IMA Journal of Management Mathematics Page 1 of 16
doi:10.1093/imaman/dpnxxx

A robustness comparison of two market network models.

Two market network models are investigated. One of them is based on the classical Pearson correlation as the measure of association between stocks returns, whereas the second one is based on the sign similarity measure of association between stocks returns. We study the uncertainty of identification procedures for the following market network characteristics: distribution of weights of edges, vertex degree distribution in the market graph, cliques and independent sets in the market graph, and the vertex degree distribution of the maximum spanning tree. We define the true network characteristics, the losses from the error of its identification by observations, and the uncertainty of identification procedures as the expected value of losses. We use elliptically contoured distribution as a model of multivariate stocks returns distribution. It is shown that identification statistical procedures based on the sign similarity are statistically robust in contrast to the procedures based on the classical Pearson correlation.

Keywords: market network analysis, random variables network, measures of association, uncertainty, risk function, distribution-free statistical procedures

1. Introduction

Mathematical models of stock market have attracted a large attention in theoretical and applied research. In particular, these models are useful for portfolio and financial risk management(Dong *et al.* (2018), I.Baltas, A.N.Yannacopoulos (2019)). One popular approach to model the stock market is related with network analysis. Recently, methods of stock market analysis based on the corresponding network models (Mantegna *et al.* (1999), Boginski *et al.* (2005)) are becoming increasingly widespread. The network model of the stock market is a complete weighted graph where vertices correspond to the stocks returns and the weights of the edges are given by some measure of association (dependence) between the stocks returns. Such model allows to investigate the hierarchical structure and clusters of the stock market (Tumminello *et al.* (2010)), the most influential stocks in the market (Hero, A., Rajaratnam, B. (2012)), the stock market dynamics (Pereira *et al.* (2019), Nguyen *et al.* (2019)), investment portfolios (Kalyagin *et al.* (2014)) and other stock market characteristics (see Chi *et al.* (2010), Marti *et al.* (2019) and Kalyagin *et al.* (2014) for a large bibliography on the subject).

To filter the most valuable information from a network model different network structures (subgraphs of the complete weighted graph) and their characteristics can be considered. Popular market network structures are: the market graph, cliques and independent sets of the market graph (Boginski *et al.* (2005)), and the maximum spanning tree (MST) of the network model (Mantegna *et al.* (1999)). Using this approach different stock markets for different countries were investigated (see, for example, Coronnello *et al.* (2005), Garas A., Argyrakis P. (2007), Huang *et al.* (2009), Jung *et al.* (2006), Tabak *et al.* (2010), Nguyen *et al.* (2019)). At the same time in these studies there is a big lack of analysis of the reliability of the findings.

Most of publications deal with Pearson correlation as the measure of association (dependence) between stocks returns. This measure is the most appropriate measure of dependence between random variables under the assumption of Gaussian distributions for the stock returns. Generally, however, the Gaussian distribution hypothesis is not confirmed by real stock market data. Elliptically contoured distributions are a natural generalization of Gaussian distributions which are widely used in financial modeling Gupta *et al.* (2013). In the case of elliptically contoured distributions other measures of association (dependence) can be more appropriate than Pearson correlation. This leads to different market

network models. The following question becomes important: how does the reliability of the results of market network analysis differ across different network models?

To answer this question we consider the market network as a random variable network (Kalyagin *et al.* (2017)) and investigate the uncertainty of statistical procedures for network structures identification. A random variable network is a pair (X,γ) where $X=(X_1,\ldots,X_N)$ is a random vector (vector of stock returns), and γ is a measure of association (dependence) between the components of the vector X (measure of association between stock returns). In our study, we consider a large class of elliptically contoured distributions of the vector X (Gupta *et al.* (2013)) and two measures of association: Pearson correlation and an alternative measure of association, sign similarity, introduced in (Bautin *et al.* (2013)). We study the reliability (uncertainty) of identification procedures for the following network characteristics: distribution of weights of edges, degree distribution in the market graph, cliques and independent sets in the market graph and the degree distribution in the maximum spanning tree. In the framework of the concept of random variable network we define the true network characteristics, the losses from the errors of its identification by observations, and the uncertainty of identification procedures as the value of the associated risk function (expected value of losses). Special attention is paid to the dependence of the uncertainty on distribution of the vector X. Procedures with uncertainty not depending on distribution are of practical interest. We call such procedures distribution-free statistical procedures (Kendall, M. G., Stuart, A. (1979)).

In our study we consider a distribution-free statistical procedure to be a robust statistical procedure in the sense that the associated risk function is robust with respect to the distribution used for the analysis. The term robustness is used in science in different senses. In robust optimization one is looking for the worst case solution of optimization problem under the condition that the parameters of the problem lie in an uncertainty set. Such approach was recently applied to robust CVaR (Conditional Value at Risk) portfolio optimization with uncertainty set described by a parallelepiped of observed values of returns (Kara *et al.* (2019)). In statistical parameter estimation robustness means weak dependence of estimations on weak perturbation of distribution (see Huber (1981), Schevlyakov, Hannu Oja (2016)). In the present paper robustness means independence or weak dependence of the risk function of network structure identification procedures on distribution from a specific class.

Robustness in the latter sense was investigated in (Bautin *et al.* (2014), Kalyagin *et al.* (2017)). In (Kalyagin *et al.* (2017)) two market network models (sign similarity network and Pearson correlation network) with elliptically contoured distribution of vector $X=(X_1,\ldots,X_N)$ were theoretically investigated. Following the paper (Kalyagin *et al.* (2017)), the sign similarity network is the random variable network where the measure $\gamma_{i,j}=\gamma(X_i,X_j)$ is a probability of sign coincidence of random variables X_i,X_j, and the Pearson correlation network is the random variable network where the measure $\gamma_{i,j}=\gamma(X_i,X_j)$ is a Pearson correlation between random variables X_i,X_j. It was proved that network models and network structures (the market graph and MST) generated by the sign similarity and Pearson correlation networks are equivalent. In addition, it was proved that statistical procedures for the market graph and MST identification are robust (distribution free) in sign similarity network. In (Bautin *et al.* (2014)) it was shown by simulations that the procedures for the market graph and the maximum spanning tree (MST) identification are not robust in the Pearson correlation network unlike procedures for the market graph and the MST identification in sign similarity network. For the simulations, the mixture of multivariate normal and Student distributions was used. The robustness of identification procedures for other network characteristics was not investigated.

The aim of the present article is to demonstrate how uncertainty of the market network analysis is related with the chosen network model. The following characteristics that are widely used in market network analysis (Marti *et al.* (2019)) are studied: distribution of weights of edges, degree distribution

in the market graph, cliques and independent sets in the market graph and degree distribution in the maximum spanning tree. It is shown that uncertainty of identification procedures in the sign similarity network does not depend on the distribution of the vector of returns (robust identification) while uncertainty of identification procedures in the Pearson correlation network essentially depends on (non robust identification). This dependence is investigated for different markets and different periods of observations.

The paper is organized as follows: in section 2 basic definitions and notations are presented; in section 3 the concepts of robustness are presented; in section 4 the loss and risk functions are proposed; in section 5 the results of robustness investigation are discussed, and in section 6 some remarks are given and the obtained results are discussed.

2. Basic definitions and notations

Our research is based on the concept of *random variables network* (Kalyagin *et al.* (2017)). A random variables network is a pair (X,γ), where $X=(X_1,\dots,X_N)$ is a random vector and γ is a measure of dependence between random variables. Denote by $\gamma_{i,j}=\gamma(X_i,X_j)$. One can consider different random variables networks according the distribution of vector X and the chosen measure of dependence γ. In the article it is assumed that the vector X has elliptically contoured distribution with a density (Anderson (2003)):

$$f(x)=|\Lambda|^{-1}g((x-\mu)'\Lambda^{-1}(x-\mu)), \tag{2.1}$$

where $x\in R^N$, $\mu\in R^N$, Λ is a symmetric positive definite matrix, $g(x)\geqslant 0$, and the function $g(x)$ satisfies the condition

$$\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}g(x'x)dx_1\dots dx_N=1,$$

where $x'x=\sum_{i=1}^N x_i^2$. This class of distributions includes in particular the multivariate normal distribution with density

$$f_{gauss}(x;\mu,\Lambda)=\frac{1}{(2\pi)^{N/2}|\Lambda|^{1/2}}exp(-\frac{1}{2}(x-\mu)'\Lambda^{-1}(x-\mu)),$$

the multivariate Student distribution with k degrees of freedom with density

$$f_{St,k}(x;\mu,\Lambda)=\frac{1}{(k\pi)^{N/2}|\Lambda|^{1/2}}\frac{\Gamma(\frac{k+N}{2})}{\Gamma(\frac{k}{2})}(1+\frac{(x-\mu)'\Lambda^{-1}(x-\mu)}{k})^{-\frac{k+N}{2}},$$

and their mixture with density

$$f_{mix}(x)=\varepsilon f_{gauss}(x;\mu,\Lambda)+(1-\varepsilon)f_{St,k}(x;\mu,\Lambda),$$

where $\varepsilon\in[0,1]$.

Two measures of dependence are considered. One of them is the Pearson correlation

$$\gamma^P_{i,j}=\rho_{i,j}=\frac{E(X_i-EX_i)(X_j-EX_j)}{\sqrt{DX_iDX_j}},$$

where EX stands for expectation, DX stands for variance.

Such measures are widely used in market network analysis (Mantegna *et al.* (1999), Boginski *et al.* (2005)).

An alternative measure of dependence is the probability of sign coincidence

$$\gamma_{i,j}^{Sg} = p^{i,j} = P((X_i - EX_i)(X_j - EX_j) > 0),$$

where $P(A)$ stands for the probability of the event A.

A random variables network generates a network model that is a complete undirected weighted graph $G = (V, \Gamma)$ where $V = \{1, 2, \ldots, N\}$ is the set of vertex associated with random variables $X_1, X_2, \ldots, X_N$ and $\Gamma = \{\gamma_{i,j} : i, j = 1, \ldots, N\}$ is the set of edge weights. It is natural to reduce the study of the network model $G = (V, \Gamma)$ to the study of its key characteristics. In graph theory, various characteristics are proposed: a threshold (market) graph, cliques and independent sets of the threshold graph, a maximum spanning tree, degrees distribution, centrality, diameter, etc. In this paper we investigate the characteristics of two most popular network structures in market network analysis, the market graph and the maximum spanning tree. More specifically,

- the threshold graph (market graph, MG) of the network model $G = (V, \Gamma)$ which is defined as an unweighted graph $G'(\gamma_0) = (V', E') : V' = V; E' \subseteq E$ (where E is a set of all edges of the network model) and $E' = \{(i, j) : \gamma_{i,j} > \gamma_0\}$, where γ_0 is a certain threshold.

- the maximum spanning tree (MST) of the network model $G = (V, \Gamma)$ which is a tree (graph without cycles) $G' = (V', E') : V' = V; E' \subset E; |E'| = |V| - 1$; such that $\sum_{(i,j) \in E'} \gamma_{i,j}$ is maximal.

In this paper we study the following characteristics:

- A distribution of the weights of edges which is defined as the function $h(x) = m$, where m is the number of edges with weights belonging to the nonoverlaping intervals $(a + (k-1)\Delta, a + k\Delta)$ for some $\Delta > 0$. To construct the function $h(x)$ the support $x \in [a, b]$ of the function $h(x)$ is divided on M nonoverlaping intervals of lengths $\Delta = \frac{b-a}{M}$. For any point x from the interval $(a + (k-1)\Delta, a + k\Delta)$ function $h(x) = m$ where m is a number of edges with weights belonging to the interval $(a + (k-1)\Delta, a + k\Delta)$. For the case of Pearson correlation the support is $[-1, 1]$ and $k = 1, \ldots, M$, whereas for the case of the probability of sign coincidence the support is $[0, 1]$ and $k = 1, \ldots, M$.

- A clique of the market graph $G = (V, E)$ is the complete subgraph of the graph G, i.e. subgraph $G' = (V', E') : V' \subset V, E' \subset E : \forall i, j \in V' \Rightarrow (i, j) \in E'$. The clique $G_1 = (V_1, E_1)$ is called the maximum clique (MC) (in size) if for any other clique $G_2 = (V_2, E_2)$: $|V_1| \geqslant |V_2|$

- An independent set (IS) of the market graph $G = (V, E)$ is an empty subgraph of G, that is, subgraph $G_1 = (V_1, E_1) : V_1 \subset V, E_1 \subset E : \forall i, j \in V_1 \Rightarrow (i, j) \notin E_1$. The independent set $G_1 = (V_1, E_1)$ is called the maximal independent set (MIS) (in size) if for any other independent set $G_2 = (V_2, E_2)$ of the graph G: $|V_1| \geqslant |V_2|$.

- A degree distribution of the market graph which is defined as the $2 \times N$ matrix, where the first row contains the possible values of the degrees of vertices $0, 1, \ldots, N-1$ and the second row contains the number of vertices ν_i of degree i, $i = 0, \ldots, N-1$.

- A degree distribution of the MST which is defined as the $2 \times N$ matrix, where the first row contains the possible values of the degrees of vertices $1, \ldots, N-1$ and the second row contains the number of vertices ν_i of degree i, $i = 1, \ldots, N-1$.

The family $\{MG(\gamma_0) : \gamma_0 \in R^1\}$ contains the most complete network model information, in particular, the network model of the market. At the same time, cliques and independent sets characterize the cluster structure of the market. In addition, the size of the maximum clique could be considered as an indicator of globalization, and the size of the maximum independent set could be considered as an indicator of the market freedom.

3. The concept of robustness

Let us assume that X has an elliptically contoured distribution with parameters μ and Λ (2.1). Let $\gamma_{i,j},\ i,\ j = 1,\ldots,N$ be the true (known) value of the dependence measure between the random variables X_i and X_j (the weight of the edge between the vertices i and j in the network model). A network model based on $\gamma_{i,j},\ i,\ j = 1,\ldots,N$ will be called a *reference* network model. The characteristics of this network model, as defined in Section 2, will be referred to as the *reference* characteristics of the network model.

Let $\hat{\gamma}_{i,j},\ i,\ j = 1,\ldots,N$ be an estimation of the dependence measure between random variables X_i and X_j, constructed from the sample $x_i(t),\ i = 1,\ldots,N;\ t = 1,\ldots,n$. A network model based on $\hat{\gamma}_{i,j},\ i,\ j = 1,\ldots,N$ will be called a *sample* network model. The characteristics of this network model will be called *sample* characteristics of the network model.

In practice the available data involve observations of stock returns $x_i(t),\ i = 1,\ldots,N;\ t = 1,\ldots,n$. The identification problems for the networks models involve the estimation of the characteristics of a network model through the available observations. For estimation of the Pearson correlation the sample Pearson correlation will be used:

$$r_{i,j} = \frac{\sum_t (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{\sqrt{\sum_t (x_i(t) - \bar{x}_i)^2 (x_j(t) - \bar{x}_j)^2}},$$

where $\bar{x}_i = \frac{1}{n}\sum_{t=1}^{n} x_i(t)$
As an estimation of the probability of sign coincidence, the frequency of sign coincidence will be used:

$$s_{i,j} = \frac{1}{n}\sum_{t=1}^{n} I_{i,j}(t),$$

where for the case of known μ

$$I_{i,j} = \begin{cases} 1, (x_i(t) - \mu_i)(x_j(t) - \mu_j) \geqslant 0 \\ 0, (x_i(t) - \mu_i)(x_j(t) - \mu_j) < 0. \end{cases}$$

For the case of unknown μ, the frequency of sign coincidence has the form

$$s_{i,j}^{mean} = \frac{1}{n}\sum_{t=1}^{n} I_{i,j}^{mean}(t),$$

where

$$I_{i,j}^{mean} = \begin{cases} 1, (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j) \geqslant 0 \\ 0, (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j) < 0. \end{cases}$$

The problem is to analyze the robustness of the estimation of the network structures characteristics using the sample Pearson correlation or the frequency of sign coincidence. To analyze robustness we

will use the approach proposed in (Kalyagin *et al.* (2017)). Consider, for example, the problem of the market graph identification. The problem of the market graph identification is to select one of the hypotheses:

$$\begin{aligned}
&H_{S_1} : \gamma_{i,j} \leqslant \gamma_0, \forall (i,j),\ i<j,\\
&H_{S_2} : \gamma_{1,2} > \gamma_0, \gamma_{i,j} \leqslant \gamma_0, \forall (i,j) \neq (1,2),\ i<j,\\
&H_{S_3} : \gamma_{1,2} > \gamma_0, \gamma_{1,3} > \gamma_0, \gamma_{i,j} \leqslant \gamma_0, \forall (i,j) \neq (1,2), (i,j) \neq (1,3),\\
&\dots\\
&H_{S_L} : \gamma_{i,j} > \gamma_0, \forall (i,j),\ i<j.
\end{aligned} \tag{3.1}$$

Hypothesis H_{S_1} corresponds to the empty market graph $G'(\gamma_0)$, hypothesis H_{S_2} corresponds to the market graph $G'(\gamma_0)$ with one edge $(1,2)$, etc., and hypothesis H_{S_L} corresponds to the complete market graph $G'(\gamma_0)$. For a network model with N vertices the number of hypotheses is $L = 2^{\frac{N(N-1)}{2}}$.

