Component Analysis in Financial Time Series

José Miguel Salgado[1,3,4], José Abílio Matos[2,4]

(1) Polithecnic Institute of Guarda, Portugal
(2) Economy Faculty, Porto University, Portugal
(3) Unit for Inland Development (UDI)*
(4) Center of Mathematics, Porto University (CMUP)
E-mail: portosalgado@gmail.com

Abstract: In this work we consider the application of some signal processing techniques to multivariate financial time series, particularly the Principal Component Analysis (PCA), the Independent Component Analysis (ICA), also known as blind source separation and the quite recent Forecastable Component Analysis (ForeCA). The key idea is not to compare their differences but more to find their "joint strenght" by joining their different views of time series. Knowing, for instance, that ICA linearly maps the observed multivariate time series into a new space of statistically independent components (ICs) we could "observe" the other two techniques at the same "time" and merge the information. We applied these techniques to two different scenarios: one, more micro, to some stocks quoted in the Portuguese Stock Market (PSI20); the other one, more macro, to study nine European stock markets.

Keywords: Data reduction, Stock Market, Pearson Correlation, Distance Covariance, Component Analysis

1. Introduction

In 1967, a seminal paper on the spectrum of empirical correlation matrices written by Marcenko and Pastur turned out to be useful in many and very different contexts like, for our purposes, finance time series. Its central objective, as a new statistical tool to analyze large dimensional data sets, only became fully relevant in the last twenty years, when the storage and handling of great amounts of data became a daily routine in financial markets.

The correlations within stock price fluctuations for different assets or markets are important because, for instance, of their direct use for risk management in the Markowitz portfolio theory. In this study, however, in a less profitability use, we are more interested in collecting the real information from stocks dependency. In practice, there are different sources of noise in the estimated correlations.

In their work, Laloux et. al., show that this accumulated noise in the correlation matrix for price fluctuations can be accounted for by using the tools from the random matrix theory (RMT). In particular, they found that the distribution of eigenvalues of empirical correlation matrix, excluding some of the largest

7th CHAOS Conference Proceedings, 7-10 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)
© 2014 ISAST

*PEst-OE/EGE/Ul4056/2014 — project financed by Fundação para a Ciência e Tecnologia (FCT)

eigenvalues, fits very well in the Marcenko-Pastur distribution of the RMT. These results strongly suggest that eigenvalues of correlation matrix falling under the Marcenko-Pastur distribution contain no genuine information about the financial markets. Extended work has been conducted to explain information contained in the deviating eigenvalues, which reveals that the largest eigenvalue corresponds to a market wide influence to all stocks and the remaining deviating eigenvalues correspond to conventionally identified business sectors. Using the same RMT method, extensive works have been performed in the correlation analysis of various stock markets.

In this work we are going to retrieve some of these results using RMT applied to two sets: first to twelve stocks quoted in the Portuguese Stock Market (PSI20) and then to nine European stock markets. Finally, we will apply Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Forestecable Component Analysis (ForeCA) to these same sets in an effort to merge the resulting information.

2. Methods

2.1 Principal Component Analysis (PCA)

PCA is defined as a statistical procedure that by means of an orthogonal transformation converts a set of observations of (possibly correlated) variables into a set of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance. The remaining components in turn have the highest variance possible under the constraint that it is orthogonal (uncorrelated with) to the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed.

PCA invention is attributed to Karl Pearson (1901) who created this as an analogue of the principal axes theorem in mechanics; it was later independently developed and named by Harold Hotelling in the 1930s. The method is mostly used as a tool in exploratory data analysis and for making predictive models.

PCA is considered the simplest of the true eigenvector-based multivariate analyses and here we will use the eigenvalue decomposition of the data covariance (or correlation) matrix, being in that way, closely related to the RMT method.

2.2 Independent Component Analysis (ICA)

The method known as independent component analysis (ICA) is also referred to as blind source separation (Herault and Jutten 1986, Jutten and Herault 1991, Comon 1994). The central assumption is that an observed multivariate time series (daily stock returns) reflect the reaction of a system (the stock market) to a few statistically independent time series. ICA seeks to extract out these independent components as well as the mixing process. Here we will follow Back and Weigend (1997) approach to express ICA in terms of other measures of the statistical independence of signals.

In financial context, ICA was proposed for the first time by Moody and Wu to separate the observational noise from the true price in a foreign exchange rate time series.

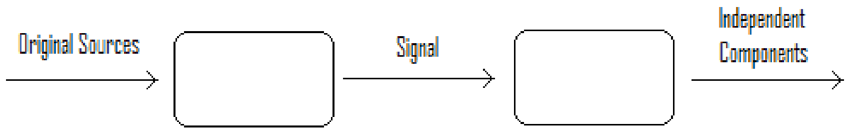

Figure 1: Schematic representation of ICA

The original sources are mixed through matrix to form the observed signal. The demixing matrix transforms the observed signal into the independent components. Figure 1 shows the most basic form of ICA. We observe a multivariate time series, consisting of values at each time step. We assume that it is the result of a mixing process.

We will consider here the application of a signal processing technique known as independent component analysis (ICA) or blind source separation. This technique can be applied to multivariate financial time series and the main idea is to linearly map the observed multivariate time series into a new space of statistically independent components.

ICA versus PCA

Independent component analysis can be contrasted with principal component analysis and so we give here a brief comparison of the two methods here. Both ICA and PCA linearly transform the observed signals into components. The key difference however, is in the type of components obtained.

PCA algorithms use only second order statistical information. On the other hand, ICA algorithms may use higher order statistical information for separating the signals. For this reason non-Gaussian signals (or at most, one Gaussian signal) are normally required for ICA algorithms based on higher order statistics. For PCA algorithms however, the higher order statistical information provided by such non-Gaussian signals is not required or used, hence the signals in this case can be Gaussian.

2.3 Forecastable Component Analysis (ForeCA)

Here we introduce Forecastable Component Analysis (ForeCA), a novel dimension reduction technique for temporally dependent signals (Goerg, 2013). Based on a new forecastability measure, ForeCA finds an optimal transformation to separate a multivariate time series into a forecastable and an orthogonal white noise space. We will the R package ForeCA, which uses a converging algorithm with a fast eigenvector solution. Applications to financial and macro-economic time series show that ForeCA can successfully discover informative structure, which can be used for forecasting as well as classification.

3. Data and Results

3.1 The PSI-20 set

The 12 stocks that we call the PSI-20 set were obtained from the PSI-20 Index which is a price index calculation based on 20 stocks obtained from the universe of Portuguese companies listed to trade on the Main Market and was designed to became the underlying element of futures and options contracts. The data used in this study are the close values and its log returns from these 12 stocks, and cover the period from January 24, 2000 to September 25, 2013 for a total of 3362 observations. In Figure2 we can see the daily values from the PSI-20 index.

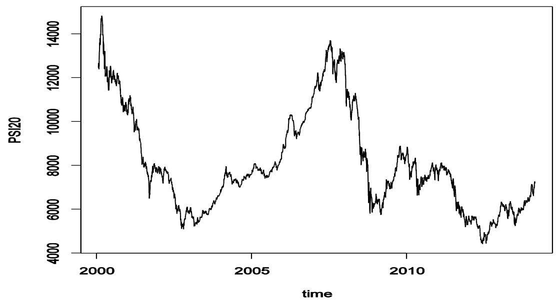

Figure 2: PSI-20 from 2001 to 2014

During the period 1997-2001 the Portuguese stock market becomes highly sensitive to fluctuations in international markets due to the integration in the euro area markets. Moreover, the reduced size of the Portuguese financial market suggests that the behavior of national stock returns depends highly or mimics the behaviors of stock returns in European and American markets. The period from January 2001 and November 2001 was characterized by economical and political instability in Europe and United States due to the introduction of euro and the high value of the dollar against the euro, some regional conflicts like the Israel- Palestinian conflict, and the September 11 with negative impacts on the financial markets, including the Portuguese stock market. In this period the PSI-20 index declined by 24,42 per cent. Between 2002 and 2007 we assisted to markets recovery, but in 2008, with the mortgage and subprime crises we saw the world markets in general, and PSI-20 in particular, going down once again. Finally, we are having some recovery signals from the beginning of 2013.

Calculating the Correlation Matrix using the Statistical Software R, we obtain for these 12 stocks:

