To appear in: *Computational Finance 1997*, The Fifth International Conference on Neural Networks in the Capital Markets. London, Dec. 15–17, 1997.

DISCOVERING STRUCTURE IN FINANCE USING INDEPENDENT COMPONENT ANALYSIS

ANDREW D. BACK
Brain Science Institute
The Institute of Physical and Chemical Research (RIKEN)
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
back@brain.riken.go.jp
www.bip.riken.go.jp/absl/back

ANDREAS S. WEIGEND
Department of Information Systems
Leonard N. Stern School of Business
New York University
44 West Fourth Street, MEC 9-74
New York, NY 10012, USA
aweigend@stern.nyu.edu
www.stern.nyu.edu/∼aweigend

Independent component analysis is a new signal processing technique. In this paper we apply it to a portfolio of Japanese stock price returns over three years of daily data and compare the results obtained using principal component analysis. The results indicate that the independent components fall into two categories, (i) infrequent but large shocks (responsible for the major changes in the stock prices), and (ii) frequent but rather small fluctuations (contributing little to the overall level of the stocks). The small number of major shocks indicate turning points in the time series and when used to reconstruct the stock prices, give good results in terms of morphology. In contrast, when using shocks derived from principal components instead of independent components, the reconstructed price does not show the same results at all. Independent component analysis is shown to be a potentially powerful method of analysing and understanding driving mechanisms in financial time series.

1 Introduction

What drives the movements of a financial time series? In this paper, we focus on a new technique which to our knowledge has not been used in any significant application to financial or econometric problems[1] called *independent component analysis* (ICA)

[1] We are only aware of Baram and Roth (1995) who use a neural network that maximizes output entropy and of Moody and Wu (1996), Moody and Wu (1997a), Moody and Wu (1997b) and Wu and Moody (1997) who apply ICA in the context of state space models for interbank foreign exchange rates to improve the

which 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 (such as daily stock returns) reflects the reaction of a system (such as the stock market) to a few statistically independent time series. ICA provides a mechanism of decomposing a given signal into statistically *independent components* (ICs).

ICA can be expressed in terms of the related concepts of entropy (Bell and Sejnowski 1995), mutual information (Amari, Cichocki and Yang 1996), contrast functions (Comon 1994) and other measures of the statistical independence of signals. For independent signals, the joint probability can be factorized into the product of the marginal probabilities. Therefore the independent components can be found by minimizing the Kullback-Leibler divergence between the joint probability and marginal probabilities of the output signals (Amari et al. 1996).

Independent component analysis can also be contrasted with principal component analysis (PCA) and so we give 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. The goal of PCA is to obtain principal components which are uncorrelated. Moreover, PCA gives projections of the data in the direction of the maximum variance. The principal components (PCs) are ordered in terms of their variances: the first PC defines the direction that captures the maximum variance possible, the second PC defines (in the remaining orthogonal subspace) the direction of maximum variance, and so forth. In ICA however, we seek to obtain statistically independent components.

PCA algorithms use only second order statistical information. On the other hand, ICA algorithms may use higher order[2] statistical information for separating the signals (see for example Cardoso 1989, Comon 1994). For this reason non-Gaussian signals (or at most, one Gaussian signal) are required for ICA. 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.

The goal of this paper is to explore whether ICA can give some indication of the underlying structure of the stock market. The hope is to find interpretable factors of instantaneous stock returns. Such factors could include news (government intervention, natural or man-made disasters, political upheaval), response to very large trades, and, of course, unexplained noise. Ultimately, we hope that this might yield new ways of analyzing and forecasting financial time series, contributing to a better understanding of financial markets.

---

separation between observational noise and the "true price."

[2]ICA algorithms based on second order statistics have also been proposed (Belouchrani, Abed Meraim, Cardoso and Moulines 1997, Tong, Soon, Huang and Liu 1990).

