We investigate the time evolution of financial cross-correlation coefficients during financial crises and compare them to what is observed in periods of stability. We choose three main events, the Dot.Com Bubble, the market crisis which followed the attacks at the Twin Towers in 2001 and the recent subprime crisis. Each of them has a different nature and a different impact on the market, which we analyze by studying separately different economic sectors. As a general trend, we observe an increase of correlation during these high volatility periods and a broadening of the distributions of correlation coefficients. We then compare the spectra of the cross-correlation matrices, calculated in different periods of three years, with the distribution of eigenvalues predicted by the Random Matrix Theory. We find that these spectra are markedly perturbated during crisis periods. Finally we show how a simple stochastic model can produce similar results.

Keywords: Correlation matrix; financial markets; agents’ behaviour.
1. The importance of financial correlation matrices

In finance, correlation matrices play a central role in the estimation of risk of a given portfolio. A portfolio is a weighted choice of stocks: in other words the fraction of the total investment of an agent in each stock of the market. Intuitively, when building a portfolio, one has to consider with particular care the probability of comovements of the stock prices, since this could lead to large losses. More formally, one can refer to the basis of Markowitz’s theory of optimal portfolio\textsuperscript{1,2}. The goal is to find the portfolio (i.e. the vector of weights $w_i$) which minimizes the risk for a given expected return. If we let $r_i$ be the expected return of the $i$\textsuperscript{th} stock, then the quantity

$$R_{(w)} = \sum_{i=1}^{N} r_i w_i$$

is the total expected return for a given $w$ vector. The total risk is analogously defined on the basis of the covariance of each couple of stocks, then the total variance (risk) of the portfolio reads

$$\sigma^2_{(w)} = \sum_{i,j=1}^{N} w_i C_{ij} w_j$$

where $C_{ij}$ is the covariance matrix. Given these definitions, the problem of optimal portfolio reduces to a linear constrained minimization problem.

As already noted by Laloux et al.\textsuperscript{3}, the empirical determination of the covariance (or correlation) matrix from the time series introduces a relevant amount of noise, which makes this empirical matrix substantially random, thus raising serious doubts on the reliability of these risk estimates. Moreover, using this matrix, a meaningful topology for the market can be defined which is strictly dependent on the correlations and time evolving\textsuperscript{4}. For this reason we find crucial to look at what happens to these matrices during periods of crisis, to observe if any signal is emerging from noise, and if this signal can be somehow used to predict bursts of volatility or to better understand the structure of the markets during unstable phases. Because the covariances among different couples of stocks are not directly comparable in magnitude, we consider a matrix which is closely related to $C_{ij}$: the correlation matrix. The correlation matrix is the matrix of Pearson’s correlation coefficients, defined as

$$\rho_{ij} = \left\langle \left( x_i(t) - \langle x_i \rangle \right) \left( x_j(t) - \langle x_j \rangle \right) \right\rangle$$

where $\langle ... \rangle$ denotes the time average and $\sigma(x_i)$ is the standard deviation of the $i$\textsuperscript{th} time series. The components of this matrix are all limited to be in the interval $[-1, 1]$ and the matrix is, by definition, a Positive Definite matrix.
2. Dataset

The dataset consists of the daily closure prices for 494 stocks listed in the S&P500 index in a 26 years period, from 1/2/1985 to 5/22/2011. The series have an high level of inhomogeneity regarding their length and the effective days of trading to which they refer. As the list of the stocks has been chosen to be the S&P500 index as it was in 2010, many of them didn’t even exist in 1985. Moreover, during such a long period, it is likely that a particular stock is not quoted in some regular trading days. Anyway, in order to construct the cross-correlation matrix, we need the series to be of the same length. One possible choice is to reduce the series by considering only the subset of days in which all the stocks were actually traded. This would result in a huge cut in our data, which would end up to be restricted to just a couple of years of trading. Our choice is then to consider adequate subsets of stocks for each period in which we want to calculate the cross-correlation matrix. As a main rule we select stocks which were quoted at least six months before the first day of the window considered. From this subset of stocks we select the closing prices for the maximum subset of days in which each of them was effectively traded. Starting from these cleaned series of prices we evaluate logarithmic returns for each day $t$ and for each stock $i$ as

$$r_{i,t} = \log\left(\frac{p_{t+1}}{p_t}\right)$$

where $p_t$ is the closing price at day $t$.

