Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Statistical procedures for the market graph construction

Q1 Alexander P. Koldanov [a], Petr A. Koldanov [a], Valeriy A. Kalyagin [a,*], Panos M. Pardalos [a,b]

[a] National Research University Higher School of Economics, Bolshaya Pecherskaya 25/12, Nizhny Novgorod, 603155, Russia
[b] Center for Applied Optimization, University of Florida, Gainesville, FL, USA

HIGHLIGHTS

- The market graph construction is investigated as a multiple decision statistical procedure.
- The multiple decision problem is reduced to the family of two-decision problems.
- Multiple decision statistical testing is reduced to generating hypothesis testing.
- The market graph construction procedure is optimal as a multiple decision statistical procedure.
- Compatibility conditions and additivity of the loss function are discussed theoretically and numerically.

ARTICLE INFO

Article history:
Received 15 November 2012
Received in revised form 4 June 2013
Accepted 4 June 2013
Available online xxxx

Keywords:
Market graph
Unbiased statistical tests
Multiple decision statistical procedures
Generating hypotheses
Additivity of the loss function
Conditional risk

ABSTRACT

The statistical analysis of the method of construction of the market graph when considered as a multiple decision statistical procedure is investigated. It is shown that under the condition of additivity of the loss function the method can be optimal in different classes of unbiased multiple statistical procedures. The results are obtained by application of the Lehmann theory of multiple decision procedures to the method of construction of the market graph. The main findings are illustrated by numerical studies of the conditional risk of multiple decision statistical procedures for different loss functions and different return distributions.

1. Introduction

Stock market models have attracted great attention in mathematical modeling since the seminal work by Markowitz (1952). An important characteristic of these models is the analysis of simultaneous behavior of the stocks based on their correlations. At the same time, the stock market can be considered as a complex network where the stocks are the nodes of the network and the correlation is a measure of similarity between them (Mantegna, 1999; Mantegna and Stanley, 2000; Tumminello et al., 2010). In the paper Boginsky et al. (2003) a new model for the analysis of the stock market was introduced. The method involves the construction of the market graph associated with the stock market and the investigation of the structural characteristics (cliques, independent sets and others) of this graph. This method was recently developed by Boginsky et al. (2005), Boginsky et al. (2006), and Huang et al. (2009); see also Bautin et al. (2013) and Vinciotti and Hashem (2013).

* Corresponding author. Tel.: +7 8314169649; fax: +7 8314169651.
E-mail address: vkalyagin@hse.ru (V.A. Kalyagin).

0167-9473/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.csda.2013.06.005

The main goal of the present paper is the investigation of the market graph construction from a statistical point of view. We consider a market graph construction as a multiple decision procedure. In our investigation we consider the class of unbiased statistical procedures in the sense of Lehmann (1951) and Lehmann and Romano (2005a). As a model of the financial market we use the Markowitz type model (Markowitz, 1952, 2010). We call a true market graph the graph with the incidence matrix with entries 0 and 1, where we have 0 if the associated correlation is less than a given threshold and 1 otherwise. In Boginsky et al. (2003) the authors use a sample market graph constructed in the same way from the sample correlations. We show that the procedure of this type can be made optimal in the class of unbiased multiple decision statistical procedures. This result is obtained by application of the Lehmann theory of multiple decision problems to the method of construction of the market graph. In the framework of the Lehmann theory (Lehmann, 1957) we discuss the following conditions of optimality for the multiple decision statistical procedures: choice of the generating hypotheses, compatibility of the family of tests for the generating hypotheses with decision space for the original hypotheses, and additivity of the loss function. We prove that the multiple decision statistical procedure for the construction of the market graph is optimal if one uses the tests of the Neyman structures. From the other hand, in practice it is more convenient to use classical unbiased correlation tests. We prove that the associated multiple decision statistical procedure is also optimal, but in a restricted class of unbiased multiple decision statistical procedures.

