

# with the multidimensional scaling

J. Tenreiro Machado · Gonçalo Monteiro Duarte ·  
Fernando B. Duarte

**Abstract** This paper applied MDS and Fourier transform to analyze different periods of the business cycle. With such purpose, four important stock market indexes (Dow Jones, Nasdaq, NYSE, S&P500) were studied over time. The analysis under the lens of the Fourier transform showed that the indexes have characteristics similar to those of fractional noise. By the other side, the analysis under the MDS lens identified patterns in the stock markets specific to each economic expansion period. Although the identification of patterns characteristic to each expansion period is interesting to practitioners (even if only in a posteriori fashion), further research should explore the meaning of such regularities and target to find a method to estimate future crisis.

**Keywords** Multidimensional scaling · Fractional calculus · Fourier transform · Economic cycle

## 1 Introduction

Modern economies have important swings in their economic activity. While in some periods most industries are growing and unemployment is low, in other periods most are operating well under capacity and unemployment is high. The reasons behind such behavior are complex and normally involve inflation, monetary policy, and business sentiment. Periods of economic prosperity are called expansions and periods of economic decline are called recessions. The combination of these movements is called the business cycle [3]. In the United States, the business cycle is formally followed by the National Bureau of Economic Research (NBER). The NBER's Business Cycle Dating Committee maintains a chronology of the US business cycle, identifying the dates of peaks and troughs that frame economic expansion or recession (Table 1) [11]. Between trough and peak, the economy is in an expansion. A recession begins just after the economy reaches a peak of activity and ends as the economy reaches its trough.

The NBER is especially well known for providing start and end dates for recessions in the United States. Normally, the financial press defines a recession as two consecutive quarters of decline in real Gross Domestic Product (GDP). However, the NBER takes a broader definition, defining a recession as “a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, in-

**Table 1** Business cycle reference dates, in U.S., since 1970

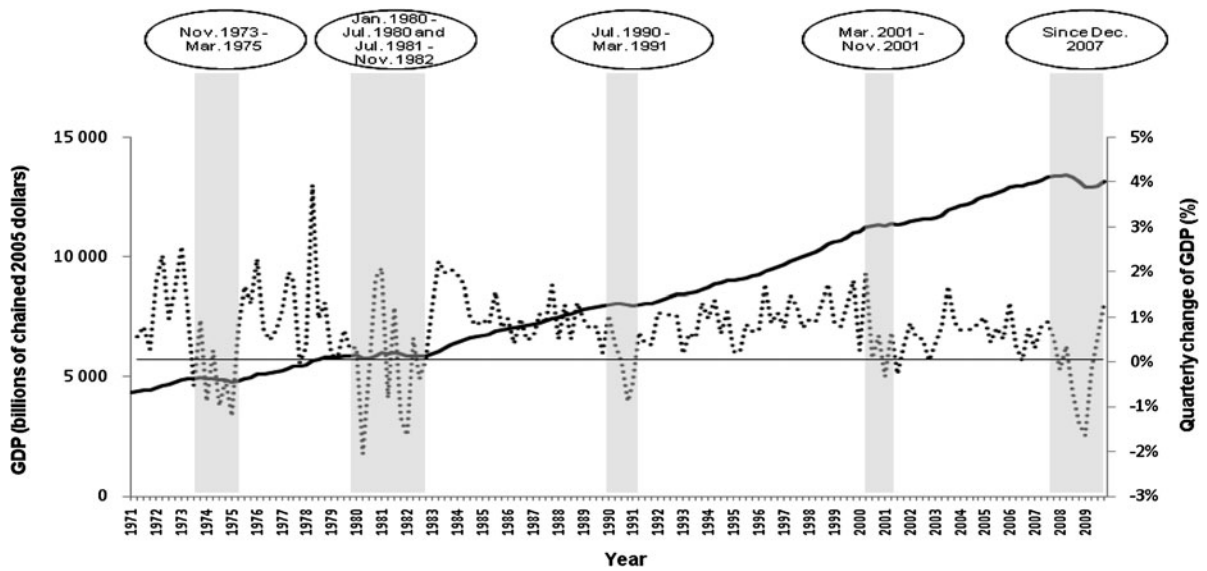
Peak	Trough
November 1973	November 1970
January 1980	March 1975
July 1981	July 1980
July 1990	November 1982
March 2001	March 1991
December 2007	November 2001

dustrial production, and wholesale-retail sales” [10]. Since the 1970s, there have been six recessions. These are now briefly explained [8, 10, 12]:

- Dec. 1969 to Nov. 1970—This recession was preceded by the second longest economic expansion in US history (Feb. 1961 to Dec. 1969). However, this growth was not sustainable and led to accelerating inflation, mainly due to the government’s unwillingness to raise taxes to finance the growing American military involvement in Vietnam and the spike in oil prices. The recession coincided with both a fiscal and monetary tightening aimed to close the budget deficits of the Vietnam War.
- Nov. 1973 to Mar. 1975—This recession was a long and deep recession, characterized for rising unemployment coinciding with rising inflation. Three factors played a large role in this recession: (a) the 1973 oil crisis, (b) the general wage and price control policies implemented in 1971 by the Nixon administration to mask inflation pressures and fight unemployment, and (c) an abnormal long decline in productivity growth.
- Jan. 1980 to Jul. 1980 and Jul. 1981—Nov. 1982—This “double dip” (or w-shaped) recession was actually the conjunction of two separate recessions interrupted by a very short (two quarter) expansion. Because the expansion was so short and the causes of both recessions were the same, we combine them in this analysis. Taken together, the double dip recession represents the deepest and longest recession in the post-war period. The primary cause of the recession was a contractionary monetary policy established by the United States Federal Reserve to control high inflation, heritage from the previous decade and the 1979 energy crisis (caused by the Iranian Revolution).
- Jul. 1990 to Mar. 1991—This recession was not particularly deep or long and can be explained by four major factors: (a) the Fed tightened monetary policy between February 1988 and May 1989 to counter a rising inflation rate; (b) an oil price shock after Iraq invaded Kuwait in August 1990 gave momentum to the starting recession; (c) there were serious solvency problems among thrift institutions due to the Savings and Loan Crisis, and (d) this situation fed a growing consumer pessimism.
- Mar. 2001 to Nov. 2001—The 1990s were the longest period of growth in American history. Such growth ended with a brief recession which was mainly provoked by: (a) the collapse of the speculative dot-com bubble, (b) a fall in business investments, and (c) the September 11th attacks.
- Since Dec. 2007—The actual recession was precipitated by a financial crisis consequent on three factors: (a) the collapse of the subprime mortgage market, (b) the securitization model by which derivatives including toxic mortgages were bundled, (c) the return to tighter monetary policy in 2005.

