



Information and hierarchical structure in financial markets

R.N. Mantegna¹

*Istituto Nazionale per la Fisica della Materia, Unità di Palermo and Dipartimento di Energetica ed Applicazioni di Fisica,
Università di Palermo, Viale delle Scienze, I-90128, Palermo, Italy*

Abstract

I investigate the information content present in the time series of stock prices of a portfolio of stocks traded in a financial market. By investigating the correlation coefficient between pairs of stocks I provide a working definition of a generalized distance between the stocks of the portfolio. This generalized distance is used to obtain an ultrametric distance matrix between the stocks. The ultrametric structure of the portfolio investigated has associated a taxonomy which is meaningful from an economic point of view. © 1999 Elsevier Science B.V. All rights reserved.

1. Introduction

In recent years a group of physicists have started to investigate financial and economic systems by using tools and methodologies which are specific to physical sciences. This choice has been motivated by several reasons. Firstly, several financial and economic systems are very good examples of *complex systems*. Secondly, starting from the eighties a huge amount of financial and economic data are recorded in a computerized form and are easily accessible for analyzes and tests. The data are monitoring in detail the considered processes and, in the case of some stock markets, the data about each elementary action of the market are available (there exists trade and quotes databases where each quote and each trade occurring in the market are recorded). Thirdly, the analysis and modeling of economic and financial systems connote fundamental and applied aspects. The fundamental aspects concern the modeling of a system composed by several sub-units which are usually interacting through nonlinear interactions in the presence of noise

and in the absence of a known conserved observable and/or symmetry properties. The applied aspects concern the rigorous definition and the best evaluation of the risk present in several economic and financial activities.

Several studies have been published in physical, economic and interdisciplinary journals. Some examples are the ones considering

- (i) the study of the statistical properties of stock price changes [1–3];
- (ii) a novel approach to the option pricing problem [4,5];
- (iii) the investigation of models of artificial financial markets [6–10] and
- (iv) the comparison between price dynamics in financial markets and velocity dynamics in turbulence [11,12].

The above examples are not exhaustive and several other problems are and could be addressed by physicists. In this lecture I try to give an answer to the following question: how to detect the existence of economic information stored in the time series of a stock price?

¹ E-mail: mantegna@ifatal.deaf.unipa.it.

2. Efficient market hypothesis

The basic stochastic model of stock price dynamics is the model assuming that the natural logarithm of the price is a diffusive process. Empirical studies performed to quantify the degree of temporal correlation in the time evolution of stock price differences have shown (see, for example, [13]) that time correlation is rather weak or absent in a time interval ranging from less than a trading day to several years. The absence of long term memory in the pairwise autocorrelation function of logarithm price changes is consistent with the assumption of the efficiency of stock markets. In an efficient market all the information available (past, present and prediction about the future) is immediately reflected in the price of the traded asset.

The modeling of the time series of discounted price of a financial good in terms of a stochastic process (specifically in terms of a martingale [14,15]) may seem paradoxical at first sight: A so important indicator, such as the price of a financial asset, is indistinguishable from a random process.

The resolution of the above paradox lays in the fact that time series which are rich of information are indeed indistinguishable from random processes. This last statement is one of the conclusions of the algorithm complexity theory. Hence, the algorithm complexity theory allows us to be consistent by simultaneously stating that (i) time series of discounted price differences are dynamical time series carrying a large amount of information and (ii) they are well modeled in terms of random processes.

3. Extracting information from time series

In this lecture I propose one possible answer to the following question: how to detect the existence of economic information stored in the time series of price of a financial good?

The main idea is to investigate several time series of stock prices in parallel. Specifically, I investigate [16, 17] the synchronous correlation coefficient between the time evolution of

$$Y_i = \ln P_i(t) - \ln P_i(t-1)$$

of a pair of stocks i and j , where i and j are the numerical labels of the stocks and $P_i(t)$ is the daily

closure price of the stock i at the day t . I compute the correlation coefficient by following the mathematical definition [18]

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}. \quad (1)$$

The average is a temporal average performed on all the trading days of the investigated time period. ρ_{ij} quantifies the correlation observed between the logarithm of price of the stock i and the logarithm of price of the stock j . ρ is varying from -1 to 1 . $\rho = 1$ indicates full correlation between Y_i and Y_j variables, $\rho = -1$ means fully anti-correlation between i and j while $\rho = 0$ is observed for an uncorrelated pair of stocks.

A portfolio of n stocks has associated an $n \times n$ correlation coefficient matrix. The matrix is symmetric with 1 in the main diagonal. The number of correlation coefficients necessary to completely describe the matrix is $n(n-1)/2$.

The portfolio of stocks discussed in this lecture is the portfolio of the 30 stocks used to compute the Dow Jones Industrial Average. By analyzing the correlation coefficients between these stocks determined during the time period from July 1989 to October 1995 and during the calendar years from 1990 to 1994, I observe [16] that a significant degree of correlation has been present between several of the pairs of stocks in the investigated periods. For example, during the calendar year of 1990, I detect a maximal correlation coefficient of 0.73 between Coca Cola and Procter & Gamble. During the same period the minimal correlation coefficient detected is 0.02.

