Banks' risk clustering using k-means: a method based on size and individual & systemic risks

Mathieu Mercadier [a,d,*], Amine Tarazi [b], Paul Armand [c], Jean-Pierre Lardy [d]

October 20, 2021

a: ESC Clermont Business School, CleRMa-UCA, 4 Boulevard Trudaine, 63000 Clermont-Ferrand, France
b: Université de Limoges, LAPE, 5 rue Félix Eboué, 87031 Limoges Cedex, France
c: Université de Limoges, XLIM UMR CNRS 7252, Département Maths Info, 123 avenue Albert Thomas, 87060 Limoges, France
d: JPLC SASU, 54 avenue de la Révolution, 87000 Limoges, France

Preliminary draft. Please do not circulate without the permission of the authors.

Abstract

This paper provides a global, transparent and dynamic decision support tool that clusters listed banks depending on their riskiness using an unsupervised learning algorithm. This entirely automatic process is updated weekly on a dedicated website, and refreshable on-demand. A large set of stand-alone and systemic risk indicators are computed and reduced to fewer representative factors through dimensionality reduction. These factors are set as features for an adjusted version of a nested k-means algorithm that handles missing data. The obtained individual banks' multi-dimensional clustering results are also aggregated per country and region. Although this procedure is designed to be editable regarding the number of banks and indicators, an empirical analysis is presented based on 161 listed banks and 85 indicators reduced to 11 main factors during the 2004-2021 period. The analysis underlines the importance of paying a particular attention to the ambiguous impact of banks' size on systemic measures.

Keywords: Decision Support Tool; Automatic Process; Banking Risks; Systemic Risk; K-means
JEL classification: G01; G20; G28; C44

---

[*] Corresponding author at: +335 55 32 81 88.
E-mail addresses:mathieu.mercadier@esc-clermont.fr (M. Mercadier), amine.tarazi@unilim.fr (A. Tarazi), paul.armand@unilim.fr (P. Armand), jp.lardy@jplc.fr (J.P. Lardy).

We are grateful to C. Hurlin and J.S. Oberoi for their constructive comments, as well as to the members of the CleRMa for further constructive comments; the usual disclaimer applies. Substantial parts of this research were carried out while Mercadier was a doctoral researcher at LAPE, Université de Limoges, France.

1 Introduction

Since the global financial crisis (GFC) of 2007-2008, banks' soundness has drawn a lot of attention from academics, regulators and politicians regarding their supervision and regulation. Over the last decade, numerous governments had to decide whether they should rescue banks (e.g., the United States' Troubled Asset Relief Program or the European Financial Stability Facility), notably to avoid the risks' propagation to the whole market. This systemic dimension led researchers to establish new indicators dedicated to quantify the systemic risk. However, it has been for over half a century that economists and practitioners have developed multiple ratios and indicators to evaluate specific risks aspects.

The objective of this work is to provide a decision support tool which groups listed banks in function of diverse representative factors built on risk indicators, including the systemic dimension. Traditionally, such tasks are carried out with econometric methods requiring various hypotheses, e.g., a weighting average of the set of indicators, which would need complex assumptions to remain dynamical. We rather select a simpler and more transparent method. Following straight on Moore's law, computing devices capacities put unsupervised learning at the center of computers' optimization techniques. Although machine learning research is currently booming (Shoham et al., 2018) the core of the algorithm used in this methodology was set up more than sixty years ago. To highlight hidden patterns in our input sample, we pick a well-known machine learning algorithm, the k-means algorithm (Lloyd, 1982)[1], which only requires specifying the number of clusters in advance. The algorithm is adjusted in our process to deal with missing data. The whole available information for every bank for every date is summarized in a set of clusters ranked over multiple risk dimensions. In addition, our approach provides the possibility to aggregate the clustering per country and region.

Both indicators' values and banks' clusters are updated weekly on a dedicated website[2], refreshable on-demand. It was notably inspired by the three following websites, dedicated to risk analysis. The Volatility Institute's website, called V-Lab[3] provides detailed ranking of banks per individual indicator, while we aggregate risk indicators to provide an overall ranking of the banks in function of the risks. Like us they provide weekly values that can be aggregated at country level but fewer systemic risk indicators are available. The Risk Management Institute[4], publicly provides two measures, one assesses the probability of default and the other gives a ranking number for banks and insurers in term of systemic

---

[1] Developed since 1957 but published in 1982.

[2] The website Banking Risk Ranking provides a global view of the results, and an additional weekly monitoring detailed report is downloadable for authorized users.

[3] See https://vlab.stern.nyu.edu/

[4] See https://nuscri.org/en/

risks derived from their size and interconnectedness. Unlike us, they do not aim to cluster banks per risk level on a weekly basis, they rather assess monthly aggregated probabilities of default per country. Finally, the Federal Reserve Bank of St. Louis[5] provides banks related solvency and risk fundamental measures at country's level, but the most recent data are at least two years old and we deal with more risk indicators.

Among the 85 risk indicators there are stand-alone metrics such as the volatility, VaR, expected shortfall, Roy's Z-scores (Roy, 1952) and credit spread approximations (Finger et al., 2002; Mercadier and Lardy, 2019) that are computed to assess banks' health. Albeit to a lesser extent, a few financial ratios are also integrated to our process as risk practitioners use them (Altman, 1968)[6]. Along with correlations and betas, other indicators are chosen to assess banks' size, liquidity and regulatory requirements. The systemic indicators born after the GFC are also calculated like the Marginal Expected Shortfall (MES) by Brownlees and Engle (2016), LRMES & SRISK by Acharya et al. (2012); Brownlees and Engle (2016), Tail Beta (Straetmans et al., 2008) and ΔCoVaR by Adrian and Brunnermeier (2016). Once the issues related to automatic extraction of financial data are handled, these indicators are computed, cleaned and summarized into representative factors through principal component analysis.

An empirical analysis is performed in the last section of this paper, detailing the choices of the number of clusters and explaining why we leant in favor of nested clustering. In addition, we provide chosen graphical examples available on the website to illustrate our point. In this context, 161 listed banks from 31 countries are ranked in function of their relative riskiness during the 2004-2021 weekly period using our method. A blend of financial market and balance sheet data is used to compute 85 indicators quantifying both individual and systemic risks. Based on economic knowledge, the number of indicators is reduced to 11 main features set as inputs of the nested k-means clustering. From an economic perspective, the centroids' analysis confirms the importance of paying a particular attention to bank size impact on systemic measures. In line with Varotto and Zhao (2018) that dealt with the overshadowing effect of bank size on systemic indicators, this methodology differentiates two cases. On the one hand, small banks tend to be overdefined as safe, according to systematic metrics, namely MES, ΔCoVaR and Tail Beta, and on the other hand SRISK is more inclined, by construction[7], to overstate big banks as risky. Lastly, we briefly discuss the similarities between the clustering and the five year Credit Default Swaps of JPMorgan Chase and Lehman Brothers.

---

[5] See https://fred.stlouisfed.org/

[6] Altman's ratio is not used in this study as it is not adapted to banks.

[7] As SRISK is expressed in US dollars.

Our aim is to build a transparent, self-contained and easily editable methodology to automatically rank worldwide listed banks based on stand-alone and systemic risk indicators on a weekly basis. A methodology designed to assist academics and practitioners to compare listed banks in terms of riskiness. In addition, the ability to aggregate per country or region may help them to identify fragile areas and prevent major future issues. Moreover, this dynamic methodology can easily be set for specific studies, with the addition or removal of banks or indicators. In addition, our outputs distinguish the more severe impact on the banking industry of the GFC compared to the current COVID-19 pandemic.

The rest of the paper is laid out as follows. Section 2 provides a brief interdisciplinary literature and website review. Section 3 introduces the process automatically extracting and handling the raw data, in terms of cleaning and outliers removal. Section 4 presents the different risk indicators used to assess specific risks and their corresponding factors. Section 5 deals with the dimensionality reduction and the adjusted clustering algorithm. Finally, section 6 provides a clustering analysis and an empirical illustration of the results.

2 Related literature and websites

Our methodology builds a bridge between two strands of the literature and more specifically two distinct research areas: the banking literature as commonly named in the area of Economics and Finance with a focus on default and systemic risks and the unsupervised learning techniques drawn from Mathematics and Computer Sciences.

Some papers deal with the banking sector using a clustering algorithm such as Moldovan and Mutu (2015) which assess the relationship between banks' default probability and their risk taking incentives for listed and unlisted European banks, using both partitional and hierarchical clustering. Knotek (2014) studies European banking sector similarities with the Ward method. A paper closely related to Soerensen and Gutiérrez (2006) which detect some basic patterns and trends in the Eurozone banking sector. In another recent article by Ercan and Sayaseng (2016), the grouping of the banking sectors based on the banking ratios show that the EU countries in similar geographic area and with higher economic partnership tend to group in similar clusters. Dardac and Giba (2011) create groups of financial crises to highlight the most efficient politics to manage these crises. Furthermore, Dardac and Boitan (2009) gather Romanian credit institutions in function of their risk profile and profitability with an agglomerative hierarchical clustering algorithm. Cernohorska (2017) highlights the similarities between Czech and Slovak banking sectors with hierarchical agglomerative clustering. Zarandi et al. (2014) use a hybrid k-means and Grey relational analysis to measure the relative efficiency of various Iranian banks.

In our study, we work with a partitional algorithm while these papers generally use hierarchical clustering. Like the Ward agglomerative algorithm (Ward, 1963) that merges, at each step, chosen pairs of clusters to summarize the information. While it is impractical for our large number of observations, we compared our k-means to Ward algorithm using sub samples as inputs and got close results. In addition, it appears in those sub cases that Ward algorithm is less reactive than k-means to new pieces of information confirming our choice. Another known unsupervised learning algorithm, the self-organized map (SOM), is not chosen because we wish more than three dimensions to rank our banks.

The authors of the previously cited studies mostly directly deals with standard financial ratios as input, such as return on assets, equity to assets ratios, etc. We rather choose to summarize our inputs into representative factors, in concordance with LeCun (2012) stating that filtering the inputs can greatly improve the results. Specifically, the data are transformed, among other modifications, with the John and Draper (1980)'s log-modulus transform and then, the number of features is reduced through principal component analysis (Pearson, 1901). In their paper, Giglio et al. (2016) build systemic indexes from estimators dimension reduction to predict macroeconomic shocks. As reported by Xu et al. (2015), k-means is more efficient in low-dimensional space. In addition to filtering noise and reducing computation time, the risk of converging toward local optima is reduced, in line with various recent papers from broad research domains (Danley, 2019; Zhu et al., 2019; Prabhu and Anbazhagan, 2011). Beyond financial ratios, raw balance sheet data, betas and correlations, individual bank risks are measured with a large number of indicators having an academic background. For instance, the Z-score, commonly attributed to Boyd and Graham (1986), Hannan and Hanweck (1988) and Boyd et al. (1993), and its market version the MZ-score (Lepetit et al., 2008; Saghi-Zedek and Tarazi, 2015) assess bank solvency. Credit spreads are approximated with the E2C developed by Mercadier and Lardy (2019) and the formula from the CreditGrades model (Finger et al., 2002). Systemic risk indicators are the Marginal Expected Shortfall, that was introduced by Brownlees and Engle (2016), its long term versions the LRMES and the SRISK (Acharya et al., 2012; Brownlees and Engle, 2016), the Delta Conditional Value-at-Risk (Adrian and Brunnermeier, 2016) and the Tail Beta (Straetmans et al., 2008; deJonghe, 2010). Besides being conventionally computed, the Value-at-Risk, the expected shortfall and the systemic indicators integrating them are computed using weighted historical methods (Boudoukh et al., 1998; Hull and White, 1998) to obtain more reactive metrics.

Many papers confirm that large banks are related to significantly higher systemic risk (deJonghe, 2010; Banbula and Iwanicz-Drozdowska, 2016; Kleinow et al., 2017). But the role of the size as a major determinant of the systemic risk is still debated (Zhang et al., 2015; Laeven et al., 2016; Sedunov, 2016; Hué et al., 2019), notably regarding its tendency to overshadow other factors contributing to banks' systemic importance (López-Espinosa et al., 2012, 2013; Weiß et al., 2014; Varotto and Zhao, 2018; Elyasiani and Jia, 2019).

Although initially introduced by Steinhaus (1956), the term "k-means" was firstly used by MacQueen (1967). The standard version of the algorithm used in this paper, was developed ten years earlier by Lloyd in 1957 at Bell Telephone Laboratories but only published in 1982 (Lloyd, 1982). This algorithm is also referred to as Lloyd-Forgy (Forgy, 1965). Inconsistent results due to the presence of local optima can be reduced by using the k-means++, an initialization method developed by Arthur and Vassilvitskii (2007), or with other techniques as shown by de Oliveira and do Carmo Nicoletti (2019). Moreover, such an algorithm requires a predetermined number of clusters to split the sample. To that end, we used the inertia based elbow curve and silhouette criteria based on the work of Rousseeuw (1987). As explained in section 5.2, we use an adjusted nested version of the algorithm (Chi et al., 2016; Niedzielski et al., 2017).

In addition to the discussed literature, we present some websites that inspired our project. Notably, the Volatility Laboratory (V-Lab) developed by the Volatility Institute from the New York University Stern School of Business under the direction of Nobel Laureate Professor Robert Engle. The Volatility Institute team of market risks specialists provides very complete and detailed information on various aspects of the risks with the following sections: volatility, correlation, systemic risk, liquidity, fixed income and more recently the climate risk analysis and a section that computes the long run Value-at-Risk. All their measures use modeling linked with conditional volatilities highly influenced by Engle's literature (Engle, 1982; Bollerslev, 1986). Their section about the systemic risk analysis particularly influenced our project.

