



## Correlation, hierarchies, and networks in financial markets

Michele Tumminello<sup>a</sup>, Fabrizio Lillo<sup>a,b</sup>, Rosario N. Mantegna<sup>a,\*</sup>

<sup>a</sup> Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, Viale delle Scienze, I-90128 Palermo, Italy

<sup>b</sup> Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

### ARTICLE INFO

#### Article history:

Received 1 September 2008

Received in revised form

31 December 2009

Accepted 14 January 2010

Available online 22 January 2010

#### JEL classification:

C32

G10

#### Keywords:

Multivariate analysis

Hierarchical clustering

Correlation based networks

Bootstrap validation

Factor models

Kullback–Leibler distance

### ABSTRACT

We discuss some methods to quantitatively investigate the properties of correlation matrices. Correlation matrices play an important role in portfolio optimization and in several other quantitative descriptions of asset price dynamics in financial markets. Here, we discuss how to define and obtain hierarchical trees, correlation based trees and networks from a correlation matrix. The hierarchical clustering and other procedures performed on the correlation matrix to detect statistically reliable aspects of it are seen as filtering procedures of the correlation matrix. We also discuss a method to associate a hierarchically nested factor model to a hierarchical tree obtained from a correlation matrix. The information retained in filtering procedures and its stability with respect to statistical fluctuations is quantified by using the Kullback–Leibler distance.

© 2010 Elsevier B.V. All rights reserved.

### 1. Introduction

Many complex systems observed in the physical, biological and social sciences are organized in a nested hierarchical structure, i.e. the elements of the system can be partitioned in clusters which in turn can be partitioned in subclusters and so on up to a certain level (Simon, 1962). The hierarchical structure of interactions among elements strongly affects the dynamics of complex systems. Therefore a quantitative description of hierarchies of the system is a key step in the modeling of complex systems (Anderson, 1972).

The analysis of multivariate data provides crucial information in the investigation of a wide variety of systems. Multivariate analysis methods are designed to extract the information both on the number of main factors characterizing the dynamics of the investigated system and on the composition of the groups (clusters) in which the system is intrinsically organized. Recently, physicists started to contribute to the development of new techniques to investigate multivariate data (Blatt et al., 1996; Hutt et al., 1999; Mantegna, 1999; Giada and Marsili, 2001; Kraskov et al., 2005; Tumminello et al., 2005; Tsafirir et al., 2005; Slonim et al., 2005). Among multivariate techniques, natural candidates for detecting the hierarchical structure of a set of data are hierarchical clustering methods (Anderberg, 1973).

The modeling of the correlation matrix of a complex system with tools of hierarchical clustering has been useful in the multivariate characterization of stock return time series (Mantegna, 1999; Bonanno et al., 2001, 2003), market index returns

\* Corresponding author. Tel.: +39 0916615074; fax: +39 0916615030.

E-mail address: [mantegna@unipa.it](mailto:mantegna@unipa.it) (R.N. Mantegna).

of worldwide stock exchanges (Bonanno et al., 2000), and volatility of stock return time series (Miccichè et al., 2003), where the estimation of statistically reliable properties of the correlation matrix is crucial for several financial decision processes such as asset allocation, portfolio optimization (Tola et al., 2008), derivative pricing, etc. We have termed the selection of statistically reliable information of the correlation matrix with the locution “filtering procedure” as in Tumminello et al. (2007a). Hierarchical clustering procedures are filtering procedures. Other filtering procedures which are popular within the econophysics community are procedures based on the random matrix theory (Laloux et al., 1999; Plerou et al., 1999; Rosenow et al., 2002; Coronello et al., 2005; Potters et al., 2005; Tumminello et al., 2007a), and procedures using the concept of shrinkage of a correlation matrix (Ledoit and Wolf, 2003; Schäfer and Strimmer, 2005; Tumminello et al., 2007b). Many others might be devised and their effectiveness tested.