2.1. *Major recent market crisis*

We choose three recent main negative events which occurred in the US stock market. These three crises are each of a different nature (two of them are endogenous, while the other is exogenous) and have different impact on the market, in intensity and width of the effects produced.

The first event is the so-called Dot.Com Bubble. This was a huge speculative bubble which mounted during the years of the sudden expansion of the Internet. In those years every company which was somewhat related to the Web was regarded by the market as a potentially explosive investment, disregarding any real economical indicator (such as profits or dividends paid). The NASDAQ index peaked on March 10, 2000 after having gained more than 100% in the precedent year. By April 2001 the index had dropped nearly 70% of its value. This was an endogenous crisis, driven mostly by euphoric expectations of the markets, and in which we expect to find strong effects on a very specific sector. We choose March 11, 2000 as the starting date of this crisis.

The second event is the market crash which followed the terroristic attacks to the Twin Towers in September 2001. This event had two main peculiarities. First it was in principle completely transversal across all the economic sectors, not being

\[^{a}\text{As the index was in 5/22/2010}\]
related to any particular economic fact. Second it was sudden, and for this reason one would expect to find very sharp transitions in the correlations distributions from the stable phase to the high volatility period. As the starting date of this crisis was chosen the first market day after September 11, 2001.

The third event is the recent Subprime Mortgages crisis. This event started as a situation which regarded only some financial institutions. After the bankrupt of Lehman Brothers, it quickly became a global economic crisis, which affected most of the market. Despite a quite fast escalation in late 2008, most analysts believe that the market was already aware of the highly risky situation in which many of the financial institutions were. For this reason we expect to find some precursors signals months before the effective start of the crisis, which we have chosen to be the date of the default of Lehman Brothers, September 11, 2008.

3. Cross-correlation coefficients distributions

The first step of our analysis consists in the study of the distribution of cross-correlation coefficients. We empirically observe their evolution in five equal windows of time of nine months each. If we mark as $t_s$ the starting day of each crisis, then we have two windows of nine months, the second of which ends at $t = t_s - 1$, then the third window in which the crisis develops and finally two more in which we observe if long term effects are present or how the markets relax to a new stable situation.

In this first analysis we consider only part of the dataset, reducing ourselves to the study of some particular economic sectors in order to highlight possible localized effects on limited sets of stocks. For example, in principle we do not expect the full market to be interested by the Dot.Com Bubble, and we expect to find dishomogeneities at least in the magnitude of the effects across different sectors. For this reason we select four groups of stocks, following the standard allocation defined by Standard&Poors\(^b\), plus a fifth, composed with the 30 stocks of the Dow Jones index. A total of 284 stocks is considered, as summarized in Table 1.

While it is clear that the financial and IT sectors are respectively the mainly interested areas for the Subprime Crisis and the Dot.Com bubble, we shall see how the energetic sector exhibited a peculiar and non-trivial behaviour during the "Twin Towers" crisis. The Dow Jones stocks are chosen to gauge the response of

### Table 1. Sectors considered

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financials</td>
<td>80</td>
</tr>
<tr>
<td>Energy</td>
<td>41</td>
</tr>
<tr>
<td>Industrials</td>
<td>58</td>
</tr>
<tr>
<td>IT</td>
<td>75</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>30</td>
</tr>
</tbody>
</table>

\(^b\)available at www.standardandpoors.com
the market as a whole to these extreme events. As a general trend we observe an increase in the mean value of the correlation coefficients in all the events studied, as well as a broadening of the distributions. The market has a clear tendency towards coordination and starts to look like all the agents are acting and thinking as one. We proceed in the analysis of each crisis.

3.1. Dot.Com Bubble

The Dot.Com Bubble was a very localized event, and this reflects sharply on the distributions of cross-correlation coefficients. All the sectors remained substantially unaffected with the exception of the IT sector. Fig. 1 shows the evolution of the distribution for the IT sector and the histograms of the variations of the single coefficients and, as an example, Fig. 2 shows the same plots for the Industrial sector. While in the case of the Industrial sector the distribution remains substantially unaltered, the IT sector histogram is clearly broadened and shifted towards larger correlations.

