Optimal Multiple Decision Statistical Procedure for Inverse Covariance Matrix

Alexander P. Koldanov and Petr A. Koldanov

Abstract A multiple decision statistical problem for the elements of inverse covariance matrix is considered. Associated optimal unbiased multiple decision statistical procedure is given. This procedure is constructed using the Lehmann theory of multiple decision statistical procedures and the conditional tests of the Neyman structure. The equations for thresholds calculation for the tests of the Neyman structure are analyzed.

Keywords Inverse covariance matrix • Tests of the Neyman structure • Multiple decision statistical procedure • Generating hypothesis

1 Introduction

A market network is constructed by means of some similarity measure between every pairs of stocks . The most popular measure of similarity between stocks of a market is the correlation between them [1–4]. The analysis of methods of market graph construction [2] from the statistical point of view was started in [5]. In [5] multiple decision statistical procedure for market graph construction based on the Pearson test is suggested. The authors of [5] note that a procedure of this type can be made optimal in the class of unbiased multiple decision statistical procedures if one uses the tests of the Neyman structure for generating hypothesis. In the present paper we use the partial correlations as a measure of similarity between stocks. In this case the elements of inverse covariance matrix are the weights of links between stocks in the market network [6].

A.P. Koldanov • P.A. Koldanov (✉)
National Research University, Higher School of Economics,
Bolshaya Pecherskaya 25/12, Nizhny Novgorod, 603155, Russia
e-mail: akoldanov@hse.ru; pkoldanov@hse.ru

The main goal of the paper is investigation of the problem of identification of inverse covariance matrix as a multiple decision statistical problem. As a result an optimal unbiased multiple decision statistical procedure for identification of inverse covariance matrix is given. This procedure is constructed using the Lehmann theory of multiple decision statistical procedures and the conditional tests of the Neyman structure. In addition the equations for thresholds calculation for the tests of the Neyman structure are analyzed.

The paper is organized as follows. In Sect. 2 we briefly recall the Lehmann theory of multiple decision statistical procedures. In Sect. 3 we describe the tests of the Neyman structure. In Sect. 4 we formulate the multiple decision problem for identification of inverse covariance matrix. In Sect. 5 we construct and study the optimal tests for testing of generating hypothesis for the elements of an inverse covariance matrix. In Sect. 6 we construct the multiple statistical procedure and consider some particular cases. In Sect. 7 we summarize the main results of the paper.

2   Lehmann Multiple Decision Theory

In this section we recall for the sake of completeness the basic idea of the Lehmann theory following the paper [5].

Suppose that the distribution of a random vector R is taken from the family $f(r, \theta) : \theta \in \Omega$, where θ is a parameter, Ω is the parametric space, r is an observation of R. We need to construct a statistical procedure for the selection of one from the set of L hypotheses, which in the general case can be stated as:

$$H_i : \theta \in \Omega_i, i = 1, \ldots, L,$$
$$\Omega_i \cap \Omega_j = \emptyset, \ i \neq j, \ \bigcup_{i=1}^{L} \Omega_i = \Omega. \tag{1}$$

The most general theory of a multiple decision procedure is the Lehmann theory [7]. The Lehmann theory is based on three concepts: generating hypothesis, generating hypothesis testing and compatibility conditions, and additivity condition for the loss function.

The multiple decision problem (1) of selecting one from the set of L hypotheses $H_i : \theta \in \Omega_i; i = 1, \ldots, L$ is equivalent to a family of M two decision problems:

$$H'_j : \theta \in \omega_j \text{ vs } K'_j : \theta \in \omega_j^{-1}, j = 1, \ldots, M, \tag{2}$$

with

$$\bigcup_{i=1}^{L} \Omega_i = \omega_j \cup \omega_j^{-1} = \Omega.$$

This equivalence is given by the relations:

$$\Omega_i = \bigcap_{j=1}^{M} \omega_j^{\chi_{i,j}}, \quad \omega_j = \bigcup_{\{i:\chi_{i,j}=1\}} \Omega_i,$$

where

$$\chi_{i,j} = \begin{cases} 1, & \Omega_i \cap \omega_j \neq \emptyset, \\ -1, & \Omega_i \cap \omega_j = \emptyset. \end{cases} \tag{3}$$

Hypotheses $H'_j : (j = 1, \ldots, M)$ are called generating hypotheses for the problem (1).

