094 | −0.056 | −0.121 | −0.043 | −0.093 | −0.082 | – | −0.116 | −0.121 | −0.141 | −0.115 | −0.133 | −0.079 | −0.110 | −0.460 |
| UBI | −0.064 | −0.056 | −0.104 | −0.087 | −0.059 | −0.067 | −0.126 | −0.087 | −0.142 | – | −0.133 | −0.165 | −0.048 | −0.081 | −0.024 | −0.063 | −0.049 |
| BPE | −0.078 | −0.040 | −0.115 | −0.063 | −0.062 | −0.066 | −0.127 | −0.093 | −0.071 | −0.092 | – | −0.143 | −0.119 | −0.119 | −0.121 | −0.127 | −0.069 |
| BPM | −0.137 | −0.117 | −0.116 | −0.054 | −0.154 | −0.141 | −0.215 | −0.122 | −0.117 | −0.109 | −0.066 | – | −0.124 | −0.117 | −0.073 | −0.105 | −0.330 |
| CMZ | −0.075 | −0.034 | −0.121 | −0.099 | −0.065 | −0.083 | −0.030 | −0.075 | −0.143 | −0.107 | −0.035 | −0.104 | – | −0.103 | −0.125 | −0.076 | −0.440 |
| DEU | −0.059 | −0.080 | −0.115 | −0.043 | −0.071 | −0.085 | −0.112 | −0.045 | −0.006 | −0.006 | −0.030 | −0.018 | −0.086 | – | −0.132 | −0.075 | −0.051 |
| CA | −0.025 | −0.045 | −0.092 | −0.046 | −0.049 | −0.062 | −0.066 | −0.026 | −0.027 | −0.079 | −0.057 | −0.094 | −0.078 | −0.049 | – | −0.064 | −0.370 |
| BNP | −0.077 | −0.056 | −0.130 | −0.040 | −0.078 | −0.060 | −0.111 | −0.080 | −0.083 | −0.081 | −0.083 | −0.114 | −0.139 | −0.154 | −0.063 | – | −0.093 |
| NAX | −0.099 | −0.066 | −0.104 | −0.091 | −0.215 | −0.054 | −0.142 | −0.083 | −0.080 | −0.095 | −0.066 | −0.109 | −0.152 | −0.151 | −0.132 | −0.100 | – |