2 ICA in General

2.1 Independent Component Analysis

ICA denotes the process of taking a set of measured signal vectors, x, and extracting from them a (new) set of statistically independent vectors, y, called the independent components or the sources. They are estimates of the original source signals which are assumed to have been mixed in some prescribed manner to form the observed signals.

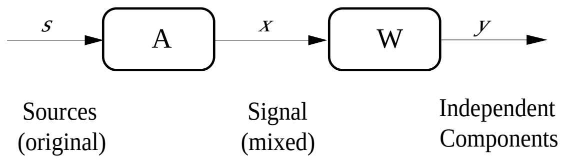


Figure 1: Schematic representation of ICA. The original sources s are mixed through matrix $\mathbf{A}$ to form the observed signal x. The demixing matrix $\mathbf{W}$ transforms the observed signal x into the independent components y.

Figure 1 shows the most basic form of ICA. We use the following notation: We *observe* a multivariate time series $\{x_i(t)\}$, $i = 1, ..., n$, consisting of n values at each time step t. We *assume* that it is the result of a mixing process

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) \quad . \tag{1}$$

Using the instantaneous observation vector $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_n(t)]$, the problem is to find a *demixing matrix* $\mathbf{W}$ such that

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{W}\mathbf{x}(t) \\ &= \mathbf{W}\mathbf{A}\mathbf{s}(t) \end{aligned} \tag{2}$$

where $\mathbf{A}$ is the unknown mixing matrix. We assume throughout this paper that there are as many observed signals as there are sources, hence $\mathbf{A}$ is a square $n \times n$ matrix. If $\mathbf{W} = \mathbf{A}^{-1}$, then $\mathbf{y}(t) = \mathbf{s}(t)$, and perfect separation occurs. In general, it is only possible to find $\mathbf{W}$ such that $\mathbf{WA} = \mathbf{PD}$ where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a diagonal scaling matrix (Tong, Liu, Soon and Huang 1991).

To find such a matrix $\mathbf{W}$, the following assumptions are made:

- The sources $\{s_j(t)\}$ are statistically independent. While it might sound strong, this is not an unreasonable assumption when one considers for example sources of very different origins ranging from foreign politics to microeconomic variables that might impact a stock price.

- At most one source has a Gaussian distribution. In the case of financial data, normally distributed signals are so rare that only allowing for one of them is not a serious restriction.

- The signals are stationary. Stationarity is a standard assumption that enters almost all modeling efforts, not only ICA.

In this paper, we only consider the case when the mixtures occur instantaneously in time. This appears to be a reasonable assumption to make since it implies that at each time instant, the observed stock price is comprised of all currently available information from all sources and is acted on immediately. It is also of interest to consider models based on multichannel blind deconvolution (Jutten, Nguyen Thi, Dijkstra, Vittoz and Caelen 1991, Weinstein, Feder and Oppenheim 1993, Nguyen Thi and Jutten 1995, Torkkola 1996, Yellin and Weinstein 1996, Parra, Spence and de Vries 1997) however we do not do this in the present paper.

2.2 Algorithms for ICA

The earliest ICA algorithm that we are aware of and one which started much interest in the field is that proposed by Herault and Jutten (1986).

Since then, a wide variety of ICA algorithms have been proposed using on-line and batch methods. In addition, various approaches for obtaining the independent components have been proposed, including: minimizing higher order moments (Cardoso 1989) or higher order cumulants (Cardoso and Souloumiac 1993), maximization of mutual information of the outputs or maximization of the output entropy (Bell and Sejnowski 1995), minimization of the Kullback-Leibler divergence between the joint and the product of the marginal distributions of the outputs (Amari et al. 1996).