Table 1: Correlation Matrix for the PSI-20 twelve stocks

$$CorM = \begin{bmatrix} 1.00 & 0.84 & 0.80 & 0.45 & 0.12 & 0.64 & -0.00 & 0.02 & 0.39 & 0.09 & 0.54 & 0.47 \\ 0.84 & 1.00 & 0.75 & 0.52 & 0.21 & 0.68 & 0.24 & 0.10 & 0.40 & -0.06 & 0.49 & 0.33 \\ 0.80 & 0.75 & 1.00 & 0.61 & 0.04 & 0.49 & -0.04 & 0.28 & 0.36 & 0.07 & 0.50 & 0.36 \\ 0.45 & 0.52 & 0.61 & 1.00 & 0.06 & 0.42 & 0.03 & 0.30 & 0.26 & -0.00 & 0.45 & 0.18 \\ 0.12 & 0.21 & 0.04 & 0.06 & 1.00 & 0.26 & 0.27 & 0.15 & 0.28 & 0.43 & 0.52 & 0.50 \\ 0.64 & 0.68 & 0.49 & 0.42 & 0.26 & 1.00 & 0.04 & -0.04 & 0.48 & 0.15 & 0.35 & 0.05 \\ -0.00 & 0.24 & -0.04 & 0.03 & 0.271 & 0.04 & 1.00 & 0.09 & 0.19 & 0.17 & -0.04 & 0.07 \\ 0.02 & 0.10 & 0.28 & 0.30 & 0.15 & -0.04 & 0.09 & 1.00 & 0.09 & 0.38 & 0.24 & 0.35 \\ 0.39 & 0.40 & 0.36 & 0.26 & 0.28 & 0.48 & 0.19 & 0.09 & 1.00 & 0.21 & 0.25 & 0.21 \\ 0.09 & -0.06 & 0.07 & -0.00 & 0.43 & 0.15 & 0.17 & 0.38 & 0.21 & 1.00 & 0.18 & 0.29 \\ 0.54 & 0.49 & 0.50 & 0.45 & 0.52 & 0.35 & -0.04 & 0.24 & 0.25 & 0.18 & 1.00 & 0.60 \\ 0.47 & 0.33 & 0.36 & 0.18 & 0.50 & 0.05 & -0.07 & 0.35 & 0.21 & 0.29 & 0.60 & 1.00 \end{bmatrix}$$

This matrix confirms some empirical ideas we had about the stocks, namely that the first and the second ones (“bes” and “bpi”) are highly correlated, which is not a surprise as these 2 stocks are from the financial sector. More surprisingly is the high correlation between each of these two and the third one (“edp”) that comes from electrical sector.

Calculating the relationship between the first three eigenvalues for 12 stocks: "bes", "bpi", "edp", "jeronimomartins", "motaengil", "novabase", "portucel", "portugaltelecom", "semapa", "sonaeC", "sonaeR" and "zonoptimus", and considering a sliding window of 20 days, we got the following:

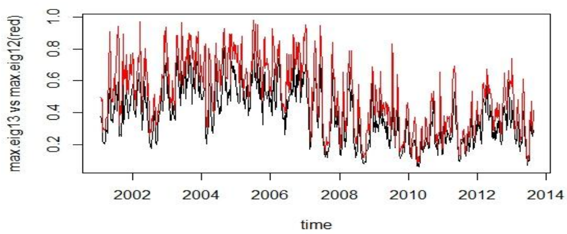

Figure 3: eig12 vs eig13 for PSI20 stocks

From Figure3 it is understandable that the ratio between the second eigenvalue (eigenvalue 2) and the first eigenvalue (eigenvalue 1), named eig12, has a global mean of 0.5, that is to say, the highest eigenvalue has a value that doubles the second one. The mean for the eig13 is, for all the period, of about 0.39, that is to say that the first eigenvalue is two and a half times higher than the third one. Looking closer at the Figure3 we can observe that these ratios diminished greatly in the last six years: for eig12 the mean goes to 0.2 and for eig13 the mean goes to 0.15. These diminishing values tell us that the stocks are more correlated.

Comparing these results with twelve random stocks we get the following figure:

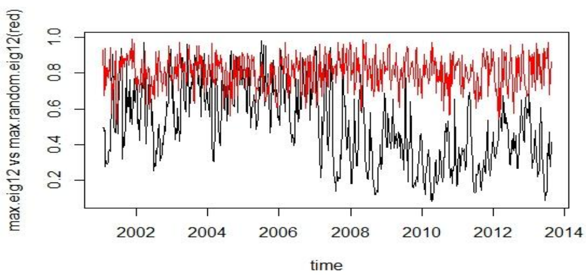

Figure 4: eig12 vs random eig12 for PSI20 stocks

Here, it is clear the difference between the eigenvalues from the random stocks (red) and the ones coming from the real stocks.

Performing Independent Component Analysis, following Back and Weigend (1997), for these stocks we can observe firstly their returns.