The equivalence between problem (1) and the family of problems (2) reduces the multiple decision problem to the testing of generating hypothesis. Any statistical procedure δ_j for hypothesis testing of H'_j can be written in the following form

$$\delta_j(r) = \begin{cases} \partial_j, & r \in X_j, \\ \partial_j^{-1}, & r \in X_j^{-1}, \end{cases} \tag{4}$$

where ∂_j is the decision of acceptance of H'_j and ∂_j^{-1} is the decision of acceptance of K'_j, X_j is the acceptance region of H'_j and X_j^{-1} is the acceptance region of K'_j (rejection region of H'_j) in the sample space. One has $X_j \cap X_j^{-1} = \emptyset$, $X_j \cup X_j^{-1} = X$, X being the sample space.

Define the acceptance region for H_i by

$$D_i = \bigcap_{j=1}^{M} X_j^{\chi_{i,j}}, \tag{5}$$

where $\chi_{i,j}$ are defined by (3), and put $X_j^1 = X_j$. Note that $\bigcup_{i=1}^{L} D_i \subset X$, but it is possible that $\bigcup_{i=1}^{L} D_i \neq X$.

Therefore, if $D_1, D_2, \ldots, D_L$ is a partition of the sample space X, then one can define the statistical procedure $\delta(r)$ by

$$\delta(r) = \begin{cases} d_1, & r \in D_1, \\ d_2, & r \in D_2, \\ \ldots \ldots \ldots \\ d_L, & r \in D_L. \end{cases} \tag{6}$$

According to [8, 9] we define the conditional risk for multiple decision statistical procedure by

$$risk(\theta, \delta) = E_\theta w(\theta, \delta(R)) = \sum_{k=1}^{L} w(\theta, d_k) P_\theta(\delta(R) = d_k), \tag{7}$$

where E_θ is the expectation according to the density function $f(x, \theta)$ and $w(\theta, d_k)$ is the loss from decision d_k under the condition that θ is true, $\theta \in \Omega$. Under the additivity condition of the loss function (see [5] for more details) the conditional risk can be written as

$$risk(\theta, \delta) = \sum_{j=1}^{M} risk(\theta, \delta_j), \quad \theta \in \Omega. \tag{8}$$

We call statistical procedure optimal in a class of statistical procedures if it has minimal conditional risk for all $\theta \in \Omega$ in this class. The main result of the Lehmann theory (see [7]) states that if Eq. (8) is satisfied and statistical procedures (4) are optimal in the class of unbiased statistical tests, then the associated multiple decision statistical procedure (6) is optimal in the class of unbiased multiple decision procedures.

3   Unbiasedness and Tests of the Neyman Structure

The class of unbiased multiple decision statistical procedures according to Lehmann [9, 10] is defined by:

$$E_\theta w(\theta, \delta(R)) \le E_\theta w(\theta', \delta(R)) \, for \, any \, \theta, \theta' \in \Omega.$$

Let $f(r; \theta)$ be the density of the exponential family:

$$f(r; \theta) = c(\theta) exp(\sum_{j=1}^{M} \theta_j T_j(r))h(r), \tag{9}$$

where $c(\theta)$ is a function defined in the parameters space, $h(r)$, $T_j(r)$ are functions defined in the sample space, and $T_j(R)$ are the sufficient statistics for θ_j, $j = 1, \ldots, M$.

Suppose that generating hypotheses (2) has the form:

$$H'_j : \theta_j = \theta_j^0 \text{ vs } K'_j : \theta_j \ne \theta_j^0, \ j = 1, 2, \ldots, M, \tag{10}$$

where θ_j^0 are fixed. For a fixed j the parameter θ_j is called information or structural parameter and $\theta_k, k \ne j$ are called nuisance parameters. According to [9] the optimal unbiased tests for generating hypotheses (10) are:

$$\delta_j = \begin{cases} \partial_j, & c_j^1(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_M) < t_j < c_j^2(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_M), \\ \partial_j^{-1}, & \text{otherwise}, \end{cases} \tag{11}$$

where $t_i = T_i(r), i = 1,\ldots,M$ and constants $c_j^1(t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_M)$, $c_j^2(t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_M)$ are defined from the equations

$$\int_{c_j^1}^{c_j^2} f(t_j,\theta_j^0 | T_i = t_i, i = 1,\ldots,M; i \neq j) dt_j = 1 - \alpha_j, \tag{12}$$

and

$$\int_{-\infty}^{c_j^1} t_j f(t_j,\theta_j^0 | T_i = t_i, i = 1,\ldots,M; i \neq j) dt_j$$

$$+ \int_{c_j^2}^{+\infty} t_j f(t_j,\theta_j^0 | T_i = t_i, i = 1,\ldots,M; i \neq j) dt_j \tag{13}$$

$$= \alpha_j \int_{-\infty}^{+\infty} t_j f(t_j,\theta_j^0 | T_i = t_i, i = 1,\ldots,M; i \neq j) dt_j,$$

where $f(t_j,\theta_j^0 | T_i = t_i, i = 1,\ldots,M; i \neq j)$ is the density of conditional distribution of statistics T_j and α_j is the level of significance of the test.

A test satisfying (12) is said to have Neyman structure. This test is characterized by the fact that the conditional probability of rejection of H_j' (under the assumption that H_j' is true) is equal to α_j on each of the surfaces $\bigcap_{k \neq j}(T_k(x) = t_k)$. Therefore the multiple decision statistical procedure associated with the tests of the Neyman structure (11), (12), and (13) is optimal in the class of unbiased multiple decision procedures.

4   Problem of Identification of Inverse Covariance Matrix

In this section we formulate the multiple decision problem for elements of inverse covariance matrix.

Let N be the number of stocks on a financial market, and let n be the number of observations. Denote by $r_i(t)$ the daily return of the stock i for the day t ($i = 1,\ldots,N; t = 1,\ldots,n$). We suppose $r_i(t)$ to be an observation of the random variable $R_i(t)$. We use the standard assumptions: the random variables $R_i(t)$, $t = 1,\ldots,n$ are independent and have all the same distribution as a random variable $R_i (i = 1,\ldots,N)$. The random vector $(R_1, R_2,\ldots,R_N)$ describes the joint behavior of the stocks. We assume that the vector $(R_1, R_2,\ldots,R_N)$ has a multivariate normal distribution with covariance matrix $\|\sigma_{ij}\|$ where $\sigma_{ij} = cov(R_i, R_j) = E(R_i - E(R_i))(R_j - E(R_j)), \rho_{ij} = (\sigma_{ij})/(\sqrt{\sigma_{ii}\sigma_{jj}}), i, j = 1,\ldots,N, E(R_i)$ is the expectation of the random variable R_i. We define a sample space as $R^{N \times n}$ with the elements $(r_i(t))$. Statistical estimation of σ_{ij} is $s_{ij} = \Sigma_{t=1}^n (r_i(t) - \overline{r}_i)(r_j(t) - \overline{r}_j)$ where $\overline{r}_i = (1/n)\Sigma_{t=1}^n r_i(t)$. The sample correlation between the stocks i and j is defined by $r_{ij} = (s_{ij})/(\sqrt{s_{ii}s_{jj}})$.

It is known [11] that for a multivariate normal vector the statistics $(\overline{r_1}, \overline{r_2}, \ldots, \overline{r_N})$ and $\|s_{ij}\|$ (matrix of sample covariances) are sufficient.

Let σ^{ij} be the elements of inverse covariance matrix $\|\sigma^{ij}\|$. Then the problem of identification of inverse covariance matrix can be formulated as a multiple decision problem of the selection of one from the set of hypotheses:

$$H_1 : \sigma^{ij} = \sigma_0^{ij}, \ i, j = 1, \ldots, N, \ i < j,$$

$$H_2 : \sigma^{12} \neq \sigma_0^{12}, \sigma^{ij} = \sigma_0^{ij}, i, j = 1, \ldots, N, (i, j) \neq (1, 2), \ i < j,$$

$$H_3 : \sigma^{13} \neq \sigma_0^{13}, \sigma^{ij} = \sigma_0^{ij}, i, j = 1, \ldots, N, (i, j) \neq (1, 3), \ i < j,$$

$$H_4 : \sigma^{12} \neq \sigma_0^{12}, \sigma^{13} \neq \sigma_0^{13}, \sigma^{ij} = \sigma_0^{ij}, i, j = 1, \ldots, N, (i, j) \neq (1, 2), \ (i, j) \neq (1, 3),$$

$$\ldots$$

$$H_L : \sigma^{ij} \neq \sigma_0^{ij}, i, j = 1, \ldots, N, \ i < j,$$

$$(14)$$

where $L = 2^M$ with $M = N(N-1)/2$.

Multiple decision problem (14) is a particular case of the problem (1). The parameter space Ω is the space of positive semi-definite matrices $\|\sigma^{ij}\|$, Ω_k is a domain in the parameters space associated with the hypothesis H_k from the set (14) $k = 1, \ldots, L$. For the multiple decision problem (14) we introduce the following set of generating hypotheses:

$$h_{i,j} : \sigma^{ij} = \sigma_0^{ij} \ \text{vs} \ k_{i,j} : \sigma^{ij} \neq \sigma_0^{ij}, \ i, \ j = 1, 2, \ldots, N, \ i < j, \qquad (15)$$

We use the following notations: $\partial_{i,j}$ is the decision of acceptance of the hypothesis $h_{i,j}$ and $\partial_{i,j}^{-1}$ is the decision of rejection of $h_{i,j}$.

