A Matrix-Variate *t* Model for Networks

Monica Billio[1]*, Roberto Casarin[1], Michele Costola[1] and Matteo Iacopini[2,3]

[1] Department of Economics, Ca' Foscari University of Venice, Venice, Italy, [2] Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [3] Tinbergen Institute, Amsterdam, Netherlands

Networks represent a useful tool to describe relationships among financial firms and network analysis has been extensively used in recent years to study financial connectedness. An aspect, which is often neglected, is that network observations come with errors from different sources, such as estimation and measurement errors, thus a proper statistical treatment of the data is needed before network analysis can be performed. We show that node centrality measures can be heavily affected by random errors and propose a flexible model based on the matrix-variate *t* distribution and a Bayesian inference procedure to de-noise the data. We provide an application to a network among European financial institutions.

1. INTRODUCTION

A network can be defined as a set of nodes and edges, which represent a relationship among the nodes (Newman et al., 2006; Newman, 2018). A wide spectrum of relational, spatial, and multivariate data from many fields, such as sociology, biology, environmental, and neuroscience, admits a natural representation as a network. In mathematical terms, a network can be represented through the notion of a graph and its properties. For an introduction to graph theory and random graphs, we refer the interested reader to Bollobás (1998) and Bollobás (2001). See Jackson (2008) for an introduction to network theory in social sciences. In this paper, we will use the two terms interchangeably. As an example, in financial networks, a node represents a firm and an edge has the interpretation of a financial relationship between two firms.

In finance, the extraction of unobserved networks from time series data has attracted the attention of many researchers since the recent financial crisis (e.g., Billio et al., 2012; Diebold and Yılmaz, 2014). A large number of different methodologies have been proposed for the estimation of financial networks from firm return series (e.g., Barigozzi and Brownlees, 2019; Bräuning and Koopman, 2020), in particular in the Bayesian approach. For example, Billio et al. (2019) propose a Bayesian non-parametric Lasso prior distribution for vector autoregressive (VAR) models, which provides a sparse estimation of the VAR coefficients and classifies the non-zero coefficients into different clusters. They extract causal networks among financial assets and find that the resulting network topologies match the features of many real-world networks. Ahelegbey et al. (2016a,b) exploit graphical models to specify both the contemporaneous and the lagged causal structures in Bayesian VAR models. In a related contribution, Bianchi et al. (2019) investigate the temporal evolution in systemic risk using a Markov-switching graphical SUR model.

The inferred network structure is intrinsically contaminated by a certain degree of estimation error, which may cumulate with other sources of errors, such as model misspecification and measurement error. Consequently, the direct use of estimated networks as inputs in network

analyses (e.g., Casarin et al., 2020; Wang et al., 2021) may result in misleading conclusions. This calls for the definition of suitable tools for cleaning the data from random disturbances, thus enabling to perform valid statistical analyses of the networks.

In this paper, we propose a new Bayesian model for network data with matrix-variate t errors which accounts for heavy tails (Tomarchio et al., 2020). The inferential procedure is based on data augmentation and conjugate prior distributions that allow for an efficient posterior sampling scheme. In addition to the studies on financial network extraction, our paper also contributes to the literature on matrix-variate models and financial connectedness.

Motivated by the increasing availability of large and multidimensional data, the use of matrix-variate distributions in time series econometrics has flourished during the last decade. The main domains where these models have been successfully applied include the classification of longitudinal datasets (Viroli, 2011), network analysis (Durante and Dunson, 2014; Zhu et al., 2017, 2019), factor analysis (Wang et al., 2019; Chen et al., 2020), stochastic volatility modeling (Gouriéroux et al., 2009; Golosnoy et al., 2012), and Gaussian dynamic linear modeling (Wang and West, 2009).

Finally, we aim to contribute to the financial economics literature by applying the proposed method for de-noising network data extracted from European firms' stock market returns. Then, we analyze the connectedness of the network and compare the results to those obtained from a direct analysis of the network raw data. Our simulation results provide an estimate of the bias in the network centrality measures induced by errors in the edges and show that the proposed approach is effective in correcting for the bias. Our empirical analysis confirms the presence of variability in the network edges. Furthermore, comparing network statistics between the raw network data and the filtered one, we find substantial evidence of differences in the most frequently used statistics, such as out-degree, eigenvector, betweenness, and closeness centrality measures.

The remaining of the paper is structured as follows: section 2 introduces a new linear model for matrix-valued data, then section 3 presents a Bayesian inference procedure. The results of an empirical analysis on real network data are illustrated in section 4. Finally, section 5 concludes.

2. A MATRIX-VARIATE t MODEL

Let $\mathcal{G}_t = (V, H_t)$, $t = 1, \ldots, T$, be a sequence of networks (Boccaletti et al., 2014), where $H_t \subset V \times V$ is the edge set and $V = \{1, \ldots, n\}$ is the set of nodes. In our application, $\mathcal{G}_t$ is a Granger network where the nodes represent institutions from different sectors and directed edges represent financial linkages. A directed edge from node j to node i represents a Granger-causal relationship from firm j to firm i, and is associated to the element $Y_{ij,t}$ of the adjacency matrix Y_t.