Like V-Lab, we provide LRMES and SRISK weekly computations and also propose to aggregate them at country's level[8]. As mentioned above, they use heavy modeling for the underlying variables likely making their indicators more accurate, as their final aim is to rank banks per indicator. While we obtain very similar values with our indicators, we calculate them in a simpler way as they are aggregated into noise reducing factors on which the clustering is done. In terms of systemic risks, V-Lab monitors the market's impact on banks while we also look at the reciprocal impact of banks on the market, for instance via the ΔCoVaR measure. Although, their strength comes from individual measure accuracy, we focus on multiple closely related measures summarized in single factors. For instance, they use a dedicated econometric process to handle asynchronous issues between firms and indexes from different time zones. We do not dig this far for our measures but

---

[8] And at regional's level in our website.

we compute multiple systemic metrics using three different indexes, one local, one global banking and one global, where the local one does not have any asynchronous issues. Our methodology only focuses on the banking sector with a sample of 161 listed banks since 2004, while V-Lab ranks 947 companies for the whole financial sector since 2000. To summarize, they provide an accurate financial ranking per few systemic indicators while we focus on clustering banks based on the many facets of risk.

Other websites provide similar information, such as the one of the non-profit credit research initiative undertaking at the Risk Management Institute of the National University of Singapore that publicly furnishes two measures, the Corporate Vulnerability Index (CVI) and their Systemically Important Financial Institution ranking (CriSIFI). The CVI is computed as the average of a country[9]'s firms' probability of default. For each country, they propose three CVI metrics, the first one is equally weighted. The second one, the CVI value weighted, is weighted by market capitalizations. Finally, they propose the tail CVI which stands for the top fifth percentile of the individual probabilities of default for a specific country. The CVI measures are updated daily on Bloomberg L.P.. Contrary to our approach, they do not aim to cluster banks per risk level, their CVI rather assesses an aggregated probability of default per country. Regarding their CriSIFI, they rank banks and insurers in term of systemic risks that are determined by their size and interconnectedness with others, which differ from V-Lab's SRISK measure. They do not display the indicators values but the rankings are updated monthly and cover 2,148 banks and insurance companies since 2000.

Among many economic indicators available on an interactive map[10], the Federal Reserve Bank of St. Louis provides some banks related solvency and risk fundamental measures. For instance, like us they compute banks' non-performing loans to Gross Loans and Z-scores. But while we provide each indicator at individual bank level, they focus on aggregated measures at country's level (or state's level for the United States) and display data that are at least two years old.

3 Programming specificities and data extraction

3.1 Programming specificities

Requiring a reactive and well-resourced database, the data are extracted from Bloomberg L.P. as a unique provider. A key point in using this software is that it comes along with an application programming interface (API) allowing developers to automatically extract financial data. Various programming languages such

---

[9] In addition, they propose the CVI measure for the S&P 500.

[10] See https://geofred.stlouisfed.org/map/

as C/C++, C# (.Net) & Java, and more recently Python & Perl are available. Regarding our project, we choose java for both its speed and portability, meaning that the compiled bytecode can be interpreted on any hardware having a compliant java virtual machine (JVM).

Then, the process is written in Python 3.6 (Python Software Foundation, 2016) and relies on the packages "numpy" by Van Der Walt et al. (2011), "pandas" by McKinney (2010) and for visual outputs on "matplotlib" by Hunter and Dale (2007). Moreover, the principal component analysis (PCA) is called from "sci-kit learn" (Pedregosa et al., 2011). Although the k-means class is available in "sci-kit learn", our own C++ version is developed[11] to deal with missing data and is called from Python as a Dynamic-link library.

3.2 Extraction methodology

Unsupervised learning algorithms are built to deal with any type of digital inputs, and many papers (Zarandi et al., 2014; Dardac and Boitan, 2009; Dardac and Giba, 2011; Soerensen and Gutiérrez, 2006; Knotek, 2014; Ercan and Sayaseng, 2016; Moldovan and Mutu, 2015) use financial report data and ratios as features. However, in this methodology, we compute risk indicators requiring both accounting and market data as inputs. An obvious limitation of comparing banks based on both types of data is that it restricts our universe to listed banks. Moreover, such information is released with different frequencies – daily for market data and quarterly, semi-annually or yearly for accounting ones.

By construction, dealing with automatically extracted listed banks' data potentially convey some issues. For instance, corporate actions issuance such as dividend payments, stock splits or equity offering might impact stocks' whole price and volume history when it is adjusted by the provider. Furthermore, mergers and acquisitions and failures modify banks' market status, thus our process checks it and removes all data after the date at which the stocks are considered delisted or acquired. In addition, banks' tickers might rarely be modified by the provider leading to non-available information and forcing us to manually identify the new corresponding ticker. Thus, each time the code is run, automatic audits are conducted on the data to highlight their consistency. Furthermore, economic releases frequencies might evolve through time and the data might contain missing points or outliers (which could be corrected at a later date by the provider). All these issues lead us to consider a full update of the database every time the process is launched.

As it is required to compute some indicators, banks' countries of domicile are set as common keys to match them with their corresponding national indexes or to

---

[11] The code dealing with the outliers is programmed in C.

convert them in US dollars. Thus, beyond banks, national indexes and currencies data, the program extracts a few descriptive fields for all Bloomberg's tickers. For each company, the extracted properties are its name, industrial sector, name & ISO code of its country of domicile, the stock currency and the bank's respective fund currency, frequency of data release, its status and potential date of suspension.

3.3 Dealing with asynchronicity and outliers

The input database is split in two, the daily data on one side and a monthly table on the other side. Once cleaned, both data sets are merged into a unique weekly table. As previously mentioned, corporate actions are dealt with through the extraction of the whole data history (cf. section 3.2) each time the process is launched. Handling the daily data is pretty straightforward as it only requires removing outliers, but it should be done carefully. Since we focus on risk, outliers must be distinguished from extreme values. Additionally, our process being dynamic, our outlier detection algorithm, explained below, works "on-the-fly" regardless of the input data set and is designed to be as simple and global as possible to handle various frequencies.

Bloomberg L.P. being specialized in market data, outliers are extremely rare in this context, but erroneous data occur more often in financial accounts. The countries present in our sample release similar financial results at different dates and frequencies. Another issue is that accounting data are released at the end of the month that can fall on weekends – only keeping business days might lead to omit some values. Studying the data, we noticed additional inconsistencies between quarterly, semiannually and yearly data. First, for specific month and year, the reference date might be different between these files for a given bank. An issue easily fixed by forcing the reference dates as the calendar's end of the month. But it becomes trickier when the corresponding values are different for a specific date. We choose to keep the closer value to the mean of the previous and consecutive values, instead of the average of these different values as at least one of them might be too extreme, negatively influencing the average. Counter-intuitively[12], these outliers can randomly be on the quarterly, semiannually or yearly file.

In short, the algorithm[13] parses each time series with two overlapped rolling windows over an automatically defined period – depending upon the number of points. This process sets values as outliers if they are beyond $\mu \pm 5\sigma$, where μ is the mean value and σ the standard deviation.

Additionally, we observe that while passing through the previous filters, some

---

[12] One can think that outliers are less expected on yearly fundamental data.

[13] The outliers' removal algorithm is available upon request to the authors.

Bloomberg's data remains odd. Actually, some values might be originally wrongly set at zero, instead of as missing value, making no economic sense. An issue neutralized by replacing the null values by missing data for some specific economic fields[14].

At this point, all banks' variables belong either to a monthly frequency or a daily one. To avoid mismatches for the indicators' computations, fundamental data sets remain on a monthly basis and only Fridays' values are kept for daily ones. Then, prior month end inputs are forward filled[15] (repeating the last available value) before weekly calculations and those expressed in national currencies are converted in US dollars.

4 Risk indicators and representative factors

4.1 Data pre-processing

According to LeCun (2012), to efficiently perform a basic machine learning algorithm, it is fundamental to pre-process the data, reduce dimensions and extract hand-crafted, domain specific features, a process called feature engineering. The success of the training is highly dependent on this chosen representation (Bengio et al., 2013). In this respect, we firstly transform the raw data in term of risk indicators that are then summarized into fewer representative factors. A first phase of feature engineering is performed for the raw data in section 3.3, through outliers' removal and market and fundamental data's dates and currencies matching. Once risk indicators are computed from raw data, a second phase of features engineering is applied to them, involving cleaning, orientation, transformation, standardization and dimensional reduction (cf. section 5.1). Although the k-means algorithm can manage directly the indicators, it is a useless time consuming process as many highly correlated indicators are handled. Besides, the process delivers close results[16] but is seven times slower and outcomes are more difficult to interpret. Therefore, dimensionality reduction is relevant to summarize the correlated indicators into representative factors. Specifically, eleven chosen factors are derived with principal component analysis from subsets of the 85 banking risk indicators introduced in this section, along with their respective factors.

It makes sense to link some indicators together, notably regarding systemic measures. In fact, many systemic indicators are built in the same way, only the

---

[14] Concerned data: bid-ask spread, probability of default, total assets, total equities, common equities, RWA, ROA market capitalization, tier 1 capital ratio, total risk-based capital ratio, common equity to total assets, non performing assets to total assets & non performing loans to total loans.

[15] Too keep as much information as possible.

[16] All results are available upon request to the authors.

underlying indexes or tuning parameters differ. However, it is less trivial for individual and size related metrics. We eventually tested an automatic algorithm called Mean Row Moments algorithm initially developed by Deutsch and Martin (1971), but it delivered unclear results. Thus, we group the indicators in factors making the most economic sense, these choices were confirmed ex-post assessing their correlation with the factor they contribute to. In the end, the factors can be classified in three main categories: individual & systemic risks measures and factors encapsulating banks' size.

4.2 Individual risk indicators

4.2.1 Financial ratios and balance sheet information

The factor called "**RatioBS**" measures banks' exposures using balance sheet ratios, such as debt to total assets, deposit to total assets, total loan to total assets, total loan to deposit, non-performing assets to total assets, non-performing-loans to total loans & reserve for loan loss to total loans. Here, the debt corresponds to the sum of long term borrowings, short term borrowings and securities sold with repurchase agreements. Deposits are the total deposits received from customers. Non-performing assets are composed of non-accrual loans, restructured loans and foreclosed real estate. Finally, non-performing-loans are those in default or close to default.

The "**CreditBS**" factor assesses banks' credit worthiness based on balance sheet information, such as common equity to total assets, return on asset, price to book ratio, Texas ratio and Z-score. The price to book ratio corresponds to a bank's market capitalization divided by its common equity. The Texas ratio was, at the origin, developed by RBC Capital Markets' analyst Gerard Cassidy to predict banks failure during the eighties recession in the state Texas. It is computed by dividing the bank's non-performing assets by the sum of the reserve for loan loss and the total equity. The Z-score is a simple and popular risk measure of a bank's probability of insolvency. It was initially proposed by Boyd and Graham (1986); Hannan and Hanweck (1988); Boyd et al. (1993), but many improvement and adjustment were developed through time, for instance, its split form as in Goyeau and Tarazi (1992). For sake of reactivity, we use the last available return on asset (ROA) and equity over total assets, contrary to their version with rolling averages. Moreover, we collect the ROA ratio corresponding to the trailing twelve months net income over the average total assets, but we compute the ratio equity to asset with the total equities and total assets. The Z-score only requires accounting data and is a decreasing function of the probability of failure.

$$Z\text{-}score_{i,t} = \frac{ROA_{i,t} + \frac{total\ equities_{i,t}}{total\ assets_{i,t}}}{\sigma_{ROA_{i,t}}}$$

Where σ_{ROA} is the ROA's standard deviation over 4 years.

We also build the "**RegCapital**" factor highlighting the main Basel's capital ratios, like the Tier 1 capital ratio, total risk-based capital ratio and risk-weighted-asset to total assets. As usual, Tier 1 capital ratio corresponds to Tier 1 capital to risk-weighted-assets. The total risk-based capital ratio, also known as capital adequacy ratio (CAR), corresponds to the total of the Tier 1 capital ratio & the supplementary capital ratio[17]. The regulatory capital factor must be read in an inverted scale (cf. section 8, table 3).

4.2.2 Credit and solvency indicators using market information

Two factors are defined to apprehend banks' riskiness using market data. The first one focuses on banks' credit quality and the other on their individual market risk.

The most straightforward input to build the "**CreditMkt**" factor would be banks' CDS spreads, however, not all banks have actively traded CDS. To make them comparable, we use two measures coming from the capital structure optional framework initially developed by Merton (1974). The probability of default seems to be an unequivocal indicator to investigate a firm's health. At first, we compute the credit spread approximation developed by RiskMetrics Group's CreditGrades model (Finger et al., 2002), based on the probability of survival.

$$\mathbb{P}(Survival)_t = \Phi\left(-\frac{A_t}{2} + \frac{log(d)}{A_t}\right) - d \cdot \Phi\left(-\frac{A_t}{2} - \frac{log(d)}{A_t}\right)$$

$$where,$$

$$d = \frac{S_0 + \bar{L}D}{\bar{L}D} \cdot \exp\left(\lambda^2\right)$$

$$A_t^2 = \left(\sigma_{S_0} \cdot \frac{S_0}{S_0 + \bar{L}D}\right)^2 \cdot t + \lambda^2$$

---

[17] The supplementary ratio for commercial banks is perpetual preferred stock ineligible for Tier 1, perpetual debt and mandatory convertible securities, qualifying senior and subordinated debt, limited life preferred stock, qualifying allowance for credit losses. The minimum ratios set by the U.S. Federal Reserve and OTC are 8% (Tier 1 + Tier 2) for both commercial banks and savings and loans.

