Financial crisis forecast : leading eigenvalues and correlations as early-warnings

Abdoul Aziz Berrada, Baptiste Goumain, Hugo Hamon

University Paris I - Panthéon Sorbonne

Abstract

In this paper, we analyze both leading eigenvalues of rolling window correlation matrices and heatmaps before, during and after 2 major crises : the Dot-com bubble and the 2008 crisis. Leading eigenvalues point out that the market changes from a mode to another, that the correlation goes into one unique direction. We show that in times of crisis, returns are highly correlated, and we try to use these results to analyze what is happening on the financial market since 2017. Finally, we explore the results to determine if these two indicators can be considered as consistent early warnings.

Keywords: Financial crises, Early warning, Correlations, Eigenvalues, Random Matrix Theory, Marchenko Pastur, Heatmaps.

Introduction

Crises of over-indebtedness are set up when everything is going well (controlled inflation, low unemployment, economic growth) and economic agents (companies, households, etc. . .) take advantage of the growth and low-interest rates to sometimes borrow beyond reason. Yet, when interest rates begin to rise as a result of monetary tightening, debt that seemed sustainable given the previously moderate interest rates, become unbearable and turn into over-indebtedness. This has been called since Hyman Minsky (1982), the "paradox of tranquillity". However, there are not only crises of over-indebtedness. Indeed, we can also evoke the recent Covid-19 crisis or the internet bubble burst in the early 2000s. A financial crisis corresponds to a massive fall in prices and is therefore not necessarily linked to indebtedness, but agents' loss of confidence in institutions. We thus wonder if these moments of "tranquillity" highlighted by Minsky can be spotted and marked as such. Based on different studies and theories such as the Modern Portfolio Theory (MPT), it can be assumed that the correlation of returns between different large companies can be the indicator that we are looking for. According to the MPT, investors can minimize their risk of investment loss by seeking to reduce the correlation between the return of selected securities.

As previously stated, many studies show that assets and financial markets are highly correlated before, during and after a financial crisis. Understanding these correlations that occur before, during and after financial crises can lead to forecast and predict them. In this paper, we aim to highlight the correlations during two major crises : the Dot-com bubble and the subprimes crisis. We then use our results to analyze what is happening on the financial market since 2017 in order to give some ideas of further studies to use eigenvalues and correlation as a consistent early-warning indicator of financial crises [1]. Some papers analyze the correlation between different stock markets [2], whereas others tend to study correlations between assets [3] and seek to detect critical events (such as economic crises) before they become fully effective. They conclude that leading eigenvalues and correlations are good indicators to predict regime changes from a non-crisis period to a crisis period and vice versa. However, they conduct their study over different prediction horizons: one day, one week, and one month at the longest, and conclude that performance deteriorates only modestly over longer prediction horizons. We will therefore try to see if we find the same results but over a longer time horizon since it is not very useful to predict a crisis one day earlier. In Didier Sornette's paper [4], one of the conclusions that emerge is that extreme events (such as crises) occur more often than one might expect. Crises do not necessarily occur under the effect of an external shock, but under the effect of smooth changes in certain variables. Therefore it is crucial to learn how to diagnose these symptoms because they are observable. We can also refer to the article by Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, and H. Eugene Stanley [5] which uses the theory of random matrices to analyze correlation matrices, which we will also do.

We chose to study companies assets correlation in times of crises, defined by a development of non-usual patterns on the market through their daily data from 1989 to 2020.

In this paper, we use the Random Matrix Theory and especially the superior eigenvalue of each correlation matrix we extract. We use the leading eigenvalue as an instability indicator [3]. It has been proven empirically that before a market crashes, these eigenvalues drastically increase in a short amount of time, leading the market from a state to another one.

Furthermore, we draw the heatmaps of correlation matrices to explore the evolution of the assets correlation in times of crises. We start 2 years before the crises, and try to highlight the growing correlation until, during, and after the crises. We first draw these heatmaps over the period of 6 months and then again over 3 months, to see if there is any difference, trying to be more precise.

Section 1 describes our dataset, the companies we chose, and how we treated the data. Section 2 introduces our methodology (eigenvalues, rolling-window matrices) and important theory points that are the foundation for the results we obtained in Section 3.

1. Data

We extracted asset prices of 30 US companies, ranged from different market capitalization. All of them are classified in the S&P. We took 10 companies of each the S&P600 Small Caps, S&P400 Mid Caps and S&P500 Large Caps which separate small, mid and large market capitalizations.
Here is the list of the companies we've taken :

S&P 600 Small Caps: 3D Systems Corporation (DDD), Glacier Bancorp, Inc. (GBCI),Getty Realty Corp. (GTY), Genesco Inc. (GCO), G-III Apparel Group, Ltd. (GIII), Frontier Communications Corporation (FTRCQ), Federal Signal Corporation (FSS), Comtech Telecommunications Corp. (CMTL), CONMED Corporation (CNMD), Cooper Tire & Rubber Company (CTB).

S&P 400 Mid Caps: New Jersey Resources Corporation (NJR), Nordson Corporation (NDSN), Oceaneering International, Inc. (OII), NVR, Inc. (NVR), NewMarket Corporation (NEU), The New York Times Company (NYT), Donaldson Company, Inc. (DCI), Dillard's, Inc. (DDS), Deluxe Corporation (DLX), MDU Resources Group, Inc. (MDU).

S&P 500 Large Caps: Apple Inc. (AAPL), Adobe Inc. (ADBE), The Walt Disney Company (DIS), Intel Corporation (INTC), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), Microsoft Corporation (MSFT), The Procter & Gamble Company (PG), UnitedHealth Group Incorporated

(UNH), Verizon Communications Inc. (VZ).

Data goes from 1989 to end of 2020. In the dataset, we have the daily prices of each of these companies assets. We can thus cover the Dot-com bubble and the 2008 crisis. As it is common, we worked on the returns of each asset, not on their price as is it initially presented in our data. We then applied the following transformation to each of the company assets :

$$r_{A,t} = log(\frac{p_{A,t}}{p_{A,t-1}})$$

where $r_{A,t}$ is the return of the company "A" at the date"t". $p_{A,t}$ is the price of its asset at the date "t" and $p_{A,t-1}$ is the price of the asset the day before, as t is a day. We then normalized the returns, so that they have all the same mean, 0, and the same standard-error, 1. De facto, we won't have to treat the heteroskedasticity problem.

2. Methodology

To obtain the leading eigenvalues, we had to build rolling window correlation matrices, to compute them over time. For each crisis, we began the study two years before the crisis. Thus, for the Dot-com bubble, which is assumed to have burst in 2000, even if there is no real consensus, we started to build rolling window matrices on 1998, until 2004. Then, we built correlation matrices on a window of 6 months, and we obtain 12 heatmaps. Then we get 24 heatmaps with the second analysis, with the window of 3 months. The heatmaps allow us to watch and analyze the evolution over time of the correlations between assets. For the 2008 crisis, we started our analysis on August 2006 until the end of 2009 and we built the same matrices. Finally, we analyze the evolution of leading eigenvalues and correlation of returns since January 2017 until December 2020.

