How to measure dynamics of stock market network?

Koldanov P.A., Semenov D.P.

National Research University Higher School of Economics,
Laboratory of Algorithms and Technologiies for Network Analysis (LATNA)
Nijni Novgorod, Russia
pkoldanov@hse.ru

April 20, 2020

Outline

1. Introduction
2. Problem statement
3. Statistical procedure
4. Procedures for CSE($\gamma_0, P^\star$) construction in different random variable network
5. Experimental results
6. Conclusion
7. Publications

Introduction

- Decision on existence of the dynamics is depends on nature of the object. Possible changes could be generated by random nature of the object.
- It is important to recognize random and dynamics changes of the object.
- To this aim it is important to learn the way to make decision with given confidence.
- The aim of the present report is to discuss the question for stock market network.

Network model

Complex system analysis - network model.

- Complete weighted graph $G = (V, \Gamma)$.
- Nodes of the graph - elements of the system.
- Weights of edges - measure $\gamma_{i,j} = \gamma(X_i, X_j)$ of dependence between the elements.

Examples: social networks, market networks, biological networks.

Network structures

Network structures - subgraphs of the network model.
$G' = (V', E') : V' \subseteq V, E' \subseteq E$

- Network structures contain useful information.
- Popular network structures in market network: maximum spanning tree (MST), threshold graph (MG), maximum cliques (MC) and maximum independent sets (MIS).
- threshold graph (TG) of network model $G = (V, \Gamma)$ - subgraph $G'(\gamma_0) = (V', E') : V' = V; E' \subseteq E, E' = \{(i,j) : \gamma_{i,j} > \gamma_0\}$, where γ_0 - given threshold.

Threshold graph identification problem

- Stocks return - random variables.
- **Problem - identify of the threshold graph by observations.**
- Threshold graph identification problem - statistical problem.
- **Threshold graph identification problem:**
  1. **measure of dependence.**
  2. **construct statistical procedure $\delta(x)$ with appropriate properties.**

Random variable network

Random variable network - pair (X, γ):

- $X = (X_1, \ldots, X_N) -$random vector, (Assump:$X = (X_1, \ldots, X_N)$ has elliptical distribution[1])
- $\gamma-$measure of dependence.
- Popular network:=Pearson network: $\gamma^P_{i,j} = \rho_{i,j} = \frac{E(X_i - E(X_i))(X_j - E(X_j))}{\sigma_i \sigma_j}$
- alternative network 1:=sign network: $\gamma^{Sg}_{i,j} = p^{i,j} = P((X_i - E(X_i))(X_j - E(X_j) > 0)$.
- alternative network 2:=Kendall network: $\gamma^{Kd}_{i,j} = 2P(X_i(1) - X_i(2)(X_j(1) - X_j(2)) > 0) - 1$

Random variable network generate network model, which is complete weighted graph $G = (V, \Gamma)$

[1] Def: elliptical density: $f(x) = |\Lambda|^{-\frac{1}{2}} g\{(x-\mu)'\Lambda^{-1}(x-\mu)\}$ where Λ symmetric positive definite matrix, $g(x) \geq 0$, and $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(y'y) dy_1 \ldots dy_N = 1$

Problem statement

- Described approach allows to construct optimal [2] and robust [3] statistical procedure for threshold graph identification in different random variables network.
- **How to construct a procedure to obtain a result with given confidence?**
- The problem is to construct the set of pair of stocks indexes $CSE = \{(i,j)\}$ which contain all $(i,j) : \gamma_{i,j} \geq \gamma_0$ with probability at least $P^\star$ or $CSE(\gamma_0, P^\star)$.
- Subset selection approach (Gupta(2002)), duality between multiple testing and selection (Finner(1994, 1996))

[2] Kalyagin V. A., Koldanov A. P., Koldanov P., Pardalos P. M. Optimal decision for the market graph identification problem in a sign similarity network // Annals of Operations Research. 2018. Vol. 266. No. 1-2. P. 313-327.

[3] V.A. Kalyagin, A.P. Koldanov, P.A. Koldanov. Robust identification in random variable networks. Journal of Statistical Planning and Inference 181 (2017) 30-40

Problem statement

Let (X, γ) - random variable network, $\gamma_{i,j}(\theta) = \gamma(X_i, X_j)$- dependence measure between X_i, X_j. Matrix of connections

$$\Gamma(\theta) = \begin{pmatrix} 1 & \gamma_{1,2}(\theta) & \dots & \gamma_{1,N}(\theta) \\ \gamma_{2,1}(\theta) & 1 & \dots & \gamma_{2,N}(\theta) \\ \dots & \dots & \dots & \dots \\ \gamma_{N,1}(\theta) & \gamma_{N,2}(\theta) & \dots & 1 \end{pmatrix}$$

Let $I = \{(i,j) : i \neq j; i,j = 1, \dots, N\}$- index set of all pair of random variables $X_1, X_2, \dots, X_N$,

$$J(\theta, \gamma_0) = \{(i,j) : \gamma_{i,j}(\theta) > \gamma_0\}; J(\theta) \subset I \tag{1}$$

index set of pair of $X_1, X_2, \dots, X_N$, which correspond to nodes with edges between these nodes under threshold γ_0 in reference threshold graph.

Problem statement

Let $x = (x_1(t), \ldots, x_N(t)), t = 1, \ldots, n$ be a sample on $(X_1, \ldots, X_N)$.

Let $\varphi_{i,j}(x) = \begin{cases} 0, & \text{edge (i,j) is included in the threshold graph} \\ 1, & \text{otherwise} \end{cases}$

Let

$$\Phi(x) = \begin{pmatrix} 0 & \varphi_{1,2}(x) & \ldots & \varphi_{1,N}(x) \\ \varphi_{2,1}(x) & 0 & \ldots & \varphi_{2,N}(x) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_{N,1}(x) & \varphi_{N,2}(x) & \ldots & 0 \end{pmatrix} \tag{2}$$

be a sample threshold graph, i.e. graph constructed by some algorithms applied to the sample.

Problem statement

Let $J(x) = \{(i,j) : \varphi_{i,j}(x) = 0, \forall i \neq j\}$-index set of pair of random variables $X_1, X_2, \ldots, X_N$, which correspond to nodes with edges between these nodes under threshold γ_0 in sample threshold graph.
The problem is to construct procedure for subset selection $J(x) \subset I$, which satisfied

$$P_\theta(J(X) \supset J(\theta, \gamma_0)) \geq P^\star, \forall \theta \in \Omega, \forall \gamma_0 \tag{3}$$

Set $J(X)$, satisfied (3), will be called CSE($\gamma_0, P^\star$).

Statistical procedure

Using duality between multiple testing and selection [Finner(1994)] problem of CSE($\gamma_0, P^\star$) construction can be considered as multiple hypotheses testing problem

$$h_{ij} : \gamma_{i,j} \geq \gamma_0 \text{ versus } k_{ij} : \gamma_{i,j} < \gamma_0 \qquad (4)$$

Note that individual hypotheses (4) satisfied the free combination condition. Therefore problems of coherence and consonance does not arise [Finner(1994)].

Let

$$\varphi_{i,j}(x) = \begin{cases} 0, & T_{i,j}(x) \geq c_{i,j} \\ 1, & T_{i,j}(x) < c_{i,j} \end{cases}$$

be the test of individual hypotheses (4).

Statistical procedure

Set of tests $\varphi_{i,j}(x)$ can be written

$$\Phi(x) = \begin{pmatrix} 0 & \varphi_{1,2}(x) & \dots & \varphi_{1,N}(x) \\ \varphi_{2,1}(x) & 0 & \dots & \varphi_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N,1}(x) & \varphi_{N,2}(x) & \dots & 0 \end{pmatrix} \quad (5)$$

According to [Finner(1994)] if multiple decision procedure (5) control $FWER \leq 1 - P^{\star}$ in strong sence then set of 0 in (5) is set of edges, included in CSE$(\gamma_0, P^{\star})$, i.e. CSE$(\gamma_0, P^{\star})= \{(i,j) : \varphi_{i,j}(x) = 0, i \neq j\}$. There are different multiple decision procedure with FWER control on strong sence. Such procedures are based on combination of individual tests which control significance level. We show by simulations that in some cases combination of individual tests without control of significance level lead to CSE $(\gamma_0, P^{\star})$ also.