The connectivity structure of a n-dimensional network $\mathcal{G}_t$ can be represented through a n-dimensional square matrix Y_t, called the adjacency matrix. Each element $Y_{uv,t}$ of the adjacency matrix is non-zero if there is an edge from institution v to institution u with $u, v \in V$, and 0 otherwise, where $u \neq v$, since self-loops are not allowed.

Unfortunately, most frequently the connectivity structure among financial institutions is not directly observable, thus requiring suitable statistical tools to extract the latent network topologies that are characterized by estimation errors. For example, our data relies on a Granger causality approach to extract network observations from financial price series, which, in turn, may be contaminated by the presence of measurement noise. Overall, these multiple sources of errors may yield an imperfect observation of the true connectivity structure, calling for the adoption of a proper de-noising procedure before performing network analyses.

We propose a matrix-variate linear stochastic model to deal with measurement and estimation errors in the adjacency matrices. The noise process is assumed to follow a matrix-variate t distribution, that accounts for potentially large deviations of the observations from the mean. The proposed model is

$$Y_t = B + E_t, \qquad E_t \overset{iid}{\sim} t_{n,n}(\nu, 0, \Sigma_1, \Sigma_2), \qquad t = 1, \ldots, T, \quad (1)$$

where $B \in \mathbb{R}^{n \times n}$ is a matrix of coefficients and $E_t \in \mathbb{R}^{n \times n}$ is a random error term. A random matrix $X \in \mathbb{R}^{p \times m}$ follows a matrix-variate t distribution, $X \sim t_{p,m}(\nu, M, \Sigma_1, \Sigma_2)$, if it has probability density function

$$P(X|\nu, M, \Sigma_1, \Sigma_2) = \frac{\Gamma_p\left(\frac{\nu+m+p-1}{2}\right)}{\pi^{\frac{mp}{2}} \Gamma_p\left(\frac{\nu+p-1}{2}\right)} |\Sigma_1|^{-\frac{m}{2}} |\Sigma_2|^{-\frac{p}{2}} \Big| I_p$$
$$+ \Sigma_1^{-1}(X - M)\Sigma_2^{-1}(X - M)' \Big|^{-\frac{\nu+m+p-1}{2}} \quad (2)$$

where $\Gamma_p(\cdot)$ is the multivariate gamma function and $|\cdot|$ denotes the matrix determinant. The matrix $M \in \mathbb{R}^{p \times m}$ is the location parameter, $\nu > 0$ is the degrees of freedom parameter, and the positive definite matrices $\Sigma_1 \in \mathbb{R}^{p \times p}$ and $\Sigma_2 \in \mathbb{R}^{m \times m}$ are scale parameters driving the covariances between each of the p rows and the m columns of X, respectively. For further details, see Chapter 4 in Gupta and Nagar (1999). Thanks to the properties of the matrix-variate t distribution, the unconditional mean and variance of Y_t are $\mathbb{E}(Y_t) = B$, if $\nu > 1$, and $\mathbb{V}ar(Y_t) = \Sigma_2 \otimes \Sigma_1/(\nu - 2)$, if $\nu > 2$, where $\otimes$ denotes the Kronecker product.

3. BAYESIAN INFERENCE

3.1. Prior Specification

In this section we describe the prior structure for the model parameters. For the coefficient matrix B, we assume a matrix normal distribution

$$B \sim \mathcal{N}_{n,n}(0, \Omega_1, \Omega_2), \quad (3)$$

where $\Omega_1 = \omega_1 I_n$ and $\Omega_2 = \omega_2 I_n$, with $\omega_1 > 0$ and $\omega_2 > 0$ fixed. A random matrix $Z \in \mathbb{R}^{p \times m}$ follows a matrix normal distribution (Gupta and Nagar, 1999, Chapter 2), denoted by $X \sim \mathcal{N}_{p,m}(M, \Sigma_1, \Sigma_2)$, if its probability density function is

$$P(X|M, \Sigma_1, \Sigma_2) = (2\pi)^{-\frac{mp}{2}} |\Sigma_1|^{-\frac{p}{2}} |\Sigma_2|^{-\frac{m}{2}}$$
$$\exp\left(-\frac{1}{2} \text{tr}\left(\Sigma_2^{-1}(X - M)'\Sigma_1^{-1}(X - M)\right)\right). \quad (4)$$