Neurally inspired algorithms have also been proposed which incorporate a nonlinear function to introduce higher order statistics (see for example Cichocki and Moszczyński 1992, Choi, Liu and Cichocki 1998, Cichocki, Unbehauen and Rummert 1994, Girolami and Fyfe 1997, Hyvärinen 1996, Karhunen 1996, Oja and Karhunen 1995). An important development recently is the field of ICA algorithms which are referred to as *natural gradient* algorthms (Amari et al. 1996, Amari 1998). A similar approach was independently derived by Cardoso and Laheld (1996) who referred to it as a *relative gradient* algorithm. This theoretically sound modification to the usual on-line updating algorithm overcomes the problem of having to perform matrix inversions at each time step and therefore permits significantly faster convergence. Another extension includes contextual ICA (Pearlmutter and Parra 1997) where a method utilizing spatial and temporal information was proposed based on maximum likelihood estimation to separate signals having colored Gaussian distributions or low kurtosis. The ICA framework has also been extended to allow for nonlinear mixing (Burel 1992, Yang, Amari and Cichocki 1997, Yang, Amari and Cichocki 1998, Lin, Grier and Cowan 1997).

A batch ICA algorithm, sometimes referred to as "decorrelation and rotation" (Pope and Bogner 1996), is given by the following two stage procedure (Bogner 1992, Cardoso and Souloumiac 1993).

1. *Decorrelation or whitening.* Here we seek to diagonalize the covariance matrix of the input signals.

2. *Rotation.* The second stage minimizes a measure of the higher order statistics which will ensure the non-Gaussian output signals are as statistically independent as possible. It can be shown that this can be carried out by a unitary rotation matrix (Cardoso and Souloumiac 1993). This second stage provides the higher order independence.

Note that this approach relies on the measured signals being non-Gaussian. For Gaussian signals, the higher order statistics are zero already and so no meaningful separation can be achieved by ICA methods.

The empirical study carried out in this paper uses the JADE (Joint Approximate Diagonalization of Eigenmatrices) algorithm (Cardoso and Souloumiac 1993) which is an efficient batch algorithm computed in stages. The first stage whitens the data by computing the sample covariance matrix, giving the second order statistics of the signals. The second stage consists of finding a rotation matrix which jointly diagonalizes eigenmatrices formed from the fourth order cumulants of the whitened data. The outputs from this stage are the independent components.

3 Analyzing Stock Returns with ICA

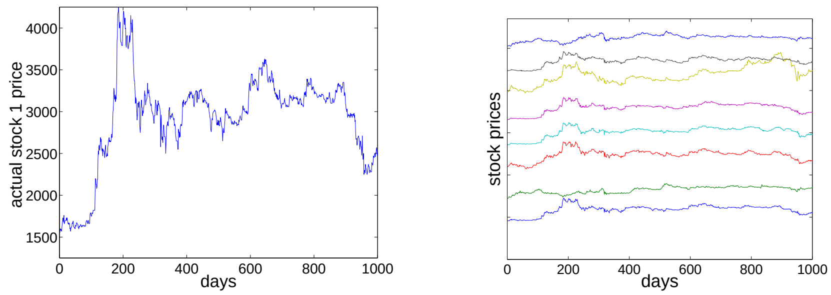
3.1 Description of the Data

To investigate the effectiveness of ICA techniques for financial time series, we apply ICA to data from the Tokyo Stock Exchange. We use daily closing prices from 1986 until 1989[3] of the 28 largest firms.

Figure 2(a) shows the stock price of the first company in our set, the Bank of Tokyo-Mitsubishi, between August 1986 and July 1988.

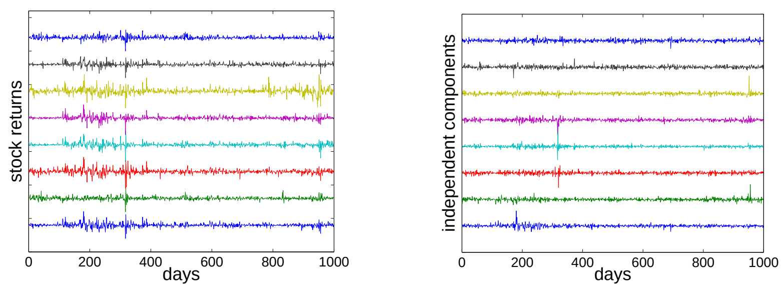
The preprocessing consists of three steps: we obtain the daily stock returns, subtract the mean of each stock, and normalize the resulting values to lie within the range $[-1, 1]$. The stock returns are obtained by taking the difference between successive values of the prices $p(t)$, as $x(t) = p(t) - p(t-1)$. Given the relatively large change in price levels over the few years of data, an alternative would have been to use relative returns, $\log(p(t)) - \log(p(t-1))$, describing geometric growth as opposed to additive growth. Figure 3 shows these normalized stock returns.