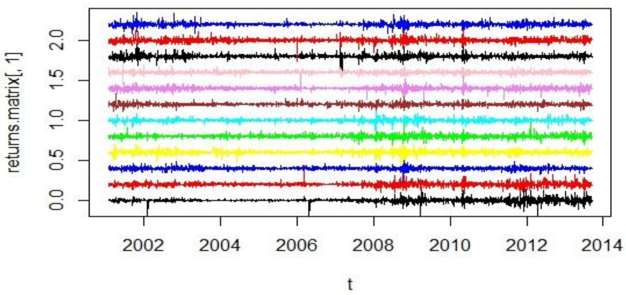

Figure 5: returns for PSI20 stocks

And then their independent components, obtained using the FastICA algorithm:

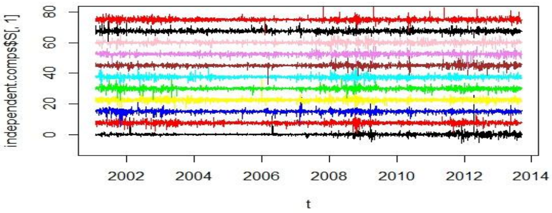

Figure 6: independent components for PSI20 stocks

Finally, we work out a little bit of ForeCA. For our 12 stocks we obtain the following results:

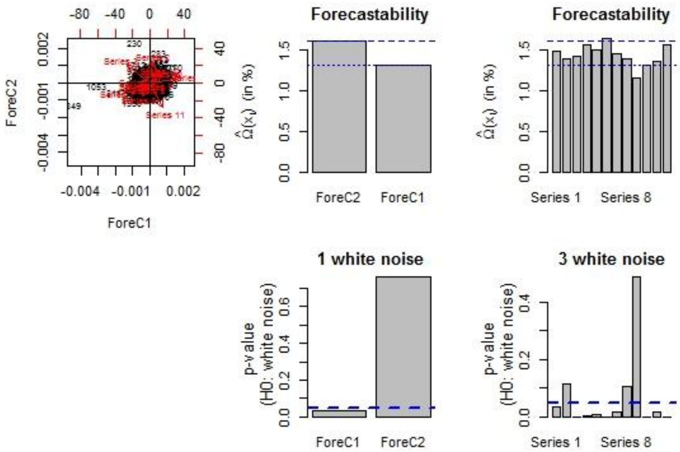

Figure 7: forecastability for PSI20 stocks

We can see that all stocks have a Forecast value over 1.0%, and, from the two calculated components the most forecastable is ForeC2. For individual stocks we must refer the sixt, that is, “Novabase”.

3.2 European Markets set

The data used in this study was taken daily for a set of nine european market indices. We analyzed the following markets: Netherlands index (AEX), Austrian index (ATX), French index (CAC 40), German index (DAX 30), British index (FTSE 100), Spanish index (IBEX 35), Italian index (MIB), Portuguese index (PSI-20) and Swiss index (SSMI) from, roughly, the year 2000 until late September 2013.

The data used in this work are the daily Close values for these nine markets for a total of 3468 observations.

In Figure8 we can observe the returns for the nine markets. It seems with no doubt that they are synchronized.
Looking at the returns helps us to look only to relative variation and not to absolute values. In fact, these markets are quite different in absolut values, as we can see from Figure8.

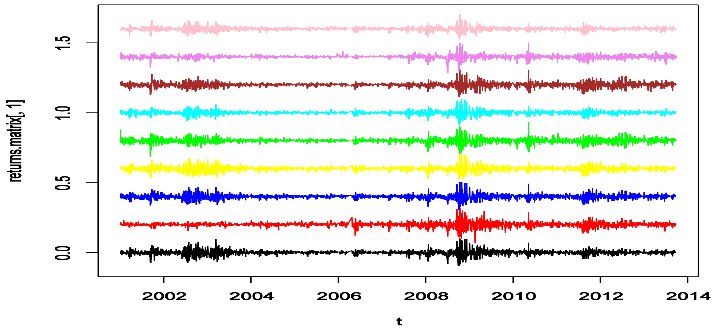

Figure 8: nine markets returns

In order to perform a study using PCA we started by calculating and relatively compare their values for the 3 highest eigenvalues from the 9 markets. In Figure [fig:eig12vseig13] we compare the relationship between the 3 major eigenvalues.

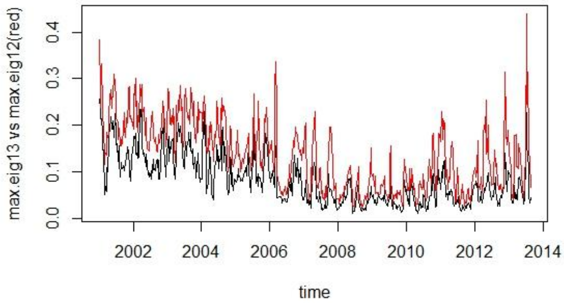

Figure 9: eig12 (red) vs eig13 for the 9 markets

We can generally say that the highest eigenvalue is getting higher over the time. It starts to be 3,3 to 5 times higher in the beginning of the XXI century and more recently became almost 10 to 15 times higher than the second. More recently, the difference between them is getting, again, smaller. From the second to the third highest we can infer a relationship of 2.

