



The weighted cross-shareholding complex network: a copula approach to concentration and control in financial markets

Roy Cerqueti^{1,2,3}  · Giulia Rotundo⁴

Received: 20 July 2021 / Accepted: 29 July 2022 / Published online: 20 August 2022
© The Author(s) 2022

Abstract

In this work, we focus on the cross-shareholding structure in financial markets. Specifically, we build ad hoc indices of concentration and control by employing a complex network approach with a weighted adjacency matrix. To describe their left and right tail dependence properties, we explore the theoretical dependence structure between such indices through copula functions. The theoretical framework has been tested over a high-quality dataset based on the Italian Stock Market. In doing so, we clearly illustrate how the methodological setting works and derive financial insights. In particular, we advance calibration exercises on parametric copulas under the minimization of both Euclidean distance and entropy measure.

Keywords Cross-shareholdings network · Entropy · Concentration and control · Copulas

The Authors thank Marcel Ausloos and Anna Maria D’Arcangelis for fruitful discussions on the data.

✉ Roy Cerqueti
roy.cerqueti@uniroma1.it

Giulia Rotundo
giulia.rotundo@uniroma1.it

¹ Department of Social and Economic Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

² School of Business, London South Bank University, London SE1 0AA, UK

³ GRANEM, Université d’Angers SFR CONFLUENCES, 49000 Angers, France

⁴ Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

1 Introduction

In 2009 Haldane, the Executive Director for Financial Stability at the Bank of England, remarked that highly interconnected financial networks might be robust-yet-fragile in the sense that “within a certain range, connections serve as shock-absorbers [and] connectivity engenders robustness” (Haldane 2013). Complex network approaches and the science of networks put well in evidence the role of the links and of the underlying network topology in the propagation of contagions and cascades (Elliott et al. 2014; Markose et al. 2012; Varela and Rotundo 2016; White 2014). Besides the cross-shareholdings, literature has examined other channels for detecting the connection among companies and their managers, posing in evidence the interlock of directorates and rich-club relationships (Cinelli 2019; Croci and Grassi 2014; D’Errico et al. 2009; Drago et al. 2015).

This paper considers a financial network where the nodes represent companies. We assume that arcs are directed, and they are weighted on the basis of the ownership of shares of companies. In particular, there is a link between two companies when one owns shares of the other. The link direction goes from the former company to the latter; its weight increases with the amount of such a share.

Following the definitions proposed in the literature, the weighted in-degree of a node gives the concentration of the related company – which offers a measure of the effective shareholders of the company’s stock. Differently, the weighted out-degree of a node drives the definition of the level of control of the company which proxies the effective amount of the stocks of the company controlled by a shareholder (Glattfelder and Battiston 2009; Pecora and Spelta 2015).

In Glattfelder and Battiston (2009), a large dataset of companies listed in national stock markets is examined. A peak for the concentration equal to 1 is detected in all the studied countries. As we will see in detail below, this means that many shareholders in the markets only control one single stock.

In Pecora and Spelta (2015), the network of the Euro Area banking market is analysed. The database showed the situation in 2012. It is quite large since data on 1534 Euro Area banks with 2298 ownership links are retrieved from Bankscope (Bureau van Dijk (BvD) Ownership Database) and cross-checked with other information (mainly Annual Reports and private communications). They detect the power law for the tails. This is relevant for detecting the characteristic of robust-yet-fragile behaviour, i.e. the network is robust against pure random fluctuations, but the contagions propagate quickly when the most relevant nodes undergo some distress. Their analysis shows that control is relatively concentrated in a few shareholders. Differently, there are only weak relationship among them.

The presence of the weights in both the concentration and the fraction of control allows performing a step ahead to the papers considering only the network topology, where integration (in-degree) and diversification (out-degree) are defined (Rotundo and D’Arcangelis 2010, 2014; Elliott et al. 2014; Garlaschelli et al. 2005).

In this respect, we have already examined the occurrence of the highest peaks of concentration for unweighted networks, hence depending strictly on the network topology. Such peaks are the most unstable configurations of the system, so they are potentially susceptible to global avalanches as an answer to an eventual trigger like

the failure or a significant fluctuation of the value of a company (Cerqueti et al. 2018; Elliott et al. 2014). Adding the weights to the network allows one to get a clearer picture of the actual relevance of network links.

The paper aims to provide a theoretical framework for exploring the joint distribution of concentration and control for a prefixed financial network and under different scenarios for the stochastic dependence between the considered quantities. Stochastic dependence is modelled through several copula functions, with particular reference to the Frechet bounds—leading to the maximum and minimum level of correlation between two random variables—the independence copula and three popular parametric Archimedean copulas—the Gumbel, the Clayton and the Frank ones. For an overview of the copulas, refer to (Joe 1997; Nelsen 2007). We use the classical Sklar's Theorem (Sklar 1959) to define the joint theoretical distribution of concentration and control starting from their univariate distributions. The employment of the considered copulas leads to a panoramic view of the stochastic dependence between the investigated financial terms. In this respect, we point out that the Pearson correlation is undoubtedly a popular measure to describe the dependence between financial variables. This explains why it is often used for building financial networks. The Pearson coefficient's relevant drawback is that it only models stochastic dependences of linear type, while financial variables usually show more complex dependence structures. Think of the tail dependence; this is a stylized fact in finance, based on the observation that the financial quantities tend to be more correlated in periods of financial distress. The usage of copulas allows deepening the analysis to unveil other kinds of dependences on the basis of the already mentioned Sklar's Theorem.

Calibration procedures on parametric copulas are also presented under different optimization criteria. In particular, we consider both an OLS best-fit approach and an entropy-based distance minimization procedure. In so doing, we can give several insights on the nature of the dependence structure between concentration and control.

The methodological proposal is tested over a high-quality empirical sample. The explored dataset contains companies from the Italian Stock Exchange (MIB30) (see Sect. 3.2 for a detailed description of the considered data). In remarkable accord with the literature, we still get a peak for the concentration in 1, which means that most companies hold shares of just one company. We also note a second smaller peak in 2, which could state the behaviour of owning shares of two different companies. We here estimate the empirical joint distribution, and we note that the copula best describing it witnesses a fair negative dependence. The measurement of the entropic distance seems to confirm such an outcome, even if remarkable deviations with the Euclidean case occur (see Sect. 4 for discussing the obtained results).

This paper is quite close to Elliott et al. (2014), where simulations are carried out to detect the joint distributions most susceptible to the spread of contagions. However, we adopt a more general perspective here since each marginal distribution of concentration and control could rise from different network configurations. The joint distribution also allows for some degrees of freedom in the network topology. Indeed, Sklar's Theorem can be used in two directions: (1) If the marginal distributions and the copula are known—i.e. one knows the marginals and the stochastic dependence between the variables—then one can obtain the joint distribution. (2) If the marginals and the joint distribution are known, then one can infer the underlying copula linking marginals

and joint distribution. In case (1), the procedure is a simple application of the copula function to the marginals, and the obtained result is the joint distribution. In case (2), one has to implement a calibration/regression exercise, where a family of parametric copulas is pre-selected, and then the distance between the empirical and the copula-based joint distribution is minimized over the copula parameters. If copula is of a nonparametric type, then one can just compute the distance between the empirical and copula-based distribution. Case (2) is precisely our approach in this paper.