The matrix normal distribution is equivalent to a multivariate normal distribution with a product-separable covariance structure, that is, $X \sim \mathcal{N}_{p,m}(M, \Sigma_1, \Sigma_2)$ is equivalent to $\mathrm{vec}(X) \sim \mathcal{N}_{pq}(\mathrm{vec}(M), \Sigma_2 \otimes \Sigma_1)$, where $\mathrm{vec}(\cdot)$ is a vectorization operator that stacks all the columns of a matrix into a column vector. Since $\Sigma_2 \otimes \Sigma_1 = (\Sigma_2/a) \otimes (a\Sigma_1)$ for any $a \neq 0$, the noise covariance matrices of the matrix-variate t distribution, Σ_1, Σ_2, are not identifiable and prior restrictions can be used to achieve identification. Nevertheless, in this paper we are interested in the variability of the errors as measured by the product $\Sigma_2 \otimes \Sigma_1$, which is always identifiable. For the noise covariances, Σ_1 and Σ_2, we assume the following hierarchical prior distribution

$$\gamma \sim \mathcal{G}a(a_\gamma, b_\gamma), \qquad \Sigma_1|\gamma \sim \mathcal{W}_n(\gamma^{-1}\Psi_1^{-1}, \kappa_1),$$
$$\Sigma_2|\gamma \sim \mathcal{IW}_n(\gamma\Psi_2, \kappa_2), \qquad (5)$$

where we use the shape-scale parametrization for the gamma distribution and the scale parametrization for the Wishart and inverse Wishart distributions, with densities

$$P(\Sigma_1|\gamma^{-1}\Psi_1^{-1}, \kappa_1) = \frac{1}{2^{n\kappa_1/2}|\gamma^{-1}\Psi_1^{-1}|^{\kappa_1/2}\Gamma_n(\frac{\kappa_1}{2})}$$
$$|\Sigma_1|^{\frac{\kappa_1-n-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\gamma\Psi_1\Sigma_1\right)\right)$$
$$P(\Sigma_2|\gamma\Psi_2, \kappa_2) = \frac{|\gamma\Psi_2|^{\kappa_2/2}}{2^{n\kappa_2/2}\Gamma_n(\frac{\kappa_2}{2})}$$
$$|\Sigma_2|^{-\frac{\kappa_2+n+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)\right),$$

where κ_1 and κ_2 are the degrees of freedom parameters and $\Gamma_n(\cdot)$ is the multivariate gamma function (see Gelman et al., 2014, Appendix A, p. 577). The common scale γ allows for various degrees of prior dependence in the unconditional joint distribution of (Σ_1, Σ_2). Finally, since the object of interest is the mean of Y_t, which is defined only for $\nu > 1$, we assume the following gamma prior distribution truncated on the interval $(1, +\infty)$

$$\nu \sim \mathcal{TG}a(a_\nu, b_\nu; 1, +\infty). \qquad (6)$$

The gamma prior distribution has been previously considered, e.g., in Geweke (1993) and Wang et al. (2011). For the use of an improper prior, see Fonseca et al. (2008). Since we are using a proper prior distribution for B, its posterior distribution is well-defined for $\nu > 0$ (Geweke, 1993), whereas the constraint $\nu > 2$ is required when using improper prior distributions.

3.2. Posterior Approximation

Denote the collection of parameters with $\theta = (B, \Sigma_1, \Sigma_2, \nu)$, and let $\mathbf{Y} = (Y_1, \ldots, Y_T)$ be the collection of all observed networks. The likelihood of the model in Equation (1) is

$$P(\mathbf{Y}|\theta) = \prod_{t=1}^T \frac{\Gamma_n\left(\frac{\nu+2n-1}{2}\right)}{\pi^{\frac{n^2}{2}}\Gamma_n\left(\frac{\nu+n-1}{2}\right)} |\Sigma_1|^{-\frac{n}{2}}|\Sigma_2|^{-\frac{n}{2}}|I_n$$
$$+ \Sigma_1^{-1}(Y_t - B)\Sigma_2^{-1}(Y_t - B)'|^{-\frac{\nu+2n-1}{2}}. \qquad (7)$$

Since the joint posterior distribution implied by the prior assumptions in Equations (3)–(6) and the likelihood in Equation (7) is not tractable, we follow a data augmentation approach. We exploit the representation of the matrix t distribution as a scale mixture of matrix normal distributions, with Wishart mixing distribution (Thompson et al., 2020). From Theorem 4.3.1 in Gupta and Nagar (1999), if $S \sim \mathcal{W}_p(\Sigma_1^{-1}, \nu + p - 1)$ and $X|S \sim \mathcal{N}_{p,m}(M, S^{-1}, \Sigma_2)$, then $X \sim t_{p,m}(\nu, M, \Sigma_1, \Sigma_2)$. Following Gelman et al. (2014) parametrization of the inverse Wishart, we obtain the equivalent representation $W = S^{-1} \sim \mathcal{IW}_p(\Sigma_1, \nu + p - 1)$ and $X|W \sim \mathcal{N}_{p,m}(M, W, \Sigma_2)$. We apply this result to $Y_t \sim t_{n,n}(\nu, B, \Sigma_1, \Sigma_2)$ and obtain the complete data likelihood

$$P(\mathbf{Y}, \mathbf{W}|\theta) = \prod_{t=1}^T \left[(2\pi)^{-\frac{n^2}{2}} |W_t|^{-\frac{n}{2}} |\Sigma_2|^{-\frac{n}{2}} \right.$$
$$\exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_2^{-1}(Y_t - B)'W_t^{-1}(Y_t - B)\right)\right)$$
$$\left. \cdot \frac{|\Sigma_1|^{\frac{\nu+n-1}{2}}}{2^{\frac{(\nu+n-1)n}{2}}\Gamma_n(\frac{\nu+n-1}{2})} |W_t|^{-\frac{\nu+2n}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma_1 W_t^{-1})\right)\right],$$
$$(8)$$

where $\mathbf{W} = (W_1, \ldots, W_T)$ is the collection of auxiliary variables, with $W_t \sim \mathcal{IW}_n(\Sigma_1, \nu + n - 1)$.

The data augmentation approach combined with our prior assumptions allows us to derive analytically the full conditional distributions of B, Σ_1, Σ_2, and $\mathbf{W}$. Since the joint posterior distribution is not tractable, we implement an MCMC approach based on a Gibbs sampling algorithm that iterates over the following steps:

1. Draw $(\nu, \mathbf{W})$ from the joint posterior distribution $P(\nu, \mathbf{W}|\mathbf{Y}, B, \Sigma_1, \Sigma_2)$ with a collapsed-Gibbs step that first samples $\nu \sim P(\nu|\mathbf{Y}, B, \Sigma_1, \Sigma_2)$ and then $\mathbf{W} \sim P(\mathbf{W}|\mathbf{Y}, B, \Sigma_1, \Sigma_2, \nu)$.
2. Draw $\mathrm{vec}(B)$ from the multivariate normal distribution $P(\mathrm{vec}(B)|\mathbf{Y}, \mathbf{W}, \Sigma_1, \Sigma_2)$.
3. Draw Σ_1 from the Wishart distribution $P(\Sigma_1|\mathbf{W}, \nu, \gamma)$.
4. Draw Σ_2 from the inverse Wishart distribution $P(\Sigma_2|\mathbf{Y}, B, \mathbf{W}, \gamma)$.
5. Draw γ from the gamma distribution $P(\gamma|\Sigma_1, \Sigma_2)$.

See the **Appendix** for further details.

3.3. Simulation Experiments

We study the effects of the network estimation errors on the network statistics. We set the size of the network to $n = 70$, assume the degrees of freedom parameter takes values $\nu = 1, 2, \ldots, 50$, and consider two experimental settings with different levels of variance in the error term E_t: (i) low variance, with $\Sigma_1 = I_n \cdot 3.0$ and $\Sigma_2 = I_n \cdot 1.2$, and (ii) high variance, with $\Sigma_1 = I_n \cdot 75.0$ and $\Sigma_2 = I_n \cdot 1.2$. The choice of the parameter settings reflects the results obtained in the empirical application.

The adjacency matrix A of the network is obtained by applying the probit transformation to the elements of B. Following common practice in the analysis of financial connectedness [e.g., see Billio et al., 2012], we fix a threshold equal to 0.05, that is
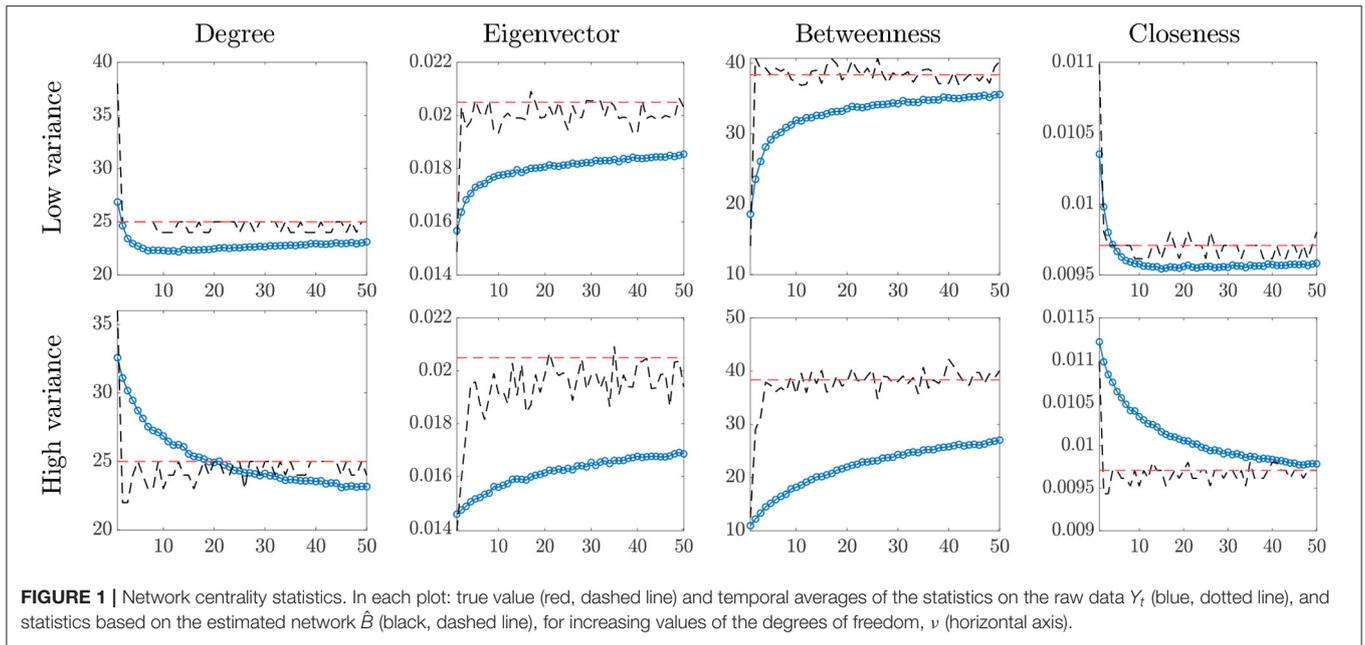
FIGURE 1 | Network centrality statistics. In each plot: true value (red, dashed line) and temporal averages of the statistics on the raw data Y_t (blue, dotted line), and statistics based on the estimated network $\hat{B}$ (black, dashed line), for increasing values of the degrees of freedom, ν (horizontal axis).