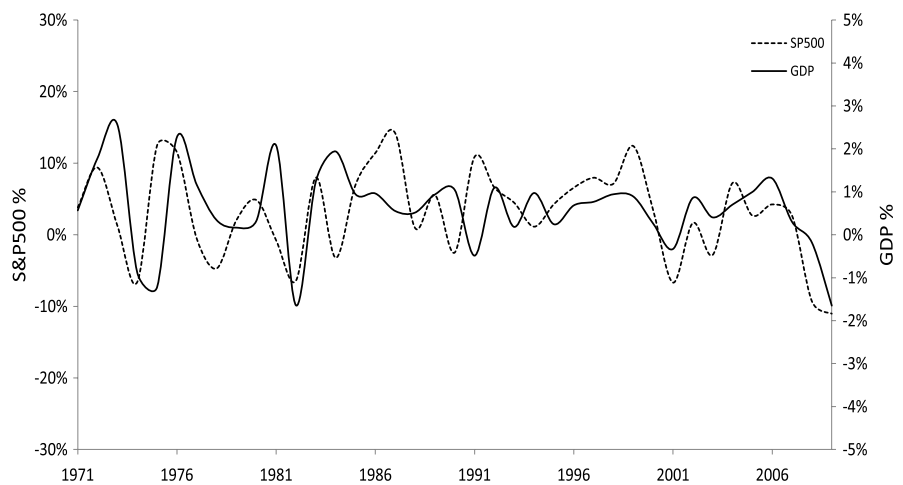
Since the 1970s, there have been six recessions (Fig. 1). Generally speaking, a recession tends to correspond to a bear market (i.e., a general decline in the stock market over a period of time) and a sustainable expansion (i.e., one that keeps inflation from rising too quickly) to a bull market (i.e., a prolonged period in which investment prices rise faster than their historical average, mostly associated with increasing investor confidence). In fact the GDP and the stock market growth tend to correlate quite well. When the economy grows, the stock markets rise and when GDP contracts, the stock market contracts. Nevertheless this correlation is not direct; stock markets tend to behave disproportionately, in a magnified way, when compared with the GDP fluctuations (i.e., when the GDP falls, the stock market falls even more; when it increases, the stock market also increases even more). This is illustrated by Fig. 2 which, as an example, compares the S&P500 [1] and GDP evolutions over time.

That is why investors normally monitor the economy, looking for inflection signs. As the inflection tends to be gradual and to take many months, the stock markets are normally faster to react than the economy and are, therefore, considered by many as a leading indicator of the business cycle. Each period of the economic cycle tends to be characterized by specific beliefs and rules, making the business cycle a good



**Fig. 1** Absolute value of the GDP (*left scale*) [2], trimestral variation of the GDP (*right scale*) and the crisis periods

**Fig. 2** Variation of the GDP (*right scale*) the S&P500 index (*left scale*)



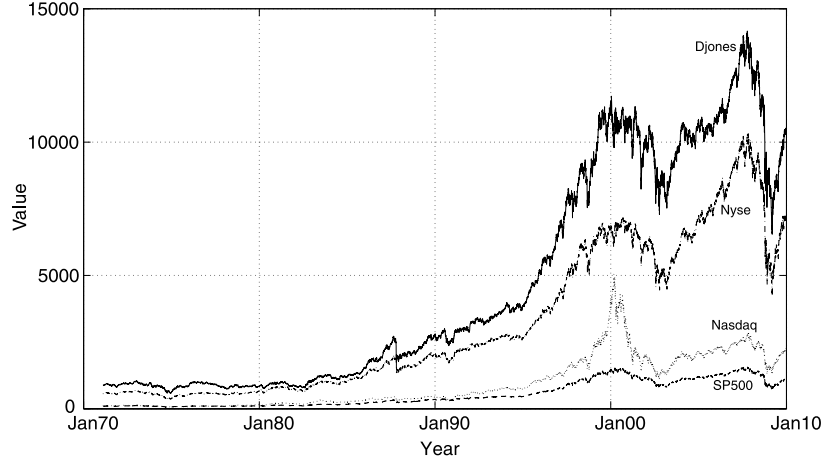
strategic framework for investment. However, during each period, these beliefs and rules can change and may lead to sudden corrections in the market. Extreme examples are stock market crashes, being the most notorious the Wall Street Crash of 1929. Nevertheless, while crashes are normally associated with bear markets, they do not necessarily go together. For instance, the important crash of 1987 did not lead to a bear market. From that day on, the beliefs, the rules, and practices associated with program trading (considered by many the cause of the crash) changed dramatically but the stock markets did not continue to decline.

The remainder of this paper is as follows. In Sect. 2, we present some fundamental concepts about the Fourier transform (FT) and the Multidimensional Scaling (MDS). In Sect. 3, we discuss the methods, perform the dynamical analysis, and present our results. Finally, in Sect. 4, we draw the main conclusions and address perspectives towards future developments.

## 2 Preliminaries

In this section, we present a review of fundamental concepts involved in the dynamical analysis.

**Fig. 3** The temporal evolution of the daily closing value for the Dow Jones, NYSE, S&P500, and NASDAQ indexes, from Feb. 1971 to Dec. 2009



## 2.1 Signals and Fourier transform

In this paper, we study numerically four US stock market indexes' (Dow Jones, NASDAQ, NYSA, and Standard & Poor's (S&P500)) signals for the period from February 8, 1971 (day in which NASDAQ began trading), to December 31, 2009.

Figure 3 depicts the time evolution of the daily closing price of the four indexes *versus* year with the well-known noisy and "chaotic-like" characteristics.

In order to examine the behavior of the signal spectrum, a power law trendline is superimposed to the Fourier Transform (FT) signal, that is, we approximate the modulus of the FT amplitude through the power law (PL):

$$\mathcal{F}\{x_k(t)\} = \int_{-\infty}^{+\infty} x_k(t)e^{-j\omega t} dt, \quad (1)$$

$$|\mathcal{F}\{x_k(t)\}| \approx p_k \omega^{q_k},$$

$$p_k \in \mathbb{R}^+, q_k \in \mathbb{R}, k = \{1, 2, 3, 4\}$$

where  $\mathcal{F}$  is the Fourier operator,  $x_k(t)$  represents the value of  $k$  index,  $t$  is time,  $\omega$  is the frequency,  $p_k$  a positive constant that depends on the signal amplitude, and  $q_k$  is the trendline slope [9]. According to the values of  $q$ , the signals can exhibit an integer or fractional order behavior.

For each signal, index  $x_k(t)$  is calculated the FT and a power trendline approximation.

Figure 4 shows the amplitude of the FT of the Dow Jones, NASDAQ, NYSE, and S&P500 indexes and the corresponding PL slope values  $q_1 = -0.84$ ,  $q_2 = -0.90$ ,  $q_3 = -0.91$ ,  $q_4 = -0.93$ , respectively.

We verify that we get a fractional spectrum in between the white and pink noises, corresponding to a considerable volatility.

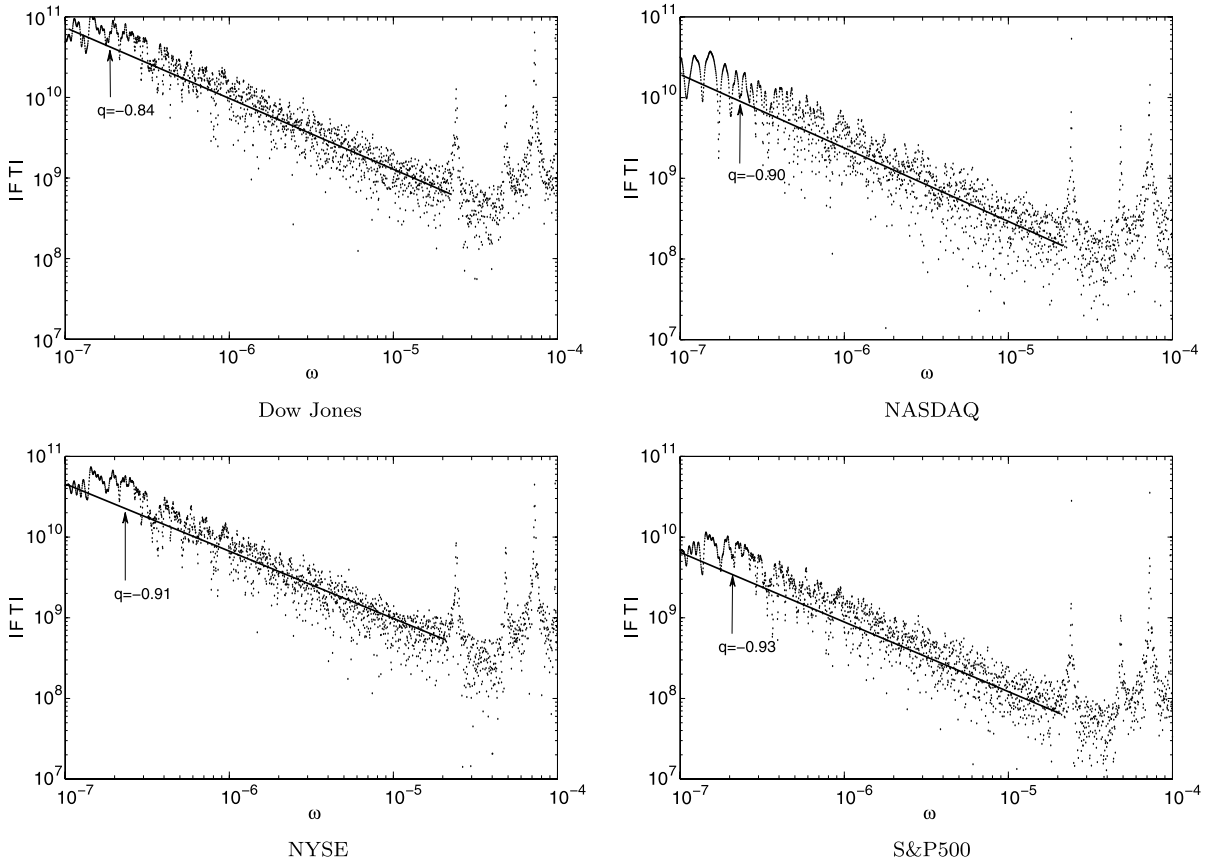
## 2.2 Multidimensional scaling

Multidimensional scaling (MDS) is a set of data analysis techniques for analysis of similarity or dissimilarity data. It is used to represent (dis)similarity data between objects by a variety of distance models.

The term similarity is used to indicate the degree of "likeness" between two objects, while dissimilarity indicates the degree of "unlikeness." MDS represents a set of objects as points in a multidimensional space in such a way that the points corresponding to similar objects are located close together, while those corresponding to dissimilar objects are located far apart. The researcher then attempts to "make sense" of the derived object configuration by identifying meaningful regions and/or directions in the space. In this article, we first introduce the basic concepts and methods of MDS. We then discuss a variety of (dis)similarity measures and the kinds of techniques to be used.

The main objective of MDS is to represent these dissimilarities as distances between points in a low dimensional space such that the distances correspond as closely as possible to the dissimilarities.

Let  $n$  be the number of different objects and let the dissimilarity for objects  $i$  and  $j$  be given by  $\delta_{ij}$ . 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  $i$  from  $\mathbf{X}$  gives the



**Fig. 4**  $|FT\{x_k(t)\}|$  and the power trendline for the indexes Dow Jones, NASDAQ, NYSE, and S&P500

coordinates for object  $i$ . Let  $d_{ij}$  be the Euclidean distance between rows  $i$  and  $j$  of  $\mathbf{X}$  defined as

$$d_{ij} = \sqrt{\sum_{s=1}^p (x_{is} - x_{js})^2} \quad (2)$$

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

$$\sigma^2 = \sum_{i=2}^n \sum_{j=1}^{i-1} w_{ij} (\delta_{ij} - d_{ij})^2 \quad (3)$$

by Kruskal [4] who was the first one to propose a formal measure for doing MDS. 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  $i > j$ . Here,  $w_{ij}$  is a user defined weight that must be nonnegative. For example, many MDS programs implicitly choose  $w_{ij} = 0$  for dissimilarities that are missing. The minimization of  $\sigma^2$  is a complex problem. Therefore, MDS programs use iterative numerical algorithms to find a matrix  $\mathbf{X}$  for which  $\sigma^2$  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 error between squared distances and squared dissimilarities.

Because Euclidean distances do not change under rotation, translation, and reflection, these operations may be freely applied to MDS solution without affecting the raw-Stress. Many MDS programs use this indeterminacy to center the coordinates so that they sum to zero dimensionwise. The freedom of rotation is often exploited to put the solution in so-called principal axis orientation. That is, the axis are rotated in such a way that the variance of  $\mathbf{X}$  is maximal along the first dimension, the second dimension is uncorrelated to the first and has again maximal variance, and so on.

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. A Shepard diagram shows both the transformation and the error. Let  $p_{ij}$  denote the proximity between objects  $i$  and  $j$ . Then a Shepard diagram plots simultaneously the pairs  $(p_{ij}, d_{ij})$  and  $(p_{ij}, \delta_{ij})$ . By connecting the  $(p_{ij}, \delta_{ij})$  points a line is obtained representing the relationship between the proximities and the disparities. The vertical distances between the  $(p_{ij}, \delta_{ij})$  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. Outliers can be detected as well as possible systematic deviations.

### 3 Dynamics of stock market indexes from a daily closing price standpoint

The data of the stock market indexes' in study consists of daily closing price  $x_k(t)$ , where  $k = \{1, 2, 3, 4\}$  identifies the stock market indexes. The data was collected from the Yahoo finance web site [1].

In order to reveal possible relationships between the signals, the MDS technique is used. For that purpose, we apply the MDS method described in Sect. 2.2 to the time correlation of the selected stock markets, as defined in the sequel.

We consider the time correlations between the daily close values. For the purpose of analyzing the time dy-

namics, each index is "sliced" into annual series, leading to 39 1-year length series.

Firstly, over each index, we compute the correlation value  $c_k(i, j)$  among the 39 annual series according the formula

$$c_k(i, j) = \left( \frac{\frac{1}{n} \sum_{u=1}^n x_i^k(u) \cdot x_j^k(t)}{\sqrt{\frac{1}{n} \sum_{u=1}^n (x_i^k(u))^2 \cdot \frac{1}{n} \sum_{u=1}^n (x_j^k(u))^2}} \right)^2 \quad (4)$$

$i, j = 1, \dots, 39$ ,  $n = 250$  days, getting four  $39 \times 39$  matrices  $\mathbf{C}_k$ ,  $k = \{1, 2, 3, 4\}$ , where each cell represents the time correlation between a pair of one-year series.

Secondly, with the same 39 annual series of each index, we calculate the FT for each one. With the FT values  $\text{Re}[\mathcal{F}\{x_k^r(t)\}] + j \text{Im}[\mathcal{F}\{x_k^r(t)\}]$  with  $k = \{1, 2, 3, 4\}$  and  $r = \{1, \dots, 39\}$  using the "distance measure" defined in the following equation:

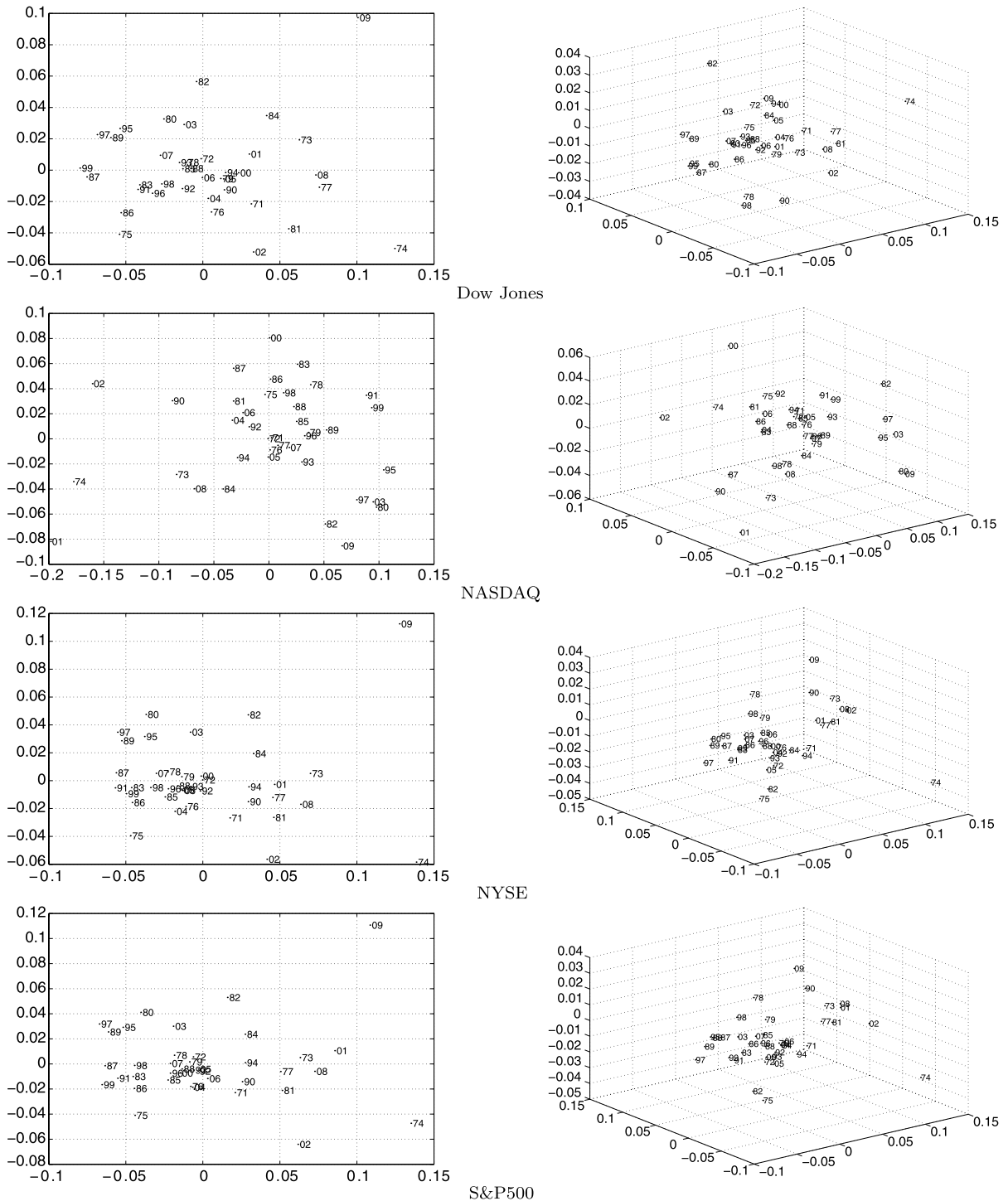
$$s_k(i, j) = \frac{\sum_{\Omega} (|\text{Re}_i^r - \text{Re}_j^r|^2 + |\text{Im}_i^r - \text{Im}_j^r|^2)}{\sum_{\Omega} (|\text{Re}_i^r|^2 + |\text{Im}_i^r|^2) \cdot \sum_{\Omega} (|\text{Re}_j^r|^2 + |\text{Im}_j^r|^2)} \quad (5)$$

where  $\Omega$  is the set of sampling frequencies for the FT calculation,  $\text{Re}_i$ ,  $\text{Im}_i$ ,  $\text{Re}_j$ , and  $\text{Im}_j$  are the values of real part and imaginary part of the FT for the index  $k = \{1, 2, 3, 4\}$ , for  $i, j = 1, \dots, 39$  series. We get an other set of four  $39 \times 39$  matrices  $\mathbf{F}_k$ ,  $k = \{1, 2, 3, 4\}$ , where each cell represents the distance between a pair of FTs.

In order to reveal possible relationships between the signals, the MDS technique is used and several distance criteria are tested. The Sammon criterion [5, 6] that tries to optimize a cost function that describes how well the pairwise distances in a data set are preserved, revealed good results, and is adopted in this work.

So, all the eight matrices are subjected to a MDS analysis [7] with the following parameters:

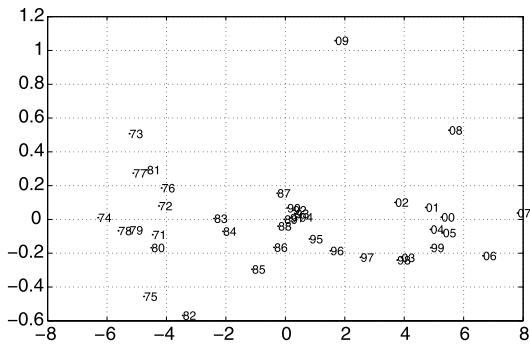
- Metric multidimensional scaling
- Uniform weighting
- Absolute scaling model
- Stress method: sammon
- Dimension of the representation space: 2–4



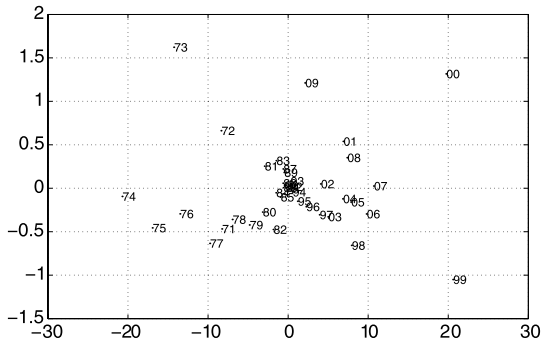
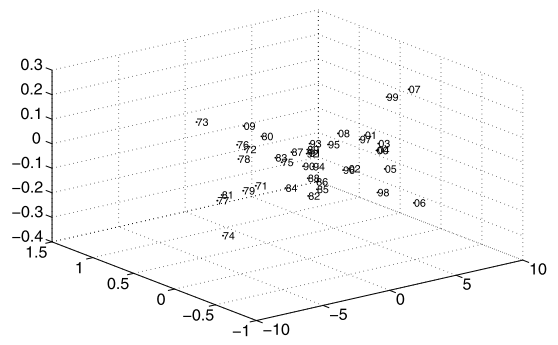
**Fig. 5** MDS for the indexes Dow Jones, NASDAQ, NYSE, and S&P500 based on the correlation distance (5): 2D (left) and 3D (right)

- Repetitions: 20
- Iterations: 200
- Convergence: 0.0001

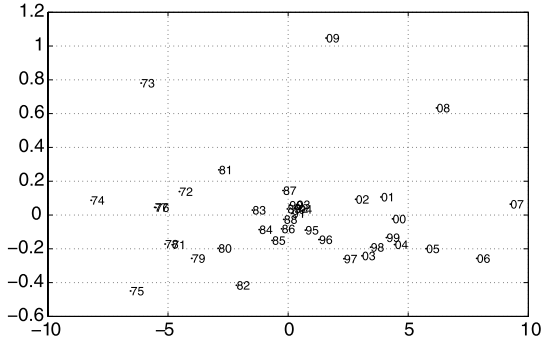
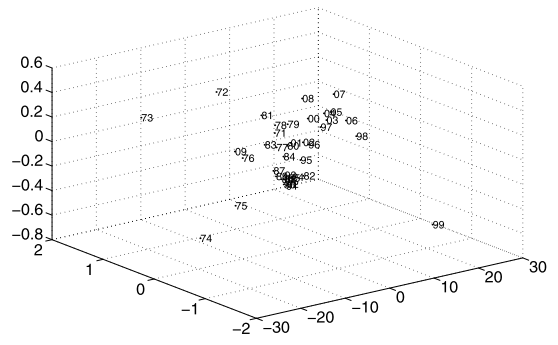
Figures 5 and 6 show the 2D and 3D MDS locus of index positioning in the perspective of the expressions (4) and (5), respectively. Figures 7 and 8 depict



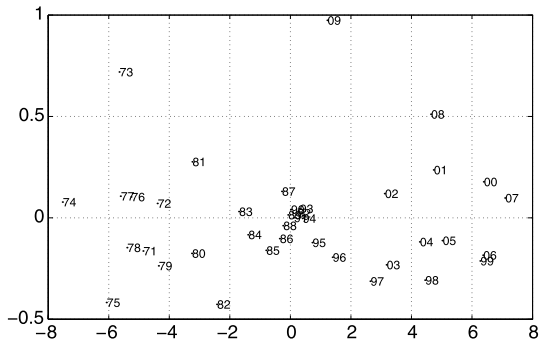
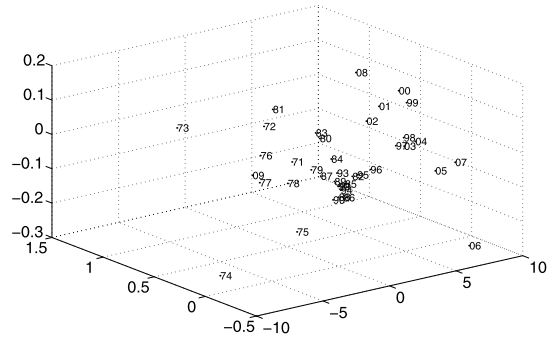
Dow Jones



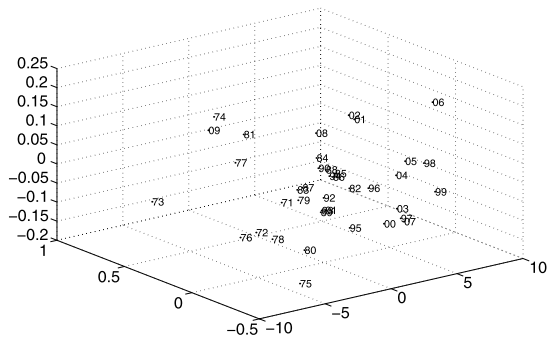
NASDAQ



NYSE



S&P500



**Fig. 6** MDS for the indexes Dow Jones, NASDAQ, NYSE, and S&P500 based on the FT distance (5): 2D (left) and 3D (right)



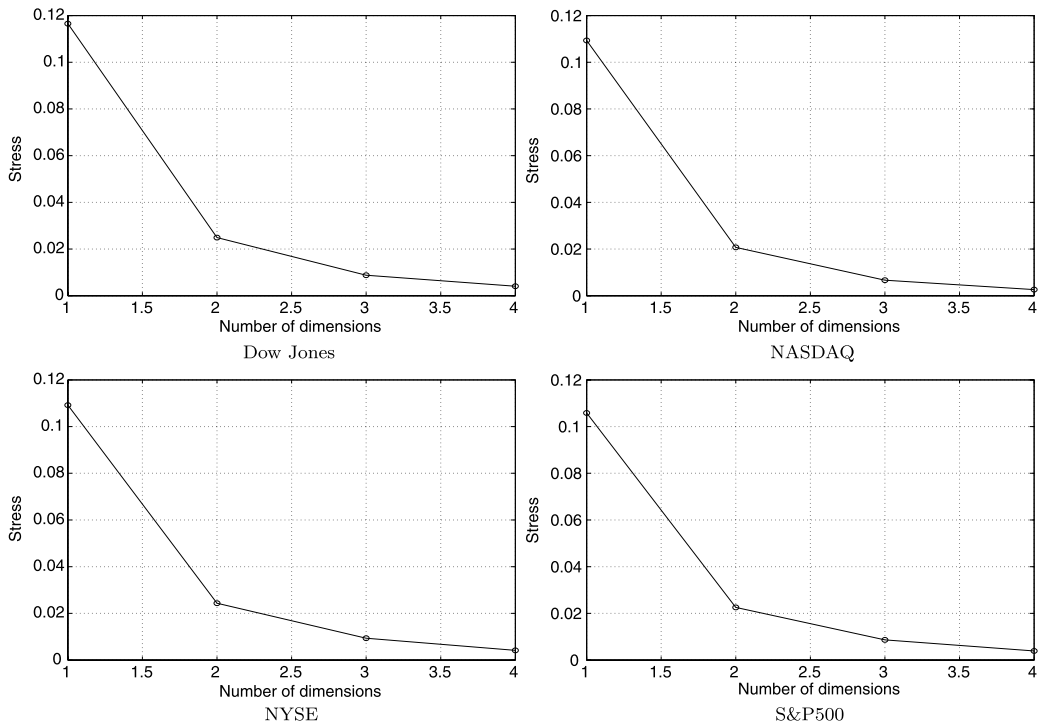


Fig. 7 Stress plot of the MDS representation vs. number of dimensions using (4)

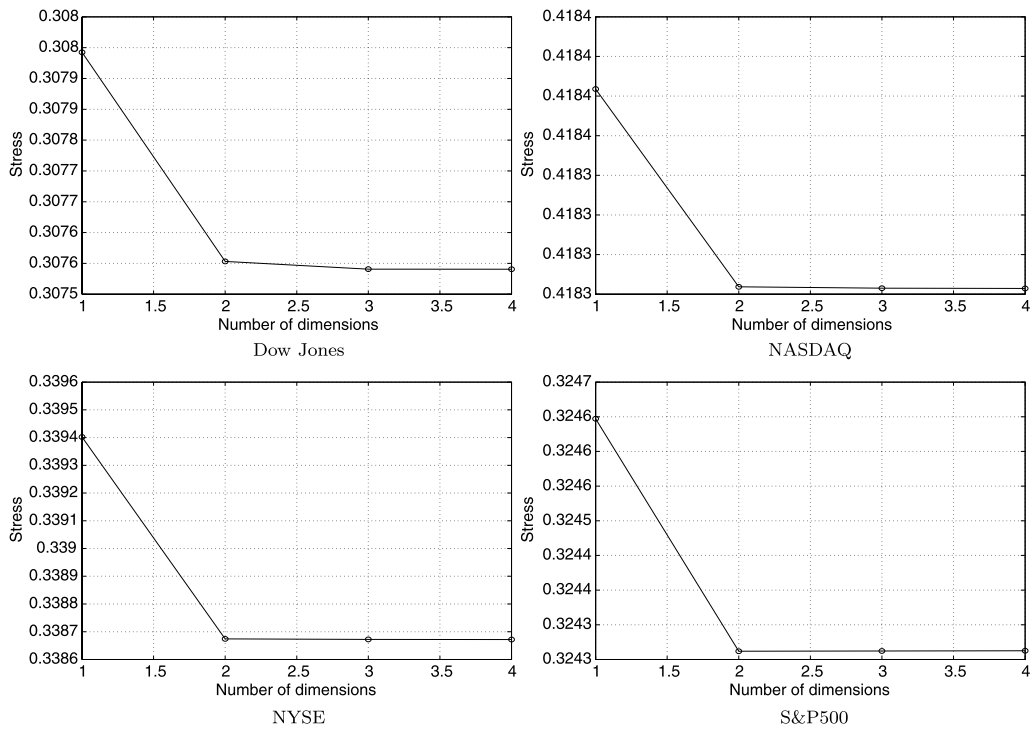
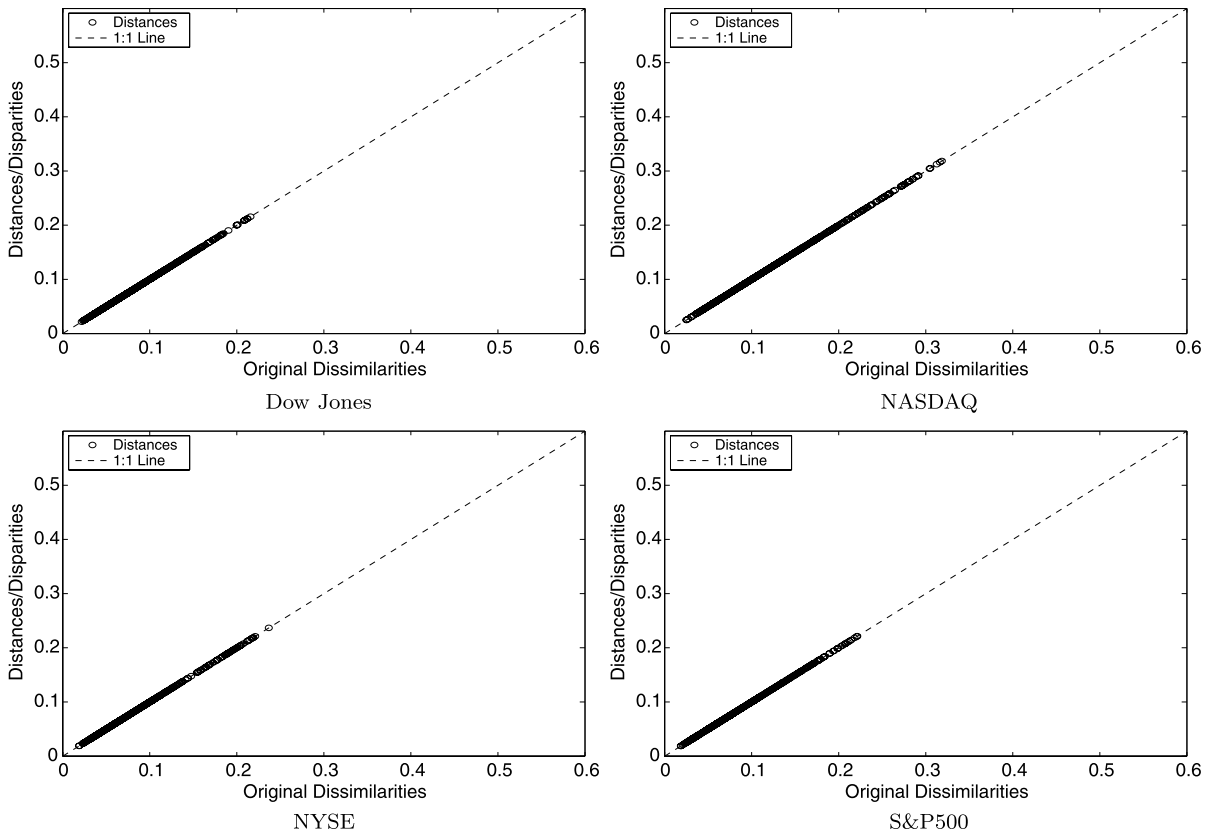