The rest of the paper is organized as follows. Section 2 provides the definition of concentration and control in the considered context of weighted financial networks. Section 3 illustrates the employed methodology and describes the analysed data. Section 4 outlines the results of the empirical analysis along with a critical discussion of them. The last section offers some conclusive remarks.

2 Technical definition of concentration and control

We consider a financial network $Net = (V, A)$, where V is the set of the N nodes which here represent companies, while $A = (a_{ij})_{i,j \in V}$ is the adjacency matrix whose entries are defined as follows:

$$\begin{aligned} a_{ij} &= 0 \text{ if the company } i \text{ does not hold shares of company } j; \\ a_{ij} &> 0 \text{ is the percentage of shares of company } j \text{ hold by company } i. \end{aligned} \quad (1)$$

By definition, network Net is weighted, and its arcs are directed. We denote by k_j^{in} and k_j^{out} the in-degree and out-degree of node j of the unweighted graph underlying Net , respectively. The quantities k_j 's are integers that count arcs without including their weights.

We adopt the definition well-outlined in Glattfelder and Battiston (2009) and define the *concentration index* for node $j \in V$ as follows:

$$s_j = \frac{(\sum_i^N a_{ij})^2}{\sum_i^N a_{ij}^2}. \quad (2)$$

We point out that s_j in (2) can be seen as the reciprocal of a disparity index. We capture how s_j in (2) works through a simple illustrative example.

Example 1 We consider a weighted network with four nodes and weights as in the four scenarios in Fig. 1.

For the equally weighted case (panel (a)), we have

$$s_j = \frac{(1/3 + 1/3 + 1/3)^2}{1/9 + 1/9 + 1/9} = 3,$$

which is exactly k_j^{in} .

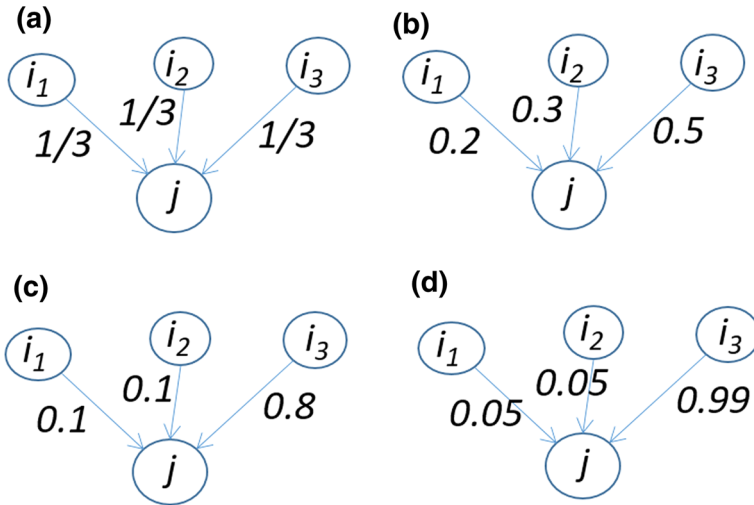


Fig. 1 Four scenarios of weights for a network with $N = 4$ nodes and arcs all incoming in the node j . Notice that (a) is the case of the equally weighted links. Panel (a) is the equally weighted case, while the other panels present different degrees of disparity among the weights (see Example 1)

Differently, if one weight is prevailing on the others (see panels (b), (c), (d)), then one has a departure from the value of k_j^{ln} as follows:

$$s_j = \frac{(0.2 + 0.3 + 0.5)^2}{(0.2)^2 + (0.3)^2 + (0.5)^2} = \frac{1}{0.38} = 2.63$$

$$s_j = \frac{(0.1 + 0.1 + 0.8)^2}{(0.1)^2 + (0.1)^2 + (0.8)^2} = \frac{1}{0.66} = 1.51$$

$$s_j = \frac{(0.005 + 0.005 + 0.99)^2}{(0.005)^2 + (0.005)^2 + (0.99)^2} = \frac{1}{0.9801} = 1.0203$$

The trivial corner case of only one incoming link in j —not shown in Fig. 1—leads to the minimum level of concentration index:

$$s_j = \frac{(1 + 0 + 0)^2}{(1)^2 + (0)^2 + (0)^2} = \frac{1}{1} = 1.$$

We notice that the distributions detected in the literature show a power-law dependence of the concentration on the in-degree (see Glattfelder and Battiston 2009). However, the quoted paper shows that even if the tails follow a power law, the bulk of the distribution is far from it; this is in line with the plots reported in Pecora and Spelta (2015).

We now present the control, according to the definition in Glattfelder and Battiston (2009), Pecora and Spelta (2015). Given a node $i \in V$, the quantity h_i is the control index of i , where

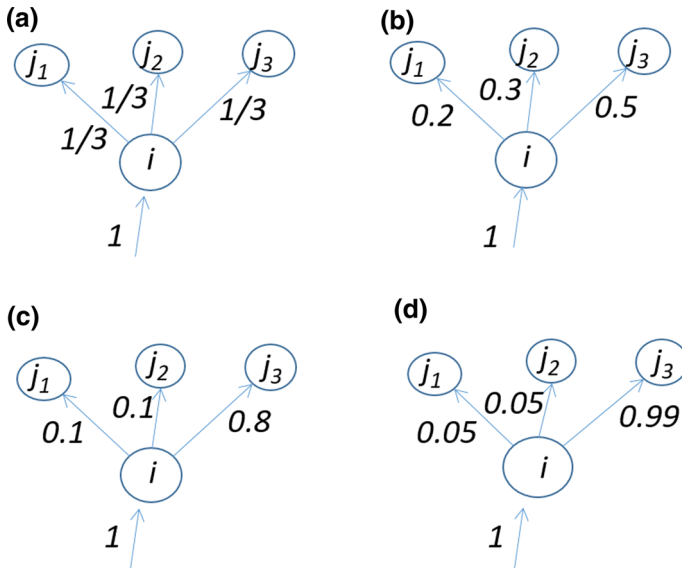


Fig. 2 Four scenarios of weights for a network containing a subnetwork with four nodes and arcs all outgoing from the node j . Notice that (a) is the case of the equally weighted links. There is only one incoming link, with a unitary weight (see Example 2)

$$h_i = \sum_{j=1}^N \frac{a_{ij}^2}{\sum_{l=1}^N a_{il}^2}. \quad (3)$$

It is interpreted as the effective number of stocks controlled by shareholder i . In fact, if $\frac{a_{ij}^2}{\sum_{l=1}^N a_{il}^2}$ is close to 1, then company i has the most part of the shares of company j .