Hypothesis $H_{S_i},\ i = 1,\dots,\ L$ could be defined by the adjacency matrix $G \in \mathscr{G}$, where $\mathscr{G}$ is the set of all possible adjacency matrices. A multiple decision statistical procedure δ for market graph identification is a map from the sample space $R^{N\times n}$ to the decision space $D = \{d_G, g \in \mathscr{G}\}$, where the decision d_G is the acceptance of hypothesis H_G, $G \in \mathscr{G}$ (market graph has adjacency matrix G). Let $S = (s_{i,j})$, $Q = (q_{i,j})$, $S, Q \in \mathscr{G}$ and

$$w(H_S; d_Q) = w(S,Q),\ \ S,Q \in \mathscr{G}$$

be the loss from the decision d_Q when hypothesis H_S is true. It is assumed that $w(S,S) = 0,\ S \in \mathscr{G}$.

The quality of statistical procedure δ is measured by the risk function (Kalyagin *et al.* (2017)). Assume the vector X has distribution from some class K. Each distribution P_θ from the class K is associated with some parameter θ from the parameter space Ω. The risk function is then defined by

$$R(S,\theta,\delta) = \sum_{Q\in\mathscr{G}} w(S,Q) P_\theta(\delta(x) = d_Q),\ \theta \in \Omega_S,$$

where Ω_S is the parametric region corresponding to hypothesis H_S (i.e the set of distributions such that the reference (true) network structure in (V,Γ) has adjacency matrix S).

It is assumed that X has elliptically contoured distribution with density (2.1). In this case, the reference market graph is defined by the matrix Λ. Therefore risk $R(S,\theta,\delta) = R(\Lambda,\mu,g,\delta)$.

The statistical procedure δ is robust if $R(\Lambda,\mu,g,\delta) = R(\Lambda,\delta)$ (Kalyagin *et al.* (2017)).

4. Loss and risk functions

Let us introduce loss and risk functions for statistical procedures of network characteristic identification (estimation) or measures of the difference between reference and sample characteristics of the corresponding network models.

Let the function $h(x)$ be the reference distribution of edge weights and $\hat{h}(x)$ be the estimate of $h(x)$. To measure the difference between $h(x)$ and $\hat{h}(x)$ we propose to use an expectation of the area S under the curve $|h(x) - \hat{h}(x)|$. Therefore, $W(h(x),\hat{h}(x)) = S(|h(x) - \hat{h}(x)|)$ and risk is $E_{\mu,\Lambda,g}(S(|h(x) - \hat{h}(x)|))$.

Let $G(\gamma_0)$ be the market graph, with k_i denoting the reference number of vertices of degree i in $G(\gamma_0)$. The estimate of k_i is denoted by $\hat{k}_i$. To measure the difference between reference and sample degrees distributions we propose to use $E_{\mu,\Lambda,g}(\sum_{i=0}^{N-1} |k_i - \hat{k}_i|)$, where N is the number vertices in $G(\gamma_0)$. Therefore, $W(k_i,\hat{k}_i) = \sum_{i=0}^{N-1} |k_i - \hat{k}_i|$ and risk is $E_{\mu,\Lambda,g}(\sum_{i=0}^{N-1} |k_i - \hat{k}_i|)$.

To measure the difference between the maximum reference clique and the maximum sample clique (independent sets) we propose to use the expectation of the power of the symmetric difference of the

vertex set of the reference clique and the sample clique (independent set) i.e. $E_{\mu,\Lambda,g}(\sum_{i=1}^{N}|v_i-\hat{v}_i|)$, where N is the number of vertices, v_i is an indicator that equals 1, if vertex i is in the reference clique (independent set) and zero otherwise, and $\hat{v}_i$ is an indicator that equals 1, if vertex i is in the sample clique (independent set) and zero otherwise. Then, $W(v_i,\hat{v}_i)=\sum_{i=1}^{N}|v_i-\hat{v}_i|$ and risk is $E_{\mu,\Lambda,g}(\sum_{i=1}^{N}|v_i-\hat{v}_i|)$.

To measure the difference of the degree distribution in the reference MST and sample MST we propose to use the probability of a correct identification of the degrees distribution in the MST.

5. Results

This section presents the robustness investigation results of two types of procedures for the characteristics of network models identification. Procedures of the first type are based on the sample Pearson correlation. Procedures of the second type are based on the frequency of sign coincidence. The robustness investigation is based on simulation of observations from a mixture distribution of the form:

$$f_{mix}(x)=\varepsilon f_{gauss}(x;\mu,\Lambda)+(1-\varepsilon)f_{St,k}(x;\mu,\Lambda),$$

where $f_{gauss}(x)$ is the N-dimensional normal distribution and $f_{St,k}(x)$ is the N-dimensional Student distribution with $k=3$ degrees of freedom. As the reference matrix Λ of the Pearson correlation network we use matrix $\Lambda=(\lambda_{ij})$ where the elements λ_{ij} are Pearson correlations calculated by real market data. For the construction of a reference sign similarity network we propose to use the relation:

$$\gamma_{i,j}^{Sg}=\frac{1}{2}+\frac{1}{\pi}arcsin(\gamma_{i,j}^{P}).$$

This relation is holds for elliptically contoured distributions (Kalyagin *et al.* (2017)).

5.1 *Reference networks*

24 models for the reference matrix Λ are used. Such models are constructed from the real data from the stock markets of Brazil, China, France, UK, Germany, India, Russia and USA. For each market the returns of $N=50$ of the most profitable stocks of the markets are analyzed for 2010, 2011, 2012 years.

In table 1 a part of the reference matrix Λ of the Pearson correlation network corresponding to the UK stock market for 2010 year is shown.

1,00	0,12	0,00	0,10	-0,05	-0,14	0,04	-0,02	-0,02	0,22
0,12	1,00	0,08	-0,03	-0,01	-0,03	-0,05	-0,03	0,05	-0,10
0,00	0,08	1,00	0,04	-0,06	0,02	0,02	-0,04	-0,04	-0,03
0,10	-0,03	0,04	1,00	-0,07	0,06	0,10	0,06	0,04	0,09
-0,05	-0,01	-0,06	-0,07	1,00	0,49	0,14	0,44	0,35	0,01
-0,14	-0,03	0,02	0,06	0,49	1,00	0,24	0,48	0,42	-0,09
0,04	-0,05	0,02	0,10	0,14	0,24	1,00	0,30	0,15	0,00
-0,02	-0,03	-0,04	0,06	0,44	0,48	0,30	1,00	0,45	0,04
-0,02	0,05	-0,04	0,04	0,35	0,42	0,15	0,45	1,00	0,01
0,22	-0,10	-0,03	0,09	0,01	-0,09	0,00	0,04	0,01	1,00

Table 1. A part of the reference matrix Λ of the Pearson correlation network. (UK, 2010 year)

Based on such reference Pearson correlation network models the reference characteristics were constructed. In total, for 8 countries, 3 years of analysis and 5 characteristics, $8 * 3 * 5 = 120$ of the reference characteristics of Pearson correlation network models were constructed. Examples of the reference characteristics of the network models are given below.

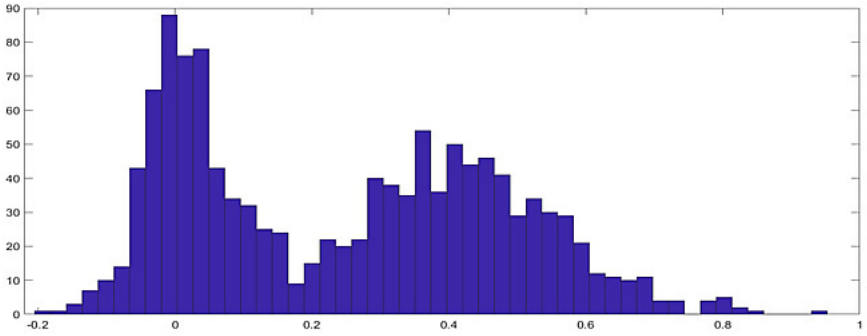

FIG. 1. Reference distribution of weights of edges in Pearson correlation network. (UK, 2010 year)

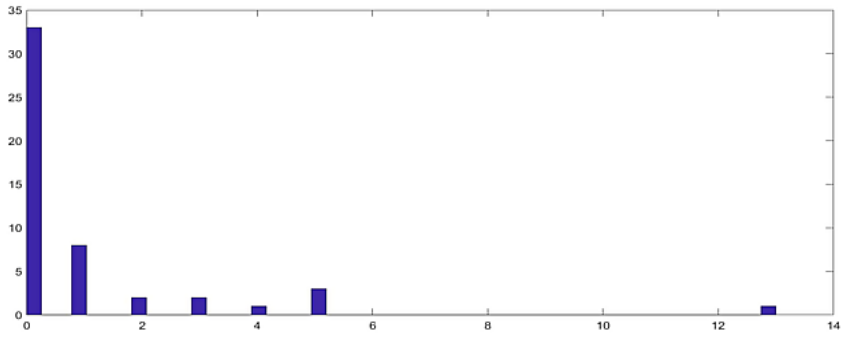

FIG. 2. Reference degree distribution in the market graph with threshold 0.3 in reference Pearson correlation network. (UK, 2010 year)

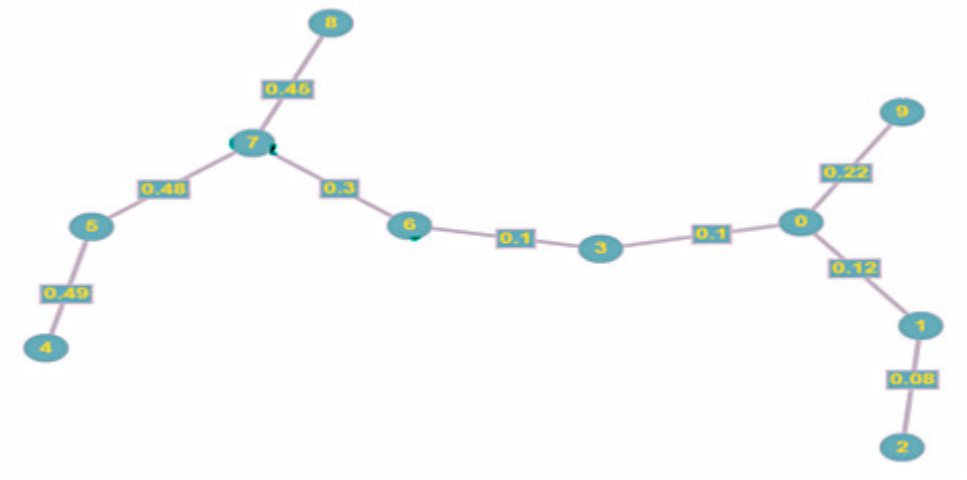

FIG. 3. MST in reference Pearson correlation network with degree distribution [1,1,1,1,2,2,2,2,3,3]. (UK, 2010 year)

In the same way the 120 characteristics in reference sign similarity network were constructed.

5.2 *Algorithm for the calculation of the risk function*

1. Choose the value of ε $(\varepsilon = 0, 0.1.0.2, ..., 1)$ and reference matrix Λ.
2. Generate n $(n = 100, 250, 1000, 10000)$ N-dimensional $(N = 50)$ random vectors with the mixture distribution.
3. Calculate the sample network characteristics.
4. Calculate the loss function (measure of difference between reference and sample network characteristics).
5. In order to estimate the robustness of network characteristic the experiment is repeated 10000 times and the estimate of the risk function is calculated.

Typical results are presented below.