3.2. Twin Tower attacks

Differently from the Dot.Com Bubble, a market crisis, which spanned all the major economic sectors, developed as a consequence of the attacks at the World Trade Center in 2001. This is clearly reflected by the cross-correlation coefficients between the 30 Dow Jones stocks. Fig. 3 shows the time evolution and points out how after 18 months the market is starting to relax to a less correlated situation.

Fig. 1. (Left) The evolution of the distribution of the cross-correlation coefficients within the IT sector in five 9-months windows. The middle window starts in the first day of the "Dot.Com" crisis (3/11/2000). (Right) Distribution of the variations of each cross-correlation coefficient between the windows considered. Red and green areas show respectively correlations decreasing or increasing of more than 0.05 with the exception of the last time window, when the "Twin Towers crisis" begins.
Time Evolution of Financial Cross-Correlation Coefficients Across Market Crisis

A quite peculiar situation is observed within the Energy sector (Fig. 4). After 9 months from the beginning of the crisis, a sharp double peak is clearly visible. A closer analysis shows that there was a set of 9 stocks which was giving little or no contribution to the right peak (see tab. 2), and that those were all energy companies with activities mainly in the territory of United States. For this reason is quite easy to understand how these companies should have been less worried by a probable war in middle east, or even get an advantage from it, in complete opposition to what one might expect from, say, a petroleum company strongly linked to middle east’s resources. The market seems to have correctly differentiated between these two classes of companies.
3.3. Subprime Crisis

Differently from the cases listed so far, in the case of Subprime Crisis the financial sector displays an high level of coordination already 18 months before the actual dramatic event of the bankrupt of Lehman Brothers (fig. 5). The huge wave of volatility that followed could either have been already foreseen by the market, which was waiting in a highly correlated state, or have been caused and amplified by the strong correlations.

4. Spectra of cross-correlation matrices

In this section we analyze the spectral properties of cross-correlation matrices over longer periods of time. As already mentioned, we have chosen to study the evolution

<table>
<thead>
<tr>
<th>Stock</th>
<th>Contribution to left peak</th>
<th>Contribution to right peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>MEE</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>MUR</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>TSO</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>WMB</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>NU</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>SWN</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>SII</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>RRC</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>CHK</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>APA</td>
<td>9</td>
<td>26</td>
</tr>
</tbody>
</table>
Fig. 5. (Left) The evolution of the distribution of the cross-correlation coefficients within the financial sector in five 9-months windows. The middle window starts in the first day of the "Sub-prime" crisis (9/11/2008). (Right) Distribution of the variations of each cross-correlation coefficient between the windows considered. Red and green areas show respectively correlations decreasing or increasing of more than 0.05

of a set of nearly 500 stocks in windows of three years, to observe what modifications occur during periods of financial instability. The number of stocks considered in each window is variable, due to the fact that some stocks were not yet traded in past times.

4.1. Random Matrix Theory

The cross-correlation matrices that we are considering are, of course, determined on the basis of finite time series. This introduces a certain amount of noise, which arises in the empirical determination of the covariances. For this reason one can expect the correlation matrix to be, at least to some extent, a random matrix. The spectral properties of the correlation matrix of \( N \) series of \( L \) random numbers, all independent and identically distributed, are known in the limit \( N \to \infty \) and \( L \to \infty \) with \( Q = L/N \) fixed and greater than 1. More precisely the eigenvalues are distributed according to the Marchenko-Pastur distribution which, if the series are rescaled to have unit variance, is given by

\[
\rho(\lambda) = \frac{Q}{2\pi} \sqrt{\frac{(\lambda_+ - \lambda)(\lambda - \lambda_-)}{\lambda}}
\]  

(5)

where

\[
\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}
\]

(6)

This distribution possesses two sharp edges, which are limiting the eigenvalues to be enclosed between two extremal, positive values. For a more complete overview of the applications of Random Matrix Theory in financial problems see Ref. 5.
4.2. Spectral properties of financial cross-correlation matrices