Also in this case, we illustrate h_i in (3) through an example.

Example 2 We consider a weighted network containing a subnetwork with four nodes and weights with values as in Fig. 1 but with a reverted direction (see the four scenarios in Fig. 2).

A simple computation gives $h_i = 0.11$ (panel (a)), $h_i = 0.38$ (panel (b)), $h_i = 0.66$ (panel (c)) and $h_i = 0.9801$ (panel (d)).

3 Data and methods

This section is devoted to illustrating the techniques used for detecting the cross-shareholding structure of the considered financial network. Moreover, it also describes the employed dataset.

3.1 Methodology

The starting point is an empirical distribution of concentration and level of control for a given set of companies. Details on the selected datasets will be reported in Sect. 3.2.

We denote by \mathcal{S} and \mathcal{H} the random variables of concentration and level of control, respectively. They follow the empirical distributions of $(s_j : j = 1 \dots, N)$ and $(h_i : i = 1 \dots, N)$ in formulas (2) and (3), respectively.

The cumulative marginal distribution functions of \mathcal{S} and \mathcal{H} will be denoted by $F_{\mathcal{S}} : \mathbb{R} \rightarrow [0, 1]$ and $F_{\mathcal{H}} : \mathbb{R} \rightarrow [0, 1]$, respectively.

The empirical joint distribution of \mathcal{S} and \mathcal{H} is $F_{\mathcal{S}, \mathcal{H}}^{\text{emp}} : \mathbb{R}^2 \rightarrow [0, 1]$.

The present analysis is organized in two steps. Copulas intervene in each of them to describe the stochastic dependence between concentration and control.

The concept of copulas is particularly suitable for our purpose. Indeed, a bivariate copula $C : [0, 1]^2 \rightarrow [0, 1]$ is a special function able to describe the dependence structure between two random variables through the classical Sklar's Theorem. We rewrite it here by conveniently adapting our notation.

Theorem 1 (Sklar's Theorem) *Let $F_{\mathcal{S}, \mathcal{H}}$ be the joint distribution function of a couple of random variables $(\mathcal{S}, \mathcal{H})$. With an intuitive notation, the marginal distribution functions are $F_{\mathcal{S}}$ and $F_{\mathcal{H}}$. Then, one can find a copula $C : [0, 1]^2 \rightarrow [0, 1]$ such that, for each $(s, h) \in \mathbb{R}^2$,*

$$F_{\mathcal{S}, \mathcal{H}}(s, h) = C(F_{\mathcal{S}}(s), F_{\mathcal{H}}(h)). \tag{4}$$

When $F_{\mathcal{S}}, F_{\mathcal{H}}$ are continuous, then C satisfying (4) is unique. Conversely, if C is a copula and $F_{\mathcal{S}}, F_{\mathcal{H}}$ are distribution functions, then $F_{\mathcal{S}, \mathcal{H}}$ in (4) is a bidimensional joint distribution function with marginal distribution functions $F_{\mathcal{S}}$ and $F_{\mathcal{H}}$.

In accord with (4), we denote by $F_{\mathcal{S}, \mathcal{H}}^C : [0, 1]^2 \rightarrow [0, 1]$ the joint distribution function coming out from the application of Sklar's Theorem 1 with a generic copula C .

By imposing specific copulas functions C , Theorem 1 assures that different types and natures of stochastic dependence can be stated between two random variables. Therefore, as we will see, we here deal with some selected copulas to describe the nature of the connection between concentration and control.

Before describing the details of the methodological approach, we introduce the considered copulas.

We consider two prominent cases: nonparametric copulas and parametric ones. The former case is used to provide information on the similarity between some peculiar cases of stochastic dependence and the real empirical dependence structure of concentration and control; the latter case allows to grasp information from the calibrated parameters of some meaningful copulas, always in the context of the similarity between the copula-based dependence and the empirical one.

For what concerns the nonparametric copulas, we consider the independence case of the product copula

$$C_I(u, v) = uv \tag{5}$$

and the Frechet bounds, which realize the maximum (upper bound, copula denoted by C_{UF}) and minimum (lower bound, copula denoted by C_{LF}) possible dependence

between the considered random variables. For the convenience of the reader, we recall the definition of the Frechet bounds:

$$C_{LF}(u, v) = \max\{u + v - 1, 0\}, \quad C_{UF}(u, v) = \min\{u, v\}. \tag{6}$$

In the parametric case, we consider three important instances of Archimedean copulas, dependent on a parameter θ :

- Gumbel Archimedean copula

$$C_G(u, v) = \exp[-((-\ln(u))^\theta + (-\ln(v))^\theta)^{1/\theta}], \quad \theta \in [1, +\infty) \tag{7}$$

- Frank Archimedean copula

$$C_F(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right], \quad \theta \neq 0 \tag{8}$$

- Clayton Archimedean copula

$$C_C(u, v) = [\max\{u^{-\theta} + v^{-\theta} - 1, 0\}]^{-1/\theta}, \quad \theta \in [-1, 0) \cup (0, +\infty) \tag{9}$$

The stochastic dependence described by the Archimedean copulas above is of a different type.

The Gumbel case is associated with asymmetric right-tail dependence, whose entity is mainly driven by the parameter θ . Frank copula can capture both positive and negative dependence cases and does not generally describe tail dependence. The Clayton case is quite similar to the Gumbel one, with tail dependence. However, Clayton copulas describe dependence on the left tail, and also in this case a more detailed specification can be derived from the analysis of the parameter θ .

We now have all the instruments for proceeding with the analysis.

To proceed, we preliminarily introduce a discretization of \mathbb{R} and \mathbb{R}^2 on the basis of the empirical sample. In detail, without losing generality, we can assume that \mathcal{S} and \mathcal{H} are discrete and take values in two discrete sets $\Gamma_{\mathcal{S}}$ and $\Gamma_{\mathcal{H}}$, respectively.

We face the problem of assessing the characteristics of the dependence structure between concentration and control. To this aim, we introduce suitable distance measures between the empirical joint distribution $F_{\mathcal{S}, \mathcal{H}}^{\text{emp}}$ and the one obtained by using the considered copulas. We propose two distance measures. The first one is the classical distance of Euclidean type, and it is defined as follows:

$$d_{\text{Euc}}(F_{\mathcal{S}, \mathcal{H}}^{\text{emp}}, F_{\mathcal{S}, \mathcal{H}}^C) = \sqrt{\sum_{(s, h) \in \Gamma_{\mathcal{S}} \times \Gamma_{\mathcal{H}}} (F_{\mathcal{S}, \mathcal{H}}^{\text{emp}}(s, h) - F_{\mathcal{S}, \mathcal{H}}^C(s, h))^2}. \tag{10}$$

The second one is the absolute deviation between the entropies of the two joint distributions, i.e.