Where $\Phi(.)$ is the standard normal cumulative distribution function. Thus,

$$CrdGrd = \frac{-\ln(\mathbb{P}(Survival)_t) \cdot (1-R)}{t}$$

The second credit indicator called E2C (Equity-to-credit) is a credit spread approximation recently derived by Mercadier and Lardy (2019). It broadly comes from the Gauss inequality applied to a Brownian motion. These indicator's parameters are the recovery rate R, the Market-Adjusted Debt (MAD) ratio[18] and the equity volatility detailed in the CreditGrades technical report (Finger et al., 2002).

$$\begin{aligned} E2C &= (1-R) \cdot \frac{4}{9} \cdot MAD \cdot \sigma_{S_0}^2 \\ where&, \\ MAD &= \frac{\bar{L}D}{S_0 + \bar{L}D} \end{aligned}$$

Except the debt which varies on lower frequencies, these hybrid indicators require market inputs available on a daily basis. Lastly, a simple market-based leverage ratio[19] is added to this factor[20].

Another market-based factor named "**IndivMktRisk**" measures banks' riskiness through nine indicators: the volatility, four VaR, three Expected Shortfall metrics and the MZ-score. In early versions, the volatility σ_i naively corresponded to the annualized daily standard deviation over a rolling year of banks' stocks returns. However, we improve the responsiveness of our process using a volatility with a forgetting factor, computing the weighted sum of squared returns normalized by $1 - \lambda$[21]:

$$\sigma_{i,t} = \sqrt{(1-\lambda) \cdot r_{i,t}^2 + \lambda \cdot \sigma_{i,t-1}^2}$$

Additionally, three other weighted historical VaR methods are used, the composite

---

[18] Including the total debt per share outstanding, the stock price and the average recovery on the debt.

[19] Debt to enterprise value ratio.

[20] Similar results (available upon request to the authors) are obtained, adding Bharath and Shumway (2008)'s naïve Merton probability of default to the "CreditMkt" factor. However, this indicator is not kept in the final version as it does not bring further insight.

[21] The decay factor λ is set at 0.98 and the initial annualized $\sigma_{i,0}$ at 0.29 for the banks (at 0.17 for the indexes), corresponding to the median of the standard deviation rolling yearly over the 2004-20 period.

VaR (Mercadier et al., in progress) that combines the Laplace and the Normal distributions, the HW VaR (Hull and White, 1998) and the BRW VaR (Boudoukh et al., 1998). The "VaR measure is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence" (Jorion, 2001). In the context of a short time-horizon, in which losses X are generally assumed as $E[X] = 0$, we write the VaR of X at confidence level α:

$$\mathbb{P}(X \geq VaR_{\alpha}(X)) = \alpha$$

In our methodology, the Value-at-Risk is the 5%-quantile of the daily returns over one year multiplied by -1 to be expressed in term of losses. In the literature, the VaR has been criticized for not properly allowing for the potential severity of the risk on a portfolio, notably linked with its lack of subadditivity (Embrechts, 2000; Acerbi and Tasche, 2002). Thus, the Expected Shortfall has become increasingly popular as an alternative risk measure. It can be defined, at confidence level α, as the expected value of loss X, conditional on the loss exceeding the Value-at-Risk:

$$ES_{\alpha}(X) = \frac{1}{\alpha} \int_{0}^{\alpha} VaR_{u}(X)\, du$$

An extension from the BRW VaR (Boudoukh et al., 1998), averaging the values beyond this point, is derived to evaluate the BRW expected shortfall. And we compute two additional Expected Shortfall measures, both based on the tail average over one year of the daily returns below the 5%-quantile and below $\sigma_S \cdot \Phi^{-1}(\alpha)$, respectively called "ES" and "ESn", where $\Phi^{-1}(.)$ is the inverse of the standard normal cumulative distribution function. Like for the Value-at-Risk, the results are expressed as losses with -1 product. The last indicator specifically designed for the banking system is derived from the Z-score introduced in section 4.2.1. The Market Z-score (Lepetit et al., 2008; Saghi-Zedek and Tarazi, 2015), is a market-based indicator updated daily. It is a decreasing function of the probability of failure that only requires stocks returns average $\bar{r}$ and standard deviations σ. Empirically, both variables are computed over the last year and expressed on a quarterly basis.

$$MZ\text{-}score = \frac{1 + \bar{r}_{i,t}}{\sigma_{i,t}}$$

4.3 Systemic risk indicators

4.3.1 Specificities

In the empirical literature (Idier et al., 2014; Weiß et al., 2014; Reboredo and Ugolini, 2015), systemic metrics generally rest upon national indexes. While being aware of major international shocks, national indexes might be impacted by more minor local events. Thus, we partly offset these idiosyncratic effects adding global and global banking indexes to the systemic factors. In fact, we define banks' systemic risk indicators with their respective national (cf. section 8), global (MSCI World) and international bank sector (MSCI World Bank) indexes. Moreover, we compute multiple systemic indicators to filter contradicting assessments as discussed by Kleinow et al. (2017).

4.3.2 Correlations and betas

"**BetaCorrel**" quantifies basic relations between a bank and its corresponding national indexes and with the MSCI World & World Bank indexes. In line with the volatility computations, a factor decay is also applied to the covariances computing the weighted sum of the cross products the returns normalized by $1 - \lambda$[22]:

$$cov(r_{i,t}, r_{m,t}) = (1 - \lambda) \cdot r_{i,t} \cdot r_{m,t} + \lambda \cdot cov(r_{i,t-1}, r_{m,t-1})$$

Firstly, two metrics are computed, Pearson correlation and beta, as usual expressed as follows: $\beta_{i,t} = cov(r_{i,t}, r_{m,t})/\sigma^2_{m,t}$. Another indicator called tail beta (Straetmans et al., 2008), a systemic measure, is calculated as banks' quantile regressions onto their corresponding national and global indexes[23]. All these indicators are computed from daily returns over a yearly rolling window.

4.3.3 Bank losses when markets are in distress

The factor called "**SysBkMkt**" appraises the impact on a specific bank when the market is in distress. It is built on eighteen measures, three MES and LRMES measures based on the quantile tail average and three MESBeta and LRMESBeta ones based on the beta between the bank and the market indexes[24]. And on three BRW MES and LRMES, extending the BRW expected shortfall (Boudoukh et al., 1998). The Marginal Expected Shortfall (MES) is a well-known systemic risk measure. It is defined by Acharya et al. (2012, 2017) as the marginal contribution of firm i to systemic risk, as measured by the Expected Shortfall of the financial system and adjusted as follows for our methodology. It corresponds to the negative value of the one day loss expected for a bank ($r_{i,t}$) if market returns ($r_{m,t}$) are less than the five percent quartile of market returns over the last year.

$$MES_{i,t} = -E[r_{i,t} \mid r_{m,t} < \mathcal{Q}_{5\%}(r_{m,t})]$$

---

[22] The decay factor λ is set at 0.98 and the initial covariance, $cov(r_{i,0}, r_{m,0})$, at $1.3 \cdot 10^{-4}$ matching the median of the covariance rolling yearly between the banks and the indexes over the 2004-20 period.

[23] Betas and tail beta are computed with a constant.

[24] The three measures are respectively against the national (cf. section 8), global and international bank sector indexes.

Benoit et al. (2017) show that "the MES of a given financial institution i is proportional to its systematic risk, as measured by its time-varying beta"; in particular, they show that one can write[25]:

$$MESBeta_{i,t} = -\beta_{it} \cdot ES(r_{m,t}) \tag{1}$$

where β_i is the beta of firm i (cf. section 4.3.2), and the Expected Shortfall corresponds to the tail average of the index daily returns below the 5%-quantile over one year. A long term adjustment of the MES by Acharya et al. (2012) is called the Long Run Marginal Expected Shortfall (LRMES). More precisely, it is defined as the firm's expected drop in equity value if the market falls by more than a given threshold within the next 6 months, it is computed as follows:

$$LRMES_{i,t} = 1 - e^{-18 \cdot MES_{i,t}}$$

And another similar measure based on Engle's V-Lab[26] uses the beta.

$$LRMESBeta_{i,t} = 1 - e^{\ln(1-d) \cdot \beta_{i,t}}$$

where $d = 0.4$ corresponds to the six-month crisis threshold (i.e. a 40% drop) for the market index decline.

4.3.4 Bank losses in US dollars when markets are in distress

As explained later (cf. section 4.4), banks' size matters. Thus, we add an "**SRISK**", a factor similarly estimating distressed markets' impact on banks but expressed in US dollars rather than in percent. Nine SRISK metrics relying either on the LRMES, its BRW version (Boudoukh et al., 1998) or on the LRMESBeta, composed this factor. The SRISK measure, introduced by Acharya et al. (2012) and Brownlees and Engle (2016), is defined as "the expected capital shortfall of a financial entity conditional on a prolonged market decline". It represents an extension of the MES, additionally embodying the liabilities and size[27] of the financial institution. Acharya et al. (2012) define it as:

$$SRISK_t(X_i) = E[k(Debt + Equity) - Equity \mid Crisis]$$

$$= kD_{it} - (1-k)[1 - LRMES_t(X_i)]E_{it}$$

with D_{it} the book value of debt, E_{it} the market value of equity and k the prudential capital ratio (commonly set at 8% and at 5.5% for Europe due to differences in accounting standards, cf. V-Lab).

---

[25] The proof of Equation (1) is given in Benoit et al. (2013, Appendix A) and Brownlees and Engle (2016).

[26] https://vlab.stern.nyu.edu/en/

[27] The banks' market capitalization.

4.3.5 Market losses when a bank is in distress

The "**SysMktBk**" factor assesses the opposite point of view measuring the impact by the market conditionally on a stress event for the bank, potentially including the contribution of the bank to the systemic risk. First, we calculate the symmetric measure to the MES and to its BRW version (Boudoukh et al., 1998), respectively called MESInv and MESBRWInv, where:

$$MESInv_{i,t} = -E[r_{m,t} \mid r_{i,t} < \mathcal{Q}_{5\%}(r_{i,t})]$$

In addition, we use the Benoit et al. (2013)'s version of the Delta Conditional Value-at-Risk (ΔCoVaR), originally introduced by Adrian and Brunnermeier (2016). According to these authors, the CoVaR corresponds to the VaR of the financial system conditionally on a specific event affecting a given firm. The ΔCoVaR is the difference between its CoVaR when the firm is, or is not, in financial distress.

$$\Delta CoVaR_{i,t}(\alpha) = -\gamma_{i,t} \cdot [VaR_{i,t}(\alpha) - VaR_{i,t}(50\%)]$$

$$\gamma_{i,t} = \rho_{i.m,t} \cdot \frac{\sigma_{m,t}}{\sigma_{i,t}}$$

Where $\alpha = 1\%$[28].
Finally, we also compute a more reactive ΔCoVaR, called ΔCoVaRComp, that relies on the composite VaR.

4.4 Size Indicators

Our project aims to deliver a global ranking of banks based on both individual and systemic risks. But the impact of bank size regarding both aspects of the risks is ambiguous. On one hand, large banks are related to significantly higher systemic risk (deJonghe, 2010; Banbula and Iwanicz-Drozdowska, 2016; Kleinow et al., 2017) but on the other hand the size may have a positive impact on a stand-alone basis. In addition, the role of the size as a major determinant of the systemic risk is unclear, sometimes accused to cloud other factors contributing to banks' systemic importance (López-Espinosa et al., 2012, 2013; Weiß et al., 2014; Varotto and Zhao, 2018; Elyasiani and Jia, 2019). Nevertheless, it is difficult to assess the additional stability offered by big banks through market-based indicators such as volatility.

Thus, we added the factor called "**Size**" regrouping fundamental data from the banks' balance sheets and the market capitalization. The fundamental features are amounts in US dollars related to the banks' balance sheet, such as common equities, total equities (common equities, minority interests and preferred equities), total

---

[28] $\alpha = 5\%$ for all other indicators.

assets, total deposits received from customers, reserves for loan losses, the total debt (cf. section 4.2.1), risk weighted assets and total loans including commercial customer & other loans. As balance sheet and market size indicators are a sign of stability, all the features are multiplied by -1 to be increasing functions of the risk[29]. The size factor must be read in an inverted scale (cf. section 8, table 3).

The stability dimension is also taken into account in "**LiquidityMkt**" which is composed of the rolling average[30] of the stock price bid-ask spreads over five days taken as percentage of the mid-price. The two others constituents of this factor are the daily volume divided by the number of shares and the daily volume multiplied by the stock price. The liquidity factor must be read in an inverted scale (cf. section 8, table 3), thus as an illiquidity one.

One could argue that this number of factors is arbitrary and that it could be reduced. However, the correlation among factors remains diversified[31], shows a lack of redundancy and as explained above, this distribution makes economic sense considering our objective.

5 Dimensionality reduction and clustering

5.1 Features engineering and dimensionality reduction

As already mentioned in section 4, a second phase of features engineering is performed on the computed risk indicators. In fact, we found remaining outliers in the indicators' values, originating from the automatically extracted raw data. Therefore, the indicators are cleaned with the following automatic quantile method:

$$f(x) = \begin{cases} x & \text{if } x \in [Q_1(x) - 15 \cdot (Q_2(x) - Q_1(x)),\ Q_3(x) + 15 \cdot (Q_3(x) - Q_2(x))], \\ NA & \text{otherwise.} \end{cases}$$

Our final target being to assess banks' risks, the indicators are "oriented" according to this point of view. In other words, when an indicator's value increases, the corresponding bank should be characterized as riskier. Therefore, the indicators that are not increasing functions of the risk are multiplied by minus one and notified with (-) sign (cf. section 8, table 3).