4. A meaningful taxonomy

In the attempt to determine from the measured correlation coefficients a meaningful taxonomy, I first define a metric. A generalized distance [17] can be given by a function of the correlation coefficient. A tentative definition is

$$d(i, j) = 1 - \rho_{ij}^2. \quad (2)$$

With this choice, in the specific investigated case $d(i, j)$ numerically fulfills the three axioms of an Euclidean metric: (i) $d(i, j) = 0$ if and only if $i = j$; (ii) $d(i, j) = d(j, i)$ and (iii) $d(i, j) \leq d(i, k) +$

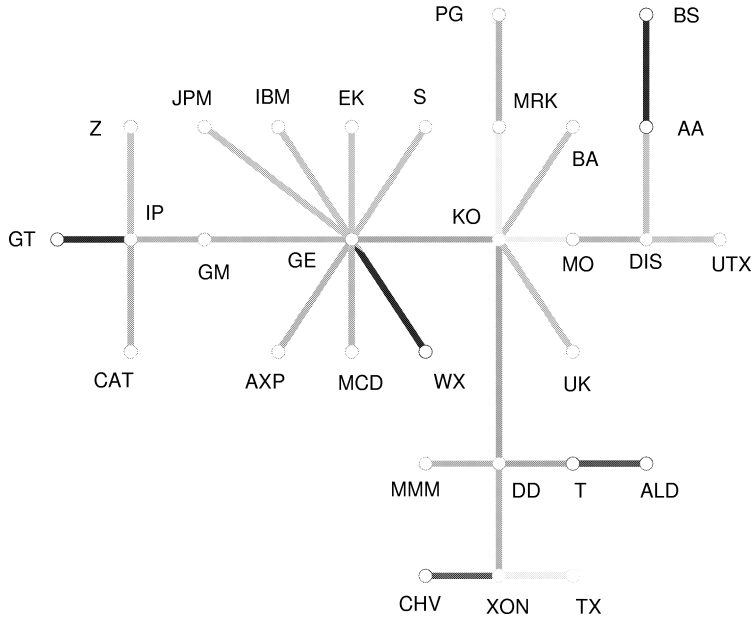


Fig. 1. Minimal spanning tree linking the 30 stocks used to compute the Dow Jones Industrial Average. The period of time used to compute the distance matrix is the calendar year 1991. The stocks are identified by their tic symbol. (AA – Alcoa, ALD – Allied Signal, AXP – American Express Co, BA – Boeing Co, BS – Bethlehem Steel, CAT – Caterpillar Inc., CHV – Chevron Corp., DD – Du Pont, DIS – Walt Disney Co., EK – Eastman Kodak Co., GE – General Electric, GM – General Motors, GT – Goodyear Tire, IBM – IBM Corp., IP – International Paper, JPM – Morgan JP, KO – Coca Cola Co., MCD – McDonald's Corp., MMM – Minnesota Mining, MO – Philips Morris, MRK – Merck & Co Inc., PG – Procter & Gamble, S – Sears Roebuck, T – AT&T, TX – Texaco Inc., UK – Union Carbide, UTX – United Tech, WX – Westinghouse, XON – Exxon Corp. and Z – Woolworth). The gray scale of segments connecting stocks groups different links with respect to the value of the distance between the stocks ($CHV-XON$ $0.60 < d(i, j) \leq 0.65$; $MRK-KO-MO$ and $XON-TX$ $0.65 < d(i, j) \leq 0.70$; $GE-KO-DD-T$ $0.70 < d(i, j) \leq 0.75$; $MO-DIS$, $PG-MRK$, $CAT-IP-GM-GE$, $AXP-GE-MCD$ and $MMM-DD-XON$ $0.75 < d(i, j) \leq 0.80$; $IP-Z$, $BA-KO-UK$, $JPM-GE$, $GE-IBM$, $EK-GE-S$ and $AA-DIS-UTX$ $0.80 < d(i, j) \leq 0.85$; $GT-IP$, $GE-WX$ and $AA-BS$ $0.85 < d(i, j) \leq 0.90$; $ALD-T$ $0.90 < d(i, j) \leq 0.95$).

$d(k, j)$, and then can be used as an Euclidean distance for all practical purposes.

With this approach, I am able to give a pragmatic definition of a metric distance between the stocks of a given portfolio. However, a metric distance is not enough to indicate a taxonomy in a unique way. Hence, I look at a more restricted topological space which has associated a unique taxonomy. This restricted topological space is the subdominant ultrametric space associated with the Euclidean distance matrix [19]. The ultrametric distance is defined by three axioms. The first two coincide with the ones defining the Euclidean distance while the triangular inequality is replaced by a stronger inequality which is named the ultrametric inequality

$$d(i, j) \leq \text{Max}\{d(i, k), d(k, j)\}. \quad (3)$$

In a space of elements where a metric is defined, it is always possible to obtain the subdominant ultrametric. An example of the procedure used to extract the subdominant ultrametric distance matrix from the Euclidean distance matrix can be found in Ref. [19]. In an ultrametric space the order of elements is not along the real line but on a hierarchical tree. The set of stocks of the portfolio analyzed can be seen as an abstract space. The n elements of this abstract space may be connected by a graph. In a connected graph with given distances between all the points of the abstract space the minimal spanning tree (MST) has $n - 1$ edges and has the minimal length of all the spanning tree with respect to the sum of the distances between two points. The structure of the hierarchical

tree is obtained using the ultrametric distance matrix and the MST associated to it.

