Independent Component Analysis in Stock Returns

Msik Ahmed Amine & Hamza Iqbi

Master Probabilités et Finance

Contents

Abstract

When we apply Independent Component Analysis (ICA) to a multivariate time series, the goal is to come up with a new space of linearly independent components and avoid difficult computation: it's simpler to multiply densities (independent components) rather than compute joint probabilities (initial case). This method is introduced in this paper applied to stock portfolios as it's a good example of multivariate time series not necessarily independent.
Applying ICA to the 28 largest japanese stocks reveals two shock types affecting prices that we will discuss later, demonstrates its superior accuracy in reconstruction over PCA and highlighting its potential in financial risk management.

1 Introduction

Independent component analysis, also known as blind source separation, aims to explain the movement of financial time series (such as daily stock returns) thanks to few statistically independent components, IC's. Note that these IC's are also time series. Several methods were presented to output these IC's: find a direction that factorize the joint probabilities, find a set of direction with minimal mutual information (that often lead to independance), minimizing the Kullback-Leiber divergence between the joint probability and marginal probabilities of the output signals. All these methods can be expressed as algorithms.

ICA can be compared to PCA: Even though they give the same type of outputs, components, PCA gives principal components, PC's, defined as components that capture the maximum variance possible when ICA gives statistically independent components. Second, ICA works with non-gaussian signals while PCA works also with gaussian signals: For Gaussian signals, the higher order statistics are zero already and so no meaningful separation can be achieved by ICA methods.

2 ICA in general

2.1 Independent component analysis

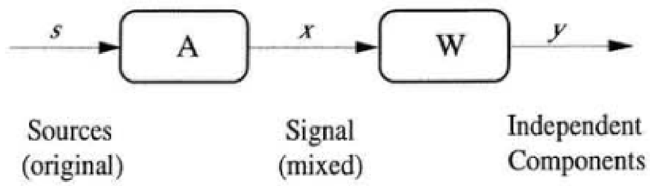

Figure 1: Schematic representation of ICA. The original sources s are mixed through matrix A to form the observed signal x. The demixing matrix W transforms the observed signal x into the independent components y.

- **A**: The unknown square mixing matrix of dimension $n \times n$, where n is the number of observed signals and sources.
- $\mathbf{s}(t)$: The vector of original source signals at time t, which are assumed to be statistically independent, of dimension $n \times 1$.
- $\mathbf{x}(t)$: The observed signal vector at time t, which is a mixture of the original sources $\mathbf{s}(t)$ through the mixing process **A**, of dimension $n \times 1$.
- **W**: The demixing matrix which, when applied to the observed signal $\mathbf{x}(t)$, estimates the independent components $\mathbf{y}(t)$, of dimension $n \times n$.
- $\mathbf{y}(t)$: The vector of independent components obtained after applying the demixing matrix **W** to the observed signal $\mathbf{x}(t)$, of dimension $n \times 1$.
- **P**: A permutation matrix which is a part of the condition $\mathbf{WA} = \mathbf{PD}$ where **D** is a diagonal scaling matrix, of dimension $n \times n$.
- **D**: A diagonal scaling matrix, as mentioned above, of dimension $n \times n$.
- **WA**: The product of the demixing matrix **W** and the mixing matrix **A**, which should equal **PD** for perfect signal separation, of dimension $n \times n$.
- t: The time index for the time series data.
- i: The index for individual signals or components, where $i = 1, \ldots, n$.
- j: An indexing variable used in summation to mix the source signals.
- n: The number of observed signals and independent sources, indicating the dimensionality of vectors and matrices where applicable.

2.2 Algorithms for ICA

This section in the paper was sort of a history of the different approaches used to implement ICA. At the earliest, some of these methods were used: minimizing higher order moments (Cardoso 1989), maximization of the output entropy (Bell and Sejnowski 1995), minimization of the Kullback-Leiber divergence (Amari et al. 1996). Then, the paper focuses on standard two-stage approach by Bogner in 1992 (also Cardoso and Soulomia1c in 1993): decorrelation (diagonalize the covriance matrix of the input signals) and rotation (goal is to minimize a measure of higher order statistics). One example of rotation method is unitary rotation matrix presented by Cardoso and Soulomia1c in 1993.

The empirical study carried out in this paper uses the JADE (Joint Approximate Diagonalization of Eigenmatrices) algorithm (Cardoso and Souloumiac 1993):

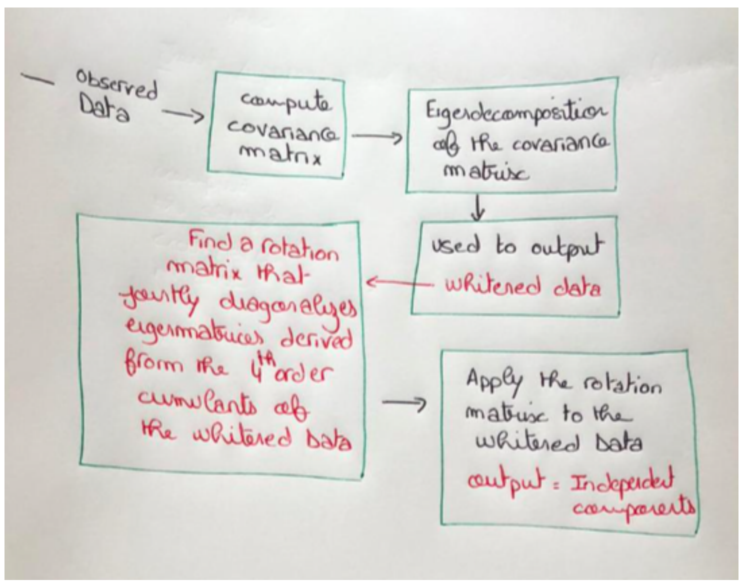

Figure 2: Schematic representation of Jade Algorithm

3 ICA in finance

3.1 Reasons to Explore ICA in Finance

The paper aims to use Independent Component Analysis (ICA) to identify underlying factors affecting instantaneous stock returns, potentially enhancing financial market analysis and forecasting.

3.2 Preprocessing and Description of the Data

This paper uses the stationary prices time series $p(t)$ to derive

$$x(t) = p(t) - p(t-1)$$

to get stationary time series more convenient to work with in this context. Then they normalize this daily stock returns. The daily data used comes from the Tokyo Stock exchange between 1986 and 1989 of the 28 largest firms listed.

4 Analyzing Stock Returns with ICA

4.1 Structure of the Independent Components

The application of ICA to the the 28 largest firms listed in the Tokyo Stock exchange suggests that only a few ICs contribute to most of the movements in the stock return, large amplitude transients in the dominant ICs contribute to the major level changes and small amplitude ICs contribute to the change in levels over short time scales, but over the whole period, there is little change in levels. Let's explain these remarks:

- **Only a few ICs contribute to most of the movements in the stock return:** he top trace in Figure 3, which is the summation of the weighted ICs, demonstrates how these few components capture the overall trend of the original stock returns.

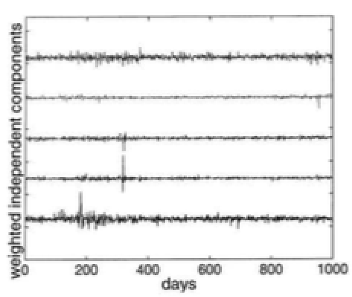

Figure 3: The four most dominant ICs after weighting by the first row of the mixing matrix A (corresponding to the Bank of Tokyo-Mitsubishi) are shown starting from the bottom trace.

- **Large amplitude transients in the dominant ICs contribute to the major level changes:** The non dominant components do not contribute significantly to level changes. This statement refers to the impact of dominant ICs that have large variations or spikes in their signals. These are the ICs that cause significant shifts in the stock price level. Figure 3 also demonstrates this point.

- **Small amplitude ICs contribute to the change in levels over short time scales, but over the whole period, there is little change in levels:** Figure 4 shows the reconstructed price obtained using the four most dominant weighted ICs and compares it to the sum of the remaining 24 nondominant weighted ICs. This implies that the contribution of the non-dominant ICs to the overall shape is only small and strenghtens the point that small amplitude ICs contribute to the change in levels more over short time scales.

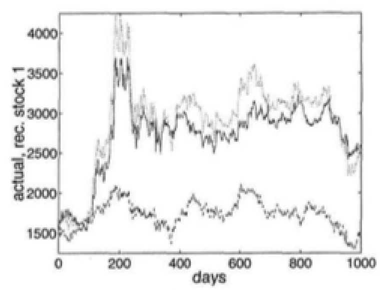

Figure 4: The dotted line on top is the original stock price. The solid line in the middle shows the reconstructed stock price using the four most dominant weighted ICs. The dashed line on the bottom shows the reconstructed residual stock price obtained by adding up the remaining 24 weighted ICs.

4.2 Thresholded ICs Characterize Turning Points

In this section, all the non dominant ICs weights are set to zero below a certain threshold. The results comparing Thresholded returns obtained from the four most dominant weighted ICs for the Bank of Tokyo-Mitsubishi indicate that ICs provide useful morphological information and can extract the turning points of original time series.

4.3 Comparison with PCA

The two main differences between Independent Component Analysis (ICA) and Principal Component Analysis (PCA) as presented in this part of the paper are:

- **Capturing Transients:** ICA is able to capture high-level transients, which are sharp, short-lived variations in the data. The transients captured by ICA are not as evident when using PCA. Sudden changes are still inportant when analyzing a time serie stock market returns.

- **Reconstruction Quality:** ICA's components, being statistically independent, can lead to a more meaningful and detailed reconstruction.

5 Application to CAC40 data

To perform an ICA on the CAC 40 index, we use the daily prices data from 15/01/2021 to 01/02/2024. In the context of financial time series analysis, such as examining the CAC 40 index, ICA can be used to identify underlying factors (economic, political, or related to specific sectors or companies within the index) that influence the movements of the index that might not be immediately apparent. Figure 5 gives an overview of the daily prices observed:

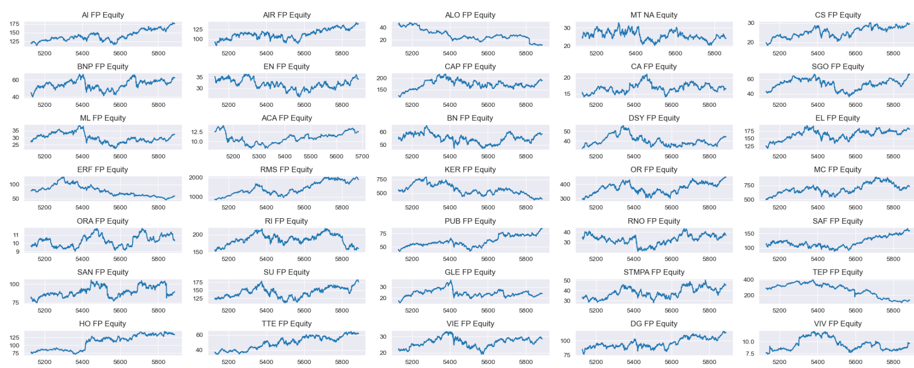

Figure 5: The daily prices of CAC 40 stocks. Some of them were removed for non-compatibility of the available data.

ICA requires observed signals to be stationary. Thus, we transform the nonstationary prices to stock returns with the following transformation:

```
returns = prices.apply(lambda x: x-x.shift(1) if x.name != 'Date' else x)
```

Figure 6: Stock returns for 5 stocks in CAC 40.

We perform an ICA on the CAC 40 stock returns:

```
from sklearn.decomposition import FastICA
ica = FastICA(n_components=min(len(returns.columns),10),
        random_state=0)
```

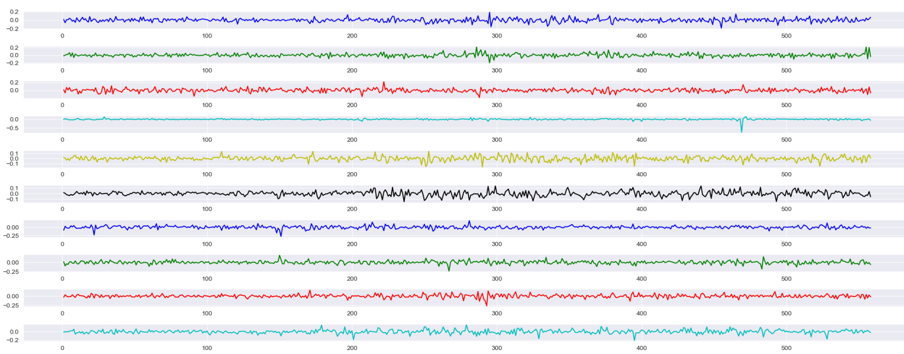

Figure 7: A visualisation of the 10 dominant ICs.