The correlation matrix of the time series of a multivariate complex system can be used to extract information about aspects of hierarchical organization of such a system. The clustering procedure is done by using the correlation between pairs of elements as a similarity measure and by applying a clustering algorithm to the correlation matrix. As a result of the clustering procedure, a hierarchical tree of the elements of the system is obtained. The correlation based clustering procedure allows also to associate a correlation based network with the correlation matrix. For example, it is natural to select the minimum spanning tree, i.e. the shortest tree connecting all the elements in a graph, as the correlation based network associated with the single linkage cluster analysis. Different correlation based networks can be associated with the same hierarchical tree putting emphasis on different aspects of the sample correlation matrix. Useful examples of correlation based networks different from the minimum spanning tree are the planar maximally filtered graph (Tumminello et al., 2005) and the average linkage minimum spanning tree (Tumminello et al., 2007c).

In correlation based hierarchical investigations, the statistical reliability of hierarchical trees and networks is depending on the statistical reliability of the sample correlation matrix. The sample correlation matrix is computed by using a finite number of records  $T$  sampling the behavior of the  $N$  elements of the system. Due to the unavoidable finiteness of  $T$ , the estimation of the sample correlation matrix presents a degree of statistical uncertainty that can be characterized under widespread statistical assumptions. Physicists (Laloux et al., 1999; Plerou et al., 1999) have contributed to the quantitative estimation of the statistical uncertainty of the correlation matrix by using tools and concepts of random matrix theory. However, theoretical results providing the statistical reliability of hierarchical trees and correlation based networks are still not available and, therefore, a bootstrap approach has been used to quantify the statistical reliability of both hierarchical trees (Tumminello et al., 2007d) and correlation based networks (Tumminello et al., 2007c).

The hierarchical tree characterizing a complex system can also be used to extract a factor model with independent factors acting on different elements in a nested way. In other words, the number of factors controlling each element may be different and different factors may act at different hierarchical levels. Tumminello et al. (2007d) have shown how to associate a hierarchically nested factor model to a system described by a given hierarchical structure.

Having available a large number of filtering procedures, researchers encounter the necessity to have a quantitative methodology able to estimate the information retained in a filtered correlation matrix obtained from the sample correlation matrix. It is also important to quantify the stability of the filtering procedure in different realizations or replicas of the process and a distance of the filtered correlation matrix from a given reference model. For all the above listed purposes, a very useful measure is the one using the Kullback–Leibler distance that was introduced in Tumminello et al. (2007a). This distance presents the important characteristics that its value quantifying the distance between a sample correlation matrix and the correlation matrix of the generating model turns out to be independent from the specific correlation matrix of the model both for multivariate Gaussian variables (Tumminello et al., 2007a) and for multivariate Student's  $t$  variables (Biroli et al., 2007; Tumminello et al., 2007b).

In the present paper we discuss in a coherent and self-consistent way (i) some filtering procedures of the correlation matrix based on hierarchical clustering and the bootstrap validation of hierarchical trees and correlation based networks, (ii) the hierarchically nested factor model, (iii) the Kullback–Leibler distance between the probability density functions of two sets of multivariate random variables and (iv) the retained information and stability of a filtered correlation matrix. We apply the discussed concepts to a portfolio of stocks traded in a financial market. The paper is organized as follows. In Section 2 we discuss how to obtain hierarchical trees and correlation based trees or networks from the correlation matrix of a complex system and we discuss the role of bootstrap in the statistical validation of hierarchical trees and correlation based networks. In Section 3 we discuss the definition and the properties of a factor model with independent factors which are hierarchically nested. In Section 4 we present an empirical application of the hierarchically nested factor model. Section 5 discusses how to quantify the information and stability of a correlation matrix by using a Kullback–Leibler distance and Section 6 presents the quantitative comparison of different filtering procedures performed with the same distance. Section 7 briefly presents some conclusions.

## 2. Correlation based hierarchical organization and networks

Hereafter we discuss a simple example illustrating two filtering procedures of a correlation matrix performed with methods of hierarchical clustering. In our approach, by using the correlation between elements as the similarity measure and by applying a given hierarchical clustering procedure, we first obtain a hierarchical tree. The information present in the hierarchical tree is completely equivalent to the information stored in the filtered matrix and, when the correlation is non-negative for each pair of elements, this matrix is positive definite (Tumminello et al., 2007d).

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات