Independent Component Analysis for Parallel Financial Time Series

Kimmo Kiviluoto and Erkki Oja

Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 2200, FIN-02015 HUT Finland

Email:Kimmo.Kiviluoto@hut.fi, Erkki.Oja@hut.fi

ABSTRACT

We apply independent component analysis (ICA) to financial time series data. The data is parallel in the sense that it represents the simultaneous cash flow at several stores belonging to the same retail chain. The ICA detects a few factors that affect the cash flow of all the stores, although each store responds to these factors in a slightly different manner. When the effect of these "fundamental factors" is removed, the impact of the actions of the management becomes more visible. Additionally, it is possible to examine the differences between the stores on the basis of their responses to the fundamental factors obtained with ICA.

KEYWORDS: independent component analysis, time series analysis, finance

1. INTRODUCTION

The problem we address in this paper can be formulated in the following way: Given financial data that reflects the cashflow of several stores belonging to the same retail chain, try to find

- the fundamental factors common to all stores that affect the cashflow data;
- the cashflow effect of the factors specific to any particular store, i.e. the effect of the actions taken at the individual stores and in its local environment;
- based on these factors, also find groups of stores that behave similarly, or have similar cashflow time series.

The problem stated above is often encountered in practice when trying to quantify the effect of managerial actions – is the observed change in the cash flow the result of the action, is it the result of some other actions such as those of the competitors, or is it just statistical fluctuation?

As the solution to the problem of finding the underlying factors, we propose the technique of Independent Component Analysis (ICA). It has been widely used for the problem of Blind Signal Separation in problems like speech separation and array sensor signal processing [2]. In financial context, the method has been proposed by Moody and Wu [4] to separate the observational noise from the "true price" in a foreign exchange rate time series. It is an intriguing question, whether this methodology would also apply to parallel financial time series.

Assume that there are some fundamental factors which affect the cashflow of all the stores, but the relative impact of these factors differs from store to store. Denote these time-varying underlying factors by $s_i(t)$, $i = 1, ..., m$ and the measured cashflow time series by $x_j(t)$, $j = 1, ..., n$ with t denoting the (discrete) time. If we now consider the effect of each factor on the measured time series to be approximately linear, we may write

$$x_j(t) = \sum_i a_{ij} s_i(t) \tag{1}$$

where the a_{ij}, describing the effect of factor $s_i(t)$ on time series $x_j(t)$, are called the mixing coefficients.

It is often more convenient to use vector - matrix notation. Denoting by $\mathbf{x}(t)$ the time series vector with elements $x_j(t)$, by $\mathbf{s}(t)$ the vector of underlying factors $s_i(t)$, and by $\mathbf{A}$ the matrix (a_{ij}), we can write

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{2}$$

If we knew matrix $\mathbf{A}$, then we could obtain $\mathbf{s}(t)$ from the available measurements $\mathbf{x}(t)$ simply by (pseudo)inverting $\mathbf{A}$. However, the basic idea of ICA is that we do not have to know matrix $\mathbf{A}$ at all, but instead can estimate the model and obtain both matrix $\mathbf{A}$ and the underlying factors $s_i(t)$ if we make the simple assumption that the factors are *statistically independent*.

Such an assumption in this specific application may not be unrealistic. For example, factors like seasonal variations due to holidays and annual variations, and factors having a sudden effect on the purchasing power of the customers like prize changes of various commodities, can be expected to have an effect on all the retail stores, and such factors can be assumed to be roughly independent of each other. Yet, depending on the policy and skills of the individual manager like e.g. advertising efforts, the effect of the factors on the cash flow of specific retail outlets are slightly different. By ICA, it is possible to isolate both the underlying factors and the effect weights, thus also making it possible to group the stores on the basis of their managerial policies using only the cash flow time series data.

In Section 2, we outline the generic solution to the ICA problem and introduce the FastICA algorithm, which is particularly suitable for this kind of application in which the number

m of the underlying factors is not known in advance. It stems from certain neural learning rules introduced by the second author. We then show in Section 3 experiments on parallel weekly cash flow measurements of 40 stores of the same retail chain over a time period of 2 1/2 years. Some conclusions are offered in Section 4.

