A geographical analysis of the systemic risk by a Compositional Data (CoDa) approach

Francesco Porro

Università degli Studi di Genova, Italy
francesco.porro@unige.it

Abstract. After the financial crises in the last decades, the systemic risk is now recognized as an unavoidable issue to be constantly monitored. In this paper the contributions to the global systemic risk of different geographic areas are investigated. The analysis is performed by considering the aggregate systemic risk of the firms located in a specific geographic area as a part of the distribution of the global systemic risk. The techniques used for a such investigation are based on Compositional Data (CoDa) methodology, a quite recent approach very useful when the relevant information convoyed by the data is in the proportions among the parts and not in their absolute values or in their sum.

Keywords: Systemic Risk, SRISK, Compositional Data, Aitchison geometry

1 Introduction

The relevance of the systemic risk has been globally emphasized by the global financial crisis of 2007-2009. Initially, there was no commonly accepted definition of the systemic risk, but now it can be identified as "the propensity of a financial institution to be undercapitalized when the financial system as a whole is undercapitalized" [6]. Many papers deal with the issue of the assessment of the systemic risk: for an overview on this topic, see for example [4], [6], and the references therein. An important measure for evaluating the systemic-risk degree associated with a single firm is the so-called SRISK, which has been firstly introduced in [1] and later extended in many other papers. More details about this topic can be found in the interesting review in [7].

The value of SRISK related to a single specific firm can be intended as the amount of money needed by the firm to rise to function normally in the case of a financial crisis (such as a very significant loss of the financial markets). Formally, at a given time t, the value of SRISK for an individual firm i is calculated by:

$$SRISK_{it} = k\,[D_{it} + (1-\phi_{it})W_{it}] - (1-\phi_{it})W_{it}, \qquad \text{where:}$$

- k is the prudential capital ratio;
- D_{it} the book value of total liabilities;
- W_{it} the market capitalization (or market value of equity);

– ϕ_{it} the Long-Run Marginal Expected Shortfall (LRMES), which corresponds to the expected drop in equity value conditional on the market falling by more than 40% within the next six months.

The SRISK can be decomposed in the difference between two quantities: the first one, which represents the *Required Capital*, is equal to $k[D_{it}+(1-\phi_{it})W_{it}]$, while the second one, given by $(1-\phi_{it})W_{it}$ corresponds to the *Available Capital*. As stated in [5], "SRISK is a function of the size of the firm, its degree of leverage, and its expected equity devaluation conditional on a market decline. SRISK is higher for firms that are larger, more leveraged, and with higher sensitivity to market declines." A negative (or null) value of SRISK for a specific firm, means that such firm can overcome the market shock with no capital injections by the government. For this reason, a measure of the financial distress for a system (with N firms) is the aggregate SRISK, that, at a given moment t, is defined by:

$$SRISK_t = \sum_{i=1}^{N} (SRISK_{it})_+ \tag{1}$$

where $(x)_+$ denotes $\max(0, x)$. This measure represents the total amount of bailing out a financial system (with N firms), conditional on a systemic event. In the computation of the aggregate SRISK, the negative amounts of SRISK do not effectively contribute in the sum in formula (1), because in a crisis it is unlikely that surplus capital will be easily mobilized through mergers or loans among firms (cfr. [4]). The fact that, financially speaking, the world is a global village and the regulators must consider "a large picture" to make decisions, is now accepted by the most part of the researchers. For this reason, an analysis about the influence of geographical factors to the systemic risk can help to monitor and to better understand the dynamics of such kind of risk.

Keeping this in mind, in this paper a geographical analysis of the composition of the systemic risk is performed. The analysis is conducted with a dataset provided by the Volatility Laboratory (V-Lab) (more information about the data can be found in Section 3). The approach used is based on the Compositional Data (CoDa) methodology, a set of quite recent techniques, which is getting more and more attention in the literature (cfr. [2] and [3]).

2 The compositional approach

The Compositional Data (CoDa) are multivariate observations where relative rather than absolute information is relevant. This means that they represent a quantitative description of the parts of some whole. The basic pillar in compositional methods are the compositions, defined as follows.

Definition 1. *A composition vector is a real-valued vector with all (strictly) positive components. A D-part composition is a class of equivalence which contains all the* compositionally equivalent *vectors in* $\mathbb{R}^D$*, where two compositions* $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_D)$ *are compositionally equivalent if there exists a positive constant* $\lambda \in \mathbb{R}^+$ *such that* $\mathbf{x} = \lambda \cdot \mathbf{y}$.

A suitable sample space for the equivalence classes is the D-part simplex $\mathbb{S}^D$, defined as:

$$\mathbb{S}^D = \{(x_1, x_2, \dots, x_D) \in \mathbb{R}^D \ : \ x_i > 0 \, \forall i; \ \sum_{i=1}^{D} x_i = c\}, \tag{2}$$

where c is a positive arbitrary constant. For further details, see [9], [8] and the references therein. Usually in compositional analysis, the vectors of proportions (which sum to 1) are used as representatives of an equivalence class: this corresponds to select $c = 1$ in the previous definition, and in the next one.