$$d_{\text{Ent}}(F_{\mathcal{S}, \mathcal{H}}^{\text{emp}}, F_{\mathcal{S}, \mathcal{H}}^C) = \left| H(F_{\mathcal{S}, \mathcal{H}}^{\text{emp}}) - H(F_{\mathcal{S}, \mathcal{H}}^C) \right|, \tag{11}$$

where $H(F_{S,\mathcal{H}}^{\text{emp}})$ and $H(F_{S,\mathcal{H}}^C)$ are the entropies of $F_{S,\mathcal{H}}^{\text{emp}}$ and $F_{S,\mathcal{H}}^C$, and they are defined as follows:

$$H(F_{S,\mathcal{H}}^{\text{emp}}) = \sum_{(s,h) \in \Gamma_S \times \Gamma_{\mathcal{H}}} F_{S,\mathcal{H}}^{\text{emp}}(s,h) \ln F_{S,\mathcal{H}}^{\text{emp}}(s,h) \tag{12}$$

and

$$H(F_{S,\mathcal{H}}^C) = \sum_{(s,h) \in \Gamma_S \times \Gamma_{\mathcal{H}}} F_{S,\mathcal{H}}^C(s,h) \ln F_{S,\mathcal{H}}^C(s,h). \tag{13}$$

Exploring distances in (10) and (11) provides several insights on the informative content of the empirical joint distribution of concentration and control in terms of their stochastic dependence. Indeed, Euclidean distance gives information on the pointwise general difference between the joint distribution functions, while the entropy is informative on the closeness of the overall shape of the distributions.

In the case of nonparametric copulas, we simply apply Theorem 1 to obtain the joint probability distribution $F_{S,\mathcal{H}}^C$ and measure the distances with the empirical distributions, according to (10) and (11). In the parametric copulas, the parameter θ is calibrated by minimizing the distance between the empirical joint distribution and the one associated with the parametric copulas so that the following problems are solved in all the considered cases of Archimedean copulas (we denote by θ^* the calibrated parameter):

$$\theta^* \in \operatorname{argmin}_{\theta} d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C) \tag{14}$$

and

$$\theta^* \in \operatorname{argmin}_{\theta} d_{\text{Ent}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C), \tag{15}$$

where $C = C_G, C_F, C_C$ and θ are taken from the variation range of the specific considered copula.

In the empirical experiments, a discussion of the value of θ^* will be carried out below for all the cases of the considered Archimedean copulas.

As conclusive remarks in this section, we stress that identifying the unique copula granting the validity of Sklar’s Theorem is a challenging task, which often does not have a definitive response. The point is that copulas can be of different shapes; moreover, we are reasonably sure that several families of copulas should still be discovered and formalized by scholars.

Our approach is then different. We take a parametric copula explaining some properties of the stochastic dependence between the considered variables, and we identify the parameters for which the distance between the empirical joint distribution and the copula-based one is minimized. In this respect, we define a copula-based joint distribution. The employed distance has a relevant, informative content. Thus, we consider two types of distance measures—the Euclidean one and the entropy—to gain insights on the dependence between the variables. The joint analysis of the calibrated parameters and the distance measure provides a clear view of the connection between concentration and control. As an additional analysis, we also take nonparametric copulas describing extreme cases of dependence and independence—the Frechet bounds and the product copula, respectively—and we assess how the joint distribution obtained

Table 1 Main statistical indicators related to concentration and control

	Mean	SD	Skewness	Kurtosis
Concentration	1.2721	0.4293	1.6566	5.3302
Control	1.0000	0.4271	0.3735	4.0847

through them is close to the empirical one, still under the guide of Euclidean and entropy measure.

3.2 The dataset

The dataset that serves as the case study gathers the cross-ownership in the Italian Stock Exchange (MIB30) on DATA, and it has already been explored in Rotundo and D'Arcangelis (2010), Cerqueti et al. (2018). It contains the cross-holdings among listed firms in the Milan Stock Market-MTA segment. It has been cross-checked through several databases: the Bureau Van Dijk, CONSOB, Bankscope, ISIS, AIDA, Datas-tream. A total of 247 stocks are listed. However, many of them are neither owned by other companies listed in the MIB30, nor play an active role in buying shares from other companies. From the perspective of the network, we remove the insulated nodes (i.e. if column i and row i are both empty, they are removed from the adjacency matrix, which diminishes its dimension). After this first pruning, the matrix A has 158 rows/columns.

We present the main descriptive statistics in Table 1 to give a clear idea of the considered dataset.

It is worth pointing out that the value of the mean of control is remarkably in line with the values detected on several stock market indexes (see, e.g. Glattfelder and Battiston 2009). Actually, the standard deviation allows seeing quite a dispersion around the mean. The skewness takes well into account the asymmetry of the distributions. The values of kurtosis add further evidence that both the distributions are quite far from the Gaussian one.

The histograms of the network weights are reported in Fig. 3. Such a figure shows that the vast majority of the companies have small ownership values.

4 Results and discussion

The analysis is carried out in a stepwise form. First, we compute the empirical marginal and joint distributions of concentration and level of control. Then, we compute the distances in (10) and (11) by considering the parametric and nonparametric cases introduced in Sect. 3.1. In the parametric case, we numerically solve problems 14 and 15 and discuss the obtained findings.

4.1 Empirical distributions of concentration and control

First of all, we introduce the empirical marginal distribution of the concentration S .

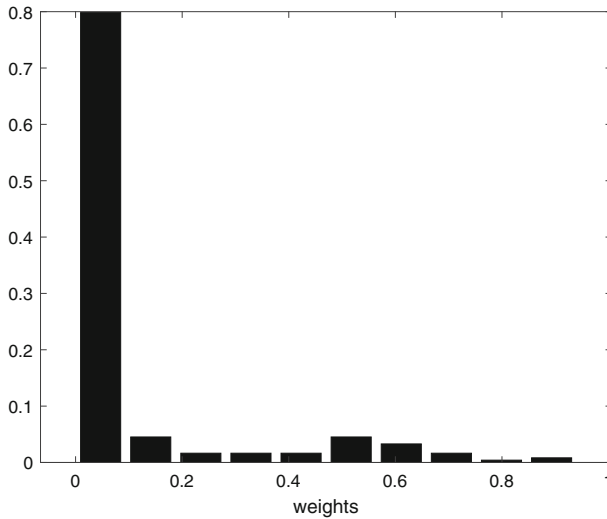


Fig. 3 Distribution of the weights of the network. The vast majority shows quite small percentages of ownership

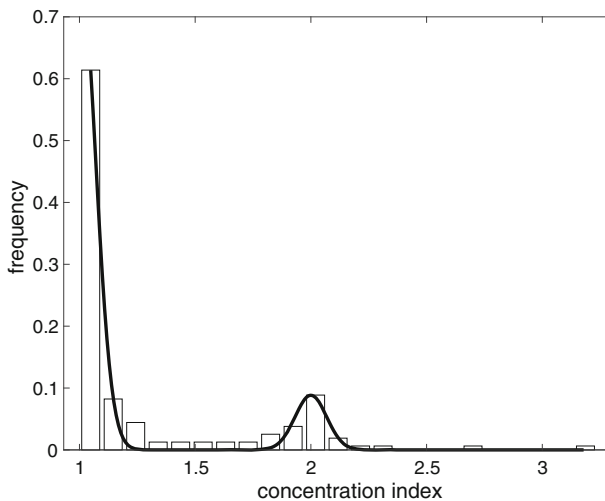


Fig. 4 Histogram of the concentration index S . We juxtapose the best fit through the function in (16). The calibrated parameters are reported in the text. Notice the bimodality of the distribution

Figure 4 shows the histogram of the concentration. There is evidence of a higher peak in 1, followed by a small peak around 2.