5 Tests of the Neyman Structure for Testing of Generating Hypothesis

Now we construct the optimal test in the class of unbiased tests for generating hypothesis (15). To construct these tests we use the sufficient statistics s_{ij} with the following Wishart density function [11]:

$$f(\{s_{k,l}\}) = \frac{[\det(\sigma^{kl})]^{n/2} \times [\det(s_{kl})]^{(n-N-2)/2} \times \exp[-(1/2) \sum_k \sum_l s_{k,l} \sigma^{kl}]}{2^{(Nn/2)} \times \pi^{N(N-1)/4} \times \Gamma(n/2) \Gamma((n-1)/2) \cdots \Gamma((n-N+1)/2)}$$

if the matrix (s_{kl}) is positive definite, and $f(\{s_{kl}\}) = 0$ otherwise. One has for a fixed $i < j$:

$$f(\{s_{kl}\}) = C(\{\sigma^{kl}\}) \times \exp[-\sigma^{ij} s_{ij} - \frac{1}{2} \sum_{(k,l) \neq (i,j); (k,l) \neq (j,i)} s_{kl} \sigma^{kl}] \times h(\{s_{kl}\}),$$

where

$$C(\{\sigma^{kl}\}) = \frac{1}{q}\det(\sigma^{kl})]^{n/2},$$

$$q = 2^{(Nn/2)} \times \pi^{N(N-1)/4} \times \Gamma(n/2)\Gamma((n-1)/2)\cdots\Gamma((n-N+1)/2),$$

$$h(\{s_{kl}\}) = [\det(s_{kl})]^{(n-N-2)/2}.$$

Therefore, this distribution belongs to the class of exponential distributions with parameters σ^{kl}.

The optimal tests of the Neyman structure (11), (12), and (13) for generating hypothesis (15) take the form:

$$\delta_{i,j}(\{s_{kl}\}) = \begin{cases} \partial_{i,j}, & c^1_{i,j}(\{s_{kl}\}) < s_{ij} < c^2_{i,j}(\{s_{kl}\}), \ (k,l) \neq (i,j), \\ \partial^{-1}_{i,j}, & s_{ij} \leq c^1_{i,j}(\{s_{kl}\}) \text{ or } s_{ij} \geq c^2_{i,j}(\{s_{kl}\}), \ (k,l) \neq (i,j), \end{cases} \quad (16)$$

where the critical values are defined from the equations

$$\frac{\int_{I\cap[c^1_{i,j};c^2_{i,j}]} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}}{\int_I \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}} = 1 - \alpha_{i,j}, \quad (17)$$

$$\int_{I\cap[-\infty;c^1_{i,j}]} s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}$$
$$+ \int_{I\cap[c^2_{i,j};+\infty]} s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij} \quad (18)$$
$$= \alpha_{i,j} \int_I s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij},$$

where I is the interval of values of s_{ij} such that the matrix (s_{kl}) is positive definite and α_{ij} is the level of significance of the tests.

Consider Eqs. (17) and (18). Note that $\det(s_{kl})$ is a quadratic polynomial of s_{ij}. Let $\det(s_{k,l}) = -C_1 s^2_{i,j} + C_2 s_{i,j} + C_3 = C_1(-s^2_{i,j} + As_{i,j} + B), C_1 > 0$ then the positive definiteness of the matrix $(s_{k,l})$ for a fixed $s_{k,l}, (k,l) \neq (i,j)$ gives the following interval for the value of $s_{i,j}$:

$$I = \{x : \frac{A}{2} - \sqrt{\frac{A^2}{4} + B} < x < \frac{A}{2} + \sqrt{\frac{A^2}{4} + B}\}. \quad (19)$$

Now we define the functions

$$\Psi_1(x) = \int_0^x (-t^2 + At + B)^K \exp(-\sigma_0^{ij} t) dt, \quad (20)$$

$$\Psi_2(x) = \int_0^x t(-t^2 + At + B)^K \exp(-\sigma_0^{ij} t) dt, \quad (21)$$

where $K = (n - N - 2)/2$. One can calculate the critical values of $c_{i,j}^1, c_{i,j}^2$ from the equations:

$$\Psi_1(c_{ij}^2) - \Psi_1(c_{ij}^1) = (1 - \alpha_{ij})(\Psi_1(\frac{A}{2} + \sqrt{\frac{A^2}{4} + B}) - \Psi_1(\frac{A}{2} - \sqrt{\frac{A^2}{4} + B})), \quad (22)$$

$$\Psi_2(c_{ij}^2) - \Psi_2(c_{ij}^1) = (1 - \alpha_{ij})(\Psi_2(\frac{A}{2} + \sqrt{\frac{A^2}{4} + B}) - \Psi_2(\frac{A}{2} - \sqrt{\frac{A^2}{4} + B})). \quad (23)$$