Performing Independent Component Analysis, following Back and Weigend (1997), for the nine markets we can observe their independent components, obtained using the FastICA algorithm:

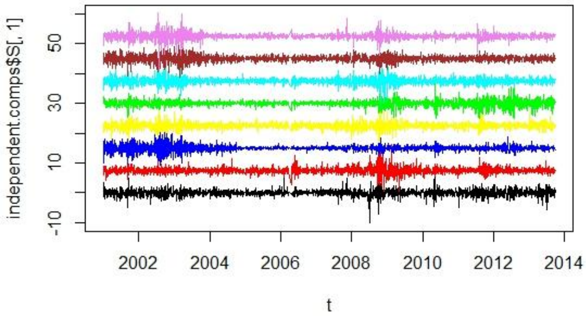

Figure 10: independent components for the 9 markets

Finally, we work out the results for markets of ForeCA. For our 9 markets we obtain the following results:

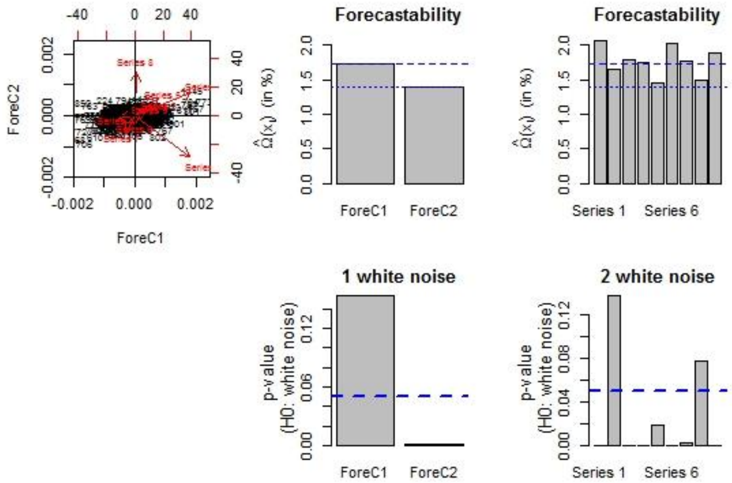

Figure 11: forecastability for PSI20 stocks

4. Conclusions

Indeed, PCA, ICA and ForeCA gives uses different but hopefully complementary views concerning the dependence and relationship between stocks or market indices.

Clearly, it is more difficult to get good insights when looking at stocks comparing with markets. These are much more correlated, something that we only see in stocks from the same sector.

Component analysis, despite being around for some time, deserves new approaches that a complementary view can offer. ForeCA gives us promising results and we hope to explore more in future work.

References

[1] Goerg, G. M. (2013). “Forecastable Component Analysis”. Journal of Machine Learning Research (JMLR) W&CP 28 (2): 64-72, 2013.

[2] J.P. Bouchaud and M. Potters. Theory of Financial Risks: from Statistical Physics to Risk Management. Cambridge University Press, Cambridge, 2001b.

[3] L. Laloux, P. Cizeau, J.P. Bouchaud, and M. Potters. Noise Dressing of Financial Correlation Matrices. Physical Review Letters, 83(7):1467–1470, 1999b.

[4] L. Laloux, P. Cizeau, and M. Potters. Random Matrix Theory and Financial Correlations. International Journal of Theoretical and Applied Finance, 3(3):391–397, 2000b.

[5] J.A.O. Matos, S.M.A. Gama, H.J. Ruskin, and J.A.M.S. Duarte. An econophysics approach to the Portuguese Stock Index–PSI-20. Physica A, 342(3-4):665–676, 2004.

[6] J.A.O. Matos, S.M.A. Gama, H.J. Ruskin, A. Sharkasi and M. Crane. Correlation of Worldwide Markets Entropies. Proceedings of the Workshop: Perspectives on Econophysics, 259:449–456, 2006c.

[7] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. Universal and Nonuniversal Properties of Cross-Correlations in Financial Time Series. Physical Review Letters, 83 (7):1471–1474, 1999b.

[8] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. A Random Matrix Theory Approach to Financial Cross Correlations. Physica A, 287:374–382, 2000.

[9] V. Plerou, P. Gopikrishnan, and B. Rosenow. Collective Behaviour of Stock Price Movement: A Random Matrix Approach. Physica A, 299:175–180, 2001b.