Statistical Uncertainty of Market Network Structures

Petr Koldanov
Department
of Applied Mathematics and Informatics
National Research University
Higher School of Economics
Nizhny Novgorod, Russia
Email: pkoldanov@hse.ru

Panos M. Pardalos
Center for Applied Optimization
University of Florida
Gainesville, FL, USA
Email: pardalos@ise.ufl.edu

Victor Zamaraev
Laboratory of Algorithms
and Technologies for Network Analysis
National Research University
Higher School of Economics
Nizhny Novgorod, Russia
Email: vzamaraev@hse.ru

***Abstract*—A common network representation of the stock market is based on correlations of time series of return fluctuations. It is well-known that financial time series have a stochastic nature. Therefore, there is uncertainty in inferences about filtered structures in market network. Thus, market network analysis needs to be complemented by estimation of uncertainty of the obtained results. However, as far as we know there are no relevant research in the literature. In the present paper we make the first step in this direction. We propose the approach to measure statistical uncertainty of different market network structures. This approach is based on conditional risk for corresponding multiple decision statistical procedures. The proposed approach is illustrated by numerical evaluation of statistical uncertainty for popular network structures. Our experimental study validates the possibility of application of the approach for comparison of uncertainty of different network structures.**

***Keywords*–Statistical uncertainty; Market network model; Conditional risk; Minimum Spanning Tree; Market Graph.**

I. Introduction

Network models of financial markets attract a growing attention last decades [1]–[8]. A common network representation of the stock market is based on correlations of return fluctuations. In such a representation, each stock corresponds to a vertex and a link between two vertices is estimated by sample correlation of corresponding returns. The obtained network is a complete weighted graph. In order to simplify the network and preserve the significant information, different filtering techniques are used in the literature.

One of the filtering procedures is the extraction of a minimal set of important links associated with the highest degree of similarity belonging to the Minimum Spanning Tree (MST) [1]. To construct the MST a greedy algorithm is used. A list of edges is sorted in descending order according to the weight and following the ordered list an edge is added to the MST if and only if it does not create a cycle. The MST was used to find a topological arrangement of stocks traded in a financial market, which has associated a meaningful economic taxonomy. This topology is useful in the theoretical description of financial markets and in search of economic common factors affecting specific groups of stocks. The topology and the hierarchical structure associated to it, is obtained by using information present in the time series of stock prices only.

The reduction to a minimal skeleton of links leads to loss of valuable information. To overcome this issue, Tumminello et al. [1] proposed to extent the MST by iteratively connecting the most similar nodes until the graph can be embedded on a surface of a given genus $g = k$. For example, for $g = 0$ the resulting graph is planar, which is called Planar Maximally Filtered Graph (PMFG). It was concluded by Tumminello et al. [1] that the method is very efficient in filtering relevant information about the connection structure both of the whole system and within obtained clusters.

Another filtering procedure, proposed by Boginski et al. [2], leads to the concept of Market Graph. A Market Graph (MG) is obtained from the original network by removing all edges with weights less than a specified threshold $\theta \in [-1, 1]$. Maximum cliques and maximum independent sets analysis of the Market Graph were used to obtain valuable knowledge about the structure of the stock market.

All these approaches use time series observations. It is well-known that financial time series have a stochastic nature. Therefore, there is uncertainty in inferences about filtered structures (MST, PMFG, MG) in market network. It is clear that the less numbers of observations one has the less this inferences are reliable. Thus, market network analysis needs to be complemented by estimation of uncertainty of the obtained results.

The main question is: how one can measure and compare uncertainty of different network structures, such as MST, PMFG, MG and others? To answer this question we propose to use the concept of statistical decision functions [9] and to consider statistical uncertainty. Within the framework of this approach, we introduce a measure of statistical uncertainty of market network structures. This allows to identify the most reliable network structures.

The paper is organized as follows. In Section II, we describe the approach and introduce the measure of statistical uncertainty of market network structures. In Section III, we give the results of the numerical simulations. In Section IV, we make concluding remarks.