The test (16) can be written in terms of sample correlations. First note that $\det(s_{kl}) = \det(r_{kl})s_{11}s_{22}\ldots s_{NN}$ where r_{kl} are the sample correlations. One has

$$\frac{\int_{J \cap [e_{i,j}^1; e_{i,j}^2]} \exp[-\sigma_0^{ij} r_{ij}\sqrt{s_{ii}s_{jj}}][\det(r_{kl})]^{(n-N-2)/2} dr_{ij}}{\int_J \exp[-\sigma_0^{ij} r_{ij}\sqrt{s_{ii}s_{jj}}][\det(r_{kl})]^{(n-N-2)/2} dr_{ij}} = 1 - \alpha_{i,j}, \quad (24)$$

$$\int_{J \cap [-\infty; e_{i,j}^1]} r_{ij}\exp[-\sigma_0^{ij} r_{ij}\sqrt{s_{ii}s_{jj}}][\det(r_{kl})]^{(n-N-2)/2} dr_{ij}$$
$$+ \int_{J \cap [e_{i,j}^2; +\infty]} r_{ij}\exp[-\sigma_0^{ij} r_{ij}\sqrt{s_{ii}s_{jj}}][\det(r_{kl})]^{(n-N-2)/2} dr_{ij} \quad (25)$$
$$= \alpha_{i,j} \int_J r_{ij}\exp[-\sigma_0^{ij} r_{ij}\sqrt{s_{ii}s_{jj}}][\det(r_{kl})]^{(n-N-2)/2} dr_{ij},$$

where $I = J\sqrt{s_{ii}s_{jj}}$, $c_{i,j}^k = e_{i,j}^k\sqrt{s_{ii}s_{jj}}$; $k = 1, 2$. Therefore, the tests (16) take the form

$$\delta_{i,j}(\{r_{kl}\}) = \begin{cases} \partial_{i,j}, & e_{i,j}^1(s_{ii}, s_{jj}, \{r_{kl}\}) < r_{ij} < e_{i,j}^2(s_{ii}, s_{jj}, \{r_{kl}\}), \\ & (k,l) \neq (i,j), \\ \partial_{i,j}^{-1}, & r_{ij} \leq e_{i,j}^1(s_{ii}, s_{jj}\{r_{kl}\}) \text{ or } r_{ij} \geq e_{i,j}^2(s_{ii}, s_{jj}\{r_{kl}\}), \\ & (k,l) \neq (i,j). \end{cases} \quad (26)$$

It means that the tests of the Neyman structure for generating hypothesis (15) do not depend on $s_{k,k}$, $k \neq i$, $k \neq j$. In particular for $N = 3$, $(i,j) = (1,2)$ one has

$$\delta_{1,2}(\{r_{k,l}\}) = \begin{cases} \partial_{1,2}, & e_{12}^1(s_{11}, s_{22}, r_{13}, r_{23}) < r_{12} < e_{12}^2(s_{11}, s_{22}, r_{13}, r_{23}), \\ \partial_{1,2}^{-1}, & r_{12} \leq e_{12}^1(s_{11}, s_{22}, r_{13}, r_{23}) \text{ or } r_{12} \geq e_{12}^2(s_{11}, s_{22}, r_{13}, r_{23}). \end{cases} \quad (27)$$

To emphasize the peculiarity of the constructed test we consider some interesting particular cases.

$n - N - 2 = 0$, $\sigma_0^{ij} \neq 0$. In this case expressions (20) and (21) can be simplified. Indeed one has in this case

$$\Psi_1(x) = \int_0^x \exp(-\sigma_0^{ij}t)dt = \frac{1}{\sigma_0^{ij}}(1 - \exp(-\sigma_0^{ij}x)),$$

$$\Psi_2(x) = \int_0^x t\exp(-\sigma_0^{ij}t)dt = -\frac{x}{\sigma_0^{ij}}\exp(-\sigma_0^{ij}x) - \frac{1}{(\sigma_0^{ij})^2}\exp(-\sigma_0^{ij}x) + \frac{1}{(\sigma_0^{ij})^2}.$$