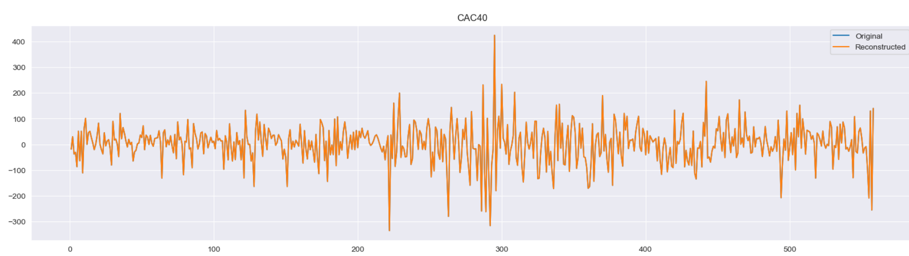

Figure 8: Reconstruction of returns.

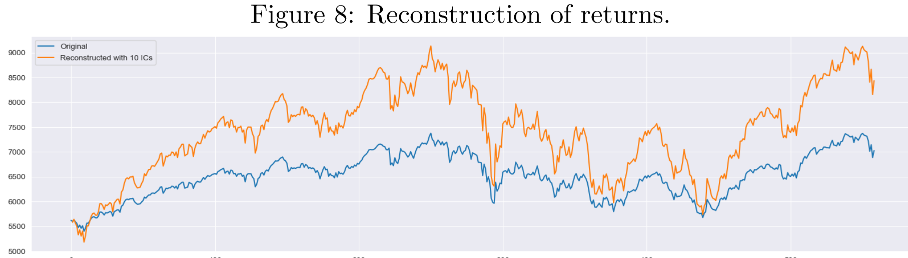

Figure 9: Reconstruction of prices.

One can see that we have a perfect reconstruction for the returns, but not for the daily prices. However, we are able to reproduce the trends observed in the day-to-day price movements. As described in the article, when one takes the dominant ICs one can rebuild the return time series as ICA gives good results when it's about capturing the overall trend of the original stock returns. However, for the price time series, one can observe that the reconstruction is not perfect even though that the trend is similar, the amplitude is never the same in the example we studied. This can be compared with the following result in the paper :

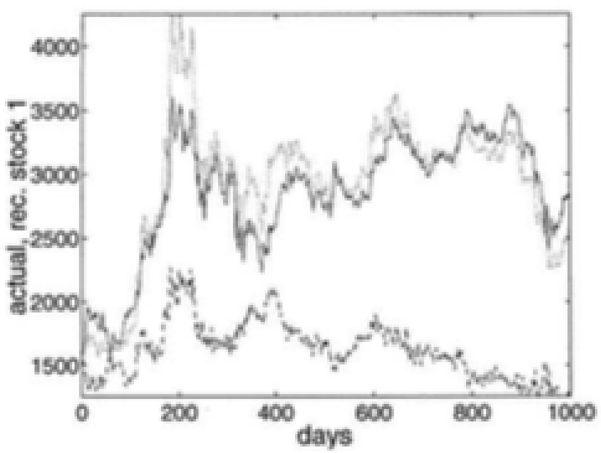

Figure 10: The solid line in the middle shows the reconstructed stock price using the four most dominant PCs. The dashed line (lowest) shows the reconstructed stock price by adding up the remaining 24 PCs. The sum of the two lines corresponds to the true price, this highest line is identical is identical to the true price (dotted line).

As expected in the paper when it comes to studying the stock prices time series, ICA behaves very good for the shape and the trends and the movements but gets a bigger error (as one can see in figure 9 by comparing the difference between the two time series) than PCA that gives usually less error but doesn't describe the trends as well.

6 Conclusion

In conclusion, the application of Independent Component Analysis (ICA) to the stock returns of the largest firms listed on the Tokyo Stock Exchange has shown that ICA works very well when it comes to capturing stock movements. The paper reveals two predominant types of independent components (ICs): those that represent infrequent but large shocks and those that characterize more frequent but smaller fluctuations. Both are important in finance to understand the underlying structure of the time series.

Moreover, the comparison between the reconstruction capabilities of ICA and PCA highlights the strengths and limitations of each one of them. When ICA works very well fro trend it also shows a bigger error when reconstructing the stock prices time series.

One possible extension is to work on a mix of PCA and ICA to maybe get the benefits and advantage sof both methods and see how we can apply it to the study of underlying structures of financial time series.