2.1. Spearman correlation coefficient

In our work, we decided to determine the correlation between the returns using the Spearman correlation coefficient. Insofar, as a correlation coefficient calculates to what extent two variables tend to change together, it is logical that we use the Spearman correlation coefficient which is called the "rank-order correlation" coefficient. In fact it calculates the correlation between 2 variables which are not linked by a linear relation. The coefficient

describes the importance and direction of the relationship between returns. It is based on their ranked values and therefore their ranks rather than on raw data. The formula for the Spearman coefficient r_s is as follows:
For a sample of size n, the n raw scores X_i and Y_i are converted to ranks rg_{X_i} and rg_{Y_i} , so :

$$r_s = \rho_{rg_X, rg_Y} = \frac{\mathrm{cov}(\mathrm{rg}_X, \mathrm{rg}_Y)}{\sigma_{rg_X} \sigma_{rg_Y}}$$

where ρ is the usual Pearson correlation coefficient applied to rank variables, $\mathrm{cov}(\mathrm{rg}_X, \mathrm{rg}_Y)$ is the covariance between the rank variables, σ_{rg_X} and σ_{rg_Y} the standard deviations and the 2 rank variables.
If all n ranks are distinct integers : $r_s = 1 - \frac{\sum d_i^2}{n(n^2-1)}$ where $d_i = rg(X_i) - rg(Y_i)$ is the difference between the two ranks of each observation. Our correlation matrix will then be formed by the set of correlation coefficients between returns 2 to 2.

2.2. Random Matrix Theory

Nowadays, it is more and more frequent to work with very large volumes of data and this in several fields, whether in Physics, Statistics, Finance as in our present case, etc... There is a great stake in the transcription of the information contained in this data. However, it quickly becomes complicated to handle and even visualize this data in more than 3 dimensions (note that we have 30 variables in our database). Indeed to estimate X a TxP matrix with T the number of observations, P the number of variables and P <<< T, we are tempted to determine the correlations between the variables and then the statistical inference allows us to have very low estimation errors. On the other hand, in the current era of Big Data, P and T can increase and P can in certain cases be in the same orders of magnitude as T and thus distort the asymptotic results and call into question all the conclusions of classical statistics. The solution to avoid these problems would then be to compress this data into a smaller set. We then appeal to the Random Matrix Theory which uses proposes to show that the spectrum of the large correlation matrices becomes stabilized when its dimension tends towards infinity. Thus the eigenvectors and eigenvalues of these large correlation matrices make it possible to reduce the dimensions of the data and at the same time to conserve as much information as possible.

In our study we will mainly use the contribution of Marchenko and Pastur (1967).

2.3. Marchenko-Pastur Distribution

In this part we will focus on the contribution of Marchenko-Pastur in the RMT. As we have just said in the previous part, Marchenko-Pastur sought to describe the behavior of the spectra of the large correlation matrices and thus to derive global properties.

Let X be a TxP random matrix.

We have $Y_t = \frac{X'X}{T}$ the sample correlation matrix and $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_K$ the eigenvalues of Y_t, the eigenvalues are random variables.

Theorem : Suppose that $\{x_{ij}\}$ are iid. random variables with mean 0 and variance σ^2.
Also assume that P,T $\rightarrow+\infty$ and P/T $\rightarrow$ c $\in$]0,+∞[.

Then the Marchenko-Pastur law has the following density function :

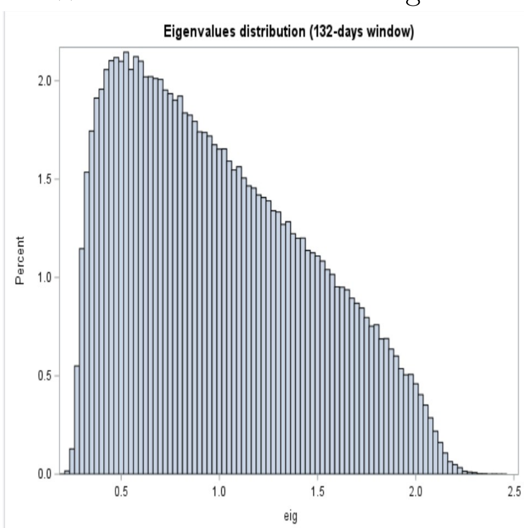
$$f_{MP}(x) = \begin{cases} \frac{1}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{cx} & \text{if } x \in [\lambda_-, \lambda_+] \\ 0 & \text{otherwise} \end{cases}$$

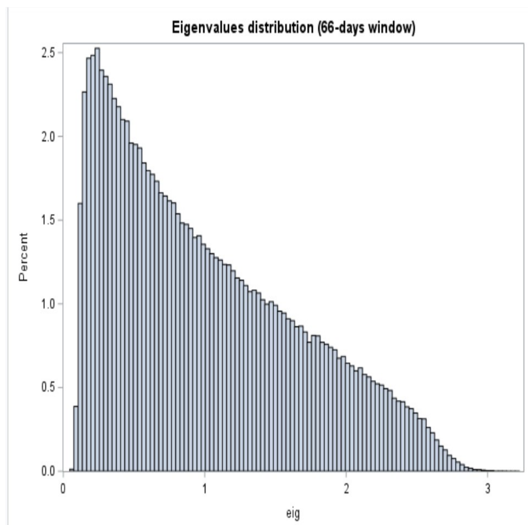
with $\lambda_\pm = \sigma^2(1 \pm \sqrt{c})^2$. Here, the constant c is the dimension to sample size ratio index and σ^2 is the scale parameter. If $\sigma^2 = 1$, the Marchenko-Pastur law is said to be the standard Marchenko-Pastur law.

For our study, we have to determine if our leading eigenvalues are significant or not. The leading eigenvalues of the rolling window correlation matrices are distributed with the Marchenko Pastur law. For the rest of this paper, to make it easier, we are going to call these eigenvalues λ_{max}. As we already mentioned it, the λ_{max}, are used as an indicator or instability. We then expect them to sharply rise in time of crisis. If they are not significant, they give no information and are therefore useless. To determine if the leading eigenvalues are significant, we simulated two Marchenko-Pastur distributions. One for the analysis on 6 months and the other for the 3 months one. For the analysis on 6 months, the size of the window is 132 days (week-ends are not included). To know the distribution of the law of Marchenko-Pastur and the associated quantiles, we simulated 10 000 normal random matrices of 132 observations and 30 variables, with a mean of zero and standard deviance of 1. For the analyse on 3 months, we did the same protocol, only the window is 66 days.

Thus we determined the correlation matrix by Spearman, the eigenvalues and then plotted the distribution of the eigenvalues.

We then have the following Marchenko-Pastur distributions:





On the left is the distribution for the study on 6 months and on the right is the one for 3 months. With these distributions, we determined the quantiles. As soon as we have eigenvalues that exceed the associated 100% quantile for the correlation matrix between returns, we consider that these eigenvalues contain information, that is to say there are significant at the 0% threshold, but if not then we'll consider that they also have noise. Here are given the quantiles of both distributions (on the same order as above) :

Quantiles (132-day window)	
Level	Quantile
100% Max	2.465632
99%	2.060007
95%	1.861752
90%	1.697906
75% Q3	1.346934
50% Median	0.929834
25% Q1	0.603549
5%	0.361368
1%	0.296207
0% Min	0.202215

Quantiles (66-day window)	
Level	Quantile
100% Max	3.2023372
99%	2.5998660
95%	2.2828071
90%	2.0172642
75% Q3	1.468166
50% Median	0.8564162
25% Q1	0.4309407
5%	0.1743425
1%	0.1192339
0% Min	0.0510303

Furthermore, to determine if our simulation truly corresponds to a Marchenko-Pastur distribution, we use the following formulas :

$$\lambda_{max} = (1 + \sqrt{c})^2$$
$$\lambda_{min} = (1 - \sqrt{c})^2$$

It gives us the 100% theorical Max quantile and the 0% theorical Min quantile of the theoretical distribution. When we apply the formulas, we then expect to find results approximately between 2.06 and 2.46 for the 6 months window and between 2.6 and 3.2 for the other one for λ_{max}. We expect to find results close to 0.296 and 0.11 for λ_{min}.

For the 6 months window, we find $(1+\sqrt{\frac{30}{132}})^2 = 2.18$ and for the 3 months window, the formula displays a result of $(1+\sqrt{\frac{30}{66}})^2 = 2.8$ for λ_{max}. For λ_{min}, we get $(1-\sqrt{\frac{30}{66}})^2 = 0.27$ and $(1-\sqrt{\frac{30}{66}})^2 = 0.1$. Thus, our simulations seem consistent with the Marchenko-Pastur distribution and we can use the 100% quantile to determine if the λ_{max} are significant or not.

3. Results

3.1. Dot-Com bubble (March 2000)

We first analyze the leading eigenvalues and heatmaps on 6 months-rolling-window correlation matrices.

Figure 1 displays the evolution of λ_{max} for the Dot-Com bubble and the evolution of the S&P500 during the period. We chose to display only the S&P500 index in all our following graphics for 2 main reasons. The first is that we don't have data about the S&P600 for the period we study since it is a more recent index. The second is that the S&P500 corresponds to almost nearly 50% of the worldwide financial market capitalization. We then consider that it is the most important and relevant index we can study, and that almost every index, including the S&P400 and 600, follow the same trends. The maximum eigenvalues inform us about the direction towards which the correlation matrix is projected, according to the studies mentioned and implied earlier, and our basic hypothesis. Again, they are only interpretable if they are significant and if their rise is actually linked to an actual event on the market. Moreover, we expect the correlations to rise at the same time as the leading eigenvalues.

We dated the bubble burst in March 2000, but there is no clear consensus about the date. The large crisis lasted from 1998 to 2002 at least. λ_{max} are the leading eigenvalues of the 6 months-rolling-window correlation matrices.

First, we notice that all the λ_{max} are significant (> 2.46, as defined earlier). We notice a first rise of the eigenvalues at the half of of 1998, 2 years before the bubble burst. It appears that at the end of 1998 the S&P500 crashed before a peak around 1200 points, just after the increase of the eigenvalues. The second relevant result is that there is a gradual rise of the λ_{max} beginning on April 2000, when the index was a high level but a few months later started to drop gradually until July 2002 when the leading eigenvalues exploded a little bit after the S&P500 lost suddenly hundreds of points. The λ_{max} seem to decrease when the index regains its value, from the beginning of 2003.

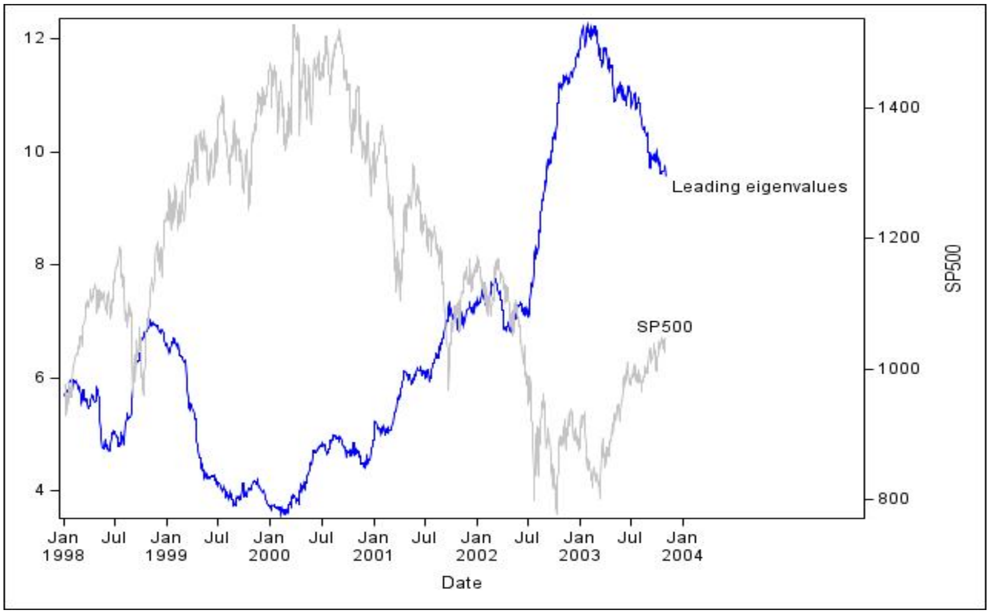


Figure 1: Leading eigenvalues on a 6-month rolling window (Dot-Com bubble)

Figure 2 illustrates the heatmaps of the correlation matrices every 6 months, 2 years before and 1 year after March 2000. Gray squares mean there is no (or little) correlation between returns.

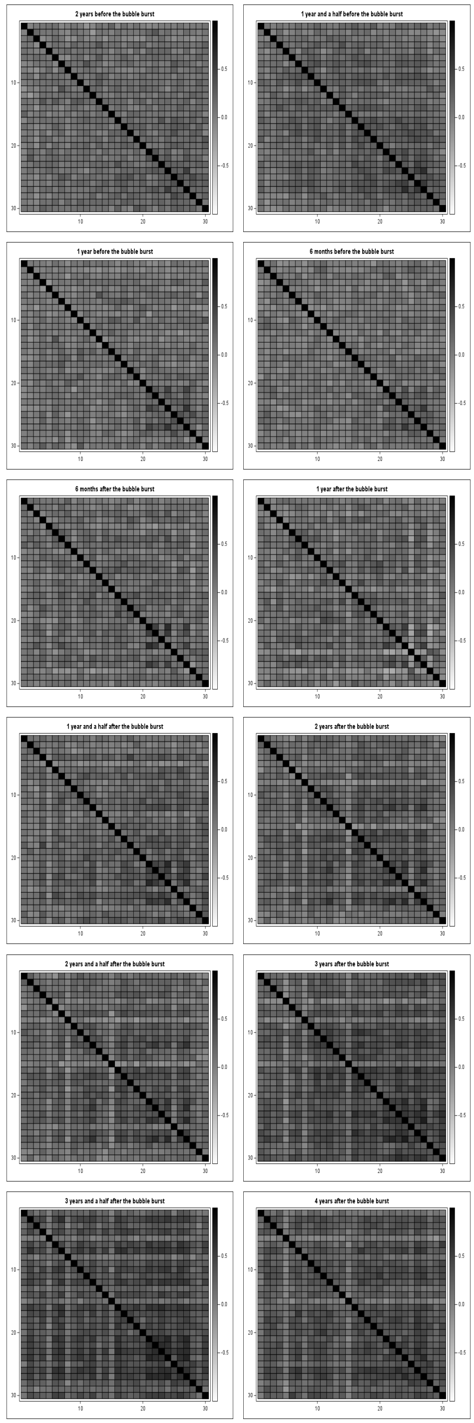


Figure 2: Heatmaps of correlation matrices for the Dot-Com bubble (every 6 months, starting 2 years before March 2000)

We can notice that there is almost no significant evolution between the different correlation matrices. We can only observe a slight correlation between the large-cap assets, which starts to grow 2 years after the bubble burst (4th line). The correlation between returns seems to reach its highest level in 2003 and 2004, when the index is already at its bottom level, 3 years afters the bubble burst. The bubble burst first happened on the NASDAQ index, on March 2000, so we could explain the non-correlation by the fact that the crisis took more time to be "transmitted" to the S&P. This non-correlation between the returns of these companies before and during the crisis (from 1998 to 2002) and the non "transmission" of the correlation from the small to the mid to the large-cap could be be explained by the fact that new technologies have evolved a lot since then, so the weight of the companies have also changed over decades.

Then, let's study the 3 months-rolling-window correlation matrices. We lead the same analyze. Figure 3 shows the evolution of λ_{max}. λ_{max} are the leading eigenvalues of the 3 months-rolling-window correlation matrices.

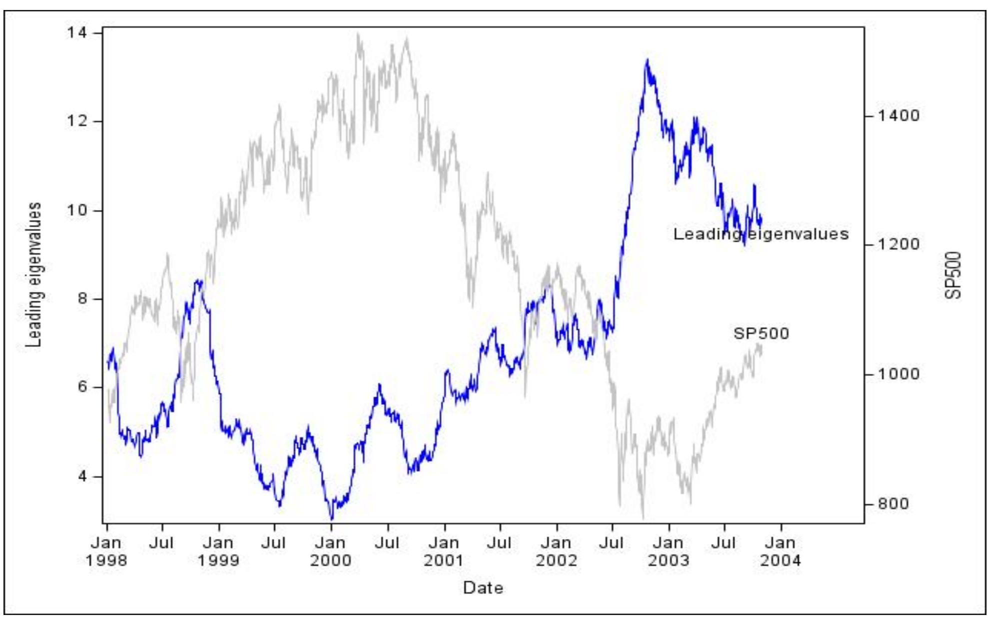


Figure 3: Leading eigenvalues on a 3-month rolling window (Dot-Com bubble)

The interpretation of these 3-month window λ_{max} is almost identical to that of the 6-month one (Figure 1), the only difference being that the volatility is higher here because we have more precise results and that an axial symmetry emerges between the two indices. We notice that between 2000 and 2002, the S&P500 is highly volatile, so the leading eigenvalues are, but there is no big event such as the ones we already mentioned, only that when the leading eigenvalues gradually increase, the index gradually decrease, meaning the is the time of the crisis.

Figure 4 shows the heatmaps of the correlation matrices every 3 months, starting 2 years before March 2000.

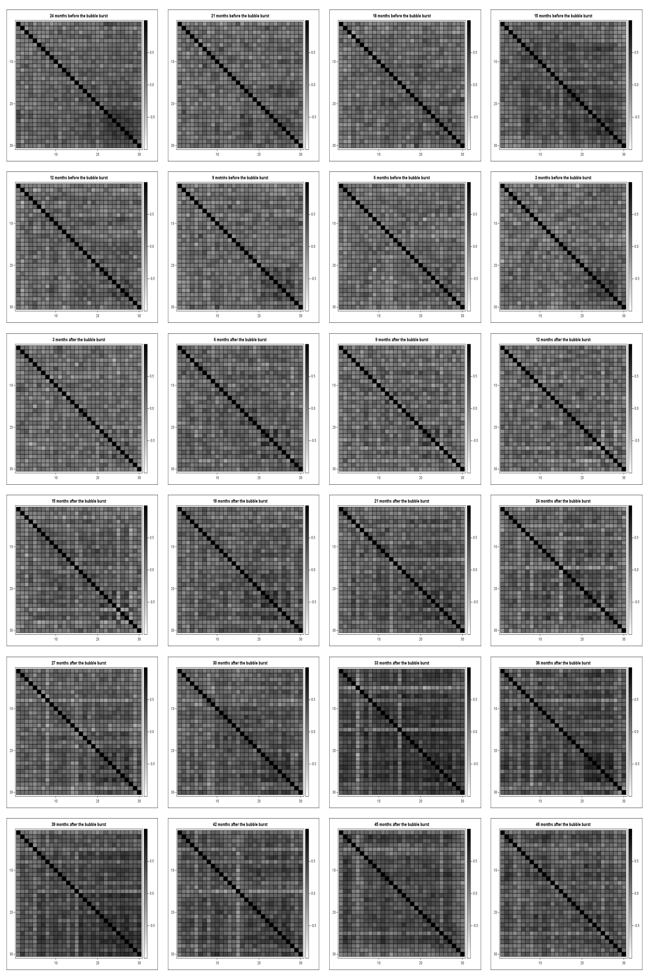


Figure 4: Heatmaps of correlation matrices for the Dot-Com bubble (every 3 months, starting 24 months before March 2000)

We discern a small correlation for big companies 24 months before March 2000. Then, there is a higher correlation for every company in our data 15 months before the bubble burst, that is to say when the S&P500 crashed at the end of 1998. We observe no or very small correlation until 33 months after March 2000, that is to say when the S&P500 reached its lowest level at the beginning of 2003 and when the leading eigenvalues were at their highest level. The correlation between assets remain high for 6 months and starts to decrease, when the index regain its value and leading eigenvalues decrease too.

In the case of the Dot-Com bubble, because of our companies that are ranked in the S&P500 and because the crisis first occured on the NASDAQ index, the leading eigenvalues and correlations are not a consistent early-warning. Indeed, the leading eigenvalues actually start to rise and explode after the S&P500 started to crash. But the index abruptly crashed years after the bubble burst on March 2000. The leading eigenvalues started to gradually rise when the index started to gradually decrease, but the correlation did not follow that trend and suddenly rose only when the crash occured. Thus, in the case of this crisis, we don't consider the leading eigenvalues as a consistent early-warning.

3.2. 2008 crisis

Figure 5 displays the evolution of λ_{max} from 2006 to 2009. We dated the beginning of the crisis in July 2007, which corresponds to the first step of the crisis, when the S&P500 started to gradually lose points before a crash, as displayed of Figure 5 : that is the beginning of the crisis. The second step is in September-October 2008, when the bubble burst. Again, there is no clear consensus about the date of the beginning of the financial crisis.

First of all, we notice that the leading eigenvalues are even more significant than the previous crisis. We observe that they start to sharply increase around July-August 2008, as expected. Indeed, the S&P500 sharply drop a few weeks later. The leading eigenvalues seem to reach their highest level at the same time the index reaches its lowest one. However, that is the second rise. Indeed, the λ_{max} first increased in 2007 as displayed on the graph. The rise of leading eigenvalues is far more gradual that the explosion in 2008, and starts more than one year before the first step of the crisis in mid-2007. At the same time, the index slowly lost its value. It is only in September 2008, a few weeks after the leading eigenvalues started to explode, that the

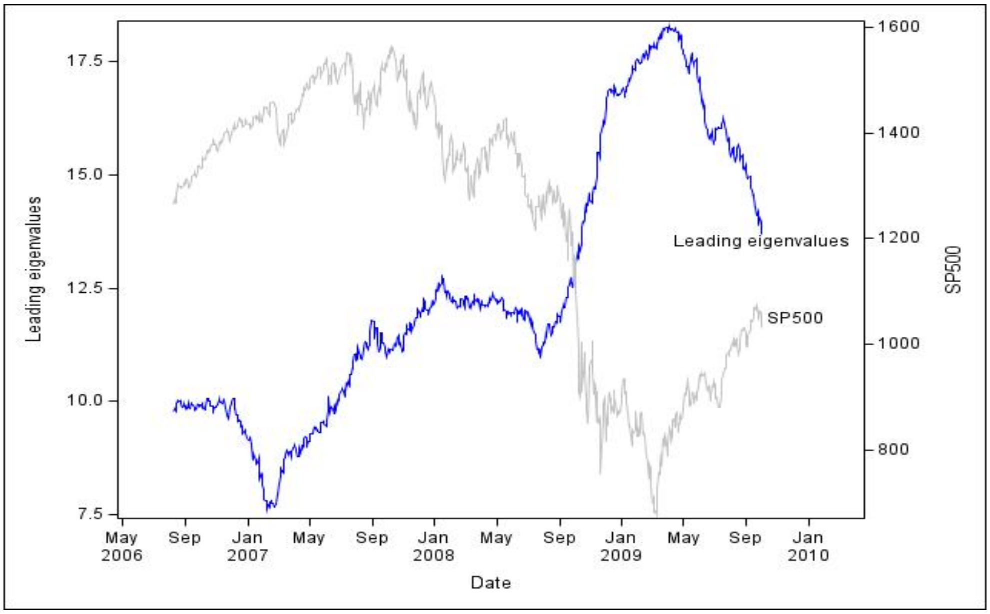


Figure 5: Leading eigenvalues on a 6-month rolling window (2008 crisis)

S&P500 lost more than 500 points in a few months, after the bubble burst in September 2008.

Figure 6 exhibits the heatmaps of the correlation matrices every 6 months, 2 years before and 1 year after October 2008. High correlations are displayed for this crisis. Indeed, there is no correlation between returns 2 years before the crisis, but we observe a rising correlation, especially for the biggest companies (from 20 to 30 on the x axis) as soon as 1 year and a half before October 2008, and even more 6 months before it. Six months after October 2008, the correlation matrix is all in black, meaning the asset returns have a degree of correlation almost equal to 1. Returns are therefore highly positively correlated. And this is verified for each companies in our data. One year after the crisis, some correlation still remains.

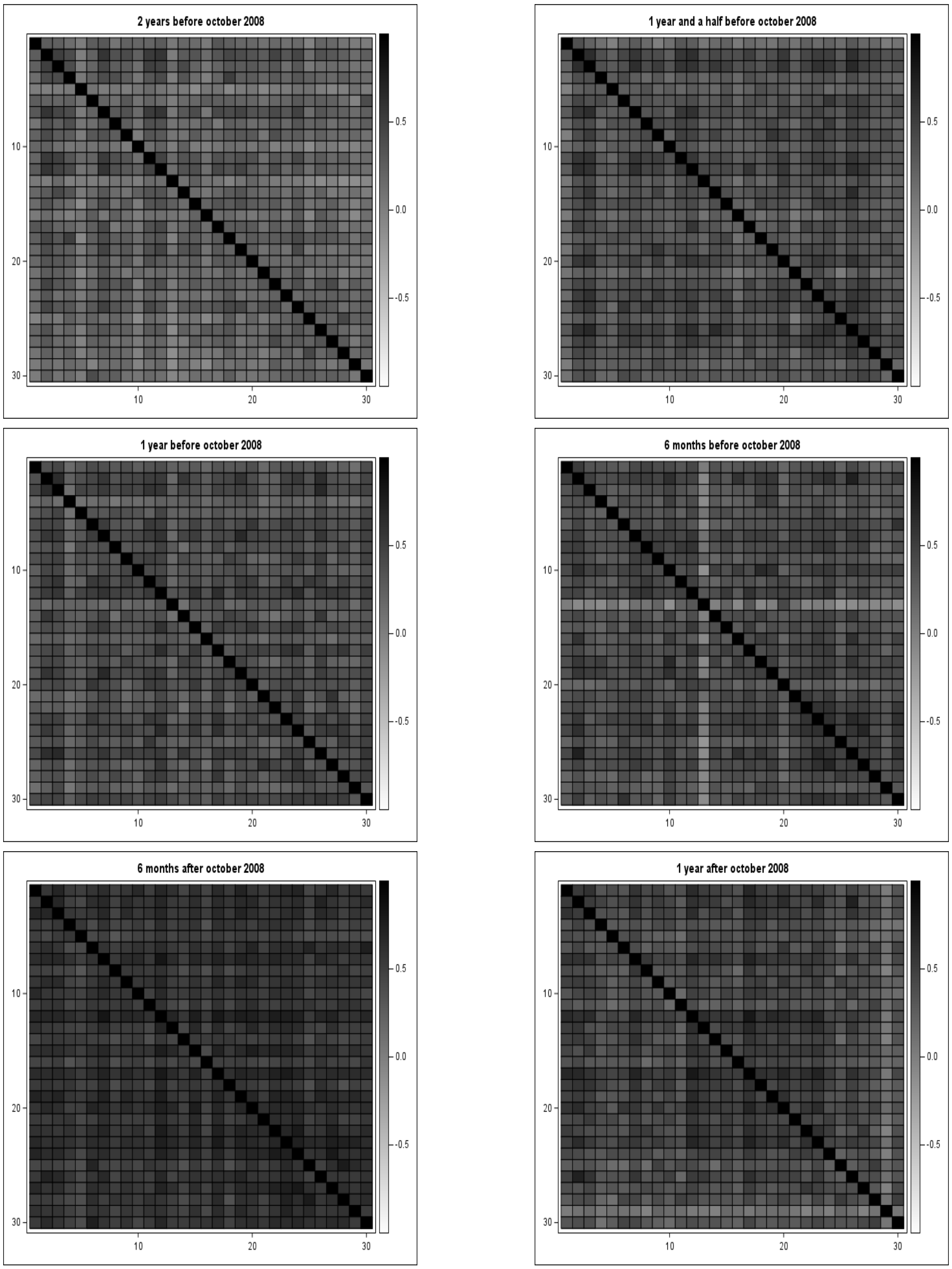


Figure 6: Heatmaps of correlation matrices for the 2008 crisis (every 6 months, starting 2 years before October 2008)

Again, we conduct the same analysis on a 3 months window. The evolution of λ_{max}, the leading eigenvalues, are displayed on Figure 7. They are highly volatile, but we draw the same results as before : there are 2 relevant rises of the leading eigenvalues and both are followed by market events, especially the second one in 2008. Moreover, we notice here that the leading eigenvalues reach their highest level a few months before the index reaches its lowest level, and that they remain high for a few months.

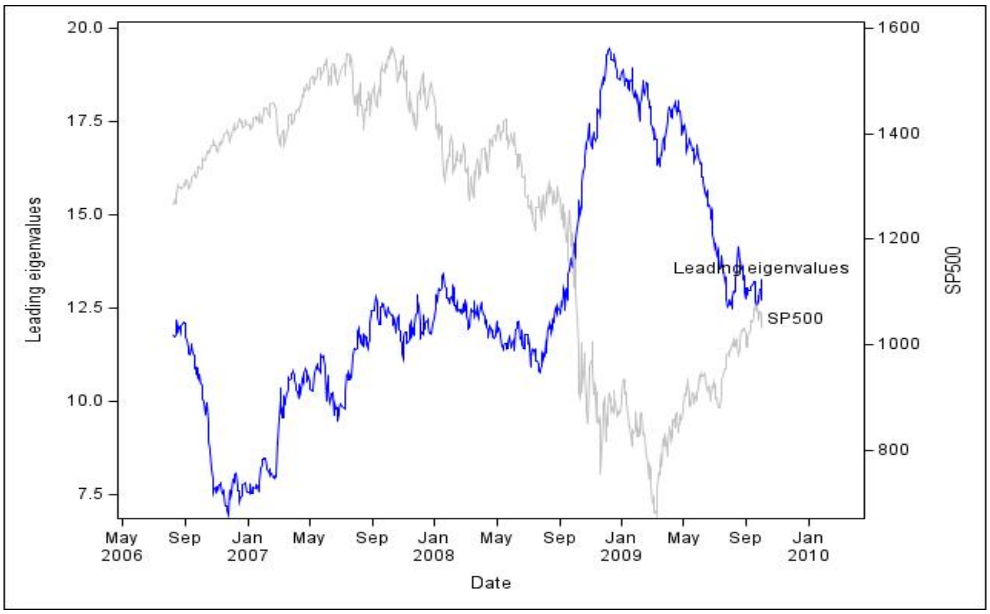


Figure 7: Leading eigenvalues on a 3-month rolling window (2008 crisis)

Figure 8 illustrates the heatmaps of the correlation matrices every 3 months, 2 years before and 1 year after October 2008. Here, we can already observe more precise results. Correlation between returns starts to occur and increase 18 months before the crisis, but still remains low. The correlation is almost at its maximum 3 months after October 2008 (that is to say from October to December), and remains very high after 6 months. It starts to decrease one year after the crisis, when the S&P500 already started to grow higher.

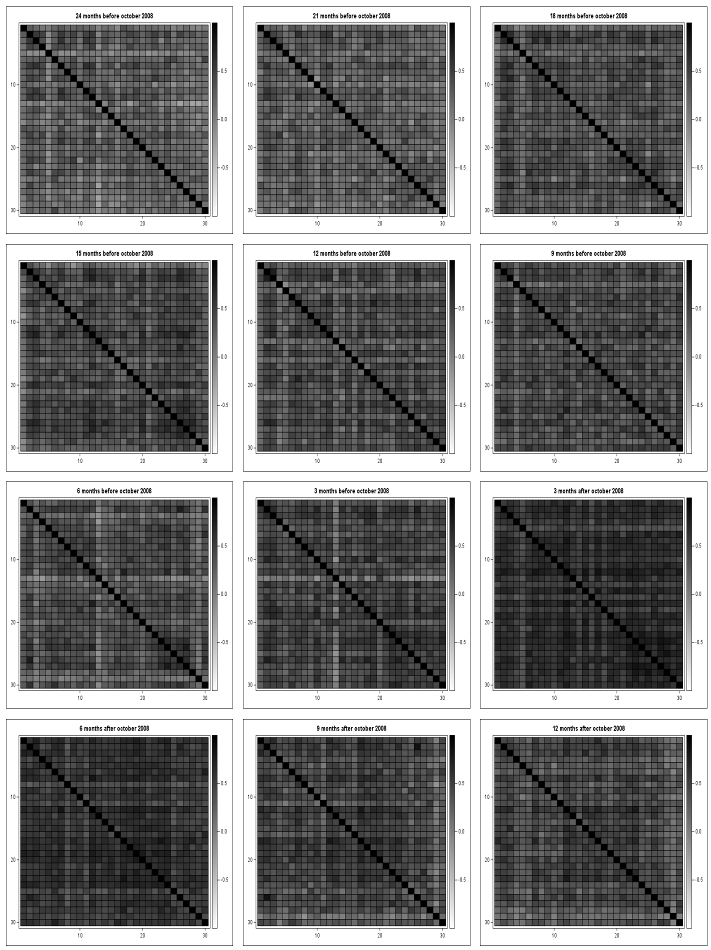


Figure 8: Heatmaps of correlation matrices for the 2008 crisis (every 3 months, starting 24 months before October 2008)

Thus, in the case of the 2008 crisis, the leading eigenvalues first raised 6 months before the beginning of the crisis, usually dated around July 2007, meaning that λ_{max} could be a very consistent early-warning for this crisis. However, this increase is followed by a very small correlation 18 months before the bubble burst, as expressed before. The second rise in 2008 is linked and a huge drop of the index value, a few weeks later. The correlation seems to be at its highest level during the crash (around October 2008). We then consider λ_{max} as a good early-warning in this case because the explosion of the value is followed by a market event and correlation between returns.

3.3. Since 2017

Finally, we analyze the situation on the market since 2017. With the study of the two previous crises, we can assume that correlation between assets starts to slowly rise up to one year before the market event and rises abruptly at the time of the crisis. It reaches its maximum during the market event and remains very high up to 6 months after the crisis. The leading eigenvalues are expected gradually rise a few months before a crisis and then to sharply increase a few weeks before or at the time of a significant crash.

First, we conduct the analyze on each semester (6 months), until the end of 2020. Figure 9 shows the evolution of leading eigenvalues. All eigenvalues are significant at our threshold. We can see that a lot of market events occured because of the high volatility. We can first notice a decrease during 2017, a year of continuous growth for the S&P500. On January 2018, the S&P (400, 500 and 600) and more generally all stocks reached record levels. Almost at the same time, the λ_{max} exploded at the end of January 2018. For the case of our index, it crashed at the end of the month (as showed on the graph). On July 2018 the S&P500 (400 and 600 too) reached their levels of January and leading eigenvalues started to drop sharply. The S&P500 hit new record level in September 2018 just before a huge drop until December 2018, losing more than 20% of its value, such as some other indexes. However, we notice that the leading eigenvalues dropped after the beginning of this trend that lasted 3 months. But the explosion of leading eigenvalues occured a few weeks before a big drop of the index value at the end of 2018. Indeed, the S&P500 lost 14% of its value during the 3 first weeks of December 2018. Then, there seem to be a false alarm in August 2019 : there is a sharp rise in leading eigenvalues but no significant market event. The stocks hit new highest records level in February 2020 just before a big crash leading to a loss of almost 34% (for the S&P500) of its value in one month. More precisely, the S&P 500 reached its peak the 19th February and the bottom level was reached the 23rd March. However, the leading eigenvalues started to rise in January 2020, but there is no explosion. The explosion of leading eigenvalues only occured at the time of the market event, in February/March, or even a little after, which is too late to be an early-warning. The market event was linked with the Covid-19 pandemic which surprised market in March 2020. Thus, the market panicked and it crashed at the same moment. We can then explain why leading eigenvalues did not explode before that event because it so unusual and unexpected that it happened too suddenly. Another reason could be that we lost information before of the 6-month rolling window, which could be too large. Since July 2020, the S&P500 is hitting new historical records. The S&P400 and 600 are following the same trends, but still no big crash or crisis occured.

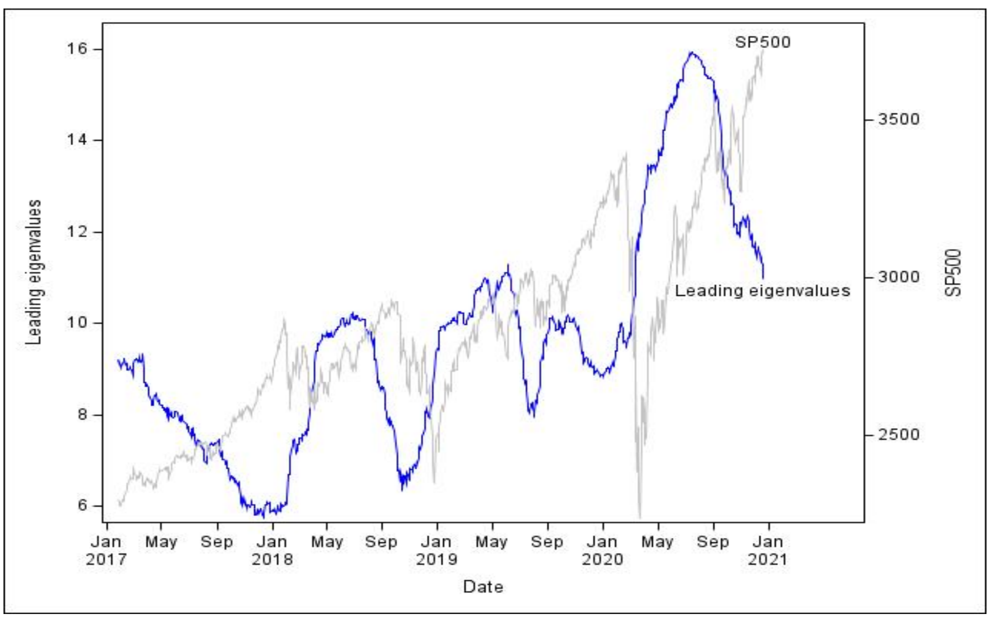


Figure 9: Leading eigenvalues since the 1st semester of 2017 (on a 6-month rolling window)

Figure 10 displays the heatmaps of every semester. We observe that there is no or low correlation during 2017, during the ascending trend of stocks. There is a very small correlation between returns for big companies during the first semester of 2018. Then, we notice a high positive correlation, especially for the small companies, on the 1st semester of 2020. Indeed, in 2020, the Covid-19 pandemic affected the entire world and led to a global economic crisis. On the second semester of 2020, the correlation is still high, but we observe some negative correlation between assets (white squares). We expect to find more precise results on the 3-month window heatmaps.