Fig. 8 Stress plot of the MDS representation vs. number of dimensions using (5)



**Fig. 9** Shepard plot of the three-dimensional MDS representation according (4)

the stress as function of the dimension of the representation space, revealing that, as usual, a high dimensional space would probably describe slightly better the “map” of the 39 series index. However, the three-dimensional representations were adopted because the graphical representation is easier to analyze while yielding a reasonable accuracy. Moreover, the resulting Sheppard plot, represented in Figs. 9–10, show that a good distribution of points around the 45-degree line is obtained.

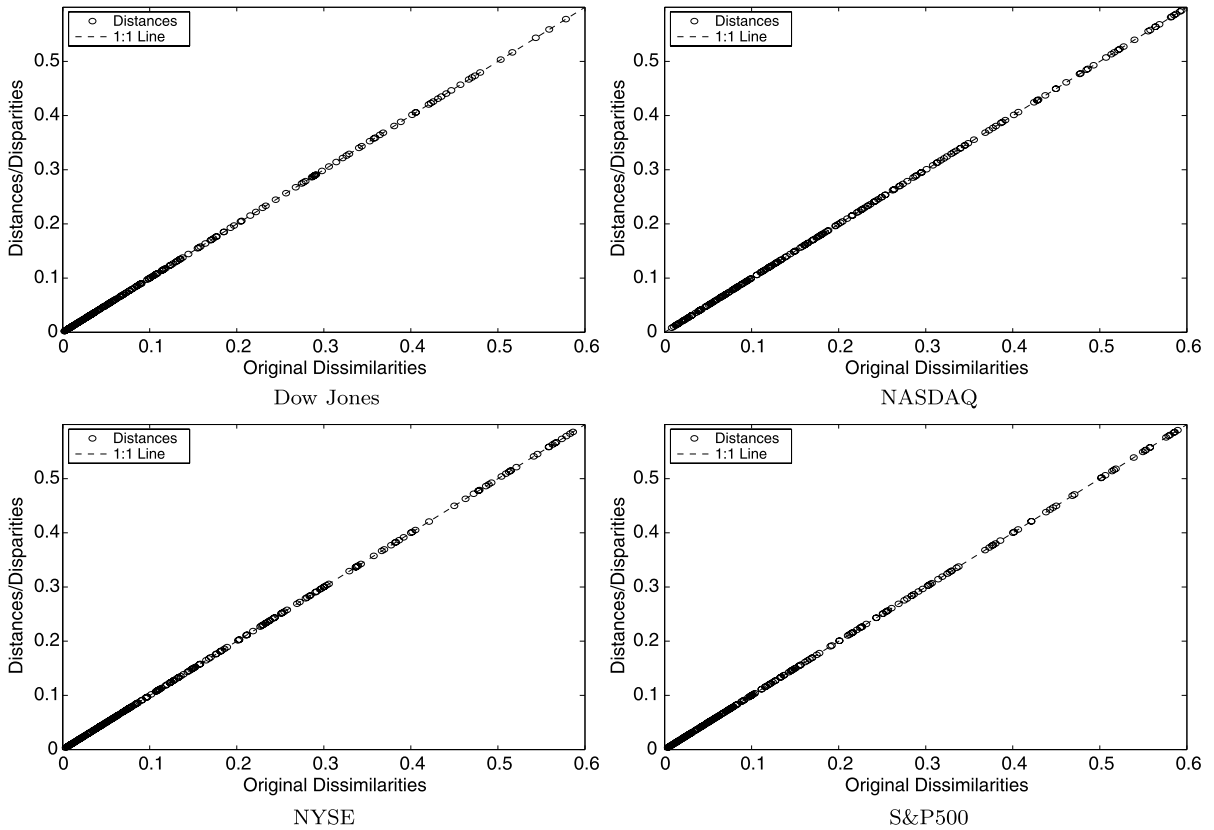
In the first case (i.e., Fig. 5), it seems that there is no clear pattern. However, in the second case (i.e., Fig. 6) a pattern emerges not only between, but also within indexes.

Analyzing Fig. 6, it is visible that the MDS curves of the four stock market indexes seem to be divided into six periods with different flows, separated by the end of an economic crisis or by a major stock market crash. This observation supports that in the stock market there are rules, behavior, and practices which are characteristic of each period of the business cycle.

Curiously, these norms seem to be translated into the MDS in the form of arc shapes in all the four stock market indexes under analysis. How can we interpret these regularities? This is an interesting question for further research and a matter of decision of the stock market handlers.

Figure 11 shows the 2D MDS chart with arcs superimposed in each cycle. We should note that probably we can find other kind of arc curves, such as spirals and ellipses that may also fit adequately. It is interesting to note that Nasdaq’s MDS curves have different patterns when compared with the curves of the other indexes. This may probably be explained by the different nature of the Nasdaq-traded companies (i.e., usually related with technology), leading to different trading practices and, therefore, to a different behavior of the index (e.g., higher volatility).