The paper is organized as follow. In Section 2 we introduce the market graph and give a formal statement of the problem. In Section 3, we recall some basic concepts from the Lehmann theory (Lehmann, 1957). In Section 4, we study (from a statistical point of view) the method of construction of the market graph. In Section 5, we present results of numerical simulation and discuss the conditional risk of described statistical procedures. In the last Section 6, we summarize the main results of the paper.

2. Market model and statement of problems

Let N be the number of stocks in the financial market, and let n be the number of observations. Denote by $p_i(t)$ the price of the stock i for the day $t (i = 1, \dots, N;\ t = 1, \dots, n)$ and define the daily return of the stock i for the period from $(t-1)$ to t as $r_i(t) = \ln(p_i(t)/p_i(t-1))$. We assume $r_i(t)$ to be a realization of the random variable $R_i(t)$. We define the sample space $R^{N\times n}$ with elements $(r_i(t))$. We consider the standard assumptions: the random variables $R_i(t), t = 1, \dots, n$ are independent with fixed i, have all the same distribution as a random variable $R_i (i = 1, \dots, N)$, and the random vector $(R_1, R_2, \dots, R_N)$ has a multivariate normal distribution with correlation matrix $\|\rho_{i,j}\|$. The sample correlation between the stocks i and j is defined by

$$s_{i,j} = \frac{\Sigma(r_i(t)-\overline{r_i})(r_j(t)-\overline{r_j})}{\sqrt{\Sigma(r_i(t)-\overline{r_i})^2}\sqrt{\Sigma(r_j(t)-\overline{r_j})^2}},$$

where $\overline{r_i} = \frac{1}{n}\sum_{t=1}^{n} r_i(t)$. It is known Anderson (2003), that for a multivariate normal vector the statistics $(\overline{r_1}, \overline{r_2}, \dots, \overline{r_N})$ and $\|s_{i,j}\|$ are sufficient.

Matrix $\|\rho_{i,j}\|$ is the matrix for the construction of the true market graph and matrix $\|s_{i,j}\|$ is the matrix for the construction of the sample market graph. Each vertex of the graph corresponds to a stock of the financial market. The edge between two vertices i and j is included in the true market graph, if $\rho_{i,j} > \rho_0$ (where ρ_0 is a threshold). For the sample graph, from the statistical point of view, it is more natural to consider the following procedure: the edge between two vertices i and j is included in the sample market graph, if $s_{i,j} > c_{i,j}$ (where $c_{i,j}$ is a threshold). We obtain a market graph construction from Boginsky et al. (2003) in the particular case where $c_{i,j} = \rho_0, i, j = 1, 2, \dots, N$.

In the present paper we study, from the statistical point of view, the procedure of construction of the sample market graph as a procedure of identification of the true market graph.

The problem of identification of the true market graph can be formulated as a multiple decision problem (in the sense of Lehmann, 1957) of the selection of one from the set of hypotheses:

$$\begin{array}{l} H_1 : \rho_{i,j} < \rho_0, \quad \forall (i,j),\ i<j, \\ H_2 : \rho_{12} > \rho_0, \qquad \rho_{i,j} < \rho_0, \quad \forall (i,j) \neq (1,2),\ i<j, \\ H_3 : \rho_{12} > \rho_0, \qquad \rho_{13} > \rho_0, \qquad \rho_{i,j} \le \rho_0, \quad \forall (i,j) \neq (1,2), (i,j) \neq (1,3), \\ \dots \\ H_L : \rho_{i,j} > \rho_0, \quad \forall (i,j),\ i<j, \end{array} \tag{1}$$

where $L = 2^M$ with $M = N(N-1)/2$. Together these hypotheses describe all possible true market graphs. In this setting, the problem is to construct a statistical procedure to choose the true market graph.

3. The Lehmann multiple decision theory

Suppose that the distribution of a random vector is taken from the parametric family $f(x,\theta) : \theta \in \Omega$, where θ is a parameter, Ω is the parametric space. We are looking for the construction of the statistical procedure for the selection of