The peak around 1 has also been detected in the dataset studied in Glattfelder and Battiston (2009). The tendency to have the ownership of just one other company could be due to the financial policy management of the company. For instance, the formal separation of a financial sector from a bigger company or the presence of ultimate owners, like for IFI PRIV (privileged shares of the Agnelli family) with respect to

IFIL (the society financing the leading companies of the Agnelli family), which are FIAT car industries and Juventus football team. However, in Glattfelder and Battiston (2009) the peak around 2 is not observed, so the intra-relationship in the MIB30 are different from the cross-relationships among countries. The best fit to the empirical distribution is provided by a mixture of two Gaussians, as follows:

$$p(x) = \sum_{i=1}^2 a_i e^{-\left(\frac{x-\bar{\mu}_i}{\sigma_i}\right)^2}. \quad (16)$$

We set $\bar{\mu}_1 = 1$ and $\bar{\mu}_2 = 2$ in accord to the visual inspection of data. This choice allows moving from a best-fit procedure with six parameters to the best fit with four parameters, with consequent improvement in the goodness of fit. The calibrated coefficients (with 95% confidence bounds) are: $a_1 = 0.786(0.7346, 0.8375)$, $a_2 = 0.0882(0.05855, 0.1178)$, $\sigma_1 = 0.09767(0.0895, 0.1058)$, $\sigma_2 = 0.09657(0.05993, 0.1332)$. We have an excellent outcome for the goodness-of-fit parameters: $SSE = 0.003358$, $R^2 = 0.9905$, $Adj. R^2 = 0.989$, $RMSE = 0.01329$.

The company with the highest concentration is GABETTI PROPERTY SOLUTIONS. It sells its shares to the highest number of other companies (ten of them), which plays a crucial role in increasing its concentration. The company works in the real estate market, and its shares are owned by companies most in buildings and energy sectors.¹

The companies with the lowest concentration are all with $s_i = 1$, which do not sell their shares in this market, so they hold 100% of their own ownership.²

We now introduce the empirical marginal distribution of the control variable \mathcal{H} .

As in the case of concentration, also the index of control shows a very pronounced peak in 1 (see Fig. 5).

Figure 5 shows the best-fit exercise with the mixture of three Gaussians, i.e.:

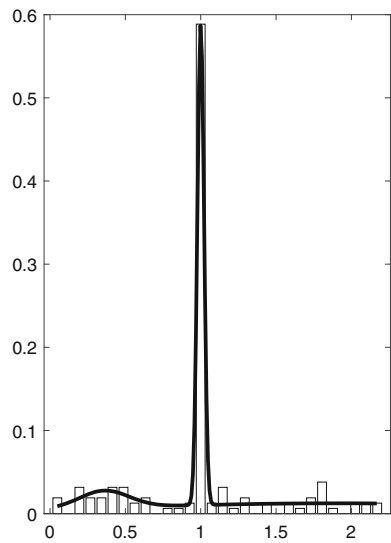
$$p(x) = \sum_{i=1}^3 a_i e^{-\left(\frac{x-\bar{\mu}_i}{\sigma_i}\right)^2}. \quad (17)$$

The best-fit exercise gives the following calibrated coefficients (with 95% confidence bounds): $a_1 = 0.5783(0.5548, 0.6018)$, $a_2 = 0.01245(0.005994, 0.01891)$,

¹ A2A AEM, ASTALDI, CEMENTIR, COBRA, DIGITAL BROS, EEMS ITALIA SPA, EL.EN. SPA, ERG, EUROFLY, FIDIA.

² 66 companies: A2A AEM, ACEGAS, ACQUE POTABILI, AEROPORTO FIRENZE, ALITALIA, ARENA, ASCOPIAVE, ASTALDI, ATLANTIA, BANCA MPS, BASTOGI, BCA IFIS, BCA ITALEASE, BCA POP INTRA, BCA POP MILANO, BCA POP SPOLETO, BCA PROFILO, BCO DESIO E BRIANZA, BCO SARDEGNA, BEGHELLI, BIALETTI, BOLZONI, BREMBO, BRIOSCHI, BUONGIORNO, CAD IT, CAIRO COMM, CARRARO, CATTOLICA ASS, CDC, CEMBRE, CENT LATTE TORINO, CIR, CLASS, CRED BERGAMASCO, CREMONINI, DAMIANI, DANIELI, DATASERVICE, DE LONGHI, DIASORIN SPA, DMT DIDITAL TECH, EDISON SPA, EMAK SPA, ENERTAD SPA, ENGINEERING - INGEGNERIA INFORMATICA - SPA, ENIA SPA, ERGO PREVIDENZA SPA, EXPRIVIA, FIAT, FILATURA DI POLLONE, FINMECCANICA, GEFRAN, GEOX, GRANITIFIANDRE, GREENERGYCAPITAL, GRUPPO COIN, IGD, IMMSI, INTERPUMP SPA, INVESTIMENTI E SVILUPPO SPA, IPI SPA, IT HOLDING, ITALCEMENTI FABBRICHE RIUNITE, ITALMOBILIARE SPA, ITWAY.

Fig. 5 Histograms of the index of control. Also in this case, we apply the best fit curve obtained through the function in (16). The calibrated parameters are given in the text. There is a clear predominance of one specific realization of the index, which presents a clearly unimodal distribution



$a_3 = 0.02065(-0.002128, 0.04342)$, $c_1 = -0.0312(-0.05275, -0.009649)$ $c_2 = 2.056(-4.096, 8.209)$, $c_3 = 0.2333(-0.08779, 0.5544)$, while we have set $\bar{\mu}_1 = 1$, $\bar{\mu}_2 = 1.9$, $\bar{\mu}_3 = 0.36$. Also in this case the goodness of fit is statistically satisfactory, with SSE: 0.002559, R-square: 0.9921, Adjusted R-square: 0.9904, RMSE: 0.01033.

The company with the highest control is ALLEANZA. It buys shares from the highest number of other companies. This high activity can be due to the financial nature of the company since it is an insurance one. On the opposite, the company with the lowest control is FILATURA DI POLLONE, which buys very small shares from three other companies, hence having a low control value.

We can define their empirical joint distribution from the marginal empirical distributions of concentration and control. Figure 6 provides a graphical representation of the joint probability distribution surface.

4.2 Distances from the empirical distribution

We now present the distance measures in (10) and (11) between the empirical distribution and the ones obtained through copulas by using Theorem 1.

We start from the distance measure d_{EUC} in (10).

In the nonparametric case, we have that $d_{EUC}(F_{S,\mathcal{H}}^{emp}, F_{S,\mathcal{H}}^C)$ is 3.933664, 2.405063 and 10.037975 when $C = C_I$ in 5, $C = C_{LF}$ and $C = C_{UF}$ in (6), respectively. This outcome suggests that the dependence structure of concentration and control is much closer to the perfect negative correlation rather than to the perfect positive one. This behaviour is in line with the results obtained on the in-degree (see Rotundo and D’Arcangelis 2010). A cross-check on the sample shows that banks and financial institutions collected in the dataset are the ones which are most holding shares of several other companies. Actually, such a strategy represents a way to provide financial support to a company by banks and financial institutions. From a different perspective,

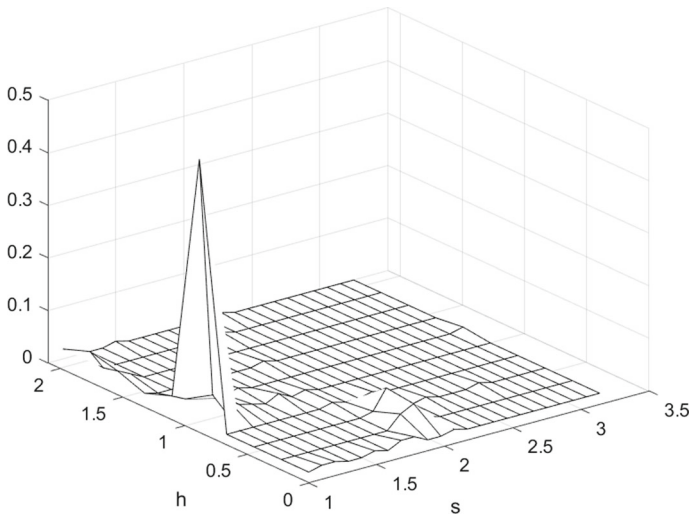


Fig. 6 Surface of the joint empirical distribution between concentration and control. Notice the peak at a low level of concentration and middle level of control

the shares of financial institutions are not bought by many different other companies. Industrial companies most show links to companies which are relevant to their business (as an example, Finmeccanica—an industrial group active in high technologies, flight, space, and defence—is linked to Ansaldo—which is among the major worldwide producers of energy plants) or, alternatively, that are offsprings (ENI and SAIPEM or ENI and SNAM Rete Gas) or that were created to separate the financial part from the core business (IFIL and Juventus or IFIL and Fiat—see Fig. 1 in Rotundo and D’Arcangelis 2010). In this sense, the behaviour is not strategic in terms of strategic investments, but it is driven by production, business and industrial needs.

Figure 7 presents the Euclidean distance (10) between the empirical joint distribution and the ones associated to the parametric Archimedean copulas in (7), (8) and (9) as the parameter θ varies. In particular, the upper panel presents the Gumbel copula (7), the middle one is the Frank copula (8), and the lower panel is the Clayton copula case in (9).

For the Gumbel copula, we have a minimum in $\theta^* = 1$. In this case, $d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C) = 3.934$ for $C = C_G$ in (7). The case of Frank copula with negative parameter exhibit a minimum $\theta^* = -18.1$, with $d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C) = 1.933$ when $C = C_F$ in (8). Still, in the Frank case—but for positive values of the parameter θ —the smallest values of the distance are taken around zero, even if there is no minimum because 0 does not belong to the definition set. However, Fig. 7 shows that $d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C) = 4.008$ when $\theta = 0.1$. Finally, the Clayton case with negative values of the parameter shows a minimum in $\theta^* = -1$ with a value of distance $d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C)$ given by 2.405. Also, in the Clayton case with positive parameters, we have the smallest distance values for θ close to zero. Even if there is no minimum, we have $d_{\text{Euc}}(F_{S,\mathcal{H}}^{\text{emp}}, F_{S,\mathcal{H}}^C) = 137.2$ for $\theta = 0.1$, with $C = C_C$.

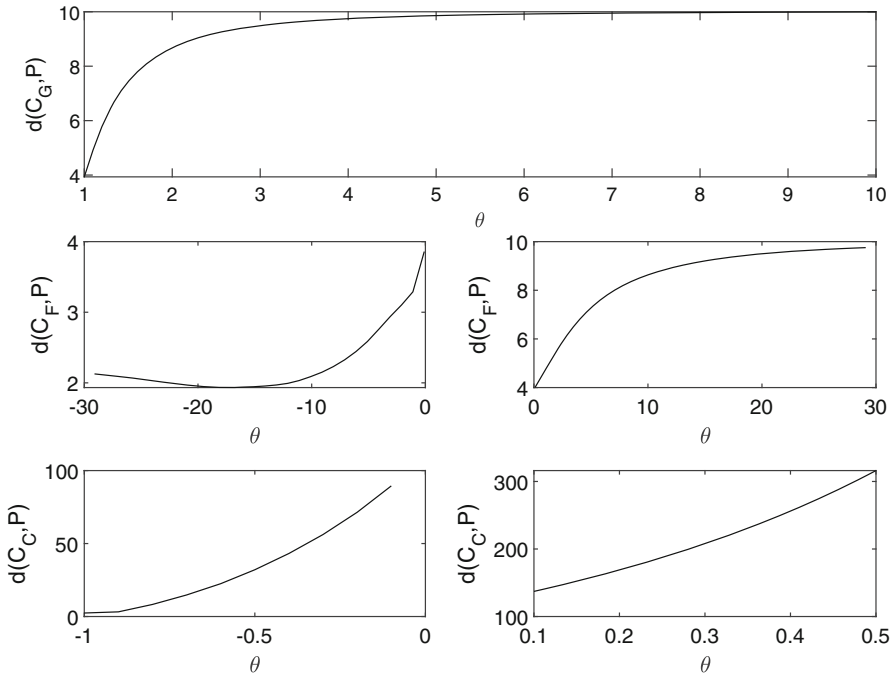


Fig. 7 Distance $d_{\text{Euc}}(F_{S, \mathcal{H}}^{\text{emp}}, F_{S, \mathcal{H}}^C)$, when $C = C_G$ in (7) (upper figure), $C = C_F$ in (8) (middle figure) and $C = C_C$ in (9) (lower figure). The two panels in the middle and lower figures are separated to emphasize the domain of the related parameters

From the perspective of copulas, the increasing values of the distance as theta increases witness that there is no tail dependence. In fact, the tail dependence of the Gumbel and the Clayton copulas would increase as theta increases, and graphical representation suggests that the distance between such copulas and the empirical joint distribution increases with θ . The Frank copula is the only one (among the chosen ones) which does not represent the tail dependence. The fact that the Frank copula is achieving the minimum (among all the copulas as θ changes) confirms the absence of tail dependence. Since the overall minimum is achieved for a negative value of θ , namely $\theta^* = -18.1$ for the Frank copula, there is evidence of a prevalence of negative dependence among the marginals (Frank is negative dependence for $\theta < 0$ and positive one for $\theta > 0$). However, such a negative dependence is “fair”, in the sense that the distance increases for $\theta < \theta^*$. This outcome confirms what was already found in the analysis of the nonparametric copulas. Moreover, it also finds confirmation in the negative correlation obtained by computing the Pearson correlation coefficient ρ between \mathcal{S} and \mathcal{H} , which is $\rho = -0.426877816804739$.