2. THE FastICA ALGORITHM

The problem of estimating the matrix $\mathbf{A}$ in eq. (2) can be somewhat simplified by performing a preliminary *prewhitening* of the observed signal vectors $\mathbf{x}(t)$. They are linearly transformed to new signal vectors $\mathbf{x}'(t)$ such that their elements are mutually uncorrelated and all have unit variance. In this transformation, their dimensionality is reduced so that elements of $\mathbf{x}'(t)$ with small variances are removed. The transformation is always possible and can be accomplished by classical Principal Component Analysis. After the transformation we have

$$\mathbf{x}'(t) = \mathbf{B}\mathbf{s}(t) \tag{3}$$

where $\mathbf{B}$, due to the prewhitening, turns out to be an orthogonal square matrix (for details, see [1]). Thus we have reduced the problem of finding an arbitrary rectangular mixing matrix $\mathbf{A}$ to the simpler problem of finding an orthogonal matrix $\mathbf{B}$, which then gives $\mathbf{s}(t) = \mathbf{B}^T\mathbf{x}'(t)$. If the i-th column of $\mathbf{B}$ is denoted $\mathbf{b}_i$, then the i-th independent factor can be computed from the prewhitened $\mathbf{x}'(t)$ process as $s_i(t) = \mathbf{b}_i^T\mathbf{x}'(t)$.

Some suggested solutions to the source separation or blind deconvolution problems use the fourth order cumulant or *kurtosis* of the signals, defined for a random variable y as

$$K(y) = Ey^4 - 3(Ey^2)^2 \tag{4}$$

For a Gaussian random variable, the kurtosis is zero; for sharper densities, it is positive, and for flatter densities, negative.

Let us now search for a linear combination of the prewhitened signals $x_i'(t)$, say, $y = \mathbf{w}^T\mathbf{x}'(t)$, such that it has maximal or minimal kurtosis. Obviously, this is meaningful only if the norm of $\mathbf{w}$ is somehow bounded, so we assume $\|\mathbf{w}\| = 1$. This can be accomplished by the recently introduced fixed point learning rule [1], dubbed the FastICA algorithm[1]. For a given sample of the prewhitened vectors $\mathbf{x}'$, the FastICA algorithm is defined as follows:

1. Take a random initial vector $\mathbf{w}_0$ of norm 1. Let $k = 1$.
2. Let $\mathbf{w}_k = E\{\mathbf{x}'(\mathbf{w}_{k-1}^T\mathbf{x}')^3\} - 3\mathbf{w}_{k-1}$. The expectation can be calculated using a sample of $\mathbf{x}'$ vectors.
3. Divide $\mathbf{w}_k$ by its norm.
4. If $\|\mathbf{w}_k - \mathbf{w}_{k-1}\|$ is not small enough, let $k = k + 1$ and go back to step 2. Otherwise, output the vector $\mathbf{w}_k$.

[1] The algorithm is available at the Web address `http://www.cis.hut.fi/projects/ica/fastica/`.

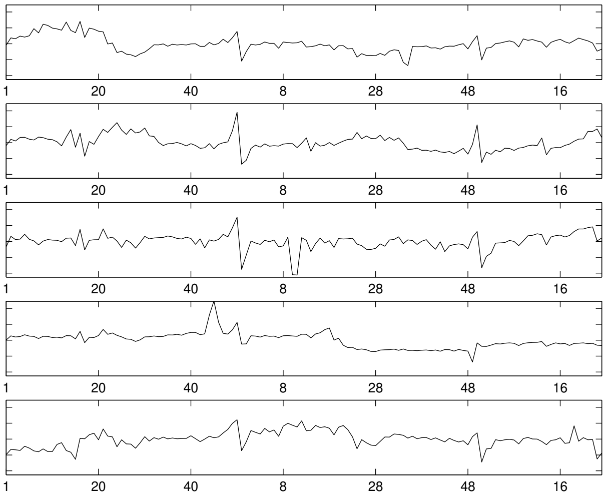

Figure 1: Five samples of the original cashflow time series (mean removed, normalized to unit standard deviation).

The final vector $\mathbf{w}_k$ given by the algorithm separates *one* of the underlying factors in the sense that $\mathbf{w}_k^T\mathbf{x}'(t)$, $t = 1, 2, ...$ equals one of the factors. To separate m factors, we run this algorithm m times. To ensure that we separate each time a different factor, we only need to add a simple projection inside the loop. Recalling that the columns of the mixing matrix $\mathbf{B}$ are orthonormal because of the prewhitening, we can separate the factors one-by-one by projecting the current estimate on the space orthogonal to the previously found columns of the mixing matrix $\mathbf{B}$.

3. EXPERIMENTS WITH 40 RETAIL STORES

The data consists of (one component of) the weekly cash flow in 40 stores that belong to the same retail chain; the cash flow measurements cover 140 weeks. Some examples of the original data are shown in figure 1.

