Commun Nonlinear Sci Numer Simulat 16 (2011) 4610–4618

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Analysis of stock market indices through multidimensional scaling

J. Tenreiro Machado [a], Fernando B. Duarte [b,*], Gonçalo Monteiro Duarte [c]

[a] Dept. of Electrical Engineering, Institute of Engineering, Porto, Portugal
[b] Dept. of Mathematics, School of Technology, Viseu, Portugal
[c] Lusofona University, Lisboa, Portugal

ARTICLE INFO

Article history:
Available online 9 May 2011

Keywords:
Multidimensional scaling
Stock market daily values
Time-varying correlation
Econophysics

ABSTRACT

We propose a graphical method to visualize possible time-varying correlations between fifteen stock market values. The method is useful for observing stable or emerging clusters of stock markets with similar behaviour. The graphs, originated from applying multidimensional scaling techniques (MDS), may also guide the construction of multivariate econometric models.

1. Introduction

Economical indexes measure the performance of segments of the stock market and are normally used to benchmark the performance of stock portfolios. This paper proposes a descriptive method which analyses possible correlations/similarities in international stock markets. Its results are expected to guide the design of statistical models aiming to test hypotheses of interest. Ultimately, the method can even lead to the postulation of new hypotheses. The study of the correlation of international stock markets may have different motivations. Economic motivations to identify the main factors which affect the behaviour of stock markets across different exchanges and countries. Statistical motivations to visualize correlations in order to suggest some potentially plausible parameter relations and restrictions.The understanding of such correlations would be helpful to the design good portfolios [16,18].

Bearing these ideas in mind the outline of our paper is as follows. In Section 2 we give the fundamentals of the multidimensional scaling (MDS) technique, which is the core of our method, and we discuss the details that are relevant for our specific application. In Section 3 we apply our method for daily data on fifteen stock markets, including major American, Asian/Pacific, and European stock markets. In Section 4 we conclude the paper with some final remarks and potential topics for further research.

2. Multidimensional scaling

Generally speaking MDS techniques develop spatial representations of psychological stimuli or other complex objects about which people make judgements (e.g., preference, relatedness), that is they represent each object as a point in a m-dimensional space. What distinguishes MDS from other similar techniques (e.g., factor analysis, cluster analysis) is that in MDS there are no preconceptions about which factors might drive each dimension. Therefore, the only data needed is a measure for the similarity between each possible pair of objects under study. The result is the transformation of the data into similarity measures which can be represented by Euclidean distances in a space of unknown dimensions [4]. The greater

* Corresponding author.
E-mail addresses: jtm@isep.ipp.pt (J.T. Machado), fduarte@estv.ipv.pt (F.B. Duarte), monteiro.duarte@gmail.com (G.M. Duarte).

1007-5704/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cnsns.2011.04.027

the similarity of two objects, the closer they are in the m-dimensional space. After having the distances between all the objects, the MDS techniques analyse how well they can be fitted by spaces of different dimensions. The analysis is normally made by gradually increasing the number of dimensions until the quality of fit (measured for example by the correlation between the data and the distance) is little improved with the addition of a new dimension. In practice a good result is normally reached well before the number of dimensions theoretically needed to a perfectly fit is reached [8,14,19,24].

In the MDS method a small distance between two points corresponds to a high correlation between two stock markets and a large distance corresponds to low or even negative correlation [17]. A correlation of one should lead to zero distance between the points representing perfectly correlated stock markets. MDS tries to estimate the distances for all pairs of stock markets to match the correlations as close as possible. MDS may thus be seen as an exploratory technique without any distributional assumptions on the data. The distances between the points in the MDS maps are generally not difficult to interpret and thus may be used to formulate more specific models or hypotheses. Also, the distance between two points should be interpreted as being the distance conditional on all the other distances. One possibility to obtain such an approximate solution is given by minimizing the stress function. The obtained representation of points is not unique in the sense that any rotation or translation of the points retains the distances [5]. To formalize MDS, we need some notation. Let n be the number of different objects and let the dissimilarity for objects i and j be given by δ_{ij}. The coordinates are gathered in an $n \times p$ matrix $\mathbf{X}$, where p is the dimensionality of the solution to be specified in advance by the user. Thus, row ir from $\mathbf{X}$ gives the coordinates for object i on dimension r. Let d_{ij} be the Euclidean distance between rows i and j of $\mathbf{X}$ defined as

