Statistical procedures for network structures identification

Petr Koldanov

National Research University Higher School of Economics,
Laboratoty of Algorithms and Technologies for Network Analysis (LATNA)
Nizhny Novgorod, Russia
Team: Valery Kalyagin, Alexander Koldanov, Panos Pardalos
pkoldanov@hse.ru

S. Peterburg, Russia, April 14, 2017

Outline

1. Introduction
2. Multiple testing procedures for market graph identification
3. Distribution free statistical procedure
4. Optimality
5. Two type of network structures
6. Multiple testing procedures for GGM identification
7. Multiple testing procedures for reduced graph identification
8. Multiple testing procedures for MG identification
9. Appendix. Publications
10. Appendix.Proof of theorem 1
11. Appendix. Data for stability
12. Appendix. Proof of the Theorem 2
13. Appendix. Holm procedure
14. Appendix. Hochberg procedure
15. Appendix. Role of measure of association
16. Appendix. Testing the symmetry conditions

General problem

- $X = (X_1, X_2, \ldots, X_N)$-random vector.
- $f(x) \in \{f(x, \theta); \theta \in \Omega\}$
- $H_i : \theta \in \Omega_i, \Omega_i \subset \Omega, i = 1, \ldots, L$
- $x(1), x(2), \ldots, x(n)$ - sample of finite size.
- Construct $\delta(x) : \mathcal{X} \to D, D = \{d_1, d_2, \ldots, d_L\}$

Network model

One way to analyze a complex system is to consider associated network model.

- Complete weighted graph $G = (V, E, \gamma)$.
- Nodes of the network model - elements of the system.
- Weights of edges in the network model are given by some measure γ of connection between elements of the system.

Examples: social networks, market networks, biological network.

Network structures

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

- Network structures contain useful information on the network model.
- Popular network structures for market network: minimum spanning tree (MST), planar maximally filtered graph (PMFG), market graph (MG), cliques and independent sets of MG.
- Popular network structures for biological network: Gaussian Graphical Model (concentration graph).

History of market network analysis

- Mantegna(1999) - MST for market network.
- Pardalos (2003) - MG for market network.
- Now there are around 3000 papers.
- Main purpose - network structure construction by numerical algorithms to real market data (stock returns) and interpretation of obtained results. Examples of interpretation.

Problem description

- Mathematical point of view - stocks returns are random variables.
- Problem - statictical uncertainty of obtained results.
- Problem of network structures identification - statistical problem.
- **Problem: construct statistical procedure $\delta(x)$ with appropriate properties to identify network structure from observations.**

Random variable network

Random variable network is a pair (X, γ):

- $X = (X_1, \ldots, X_N)-$random vector,
- $\gamma-$measure of association.

Example - market network (nodes correspond to the stocks, behaviour of stocks is described by returns (portfolio theory))

- 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:=Sign similarity network: $\gamma^{Sg}_{i,j} = p^{i,j} = P((X_i - E(X_i))(X_j - E(X_j) > 0)$.

Any random variable network generate network model. Network model is complete weighted graph $G = (V, E, \gamma)$

Problem statement

- (X, γ)-random variable network, $G = (V, E, \gamma)$-generated network model.
- $G' = (V', E') : V' \subseteq V, E' \subseteq E$ - network structure.
- It is known that distribution of X from class $\mathcal{K} = \{(f(x, \theta), \theta \in \Omega)\}$.
- Let $S = (s_{i,j})$, $S \in \mathcal{G}$ - set of all adjacency matrices.
- $H_S : \theta \in \Omega_S$-hypothesis that network structure has adjacency matrix $S, S \in \mathcal{G}$.
- Observation $X(t) = (X_1(t), \ldots, X_N(t)), t = 1, \ldots, n$

Problem: construct statistical procedure $\delta(x)$ with appropriate properties to identify network structure from observations i.e. to select one from disjoint hypotheses $H_S, S \in \mathcal{G}$.

Quality of statistical procedures for network structure identification

- Statistical procedure $\delta(x) = \begin{cases} d_Q, & x \in D_Q \end{cases}$
  $\bigcup_{Q \in \mathcal{G}} D_Q = \mathcal{X}$ is the partition of sample space.
- $\delta(x) = d_Q$ - decision, that network structure has adjacency matrix $Q, Q \in \mathcal{G}$.
- $w(H_S; d_Q) = w(S, Q)$ - loss from the decision d_Q when the hypothesis H_S is true, $w(S, S) = 0, S \in \mathcal{G}$.
- Risk function of statistical procedure $\delta(x)$ is defined by

$$Risk(S, \theta; \delta) = \sum_{Q \in \mathcal{G}} w(S, Q) P_\theta(\delta(x) = d_Q), \quad \theta \in \Omega_S, S \in \mathcal{G}$$

$P_\theta(\delta(x) = d_Q)$ - the probability that decision d_Q is taken while the true decision is d_S.

Problem statement

- to investigate uncertainty of statistical procedures for network structures identification
- to construct optimal, in some sense, procedure
- to construct distribution free, in some sense, procedure

Distribution free statistical procedure.

Let $\mathcal{K}$ - class of distributions of X, such that network models, generated by (X, γ) coincide $\gamma(X_i^{(1)}, X_j^{(1)}) = \gamma(X_i^{(2)}, X_j^{(2)})$. Then network structures also coincide.
But properties of statistical procedures for network structures identification may depend on distribution of $X \in \mathcal{K}$.
Problem: construct distribution free statistical procedure for network structure identification.

Definition: statistical procedure δ is distribution free in class $\mathcal{K}$, if risk function $Risk(S, \theta, \delta)$ does not depend from distribution of vector X from class $\mathcal{K}$ for any S.

Example: class $\mathcal{K}$ of elliptical distributions

Most common models of stock market is the class of elliptical distributions.

$$f(x;\theta) = |\Lambda|^{-\frac{1}{2}} g\{(x-\mu)'\Lambda^{-1}(x-\mu)\} \tag{1}$$

where $\theta = (\mu, \Lambda, g)$, $\mu \in R^N$, Λ - symmetric positive definite matrix, $g(x) \geq 0$, and

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y'y)dy_1 \dots dy_N = 1$$

In the following we assume that μ is known.

Example: class $\mathcal{K}$ of elliptical distributions

Let $\mathcal{K}(\Lambda)$ be the subclass of class of elliptical distributions with fixed Λ. It is known $\gamma^P_{i,j} = \frac{\lambda_{i,j}}{\sqrt{\lambda_{i,i}\lambda_{j,j}}}$.

Then network models, generated by $(X, \gamma^P), X \in \mathcal{K}(\Lambda)$, coincide.

Lemma 1: Probabilities $\gamma^{Sg}_{i,j} = P(X_i X_j > 0)$ are defined by the matrix Λ and does not depend from g.

Then network models, generated by $(X, \gamma^{Sg}), X \in \mathcal{K}(\Lambda)$, coincide.

Multiple testing statistical procedures for market graph (MG) identification

Individual edge hypotheses:

$$h_{ij} : \gamma_{ij} \leq \gamma_0 \text{ vs } k_{ij} : \gamma_{ij} > \gamma_0.$$

Individual tests:

$$\varphi_{ij}(X) = \begin{cases} 1, & t_{ij}(X) > c_{ij} \\ 0, & t_{ij}(X) \leq c_{ij} \end{cases}$$

Multiple testing statistical procedure: statistical procedure, based on statistics of individual tests.