Pearson network

In pearson network measure γ is pearson correlation $\gamma_{i,j} = \rho_{i,j}$. Test of individual hypotheses (4) is:

$$\varphi_{ij}^{P}(x) = \begin{cases} 0, & \frac{1}{2\sqrt{n-3}} \left(\ln \frac{1+r_{i,j}}{1-r_{i,j}} - \ln \frac{1+\gamma_0^P}{1-\gamma_0^P} \right) \geq c_{i,j}^P \\ 1, & \frac{1}{2\sqrt{n-3}} \left(\ln \frac{1+r_{i,j}}{1-r_{i,j}} - \ln \frac{1+\gamma_0^P}{1-\gamma_0^P} \right) < c_{i,j}^P \end{cases} \qquad (6)$$

If vector $X = (X_1, \dots, X_N)$ has multivariate normal distribution $N(\mu, \Sigma)$ and $c_{i,j}^P$ - $\frac{1-P^\star}{C_N^2}$ quantile of the distribution, then (5) with tests (6) is Bonferonni multiple hypotheses (4) testing procedure with control $FWER \leq 1 - P^\star$ in strong sence. Then

$$\mathrm{CSE}_{SS}^{P}(\gamma_0^P, P^\star) = \{(i,j) : \varphi_{ij}^P(x) = 0, i \neq j\}.$$

Sign network

Individual test for testing (4) in sign network is

$$\varphi_{i,j}^{Sg} = \begin{cases} 0, & \hat{p}^{i,j} \geq c_{i,j}^{Sg} \\ 1, & \hat{p}^{i,j} < c_{i,j}^{Sg} \end{cases} \quad (7)$$

where

$$\hat{p}^{i,j} = \sum_{t=1}^{n} I_{ij}(t)$$

$$I_{ij}(t) = \begin{cases} 1, & (x_i(t) - \mu_i)(x_j(t) - \mu_j) > 0 \\ 0, & \text{otherwise} \end{cases}$$

As follows from [Kalyagin(2017)] if vector $X = (X_1, \ldots, X_N)$ has elliptical distribution $ECD_N(\mu, \Sigma, g)$ and $c_{i,j}^{Sg}$ is $\frac{1-P^\star}{C_N^2}$-quantile of binomial distribution $b(n, \gamma_0^{Sg})$, where $\gamma_0^{Sg} = \frac{1}{2} + \frac{1}{\pi} \arcsin \gamma_0^P$, then procedure (5) with tests (7) is Bonferonni procedure with $FWER \leq 1 - P^\star$ control in strong sence. Then

$$\text{CSE}_{SS}^{Sg}(\gamma_0^{Sg}, P^\star) = \{(i,j) : \varphi_{ij}^{Sg}(x) = 0, i \neq j\}.$$

Kendall network

Individual test for testing (4) for $\gamma_0 = 0$ in Kendall network (X, γ^{Kd}) is:

$$\varphi_{ij}^{Kd} = \begin{cases} 0, & \sqrt{\frac{9n(n-1)}{2(2n+5)}}\left(T_{ij}^{Kd} - \gamma_0^{Kd}\right) \geq c_{i,j}^{Kd} \\ 1, & \sqrt{\frac{9n(n-1)}{2(2n+5)}}\left(T_{ij}^{Kd} - \gamma_0^{Kd}\right) < c_{i,j}^{Kd} \end{cases} \quad (8)$$

where $\gamma_0^{Kd} = \frac{2}{\pi} \arcsin \gamma_0^P$,

$$T_{ij}^{Kd} = \frac{1}{n(n-1)} \sum_{s=1}^{n} \sum_{t=1}^{n} sign\left((x_i(t) - x_i(s))(x_j(t) - x_j(s))\right)$$

From [Daniels(1947)], [Hoffding(1947)] it follow for vector $X = (X_1, \ldots, X_N)$ with independent components and $c_{i,j}^{Kd} - \frac{1-P^\star}{C_N^2}$-quantile of standard normal, that for $n \to \infty$ procedure (5), (8) is Bonferonni procedure with $FWER \leq 1 - P^\star$ control in strong sence. Then

$$\text{CSE}_{SS}^{Kd}(\gamma_0^{Kd}, P^\star) = \{(i,j) : \varphi_{ij}^{Kd}(x) = 0, i \neq j\}.$$

Experimental results.