As depicted in figures 2 and 3, the FastICA algorithm finds several clearly different fundamental factors hidden in the original data. In figure 2, the prewhitening was performed so that the original signal vectors were projected to the subspace spanned by their first four principal components; in this projection, 77% of the energy of the original signals was retained. In figure 3, the signals were projected to the subspace spanned by the five first principal components, retaining 80% of the energy. The number of the fundamental factors (or IC's) found by FastICA is then equal to the dimension of the projected signals.

Note the strong similarity between the fundamental factors in figures 2 and 3 – the only more visible difference between them is that the first factor in figure 2 is split into the two factors on the two upmost rows of figure 3. A similar phenomenon was found to occur when we experimented with a higher number of ICA's: one factor is split into two, and the remaining factors change very little. The factors are thus rather stable.

The factors have clearly different interpretations. In figure 2, the upmost factor follows the sudden changes that are caused by holidays etc.; the most prominent example is the Christmas time. The factor on the bottom row, on the other hand, reflects the slower seasonal variation, with the effect of the summer holidays clearly visible. The factor on the third row could represent a still slower variation, something resembling a trend. The last factor, on the second row, is different from the others; it might be that this factor follows mostly the relative competitive position of the retail chain with respect to its competitors, but other interpretations are also possible.

Figure 4 shows the residual time series for the five stores, the original data of which is shown in figure 1. In these residual time series, the fundamental factors that are, to a varying degree, common to all the stores are removed by regression from the original time series. If the model works, it is these residual time series that best capture the effect of the actions taken by the management of individual stores.

Finally, in figure 5 the stores are displayed on a Self-Organizing Map (SOM) [3]: the map units are labeled with the store numbers that are mapped to them. The nearer two stores are located on the map, the more they resemble each other in their responses to the fundamental factors. This follows because the map was trained using the rows of the mixing matrix **A**, which reflect each store's response to the fundamental factors, and because the SOM tends to preserve the topological relations between the input vectors. As examples of the cashflow time series in the different parts of the map, see figure 6. Note how the time series of stores 26 and 35 on the two bottom rows appear almost identical to each other and are clearly different from the other two time series; contrast this to the locations of these stores on the SOM in figure 5.

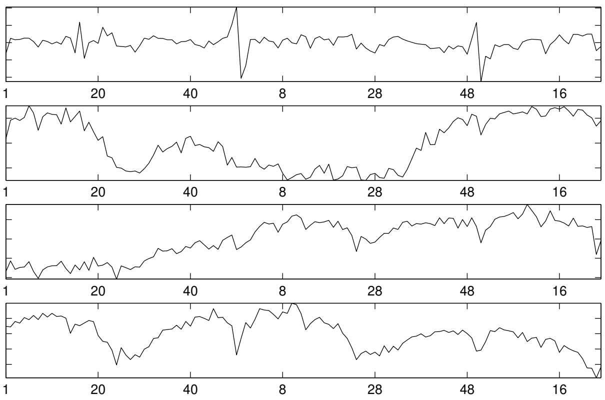

Figure 2: Four fundamental factors found with the ICA.

4. DISCUSSION

We presented a preliminary analysis of a set of parallel time series, trying to isolate the fundamental factors common to all of them, as well as the effect that these factors have on each time series. The Independent Component Analysis (ICA), together with the Self-Organizing Map (SOM) used to find relative similarities between the time series, are both exploratory data analysis methods whose results must be carefully inspected by a domain expert. The ICA method clearly reveals factors that cannot be found by techniques relying on second-order statistics only (say, principal component or factor analysis) and thus may have novel significance. The factors are believed to capture some real effects, and their stability with varying compression rate gives further support to this hypothesis. However, both the factors and the mixing effects must in future works be submitted to further analysis to show their real usefulness.

A possible direction for future research would be to apply suitable filtering on the fundamental factors. Say, the factor that seems to capture seasonal change could be smoothed using a kernel that has a width of 6 weeks, whereas the trend factor would be smoothed using much wider kernel. The seasonal factor might also be averaged over several years. This way our method would lead to the classical time series analysis technique of decomposing the time series to trend, seasonal etc. components, helping to find educated guesses for the length scales of different components. Fitting these filtered fundamental factors to the data, also the grouping obtained with SOM might improve; here different basis functions should probably be given different weights.