Table 2 presents the main economic periods and crisis in period under this study [11, 12].



**Fig. 10** Shepard plot of the three-dimensional MDS representation according to (5)

**Table 2** Crises and normal economic periods from 1970

Period	Economic context
1970–1972	Expansion
1973–1975	Recession
1975–1980	Expansion
1980–1982	Recession
1983–1987	Expansion
1987–1988	Recession
1988–1990	Expansion
1990–1992	Recession
1993–2000	Expansion
2001–2002	Recession
2003–2007	Expansion
2007–	Recession

#### 4 Conclusions

This paper analyzed the evolution of four important stock market indexes. An initial evaluation using the

Fourier transform showed that the indexes have characteristics similar to those of fractional noise, somehow in between the white and pink noises. For the purpose of establishing a tool taking into consideration both the time dynamics and the statistical properties, it was adopted a modified version of the MDS. In fact, MDS does not see directly the time dynamics of the variable which have only an indirect influence in the final map through the signal comparison measure. For embedding time dynamics into the MDS, that is, for having a dynamical MDS map, the stock indexes were divided in time slices. The 1-year windows of the stock market indexes were then compared through two distinct measures, namely the correlation in the time domain and the distance in the frequency domain. The second measure showed to be more successful in revealing economic periods and crisis. In the four indexes, the MDS maps of the time slices depicted clearly regions, with arc shapes, corresponding to well-known economic periods. While this a posteriori analysis proves to be correct, several new ques-

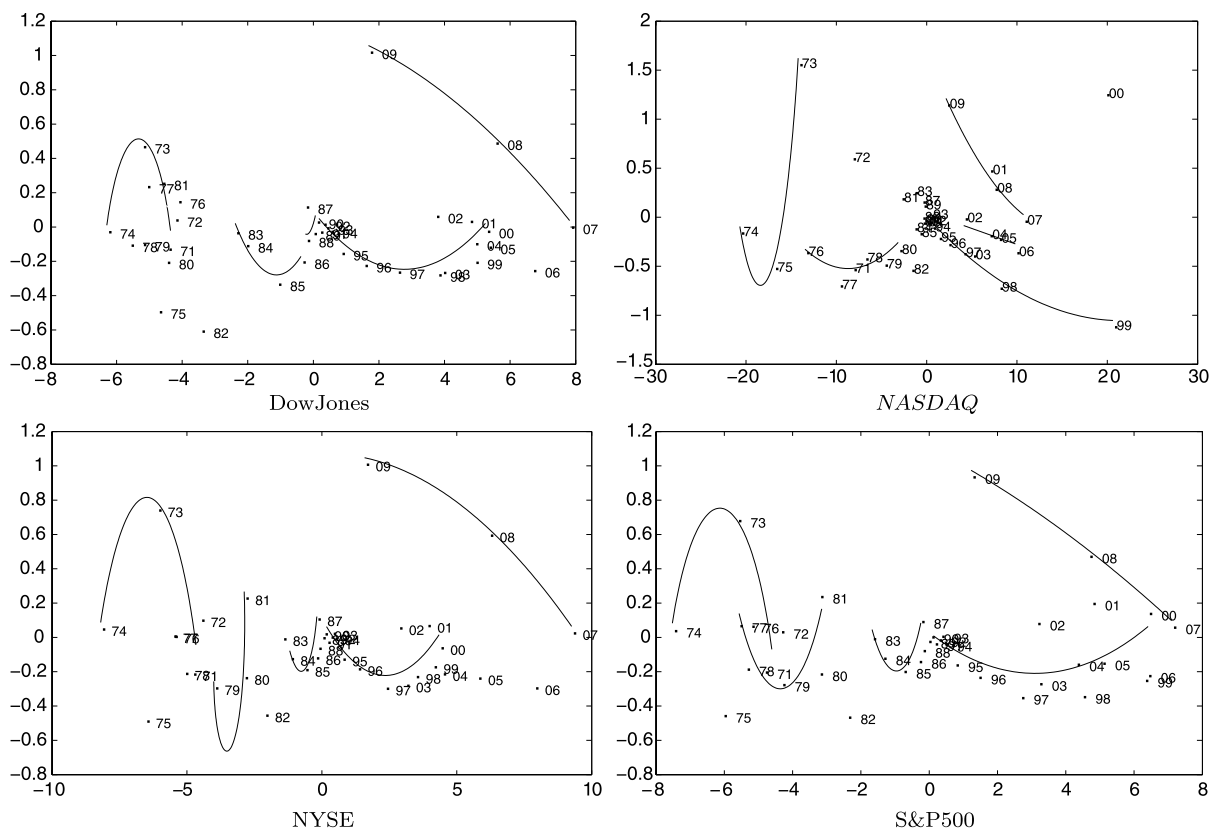


Fig. 11 MDS and arc curves for all indexes with FT distance in (5)

tions emerge, such as the meaning of the shape of the arcs in each region and if the possibility of using the method for estimating, may be “guessing,” future economic periods.

## References

1. <http://finance.yahoo.com>
2. U.S. Department of Commerce, Bureau of Economic Analysis: <http://www.bea.gov/national/index.htm#gdp>
3. Burns, Arthur F., Wesley, C. Mitchell: Measuring Business Cycles. National Bureau of Economic Research, Washington (1946)
4. Kruskal, J.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27 (1964). <http://dx.doi.org/10.1007/BF02289565>
5. Sammon, J.W.: A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.* **18**(5), 401–409 (1969)
6. Seber, G.A.F.: *Multivariate Observations*. Wiley, New York (1986)
7. Cox, T., Cox, M.: *Multidimensional Scaling*. Chapman & Hall/CRC, New York (2001)
8. Knoop, T.A.: *Recessions and Depressions: Understanding Business Cycles*. Praeger, New York (2004)
9. Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, E.: A unified econophysics explanation for the power-law exponents of stock market activity. *Physica A* **382**, 81–88 (2007)
10. Bordo, M.D., Landon-Lane, J.: Determination of the December 2007 peak in economic activity. Working Paper Series (2008). <http://www.nber.org/dec2008.pdf>
11. National Bureau of Economic Research: USA business cycle expansions and contractions (2008). <http://www.nber.org/cycles.html>
12. Bordo, M.D., Landon-Lane, J.: Exits from recessions: the u.s. experience 1920–2007. In: Reinhart, V. (ed.) *Proc. No Way Out: Government Response to the Financial Crisis* (2009)