Experimental results for CSE($\gamma_0, P^\star$) construction in random variable network $(X, \gamma^P), (X, \gamma^{Sg}), (X, \gamma^{Kd})$ for $X = (X_1, \ldots, X_N)$ $N = 5, 10, 20$ sample size $n = 100, 500, 1000$. Samples are modeled from

$$f(x) = \epsilon f_{Norm}(\mu, \Sigma) + (1 - \epsilon) f_{St}(\mu, \Sigma)$$

where $f_{Norm}(\mu, \Sigma)$–density of multivariate normal distribution, $f_{St}(\mu, \Sigma)$–density of Student distribution with 3 degree of freedom. Parameters μ, Σ are estimated by data of 5, 10, 20 most liquid stocks from USA stock market (tickers are GE, BAC, F, T, CHK, TWTR, ABEV, PBR, PFE, VALE, WFC, BABA, SWN, ITUB, FCX, ORCL, C, VZ, NOK, JCP) from the period 1.01.2018 to 1.01.2019.

Questions

1. Dependence of confidence probability that CSE($\gamma_0, P^\star$) contains all edges of network model with weights at least γ_0, from chosen measure (γ^P, γ^{Sg}, γ^{Kd}), value of mixture parameter ϵ, value of γ_0 (which define the number of edges in the true graph, for example, for $\gamma_0 = 0.1, N = 20$ number of edges is 174) and sample size.
2. Dependence of mean number of edges in CSE($\gamma_0, P^\star$) from measure (γ^P, γ^{Sg}, γ^{Kd}), value of mixture parameter ϵ, value of γ_0 and sample size.
3. Dependence of mean number of edges of the true graph not belonging to CSE($\gamma_0, P^\star$) from measure (γ^P, γ^{Sg}, γ^{Kd}), value of mixture parameter ϵ, value of γ_0 and sample size.

Dependence of confidence probability. Pearson network

Table 1 show, that for Pearson network probability $P^{\star} = 0.9$ that all edges $(i, j) : \gamma_{i,j} \geq \gamma_0$ would belong to CSE$(\gamma_0, P^{\star})$ does not controlled for changing mixture parameter ϵ for the vector X of size $(N = 20)$, $\gamma_0 = 0.1, 0.3$ and sample size $n = 100$ for the case $\epsilon = 0.8$.

γ_0 / ϵ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.40	0.50	0.54	0.62	0.71	0.71	0.80	0.82	0.89	0.94
0.3	0.44	0.52	0.62	0.58	0.72	0.78	0.75	0.86	0.87	0.95
0.5	0.85	0.88	0.90	0.88	0.90	0.91	0.93	0.95	0.97	0.98
0.7	0.97	0.98	0.97	0.97	0.97	0.97	0.99	0.99	0.99	0.99

Table: Estimation of confidence probability CSE$(\gamma_0, P^{\star})$ in Pearson network $(N = 20, n = 100)$.

Dependence of confidence probability. Sign network

Table 2 show that confidence probability CSE($\gamma_0, P^\star$) in sign network does not depend from mixture parameter. Moreover experimental results shows that the probability is strongly higher than specified value.

γ_0 / ϵ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.99	0.99	0.98	0.99	0.98	1.00	0.99	0.98	0.99	0.99
0.3	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	1.00
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
0.7	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table: Estimation of confidence probability CSE($\gamma_0, P^\star$) in sign network ($N = 20, n = 100$).

Dependence of confidence probability. Kendall network

Properties of Kendall test (8) with respect to mixture parameter ϵ was studied in [Koldanov(2019(2))]. It was shown that significance level of test (8) depend from mixture parameter ϵ. It is interesting to note that table 3 show that confidence probability CSE($\gamma_0, P^\star$) in Kendall network does not depend from mixture parameter. Moreover experimental results shows that the probability is strongly higher than specified value.

γ_0 /ϵ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.97	0.98	0.99	0.98	0.97	0.99	0.99	0.97	0.98	0.99
0.3	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99
0.5	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99
0.7	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00

Table: Estimation of confidence probability CSE($\gamma_0, P^\star$) in Kendall network ($N = 20, n = 100$).

Mean number of edges. Pearson network

In table 4 average size of CSE($\gamma_0, P^\star$) in Pearson network is presented. The CSE($\gamma_0, P^\star$) is constructed by $n = 100$ observations vector X size is $N = 20$ with respect to threshold γ_0 and mixture parameter ϵ.