Starting from the 30×30 Euclidean matrix, I obtain the MST and the hierarchical tree associated to the 30 stocks of the portfolio for the different periods investigated. To illustrate the results, in Fig. 1 I show the MST observed by investigating the calendar year of 1991.

The MST and the associated hierarchical tree [17] show the existence of clusters of stocks which are meaningful from an economic point of view. The strongest links are observed between the three oil companies (Chevron, Exxon and Texaco). Interesting links are also observed between consumer product companies like Coca Cola, Philips Morris and Merck & Co. The obtained taxonomy is able to group stocks which are homogeneous with respect to the economic activity of the firms.

5. Summary

Our analysis shows that starting from the time series of stock prices present in a given portfolio it is possible to build up a procedure that provide us a taxonomy of the considered stocks. In the investigation performed by considering the 30 stocks of the Dow Jones Industrial Average, the stocks are organized in a taxonomy which is meaningful from an economic point of view. Specifically, the cluster of the oil companies and the cluster of consumer product companies are clearly observed. The link between the two clusters of companies is realized throughout Du Pont Company namely a company which is producing custom oriented chemical products.

The observation of a meaningful taxonomy obtained starting from a quantitative analysis of the time series of stock prices is fully consistent with the picture I briefly described in the introduction of this lecture, namely that the time series of stock price in a stock market are time series rich of information that are hardly distinguishable from random processes. The encoded information is in part retrievable if one uses efficient strategies. This kind of studies can be used to detect the common economic factors which

are driving the temporal evolution of groups of stocks traded in stock markets.

Acknowledgements

I wish to thank INFN and MURST for financial support.

References

- [1] R.N. Mantegna, *Physica A* 179 (1991) 232.
- [2] R.N. Mantegna, H.E. Stanley, *Phys. Rev. Lett.* 73 (1994) 2946; R.N. Mantegna, H.E. Stanley, in: *Lévy Flights and Related Topics in Physics*, M.F. Shlesinger, G.M. Zaslavsky, U. Frisch (Eds.) (Springer, Berlin, 1995) 300–312.
- [3] R.N. Mantegna, H.E. Stanley, *Nature* 376 (1995) 46.
- [4] J.-P. Bouchaud, D. Sornette, *J. Phys. I (France)* 4 (1994) 863.
- [5] E. Aurell, R. Baviera, O. Hammarlid, M. Serva, A. Vulpiani, *Gambling and Pricing of Derivatives*, Preprint (1998).
- [6] H. Takayasu, H. Miura, T. Hirabayashi, K. Hamada, *Physica A* 184 (1992) 127.
- [7] P. Bak, K. Chen, J.A. Scheinkman, M. Woodford, *Ricerche Economiche* 47 (1993) 3.
- [8] M. Levy, S. Solomon, *Int. J. Mod. Phys. C* 7 (1996) 595.
- [9] S. Galluccio, Y.-C. Zhang, *Phys. Rev. E* 54 R4516 (1996).
- [10] P. Bak, M. Paczuski, M. Shubik, *Physica A* 246 (1997) 430.
- [11] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, *Nature* 381 (1996) 767.
- [12] R.N. Mantegna, H.E. Stanley *Nature* 383 (1996) 587; R.N. Mantegna, H.E. Stanley, in: *Proc. International Conference on Pattern Formation in Fluids and Materials*, *Physica A* 239 (1997) 255–266.
- [13] P.H. Cootner (Ed.), *The Random Character of Stock Market Prices* (MIT Press, Cambridge, MA) 1964.
- [14] J.L. Doob, *Stochastic Processes* (John Wiley, New York, 1953).
- [15] P.A. Samuelson, *Industrial Management Review* 6 (1965) 41; reproduced as Ch. 198 in: P.A. Samuelson, *Collected Scientific Papers*, Vol. III (MIT Press, Cambridge, 1972).
- [16] R.N. Mantegna, in: *Proc. Applied Nonlinear Dynamics and Stochastic Systems Near the Millenium*, J.B. Kadtko, A. Bulsara (Eds.) (AIP Press, Woodbury, New York, 1997); R.N. Mantegna, in: *Proc. Workshop on Econophysics*, J. Kertesz, I. Kondor (Eds.) (Kluwer, New York, in press).
- [17] R.N. Mantegna, *Hierarchical structure in financial markets*, *Cond-Mat* 9802256.
- [18] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971).
- [19] R. Rammal, G. Toulouse, M.A. Virasoro, *Rev. Mod. Phys.* 58 (1986) 765.