[3]We chose a subset of available historical data on which to test the method. This allows us to reserve subsequent data for further experimentation.



(a) (b)

Figure 2: (a) The price of the Bank of Tokyo-Mitsubishi stock for the period 8/86 until 7/88. This bank is one of the largest companies traded on the Tokyo Stock Exchange. (b) The largest eight stocks on the Tokyo Stock Exchange over the same period, offset for clarity. The lowest line displays the price of the Bank of Tokyo-Mitsubishi.



(a) (b)

Figure 3: (a) The stock returns (differenced time series) of the first eight stocks for the period 8/86 until 7/88. The large negative return at day 317 corresponds to the crash of 19 October 1987. The lowest line again corresponds to the Bank of Tokyo-Mitsubishi. The question is: can ICA reveal useful information about these time series? (b) The first eight ICs, resulting from the ICA of all 28 stocks.

3.2 Structure of the Independent Components

We performed ICA on the stock returns using the JADE algorithm (Cardoso and Souloumiac 1993) described in Section 2.2. In all the experiments, we assume that the number of stocks equals the number of sources supplied to the mixing model.

In the results presented here, all 28 stocks are used as inputs in the ICA. However for clarity, the figures only display the first few ICs. Figure 3(b) shows a subset of eight ICs obtained from the algorithm. Note that the goal of statistical independence forces the 1987 crash to be carried by only a few components.

We now present the analysis of a specific stock, the Bank of Tokyo-Mitsubishi. The contributions of the ICs to any given stock can be found as follows.

For a given stock return, there is a corresponding row of the mixing matrix $\mathbf{A}$ used to weight the independent components. By multiplying the corresponding row of $\mathbf{A}$ with the ICs, we obtain the weighted ICs. We define *dominant* ICs to be those ICs with the largest maximum signal amplitudes. They have the largest effect on the reconstructed stock price. In contrast, other criteria, such as the variance, would focus not on the largest value but on the average.

Figure 4(a) weights the ICs with the the first row of the mixing matrix which corresponds to the Bank of Tokyo-Mitsubishi. The four traces at the bottom show the four most dominant ICs for this stock.

From the usual mixing process given by Eq. (1), we can obtain the reconstruction of the ith stock return in terms of the estimated ICs

$$\hat{x}_i(t-j) \quad = \quad \sum_{k=1}^{n} a_{ik} y_k(t-j) \quad j = 0, .., N-1 \tag{3}$$

where $y_k(t-j)$ is the value of the kth estimated IC at time $t-j$ and a_{ik} is the weight in the ith row, kth column of the estimated mixing matrix $\mathbf{A}$ (obtained as the inverse of the demixing matrix $\mathbf{W}$). We define the weighted ICs for the ith observed signal (stock return) as

$$y_{ik}(t-j) \quad = \quad a_{ik} y_k(t-j) \quad k = 1, .., n; j = 0, .., N-1. \tag{4}$$

In this paper, we rank the weighted ICs with respect to the first stock return. Therefore, we multiply the ICs with the first row of the mixing matrix and so we use a_{1k} $k = 1, .., n$ to obtain the weighted ICs. The weighted ICs are then sorted[4] using an L_∞ norm, since we are most interested in showing just those ICs which cause the maximum price change in a particular stock.

The ICs obtained from the stock returns reveal the following aspects:

- 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. The nondominant components do not contribute significantly to level changes.
- 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(b) 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.