$a_{ij} = \mathbb{I}(\Phi(b_{ij} < 0.05))$, where a_{ij} and b_{ij} are the (i, j)-th elements of A and B, respectively, $\mathbb{I}(p)$ is the indicator function, taking value 1 when p is true and 0 otherwise, and Φ denotes the cdf of the standard normal distribution.

We focus on four measures of node centrality commonly used in network analysis (Newman, 2018, Chapter 7): out-degree, d_i^{out}, eigenvector centrality, c_i^e, closeness centrality, c_i^c, and betweenness centrality, c_i^b. To define these measures, we first introduce some notation. A path is a sequence of edges which joins a sequence of distinct vertices, and a node s is said to be reachable from node t if there exists a path which starts with t and ends with s.

The out-degree on node i is the total number of outgoing connections, that is

$$d_i^{out} = \sum_{j=1}^{n} Y_{ji,t}.$$

The eigenvector centrality describes the influence of a node in a network by accounting for the centrality of all the other nodes in its neighborhood. For each node i, it is defined as

$$c_i^b = \frac{1}{\lambda} \sum_{v \in N_i} a_{iv} c_v^b, \tag{9}$$

where the score c_i^b is related to the score of its neighborhood $N_i = \{v \in V; a_{iv} = 1\}$ and λ is an eigenvalue of the adjacency matrix A. The closeness centrality accounts for connectivity patterns by indicating how easily a node can reach other nodes. For each node i, it is given by

$$c_i^c = \frac{n-1}{\sum_{v \in V, v \neq i} l(i, v)} \tag{10}$$

where $l(i, v)$ is the length of the shortest path between i and v. A related measure is the betweenness centrality, which indicates how relevant a node is in terms of connecting other nodes in the graph. Let $n(u, v)$ be the number of shortest paths P_{uv}^* from u to v and $n_i(u, v)$ be the cardinality of the set $\{P_{uv}^* : i \in P_{uv}^*\}$, that is the number of shortest paths from u to v going through the node i. Then the betweenness centrality for node i is

$$c_i^b = \sum_{u \neq v, i \notin \{u, v\}} \frac{n_i(u, v)/n(u, v)}{(n-1)(n-2)}. \tag{11}$$

Figure 1 shows the network statistics for the true network A, the estimated network based on $\hat{B}$, and the empirical averages of the statistics based on the raw data Y_t. The main findings are summarized below:

- for heavier-tailed noise distribution (i.e., smaller ν), the bias of the empirical network statistics increases;
- the bias increases with the scale of the noise, especially for the eigenvector and betweenness centrality measures;
- as the degrees of freedom increase, the noise distribution converges to a matrix normal and the empirical averages approach the true value;
- the model proposed in Equation (1) yields correct estimates of the metrics for all values of ν, even in presence of heavy tails, with higher dispersion as the scale of the noise increases.

Overall, these findings indicate that a direct implementation of network analysis in presence of noisy measurements may lead to misguiding conclusions. This issue can be addressed by using the proposed methodology to de-noise the network data.