Finally one has the system of two equations for defining constants c_{ij}^1, c_{ij}^2:

$$\exp(-\sigma_0^{ij}c_{ij}^1) - \exp(-\sigma_0^{ij}c_{ij}^2)$$

$$= (1-\alpha_{ij})\{\exp(-\sigma_0^{ij}(\frac{A}{2} - \sqrt{\frac{A^2}{4}+B})) - \exp(-\sigma_0^{ij}(\frac{A}{2} + \sqrt{\frac{A^2}{4}+B}))\},\tag{28}$$

$$c_{ij}^1\exp(-\sigma_0^{ij}c_{ij}^1) - c_{ij}^2\exp(-\sigma_0^{ij}c_{ij}^2)$$

$$= (1-\alpha_{ij})\{(\frac{A}{2} - \sqrt{\frac{A^2}{4}+B})\exp(-\sigma_0^{ij}(\frac{A}{2} - \sqrt{\frac{A^2}{4}+B}))\tag{29}$$

$$-(\frac{A}{2} + \sqrt{\frac{A^2}{4}+B})\exp(-\sigma_0^{ij}(\frac{A}{2} + \sqrt{\frac{A^2}{4}+B}))\}.$$

$\underline{\sigma_0 = 0}$. In this case the critical values are defined by the system of algebraic equations (22) and (23) where the functions Ψ_1, Ψ_2 are defined by

$$\Psi_1(x) = \int_0^x (-t^2 + At + B)^K dt,$$

$$\Psi_2(x) = \int_0^x t(-t^2 + At + B)^K dt,$$

In this case the tests of the Neyman structure have the form

$$\delta_{i,j}(\{r_{k,l}\}) = \begin{cases} \partial_{i,j}, & e_{i,j}^1(\{r_{kl}\}) < r_{ij} < e_{i,j}^2(\{r_{kl}\}), \ (k,l) \neq (i,j), \\ \partial_{i,j}^{-1}, & r_{ij} \leq e_{i,j}^1(\{r_{kl}\}) \text{ or } r_{ij} \geq e_{i,j}^1(\{r_{kl}\}), \ (k,l) \neq (i,j). \end{cases}\tag{30}$$

$\underline{n - N - 2 = 0, \ \sigma_0 = 0}$. In this case one has

$$c_{i,j}^1 = \tfrac{A}{2} - (1-\alpha_{ij})\sqrt{\tfrac{A^2}{4}+B},$$

$$c_{i,j}^2 = \tfrac{A}{2} + (1-\alpha_{ij})\sqrt{\tfrac{A^2}{4}+B}.\tag{31}$$

6 Multiple Statistical Procedure Based on the Tests of the Neyman Structure

Now it is possible to construct the multiple decision statistical procedure for problem (14) based on the tests of Neyman structure. Then the multiple decision statistical procedure (6) takes the form:

$$
\delta = \begin{cases}
d_1, \; c_{ij}^1(\{s_{kl}\}) < r_{ij} < c_{ij}^2(\{s_{kl}\}), \\
\quad i < j, i, j, k, l = 1, \dots, N, (k,l) \neq (i,j), \\
d_2, \; r_{12} \leq c_{12}^1(\{s_{kl}\}) \text{ or } r_{12} \geq c_{12}^2(\{s_{kl}\}), c_{ij}^1(\{s_{kl}\}) < r_{ij} < c_{ij}^2(\{s_{kl}\}), \\
\quad i < j, i, j, k, l = 1, \dots, N, (k,l) \neq (i,j), \\
\dots \; \dots\dots\dots \\
d_L, \; r_{ij} \leq c_{ij}^1(\{s_{kl}\}) \text{ or } r_{ij} \geq c_{ij}^2(\{s_{kl}\}), \\
\quad i < j, i, j, k, l = 1, \dots, N, (k,l) \neq (i,j),
\end{cases} \tag{32}
$$

where $c_{ij}(\{s_{kl}\})$ are defined from the Eqs. (17) and (18).

One has $D_k = \{r \in R^{N \times n} : \delta(r) = d_k\}$, $k = 1, 2, \dots, L$. It is clear that

$$
\bigcup_{k=1}^{L} D_k = R^{N \times n}.
$$

Then $D_1, D_2, \dots, D_L$ is a partition of the sample space $R^{N \times n}$. The tests of the Neyman structure for generating hypothesis (15) are optimal in the class of unbiased tests. Therefore if the condition of the additivity (8) of the loss function is satisfied, then the associated multiple decision statistical procedure is optimal. For discussion of additivity of the loss function see [5].

We illustrate statistical procedure (32) with an example.