γ_0 / ϵ	0	0.2	0.4	0.6	0.8	1
0.1(174)	184.40	186.16	187.38	187.82	189.01	189.93
0.3(72)	169.24	173.83	175.93	177.85	181.15	184.26
0.5(12)	117.50	118.55	118.85	119.15	122.01	120.15
0.7(3)	33.22	31.37	29.06	26.01	21.90	19.21

Table: Average size of CSE($\gamma_0, P^\star$) in Pearson network.

In table 5 average size of CSE($\gamma_0, P^\star$) in sign network is presented. The CSE($\gamma_0, P^\star$) is constructed by $n = 100$ observations vector X size is $N = 20$ with respect to threshold γ_0 and mixture parameter ϵ.

γ_0 / ϵ	0	0.2	0.4	0.6	0.8	1
0.1(174)	189.93	189.95	189.93	189.93	189.96	189.95
0.3(72)	188.52	188.4	188.5	188.57	188.43	188.47
0.5(12)	165.21	164.21	165.08	165.11	164.01	166.05
0.7(3)	76.07	76.24	76.5	77.34	76.72	75.69

Table: Average size of CSE($\gamma_0, P^\star$) in sign network

Mean number of edges. Kendall network

In table 6 average size of CSE($\gamma_0, P^\star$) in Kendall network is presented. The CSE($\gamma_0, P^\star$) is constructed by $n = 100$ observations vector X size is $N = 20$ with respect to threshold γ_0 and mixture parameter ϵ.

γ_0 / ϵ	0	0.2	0.4	0.6	0.8	1
0.1(174)	189.94	189.93	189.94	189.95	189.96	189.95
0.3(72)	188.37	188.54	188.38	188.39	188.56	188.62
0.5(12)	165.14	164.85	165.46	163.84	164.64	164.1
0.7(3)	77.93	75.68	77.73	77.38	76.19	77.2

Table: Average size of CSE($\gamma_0, P^\star$) in Kendall network

Mean number of true edges in CSE($\gamma_0, P^\star$). Pearson network

In table 7 mean number of true edges in CSE($\gamma_0, P^\star$) in Pearson network is presented. The CSE is constructed by $n = 100$ observations vector X size is $N = 20$ with respect to threshold γ_0 and mixture parameter ϵ. Obtained results shows maximal mean error is equal to 4. This mean that maximal number of true edges which does not belong to CSE($\gamma_0, P^\star$) in Pearson network is equal to 4.

γ_0 / ϵ	0	0.2	0.4	0.6	0.8	1
0.1	169.80	171.27	172.23	172.42	173.39	173.99
0.3	69.66	70.59	70.82	71.19	71.55	72.00
0.5	11.76	11.85	11.84	11.88	11.95	12.00
0.7	2.95	2.95	2.97	2.98	2.98	3.00

Table: Mean number of true edges in CSE($\gamma_0, P^\star$). Pearson network

Conclusion

- Individual test in sign network does control significance level for $0 < \epsilon < 1$.
- Individual test in Kendall network does not control significance level for $0 < \epsilon < 1$.
- However sufficient sets for edges in sign and Kendall network do control $P^\star$.
- Sufficient sets for edges in Pearson network do not control $P^\star$.
- However maximal number of true edges which does not belong to $CSE(\gamma_0, P^\star)$ in Pearson network is small (equal to 4).
- CSE in Pearson network are smaller than CSE in sign and kendall networks.

Publications.

- Gupta, S.S.
  MULTIPLE DECISION PROCEDURES. Theory and Methodology of Selecting and Ranking Populations/S.S. Gupta, S. Panchapakesan
  SIAM, 2002. —568 pp.
- H. Finner, G. Giani Closed subset selection procedures for selecting good populations. Journal of Statistical Planning and Inference 38 (1994) 179-200.
- V.A. Kalyagin, A.P. Koldanov, P.A. Koldanov. Robust identification in random variable networks. Journal of Statistical Planning and Inference 181 (2017) 30–40
- W. Hoffding. On the Distribution of the Rank Correlation Coefficient When the Variates are Not Independent. Biometrika. 1947. Vol. 34. No. 3/4 pp. 183-196

Publications.

H. E. Daniels and M. G. Kendall. The Significance of Rank Correlations Where Parental Correlation Exists. Biometrika. 1947. Vol. 34. No. 3/4. pp. 197-208.

Koldanov P. Statistics of individual tests for market graph identification in market network. // Advances in Computer Science Research. 2019. P. 50-55.

THANK YOU FOR YOUR ATTENTION!