- Single step procedures (Bonferroni and others)
- Stepwise procedures (Holm, Hochberg and others)

Example 1.Pearson network

Individual hypotheses (Pearson measure): $h_{ij} : \rho_{i,j} \leq \rho_0$ vs $k_{ij} : \rho_{i,j} > \rho_0$

- $\partial^P_{i,j}(x_i, x_j) = \begin{cases} 1, & \frac{r_{i,j}-\rho_0}{\sqrt{1-r^2_{i,j}}} > c^{St}_{i,j} \\ 0, & \frac{r_{i,j}-\rho_0}{\sqrt{1-r^2_{i,j}}} \leq c^{St}_{i,j} \end{cases}$ UMP in the class of invariant tests
- $\varphi^P_{i,j}(x_i, x_j) = \begin{cases} 1, & z_{i,j} > c_{i,j} \\ 0, & z_{i,j} \leq c_{i,j} \end{cases}$ asymptotically optimal in the class of invariant tests
  where $z_{i,j} = \sqrt{n}\left(\frac{1}{2}\ln\left(\frac{1+r_{i,j}}{1-r_{i,j}}\right) - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right)$,
- $c_{i,j}$ is $(1-\alpha_{ij})$-quantile of standart normal distribution $N(0,1)$,
- $c^{St}_{i,j}$ is $(1-\alpha_{ij})$-quantile of Student distribution t_{n-1},
- $\alpha_{i,j}$ is the given significance level for individual edge i, j test,

$$r_{i,j} = \frac{\sum_{t=1}^{n}(x_i(t)-\overline{x_i})(x_j(t)-\overline{x_j})}{\sqrt{\sum_{t=1}^{n}(x_i(t)-\overline{x_i})^2 \sum_{t=1}^{n}(x_j(t)-\overline{x_j})^2}}$$

Example 1.Pearson network. Multiple testing procedures

- Single step rules

$$\partial^P(x) = \begin{pmatrix} 1, & \partial^P_{1,2}(x), & \dots, & \partial^P_{1,N}(x) \\ \partial^P_{2,1}(x), & 1, & \dots, & \partial^P_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \partial^P_{N,1}(x), & \partial^P_{N,2}(x), & \dots, & 1 \end{pmatrix}.$$

$$\delta^P(x) = \begin{pmatrix} 1, & \varphi^P_{1,2}(x), & \dots, & \varphi^P_{1,N}(x) \\ \varphi^P_{2,1}(x), & 1, & \dots, & \varphi^P_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi^P_{N,1}(x), & \varphi^P_{N,2}(x), & \dots, & 1 \end{pmatrix}.$$

- Stepwise rules (Holm, Hochberg procedures with individual tests $\partial^P_{i,j}; \varphi^P_{i,j}$)

Example 2. Sign similarity network

- Individual hypotheses: $h_{ij} : p^{i,j} \leq p_0$ vs $k_{ij} : p^{i,j} > p_0$
- $I_{i,j}(t) = \begin{cases} 1, & sign(x_i(t) - \mu_i) = sign(x_j(t) - \mu_j) \\ 0, & \text{else} \end{cases}$
- Define $T_{i,j}^{Sg} = \sum_{t=1}^{n} I_{i,j}(t)$,
- $\varphi_{i,j}^{Sg} = \begin{cases} 0, & T_{i,j}^{Sg} \leq c_{i,j} \\ 1, & T_{i,j}^{Sg} > c_{i,j} \end{cases}$,
  where $c_{i,j}$ is defined from equation: $\sum_{k=c_{i,j}}^{n} \frac{n!}{k!(n-k)!}(p_0)^k(1-p_0)^{n-k} \leq \alpha$

Example 2. Sign similarity network

- Multiple decision single step (Bonferroni type) procedure

$$\delta^{Sg}(x) = \begin{pmatrix} 1, & \varphi^{Sg}_{1,2}(x), & \dots, & \varphi^{Sg}_{1,N}(x) \\ \varphi^{Sg}_{2,1}(x), & 1, & \dots, & \varphi^{Sg}_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi^{Sg}_{N,1}(x), & \varphi^{Sg}_{N,2}(x), & \dots, & 1 \end{pmatrix}.$$

- Holm, Hochberg procedures with the use of statistics $T^{Sg}_{i,j}$

δ^{Sg} is distribution free

Theorem 1. Let random vector $(X_1, \ldots, X_N)$ has elliptically contoured distribution with density

$$f(x;\theta) = |\Lambda|^{-1/2} g((x-\mu)'\Lambda(x-\mu))$$

Then

1. the risk functions $R(S, \theta; \delta_B^{Sg})$, $R(S, \theta; \delta_H^{Sg})$, $R(S, \theta; \delta_{Hg}^{Sg})$ are defined by the matrix Λ and does not depend on the function g for any loss function w, $S-$ adjacency matrix of MG.
2. the risk function $R(H_{MST}, \theta; \delta^{Sg})$ of Kruskal procedure for MST construction are defined by the matrix Λ and does not depend on the function g for any loss function.

δ^{Sg} is distribution free

Lemma 2. Let random vector $(X_1, \ldots, X_N)$ has elliptically contoured distribution. Then the probabilities

$$p(i_1, \ldots, i_N) := P_\Lambda(i_1(X_1 - \mu_1) > 0, \ldots, i_N(X_N - \mu_N) > 0)$$

are defined by the matrix Λ and does not depend on the function g for any $i_k \in \{-1, 1\},\ k = 1, 2, \ldots, N$.

Lemma 3. Let random vector $(X_1, \ldots, X_N)$ has elliptically contoured distribution with density

$$f(x; \mu, \Lambda) = |\Lambda|^{-1/2} g((x - \mu)' \Lambda (x - \mu))$$

Then joint distribution of the statistics $T_{i,j}^{Sg}$ $(i, j = 1, 2, \ldots, N; i \neq j)$ are defined by the matrix Λ and does not depend on the function g.

Comparison

1. Statistical procedures, based on sample Pearson correlations, are not robust in the class of elliptical distributions with fixed Λ.
2. Procedures, based on sample signs correlations, are distribution free in the class of elliptical distributions with fixed Λ.

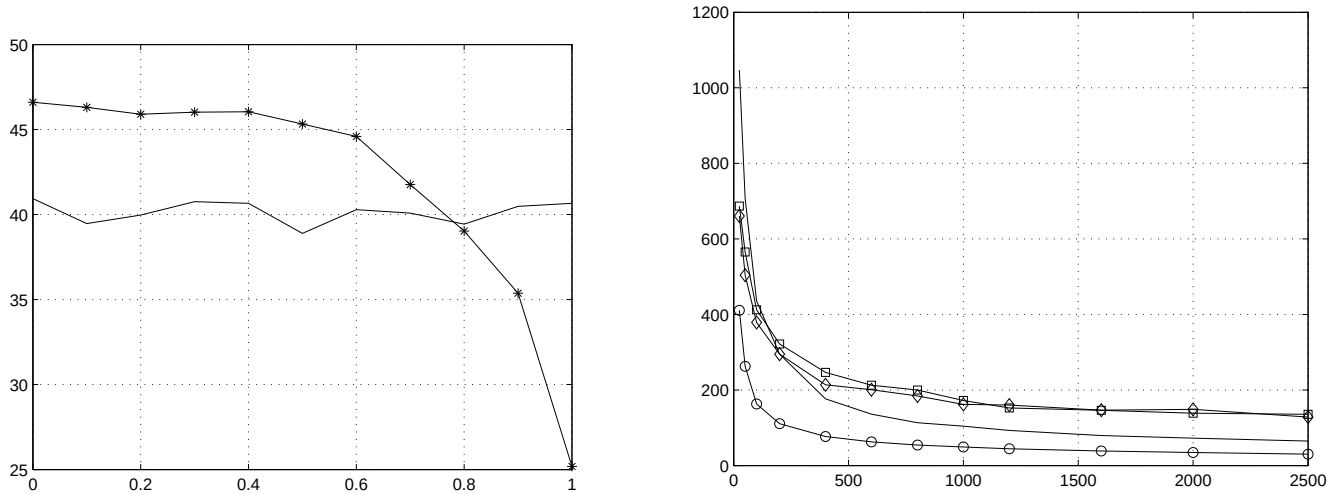

Figure: 2: Risk function for MG, $\rho_0 = 0.64$. Left - $n = 400$, star line - δ^P, line - δ^S, right: circle - $\gamma = 1$, δ^P; diamond - $\gamma = 0,5$, δ^P; square - $\gamma = 0$, δ^P, line - δ^S.The model is the mixture distribution consisting of multivariate normal distribution and multivariate Student distribution with 3 degree of freedom.

Advantages of sign similarity network

Sign similarity network:

- easy to interpretation;
- allow of generalization to any number of random variables;
- statistical procedures in sign similarity network are distribution free;
- distribution free statistical procedures in sign similarity network can be applied for network structure identification in other network models.

Comparison of these two networks

If X has elliptical distribution (1) then network structures in Pearson correlation network are equivalent to network structures in sign similarity network.

Theorem 2: Let vector $X = (X_1, \ldots, X_N)$ has elliptical distribution (1). Then:

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

Corollary 1: MST in network model generated by Pearson correlation network coincide with MST in network model generated by sign similarity network.

Corollary 2: MG with threshold ρ_0 in network model generated by Pearson correlation network coincide with MG with threshold $p_0 = \frac{1}{2} + \frac{1}{\pi} \arcsin(\rho_0)$ in network model generated by sign similarity network.

$\gamma_{i,j}^{P} = 0.1 \Leftrightarrow \gamma_{i,j}^{Sg} = 0.53; \gamma_{i,j}^{P} = 0.6 \Leftrightarrow \gamma_{i,j}^{Sg} = 0.705$

Optimality. Additive loss function

For MG or GGM identification problem it is natural to consider loss functions which are additive.
$a_{i,j}$ - individual loss from false inclusion of edge (i,j) in MG or GGM.
$b_{i,j}$- individual loss from false non inclusion of the edge (i,j).
Let

$$l_{i,j}(S,Q) = \begin{cases} a_{i,j}, & \text{if } s_{i,j}=0, q_{i,j}=1, \\ b_{i,j}, & \text{if } s_{i,j}=1, q_{i,j}=0, \\ 0, & \text{else} \end{cases}$$

Loss function is additive:

$$w(S,Q) = \sum_{i=1}^{N}\sum_{j=1}^{N} l_{i,j} = \sum_{\{i,j:s_{i,j}=0;q_{i,j}=1\}} a_{i,j} + \sum_{\{i,j:s_{i,j}=1;q_{i,j}=0\}} b_{i,j}$$

Problem statement. Individual hypotheses

- Let $\beta = (i, j)$ - edge of network model ($\beta \in E, \beta = 1, \ldots, C_N^2$).
- Let $\omega_{i,j}^{-1} = \omega_\beta^{-1}$ be the set of parameters θ, such that $(i, j) \in G'$, ω_β be the set of parameters θ, such that $(i, j) \notin G'$.
- Define $h_\beta : \theta \in \omega_\beta^{-1}$ vs $k_\beta : \theta \in \omega_\beta$
- $\Omega_S = (\bigcap_{i,j:s_{i,j}=1} \omega_{i,j}^{-1}) \cap (\bigcap_{i,j:s_{i,j}=0} \omega_{i,j})$

Two type of network structures

- Network structures with arbitrary number of elements of network model (MG, GGM):

$$\begin{aligned}
&\Omega_{0,0,\ldots,0} = \bigcap_{\beta\in\{1,\ldots,C_N^2\}} \omega_\beta \\
&\Omega_{1,0,\ldots,0} = \omega_1^{-1} \cap \bigcap_{\beta\in\{2,\ldots,C_N^2\}} \omega_\beta \\
&\ldots \\
&\Omega_{1,1,\ldots,1} = \bigcap_{\beta\in\{1,\ldots,C_N^2\}} \omega_\beta^{-1}
\end{aligned} \tag{3}$$

- Network structures with fixed number of elements of network model (MST): Let E_1-set of edges of MST_1, E_2-set of edges of $MST_2, \ldots,$ E_L-set of edges of $MST_L, |E_i| = N-1, \forall i = 1,\ldots,L$

$$\begin{aligned}
&H_{E_1} = \bigcap_{\beta\in\{E_1\}} \omega_\beta^{-1} \cap \bigcap_{\beta\in\{1,\ldots,C_N^2\}-\{E_1\}} \omega_\beta \\
&H_{E_2} = \bigcap_{\beta\in\{E_2\}} \omega_\beta^{-1} \cap \bigcap_{\beta\in\{1,\ldots,C_N^2\}-\{E_2\}} \omega_\beta \\
&\ldots \\
&H_{E_L} = \bigcap_{\beta\in\{E_L\}} \omega_\beta^{-1} \cap \bigcap_{\beta\in\{1,\ldots,C_N^2\}-\{E_L\}} \omega_\beta
\end{aligned} \tag{4}$$

Statistical procedures for network structure identification with arbitrary number of elements

Let $\varphi_{i,j} = \begin{cases} 1, & x \in A_{i,j}^{-1} \\ 0, & x \in A_{i,j} \end{cases}$ be the test of hypothesis $h_{i,j} = h_\beta$.

Define $\delta(x) = \begin{cases} d_Q, & x \in D_Q = \bigcap_{i,j} A_{i,j}^{\kappa_{Q;i,j}} \end{cases}$, $\kappa_{Q;i,j} = \begin{cases} -1, & q_{i,j} = 1 \\ 1, & q_{i,j} = 0 \end{cases}$

Theorem 1 Let loss function be additive and $a_\beta = a, b_\beta = b, \forall \beta = 1, \ldots, C_N^2$. Then for network structures with arbitrary number of elements

$$R(S, \theta, \delta) = \sum_\beta r(h_\beta, \theta, \varphi_\beta) = aE_\theta(X_1(\delta, S)) + bE_\theta(X_2(\delta, S))$$

where $X_1(X_2)$-number of incorrectly included (non included) elements of network model in network structure.

Statistical procedures for network structure identification with given number of elements

Let $\varphi_{i,j} = \begin{cases} 1, & x \in A_{i,j}^{-1} \\ 0, & x \in A_{i,j} \end{cases}$ be the test of hypothesis h_β.

Define $\delta(x) = \begin{cases} d_Q, & x \in D_Q = \bigcap_{i,j} A_{i,j}^{\kappa_{Q;i,j}} \end{cases}$, $\kappa_{Q;i,j} = \begin{cases} -1, & q_{i,j} = 1 \\ 1, & q_{i,j} = 0 \end{cases}$

Theorem 2 Let loss function be additive, $a_\beta = a, b_\beta = b, \forall \beta = 1, \ldots, C_N^2$ and set of tests $\varphi_{i,j}$ is compatible i.e

$$\sum_{(\kappa_{i\beta_1}, \kappa_{i\beta_2}, \ldots, \kappa_{i\beta_K}): \kappa_{i\beta_{i_1}} = \ldots = \kappa_{i\beta_{i_M}} = -1, \kappa_{i\beta_{i_{M+1}}} = \ldots = \kappa_{i\beta_{i_N}} = 1} P(\bigcap_\beta A_\beta^{\kappa_{i\beta}}) = 1$$

or

$$P(\bigcap_\beta A_\beta^{\kappa_{i\beta}}) = 0 \quad \sum_{\beta: \kappa_{i\beta} = -1} \kappa_{i\beta} \neq -M$$

Then for network structures with given number M of elements

$$R(S, \theta, \delta) = \sum_\beta r(h_\beta, \theta, \varphi_\beta) = (a + b) E_\theta(X_1(\delta, S))$$

Quality of statistical procedures. Unbiasedness.

Decision function $\delta(x)$ is said to be w-unbiased if for all θ, θ'

$$E_\theta w(\theta', \delta(X)) \geq E_\theta w(\theta, \delta(X))$$

"δ is unbiased if on the average $\delta(X)$ comes closer to the correct decision than to any wrong one" In our case it can be written as

$$\sum_{Q \in \mathcal{G}} w(S, Q) P(\delta(x) = d_Q / H_S) \leq \sum_{Q \in \mathcal{G}} w(S', Q) P(\delta(x) = d_Q / H_S),$$

$\forall S, S' \in \mathcal{G}$

- for network structures with arbitrary number of elements
  $aE_\theta(X_1(\delta, S)) + bE_\theta(X_2(\delta, S)) \leq aE_\theta(X_1(\delta, S')) + bE_\theta(X_2(\delta, S'))$
- for network structures with given number of elements
  $E_\theta(X_1(\delta, S)) \leq E_\theta(X_1(\delta, S'))$
  Does not depend on a, b.

General theorem for network structure with arbitrary number of elements

Individual edge hypotheses:

$$h_{ij} : \gamma_{ij} \leq \gamma_0 \text{ vs } k_{ij} : \gamma_{ij} > \gamma_0.$$

Individual tests:

$$\varphi_{ij}(x) = \begin{cases} 1, & t_{ij}(x) > c_{ij} \\ 0, & t_{ij}(x) \leq c_{ij} \end{cases}$$

$$\Phi(x) = \begin{pmatrix} 1, & \varphi_{12}(x), & \dots, & \varphi_{1N}(x) \\ \varphi_{21}(x), & 1, & \dots, & \varphi_{2N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N1}(x), & \varphi_{N2}(x), & \dots, & 1 \end{pmatrix}. \quad (5)$$

Define multiple statistical procedure for network structure identification

$$\delta(x) = d_G, \text{ iff } \Phi^{opt}(x) = G \quad (6)$$

General theorem for network structure with arbitrary number of elements

Theorem 3: Let loss function be additive and tests $\varphi_{ij}(x)$ are UMP in the class of unbiased. Then statistical procedure $\Phi(x)(5)$ is optimal in the class of unbiased statistical procedures.

Sketch of proof:

- $\varphi_{ij}(x)$ - unbiased $\Rightarrow R(s_{i,j}, \varphi_{ij}(x)) \leq R(s'_{i,j}, \varphi_{ij}(x))$.
  Loss function is additive $\Rightarrow R(H_S, \delta) = \sum_{i,j=1}^{N} R(s_{i,j}, \varphi_{ij})$. $\forall S, S'$

$$\sum_Q w(S, Q) P_\theta(\delta(x) = d_Q | H_S) \leq \sum_Q w(S', Q) P_\theta(\delta(x) = d_Q | H_S)$$

  Then $\delta(x)$ is unbiased.

General theorem for network structure with arbitrary number of elements

- Let $\delta'(x)$ is other unbiased procedure. Then $\delta'(x)$ defines the partition of sample space by L parts $D_G = \{x : \delta'(x) = G\}$. Let $A_{i,j} = \bigcup_{G:g_{i,j}=0} D_G, A_{i,j}^{-1} = \bigcup_{G:g_{i,j}=1} D_G$. Define
$$\varphi'_{i,j} = \begin{cases} 0, & x \in A_{i,j} \\ 1, & x \in A_{i,j}^{-1} \end{cases}$$

$$\sum_Q w(S,Q) P_\theta(\delta'(x) = d_Q | H_S) \leq \sum_Q w(S',Q) P_\theta(\delta'(x) = d_Q | H_S)$$

Let $H_S, H_{S'}$ such that $s_{i,j} \neq s'_{i,j}, s_{j,i} \neq s'_{j,i}$.
$\varphi_{i,j}(x)$-UMP in the class of unbiased, then
$R(s_{i,j}, \varphi_{i,j}(x)) \leq R(s_{i,j}, \varphi'_{i,j}(x))$.
Then $R(H_S, \delta) \leq R(H_S, \delta')$

Multiple testing procedures for GGM identification

Let vector $(X_1, X_2, \ldots, X_N)$ has multivariate normal distribution $N(\mu, \Sigma)$.
$H_{\beta_1,\beta_2,\ldots,\beta_M}$- hypothesis, that GGM has edges $\beta_1, \beta_2, \ldots, \beta_M$,$\beta_k = (i_k, j_k)$.
Individual edge hypotheses:

$$h_{\beta_k} : \rho^{i_k j_k} = 0 \text{ vs } k_{\beta_k} : \rho^{i_k j_k} \neq 0.$$

Individual tests:

$$\varphi_{ij}^{opt} = \begin{cases} 0, & |r^{i,j}| < 1 - 2c_{\alpha/2}^{\beta} \\ 1, & |r^{i,j}| > 1 - 2c_{\alpha/2}^{\beta} \end{cases} \tag{7}$$

where $c_{\alpha/2}^{\beta}$ is the $\alpha/2$-quantile of Beta distribution $Be(\frac{n-N}{2}, \frac{n-N}{2})$

Multiple testing procedures for GGM identification

$$\Phi^{opt}(x) = \begin{pmatrix} 0, & \varphi_{12}^{opt}(x), & \dots, & \varphi_{1N}^{opt}(x) \\ \varphi_{21}^{opt}(x), & 0, & \dots, & \varphi_{2N}^{opt}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N1}^{opt}(x), & \varphi_{N2}^{opt}(x), & \dots, & 0 \end{pmatrix}. \tag{8}$$

Define multiple statistical procedure for concentration graph identification

$$\delta^{opt}(x) = d_G, \text{ iff } \Phi^{opt}(x) = G \tag{9}$$

Theorem 4: Multiple decision statistical procedure $\Phi^{opt}(x)$ is optimal in the class of unbiased statistical procedures under additive loss function.

Multiple testing procedures for GGM identification

Lemma 1: Optimal in the class of w-unbiased statistical level α test for individual edge hypotheses is:$H_{ij} : \rho^{i,j} = 0$ against $K_{ij} : \rho^{i,j} \neq 0$ is:

$$\varphi_{ij}^{opt} = \begin{cases} 0, & \frac{|as_{ij} - \frac{b}{2}|}{\sqrt{\frac{b^2}{4} + ac}} < 1 - 2c_{\alpha/2}^{\beta} \\ 1, & \frac{|as_{ij} - \frac{b}{2}|}{\sqrt{\frac{b^2}{4} + ac}} > 1 - 2c_{\alpha/2}^{\beta} \end{cases} \tag{10}$$

where $\det(s_{kl}) = -as_{ij}^2 + bs_{ij} + c$, $c_{\alpha/2}^{\beta}$ is the $\alpha/2$-quantile of Beta distribution $Be(\frac{n-N}{2}, \frac{n-N}{2})$. $(a = a(\{s_{kl}\}), b = b(\{s_{kl}\}), c = c(\{s_{kl}\}))$.
Lemma 2: Test (7) is equivalent to UMPU test (10) for testing $\rho^{i,j} = 0$ vs $\rho^{i,j} \neq 0$.

GGM selection. Statistical approach

Dempster(1972), Edwards(2000), Drton(2003,2005,2007)
Let $X(t), t = 1, 2, \ldots, n$ be a observations.

- Calculate $r^{i,j} = \frac{s^{i,j}}{\sqrt{s^{i,i} s^{j,j}}}$-sample partial correlation.
- Apply multiple hypotheses testing procedures (Holm, Hochberg and so on) for set of hypotheses $h_{i,j} : \rho^{i,j} = 0$ vs alternative $k_{i,j} : \rho^{i,j} \neq 0$.
- Drawbacks - control FWER only, asymptotic results.

Optimal statistical procedure for reduced graph identification

Let vector $(X_1, X_2, \ldots, X_N)$ has multivariate normal distribution $N(\mu, \Sigma)$.
$H_{i_1,i_2,\ldots,i_M}$- hypothesis, that reduced graph has vertex $i_1, i_2, \ldots, i_M$.
Individual vertex hypotheses:

$$h_i : \mu_i \leq \mu_0 \text{ vs } k_i : \mu_i > \mu_0$$

Individual tests:

$$\varphi_i(x) = \begin{cases} 1, & U_i(x) > c_i \\ 0, & U_i(x) \leq c_i \end{cases}$$

$$U_i(x) = \sqrt{n}\frac{(\overline{x}_i - \mu_0)}{\sqrt{\sigma_{ii}}}, \tag{11}$$

Statistical procedure for reduced graph identification:

$$\delta(x) = (\varphi_1(x), \varphi_2(x), \ldots, \varphi_N(x)). \tag{12}$$

Optimal statistical procedure for reduced graph identification

Theorem 5: Let random vector $(X_1, \ldots, X_N)$ has a multivariate normal distribution $N(\mu, \Sigma)$ with unknown μ and known diag(Σ) . If $U_i(x)$ is defined by (11) then statistical procedure (12) for problem of reduced graph identification is optimal in the class of W-unbiased multiple decision statistical procedures (where W is additive loss function) and c_i is $(1 - \alpha_i)$-quantile of standard normal distribution.

Optimal statistical procedure for reduced graph identification

Lemma 3: Let random vector $(X_1, \ldots, X_N)$ has a multivariate normal distribution $N(\mu, \Sigma)$, where $\mu = (\mu_1, \ldots, \mu_N)$ is unknown vector, $\Sigma = ||\sigma_{ij}||$ is known matrix. For testing hypothesis $h_1 : \mu_1 \leq \mu_0$ against $k_1 : \mu_1 > \mu_0$ an optimal unbiased test has the form:

$$\varphi_i(x) = \begin{cases} 0, & \overline{x}_i \leq c_i(T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N) \\ 1, & \overline{x}_i > c_i(T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N) \end{cases} \tag{13}$$

where c_i for a given α_i is defined from $P(\overline{x}_i > c_i | T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N) = \alpha_i$.

$$\begin{aligned} T_1 &= \sigma^{11}\overline{x_1} + \sigma^{12}\overline{x_2} + \ldots + \sigma^{1N}\overline{x_N} \\ &\ldots \\ T_k &= \sigma^{k1}\overline{x_1} + \sigma^{k2}\overline{x_2} + \ldots + \sigma^{kN}\overline{x_N} \\ &\ldots \\ T_N &= \sigma^{N1}\overline{x_1} + \sigma^{N2}\overline{x_2} + \ldots + \sigma^{NN}\overline{x_N} \end{aligned} \tag{14}$$

Optimal statistical procedure for reduced graph identification

Lemma 4: Let random vector $(X_1, \ldots, X_N)$ has a multivariate normal distribution $N(\mu, \Sigma)$, where $\mu = (\mu_1, \ldots, \mu_N)$ is unknown vector, $\Sigma = ||\sigma_{ij}||$ is known matrix. The random variables $\overline{x}_i$ and $T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N$ are independent.
Lemma 3 implies that the optimal test has a Neyman structure and lemma 4 implies that this test can be written as:

$$\varphi_i(x) = \begin{cases} 0, & U_i(x) = \frac{\sqrt{n}(\overline{x_i} - \mu_0)}{\sqrt{\sigma_{ii}}} \leq c_i \\ 1, & U_i(x) > c_i \end{cases} \tag{15}$$

Therefore, (12) is optimal in the class of W-unbiased multiple decision statistical procedures. Note that optimal multiple decision statistical procedure (12) depends on diagonal elements of covariance matrix Σ only. □

Market network. Known results

Let $X(t), t = 1, 2, \ldots, n$ be a observations (daily returns) of stocks $X(t) = (X_1(t), \ldots, X_N(t))$

Traditional approach to MG identification:

- calculate $r_{i,j}$
- if $r_{i,j} > \theta$ (where θ is a given threshold), add edge (i,j) in MG

Traditional approach to MST identification:

- calculate $r_{i,j}$
- order $r_{i,j}$
- apply Kruscal algorithm to MST construction

Drawback:

- statistical problem statement is absent.
- statistical properties are unknown.

Multiple testing statistical procedures for MG identification

Individual edge hypotheses:

$$h_{ij} : \gamma_{ij} \leq \gamma_0 \text{ vs } k_{ij} : \gamma_{ij} > \gamma_0.$$

Individual tests:

$$\varphi_{ij}(X) = \begin{cases} 1, & t_{ij}(X) > c_{ij} \\ 0, & t_{ij}(X) \leq c_{ij} \end{cases}$$

Multiple testing statistical procedure:

$$\Phi(x) = \begin{pmatrix} 0, & \varphi_{12}(x), & \dots, & \varphi_{1N}(x) \\ \varphi_{21}(x), & 0, & \dots, & \varphi_{2N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N1}(x), & \varphi_{N2}(x), & \dots, & 0 \end{pmatrix}. \tag{16}$$

Example 1.Pearson network

Individual hypotheses (Pearson measure): $h_{ij} : \rho_{i,j} \le \rho_0$ vs $k_{ij} : \rho_{i,j} > \rho_0$

- $\partial^P_{i,j}(x_i, x_j) = \begin{cases} 1, & \frac{r_{i,j}-\rho_0}{\sqrt{1-r^2_{i,j}}} > c^{St}_{i,j} \\ 0, & \frac{r_{i,j}-\rho_0}{\sqrt{1-r^2_{i,j}}} \le c^{St}_{i,j} \end{cases}$ UMP in the class of invariant tests
- $c^{St}_{i,j}$ is $(1-\alpha_{ij})$-quantile of Student distribution t_{n-1},
- Multiple testing procedures

$$\partial^P(x) = \begin{pmatrix} 0, & \partial^P_{1,2}(x), & \dots, & \partial^P_{1,N}(x) \\ \partial^P_{2,1}(x), & 0, & \dots, & \partial^P_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \partial^P_{N,1}(x), & \partial^P_{N,2}(x), & \dots, & 0 \end{pmatrix}.$$

Example 2. Sign similarity network

- Individual hypotheses: $h_{ij} : p^{i,j} \leq p_0$ vs $k_{ij} : p^{i,j} > p_0$
- $$\varphi_{i,j}^{Sg} = \begin{cases} 0, & T_{i,j}^{Sg} \leq c_{i,j} \\ 1, & T_{i,j}^{Sg} > c_{i,j} \end{cases},$$

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

  $$I_{i,j}(t) = \begin{cases} 1, & sign(x_i(t)) = sign(x_j(t)) \\ 0, & \text{else} \end{cases}$$

  $c_{i,j}$ is defined from equation: $\sum_{k=c_{i,j}}^{n} \frac{n!}{k!(n-k)!}(p_0)^k(1-p_0)^{n-k} \leq \alpha$

Example 2. Sign similarity network

Multiple decision single step procedure

$$\delta^{Sg}(x) = \begin{pmatrix} 0, & \varphi^{Sg}_{1,2}(x), & \dots, & \varphi^{Sg}_{1,N}(x) \\ \varphi^{Sg}_{2,1}(x), & 0, & \dots, & \varphi^{Sg}_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi^{Sg}_{N,1}(x), & \varphi^{Sg}_{N,2}(x), & \dots, & 0 \end{pmatrix}.$$

Optimality of ∂^P

Theorem 6: Let loss function w be additive, individual test statistics $t_{i,j}$ depends only on observations $X_i(t), X_j(t)$ and vector $X = (X_1, \ldots, X_N)$ has a multivariate normal distribution. Then for single step statistical procedure ∂^P for threshold graph identification ($\rho_0 = 0$) in Pearson correlation network one has $Risk(S, \partial^P) \leq Risk(S, \partial)$ for any adjacency matrix S and any $w-$unbiased δ.

Theorem 7: Let loss function w be additive, individual test statistics $t_{i,j}$ depends only on observations $X_i(t), X_j(t)$ and vector $X = (X_1, \ldots, X_N)$ has a multivariate normal distribution. Then for single step statistical procedure ∂^P for threshold graph identification in Pearson correlation network one has $Risk(S, \partial^P) \leq Risk(S, \delta)$ for any adjacency matrix S and any invariant δ.

Assumption of normality can not be removed.

Optimality of δ^{Sg}

Theorem 8: Let loss function w be additive, individual test statistics $t_{i,j}$ depends only on $I_{i,j}(t)$, $E(X_i)$ are known $\forall i = 1, \ldots, N$ and distribution of vector $X = (X_1, \ldots, X_N)$ satisfy the symmetry condition below.
Then for single step statistical procedure δ^{Sg} for threshold graph identification in sign similarity network one has $Risk(S, \delta^{Sg}) \leq Risk(S, \delta)$ for any adjacency matrix S and any $w-$unbiased δ.
Symmetry condition:

$$p_{1,1}^{i,j} = p_{-1,-1}^{i,j}; p_{1,-1}^{i,j} = p_{-1,1}^{i,j}$$

$$p_{1,1}^{i,j} = P(X_i - E(X_i) > 0, X_j - E(X_j) > 0)$$

$$p_{-1,-1}^{i,j} = P(X_i - E(X_i) \leq 0, X_j - E(X_j) \leq 0)$$

$$p_{-1,1}^{i,j} = P(X_i - E(X_i) \leq 0, X_j - E(X_j) > 0)$$

$$p_{1,-1}^{i,j} = P(X_i - E(X_i) > 0, X_j - E(X_j) \leq 0)$$

Our publications on market network analysis.

- Koldanov A.P., Koldanov P.A., Kalyagin V.A., Pardalos P.M. Statistical Procedures for the Market Graph Construction, Computational Statistics & Data Analysis, v.68, pp.17-29 (2013).
- Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. Simple measure of similarity for the market graph construction, Computational Management Science, Volume 10, Issue 2 (2013), Page 105-124.
- Bautin G.A., Kalyagin V.A., Koldanov A.P. Comparative analysis of two similarity measures for the market graph construction, Models, Algorithms and Technologies for Networks Analysis, Springer Proceedings on Mathematics and Statistics, v.59 (2013), pp.29-41.
- Koldanov A. P., Koldanov P.A: Optimal Multiple Decision Statistical Procedure for Inverse Covariance Matrix. Springer Optimization and Its Applications. Vol. 87 (2014), pp 205-216.

Our publications on market network analysis.

- Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M., Zamaraev V.A.: Measures of uncertainty in market network analysis, Physica A. Vol 413 (2014), pp.59-70.
- Vizgunov A.N., Goldengorin B.I., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P. M., Network approach for the Russian stock market, Computational Management Science, Springer-Verlag, Volume 11, Issue 1 (2014), Page 45-55.
- Kalyagin V.A., Koldanov P.A., Zamaraev V.A. Network structures uncertainty for different markets, Springer Optimization and Its Applications. 2014. Vol. 100. P. 181-197.
- Bautin G.A., Koldanov A.P., Pardalos P.M. Robustness of sign correlation in market network analysis, Springer Optimization and Its Applications. 2014. Vol. 100. P. 25-33.

Our publications on market network analysis.

- Kalygin V.A. Koldanov A.P. and Pardalos P.M. A General Approach to Network Analysis of Statistical Data Sets, in: Learning and Intelligent Optimization. / . .: P. M. Pardalos, M. Resende, C. Vogiatzis, J. Walteros. Vol. 8426: Lectute Notes in Computer Science. Switzerland : Springer, 2014. P. 88-97.
- Koldanov P.A. Bautin G.A. Multiple decision problem for stock selection in market network, Vol. 8426: Lectute Notes in Computer Science. Switzerland : Springer, 2014. P. 98-110.
- Kalyagin V. Koldanov A. Koldanov P. Zamaraev V. Market Graph and Markovitz Model, Optimization in Science and Engineering, 2014, pp 293-306
- Latyshev A., Koldanov P. Testing of connections between Pearson and sign correlations in market network, in: Models, Algorithms, and Technologies for Network Analysis / From the 4th International Conference on Network Analysis / . .: V. A. Kalyagin, P. Koldanov, P. M. Pardalos. NY : Springer International Publishing, 2016. P. 175-182.

Our publications on market network analysis.

- Koldanov P., Komissarova A. Statistical uncertainty of minimum spanning tree in market network, in: Models, Algorithms, and Technologies for Network Analysis / From the 4th International Conference on Network Analysis / . .: V. A. Kalyagin, P. Koldanov, P. M. Pardalos. NY : Springer International Publishing, 2016. P. 157-164.
- Koldanov P., Kalyagin V. A., Bautin G. A. On some statistical procedures for stock selection problem // Annals of Mathematics and Artificial Intelligence. 2016. Vol. 76. No. 1. P. 47-57.
- Koldanov P., Lozgacheva N. MULTIPLE TESTING OF SIGN SYMMETRY FOR STOCK RETURN DISTRIBUTIONS // International Journal of Theoretical and Applied Finance. 2016. Vol. 19. No. 8. P. 1650049-1-1650049-14.

Our publications on market network analysis.

- Kalyagin V. A., Koldanov A. P., Petr A. Koldanov. Robust identification in random variables networks // Journal of Statistical Planning and Inference. 2017. Vol. 181, P. 30-40.
- Koldanov P., Koldanov A. P., Kalyagin V. A., Pardalos P. M. Uniformly most powerful unbiased test for conditional independence in Gaussian graphical model // Statistics & Probability Letters, 2017, Vol. 122, P. 90-95.

THANK YOU FOR YOUR ATTENTION!

Proof of lemma 2

Lemma 2. Probabilities
$p(i_1, i_2, \ldots, i_N) := P_\Lambda(i_1 X_1 > 0, i_2 X_2, \ldots, i_N X_N > 0)$ are defined by the matrix Λ and does not depend on the function g.

Proof

$$P(i_1 X_1 > 0, i_2 X_2, \ldots, i_N X_N > 0) = \int_{i_k x_k > 0, k=1,2,\ldots,N} |\Lambda|^{-\frac{1}{2}} g(x'\Lambda x) dx_1 \ldots dx_N \tag{17}$$

Matrix Λ is positive definite, therefore there exists a matrix C such that $C'\Lambda C = I$. Put $y = C^{-1}x$. Then $x = Cy$ and

$$\int_{i_k x_k > 0, k=1,2,\ldots,N} |\Lambda|^{-\frac{1}{2}} g(x'\Lambda x) dx_1 \ldots dx_N = \int_D g(y'y) dy_1 \ldots dy_N \tag{18}$$

where D is given by

$$0 < i_k(c_{k,1} y_1 + c_{k,2} y_2 + \ldots + c_{k,N} Y_N) < \infty, \quad k = 1, 2, \ldots, N \tag{19}$$

Proof of lemma 2

Vector y can be written in polar coordinates as:

$$\begin{aligned} y_1 &= r\sin(\theta_1) \\ y_2 &= r\cos(\theta_1)\sin(\theta_2) \\ y_3 &= r\cos(\theta_1)\cos(\theta_2)\sin(\theta_3) \\ &\dots \\ y_{N-1} &= r\cos(\theta_1)\cos(\theta_2)\dots\cos(\theta_{N-2})\sin(\theta_{N-1}) \\ y_N &= r\cos(\theta_1)\cos(\theta_2)\dots\cos(\theta_{N-2})\cos(\theta_{N-1}) \end{aligned} \tag{20}$$

where $-\frac{\pi}{2} \le \theta_i \le \frac{\pi}{2}, i = 1\dots, N-2; -\pi \le \theta_{N-1} \le \pi, 0 \le r \le \infty$ The Jacobian of the transformation (20) is

$$r^{N-1}\cos^{N-2}(\theta_1)\cos^{N-3}(\theta_2)\dots\cos(\theta_{N-2})$$

In polar coordinates region (19) is transformed to the region $D' \times R_+^1$ where D' given by $(k = 1, 2, \dots, N)$:

$$0 < i_k(c_{11}\sin(\theta_1) + \dots + c_{1N}\cos(\theta_1)\cos(\theta_2)\dots\cos(\theta_{N-2})\cos(\theta_{N-1})) < \infty \tag{21}$$

Proof of lemma 2

Then $p(i_1, i_2, \ldots, i_N)$ can be written as

$$\int_{D'} \int_0^{\infty} r^{N-1} \cos^{N-2}(\theta_1) \cos^{N-3}(\theta_2) \ldots \cos(\theta_{N-2}) g(r^2) dr d\theta_1 \ldots d\theta_{N-1} =$$

$$= \int_{D'} \cos^{N-2}(\theta_1) \cos^{N-3}(\theta_2) \ldots \cos(\theta_{N-2}) d\theta_1 \ldots d\theta_{N-1} \int_0^{\infty} r^{N-1} g(r^2) dr$$

It is known that

$$\int_0^{\infty} r^{N-1} g(r^2) dr = \frac{1}{C(N)}$$

where

$$C(N) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \ldots \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\pi}^{\pi} \cos^{N-2}(\theta_1) \cos^{N-3}(\theta_2) \ldots \cos(\theta_{N-2}) d\theta_1 \ldots d\theta_{N-1}$$

Region D' is defined by the matrix Λ and does not depend on the function g. Then $p(i_1, i_2, \ldots, i_N)$ are defined by the matrix Λ and does not depend on the function g.