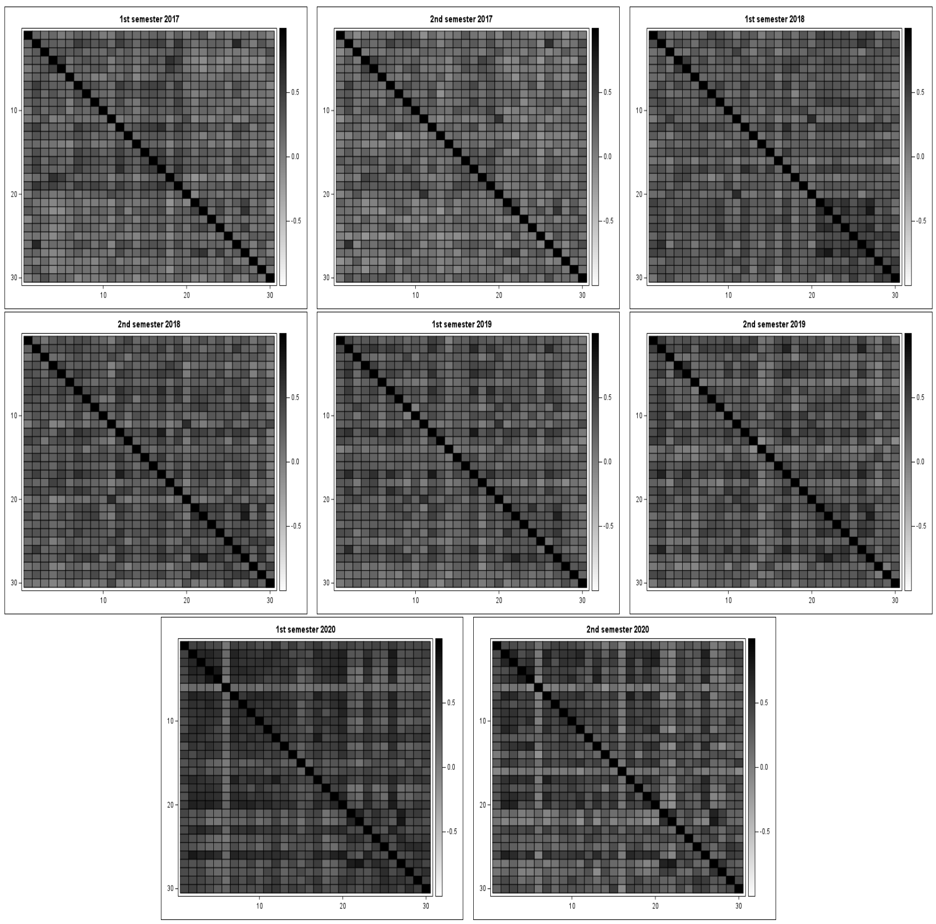


Figure 10: Heatmaps of correlation matrices since 1st semester of 2017 (every 6 months)

Now, the last analysis gives us a more precise point of view about what happened on the market with Figures 11 and 12. We observe a positive correlation on the first quarter of 2018. With the previous analysis, we were only able to conclude that a market event occured during the first semester of 2018, especially for big companies. Now, we can assert that the market event happened during the first quarter of 2018, which corresponds to our explanation : the S&P (400, 500 and 600) hit high levels and then crashed. Moreover, on this heatmap (2nd line and 2nd column), we can point out that the correlation is higher than the correlation on the 6-months analysis. Then, on the last quarter of 2018, the correlation is high again, which corresponds to the huge drop from September to December 2018. On the last quarter of 2019, correlation between returns was very high too as the first quarter of 2020 is. We observe here a peculiar phenomenon : returns of big companies seem to be correlated only (or at least more) with big companies, and returns of small companies (companies from 0 to 20 on the x axis) of our data seem to be correlated only (or at least more) with small companies. The period corresponds to the big rise before the crash in March 2020. Finally, for the rest of 2020, there seem to be a higher correlation than usual (compared to 2017 for example).

About the evolution of the leading eigenvalues, we draw almost the same conclusions as before : there is the gradual decrease in 2017, and explosion on January 2018, just before the huge drop on every stocks at the end of the month. Another explosion on September/October 2018, just before the huge crash of the index on the last quarter of 2018. Finally, the most impressive explosion of the leading eigenvalues occurs on January 2020, a few weeks before a huge crash of, again, every stocks, including the S&P500, on March 2020. Thus, the 3-month window analysis gives us a more precise result : with the 6-month window analysis we observe the explosion of leading eigenvalues too late to be an early-warning, in March 2020. Here, the explosion occurs a few weeks before the crash.

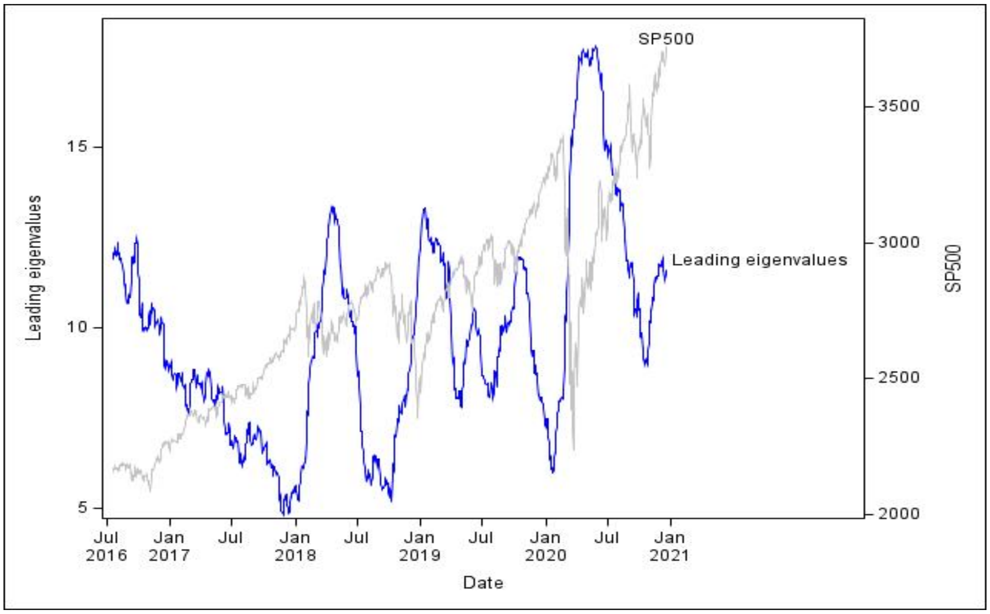


Figure 11: Leading eigenvalues since the 1st semester of 2017 on a 3-month rolling window

Since 2017, the rises of leading eigenvalues seem to be linked to a rise of correlations between returns. Indeed, we observe that it is true for the market events we mentioned. We then consider our indicators as a consistent early-warning in the case of the 3-month window analysis.