Let us now move to the analysis of the entropy distance measure d_{Ent} in (11).

First of all, we have that $H(F_{S, \mathcal{H}}^{\text{emp}}) = 1.368666$.

We can remark that the entropy of the case study $H(F_{S, \mathcal{H}}^{\text{emp}})$ is in the middle of the range of variation of the entropy in our case—we calculate that the maximum value

for the entropy is 5.5, being $N = 23 \times 11 = 253$. Therefore, the system shows some degree of organization, although it is still far from a globally disordered system.

The entropies in the nonparametric case are $H(F_{S,\mathcal{H}}^C) = 1.96713208$ when $C = C_I$ in (5), $H(F_{S,\mathcal{H}}^C) = 1.09578522$ and $H(F_{S,\mathcal{H}}^C) = 1.70745533$ when $C = C_{LF}$ and $C = C_{UF}$ in (6), respectively. Coming to the differences, we have $d_{Ent}(F^{\text{emp}}, F^{C_P}) = 0.59846570$, $d_{Ent}(F^{\text{emp}}, F^{C_{LF}}) = 0.27288117$, $d_{Ent}(F^{\text{emp}}, F^{C_{UF}}) = 0.33878895$.

Therefore, the joint distribution obtained through the Lower Frechet copula is the closest one to the empirical distribution, according to distance d_{Ent} in (11), on the basis of the comparison of the values $d_{Ent}(F^{\text{emp}}, F^{C_P})$, $d_{Ent}(F^{\text{emp}}, F^{C_{LF}})$ and $d_{Ent}(F^{\text{emp}}, F^{C_{UF}})$. This confirms the prevalence of a negative correlation between concentration and control, even when we consider the overall shape of the distribution.

Parametric copulas are graphically presented in Fig. 8.

In the Gumbel case of $C = C_G$, there is a maximum for the distance in $\theta = 1.5600$, $d_{Ent}(F^{\text{emp}}, F^G) = 0.7134$. There is no minimum. However, for $\theta = 20.0000$, $d_{Ent}(F^{\text{emp}}, F^G) = 0.3809$.

On the Frank copula: there is no minimum and no maximum on the negative values of θ . For the positive values of θ , there is a maximum for $\theta = 5.5100$ given by $d_{Ent}(F^{\text{emp}}, F^F) = 0.7391$.

On the Clayton copula, there is a maximum for the Clayton copula in $\theta = 5.2300$, with $d_{Ent}(F^{\text{emp}}, F^C) = 0.7281$. For the negative values of θ , we have no minimum. However, for $\theta = -0.9$, we have $d_{Ent}(F^{\text{emp}}, F^C) = 0.039$.

In the Gumbel case, the behaviour of the distance d_{Ent} is radically different to the one of d_{Euc} , with a small deviation for $\theta \rightarrow \infty$. This means that the empirical and copula-based distributions have similar probabilistic shapes when a high dependence on the tail is considered.

Similar comments hold for the Clayton copula: low values of the distances are achieved for $\theta \rightarrow \infty$, which means similarity between the disorder generated by the investigated distributions in the case of strong positive dependence. We recall that for $\theta \rightarrow \infty$, the Clayton copula tends to the Upper Frechet, where the dependence is on the left tail (the negative one).

Both the Clayton and the Gumbel copulas confirm that strong dependence on the tails leads the empirical distribution as close as possible—in terms of entropy measures—to those generated by the copulas. Moreover, Clayton also allows assessing a minimum of the entropy distance for θ , close to -1 , that would be associated with the Lower Frechet case. Then, the investigated joint distributions tend to have similar shapes also in the presence of a strong negative correlation between concentration and control.

The Frank copula shows a maximum of the entropic distance for values of θ positive but close to 0, and the smallest one for $\theta \rightarrow -\infty$. Also the case $\theta \rightarrow +\infty$ gives small distance values. Remembering that positive (negative) values of the θ in the Frank copulas stand for positive (negative) dependence, we get that strong positive and negative dependence between concentration and control leads to a shape similarity

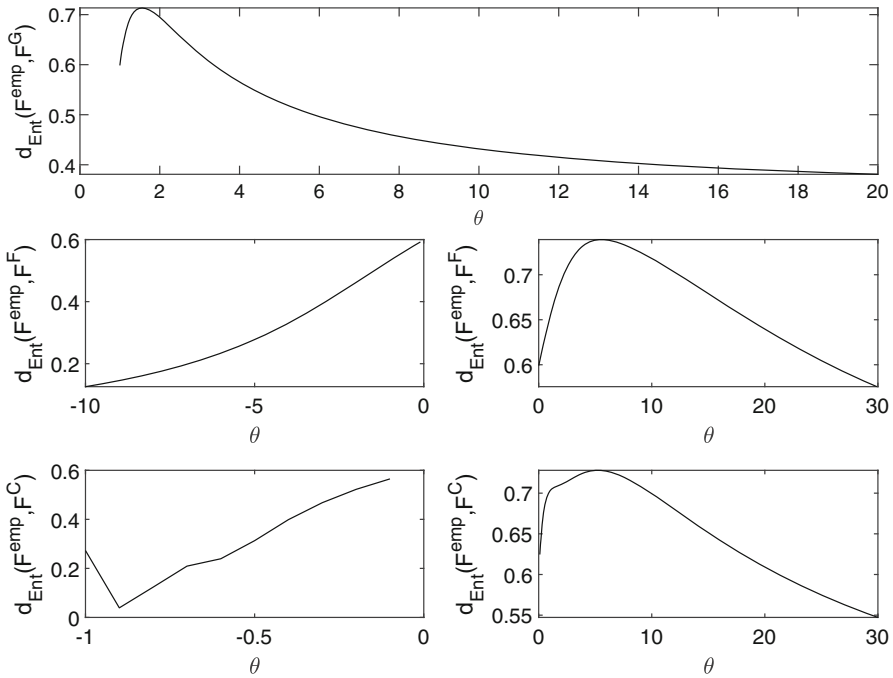


Fig. 8 Distance $d_{Ent}(F_S^{\text{emp}}, F_S^C)$, when $C = C_G$ in (7) (upper figure), $C = C_F$ in (8) (middle figure) and $C = C_C$ in (9) (lower figure). Also in this case, the two panels in the middle and lower figures are taken as distinct entities to emphasize the domain of the related parameters

with the empirical distribution—with a preference for the negative dependence. We also observe that such dependence is “fair”, not on the tails.