Proof of lemma 3

Lemma 3. Joint distribution of the statistics $T_{i,j}^{Sg}$ are defined by the matrix Λ and does not depend on the function g.

Proof.

- $T_{i,j}^{Sg} = \frac{n}{2} + \frac{1}{2}\sum_{t=1}^{n} \text{sign}(X_i(t))\text{sign}(X_j(t))$
- from the lemma 2 - joint distribution of the $\text{sign}(X) = (\text{sign}(X_1), \text{sign}(X_2), \ldots, \text{sign}(X_N))$ is defined by the matrix Λ and does not depend on the function g.
- random vectors $\text{sign}(X(t))$, $t = 1, 2, \ldots, n$ are independent and identically distributed.
- then the joint distribution $\text{sign}(X_i(t))$, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, n$ is defined by the matrix Λ and does not depend on the function g.
- then joint distribution of statistics $T_{i,j}^{Sg}$, $i, j = 1, 2, \ldots, N; i < j$ does not depend on the function g.

Data for stability

1. We consider the real-world data from USA stock market. We take $N = 83$ largest by capitalization companies and consider the daily returns of these companies for the period from 03.01.2011 up to 31.12.2013, total 751 observations.
2. We calculate correlation matrix Σ by this data and consider the matrix Σ as true matrix. Structures of the matrix are considered as true structures.
3. We simulate a certain number of observation (n) using the mixture distribution. The mixture distribution is constructed as follow - random vector $X = (X_1, \ldots, X_N)$ takes value from $N(0, \Sigma)$ with probability γ and from $t_3(0, \Sigma)$ with probability $1 - \gamma$.
4. We estimate the matrix Σ using the chosen association measure (Pearson $\rho_{i,j}$ or probability $p^{i,j}$).
5. We construct the sample threshold graph basing on the estimation and compare it to the true threshold graph.

Appendix: Proof of the Theorem 2

Theorem 2: Let vector $X = (X_1, \ldots, X_N)$ has elliptical distribution (1). Then:

$$\gamma_{i,j}^{Sg} = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{\lambda_{i,j}}{\sqrt{\lambda_{i,i}\lambda_{j,j}}} = \frac{1}{2} + \frac{1}{\pi} \arcsin \gamma_{i,j}^{P} \tag{22}$$

Prove: It is known $E(X) = \mu$. Without loss of generality let $\mu = 0$. Define matrix $A = (a_{i,j}) = \Lambda^{-1}$. Density of random vector (X_i, X_j) has the form:

$$f(x_i, x_j) = |A^{-1}|^{-\frac{1}{2}} g(a_{i,i}x_i^2 + 2a_{i,j}x_i(x_j + a_{j,j}x_j^2)$$

The prove is based on the following lemma:

Proof of the Theorem 2

Lemma 1: Probability $\gamma_{i,j}^{Sg} = P(X_i X_j > 0)$ defined by the matrix Λ and does not depend from g.

Prove: Matrix $A_{i,j} = \begin{pmatrix} a_{i,i} & a_{i,j} \\ a_{i,j} & a_{j,j} \end{pmatrix}$ is positive definite, then exists $C = \begin{pmatrix} c_{i,i} & c_{i,j} \\ c_{j,i} & c_{j,j} \end{pmatrix}$ such that $C'AC = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Define $U = c_{i,i}X_i + c_{i,j}X_j$, $V = c_{j,i}X_i + c_{j,j}X_j$. Then random vector (U, V) has distribution with density $f(u, v) = g(u^2 + v^2)$. Then

$$P(X_i > 0, X_j > 0) = P(\frac{c_{ii}}{c_{jj}}V < U < \frac{c_{ij}}{c_{jj}}V) + P(\frac{c_{ij}}{c_{jj}}V < U < \frac{c_{ii}}{c_{jj}}V) =$$

$$= \begin{cases} \frac{arctg(c_{ij}/c_{jj}) - arctg(c_{ii}/c_{jj})}{2\pi}, & \frac{c_{ij}}{c_{jj}} > \frac{c_{ii}}{c_{jj}} \\ \frac{arctg(c_{ii}/c_{jj}) - arctg(c_{ij}/c_{jj})}{2\pi}, & \frac{c_{ij}}{c_{jj}} < \frac{c_{ii}}{c_{jj}} \end{cases}$$

Then $P(X_i > 0, X_j > 0)$ does not depend from g. Similarly $P(X_i < 0, X_j < 0)$ does not depend from g. Then $P(X_i X_j > 0) = P(X_i > 0, X_j > 0) + P(X_i < 0, X_j < 0)$ does not depend from g.

Holm procedure

- Step 1: If $\max_{i,j=1,\dots,N} T_{i,j} \le c_1^H$ then accept all hypotheses $h_{i,j}, i,j = 1,2,\dots,N$, else if $\max_{i,j=1,\dots,N} T_{i,j} = T_{i_1,j_1}$ then reject hypothesis h_{i_1,j_1} and go to step 2.
- ...
- Step K: Let $I = \{(i_1,j_1),(i_2,j_2),\dots,(i_{K-1},j_{K-1})\}$ be the set of indexes of previously rejected hypotheses. If $\max_{(i,j)\notin I} T_{i,j} \le c_K^H$ then accept all hypotheses $h_{i,j}$, $(i,j) \notin I$, else if $\max_{(i,j)\notin I} T_{i,j} = T_{i_K,j_K}$ then reject hypothesis h_{i_K,j_K} and go to step (K+1).
- ...
- Step M: Let $I = \{(i_1,j_1),\dots,(i_{M-1},j_{M-1})\}$ be the set of indexes of previously rejected hypotheses. Let $(i_M,j_M) \notin I$. If $T_{i_M,j_M} \le c_M^H$ then accept the hypothesis h_{i_M,j_M}, else reject hypothesis h_{i_M,j_M} (reject all hypotheses).

For a given significance level α the critical values c_K^H for Holm procedure are given by $F_{\gamma_0}(c_K^H) = 1 - \dfrac{\alpha}{M-K+1}, \quad K = 1,2,\dots,M$

Hochberg procedure

- Step 1: If $T_{i_1,j_1} = \min_{i,j=1,\dots,N} T_{i,j} > c_1^{Hg}$ then reject all individual hypotheses $h_{i,j}$, else accept hypothesis h_{i_1,j_1} and go to step 2.
- ...
- Step K: Let $I = \{(i_1,j_1),\dots,(i_{K-1},j_{K-1})\}$ be the set of indexes of previously accepted hypotheses. If $T_{i_K,j_K} = \min_{i,j=1,\dots,N;(i,j)\notin I} T_i(x) > c_K^{Hg}$ then reject all hypotheses $h_{i,j}$, $(i,j) \notin I$, else accept hypothesis h_{i_K,j_K} and go to step (K+1).
- ...
- Step M: Let $I = \{(i_1,j_1),\dots,(i_{M-1},j_{M-1})\}$ be the set of indexes of previously accepted hypotheses. Let $(i_M,j_M) \notin I$. If $T_{i_M,j_M} > c_M^{Hg}$ then reject the hypothesis h_{i_M,j_M} else accept the hypothesis h_{i_M,j_M} (accept all hypothesis).

For a given significance level α the critical values c_K^{Hg} for Hochberg procedure are given by $F_{\gamma_0}(c_K^{Hg}) = 1 - \frac{\alpha}{K}, \quad K = 1,2,\dots,M$

Role of measure of association

Therefore the following statistical procedure for threshold graph identification in Pearson correlation network will be distribution free:

- fix a threshold ρ_0.
- Take δ a threshold graph identification statistical procedure in sign similarity network distribution free in the class of elliptically contoured distributions.
- Apply statistical procedures δ for threshold graph identification with the threshold

$$p_0 = \frac{1}{2} + \frac{1}{\pi} \arcsin \rho_0$$

.