Let $N = 3$. In this case problem (14) is the problem of the selection of one from eight hypotheses:

$$
\begin{aligned}
H_1 &: \sigma^{12} = \sigma_0^{12}, \; \sigma^{13} = \sigma_0^{13}, \; \sigma^{23} = \sigma_0^{23}, \\
H_2 &: \sigma^{12} \neq \sigma_0^{12}, \; \sigma^{13} = \sigma_0^{13}, \; \sigma^{23} = \sigma_0^{23}, \\
H_3 &: \sigma^{12} = \sigma_0^{12}, \; \sigma^{13} \neq \sigma_0^{13}, \; \sigma^{23} = \sigma_0^{23}, \\
H_4 &: \sigma^{12} = \sigma_0^{12}, \; \sigma^{13} = \sigma_0^{13}, \; \sigma^{23} \neq \sigma_0^{23}, \\
H_5 &: \sigma^{12} \neq \sigma_0^{12}, \; \sigma^{13} \neq \sigma_0^{13}, \; \sigma^{23} = \sigma_0^{23}, \\
H_6 &: \sigma^{12} = \sigma_0^{12}, \; \sigma^{13} \neq \sigma_0^{13}, \; \sigma^{23} \neq \sigma_0^{23}, \\
H_7 &: \sigma^{12} \neq \sigma_0^{12}, \; \sigma^{13} = \sigma_0^{13}, \; \sigma^{23} \neq \sigma_0^{23}, \\
H_8 &: \sigma^{12} \neq \sigma_0^{12}, \; \sigma^{13} \neq \sigma_0^{13}, \; \sigma^{23} \neq \sigma_0^{23}.
\end{aligned} \tag{33}
$$

Generating hypotheses are:

$h_{1,2} : \sigma^{12} = \sigma_0^{12}$ vs $k_{1,2} : \sigma^{12} \neq \sigma_0^{12}$, σ^{13}, σ^{23} are the nuisance parameters.

$h_{1,3} : \sigma^{13} = \sigma_0^{13}$ vs $k_{1,3} : \sigma^{13} \neq \sigma_0^{13}$, σ^{12}, σ^{23} are the nuisance parameters.

$h_{2,3} : \sigma^{23} = \sigma_0^{23}$ vs $k_{2,3} : \sigma^{23} \neq \sigma_0^{23}$, σ^{12}, σ^{13} are the nuisance parameters.

In this case multiple statistical procedure for problem (33) (if $\sigma_0 \neq 0$) is:

$$
\delta = \begin{cases}
d_1, & c_{12}^1 < r_{12} < c_{12}^2, & c_{13}^1 < r_{13} < c_{13}^2, & c_{23}^1 < r_{23} < c_{23}^2, \\
d_2, & r_{12} \leq c_{12}^1 \text{ or } r_{12} \geq c_{12}^2, & c_{13}^1 < r_{13} < c_{13}^2, & c_{23}^1 < r_{23} < c_{23}^2, \\
d_3, & c_{12}^1 < r_{12} < c_{12}^2, & r_{13} \leq c_{13}^1 \text{ or } r_{13} \geq c_{13}^2, & c_{23}^1 < r_{23} < c_{23}^2, \\
d_4, & c_{12}^1 < r_{12} < c_{12}^2, & c_{13}^1 < r_{13} < c_{13}^2, & r_{23} \leq c_{23}^1 \text{ or } r_{23} \geq c_{23}^2, \\
d_5, & r_{12} \leq c_{12}^1 \text{ or } r_{12} \geq c_{12}^2, & r_{13} \leq c_{13}^1 \text{ or } r_{13} \geq c_{13}^2, & c_{23}^1 < r_{23} < c_{23}^2, \\
d_6, & c_{12}^1 < r_{12} < c_{12}^2, & r_{13} \leq c_{13}^1 \text{ or } r_{13} \geq c_{13}^2, & r_{23} \leq c_{23}^1 \text{ or } r_{23} \geq c_{23}^2, \\
d_7, & r_{12} \leq c_{12}^1 \text{ or } r_{12} \geq c_{12}^2, & c_{13}^1 < r_{13} < c_{13}^2, & r_{23} \leq c_{23}^1 \text{ or } r_{23} \geq c_{23}^2, \\
d_8, & r_{12} \leq c_{12}^1 \text{ or } r_{12} \geq c_{12}^2, & r_{13} \leq c_{13}^1 \text{ or } r_{13} \geq c_{13}^2, & r_{23} \leq c_{23}^1 \text{ or } r_{23} \geq c_{23}^2.
\end{cases}
\tag{34}
$$

The critical values $c_{12}^k = c_{12}^k(r_{13}, r_{23}, s_{11}s_{22})$, $c_{13}^k = c_{13}^k(r_{12}, r_{23}, s_{11}s_{33})$, $c_{23}^k = c_{23}^k(r_{12}, r_{13}, s_{22}s_{33})$; $k = 1, 2$ are defined from Eqs. (24) and (25). If $n = 5$; $\sigma_0^{ij} \neq 0$; $i, j = 1, 2, 3$, then the critical values c_{ij}^k; $k = 1, 2$ are defined from (28) and (29).