$$d_{ij} = \sqrt{\sum_{r=1}^{p} (x_{ir} - x_{jr})^2} \tag{1}$$

that is, the length of the shortest line connecting points i and j on dimension r. The objective of MDS is to find a matrix $\mathbf{X}$ such that d_{ij} matches δ_{ij} as closely as possible. This objective can be formulated in a variety of ways but here we use the raw-Stress σ^2,

$$\sigma^2 = \sum_{i=2}^{n} \sum_{j=1}^{i-1} w_{ij} (\delta_{ij} - d_{ij})^2 \tag{2}$$

proposed by Kruskal [13], who was the first one to propose a formal measure for doing MDS, where w_{ij} is a user defined weight that must be nonnegative. This measure is also referred to as the least-squares MDS model. Note that due to the symmetry of the dissimilarities and the distances, the summation only involves the pairs i, j where $i > j$. For example, many MDS programs implicitly choose $w_{ij} = 0$ for dissimilarities that are missing. The minimization of σ^2 is a complex problem. Therefore, MDS programs use iterative numerical algorithms to find a matrix $\mathbf{X}$ for which σ^2 is a minimum. In addition to the raw Stress measure there exist other measures for doing Stress. One of them is normalized raw Stress, which is simply raw Stress divided by the sum of squared dissimilarities. The advantage of this measure over raw Stress is that its value is independent of the scale and the number of dissimilarities. The second measure is Kruskal's Stress-1 which is equal to the square root of raw Stress divided by the sum of squared distances. A third measure is Kruskal's Stress-2, which is similar to Stress-1 except that the denominator is based on the variance of the distances instead of the sum of squares. Another measure that seems reasonably popular is called S-Stress and it measures the sum of squared errors between squared distances and squared dissimilarities.

In order to assess the quality of the MDS solution we can study the differences between the MDS solution and the data. One convenient way to do this is by inspecting the so-called Shepard diagram [21]. A Shepard diagram shows both the transformation and the error. Let p_{ij} denotes the proximity between objects i and j. Then, a Shepard diagram plots simultaneously the pairs (p_{ij}, d_{ij}) and (p_{ij}, δ_{ij}). By connecting the (p_{ij}, δ_{ij}) points a line is obtained representing the relationship between the proximities and the disparities. The vertical distances between the (p_{ij}, δ_{ij}) points and (p_{ij}, d_{ij}) are equal to $\delta_{ij} - d_{ij}$, that is, they give the errors of representation for each pair of objects. Hence, the Shepard diagram can be used to inspect both the residuals of the MDS solution and the transformation.

3. Analysis of stocks markets

In this section we study numerically the fifteen selected stock markets, including six American markets, six European markets and three Asian/Pacific markets.

Our data consist of the h daily close values of $s = 15$ stock markets from January 2, 2000, up to December 31, 2009, to be denoted as $x_i(t)$, $1 \leqslant t \leqslant h$, $i = 1, \ldots, s$. The stock markets are listed in Table 1.

The data are obtained from data provided by Yahoo Finance web site [12], and they measure indexes in local currencies. Fig. 1 depicts the time evolution, of daily, closing price of the fifteen stock markets versus year with the well-know noisy and ''chaotic-like'' characteristics.

Assuming that financial index prices are random variables one of the most important analyses' parameter of the financial indexes it is the *volatility* [3]. Volatility measures variability or dispersion about a central tendency. Normally is defined as the deviation from their mean. The historical volatility is the volatility of a series of index prices where we look back over the historical price. The historical volatility estimate, for each index i, is given by