II. Measure of statistical uncertainty

Let N be a number of stocks, n be a number of days of observations. In our study financial instruments are character-

 ISBN: 978-1-61208-358-2

ized by daily returns of the stocks. Stock k return for day t is defined as

$$R_k(t) = \ln \frac{P_k(t)}{P_k(t-1)}, \qquad (1)$$

where $P_k(t)$ is the price of stock k on day t. We assume that for fixed k, $R_k(t)$, $t = 1, \ldots, n$, are independent random variables with the same distribution as R_k (i.i.d.) and the random vector $R = (R_1, \ldots, R_N)$ has multivariate distribution with correlation matrix

$$||\rho_{ij}|| = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1N} \\ \cdots & \cdots & \cdots \\ \rho_{N1} & \cdots & \rho_{NN} \end{pmatrix}. \qquad (2)$$

For this model we introduce the *reference network*, which is a complete weighted graph with N nodes and weight matrix $||\rho_{ij}||$. For the reference network one can consider corresponding reference structures, e.g., reference MST, reference PMFG, reference Market Graph and others.

Let $r_k(t), k = 1, \ldots, N, t = 1, \ldots, n$, be the observed values of returns. Define the *sample covariance*

$$s_{ij} = \frac{1}{n-1} \sum_{t=1}^{n} (r_i(t) - \overline{r}_i)(r_j(t) - \overline{r}_j), \qquad (3)$$

and *sample correlation*

$$r_{ij} = \frac{s_{i,j}}{\sqrt{s_{i,i}s_{j,j}}} \qquad (4)$$

where $\overline{r}_i = \frac{1}{n} \sum_{t=1}^{n} r_i(t)$. Using the sample correlations we introduce the (*n-period*) *sample network*, which is a complete weighted graph with N nodes and weight matrix $||r_{ij}||$. For the sample network one can consider the corresponding sample structures, e.g., sample MST, sample PMFG, sample Market Graph and others.

To handle statistical uncertainty we propose to compare the sample network with the reference network. Our comparison will be based on conditional risk connected with possible losses. The associated loss function is defined following Koldanov et al. [10] within the framework of multiple decision theory [11].

For a given structure $\mathcal{S}$, we introduce a set of hypothesis:

- h_{ij}: edge between vertices i and j is not included in the reference structure $\mathcal{S}$;
- k_{ij}: edge between vertices i and j is included in the reference structure $\mathcal{S}$.

To measure the losses, we consider two types of errors:

Type I error: edge is included in the sample structure when it is absent in the reference structure;
Type II error: edge is not included in the sample structure when it is present in the reference structure.

Let a_{ij} be the loss associated with the error of the first kind and b_{ij} the loss associated with the error of the second kind for the edge (i, j). According to the statistical decision theory [9] and taking into account additivity of the loss function [10], [11] we define the conditional risk for a given structure $\mathcal{S}$ as

$$\mathcal{R}(\mathcal{S}, n) = \sum_{1 \le i < j \le N} [a_{ij} P_n(d_{k_{ij}}|h_{ij}) + b_{ij} P_n(d_{h_{ij}}|k_{ij})], \quad (5)$$

where $P_n(d_{k_{ij}}|h_{ij})$ is the probability of rejecting hypothesis h_{ij} when it is true and $P_n(d_{h_{ij}}|k_{ij})$ is the probability of accepting hypothesis h_{ij} when it is false. Conditional risk is appropriate to evaluate the quality of different statstical procedures of identification of given structure. In this paper, we consider the case where $a_{ij} = 1/2M_1$ and $b_{ij} = 1/2M_2$. In this case the conditional risk is equivalent to per-family error rate (PFE) type error [12], which we call *fraction of error*:

$$\mathcal{E}(\mathcal{S}, n) = \sum_{1 \le i < j \le N} \left[\frac{1}{2M_1} P_n(d_{k_{ij}}|h_{ij}) + \frac{1}{2M_2} P_n(d_{h_{ij}}|k_{ij}) \right], \quad (6)$$

where M_1 – is a maximal possible number of type I errors and and M_2 – is a maximal possible number of type II errors.

We say that structure $\mathcal{S}_1$ *is more stable than structure* $\mathcal{S}_2$ if $\mathcal{E}(\mathcal{S}_1, n) < \mathcal{E}(\mathcal{S}_2, n)$ for any number of observations n. In other words statistical uncertainty of structure $\mathcal{S}_1$ is less than statistical uncertainty of structure $\mathcal{S}_2$ if $\mathcal{E}(\mathcal{S}_1, n_1) = \mathcal{E}(\mathcal{S}_2, n_2)$ implies $n_1 < n_2$. We define the *$\mathcal{E}$-measure of statistical uncertainty of structure $\mathcal{S}$ (of level $\mathcal{E}_0$)* as the number of observations $n_{\mathcal{E}}$ such that $\mathcal{E}(\mathcal{S}, n_{\mathcal{E}}) = \mathcal{E}_0$, where $\mathcal{E}_0$ is given value.

III. Results

To illustrate our approach, we consider the network with $N = 250$ nodes and $R \sim N((0, \ldots, 0), ||\rho_{ij}^{US}||), i, j = \overline{1, N}$, where the correlation matrix $||\rho_{ij}^{US}||$ consists of pairwise correlations of daily returns of a set of 250 randomly chosen financial instruments traded in the US stock markets over a period of 365 consecutive trading days in 2010-2011. We use the matrix $||\rho_{ij}^{US}||$ as a weight matrix for our reference network. We will refer to it as the *US reference network*. Note that the only reason of this choice of the stocks is to validate our approach on a correlation matrix based on real data.

To construct the n-period sample network we simulate the sample $x_{11}, \ldots, x_{1N}, \ldots, x_{n1}, \ldots, x_{nN}$ from multivariate normal distribution $N((0, \ldots, 0), ||\rho_{ij}^{US}||), i, j = \overline{1, N}$, $N = 250$. To measure statistical uncertainty of network structure $\mathcal{S}$ we use fraction of errors $\mathcal{E}(\mathcal{S}, n)$, which we estimate in the following way:

1) In the US reference network, find reference structure $\mathcal{S}$.
2) Simulate sample $x_{11}, \ldots, x_{1N}, \ldots, x_{n1}, \ldots, x_{nN}$.
3) Calculate estimations r_{ij} of parameters ρ_{ij}^{US}.
4) In sample network (with weight matrix $||r_{ij}||$), find sample structure $\mathcal{S}$.
5) Calculate fraction of errors of type I, fraction of errors of type II and total fraction of error.
6) Repeat many times steps 1-5 and calculate $\mathcal{E}(\mathcal{S}, n)$.

In our experiments, we choose a level of statistical uncertainty $\mathcal{E}_0 = 0.1$.

A. Statistical uncertainty of MST

Observe that if total fraction of error $X = 0$ then reference MST and sample MST are equal; if total fraction of error $X = 1$ then reference MST and sample MST are completely different, i.e., have no common edges. The latter situation may hold for several sample MSTs under fixed reference MST. For Minimum Spanning Tree, one has $M_1 = M_2 = N - 1$, where N is a number of vertices in considered network. Note that in MST a number of errors of type I X_1 is equal to a number of errors of type II X_2. Measures of statistical uncertainty for MST can be defined from the equation:

$$\frac{1}{2(N-1)} \sum_{1 \leq i < j \leq N} [P_n(x_1^{ij} = 1) + P_n(x_2^{ij} = 1)] =$$

$$= \frac{1}{(N-1)} \sum_{1 \leq i < j \leq N} P_n(x_1^{ij} = 1) = \mathcal{E}_0, \qquad (7)$$

where $x_1^{ij} = 1$ if edge (i, j) is incorrectly included into sample structure and $x_1^{ij} = 0$ otherwise and $x_2^{ij} = 1$ if edge (i, j) is incorrectly not included into sample structure and $x_2^{ij} = 0$ otherwise. One has