- Consider obtained graph as the threshold graph in Pearson correlation network.

In particular one can construct single step, Holm and Hochberg distribution free statistical procedures for threshold graph identification in Pearson correlation network.

Symmetry conditions. Tests for individual hypotheses

Individual hypotheses:
$h_1^{i,j} : p_{1,1}^{i,j} = p_{-1,-1}^{i,j}$ vs $k_1^{i,j} : p_{1,1}^{i,j} \neq p_{-1,-1}^{i,j};\ i,j = 1,\dots,N;\ i \neq j$
Statistics $T_{1,1}^{i,j} = \sum_{t=0}^{n} T_{1,1}^{i,j}(t), \quad T_{-1,-1}^{i,j} = \sum_{t=0}^{n} T_{-1,-1}^{i,j}(t),$

$$T_{1,1}^{i,j}(t) = \begin{cases} 1, & X_i(t) \geq 0, X_i(t) \geq 0 \\ 0, & else \end{cases}$$

$$T_{-1,-1}^{i,j}(t) = \begin{cases} 1, & X_i(t) < 0, X_i(t) < 0 \\ 0, & else \end{cases}$$

Symmetry conditions. Tests for individual hypotheses

Individual hypotheses:
$h_2^{i,j} : p_{1,-1}^{i,j} = p_{-1,1}^{i,j}$ vs $k_2^{i,j} : p_{1,-1}^{i,j} \neq p_{-1,1}^{i,j};\ i,j = 1,\dots,N;\ i \neq j$
Statistics $T_{1,-1}^{i,j} = \sum_{t=0}^{n} T_{1,-1}^{i,j}(t), \quad T_{-1,1}^{i,j} = \sum_{t=0}^{n} T_{-1,1}^{i,j}(t)$

$$T_{1,-1}^{i,j}(t) = \begin{cases} 1, & X_i(t) \geq 0, X_i(t) < 0 \\ 0, & else \end{cases}$$

$$T_{-1,1}^{ij}(t) = \begin{cases} 1, & X_i(t) < 0, X_i(t) \geq 0 \\ 0, & else \end{cases}$$

Optimal tests for individual hypotheses testing $h^1_{i,j}$

Exponential form for the joint distribution of statistics $T_{k,l}$:

$$P(T_{1,1} = k_1, T_{-1,-1} = k_2, T_{1,-1} = k_3, T_{-1,1} = k_4) =$$

$$= C \exp\{k_1 \ln \frac{p_{1,1}}{p_{-1,-1}} + (k_1 + k_2) \ln \frac{p_{-1,-1}}{p_{-1,1}} + k_3 \ln \frac{p_{1,-1}}{p_{-1,1}}\}$$

where

$$C = \frac{n!}{k_1!k_2!k_3!k_4!}(1 - p_{1,1} - p_{-1,-1} - p_{1,-1})^n$$

Then uniformly most powerful test for testing hypothesis $h^1_{i,j}$ has Neymann structure and can be written as:

$$\varphi^1_{i,j} = \begin{cases} 0, & C_1(k, k_3) < k_1 < C_2(k, k_3) \\ 1, & else \end{cases} \tag{23}$$

where k_1, k_2, k_3, k_4 are the observed values of statistics $T^{i,j}_{1,1}$, $T^{i,j}_{-1,-1}$, $T^{i,j}_{1,-1}$, $T^{i,j}_{-1,1}$, $k = k_1 + k_2$.

Optimal tests for individual hypotheses testing $h^1_{i,j}$

The constants C_1, C_2 are defined from conditional distribution of statistic $T_{1,1}$ under conditions $T_{1,1} + T_{-1,-1} = k$, $T_{1,-1} = k_3$ and assumption that the hypothesis $h^1_{i,j}$ is true. One has

$$P(T_{1,1} = k_1 | T_{1,1} + T_{-1,-1} = k, T_{1,-1} = k_3) =$$

$$= \frac{P(T_{1,1} = k_1, T_{-1,-1} = k - k_1, T_{1,-1} = k_3)}{P(T_{1,1} + T_{-1,-1} = k, T_{1,-1} = k_3)}$$

$$P(T_{1,1}+T_{-1,-1} = k, T_{1,-1} = k_3) = \sum_{i=0}^{k} P(T_{1,1} = i, T_{-1,-1} = k-i, T_{1,-1} = k$$

$$= \frac{n!}{k_3!(n-k_3-k)!k!} p_{1,-1}^{k_3} (p_{1,1}+p_{-1,-1})^k (1-p_{1,1}-p_{-1,-1}-p_{1,-1})^{n-k_3-k}$$

$$P(T_{1,1} = k_1, T_{-1,-1} = k - k_1, T_{1,-1} = k_3) =$$

$$= \frac{n!}{k_1!k_3!(k-k_1)!(n-k_3-k)!} p_{1,1}^{k_1} p_{1,-1}^{k_3} p_{-1,-1}^{k-k_1} (1-p_{1,1}-p_{-1,-1}-p_{1,-1})^{n-k_3-k}$$

Optimal tests for individual hypotheses testing $h^1_{i,j}$

$$P(T_{1,1} = k_1 | T_{1,1} + T_{-1,-1} = k, T_{1,-1} = k_3) =$$

$$= C_k^{k_1} \left(\frac{p_{1,1}}{p_{1,1} + p_{-1,-1}} \right)^{k_1} \left(\frac{p_{-1,-1}}{p_{1,1} + p_{-1,-1}} \right)^{k-k_1}$$

Under $h^1_{i,j}$ one has $p_{1,1} = p_{-1,-1}$. Optimal test is

$$\varphi^1_{i,j} = \begin{cases} 0, & C_1(k) < k_1 < C_2(k) \\ 1, & else \end{cases} \tag{24}$$

where $C_1(k)$ and $C_2(k)$ are defined by

$$C_1(k) = \max\{C : (\frac{1}{2})^k \sum_{i=0}^{C} C_k^i \leq \frac{\alpha}{2}\}$$

$$C_2(k) = \min\{C : (\frac{1}{2})^k \sum_{i=C}^{k} C_k^i \leq \frac{\alpha}{2}\}$$

p-values

The p-value of the test can be calculated by

$$p^1_{i,j} = 2\min\{(\frac{1}{2})^k \sum_{i=k_1}^{k} C^i_k,\ (\frac{1}{2})^k \sum_{i=0}^{k_1} C^i_k\} \tag{25}$$

On the same way one can construct the uniformly most powerful test for the hypothesis $h^2_{i,j}$. The test can be written as

$$\varphi^2_{i,j} = \begin{cases} 0, & C_1(m) < k_3 < C_2(m) \\ 1, & else \end{cases} \tag{26}$$

where $m = k_3 + k_4$ The p-value of the test (26) can be calculated by

$$p^2_{i,j} = 2\min\{(\frac{1}{2})^m \sum_{i=k_3}^{m} C^i_m,\ (\frac{1}{2})^m \sum_{i=0}^{k_3} C^i_m\} \tag{27}$$

Note that by construction all individual tests are distribution free uniformly most powerful tests of Neymann structure.

Rejection graph

We select 100 stocks from US market with a highest trading volume during the period of 8 years, from 01.01.2006 to 31.12.2013. We compare results for different periods of observations: 8 periods of 1 year each, 4 periods of 2 years each, 2 periods of 4 years each and 1 period of 8 years. Significance level of multiple tests are set to $\alpha = 0,1$ and $\alpha = 0,5$. To describe the results of multiple testing we introduce a *rejection graph*. Edge (i, j) is included in the rejection graph for hypotheses h^1 iff the hypothesis $h^1_{i,j}$ is rejected by multiple testing procedure. Nodes of the rejection graph are vertices adjacent to these edges.

Rejection graph

The Figure illustrates the structure of the rejection graph for the year 2006, $\alpha = 0.5$, US market.