Definition 2. *The closure (to c) of the D-part composition* $\mathbf{x} = (x_1, x_2, \dots, x_D)$ *is given by:*

$$\mathscr{C}(\mathbf{x}) = \left(\frac{c \cdot x_1}{\sum_{i=1}^{D} x_i}, \frac{c \cdot x_2}{\sum_{i=1}^{D} x_i}, \dots, \frac{c \cdot x_D}{\sum_{i=1}^{D} x_i} \right).$$

Starting from these initial two definitions, it is possible to create a coherent geometry, called Aitchison geometry on the simplex, which allows a deep analysis of compositional data (see [9] for further details).

A typical compositional dataset $\mathbf{X}$ is a sample of n observations of D-part compositions $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, with $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD}), \ i = 1, 2, \dots, n$. Since in such a dataset, the standard statistical descriptive measures, based on the real Euclidean structure, applied to compositional data may lead to erroneous conclusions (see for example [9]), an alternative set of descriptive measures based on the Aitchison geometry can more properly be used. In the following, just the most common two are reported.

Definition 3. *An indicator of central tendency for the compositional dataset* $\mathbf{X}$ *is the closed geometric mean. This vector is called* center*, and it is defined as*

$$cen(\mathbf{X}) = \mathscr{C}(g_1, g_2, \dots, g_D),$$

where g_j *denotes the geometric mean of the* n *observations related to the* j*-th component of the vectors in* $\mathbf{X}$*:* $g_j = \left(\prod_{i=1}^{n} x_{ij}\right)^{1/n}, \ j = 1, 2, \dots, D.$

The dispersion in a compositional dataset $\mathbf{X}$, can be described by the variation matrix, defined by:

$$T = \begin{pmatrix} 0 & t_{12} & \cdots & t_{1D} \\ t_{21} & 0 & \cdots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & \cdots & 0 \end{pmatrix}, \quad \text{where } t_{ij} = var\left(\ln \frac{x_i}{x_j}\right).$$

For measuring the variability of a dataset by a single value, the following definition of *Total Variance* has been introduced.

Definition 4. *The Total Variance of the compositional sample* $\mathbf{X}$ *is a measure of its global dispersion. It is based on the entries of the variation matrix* T*:*

$$TotVar(\mathbf{X}) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} var\left(\ln \frac{x_i}{x_j}\right) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} t_{ij}.$$

As the information convoyed by the compositions is relative, an usual practice is to apply transformations, mapping the compositions into real vectors in order to exploit the usual Euclidean structure. In the literature there are several transformations based on the logratios: the additive logratio (*alr*), the centered logratio (*clr*), and the isometric logratio (*ilr*). Unfortunately, even a simple overview of these transformations is out of scope of this paper: the interested reader can see [9] and [8], among the others, for further details. For the provided analysis in the following, it can just be reported that the *clr*-transformation is basically characterized by two important properties: the first one is that it does not change the number of parts, since a D composition is mapped in a vector in $\mathbb{R}^D$. The second one is that it preserves the distances and the angles: this implies that the Aitchison distance in the simplex of two compositions is equal to the distance of the corresponding transformed vectors in $\mathbb{R}^D$ (see [9] for details). This feature is fundamental in exploratory analyses based on metrics, like *clr*-biplots and ternary or De Finetti diagrams. The definition of the centered logratio transformation, is the following one.

Definition 5. *The centered logratio transformation (clr) of a composition* $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ *is given by*

$$clr(\mathbf{x}) = \ln\left(\frac{x_1}{g_m(\mathbf{x})}, \frac{x_2}{g_m(\mathbf{x})}, \ldots, \frac{x_D}{g_m(\mathbf{x})}\right),$$

where $g_m(\mathbf{x})$ *denotes the geometric mean of the D parts:* $g_m(\mathbf{x}) = \left(\prod_{i=1}^{D} x_i\right)^{1/D}$.

In every compositional data analysis, graphics are usually used to visualize and interpret the data. The most common one is the PCA biplot. Generally speaking, the biplots permit the representation of a rank-2 approximation of the data, and they are based on the Single Value Decomposition (SVD) of the centered (or standardized) data matrix. In Compositional Data analysis, there are two basic kinds of biplots: the *form biplot*, which favours the display of the units, and the *covariance biplot*, which favours the display of the variables. More details on them can be found in [9], and in [8].

3 Application

The data used in the application are provided by the Volatility Laboratory (V-Lab) of the New York University Stern School of Business, and they are available at the website: https://vlab.stern.nyu.edu. The dataset consists of the value of the aggregate SRISK for four macro-areas: Africa, Americas, Europe, and Asia. Such values are related to each quarter from 2014 to 2021: the value for each quarter is the value of SRISK in the last month of the period (March, June, September, and December). The only exception regards the year 2021: since at the moment, the SRISK detection of December is not available, the last quarter of such year is not considered. The aggregate SRISK of each macro-area is based on the SRISK of all the firms in such geographical area, monitored by the V-Lab team, therefore enough big to have a systemic relevance. In practice, the dataset can be seen as a compositional sample of 31 different 4-compositions, grouped in 8 different years (2014-2021). The left panel of Table 1 reports the sample center and the right panel, the variation matrix of the dataset. The center highlights an important predominance of Asia, showing a geometric mean that exceeds more than the 50% SRISK share over the sample period. The right panel shows that the maximum variability is associated with Africa, and the Total Variance is equal to 0.2890. In order to identify patterns in the data,

Table 1. The center of the 4-part composition (left panel) and the Logratio variances of the 4-part composition for the macro-areas (right panel).

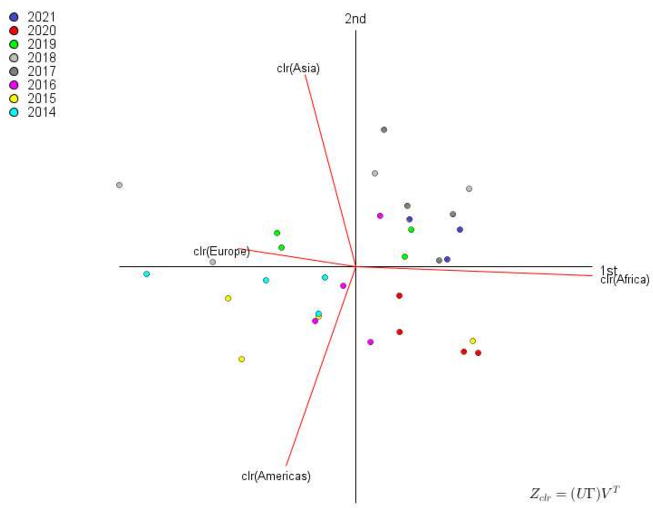
Macro-Areas	Center
Africa	0.0031
Americas	0.1299
Asia	0.5262
Europe	0.3408

	Africa	Americas	Asia	Europe	Clr variances
Africa	0	0.2751	0.2525	0.3219	0.1061
Americas	0.2751	0	0.1438	0.0840	0.0606
Asia	0.2525	0.1438	0	0.0786	0.0629
Europe	0.3219	0.0840	0.0786	0	0.0594
				Total Variance	0.2890

Figure 1 shows the covariance biplot. The first two Principal Components are representative of the dataset variability, since the first one explains the 66.43% of the total variance, and the second one adds an other 24.83%, bearing the proportion of the cumulative explained variance at more than 90% (91.26%). By using some interpretation rules of the compositional biplot, some remarks can be achieved from Figure 1. The first one is related to the interpretation of the Principal Components: the first component discriminates Africa from the other macro-areas, likely because it captures a sort of "evolution degree" of the financial markets in the macro-area; the second one sharply distinguishes between Asia and Americas. By an observation of the length of the four rays (the segments joining each vertex to the center), it can be stated that the smallest one is related to Europe, suggesting that the logratio of such macro-area has the smallest contribution to the total variability. The links (segments joining two vertices) corresponding to the pairs Asia-Americas and Europe-Africa are nearly orthogonal: this suggests that the corresponding logratios should be checked for zero correlation. The observation that the four vertices are very spread out indicates that all the variances of the logratios are very far from being null, highlighting the lack of proportionality of the parts. Since the projections of the points of the quarters in 2014 on the link Asia-Americas are (quite) close to the projection of the center on the same link, it can be state that such quarters have the value of the logratio corresponding to Asia and Americas (quite) equal to its average on the whole dataset. The projections of the points of years 2014, 2015, 2018 and 2020 are (quite) far from the projection of the center on the link Europe-

Africa: this shows that the values of the corresponding logratios are (quite) very different from the average on the whole dataset. All these findings can be considered coherent with the results of other analysis, but their added value is that they came from the application of the quite recent CoDa methodology.

Fig. 1. The Covariance Biplot of the dataset of the four macro-areas.



References

1. Acharya, V., Richardson, M.: Restoring financial stability: how to repair a failed system. John Wiley & Sons (2009)
2. Aitchison, J., The statistical analysis of compositional data. J R Stat Soc Series B Stat Methodol 44(2), 139-160 (1982)
3. Aitchison, J.: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd, London, UK (1986)
4. Benoit, S., Colliard, J.E., Hurlin, C., Pérignon, C. Where the risks lie: A survey on systemic risk. Rev Financ 21(1), 109-152 (2017)
5. Brownlees, C., Engle, R.F. SRISK: A conditional capital shortfall measure of systemic risk. Rev Financ Stud 30(1), 48-79 (2017)
6. Engle, R., Jondeau, E., Rockinger, M. Systemic Risk in Europe. Rev Financ 19(1), 145-190 (2015)
7. Engle, R., Systemic Risk 10 Years Later. Annu Rev Financ Econ 10, 125-152 (2018)
8. Filzmoser, P., Hron, K., Templ, M. Applied compositional data analysis. Switzerland: Springer Nature (2018)
9. Pawlowsky-Glahn, V., Egozcue, J., Tolosana-Delgado, R. Modeling and analysis of compositional data. John Wiley & Sons (2015)