5.3 *Distribution of edge weights*

In figures 4 and 5 the measure of difference between reference and sample distributions of the edge weights as a function of the mixture parameter ε is presented. The results of the experiments show that this measure is robust to a change of ε in the case when the distribution of edge weights is estimated by the frequency of signs coincidence (solid and dash-doted lines) unlike the case when it is estimated by the sample Pearson correlation (dashed line). For $\varepsilon < 0.6$ $(\varepsilon > 0.6)$ procedures based on the frequency of sign coincidence lead to less (more) errors compared to procedures based on sample Pearson correlations. This conclusion is holds for both cases of known and unknown shift parameter.

Similar results for the distribution of the edge weights were obtained for reference models of other markets.

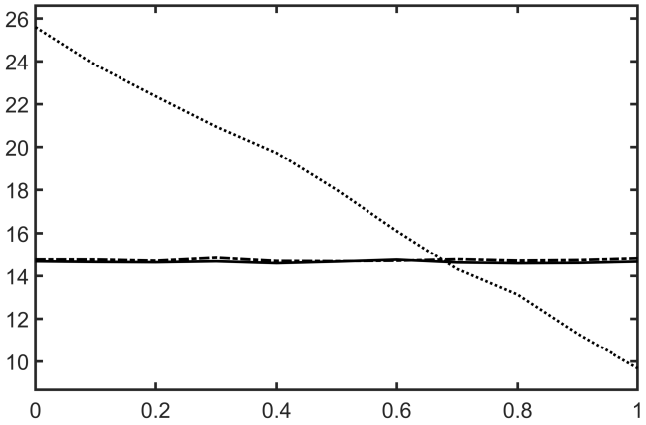

FIG. 4. The dependence of the difference measure between the reference distribution of weights of edges and it's estimation from ε. France, 2012. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

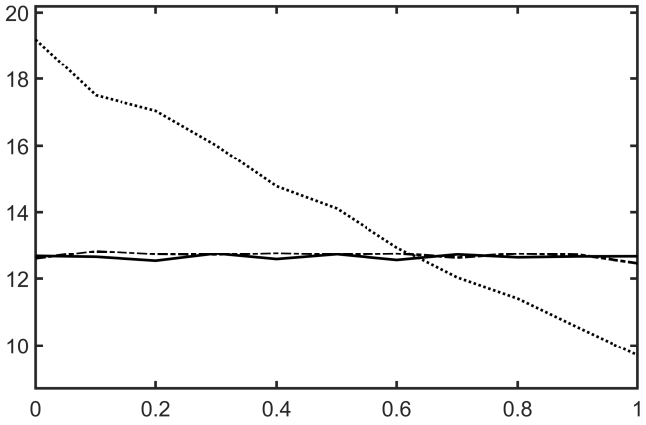

FIG. 5. The dependence of the difference measure between the reference distribution of weights of edges and it's estimation from ε. India, 2011. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

5.4 *Degree distribution in the market graph*

Figures 6 and 7 present the measure of difference between reference and sample degree distributions in the market graphs as function of mixture parameter. The results of the experiments show that this measure is robust to a change of ε when the degree distribution is estimated by the frequency of

sign coincidence (solid and dash-doted lines) unlike the sample Pearson correlation (dashed line). The obtained results show that for $\gamma_0 = 0.3$ and $\varepsilon < 0.6$ ($\varepsilon > 0.6$) procedures based on the frequency of sign coincidence lead to less (more) errors compared to procedures based on the sample Pearson correlations. However, for $\gamma_0 = 0.1$ procedures based on the frequency of sign coincidence lead to less (more) errors compared to procedures based on sample Pearson correlations for $\varepsilon < 0.4$ ($\varepsilon > 0.4$). These conclusions are hold for both cases of known and unknown shift parameter. Similar results for the degree distribution were obtained for reference models of other markets.

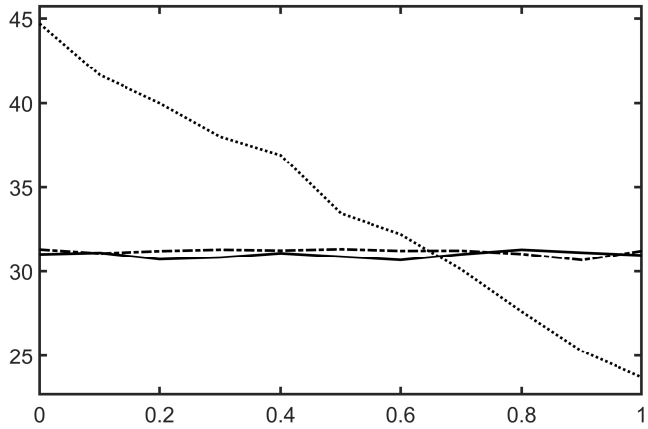

FIG. 6. The dependence of the difference measure between the reference degree distribution in market graph and it's estimation from ε. $\gamma_0 = 0.3$. Brazil, 2011. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

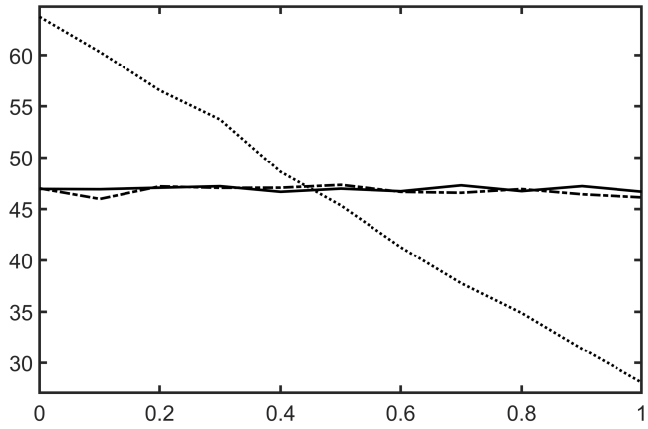

FIG. 7. The dependence of the difference measure between the reference degree distribution in market graph and it's estimation from ε. $\gamma_0 = 0.1$. USA, 2012. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

5.5 *Cliques and Independent sets*

Figure 8 and 9 present measure of difference between reference and sample independent sets and between reference and sample cliques as function of ε. The results of the experiments show that this measure is robust to a change of ε when independent sets and cliques are estimated by the frequency of signs coincidence (solid and dash-doted lines) unlike the sample Pearson correlation (dashed line). For $\gamma_0 = 0.5$ and $\varepsilon < 0.5$ ($\varepsilon > 0.5$) identification procedures for independent sets based on the frequency of sign coincidence lead to less (more) errors compared to procedures based on sample Pearson correlations. For $\gamma_0 = 0.5$ and $\varepsilon < 0.6$ ($\varepsilon > 0.6$) identification procedures for cliques based on the frequency of sign coincidence lead to less (more) errors compared to procedures based on sample Pearson correlations. These conclusions are hold for both cases of known and unknown shift parameter. Similar results were obtained for reference models of other stock markets.

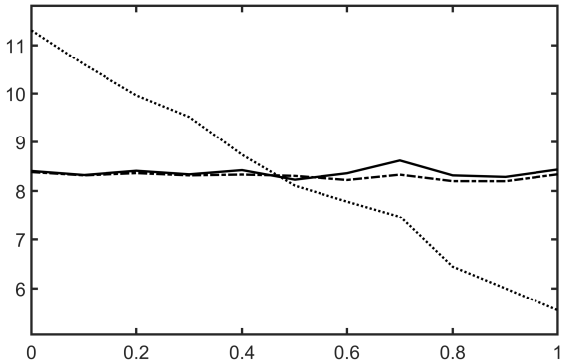

FIG. 8. Measure of difference between the reference independent set and it's estimation as function of ε. $\gamma_0 = 0.5$. Germany, 2010. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

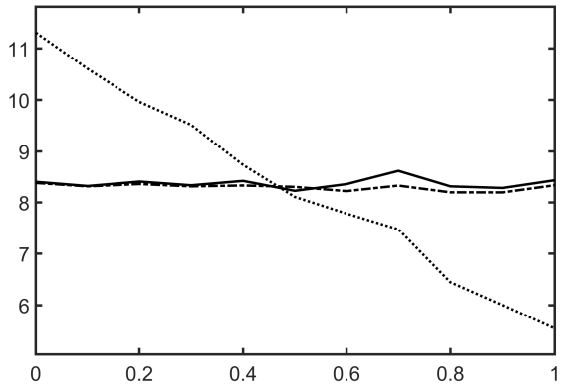

FIG. 9. Measure of difference between the reference clique and it's estimation as function of ε. $\gamma_0 = 0.5$. Great Britain, 2012. $n = 250$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

5.6 *Degree distribution in the maximum spanning tree*

Figures 10 and 11 present the measure of difference between reference and sample degree distribution of MST as function of ε. The results of the experiments show that this measure is robust to a change in the parameter of the mixture ε when the degree distribution in the MST is estimated by the frequency of signs coincidence (solid and dash-doted lines) unlike the sample Pearson correlation (dashed line). For $\varepsilon < 0.6$ ($\varepsilon > 0.6$) procedures based on the frequency of signs coincidence lead to more (less) probability of true decision compared to procedures based on sample Pearson correlations. This conclusion is holds for both cases of known and unknown shift parameter. Similar results were obtained for reference models of other stock markets.

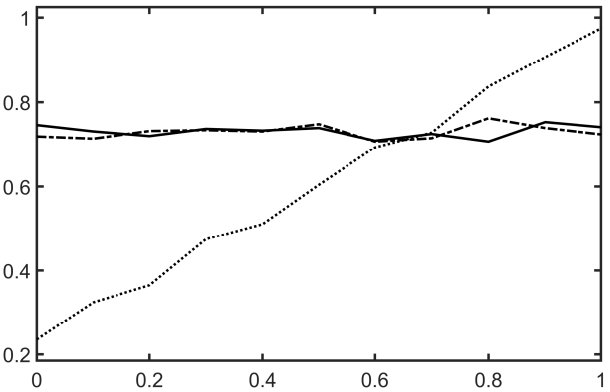

FIG. 10. Measure of difference between the degree distribution in reference MST and it's estimation from ε. China, 2012. $n = 10000$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

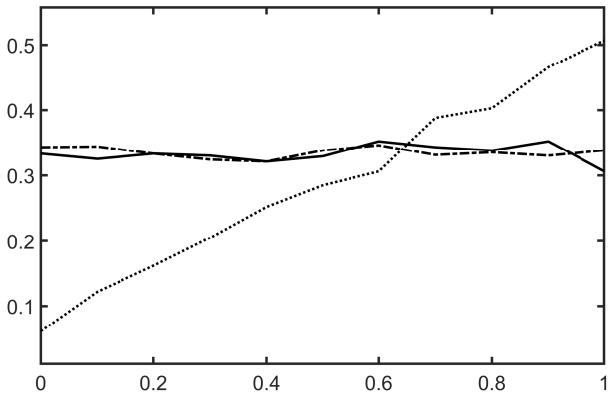

FIG. 11. Measure of difference between the degree distribution in reference MST and it's estimation from ε. Russia, 2012. $n = 10000$ observations. Solid line - frequency of sign coincidence with known μ, dash-doted line - frequency of sign coincidence with unknown μ, dashed line - sample Pearson correlation.

6. Concluding remarks

The paper proposed a methodology of uncertainty analysis for the following market network characteristics: distribution of weights of edges, vertex degree distribution in the market graph, cliques and independent sets in the market graph, and the vertex degree distribution of the maximum spanning tree. This methodology is applied for different markets and different periods of observations. The presented results show that identification procedures based on the sample Pearson correlation depends on the returns distribution (are not robust or distribution free). On the other hand, the procedures based on sample sign similarity (frequency of sign coincidence) are robust (distribution free). Therefore, the procedures based on sample sign similarity are more useful when there is no information of the distribution of stock market returns. Another application of sample sign similarity is related with the mean-variance optimal portfolio in the case of unknown distribution. The estimation of Pearson correlations by sample sign similarity is robust with respect to the distribution of returns and therefore is more suitable for portfolio management. The theoretical foundation of the obtained results can be based on the approach developed in Kalyagin *et al.* (2017) and is outside the scope of this paper. It will be a subject of further publications.

ACKNOWLEDGEMENTS The results of sections 1 - 4 of the article were prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE). The results of section 5 were obtained with the support of the RFFI grant 18-07-00524 and RFFI grant 19-31-90088.

References

ANDERSON T.W. (2003). An Introduction to Multivariate Statistical Analysis. NewYork: Wiley-Interscience.

BALTAS, I., YANNACOPOULOS, A.N. (2019) Portfolio management in a stochastic factor model under the existence of private information // IMA Journal of Management Mathematics, 30, P. 77-103.

BAUTIN G.A., KOLDANOV A.P., PARDALOS P.M. (2014). Robustness of Sign Correlation in Market Network Analysis // Springer Optimization and Its Applications. Vol. 100. P. 25-33.

BAUTIN G.A., KALYAGIN V.A., KOLDANOV A.P., KOLDANOV P.A., PARDALOS P.M. (2013). Simple Measure of Similarity for the Market Graph Construction // Computational Management Science. Vol. 10. P. 105-124.

BOGINSKI V., BUTENKO S., PARDALOS P.M. (2005). Statistical Analysis of Financial Networks // Journal Computational Statistics and Data Analysis. Vol. 48 (2). P. 431-443.

CHI K., TSE, LIU, JING, AND LAU, FRANCIS C. M. (2010). A Network Perspective of the Stock Market // Journal of Empirical Finance 17(4): P. 659-667.

CORONNELLO C., TUMMINELLO M., LILLO F., MICCICHI S., MANTEGNA R.N. (2005). Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange // Acta Physica Polonica B. Vol.36. P. 2653-2679.

DONG, Y., WANG, G., CHUENYUEN, K. (2018) Correlated default models driven by a multivariate regime-switching shot noise process // IMA Journal of Management Mathematics, 29, P. 351-375.

GARAS A., ARGYRAKIS P. (2007). Correlation Study of the Athens Stock Exchange // Physica A. Vol. 380. P. 399-410.

GUPTA F.K.(2013) Elliptically Contoured Models in Statistics and Portfolio Theory // F.K. Gupta, T. Bodnar, T. Varga Springer.

JUNG W.S., CHAE S., YANG J.S., MOON H.T. (2006). Characteristics of the Korean Stock Market Correlations // Physica A: Statistical Mechanics and its Applications, Vol. 361.Issue 1, P. 263-271.

HERO, A. AND RAJARATNAM, B. (2012) Hub discovery in partial correlation graphs // IEEE Trans. Inform. Theor., 58, P. 6064-6078.

HUANG W.-Q., ZHUANGX.-T., YAOS. (2009). A Network Analysis of the Chinese Stock Market // Physica A: Statistical Mechanics and its Applications, Vol. 388. P.2956-2964.

HUBER P.J. (1981). Robust Statistics. Wiley, New York.

KALYAGIN V., KOLDANOV A., KOLDANOV P., ZAMARAEV V. (2014). Market Graph and Markowitz Model, in: Optimization oi Science and Engineering (In Honor of the 60th Birthday of Panos M. Pardalos). NY : Springer. Ch. 15. P. 301-313.

KALYAGIN V.A., KOLDANOV A.P., KOLDANOV P.A. (2017). Robust Identification in Random Variables Networks // Journal of Statistical Planning and Inference. Vol. 181, P. 30-40.

KARA G., OZMEN A. AND WEBER G. W. (2019) Stability advances in robust portfolio optimization under parallelepiped uncertainty // Central European Journal of Operations Research, 27(1), 241261. 27: 241.

KENDALL, M. G. (MAURICE GEORGE)., STUART, A. (1979). The advanced theory of statistics (4th ed). Griffin, London

MANTEGNA R.N. (1999). Hierarchical Structure in Financial Market // European Physical Journal. Series B. Vol. 11. P. 193-197.

MARTI, G., NIELSEN, F., BIKOWSKI, M., DONNAT, P.(2019) A review of two decades of correlations, hierarchies, networks and clustering in financial markets, arXiv:1703.00485a

NETWORK MODELS IN ECONOMICS AND FINANCE. (V.A. KALYAGIN, P.M. PARDALOS, M.R. THEMISTOCLES EDS.) (2014) // Springer Optimization and Its Applications, v. 100.

NGUYEN, Q.; NGUYEN, N.K.K.; NGUYEN, L.H.N. (2019) Dynamic topology and allometric scaling behavior on the Vietnamese stock market // Physica A: Statistical Mechanics and its Applications, 514, P. 235-243.

PEREIRA EJAL, FERREIRA PJS, DA SILVA MF, MIRANDA JGV, PEREIRA HBB (2019) Multiscale network for 20 stock markets using DCCA // Physica A: Statistical Mechanics and its Applications, Volume 529, 121542.

SHEVLYAKOV G. L., OJA H. (2016) Robust Correlation: Theory and Applications. Wiley Series in Probability and Statistics.

TABAK B.M., THIAGO R.S., CAJUEIRO D.O. (2010). Topological Properties of Stock Market Networks: The Case of Brazil // Physica A: Statistical Mechanics and its Applications. Vol. 389. P. 3240-3249.

TUMMINELLO M., LILLO F., MANTEGNA R.N. (2010). Correlation, Hierarchies and Networks in Financial Markets // Journal of Economic Behavior Organization. Vol. 75. P. 40-58.