FIGURE 2 | Graphical representation of the de-noised network (left) and raw networks at $t_1 = 1$ (middle) and $t_2 = 105$ (right). Black dots represent financial firms and gray arcs represent directed edges (clockwise orientation). The red dot stands for the most central institution according to degree centrality. In each plot, the size of a node is proportional to its out-degree.
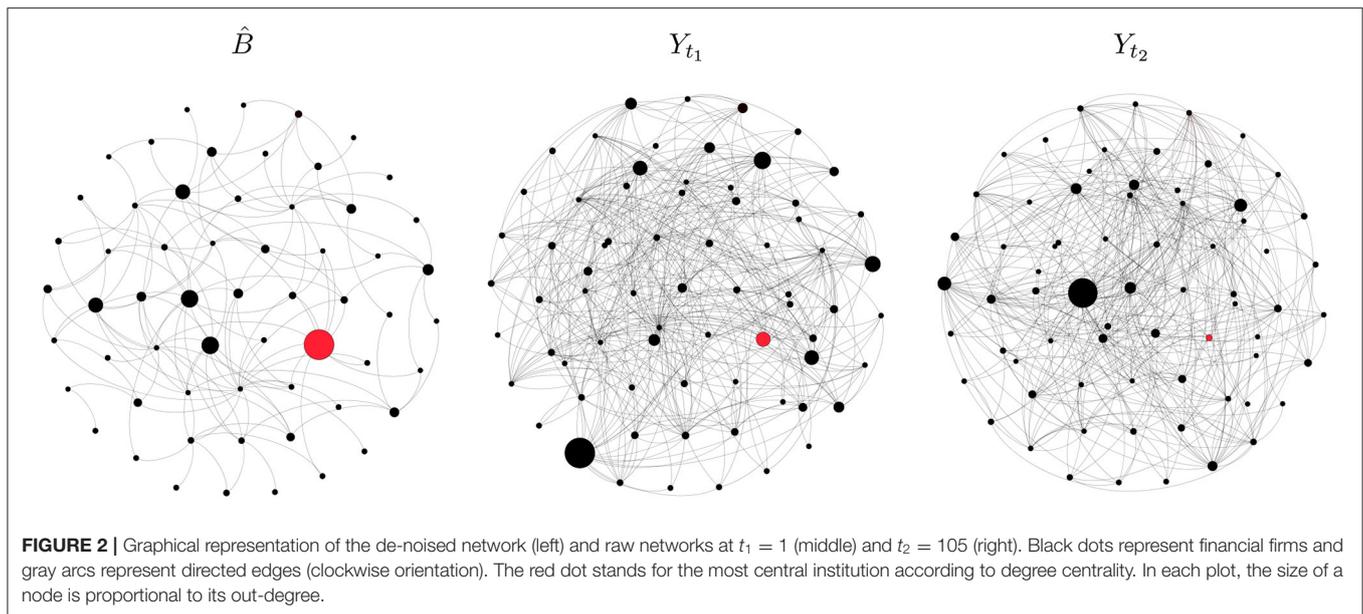
4. EMPIRICAL ANALYSIS

In this section, we provide a description of the raw time series data and of the methodology used to extract the financial networks. Then, we present the benefits of the proposed model for computing the summary statistics most widely used in applied network analysis.

4.1. Data Description

We consider the stock prices of the 70 European firms with the largest market capitalization (source: Bloomberg and Eikon/Datastream). The dataset includes 28 German, 37 French, and five Italian firms, belonging to 11 GICS sectors: Communication Services (four firms), Consumer Discretionary (15 firms), Consumer Staples (six firms), Energy (two firms), Financials (11 firms), Health Care (six firms), Industrials (10 firms), Information Technology (five firms), Materials (two firms), Real Estate (three firms), Utilities (five firms), and Food and Beverages (one firm)[1].

Data are sampled from the 4th of January 2016 to the 31st of December 2019, at weekly frequency (Friday-Friday). The period after the outbreak of the COVID-19 is excluded from the analysis due to the break induced on the network structure.

We extract the network sequence using the pairwise Granger-causality test (e.g., see Billio et al., 2012) on a rolling window of 104 weekly logarithmic returns (i.e., 2 years). The auxiliary regression used is

$$\begin{cases} x_{i,t} = \beta_{11} x_{i,t-1} + \beta_{12} x_{j,t-1} + \varepsilon_{it} \\ x_{j,t} = \beta_{21} x_{i,t-1} + \beta_{22} x_{j,t-1} + \varepsilon_{jt} \end{cases} \quad (12)$$

where $i, j = 1, \ldots, n$, for $i \neq j$. Each entry (i, j) of the matrix Y_t, denoted by $Y_{ij,t}$, is the *p*-value of the Granger test statistic. The

element $Y_{ij,t}$ represents the probability that $x_{j,t}$ Granger-causes $x_{i,t}$. We estimate a total of 105 adjacency matrices, for the period from the 29th of December 2017 to the 31st of December 2019[2].

4.2. Results

In this section, we apply the model and inference proposed in sections 2, 3 to estimate the impact of the risk factors on the European financial network. We run the Gibbs sampler for 5,000 iterations after discarding the first 2,000 as burn-in[3].

Figure 2 shows the de-noised directed network $\hat{B}$ and two elements of the raw series. In each plot, a node represents a firm and its size is proportional to its out-degree. The red dot indicates the most central node in $\hat{B}$, that is the node with the highest out-degree in $\hat{B}$, and directed edges are represented by clockwise-oriented arcs. The node with the highest out-degree is a financial firm, whereas the most central node according to the other measures belongs to the energy sector. The largest black dot in each plot of **Figure 2** represents the most degree-central institution, that is the one with highest out-degree. As shown by the position of the largest dot in the middle and right plots, the most central node varies over the sample. This supports the claim that observation contaminated by noise, if not properly filtered, may alter network analyses.

The posterior density of the degrees of freedom in the left plot of **Figure 3** suggests that the noise distribution is close to the Gaussian. Nonetheless, as shown in the simulation experiments, our approach is able to provide more accurate estimates of the network measures as compared to the empirical averages. This

---

[1]The list of the firms, the countries, and the information about their GICS sectors and industries are available upon request to the Authors.

[2]The estimation algorithm has been parallelized and implemented in MATLAB on the High Performance Computing (HPC) cluster at Ca' Foscari University. Each node has two CPUs Intel Xeon with 20 cores 2.4 GHz and 768 GB of RAM.

