Empirical properties of the variety of a financial portfolio and the single-index model

F. Lillo and R.N. Mantegna
Max-Planck Institut f"ur Physik komplexer Systeme, N"othnizer Str. 38, 01187 Dresden, Germany
and
Istituto Nazionale per la Fisica della Materia, Unità di Palermo, Viale delle Scienze, 90128 Palermo, Italy

Received 30 August 2000

Abstract. We investigate the variety of a portfolio of stocks in normal and extreme days of market activity. We show that the variety carries information about the market activity which is not present in the single-index model and we observe that the variety time evolution is not time reversal around the crash days. We obtain the theoretical relation between the square variety and the mean return of the ensemble return distribution predicted by the single-index model. The single-index model is able to mimic the average behavior of the square variety but fails in describing quantitatively the relation between the square variety and the mean return of the ensemble distribution. The difference between empirical data and theoretical description is more pronounced for large positive values of the mean return of the ensemble distribution. Other significant deviations are also observed for extreme negative values of the mean return.

PACS. 02.50.Ey Stochastic processes – 05.40.-a Fluctuation phenomena, random processes, noise, and Brownian motion – 89.90.+n Other topics in areas of applied and interdisciplinary physics

1 Introduction

Financial markets can be regarded as model complex systems [1]. They are open systems composed of many non-equivalent sub-units interacting in a nonlinear way. They are continuously monitored and a huge amount of carefully recorded financial data are now accessible for analysis and modeling of market microstructure. This allows to perform empirical analyses elucidating statistical regularities that can be used to test models of financial activities [2–4]. These tests provide information about the strengths and weaknesses of the various models pointing out the aspects that need to be improved to obtain a better model.

Stylized facts observed in financial markets mainly refer to the statistical properties of asset returns and volatility and to the degree and nature of cross-correlation between different assets traded synchronously or quasi synchronously and belonging to given portfolios. Recently, we have proposed to model the different behavior observed in the stock returns of a portfolio by considering the statistical properties (shape, moments, etc.) of the ensemble return distribution of stocks simultaneously traded in a market. Our studies [5–7] have shown that the statistical properties of the ensemble return distribution are roughly conserved in normal days of activity of the market whereas during crash and rally they change in a systematic way.

The single-index model [8,9] is not adequate to model some of these findings. Specifically, it fails in describing the statistical properties of the standard deviation (called by us variety) of the ensemble return distribution and it misses to quantitatively reproduce the symmetry breaking of the empirical return distributions observed during crash and rally days [6]. On the other hand the same model is a rather attractive because it describes pretty well several stylized facts related to the first moment of the ensemble return distribution.

In the present study we investigate the empirical behavior of the variety with respect to the theoretical predictions of the single-index model. We find that variety is only mimicked at a “zero-order” by the single-index model and significant discrepancies are observed in the statistical properties of this variable both in extreme days and in periods of normal activity of the market.

2 The ensemble return distribution in normal and extreme days

In our empirical analysis we investigate the statistical properties of the ensemble return distribution obtained for a portfolio of stocks traded in a financial market. The investigated market is the New York Stock Exchange (NYSE) during the 12-year period from January 1987 to
December 1998. This time period comprises 3032 trading days. Here we present empirical analyses of two different sets of stocks. The first is the set of all the stocks traded in the NYSE. For this statistical ensemble the number of stocks is not fixed because the total number of assets $n$ traded in the NYSE is rapidly increasing in the investigated time period and ranges from 1128 in 1987 to 2788 in 1998. The second set is the set of 1071 stocks which are continuously traded in the NYSE in the considered period. The total number of financial records processed exceeds 6 millions.

The variable investigated in our analysis is the daily price return, which is defined as

$$ R_i(t) = \frac{Y_i(t) - Y_i(t-1)}{Y_i(t-1)}, $$

where $Y_i(t)$ is the closure price of $i$th asset at day $t$ ($t = 1, 2, \ldots$). In our study we consider only the trading days and we remove the weekends and the holidays from the data set. Moreover we do not consider price returns which are in absolute values greater than 50% because some of these returns might be attributed to errors in the database and may affect in a considerable way the statistical analyses. For each set of stocks, we extract the $n$ returns of the $n$ stocks at each trading day and we consider the probability density function (pdf) of price returns. The distribution of these returns describes the general activity of the market at the selected trading day. In the periods of normal activity of the market, the central part of the distribution is conserved for a long time. In these periods the shape of the distribution is systematically non-Gaussian and approximately symmetrical [7]. During crashes and rallies the ensemble return distribution changes abruptly shape. In a previous study [6] we have shown that the ensemble return distribution becomes asymmetric in critical days. Specifically, in crash days the ensemble return distribution becomes negatively skewed whereas in rally days the distribution becomes positively skewed. The change of the symmetry properties is not the only change of the pdf observed in crash and rally days. In fact during critical days the central moments of the pdf assume values rather different from the typical ones.

To illustrate the change of the distribution in crash and rally days and in the nearby time periods we select the three biggest crashes present during the time period of our database. They correspond to – (i) the black Monday crash of 19th October 1987 when the Standard and Poor’s 500 index had a $-20.4\%$ return, (ii) the crash of 27th October 1997 when the Standard and Poor’s 500 index had a $-6.9\%$ return, and (iii) the crash occurring at 31st August 1998 when the Standard and Poor’s 500 index had a $-6.8\%$ return. Related to these crash days there are also relevant rally days. This is because the days of greatest rallies of our database occur just one or few days after crashes. In the 1987 time period, in addition to the rally days, a second crash of $-8.3\%$ of the Standard and Poor’s 500 index occurred at 26th October 1987. The statistical behavior of stock market indices during crashes has also been investigated under a different perspective in reference [10].

Figure 1 shows the contour plot of the logarithm of ensemble return distribution in a 200 trading days time interval centered at 19 October 1987 (top panel), 27 October 1997 (middle panel), and 31 August 1998 (bottom panel). In all the three panels we set the value 0 in the abscissa at the crash day. The contour plots are obtained for equidistant intervals of the logarithmic probability density. The brightest area of the contour plots corresponds to the most probable value.

**Fig. 1.** Contour plots of the logarithm of the ensemble return distribution in a 200 trading days time interval centered at 19 October 1987 (top panel), 27 October 1997 (middle panel), and 31 August 1998 (bottom panel). In all the three panels we set the value 0 in the abscissa at the crash day. The contour plots are obtained for equidistant intervals of the logarithmic probability density. The brightest area of the contour plots corresponds to the most probable value.