$$X_1 = \sum_{1 \leq i < j \leq N} x_1^{ij}; \quad X_2 = \sum_{1 \leq i < j \leq N} x_2^{ij}; \qquad (8)$$

$$X = \frac{1}{2} \left(\frac{X_1}{M_1} + \frac{X_2}{M_2} \right). \qquad (9)$$

Results of the study of statistical uncertainty of MST are presented in Figure 1. As one can see, the condition $\mathcal{E}(\text{MST}, n) \leq 0.1$ is achieved when the number of observed periods $n_{\mathcal{E}}$ is more than 10 000. Note that when $n = 1000$ sample and reference MSTs have only 70% of common edges. Moreover, by further increasing the number of observations does not lead to considerable decrease of statistical uncertainty of MST.

B. Statistical uncertainty of PMFG

Observe that $X = 0$ means that reference PMFG and sample PMFG are equal; $X = 1$ means that reference PMFG and sample PMFG are completely different, i.e., have no common edges. The latter situation may hold for several sample PMFGs under fixed reference PMFG. For Planar Maximally Filtered Graph, one has $M_1 = M_2 = 3N - 6$, where N is a number of vertices in considered network. For each edge (i, j) such that $x_1^{ij} = 1$ there is an edge (k, s) with $x_2^{ks} = 1$ and vice versa. It means that in PMFG a number of errors of type I is equal to a number of errors of type II, i.e., $X_1 = X_2$. Since M_1 and M_2 are constants, both measures of statistical uncertainty for PMFG are equivalent and can be defined from the equation:

$$\frac{1}{(3N-6)} \sum_{1 \leq i < j \leq N} P_n(x_1^{ij} = 1) = \mathcal{E}_0. \qquad (10)$$

Results of the study of statistical uncertainty of PMFG are presented in Figure 1. As one can see, the condition $\mathcal{E}(\text{PMFG}, n) \leq 0.2$ is not achieved even when the number of observed periods $n_{\mathcal{E}}$ is equal to 10 000.

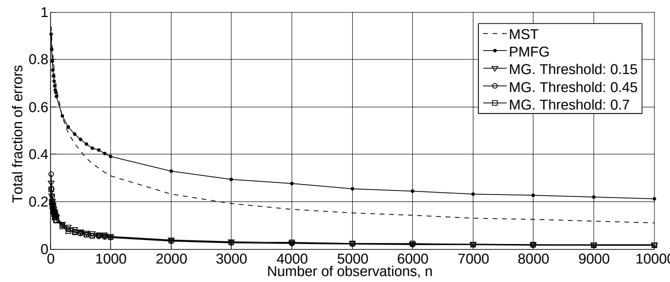

Figure 1. Total fraction of errors in PMFG, MST and MG.

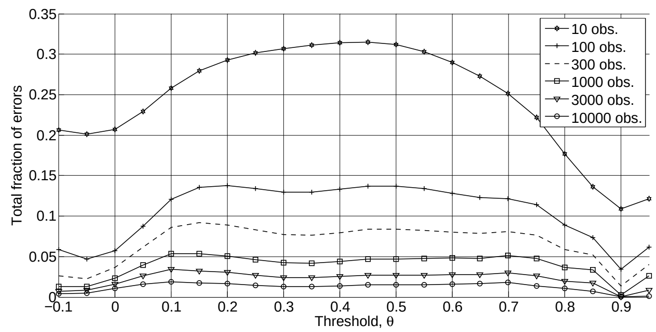

Figure 2. Total fraction of errors in Market Graphs.

C. Statistical uncertainty of MG

Observe that $X = 0$ means that reference MG and sample MG are equal; $X = 1$ means that sample MG is complement to reference MG. Let us pay attention that the latter situation for Market Graph is possible in only one case, in contrast to MST. For Market Graph one has $M_1 = \binom{N}{2} - M, M_2 = M$, where N is the number of the vertices in the considered network and M is the number of edges in the given reference Market Graph. Since M_1 and M_2 are constants, both measures of statistical uncertainty for MG are equivalent and can be defined from the equation:

$$\frac{1}{2} \sum_{1 \leq i < j \leq N} \left[\frac{1}{\binom{N}{2} - M} P_n(x_1^{ij} = 1) + \frac{1}{M} P_n(x_2^{ij} = 1) \right] = \mathcal{E}_0. \qquad (11)$$

Results of the study of statistical uncertainty of MG are presented in Figures 1 and 2. As one can see, the condition $\mathcal{E}(\text{MG}, n) \leq 0.1$ is achieved under the number of observed periods $n_{\mathcal{E}} = 300$ for all thresholds $\theta \in [-0.1, 1]$, which is much more reasonable than the statistical uncertainty of MST.

IV. Conclusion

In the present paper, we introduced the measure of statistical uncertainty for different network structures, which is based on average fraction of errors known as per-family error rate in the theory of multiple comparison statistical procedures [12]. This measure is the particular case of conditional risk.

To illustrate our approach we consider the network where the correlation matrix consists of pairwise correlations of daily returns of a set of 250 randomly chosen financial instruments traded in the US stock markets over a period of 365 consecutive trading days in 2010-2011.

 ISBN: 978-1-61208-358-2

Our experimental study validates the possibility of application of the approach for comparison of uncertainty of different network structures. In particular, in our experiments, Market Graph is more reliable with respect to statistical uncertainty than Minimum Spanning Tree, which in turn is more reliable than Planar Maximally Filtered Graph.

REFERENCES

[1] M. Tumminello, T. Aste, T. Di Matteo, and R. Mantegna, “A tool for filtering information in complex systems,” Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 30, 2005, pp. 10 421–10 426.

[2] V. Boginski, S. Butenko, and P. M. Pardalos, “Statistical analysis of financial networks,” Computational Statistics & Data Analysis, vol. 48, no. 2, 2005, pp. 431–443.

[3] M. A. Djauhari, “A robust filter in stock networks analysis,” Physica A: Statistical Mechanics and its Applications, vol. 391, no. 20, 2012, pp. 5049–5057.

[4] M. A. Djauhari and G. S. Lee, “Minimal spanning tree problem in stock networks analysis: An efficient algorithm,” Physica A: Statistical Mechanics and its Applications, vol. 392, 2013, pp. 2226–2234.

[5] S. Li, J. He, and Y. Zhuang, “A network model of the interbank market,” Physica A: Statistical Mechanics and its Applications, vol. 389, no. 24, 2010, pp. 5587–5593.

[6] A. Namaki, G. Jafari, and R. Raei, “Comparing the structure of an emerging market with a mature one under global perturbation,” Physica A: Statistical Mechanics and its Applications, vol. 390, no. 17, 2011, pp. 3020–3025.

[7] G. A. Bautin, V. A. Kalyagin, A. P. Koldanov, P. A. Koldanov, and P. M. Pardalos, “Simple measure of similarity for the market graph construction,” Computational Management Science, vol. 10, 2013, pp. 105–124.

[8] G.-J. Wang, C. Xie, S. Chen, J.-J. Yang, and M.-Y. Yang, “Random matrix theory analysis of cross-correlations in the us stock market: Evidence from pearson correlation coefficient and detrended cross-correlation coefficient,” Physica A: Statistical Mechanics and its Applications, vol. 392, 2013, pp. 3715–3730.

[9] A. Wald, Statistical decision functions. Oxford, England: Wiley, 1950.

[10] A. P. Koldanov, P. A. Koldanov, V. A. Kalyagin, and P. M. Pardalos, “Statistical procedures for the market graph construction,” Computational Statistics & Data Analysis, vol. 68, 2013, pp. 17–29.

[11] E. L. Lehmann, “A theory of some multiple decision problems,” The Annals of Mathematical Statistics, 1957, pp. 1–25.

[12] Y. Hochberg and A. C. Tamhane, Multiple comparison procedures. John Wiley & Sons, Inc., 1987.

 ISBN: 978-1-61208-358-2