The spectral properties of financial cross-correlation matrices show a remarkable similarity with the predictions of the Random Matrix Theory, thus sustaining the hypothesis that their structure is dominated by the noise caused by the finiteness of the time series. This approach is anyway helpful in determining what aspects of the spectra of these matrices are real signals from the markets. As already noted by Plerou et al. there are some eigenvalues that lay way beyond the maximum predicted value, with one of them being much larger than the others. These eigenvalues correspond, respectively, to modes tracking the synchronism inside particular sectors or the movement of the full market. As shown in Fig. 6 in periods of stability the matrix is very close to be random, except for the few eigenvalues mentioned. In periods of crisis the spectrum changes noticeably, and the hypothesis of a quasi-random matrix does not hold anymore.

4.3. A simple stochastic model

We proceed now in describing a minimal stochastic model, which is able to produce time series with a correlation matrix with spectral properties very similar to those of the real matrices. The basic idea is that during periods of instability traders start to attribute an increasing importance to the collective movements of the market. This is a self reinforcing mechanism, being a strategy which validates itself: an always larger number of agents will be acting in synchrony, producing an increasingly stronger signal. If this mechanism is applied in a not completely blind way, traders should also be able to differentiate between different stocks, and to consider a different level of reaction to the global stimulus for each of them. We have seen

![Fig. 6.](image-url)
in the previous sections that this is indeed the case. Moreover, some efforts have already been made to show how the effective number of agents playing different strategies on the market is a key feature for some agent based models to exhibit volatility bursts\textsuperscript{7,8,9}. The model consists in a simple procedure:

(1) A random series $\mu$ of length $L$ is extracted out of a gaussian distribution with zero mean and unit variance.

(2) For each of the $N$ time series $\alpha_j$ ($j = 1, 2, ..., N$) to be generated, a random number $p_j \in [P_m, P_M]$ is chosen out of a uniform distribution.

(3) The $i$–th element of each series $\alpha_j$ is chosen to be equal to $\mu_i$ with probability $p_j$, or is extracted from a standard gaussian distribution with probability $1 - p_j$.

In this way we build artificial time series with a common correlation structure. The only adjustable parameters are $P_M$ and $P_m$, which somehow define the range of variability of the correlations of the $\alpha_j$ with $\mu$ and so among themselves. Fig. 7 shows how this simple method can produce correlation matrices quite similar to the ones observed in real markets in periods of crisis. Also the spectra display a remarkable agreement, not only with the peak present in the area of low eigenvalues, but also in the presence and magnitude of the largest eigenvalue which is found to be of order $N \langle \rho \rangle$, with $\langle \rho \rangle$ being the average correlation coefficient across all the market (i.e. the average over all the elements of the upper or lower triangle of the correlation matrix), both in the real cases and in the model. Such a close agreement with this very simple model seems to suggest that, in periods of high volatility, the market is effectively behaving as if the values of all stocks were strongly linked to a single source of information.

![Fig. 7. On first row a comparison between a purely random correlation matrix, two real financial correlation matrix, and a matrix generated with the proposed model with $P_m = 0.6$ and $P_M = 0.8$. On the second row the corresponding eigenvalues distribution are shown with the Marchenko-Pastur distributions with appropriate parameters. The largest eigenvalue is not shown.](image-url)
5. Conclusions

We have found a regular trend of growth of the correlation coefficients in periods of crisis. Anyway, even when looking closely into economic sectors, the correlations may have a wide range of variability, clustering in some cases, with a strong relation with real economical conditions of firms. In some cases strong correlations are observed even before the effective explosion of volatility, when the information on a possible instability of the market is sufficiently spread between investors (e.g. in the case of the Subprime Crisis).

When observing the spectral properties of the correlation matrices, interestingly we have found that they become far from random in crisis periods. This is a signal which needs to be interpreted. We have then built a minimal model, in which the key feature is the response of the market to a single source of information. When the strenght of the reaction to the information is on average strong enough, and sufficiently heterogeneous between different stocks, this model reproduces the observed spectra with surprising agreement, thus indicating that the few ingredients present in the model are probably a good representation of what is really happening.

Acknowledgments

We acknowledge partial support from the PNR national project CRISIS-lab and EU FET Open project FOC nr. 255987

References