The log-modulus transformation, developed by (John and Draper, 1980), is applied to all the risk indicators. It contributes via the logarithm function to pull in the distribution's tails and via the modulus transformation to make the shape

---

[29] Further explanation available in section 5.1.

[30] A trading day contributes to the output if there are at least ten valid bid-ask spread points on that day. This field returns a value if at least three trading days are eligible to contribute.

[31] Only 4 out of 55 pairwise correlations are above 70%.

more Normal and is expressed as follows:

$$LnMod(x) = sign(x) \times \ln(\mid x \mid + 1)$$

Finally, the indicators are standardized, subtracting the mean and dividing by the standard deviation[32].

Once the data are cleaned, oriented as increasing functions of the risk, transformed via log-modulus and standardized, we group the indicators of a specific factor into a dedicated subsample. In this subset, we remove all rows only having missing values (nan), then we partly fill the remaining nan. At first, we apply a fill forward method until the next available data and then, we do a backward filling only over a maximum of 12 weeks.

The eleven factors are then derived from the individual, systemic and size indicators using a famous dimensionality reduction algorithm, the principal component analysis (Pearson, 1901). Firstly, we compute the eigen values and eigen vectors from their correlation matrix. Then, we multiply the first eigen vector (a.k.a. principal component) – corresponding to highest eigen value, with the standardized matrix to project this matrix onto the first factor. Keeping the orientation as an increasing function of the risk, we change the output vectors' signs, if needed, according to their correlation with the sums of input matrices' rows and normalize the results by dividing by the square root of the eigen value.

All indicators and factors explained variances are summarized in table 3 (section 8). These eleven factors are then set as input features for the k-means algorithm.

5.2 Adjusted nested k-means clustering

Considering our objective, being able to correctly identify past bank failures would have been very convenient to label our data. However, there is, for instance, still debate on whether all US firms benefiting from the Troubled Asset Relief Program (TARP) of the United States government, would have defaulted otherwise. Thus, to avoid being biased by such unreliable information, we do not supervise the learning. Instead, we choose an unsupervised learning algorithm, the k-means clustering (Steinhaus, 1956; MacQueen, 1967; Lloyd, 1982), which has in our perspective some good properties. This method is transparent and simple while being widely used because of its convenient speed. Other more complex algorithms – e.g. derived from neural networks, are currently under critics because even though their results are impressive, they are inscrutable.

---

[32] A Uscore was also performed: $\frac{x - 0.5 \cdot (\max(x) + \min(x))}{0.5 \cdot (\max(x) - \min(x))}$, but ended up inconclusive. In fact, standardization with mean and standard deviation is more consistent with the k-means Euclidean norm criteria.

The k-means algorithm clusters the input data set through a fixed a priori number of clusters, noted k. The biggest issues are precisely related to the choice of the hyper-parameter k, along with the choice of the centroids at initiation which might lead to different inefficient local optimums. The number of cluster is determined according to the well-known elbow curve method. It consists in running the k-means algorithm on the input matrix for a range of number of clusters k is $\{1, 2, ..., K\}$. For each k, the inertia, measuring the distance between the observations and their clusters centers, is computed.

$$Inertia = \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where c_j corresponds to a points closest centroid and j is $\{1, 2, ..., k\}$. Another metric, the silhouette criterion provides a measure of how well each point lies within a cluster (Rousseeuw, 1987). Defined as one minus the average distance from their respective closest centroid c_j over the same points distance from their second closest one c'_j, where j is $\{1, 2, ..., k\}$:

$$Silhouette = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\left\| x_i^{(j)} - c_j \right\|^2}{\left\| x_i^{(j)} - c'_j \right\|^2}$$

The choice of the centroids at initiation is less trivial. For instance, de Oliveira and do Carmo Nicoletti (2019) compares five initialization algorithms. Notably, they find interesting the k-means++ improvement through the Single Pass Seed Selection (SPSS) introduced by Pavan et al. (2011). The k-means++, developed by Arthur and Vassilvitskii (2007), is an initialization algorithm that is in general quickest and more robust than randomly selecting k points as initial clusters that might lead to inefficient local optima. Broadly speaking, once the first centroid is randomly determined as one point of the data set, the next ones are more likely to be chosen as their distances from the previous chosen centroids are large. The SPSS initiation methods differs for the choice of the first centroid, instead of being randomly chosen, it corresponds to the highest density point. However, we do not apply the SPSS because computing all distances on a huge data set as ours is slow. Preferring the k-means++, we might end up with various non obviously optimal outputs, thus we launch the process 60 times – which is still far quicker than the SPSS version. For sake of consistency, the first point is chosen randomly for some runs and is fixed for others. For instance, the initial centroid can be the closest or farthest point from the data set's average, median, min, max, first or third quantile. Then, alternatively the four other points are either the farthest from each other or chosen randomly.

Once the initialization centroids are determined, it goes back to a standard k-means clustering, associating each point to the nearest centroid, creating a first set of clusters. Then k new centroids are computed as barycenters of those clusters and each point is associated to the nearest new centroid, once this loop starts, it lasts until the centroids' locations do not change anymore. For n observations and k clusters, this algorithm aims at minimizing the following objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \cdot \mathbf{1}_{x_i^{(j)}}$$

where $\left\| x_i^{(j)} - c_j \right\|$ indicates the distance between a data point and its respective cluster centers, $x_i, c_j \in \mathcal{R}^d$ and $\mathbf{1}_{x_i^{(j)}}$ equals 1 if x_i belongs to cluster j and 0 otherwise. Please find below a quick pseudocode explaining our adjusted version of the k-means algorithm.

```
1. Initialize clusters centroids $c_1, c_2, ..., c_k \in \mathcal{R}^d$.
2. Repeat {
     Assign each observation to the cluster which has the closest centroid.
     For every i, $x_i^{(j)}$ is attached to cluster $j$ that minimizes $\|x_i - c_j\|^2$.

     Compute the new mean (=centroid) for each cluster.
     For every j, set
     $c_j := \frac{\sum_{i=1}^{n} x_i^{(j)} \cdot \mathbf{1}_{x_i^{(j)}}}{\sum_{i=1}^{n} \mathbf{1}_{x_i^{(j)}}}$
   }
   Until convergence
```

As explained before this algorithm is run 60 times and the chosen version is the one with the lowest inertia. To provide consistent labels to the clusters, we sum all individual risks factors' coordinates[33] (as they are oriented as increasing function of the centroids, cf. section 5.1) and set as 1 the higher number hence riskier and as k the less risky.

The main issue in our case is that our sample has missing data. Therefore, we developed our own k-means program to address this issue. The original k-means

---

[33] The factors assessing the size or systemic dimensions are removed from the sum.

algorithm does not handle missing data, thus machine learning professionals must choose between loosing information through the removal of observations having missing data or filling the missing points, a decision involving strong assumptions about the missingness patterns. However a method relying on the majorization-minimization algorithm called k-pod was developed to handle missing data by Chi et al. (2016). An R version[34] of their algorithm is publicly availabe, while we found very close results with it, the time of computations for a large[35] table is around 900 times slower.

6 Applied clustering analysis and empirical illustration

6.1 Number of clusters choice

Our data set is built on 161 international listed banks from 31 countries, among the biggest in term of total assets based on Barth and Wihlborg (2016, 2017), more details are available in section 8. The study spans from January 2nd, 2004 to September 24th, 2021, including both crisis & post-crisis periods and the raw sample is composed of balance sheet and market data.

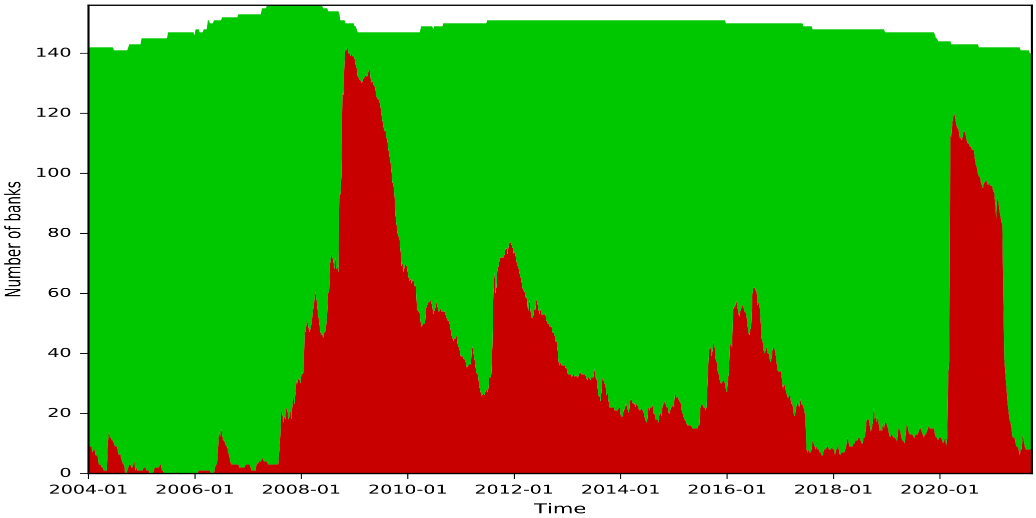
In the final part of our methodology, the factors are inputted in an adjusted nested version of the k-means algorithm. Firstly as explained in section 5.2, the number of clusters must be determined. From the inertias plotted on an arm shaped line chart, the chosen value located in the elbow corresponds to k=2, where the major breakpoint happens. A choice confirmed by the silhouette criterion[36].

---

[34] See e.g. the R's package 'kpodclustr' by Chi and Chi (2015).

[35] $10^5 \times 11$.

[36] Both elbow curves respectively based on the inertia and the silhouette criteria are available upon request to the authors.



Notes. This graph shows how banks are clustered through time. The results notably highlight crisis periods with higher number of banks within the riskiest cluster (red).

Figure 1: Clustering timeline (k=2)

Running a two clusters k-means algorithm as suggested by these metrics, we end up with an historical timeline (cf. fig. 1) highlighting four periods globally riskier than the others. The first three points in time match those highlighted by the probability of crisis graph from the recent paper by Engle and Ruan (2019). The first period, the riskier one, corresponds to the 2008 GFC originating in the United States and the next ones are respectively linked to the 2012 European sovereign debt crisis and the 2016 Asian debt crisis. The last turbulent period depicts the current COVID-19 pandemic. While it seems to efficiently split, for a given date, the riskiest banks (in red, bottom) from the safest ones (in green, top), it is sensible to propose a more detailed analysis.

A higher granularity can easily be reached launching a k-means algorithm in each cluster – a concept called nested k-means (cf. Niedzielski et al. (2017)). As usual, we start by finding the two corresponding values of k based on the inertia and silhouette methods. According to the breakpoints of the corresponding elbow curves[37], we respectively split the riskier cluster in two subclusters and the safest one in three, which leads to an overall k=5, but more explicitly written k=2→2|3[38]. A reasonable choice according to another metric determining the number of clusters known as the score function (SF), developed by Saitta et al. (2008)[39]. Beyond being in line with the elbow curves, the nested algorithm provides steadier results than

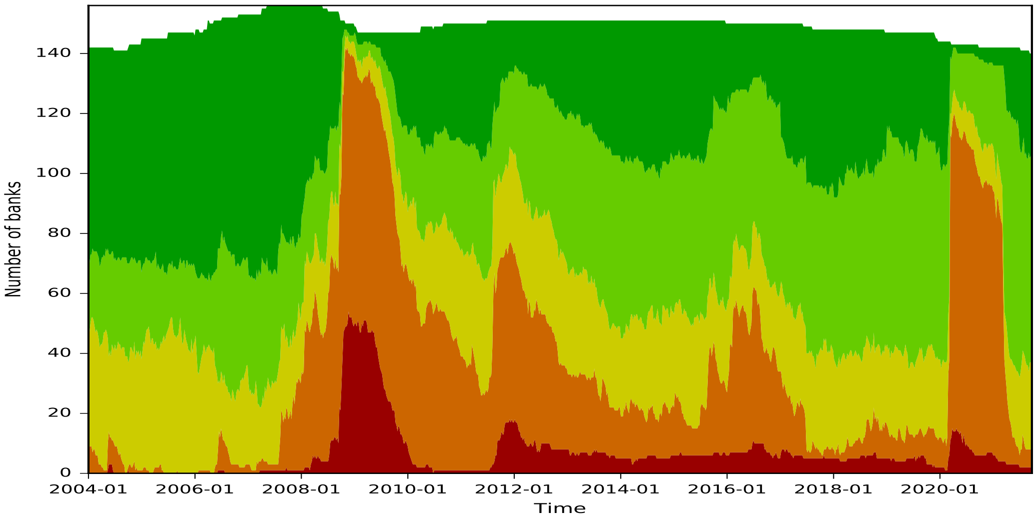
---

[37] Available upon request to the authors.

[38] The nested k-means requires 2 clusters at first and then respectively in 2 and 3 clusters.

[39] Available upon request to the authors.

the standard k=5 version[40] when a factor is removed.



Notes. This graph shows how banks are clustered through time with a higher granularity than in fig. 1. The results still highlight crisis periods but with a consistent lower number of banks within the riskiest cluster (red, bottom).

Figure 2: Clustering timeline (k=2→2|3)

By construction, as it is generated by a nested algorithm, the same overall shape is observed whether using k=2 or k=2→2|3[41]. However, it is more shaded for k=2→2|3, where the red and orange clusters split the previous riskier cluster and the other colors divide the safest one. According to fig. 2, the increase in granularity outputs a reassuring lower number of extremely risky banks even during the GFC compared to fig. 1, although higher than in other periods. Additionally, the European crisis seems to have reached its maximum in 2012, which is consistent with the reality. Comparing the two riskiest clusters, our methodology distinguishes a more severe impact on the banking industry of the GFC –a banking crisis –from the current COVID-19 pandemic –not directly related to the banking sector. More specifically, the banking sector, although still under pressure, was quickly supported by massive support programs launched worldwide by central banks, e.g. the Fed's monetary policy measures or the ECB's Pandemic Emergency Purchase Programme (Rizwan et al., 2020; Borri and di Giorgio, 2021).

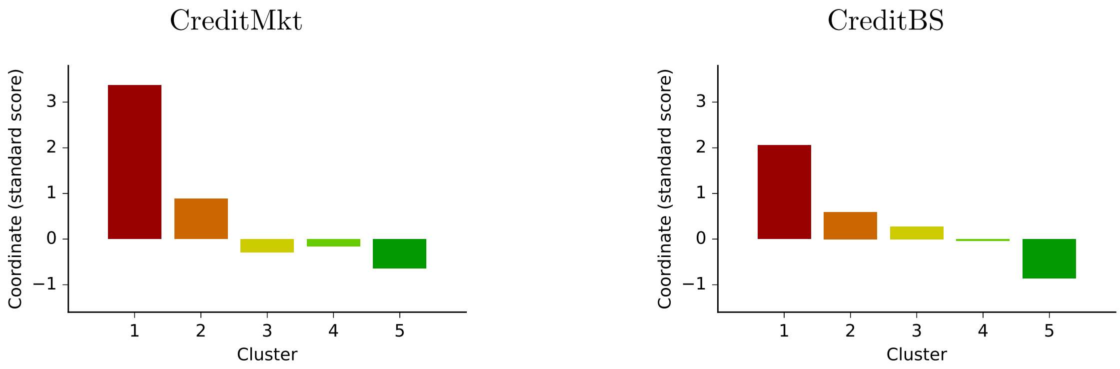
---

[40] The standard method only provides a non-significant gain in the inertia.

[41] Adding a final step to move the points back to the closest centroids gives similar results and only around 2% of the points are concerned.

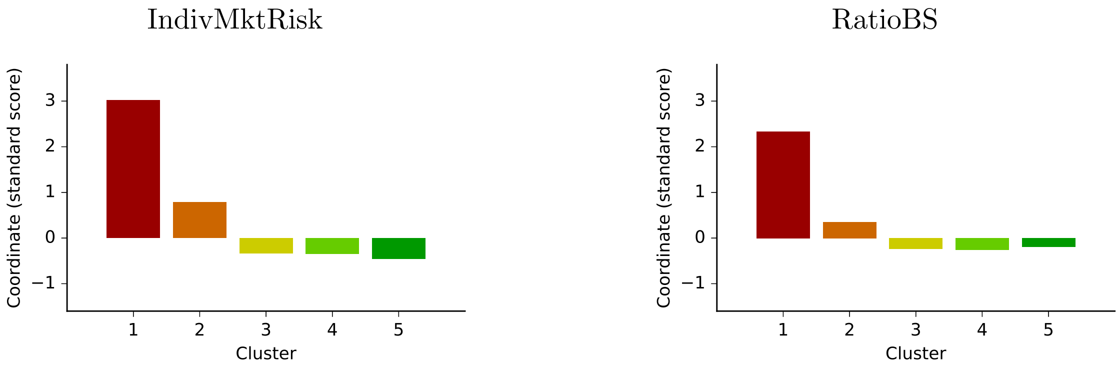
6.2 Centroids' analysis

Beyond the previously introduced historical timeline, our automatic process provides a weekly update of banks' clustering. However, from an economic point of view, the analysis of the centroids matters. As introduced in the end of section 5.2, the k-means algorithm only creates k classes regardless of their meaning. Thus, we set the clusters order in function of the individual risk factors, as they are derived from the most simple and transparent indicators.



Notes. These bar charts display the standardized coordinates of the clusters' centroids ordered from the riskiest (red, lhs) to the safest (green, rhs) per credit related factors. The results notably emphasize the riskiest banks' contribution to the first cluster.

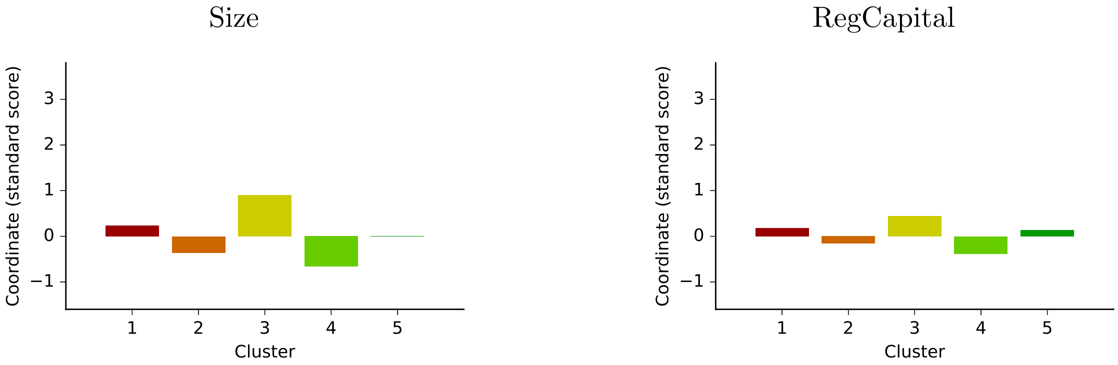
Figure 3: Credit related individual risks centroids



Notes. These bar charts display the standardized coordinates of the clusters' centroids ordered from the riskiest (red, lhs) to the safest (green, rhs) per individual risk factors. The results notably emphasize the riskiest banks' contribution to the first cluster.

Figure 4: Other individual risks centroids

On figs. 3 and 4, the two worst clusters stick out, and above all emphasizing the riskiest banks' contribution to the first one. In addition, the most solvent institutions stand out in the fifth group, except for the balance sheet ratios factor. However there is no obvious conclusion regarding the third and fourth clusters. The next paragraph is focused on the "Size" factor, as it is linked to the remaining analyses.



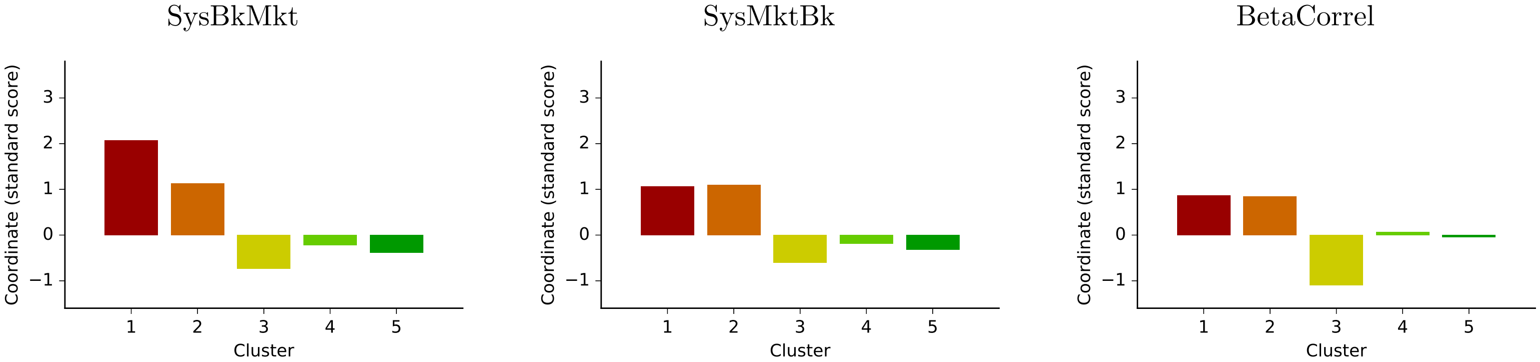
Notes. These bar charts display the standardized coordinates of the clusters' centroids ordered from the riskiest (red, lhs) to the safest (green, rhs) per size and regulatory capital factors (inverted scales). Above the fact that the smallest and biggest firms are spread within the middle clusters, the results underline that the size does not impact the most extreme clusters.

Figure 5: Size centroids

On fig. 5 (left), the smallest firms are concentrated in the third cluster[42] and the biggest ones in the second but mostly in the fourth one. This is an argument in favor of a certain stability brought by big banks. Above all, the riskiest and safest clusters are not impacted by the size, an argument supporting the exclusion of banks' size from the major determinants of their risks, in line with Varotto and Zhao (2018). According to the scale, the regulatory capital factor has a limited impact on the risk and groups banks like the size factor.

---

[42] Let us remember that the indicators composing the size factors are multiplied by -1.



Notes. These bar charts display the standardized coordinates of the clusters' centroids ordered from the riskiest (red, lhs) to the safest (green, rhs) per systemic factors. Beyond the beta's influence clustering the systemically safest banks in the middle cluster, the results underline the riskiest banks' contribution to the first two clusters.

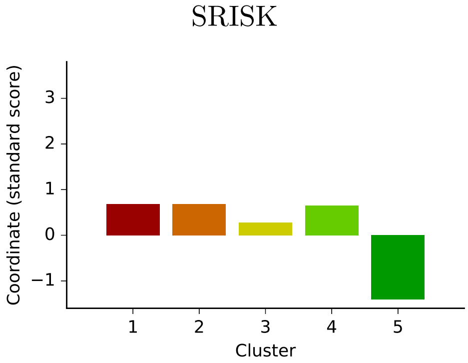
Figure 6: Systemic risks centroids

While the two worst clusters still emerge for systemic measures[43], the less systemically risky banks belong to the third cluster and not anymore to the fifth. According to the importance of "BetaCorrel" and from a detailed analysis using the algorithm directly on the 85 indicators, the systemic factors' worst third cluster seems to be particularly related to systematic (beta-based) indicators. By extrapolating from the previous paragraph's size analysis, small banks appear to have, as expected, less relations with the market – through lower betas. A phenomenon also visible, to a lower extent in the third group, for the other systemic factor, the SRISK. But highlighting that bank sizes and betas are not sufficient determinants of their overall riskiness[44].

---

[43] A difference between the worst two clusters of the "SysBkMkt" & "SysMktBk" factors (fig. 6) leads us to emphasize that the large banks from the second riskiest cluster seem to contribute the most to the systemic risk. A phenomenon accentuated by the addition of the COVID-19 pandemic data (in line with Borri and di Giorgio (2021)).

[44] As they do not belong to the safest cluster.



Notes. This bar chart displays the standardized coordinates of the clusters' centroids ordered from the riskiest (red, lhs) to the safest (green, rhs) per SRISK factor. The result notably highlights the safety brought by low SRISK.

Figure 7: SRISK centroids

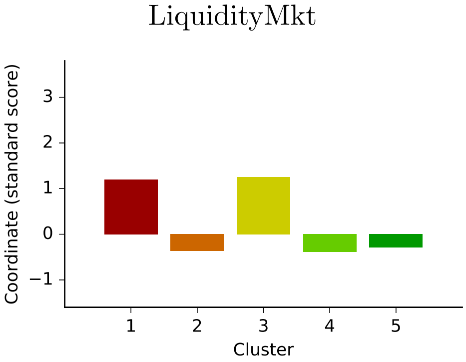
The SRISK is a systemic measure, expressed in US dollars, having common components with the size factor, namely the debt and the market capitalization. The noteworthy contribution here is the safety brought by banks having a low SRISK[45]. While this factor is globally well ordered, the biggest banks from the fourth cluster seem to provide relatively high values. Leading us to the same conclusion as in Varotto and Zhao (2018)'s study which points out that systemic indicators primarily driven by firm size lead to an overriding concern for big firms. In their paper, they fix the size effect with a standardized measure called sSRISK dividing the SRISK[46] by the Total Assets. While the overall results are close, the sSRISK actually matches the individual risk factors distribution[47]. However, our methodology still deals with the original SRISK indicators as we wish to limit the number of assumptions and work with widely tested measures. The size effect is handled by the addition of size factors.

---

[45] Possibly a negative one.

[46] Floored at 0.

[47] According to results available upon request to the authors.



Notes. This bar chart displays the standardized coordinates of the clusters' centroids ordered from the riskiest (red) to the safest (green) per liquidity factor (inverted scale, illiquidity). The result notably emphasizes that small and highly risky banks tend to be illiquid.

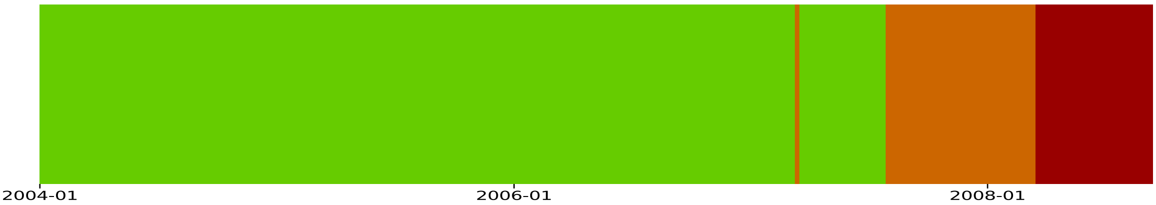
Figure 8: Market liquidity centroids

Lastly, the illiquid banks have more odds to belong either to the riskiest cluster or the third one, grouping small banks according to the size analysis. As expected, small banks and extremely risky ones tend to be illiquid.

Aware of our data sample and methodology limits, we conclude in saying that MES, ΔCoVaR and Tail Beta measures tend to overstate small banks as the safest ones by default. The SRISK has rather a tendency to overstate big institutions as riskier.

6.3 Illustrations with major American banks

As an illustration, we graphically compare two major American banks, the bankrupted Lehman Brothers and JPMorgan Chase, with their respective five year Credit Default Swaps (CDS)[48].



Notes. This graph presents in which cluster Lehman Brothers belongs through time. It notably highlights its presence in the worst cluster as it gets closer to bankruptcy.

Figure 9: Lehman Brothers Output

---

[48] In our methodology, traded 5y CDS are not set as inputs.

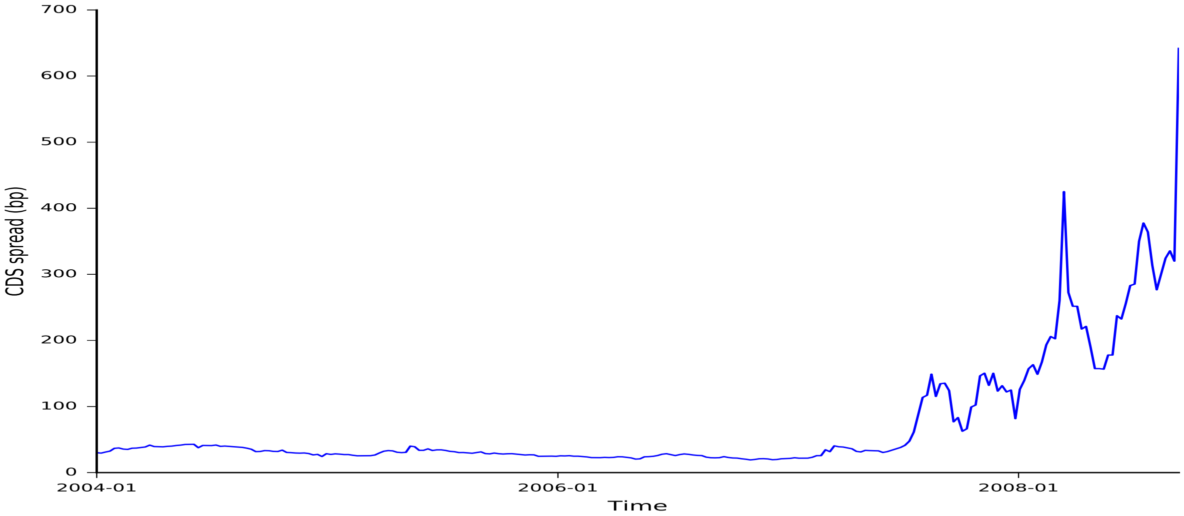
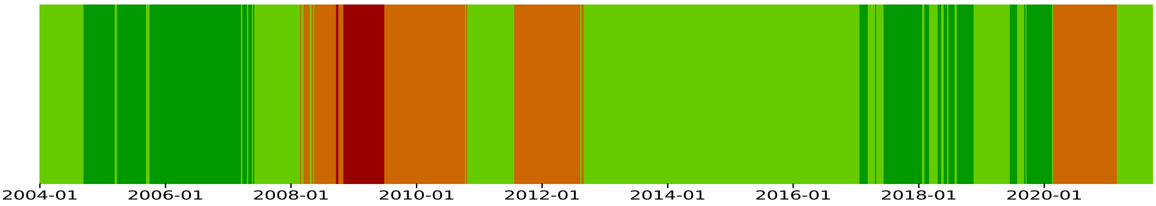


Figure 10: Lehman Brothers 5y CDS

Lehman Brothers' collapsing period closely linked to the subprime mortgage crisis clearly appears on both graphs. It highlights the path to bankruptcy in 2008.



Notes. This graph presents in which cluster JPMorgan Chase belongs through time. It notably highlights its presence in the worst cluster during the GFC.

Figure 11: JPMorgan Chase Output

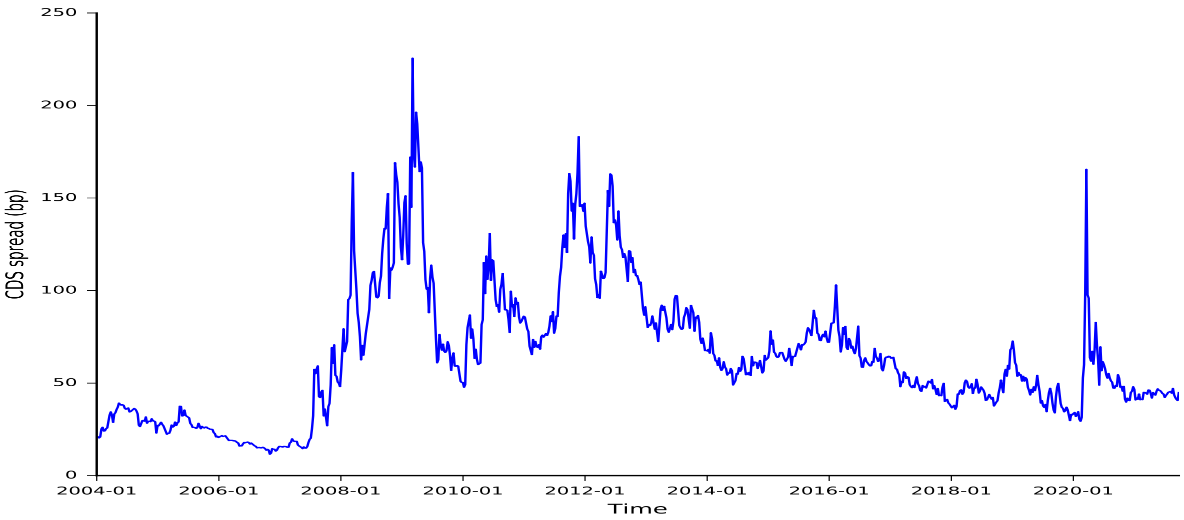


Figure 12: JPMorgan Chase 5y CDS

JPMorgan Chase reached its riskiest moments right after the Lehman Brothers failure, it is marked by both the red period from our clustering, matching the 5y CDS highest values. Other less risky periods arise around 2012 and 2020, corresponding as well to the derivative values.

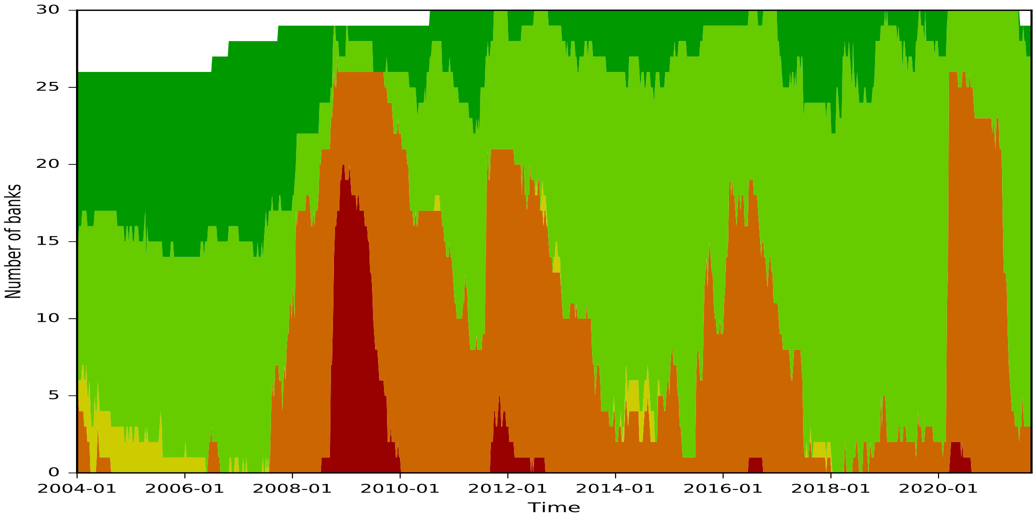
Finally, our methodology provides an aggregated version of the clustering per country and per region (see the United States example in section 8). To do so, weighted averages of banks' cluster numbers are computed in terms of total assets per country and region. The website (cf. section 1, footnote 3) disseminates the results that are updated every week with a dedicated detailed report available for authorized users. Beyond the historical graph (cf. fig. 2) and each bank, country and region's time frame (as above), the report includes an exhaustive listing of the banks per cluster, as of the most recent date.

6.4 Focus on G-SIBs

Although more than 60% of the Global Systemically Important Banks (G-SIBs[49]) were in the worst cluster during the GFC, they were less than 20% during the European debt crisis and less than 10% in the recent COVID-19 pandemic (cf. fig. 13). It seems that financial constraints established by regulators work fine. But this argument can be challenged by the fact that the number of banks in the first two clusters remains stable during crisis periods. A phenomenon likely related to the banks' size.

---

[49] Composed of thirty banks, chosen from the November 11th, 2020 list of G-SIBs `https://www.fsb.org/2020/11/2020-list-of-global-systemically-important-banks-g-sibs/`. Groupe BPCE not being listed, Natixis is set as proxy.



Notes. This graph shows how Global Systemically Important Banks are clustered through time. The results notably highlight crisis periods with higher number of banks within the riskiest cluster (red).

Figure 13: G-SIBs' clustering timeline (k=2→2|3)

6.5 Clusters' behavior

Finally, we draw conclusions from the overall behavior of the clusters. First, table 1 summarizes the average of consecutive weeks banks spend in a specific cluster.

Cluster	1	2	3	4	5
Number of weeks mean (SD)	30.1 (58)	23.9 (34.2)	25 (75.5)	21.3 (43.7)	45.3 (83.7)

Table 1: Average number of consecutive weeks spent by banks in each cluster

Except for the two extreme clusters – in which banks seems to remain for longer time periods – banks generally spend on average around six months in a specific cluster.

Additionally, one can be interested in the probability a bank remains in the same cluster or transfers to another one. Thus, table 2 is the weekly transition matrix of a Markov chain for all banks over the entire time length.

Cluster	1	2	3	4	5
1	96.884%	3.004%	0.112%	0.000%	0.000%
2	0.713%	95.842%	0.478%	2.680%	0.287%
3	0.043%	0.553%	96.114%	2.295%	0.995%
4	0.003%	1.721%	1.581%	95.469%	1.226%
5	0.005%	0.337%	0.613%	1.166%	97.878%

Notes. $\mathbb{P}_{i,j}$ is the probability of transitioning from i to j.

Table 2: Clusters' weekly transition matrix

It is extremely rare that a bank is transferred by more than two clusters (i.e. 4 to 1 or 2 to 5, etc.), and the more likely transitions are either by one cluster or, for a large bank between clusters 2 and 4. Finally, small banks in the third cluster are more likely to become safer than riskier.

7 Conclusion

In this paper we provide an automatically updated decision support tool clustering listed banks in function of representative risk factors. Many financial ratios, stand-alone and systemic risk indicators, combining balance sheet and market information, are computed and summarized into factors. A special attention is paid to the ambiguous impact of bank size on systemic risk measures. Notably, the size does not impact the two most extreme clusters, partially supporting Varotto and Zhao (2018)'s argument of size not being a persistent determinant of systemic risk. The entire methodology is detailed including the dynamic data extraction, cleaning procedures, indicators and factors calculations and the algorithms choice and setting. Notably, we picked a nested k-means algorithm adjusted to handle missing values. The resulting bank clustering is also aggregated per country and region to identify zones of fragility. Inspired by the V-Lab, our results are updated weekly, refreshable on-demand, on a dedicated website. Being easily editable, via the addition or removal of banks, indicators or factors[50], our dynamic methodology can be set for specific purposes. Additionally, we emphasize that our outputs distinguish a more severe impact on the banking industry of the GFC compared to the current COVID-19 pandemic.

Several improvements can be imagined for such a risk clustering of banks. For low explained variance, more than one principal component could be dealt with. Furthermore, other clustering algorithms could be tested such as a fuzzy k-means

---

[50] E.g. additional liquidity measures (Amihud, 2002; Goyenko et al., 2009) or operational risk indicators (Sanford and Moosa, 2015; Andersen et al., 2016)

(Dunn, 1973; Bezdek, 1981), which gives the probability for each observation to belong to the various clusters[51].

References

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487–1503.

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

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

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

Altman, E. L. (1968). Financial ratios, discriminant analysis and the prediciton of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5:31–56.

Andersen, L. B., Häger, D., and Vormeland, H. B. (2016). Causal analysis of operational risk for deriving effective key risk indicators. *Journal of Risk Management in Financial Institutions*, 9(3):289–304.

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. pp. 1027-1035.

Banbula, P. and Iwanicz-Drozdowska, M. (2016). The systemic importance of banks - name and shame seems to work. *Finance Research Letters*, 18:297–301.

Barth, J. R. and Wihlborg, C. (2016). Too big to fail and too big to save: Dilemmas for banking reform. *National Institute Economic Review*, 235(1):27–39.

Barth, J. R. and Wihlborg, C. (2017). Too big to fail: Measures, remedies, and consequences for efficiency and stability. *Financial Markets, Institutions & Instruments*, 26(4):175–245.

---

[51] The differences and advantages of fuzzy clustering over standard k-means are discussed in various recent papers (Heil et al., 2019; Kindhi et al., 2019; González-Fernández et al., 2019).

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence IEEE Transactions on*, 35(8):1798–1828.

Benoit, S., Colletaz, G., Hurlin, C., and Pérignon, C. (2013). A theoretical and empirical comparison of systemic risk measures. *HEC Paris Research Paper*, (FIN-2014-1030).