[4] ICs can by sorted in various ways. For example, in the implementation of the JADE algorithm Cardoso and Souloumiac (1993) used a Euclidean norm to sort the rows of the demixing matrix $\mathbf{W}$ according to their contribution across all signals.

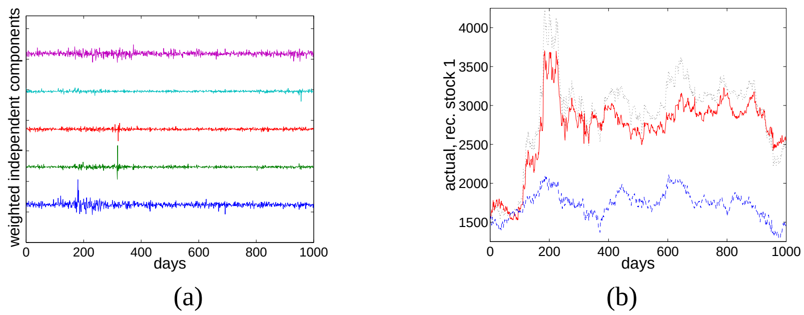


Figure 4: (a) The four most dominant weighted ICs (corresponding to the Bank of Tokyo-Mitsubishi) are shown starting from the bottom trace. The top trace is the summation of the remaining 24 least dominant ICs for this stock. (b) *Top* (dotted) line: original stock price. *Middle* (solid) line: reconstructed stock price using the four most dominant weighted ICs. *Bottom* (dashed) line: reconstructed residual stock price obtained by from remaining 24 weighted ICs. Note that the major part of the true 'shape' comes from the most dominant components; the contribution of the non-dominant ICs to the overall shape is only small.

3.3 Thresholded ICs Characterize Turning Points

The preceding section discussed the effect of a lossy reconstruction of the original prices, obtained by considering the cumulative sums of only the first few dominant ICs. In this section we examine the reconstructed prices using the dominant ICs after they have been thresholded to remove all values smaller than a certain level.

The thresholded reconstructions are described by

$$\bar{x}_i(t-j) = \sum_{k=1}^{n} g\left(\bar{y}_{ik}(t-j)\right) \quad j = 0,..,N-1, \tag{5}$$

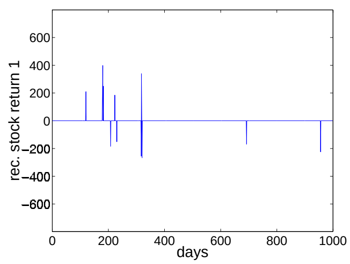
$$g(u) = \begin{cases} u & |u| \geq \xi \\ 0 & |u| < \xi \end{cases} \tag{6}$$

where $\bar{x}_i(t-j)$ are the returns constructed using thresholds, $g(\cdot)$ is the threshold function and ξ is the threshold value. The threshold was set arbitrarily to a value which excluded almost all of the lower level components.

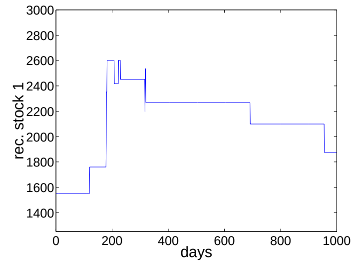
The reconstructed stock prices are found as

$$\begin{aligned} \hat{p}_i(j+1) &= \hat{p}_i(j) + \bar{x}_i(j) \quad j = t-N,..,t-1 \\ \hat{p}_i(t-N) &= p_i(t-N) \end{aligned} \tag{7}$$

For the first stock, the Bank of Tokyo-Mitsubishi, $p_1(t-N) = 1550$.



(a)



(b)

Figure 5: ICA results for the Bank of Tokyo-Mitsubishi: (a) thresholded returns constructed from the four most dominant weighted ICs, (b) reconstructed prices obtained by computing the cumulative sum of only the thresholded values. Note that the price for the 1,000 points plotted is still characterized well by only a few innovations.