If $\sigma_0^{ij} = 0, \forall i, j$ and $n = 5$, then tests (30) for generating hypothesis depend on the sample correlation only. Therefore the corresponding multiple statistical procedure with L decisions depends only on the sample correlation too. This procedure is (34) where constants $c_{12}^k = c_{12}^k(r_{13}, r_{23}), c_{13}^k = c_{13}^k(r_{12}, r_{23}), c_{23}^k = c_{23}^k(r_{12}, r_{13}); k = 1, 2$.

In this case

$$
I_{1,2} = (r_{13}r_{23} - G_{1,2}; r_{13}r_{23} + G_{1,2}),
$$

$$
I_{1,3} = (r_{12}r_{23} - G_{1,3}; r_{12}r_{23} + G_{1,3}),
$$

$$
I_{2,3} = (r_{12}r_{13} - G_{2,3}; r_{12}r_{13} + G_{2,3}),
$$

where

$$
G_{1,2} = \sqrt{(1 - r_{13}^2)(1 - r_{23}^2)},
$$

$$
G_{1,3} = \sqrt{(1 - r_{12}^2)(1 - r_{23}^2)},
$$

$$
G_{2,3} = \sqrt{(1 - r_{12}^2)(1 - r_{13}^2)},
$$

and the critical values are

$$
c_{12}^1 = r_{13}r_{23} - (1 - \alpha_{12})G_{1,2},
$$

$$
c_{12}^2 = r_{13}r_{23} + (1 - \alpha_{12})G_{1,2},
$$

$$
c_{13}^1 = r_{12}r_{23} - (1 - \alpha_{13})G_{1,3},
$$

$$
c_{13}^2 = r_{12}r_{23} + (1 - \alpha_{13})G_{1,3},
$$

$$c_{23}^1 = r_{12}r_{13} - (1 - \alpha_{23})G_{2,3},$$

$$c_{23}^2 = r_{12}r_{13} + (1 - \alpha_{23})G_{2,3}.$$

Note that in this case test (34) has a very simple form.

7 Concluding Remarks

Statistical problem of identification of elements of inverse covariance matrix is investigated as multiple decision problem. Solution of this problem is developed on the base of the Lehmann theory of multiple decision procedures and theory of tests of the Neyman structure. It is shown that this solution is optimal in the class of unbiased multiple decision statistical procedures. Obtained results can be applied to market network analysis with partial correlations as a measure of similarity between stocks returns.

References

[1] Mantegna, R.N.: Hierarchical structure in financial market. Eur. Phys. J. series B **11**, 193–197 (1999)

[2] Boginsky V., Butenko S., Pardalos P.M.: On structural properties of the market graph. In: Nagurney, A. (ed.) Innovations in Financial and Economic Networks. pp. 29–45. Edward Elgar Publishing Inc., Northampton (2003)

[3] Boginsky V., Butenko S., Pardalos P.M.: Statistical analysis of financial networks. Comput. Stat. Data. Anal. **48**, 431–443 (2005)

[4] M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, H.A.: Tool for Filtering Information in Complex Systems. Proc. Natl. Acad. Sci. Uni. States Am. **102** 30, 10421–10426 (2005)

[5] Koldanov, A.P., Koldanov, P.A., Kalyagin, V.A., Pardalos, P.M.: Statistical Procedures for the Market Graph Construction. Comput. Stat. Data Anal. **68**, 17–29, DOI: 10.1016/j.csda.2013.06.005. (2013)

[6] Hero, A., Rajaratnam, B.: Hub discovery in partial correlation graphical models. IEEE Trans. Inform. Theor. **58**–l(9), 6064–6078 (2012)

[7] Lehmann E.L.: A theory of some multiple decision procedures 1. Ann. Math. Stat. **28**, 1–25 (1957)

[8] Wald, A.: Statistical Decision Function. Wiley, New York (1950)

[9] Lehmann E.L., Romano J.P.: Testing Statistical Hypothesis. Springer, New York (2005)

[10] Lehmann E.L.: A general concept of unbiasedness. Ann. Math. Stat. **22**, 587–597 (1951)

[11] Anderson T.W.: An Introduction to Multivariate Statistical Analysis. 3 edn. Wiley-Interscience, New York (2003)