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

Bezdek, J. C. (1981). *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum.

Bharath, S. T. and Shumway, T. (2008). Forecasting default with the merton distance to default model. *The Review of Financial Studies*, 21:1339–1369.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Borri, N. and di Giorgio, G. (2021). Systemic risk and the covid challenge in the european banking sector. *Journal of Banking & Finance*, In Press:1–13.

Boudoukh, J., Richardson, M., and Whitelaw, R. F. (1998). The best of both worlds: A hybrid approach to calculating Value at Risk. *Risk*, 11:64–67.

Boyd, J. H. and Graham, S. L. (1986). Risk, regulation, and bank holding company expansion into nonbanking. *Quarterly Review Federal Reserve Bank of Minneapolis*, 10(2):2–17.

Boyd, J. H., Graham, S. L., and Hewitt, R. S. (1993). Bank holding company mergers with nonbank financial firms: effects on the risk of failure. *Journal of Banking & Finance*, 17:43–63.

Brownlees, C. and Engle, R. F. (2016). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *The Review of Financial Studies*, 30(1):48–79.

Cernohorska, L. (2017). The efficiency of selected banking sectors in the european union. *Financial Assets and Investing*, 8(1):5–18.

Chi, J. T. and Chi, E. C. (2015). kpodclustr. *R Cran*.

Chi, J. T., Chi, E. C., and Baraniuk, R. G. (2016). k-pod: A method for k-means clustering of missing data. *The American Statistician*, 70(1):91–99.

Danley, B. (2019). Forest owner objectives typologies: Instruments for each owner type or instruments for most owner types? *Forest Policy and Economics*, 105:72–82.

Dardac, N. and Boitan, I. A. (2009). A cluster analysis approach for banks´ risk profile: The romanian evidence. *European Research Studies Journal*, 12(1):109–118.

Dardac, N. and Giba, A. (2011). Systemic financial crisis: A cluster analysis. *European Research Studies Journal*, 14(2):53–64.

de Oliveira, A. F. and do Carmo Nicoletti, M. (2019). Favoring the k-means algorithm with initialization methods. In *International Conference on Intelligent Systems Design and Applications*, pages 21–31.

deJonghe, O. (2010). Back to the basics in banking? a micro-analysis of banking system stability. *Journal of financial intermediation*, 19(3):387–417.

Deutsch, S. B. and Martin, J. J. (1971). An ordering algorithm for analysis of data arrays. *Operations Research*, 19(6):1350–1362.

Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3:32–57.

Elyasiani, E. and Jia, J. J. (2019). Relative performance and systemic risk contributions of small and large banks during the financial crisis. *The Quarterly Review of Economics and Finance.*

Embrechts, P. (2000). Extreme value theory: Potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, 6(1):449–456.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Ruan, T. (2019). Measuring the probability of a financial crisis. In *Proceedings of the National Academy of Sciences.*

Ercan, H. and Sayaseng, S. (2016). The cluster analysis of the banking sector in europe. In *Economics and Management of Global Value Chains*, pages 111–127.

Finger, C. C., Lardy, J. P., Finkelstein, V., Pan, G., Ta, T., and Tierney, J. (2002). CreditGrades. Technical report, RiskMetrics Group.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classification. *Biometrics*, 21:768–769.

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

González-Fernández, A. B., Rodríguez-Pérez, J. R., Sanz-Ablanedo, E., Valenciano, J. B., and Marcelo, V. (2019). Delineating vineyard zones by fuzzy k-means algorithm based on grape sampling variables. *Scientia Horticulturae*, 243:559–566.

Goyeau, D. and Tarazi, A. (1992). Evaluation du risque de défaillance bancaire en europe. *Revue d'Economie Politique*, 102:249–280.

Goyenko, R. Y., Holden, C. W., and Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92:153–181.

Hannan, T. and Hanweck, G. (1988). Bank insolvency risk and the market for large certificates of deposit. *Journal of Money, Credit and Banking*, 20(2):203–211.

Heil, J., Häring, V., Marschner, B., and Stumpe, B. (2019). Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with west african soils. *Geoderma*, 337:11–21.

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

Hull, J. and White, A. (1998). Incorporating volatility updating into the historical simulation method for VaR. *Journal of Risk*, 1:5–19.

Hunter, J. and Dale, D. (2007). The matplotlib users guide. Technical report, R Cran.

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

John, J. A. and Draper, N. R. (1980). An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):190–197.

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*. 3 edition.

Kindhi, B. A., Sardjono, T. A., Purnomo, M. H., and Verkerke, G. J. (2019). Hybrid k-means, fuzzy c-means, and hierarchical clustering for dna hepatitis c virus trend mutation analysis. *Expert Systems with Applications*, 121:373–381.

Kleinow, J., Moreira, F., Strobl, S., and Vähämaa, S. (2017). Measuring systemic risk: A comparison of alternative market-based approaches. *Finance Research Letters*, 21:40–46.

Knotek, P. (2014). Banking sectors in emu - cluster analysis. *European Scientific Journal*, 10(34).

Laeven, L., Ratnovski, L., and Tong, H. (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of banking and finance*, 69(S1):S24–S34.

LeCun, Y. (2012). *Learning invariant feature hierarchies. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7583 LNCS, PART 1.*

Lepetit, L., Nys, E., Rous, P., and Tarazi, A. (2008). Bank income structure and risk: An empirical analysis of european banks. *Journal of Banking and Finance*, 32:1452–1467.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

López-Espinosa, G., Moreno, A., Rubia, A., and Valderrama, L. (2012). Short-term wholesale funding and systemic risk: A global covar approach. *Journal of Banking & Finance*, 36(12):3150–3162.

López-Espinosa, G., Rubia, A., Valderrama, L., and Antón, M. (2013). Good for one, bad for all: Determinants of individual versus systemic risk. *Journal of Financial Stability*, 9(3):287–299.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Volume 1: Statistics, 281–297.

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the ninth Python in science conference*, 445:51–56.

Mercadier, M. and Lardy, J.-P. (2019). Credit spread approximation and improvement using random forest regression. *European Journal of Operational Research*, 277(1):351–365.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470.

Moldovan, D. and Mutu, S. (2015). A cluster analysis on the default determinants in the european banking sector. In *Business Information Systems Workshops - BIS 2015 International Workshops, Poznań, Poland, June 24-26, 2015, Revised Papers*, pages 64–74.

Niedzielski, T., Jurecka, M., Stec, M., Wieczorek, M., and Mizinski, B. (2017). The nested k-means method: A new approach for detecting lost persons in aerial images acquired by unmanned aerial vehicles. *Journal of Field Robotics*, 34(8):1395–1406.

Pavan, K. K., Rao, A. A., Rao, A. V. D., and Sridhar, G. R. (2011). Robust seed selection algorithm for k-means type algorithms-optimal centroids using high density object. *International Journal of Computer Science & Information Technology*, 3(5):147–163.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Prabhu, P. and Anbazhagan, N. (2011). Improving the performance of k-means clustering for high dimensional data set. *International Journal on Computer Science and Engineering*, 3(6):2317–2322.

Python Software Foundation (2016). Python 3.6.0 documentation.

Reboredo, J. C. and Ugolini, A. (2015). Systemic risk in european sovereign debt markets: A covar-copula approach. *Journal of International Money and Finance*, 51:214–244.

Rizwan, M. S., Ahmad, G., and Ashraf, D. (2020). Systemic risk: The impact of COVID-19. *Finance Research Letters*, 36:101682 (1–7).

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica*, 20:431–449.

Saghi-Zedek, N. and Tarazi, A. (2015). Excess control rights, financial crisis and bank profitability and risk. *Journal of Banking and Finance*, 55:361–379.

Saitta, S., Raphael, B., and Smith, I. F. C. (2008). A comprehensive validity index for clustering. *Intelligent Data Analysis*, 12(6):529–548.

Sanford, A. and Moosa, I. (2015). Operational risk modelling and organizational learning in structured finance operations: a bayesian network approach. *Journal of the Operational Research Society*, 66:86–115.

Sedunov, J. (2016). What is the systemic risk exposure of financial institutions? *Journal of financial stability*, 24(C):71–87.

Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J., Grosz, B., and Bauer, Z. (2018). The ai index 2018 annual report. Technical report, AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA.

Soerensen, C. K. and Gutiérrez, J. M. P. (2006). Euro area banking sector integration using hierarchical cluster analysis techniques. In *European Central Bank: Working Paper Series*, number 627.

Steinhaus, H. (1956). Sur la division des corps matériels en parties. *Bulletin de l'académie Polonaise des sciences*, 4(12).

Straetmans, S. T. M., Verschoor, W. F. C., and Wolff, C. C. P. (2008). Extreme us stock market fluctuations in the wake of 9/11. *Journal of Applied Econometrics*, 23(1):17–42.

Varotto, S. and Zhao, L. (2018). Systemic risk and bank size. *Journal of International Money and Finance*, 82:45–70.

Walt, S. V. D., Colbert, S., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13:22–30.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58(301):236–244.

Weiß, G. N., Neumann, S., and Bostandzic, D. (2014). Systemic risk and bank consolidation: International evidence. *Journal of Banking and Finance*, 40:165–181.

Xu, Q., Ding, C., Liu, J., and Luo, B. (2015). PCA-guided search for k-means. *Pattern Recognition Letters*, 54:50–55.

Zarandi, S. A. T. P., Shahabi, V., Emami, M., Rezaee, A., and Hassanloo, M. (2014). Ranking banks using k-means and grey relational method. *Management Science Letters*, 4(10):2319–2324.

Zhang, Q., Vallascas, F., Keasey, K., and Cai, C. X. (2015). Are market-based measures of global systemic importance of financial institutions useful to regulators and supervisors? *Journal of Money, Credit and Banking*, 47(7):1403–1442.

Zhu, C., Idemudia, C. U., and Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and k-means techniques. *Informatics in Medicine Unlocked*.

8 Appendix

Complete list of banks

Australia: Aust And Nz Bank (ANZ AU Equity), Commonw Bk Austr (CBA AU Equity), Macquarie Group (MQG AU Equity), Natl Aust Bank (NAB AU Equity), Westpac Banking (WBC AU Equity), **Austria**: Erste Group Bank (EBS AV Equity), Raiffeisen Bank (RBI AV Equity), **Belgium**: Dexia Sa (DEXB BB Equity), Kbc Group (KBC BB Equity), **Brazil**: Banco Do Brasil (BBAS3 BZ Equity), Bradesco Sa-Pref (BBDC4 BZ Equity), Itau Uniban-Pref (ITUB4 BZ Equity), **Britain**: Alliance & Leice (AL/ LN Equity), Barclays Plc (BARC LN Equity), Bradford & Bing (BB/ LN Equity), Hbos Plc (HBOS LN Equity), Hsbc Holdings Pl (HSBA LN Equity), Lloyds Banking (LLOY LN Equity), Natwest Group Plc (NWG LN Equity), Standard Charter (STAN LN Equity), **Canada**: Bank Of Montreal (BMO CN Equity), Bank Of Nova Sco (BNS CN Equity), Can Impl Bk Comm (CM CN Equity), Natl Bk Canada (NA CN Equity), Royal Bank Of Ca (RY CN Equity), Toronto-Dom Bank (TD CN Equity), **China**: Agricultural-A (601288 CH Equity), Bank Of China-A (601988 CH Equity), China Const Ba-A (601939 CH Equity), China Merch Bk-A (600036 CH Equity), China Minsheng-A (600016 CH Equity), Ind & Comm Bk-A (601398 CH Equity), Ping An Bank-A (000001 CH Equity), Shang Pudong-A (600000 CH Equity), **Denmark**: Danske Bank A-S (DANSKE DC Equity), Jyske Bank-Reg (JYSK DC Equity), **Finland**: Nordea Bank Abp (NDA SS Equity), **France**: Bnp Paribas (BNP FP Equity), Credit Agricole (ACA FP Equity), Credit Industrie (CC FP Equity), Natixis (KN FP Equity), Soc Generale Sa (GLE FP Equity), **Germany**: Aareal Bank Ag (ARL GR Equity), Commerzbank (CBK GR Equity), Deutsche Bank-Rg (DBK GR Equity), Deutsche Postban (DPB GR Equity), **Greece**: Alpha Bank Ae (ALPHA GA Equity), Attica Bank Sa (TATT GA Equity), Eurobank Ergasia (EUROB GA Equity), Piraeus Bank (TPEIR GA Equity), **Hong Kong**: Bank East Asia (23 HK Equity), Boc Hong Kong Ho (2388 HK Equity), Chong Hing Bank (1111 HK Equity), Hang Seng Bank (11 HK Equity), **India**: Axis Bank Ltd (AXSB IN Equity), Bank Of Baroda (BOB IN Equity), Bank Of India (BOI IN Equity), Canara Bank (CBK IN Equity), Hdfc Bank Ltd (HDFCB IN Equity), Icici Bank Ltd (ICICIBC IN Equity), Idbi Bank Ltd (IDBI IN Equity), Indian Overseas (IOB IN Equity), Punjab Natl Bank (PNB IN Equity), State Bank Ind (SBIN IN Equity), Syndicate Bank (SNDB IN Equity), Union Bank India (UNBK IN Equity), **Ireland**: Bank Of Ireland (BIRG ID Equity), Irish Bank Resol (ANGL ID Equity), **Israel**: Bank Hapoalim (POLI IT Equity), Bank Leumi Le-Is (LUMI IT Equity), Fibi Bank (FIBI IT Equity), Israel Discoun-A (DSCT IT Equity), Mizrahi Tefahot (MZTF IT Equity), Union Bank Israe (UNON IT Equity), **Italy**: Banca Monte Dei (BMPS IM Equity), Intesa Sanpaolo (ISP IM Equity), Mediobanca (MB IM Equity), Unicredit Spa (UCG IM Equity), **Japan**: Chiba Bank Ltd (8331 JP Equity), Daiwa Secs Grp (8601 JP Equity), Gunma Bank Ltd (8334 JP Equity), Mitsubishi Ufj F (8306 JP Equity), Mizuho Financial (8411 JP Equity), Nomura Holdings (8604 JP Equity), Oita Bank Ltd (8392 JP Equity), Resona Holdings (8308 JP Equity), Shiga Bank Ltd (8366 JP Equity), Shizuoka Bank (8355 JP Equity), Smfg (8316 JP Equity), The Bank Iwate (8345 JP Equity), Yamagata Bank (8344 JP Equity), **Netherlands**: Ing Groep Nv (INGA NA Equity), **Norway**: Dnb Asa (DNB NO Equity), Sparebank 1 Sr B (SRBNK NO Equity), **Poland**: Bos (BOS PW Equity), Handlowy (BHW PW Equity), Mbank Sa (MBK PW Equity), Pekao (PEO PW Equity), Pkobp (PKO PW Equity), Santander Bank (SPL PW Equity), **Portugal**: Banco Bpi Sa-Reg (BPI PL Equity), Banco Com Port-R (BCP PL Equity), **Russia**: Sberbank (SBER RM Equity), **Singapore**: Dbs Group Hldgs (DBS SP Equity), Ocbc Bank (OCBC SP Equity), United Overseas (UOB SP Equity), **South Korea**: Dgb Financial Gr (139130 KS Equity), Industrial Bank (024110 KS Equity), Shinhan Financia (055550 KS Equity), **Spain**: Banco Popular (POP SM Equity), Banco Sabadell (SAB SM Equity), Banco Santander (SAN SM Equity), Bankinter (BKT SM Equity), Bbva (BBVA SM Equity), Caixabank Sa (CABK SM Equity), **Sweden**: Seb Ab-A (SEBA SS Equity), Svenska Han-A (SHBA SS Equity), Swedbank Ab-A (SWEDA SS Equity), **Switzerland**: Banq Cant Gen-Br (BCGE SW Equity), Credit Suiss-Reg (CSGN SW Equity), Ubs Group Ag (UBSG SW

Equity), Vontobel Hldg-R (VONN SW Equity), **Taiwan**: Chang Hwa Bank (2801 TT Equity), Far Eastern Intl (2845 TT Equity), Fubon Financial (2881 TT Equity), **Turkey**: Akbank (AKBNK TI Equity), Denizbank As (DENIZ TI Equity), Garanti (GARAN TI Equity), Halkbank (HALKB TI Equity), Is Bankasi (ISCTR TI Equity), Vakibank (VAKBN TI Equity), Yapi Kredi (YKBNK TI Equity), **United States**: Bank Ny Mellon (BK US Equity), Bank Of America (BAC US Equity), Bear Stearns Cos (2942331Q US Equity), Capital One Fina (COF US Equity), Citigroup Inc (C US Equity), Comerica Inc (CMA US Equity), Countrywide Fina (CFC US Equity), Fifth Third Banc (FITB US Equity), First Horizon Na (FHN US Equity), Goldman Sachs Gp (GS US Equity), Huntington Banc (HBAN US Equity), Jpmorgan Chase (JPM US Equity), Keycorp (KEY US Equity), Lehman Bros Hldg (LEHMQ US Equity), M&T Bank Corp (MTB US Equity), Merrill Lynch (MER US Equity), Morgan Stanley (MS US Equity), Netbank Inc (NTBKQ US Equity), Northern Trust (NTRS US Equity), Pnc Financial Se (PNC US Equity), Regions Financia (RF US Equity), Schwab (Charles) (SCHW US Equity), State St Corp (STT US Equity), Suntrust Banks (STI US Equity), Synovus Finl (SNV US Equity), Truist Financial Corp (TFC US Equity), Us Bancorp (USB US Equity), Wamu Inc (WAMUQ US Equity), Wells Fargo & Co (WFC US Equity)

Country	**Index**
Australia	AS51 Index
Austria	ATX Index
Belgium	BEL20 Index
Brazil	IBOV Index
Britain	UKX Index
Canada	SPTSX Index
China	SHCOMP Index
Denmark	KFX Index
Finland	HEX Index
France	CAC Index
Germany	DAX Index
Greece	ASE Index
Hong Kong	HSI Index
India	SENSEX Index
Ireland	ISEQ Index
Israel	TA-35 Index
Italy	FTSEMIB Index
Japan	NKY Index
Netherlands	AEX Index
Norway	OBX Index
Poland	WIG20 Index
Portugal	PSI20 Index
Russia	IMOEX Index
Singapore	STI Index
South Korea	KOSPI2 Index
Spain	IBEX Index
Sweden	OMX Index
Switzerland	SMI Index
Taiwan	TWSE Index
Turkey	XU100 Index
United States	SPX Index

The studied universe is as follow:

- 161 listed banks out of which some are delisted or acquired: Netbank, Bear Stearns, Countrywide Financial, Lehman Brothers, Washington Mutual, Bradford & Bingley, Alliance & Leicester, Merril Lynch, HBOS, Irish Bank Resolution, Deutsche Postbank, Banco Popular Español, Crédit Industriel et Commercial, Banco BPI, Dexia, SunTrust Banks, Denizbank, Syndicate Bank, Union Bank of Israel, Natixis & Chong Hing Bank.

- 31 national indices (Number of banks per country): Australia (5), Austria (2), Belgium (2), Brazil (3), Britain (8), Canada (6), China (8), Denmark (2), Finland (1), France (5), Germany (4), Greece (4), Hong Kong (4), India (12), Ireland (2), Israel (6), Italy (4), Japan (13), Netherlands (1), Norway (2), Poland (6), Portugal (2), Russia (1), Singapore (3), South Korea (3), Spain (6), Sweden (3), Switzerland (4), Taiwan (3), Turkey (7), United States (29) and the MSCI World & MSCI World Bank.

- 20 Currencies + 1 currency index (DXY) standing for the USD.

- 926 dates from January 2$^{\text{nd}}$, 2004 until September 24$^{\text{th}}$, 2021.

Indicators and factors specificities

Factors	**Explained Variance**	**Indicators**	**Description**
BetaCorrel	65%	Correl	correl with national stock index
BetaCorrel		CorrelGlb	correl with MXWO
BetaCorrel		CorrelGlbBk	correl with MXWO-bank
BetaCorrel		Beta	Beta with national stock index
BetaCorrel		BetaGlb	Beta with MXWO
BetaCorrel		BetaGlbBk	Beta with MXWO-bank
BetaCorrel		TailBeta	Tail Beta with national stock index
CreditBS	48%	TexasRatio	Texas ratio
CreditBS		Z-score (-)	Z-score
CreditBS		PBRatio (-)	Price-to-Book ratio
CreditBS		ROA (-)	Return on Assets
CreditBS		ComEqy2Ast (-)	Common Equities to Total Assets
CreditMkt	73%	CrdGrd	CDS Approximation
CreditMkt		E2C	CDS Approximation
CreditMkt		Debt2EV	Debt to Enterprise Value
IndivMktRisk	92%	Volatility	annualized volatility
IndivMktRisk		VaR	1y historical Value-at-Risk
IndivMktRisk		VaRComp	Composite Value-at-Risk
IndivMktRisk		VaRHW	Hull & White Value-at-Risk
IndivMktRisk		VaRBRW	Boudoukh et al. Value-at-Risk
IndivMktRisk		ES	1y historical Expected Shorfall
IndivMktRisk		ESBRW	Boudoukh et al. Expected Shorfall
IndivMktRisk		ESn	Normal Expected Shorfall
IndivMktRisk		MZ-score (-)	Market Z-score
RatioBS	41%	Debt2Ast	Total Debts to Total Assets
RatioBS		Dpst2Ast (-)	Deposits to Total Assets
RatioBS		Loan2Ast	Total Loans to Total Assets
RatioBS		Loan2Dpst	Total Loans to Deposits
RatioBS		NPA2Ast	Non Preforming Assets to Total Assets
RatioBS		NPL2Loan	Non Preforming Loans to Total Loans
RatioBS		LLRsrv2Loan	Reserve for Loan Loss to Total Loans
SysBkMkt	77%	MES	MES with national stock index
SysBkMkt		MESGlb	MES with MXWO
SysBkMkt		MESGlbBk	MES with MXWO-bank
SysBkMkt		MESBRW	MES with national stock index
SysBkMkt		MESBRWGlb	MES with MXWO
SysBkMkt		MESBRWGlbBk	MES with MXWO-bank
SysBkMkt		MESBeta	MES with national stock index
SysBkMkt		MESBetaGlb	MES with MXWO
SysBkMkt		MESBetaGlbBk	MES with MXWO-bank
SysBkMkt		LRMES	6M MES with national stock index
SysBkMkt		LRMESGlb	6M MES with MXWO
SysBkMkt		LRMESGlbBk	6M MES with MXWO-bank
SysBkMkt		LRMESBRW	6M MES with national stock index
SysBkMkt		LRMESBRWGlb	6M MES with MXWO
SysBkMkt		LRMESBRWGlbBk	6M MES with MXWO-bank
SysBkMkt		LRMESBeta	6M MES with national stock index
SysBkMkt		LRMESBetaGlb	6M MES with MXWO
SysBkMkt		LRMESBetaGlbBk	6M MES with MXWO-bank

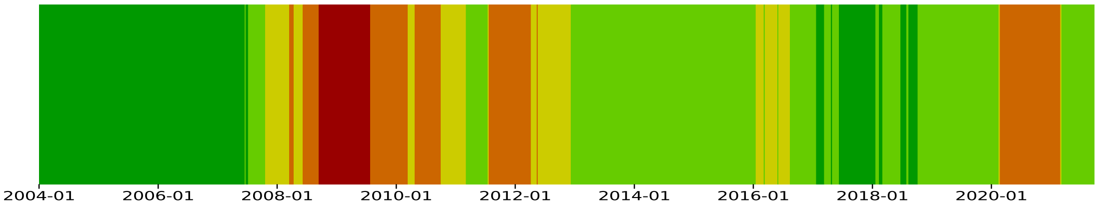
Table 3: Summary Table-1

Factors	**Explained Variance**	**Indicators**	**Description**
SysMktBk	78%	MESInv	MES from nat. stock index on Bank
SysMktBk		MESInvGlb	MES from MXWO on Bank
SysMktBk		MESInvGlbBk	MES from MXWO-bank on Bank
SysMktBk		MESBRWInv	MES from nat. stock index on Bank
SysMktBk		MESBRWInvGlb	MES from MXWO on Bank
SysMktBk		MESBRWInvGlbBk	MES from MXWO-bank on Bank
SysMktBk		dCoVaR	ΔCoVaR with national stock index
SysMktBk		dCoVaRGlb	ΔCoVaR with MXWO
SysMktBk		dCoVaRGlbBk	ΔCoVaR with MXWO-bank
SysMktBk		dCoVaRComp	ΔCoVaR with national stock index
SysMktBk		dCoVaRCompGlb	ΔCoVaR with MXWO
SysMktBk		dCoVaRCompGlbBk	ΔCoVaR with MXWO-bank
SRISK	90%	SRISKMES	SRISK with national stock index
SRISK		SRISKMESGlb	SRISK with MXWO
SRISK		SRISKMESGlbBk	SRISK with MXWO-bank
SRISK		SRISKMESBRW	SRISK with national stock index
SRISK		SRISKMESBRWGlb	SRISK with MXWO
SRISK		SRISKMESBRWGlbBk	SRISK with MXWO-bank
SRISK		SRISKBeta	SRISK with national stock index
SRISK		SRISKBetaGlb	SRISK with MXWO
SRISK		SRISKBetaGlbBk	SRISK with MXWO-bank
RegCapital (i.s.)	71%	Cap2RskCap (-)	Total capital to Risk based Capital
RegCapital (i.s.)		T1CapRatio (-)	Tier 1 Risk-Based Capital Ratio
RegCapital (i.s.)		RWA2Ast	RWA to Total Assets
Size (i.s.)	82%	LLRsrv (-)	Reserve for Loan Loss
Size (i.s.)		Deposit (-)	Deposits
Size (i.s.)		ComEqy (-)	Common Equities
Size (i.s.)		TotEqy (-)	Total Equities
Size (i.s.)		TotAst (-)	Total Assets
Size (i.s.)		RWA (-)	Risk-Weighted Assets
Size (i.s.)		TotLoan (-)	Total Loans
Size (i.s.)		TotDebt (-)	Total Debts
Size (i.s.)		MktCap (-)	Market Capitalization
LiquidityMkt (i.s.)	61%	BidAskSprd	Bid-Ask spread
LiquidityMkt (i.s.)		VolmNbSh (-)	Volume / Number of Shares
LiquidityMkt (i.s.)		VolmPx (-)	Volume * Stock prices

Notes. This table introduces each factor's name & explained variance and its members' (indicators) name & description. It is worth noting that a factor's (i.s.) refers to inverted scale (e.g. large size factor for small banks), and an indicator's (-) sign means that it has been multiplied by -1 to be an increasing function of the risk.

Table 4: Summary Table-2

The United States aggregated clustering



Notes. This graph presents in which cluster the United States banks' weighted average per total assets belongs through time. It notably highlights its presence in the worst cluster during the GFC and the debt crisis to a lower extent.

Figure 14: The United States Output