[3]The total computing time for the empirical application is about 22 h with code written in MATLAB and run on an Intel Xeon with 20 cores 2.4 GHz and 768 GB of RAM.
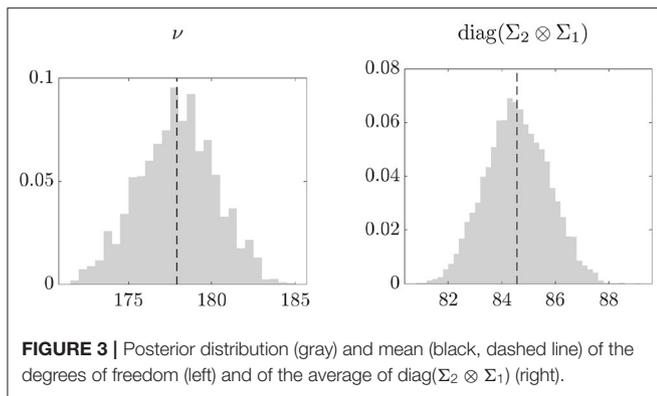
FIGURE 3 | Posterior distribution (gray) and mean (black, dashed line) of the degrees of freedom (left) and of the average of diag($\Sigma_2 \otimes \Sigma_1$) (right).

FIGURE 4 | Network centrality statistics. In each plot: posterior distribution (gray) and mean (black, dashed line) of the statistics, and temporal averages of the statistics on the raw data Y_t (red, dashed line).

is particularly evident for the eigenvector centrality, where the empirical averages are sensitive to errors compatible with a *t* distribution with large degrees of freedom (see **Figure 1**). The right plot shows the posterior distribution of the average of the elements on the main diagonal of $\Sigma_2 \otimes \Sigma_1$. The estimated average variance is 84.5, thus providing evidence of high variability in the network observations.

As described in section 1, the presence of noise in the data can invalidate network analyses, such as the identification of the most central institution. Motivated by this fact, we assess the importance of de-noising the network data by computing the network centrality measures on the raw data and on the de-noised network obtained using the method in section 2. The results are shown in **Figure 4**, which reports the posterior distribution of the centrality measures computed on the de-noised network (gray, with the black line representing the posterior mean) and the temporal average of the statistics computed on the raw data (red line). We find that all centrality measures of the de-noised data differ from the temporal average of the raw data; in particular, the eigenvector and betweenness centrality based on the raw data are underestimated, while the closeness centrality is overestimated. Overall, these findings provide evidence that the presence of noise in network data may jeopardize the validity of the ensuing analyses of the network structure.

5. CONCLUSIONS

A common, though often neglected, aspect of network analysis is that observations for networks might come with errors from different sources, such as estimation and measurement errors. We show that noise may invalidate the study of the network topology, such as the measurement of node centrality.

We have introduced a new matrix-variate regression framework that allows for heavy-tailed matrix-variate *t* errors to address this issue. The model is applied to filter out the noise from network data as a preliminary step before investigating the connectedness structure. In the presence of heavy-tailed error distributions or big scales of the variance of the noise, the proposed approach has superior performance compared to the temporal averages of the network statistics. Finally, we have
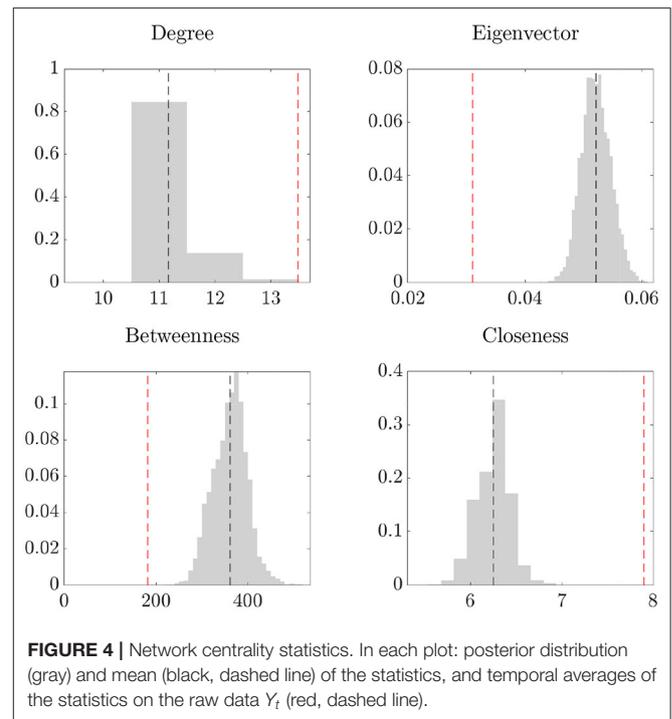
applied the model to a sequence of estimated networks among European firms and find evidence of large error variance that affects the centrality measures.

More generally, our approach can be implemented to obtain robust network inference or fit benchmark random network models, such as those proposed by Erdőrigos-Rény and Albert-Barabási, thus representing a valuable tool for researchers investigating networks.