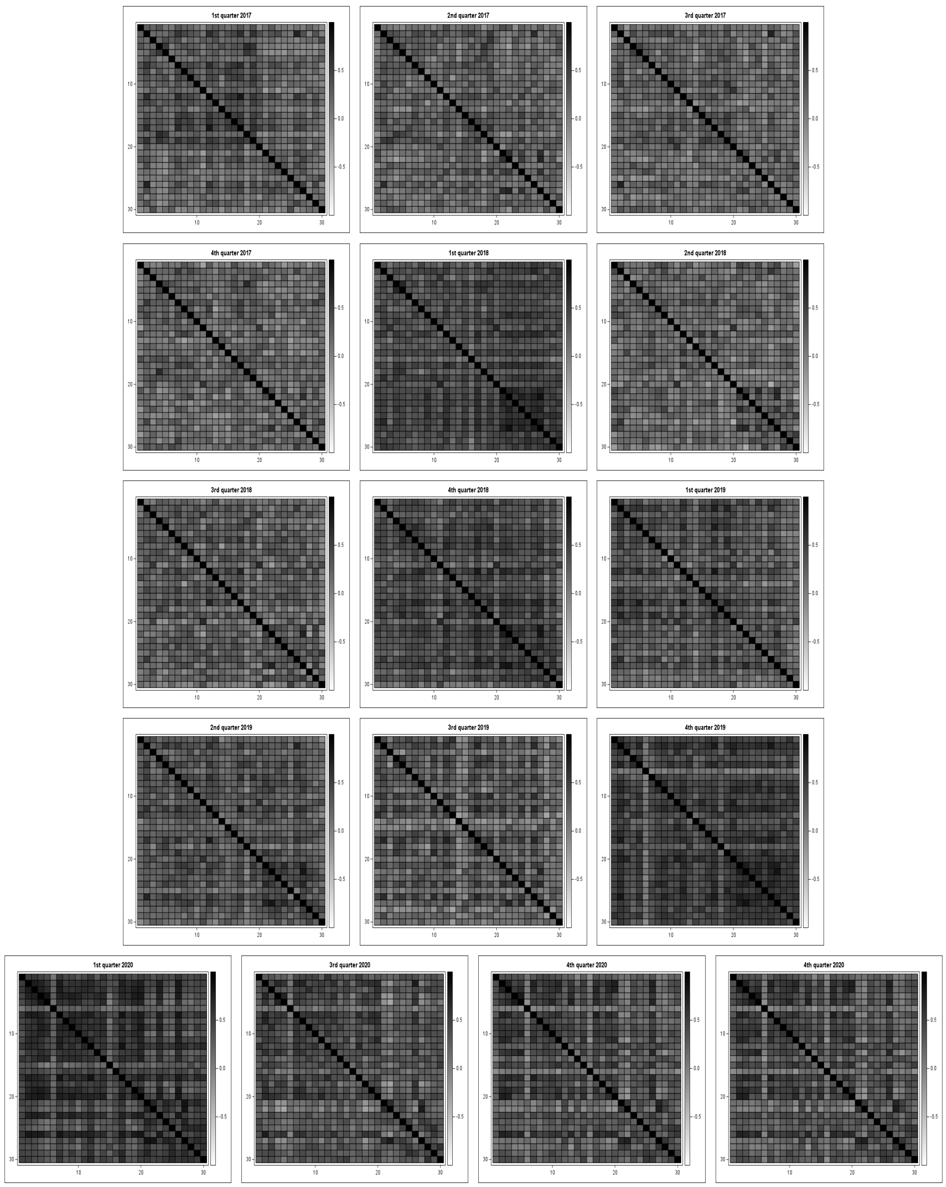


Figure 12: Heatmaps of correlation matrices since 1st semester of 2017 (every 3 months)

4. Conclusion

In this paper, we analyzed the relationship between the correlation of returns for 30 companies : ten in the S&P 400, ten in the S&P 500, and ten in the S&P 600. We then calculated correlation matrices calculated correlation matrices of three and six months throughout two years before the crisis, and up to 4 years after to examine the evolution of the correlations. The 3-month analysis allowed us to draw more precise conclusions. We also calculated the eigenvalues of the rolling-window matrices to extract the maximum eigenvalue and thus observe the evolution of these matrices over the entire period. It appeared that all our leading eigenvalues are significant.

The first focus of our study was the Internet crisis of the early 2000s. However, we could not draw any conclusion from it, because the correlation was very low, almost null, and was only significant years after the bubble burst in 2000. Moreover, the leading eigenvalues did not seem to explode before an extreme event. The possible reason for the failure of our analysis to predict this crisis is that we took companies that are currently in the S&P 400, 500, and 600, even though we do not have data about the SP600 for the period that we are studying, since it is a more recent index. Another explanation could be due to incorrect dating, since there is no consensus or exact date for this crisis. As mentioned before, the bubble first burst on the NASDAQ and therefore the crisis took time to spread to our index, even if it started to drop in 2000, it did only suddenly crash in 2002.

Then we focused our analysis on the 2008 crisis. This time, while the results were more conclusive, we were able to observe critical transition phases, such as when the leading eigenvalues first raised 6 months before the crisis began in July 2007. They then exploded at the time of the market event. Correlations are at their maximum level at this time and until up to 6 months year after the crisis.

Continually, we studied the situation from 2017 to the end of December 2020, trying to conclude if leading eigenvalues and correlations are a consistent early warning for market events or not. Although the year 2020 experienced a particular crisis because of the Covid-19 pandemic, and the containment of the population. As can be seen in Figure 11, the SP 500 losing 31.9% of its value on March 23, 2020, the maximum eigenvalues exploded from about 6 to 18 in almost a few weeks and this several weeks before the end of March, but only on the 3-month window analysis. It, therefore, appears that even in this particular context our maximum leading eigenvalues are a consistent early warning. Moreover, the strong correlation observed at the end of 2019 and beginning of 2020 on the heatmaps reinforces this hypothesis.

We, therefore, find, as in Hankyu Moon and Tsai-Ching article [3], that the leading eigenvalues increase significantly before large-scale market events making leading eigenvalues a good indicator for the 2008 crisis and some crashes since 2017, especially the Covid-19 crisis. In these two cases, the correlation is linked with the rise of leading eigenvalues. However, in the case of the Dot-Com bubble, we do not consider these indicators as consistent early-warnings.

The leading eigenvalues do predict some crises, as we have seen they explode when there is a crisis and there is an increase in correlations, but the problem is that this does not provide an early enough warning advance in some cases we study. However, to truly determine if leading eigenvalues and return correlations are a consistent early-warning of financial crisis, this study should especially be carried out on other crises and with other companies as mentioned above. Because our study only works for the 2008 crisis and 2019-2020, so we are far from being able to generalize our result. Indeed, our work can be extended in several ways. To begin, we could lead the same study with companies in Asia and Europe, to confirm our supposition, like in a mathematical recurrence. Here we have made the initialization, so we could try to make the heredity of this hypothesis. Then we could look at the correlation between the international stock markets and try to see if it increases during periods of high volatility.

We could also reduce the window to only 1 month like in the Hankyu Moon and Tsai-Ching paper [3] to possibly have even more precise results, because we could see that it was already more precise and overall better on 3 months than 6. Indeed, it was less smooth and therefore we lost less information, although it is little more volatile.

Finally, it seems important to question our threshold for the leading eigenvalues. We should have observed periods of non-significant leading eigenvalues. However, all our leading eigenvalues are significant, meaning that we miss and lose some information.

References

[1] Kornprobst Antoine and Douady Raphael. A practical approach to financial crisis indicators based on random matrices. 2015.

[2] Leonidas Sandoval Junior & Italo De Paula Franca. Correlation of financial markets in times of crisis. 2011.

[3] Hankyu Moon & Tsai-Ching Lu. Network catastrophe: Self-organized patterns reveal both the instability and the structure of complex networks. *Scientific Reports*, 2015.

[4] Didier Sornette. Dragon-kings, black swans and the prediction of crises. *International Journal of Terraspace Science and Engineering*, 2009.

[5] Bernd Rosenow Luís A. Nunes Amaral Vasiliki Plerou, Parameswaran Gopikrishnan and H. Eugene Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters*, 1999.