Looking at the overall analysis, we can conclude that the empirical joint distribution of concentration and control seems to come from a structure of negative dependence.

5 Conclusions

The present paper examined the stochastic dependence induced by cross-shareholdings in a stock market. In line with the literature, the indices of concentration and control describe the effectiveness of the number of shares sold (concentration) or bought (control). In the case study that serves to outline the analysis, some features already found in literature emerged, like the peak for the concentration at 1, meaning that many companies do not sell their shares in the examined market, leaving room for externalities. We are interested in understanding the type of stochastic dependence that gives rise to the similarity between the empirical distribution and a theoretically obtained one. The proposed empirical instance employs high-quality data coming from the Italian Stock Exchange. To increase the informative content of the analysis, we present the cases of Euclidean distance and entropic distance between the investigated distribution and the empirical one. The former case provides information about the averaged similarity

of the terms of the distribution. At the same time, the latter deals with the likeness of the disorder related to the overall distributions. We relied on both parametric and non-parametric copulas to model the dependence structure. What emerges is a substantial difference when dealing with entropic and Euclidean distance, even if it seems that the negative dependence between concentration and control provides a more satisfactory description of the empirical sample.

In this context, fitting the empirical distribution with a continuous function pursues two main targets. On one side, the empirical marginal distributions are well-fitted by some mixtures of Gaussians. Therefore, one can reasonably employ such continuous functions for applying Sklar's Theorem, with the relevant benefit of not being constrained to take only the observed values for describing the empirical distributions. This is exactly the approach followed in the present paper. On the other side, the considered functions can be suitably perturbed in the light of modelling modifications of the empirical distributions of concentration and control. The perturbed functions can then be used to apply Sklar's Theorem to derive information on new configurations of such variables. The interest in working with a joint distribution different from the actual one raises by the fact that regulations and/or economic factors may impact the overall topology of the network. In fact, even well-consolidated financial markets may change their topology. This second aspect is not covered in the current version of the paper—it is well-beyond our target. However, we plan to develop it in future studies.

Our model presents two main limitations: on one side, the analysis has been carried out on the basis of some specific families of copulas. This is restrictive in that there is an endless debate on such methodological instruments with a long list of classes of copulas. However, the considered instances are particularly meaningful; they describe paradigmatic cases of stochastic dependence, which are suitable for modelling the correlation between financial variables; on the other side, empirical results are related to the special case of the Italian Stock Exchange. However, the proposed method is versatile; it can be employed in other empirical samples.

Interestingly, since concentration and control are an extension of in- and out-degree, this study can be seen as an extension of the concept of assortativity, which was first defined as the correlation among the node degrees. Indeed, copulas do not limit to the empirical dataset but consider a broader range of dependence structures. This challenging extension of our study is already on our research agenda.

In terms of network topology, we point out that the raw links give their structure. The network has already been detected as close to the scale-free one (see Fig. 1 in Rotundo and D'Arcangelis 2010). However, the analysis of mesoscale and communities is out of the scope of the present paper. This is another challenging task that deserves a focused research paper.

Finally, we also stress that papers examining only the out- and in- degree (see, e.g. Garlaschelli et al. 2005) and the studies on the minimum spanning tree calculated only on the correlation networks (see, e.g. Bonanno et al. 2003; Onnela et al. 2003) suggest the restriction of market investments to be the common features of markets in recession times. This is left for future studies on time-evolving networks.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bonanno G, Caldarelli G, Lillo F, Mantegna RN (2003) Topology of correlation-based minimal spanning trees in real and model markets. *Phys Rev E* 68(4):046130
- Cerqueti R, Rotundo G, Ausloos M (2018) Investigating the configurations in cross-shareholding: a joint copula-entropy approach. *Entropy* 20(2):134
- Cinelli M (2019) Generalized rich-club ordering in networks. *J Complex Netw* 7(5):702–719
- Cinelli M, Ferraro G, Iovanella A, Rotundo G (2019) Assessing the impact of incomplete information on the resilience of financial networks. *Ann Oper Res* 299:721–745
- Croci E, Grassi R (2014) The economic effect of interlocking directorates in Italy: new evidence using centrality measures. *Comput Math Organ Theory* 20(1):89–112
- D'Errico M, Grassi R, Stefani S, Torriero A (2009) Shareholding networks and centrality: an application to the Italian financial market. In: *Networks, topology and dynamics*, Springer, Berlin, pp 215–228
- Drago C, Ricciuti R, Santella P (2015) An attempt to disperse the Italian interlocking directorship network: analyzing the effects of the 2011 Reform. SSRN 2580700
- Elliott M, Golub B, Jackson MO (2014) Financial networks and contagion. *Am Econ Rev* 104(10):3115–53
- Gao J, Barzel B, Barabási AL (2016) Universal resilience patterns in complex networks. *Nature* 530(7590):307–312
- Garlaschelli D, Battiston S, Castri M, Servedio VD, Caldarelli G (2005) The scale-free topology of market investments. *Physica A* 350(2–4):491–499
- Glattfelder JB, Battiston S (2009) Backbone of complex networks of corporations: the flow of control. *Phys Rev E* 80:036104
- Haldane AG (2013) Rethinking the financial network. In: *Fragile stabilität–stabile fragilität*, Springer VS, Wiesbaden, pp 243–278
- Joe H (1997) *Multivariate models and multivariate dependence concepts*. CRC Press, Cambridge
- Markose S, Giansante S, Shaghghi AR (2012) Too interconnected to fail financial network of US CDS market: topological fragility and systemic risk. *J Econ Behav Organ* 83(3):627–646
- Nelsen RB (2007) *An introduction to copulas*. Springer Science & Business Media
- Onnela JP, Chakraborti A, Kaski K, Kertesz J (2003) Dynamic asset trees and Black Monday. *Physica A* 324(1–2):247–252
- Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86(14):3200
- Pecora N, Spelta A (2015) Shareholding relationships in the Euro Area banking market: a network perspective. *Physica A* 434:1–12
- Rotundo G, D'Arcangelis AM (2010) Ownership and control in shareholding networks. *J Econ Interac Coord* 5(2):191–219
- Rotundo G, D'Arcangelis AM (2014) Network of companies: an analysis of market concentration in the Italian stock market. *Qual Quant* 48(4):1893–1910
- Rotundo G, D'Arcangelis AM (2010) Network analysis of ownership and control structure in the Italian Stock market. *Adv Appl Stat Sci* 2(2):255–273
- Sklar M (1959) Fonctions de repartition en dimensions et leurs marges. *Publ Inst Stat Univ Paris* 8:229–231
- Varela LM, Rotundo G (2016) Complex network analysis and nonlinear dynamics. In: *Complex networks and dynamics*, Springer, Cham, pp 3–25

White LJ (2014) The basics of “too big to fail”. http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2405433

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.