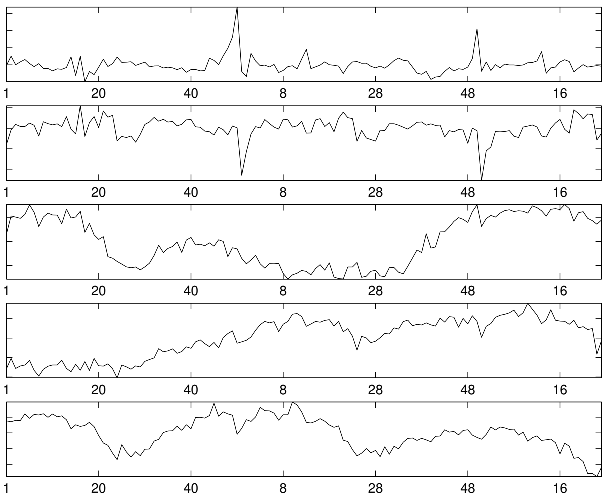

Figure 3: Five fundamental factors found with the ICA.

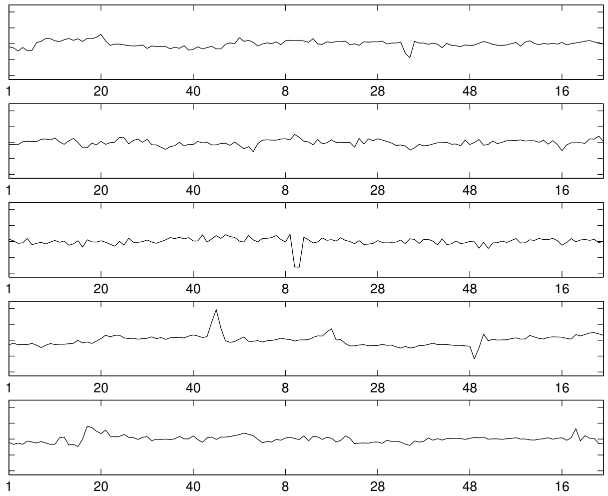

Figure 4: The residual cashflow time series, obtained by subtracting the effect of the five fundamental factors from the original time series (same scale on vertical axis as in Fig. 1).

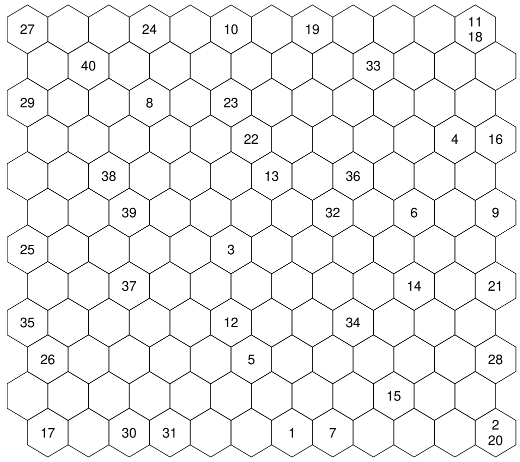

Figure 5: Location of the rows of the mixing matrix **A** on a self-organizing map.

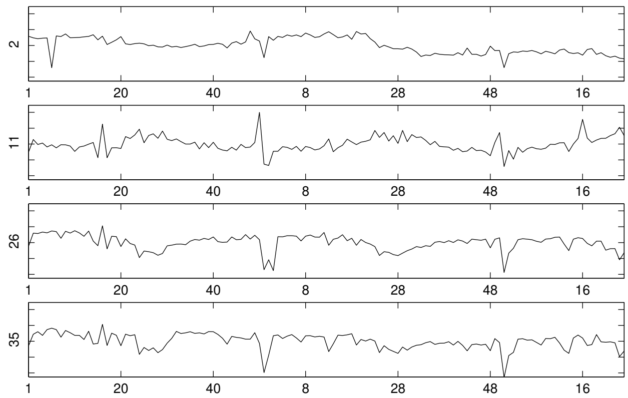

Figure 6: Examples of the cashflow time series mapped to different parts of the SOM; number on the left corresponds to the label on the SOM. Note how the time series 26 and 35 that are close to each other on the map look much more similar to each other than to the other two time series.

5. References

[1] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[2] J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, and E. Oja. Applications of neural blind separation to signal and image processing. In *Proc. 1997 Int. Conf. on Acoustics, Speech, and Signal Proc. ICASSP-97*, Munich, Germany, April 1997.

[3] T. Kohonen. *Self-Organizing Maps.* Springer Series in Information Sciences 30. Springer, Berlin Heidelberg New York, 1995.

[4] J. Moody and L. Wu. What is the "true price"? – State space models for high frequency FX data. In A. S. Weigend, Y. Abu-Mostafa, and A.-P. N. Refenes, editors, *Decision Technologies for financial engineering – proceedings of hte NNCM '96*, volume 7 of *Progress in Neural Processing*, pages 346–358. World Scientific, 1997.