In this paper, we focus on the case where all the noisy observations contain the same set of nodes, meaning that there is no uncertainty on the network nodes. An interesting extension would be to consider observed networks having different sizes. We leave this for future research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2021.674166/full#supplementary-material

REFERENCES

Ahelegbey, D. F., Billio, M., and Casarin, R. (2016a). Bayesian graphical models for structural vector autoregressive processes. *J. Appl. Econometr.* 31, 357–386. doi: 10.1002/jae.2443

Ahelegbey, D. F., Billio, M., and Casarin, R. (2016b). Sparse graphical vector autoregression: a Bayesian approach. *Ann. Econ. Stat.* 333–361. doi: 10.15609/annaeconstat2009.123-124.0333

Barigozzi, M., and Brownlees, C. (2019). Nets: network estimation for time series. *J. Appl. Econometr.* 34, 347–364. doi: 10.1002/jae.2676

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with Markov switching graphical sur models. *J. Econometr.* 210, 58–74. doi: 10.1016/j.jeconom.2018.11.005

Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *J. Econometr.* 212, 97–115. doi: 10.1016/j.jeconom.2019.04.022

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J. Financ. Econ.* 104, 535–559. doi: 10.1016/j.jfineco.2011.12.010

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544, 1–122. doi: 10.1016/j.physrep.2014.07.001

Bollobás, B.(1998). *Modern Graph Theory.* New York, NY: Springer.

Bollobás, B. (2001). *Random Graphs.* Cambridge: Cambridge University Press.

Bräuning, F., and Koopman, S. J. (2020). The dynamic factor network model with an application to international trade. *J. Econometr.* 216, 494–515. doi: 10.1016/j.jeconom.2019.10.007

Casarin, R., Iacopini, M., Molina, G., ter Horst, E., Espinasa, R., Sucre, C., et al. (2020). Multilayer network analysis of oil linkages. *Econometr. J.* 23, 269–296. doi: 10.1093/ectj/utaa003

Chen, E. Y., Tsay, R. S., and Chen, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *J. Am. Stat. Assoc.* 115, 775–793. doi: 10.1080/01621459.2019.1584899

Diebold, F. X., and Yılmaz, K. (2014). On the network topology of variance decompositions: measuring the connectedness of financial firms. *J. Econometr.* 182, 119–134. doi: 10.1016/j.jeconom.2014.04.012

Durante, D., and Dunson, D. B. (2014). Bayesian dynamic financial networks with time-varying predictors. *Statist. Probab. Lett.* 93, 19–26. doi: 10.1016/j.spl.2014.06.015

Fonseca, T. C., Ferreira, M. A., and Migon, H. S. (2008). Objective bayesian analysis for the student-t regression model. *Biometrika* 95, 325–333. doi: 10.1093/biomet/asn001

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis.* CRC Press.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *J. Appl. Econometr.* 8, S19–S40. doi: 10.1002/jae.3950080504

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *J. Econometr.* 167, 211–223. doi: 10.1016/j.jeconom.2011.11.004

Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *J. Econometr.* 150, 167–181. doi: 10.1016/j.jeconom.2008.12.016

Gupta, A. K., and Nagar, D. K. (1999). *Matrix Variate Distributions.* Boca Raton, FL: CRC Press.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton, NJ: Princeton University Press.

Newman, M. (2018). *Networks.* Oxford: Oxford University Press.

Newman, M. E., Barabási, A. L. E., and Watts, D. J. (2006). *The Structure and Dynamics of Networks.* Princeton, NJ: Princeton University Press.

Thompson, G. Z., Maitra, R., Meeker, W. Q., and Bastawros, A. F. (2020). Classification with the matrix-variate-t distribution. *J. Comput. Graph. Statist.* 29, 668–674. doi: 10.1080/10618600.2019.1696208

Tomarchio, S. D., Punzo, A., and Bagnato, L. (2020). Two new matrix-variate distributions with application in model-based clustering. *Comput. Statist. Data Anal.* 152:107050. doi: 10.1016/j.csda.2020.107050

Viroli, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Statist. Comput.* 21, 511–522. doi: 10.1007/s11222-010-9188-x

Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Econometr.* 208, 231–248. doi: 10.1016/j.jeconom.2018.09.013

Wang, H., and West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika* 96, 821–834. doi: 10.1093/biomet/asp049

Wang, J. J., Chan, J. S., and Choy, S. B. (2011). Stochastic volatility models with leverage and heavy-tailed distributions: a Bayesian approach using scale mixtures. *Comput. Statist. Data Anal.* 55, 852–862. doi: 10.1016/j.csda.2010.07.008

Wang, Y. R., Li, L., Li, J. J., Huang, H., et al. (2021). Network modeling in biology: statistical methods for gene and brain networks. *Statist. Sci.* 36, 89–108. doi: 10.1214/20-STS792

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *Ann. Statist.* 45, 1096–1123. doi: 10.1214/16-AOS1476

Zhu, X., Wang, W., Wang, H., and Härdle, W. K. (2019). Network quantile autoregression. *J. Econometr.* 212, 345–358. doi: 10.1016/j.jeconom.2019.04.034