The thresholded returns of the four most dominant ICs are shown in Figure 5(a), and the stock price reconstructed from the thresholded return values are shown in Figure 5(b). The figures indicate that the thresholded ICs provide useful morphological information and can extract the turning points of original time series.

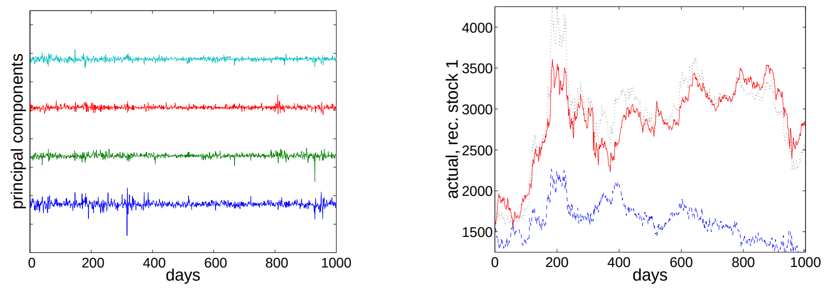
3.4 Comparison with PCA

PCA is a well established tool in finance. Applications range from Arbitrage Pricing Theory and factor models to input selection for multi-currency portfolios (Utans, Holt and Refenes 1997). Here we seek to compare the performance of PCA with ICA using singular value decomposition (SVD).

The results from the PCs obtained from the stock returns reveal the following aspects:

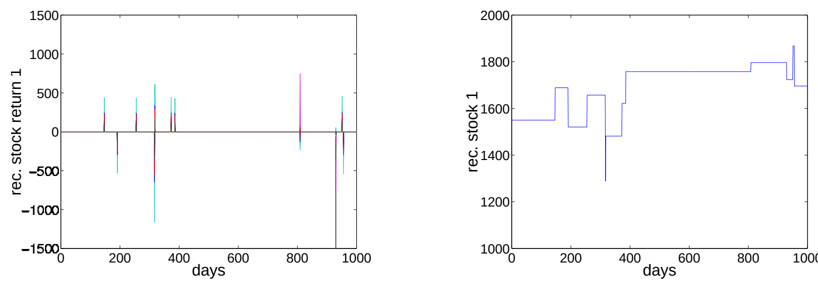
- The distinct shocks which were identified in the ICA case are much more difficult to observe.
- While the first four PCs are by construction the best possible fit in a quadratic error sense to the data, they do not offer the same insight in structure of the data compared to the ICs.
- The dominant transients obtained from the PCs, ie., after thresholding, do not lead to the same overall shape of the stock returns as the ICA approach. Hence we cannot make the same conclusions about high level and low level signals in the data. The effect of thresholding is shown in Figure 7.

For the experiment reported here, the four most dominant PCs are the same, whether ordered in terms of variance or using the L_∞ norm as in the ICA case. Beyond that the orders change.



(a) (b)

Figure 6: For the Bank of Tokyo-Mitsubishi: (a) the four most dominant PCs corresponding to stock returns, (b) *Top* (dotted) line: original stock price. *Middle* (solid) line: reconstructed stock price using the four most dominant PCs. *Bottom* (dashed) line: reconstructed residual stock price obtained by from remaining 24 PCs. The sum of the two lower lines corresponds to the true price. In this case, the error is smaller than that obtained when using ICA. However, the overall shape of the stock is not reconstructed as well by the PCs.



(a) (b)

Figure 7: PCA results for the Bank of Tokyo-Mitsubishi: (a) thresholded returns constructed from the four most dominant PCs, (b) reconstructed prices. In this case, the model does not capture the large transients observed in the ICA case and fails to adequately approximate the shape of the original stock price curve.

Figure 7(b) shows the reconstructed stock price from the thresholded returns are a poor fit to the overall shape of the original price. This implies that key high level transients that were extracted by ICA are not obtained through PCA.

In summary, while PCA also decompose the original data, the PCs do not possess the high order independence obtained of the ICs. A major difference emerges when only the largest shocks of the estimated sources are used. While the cumulative sum up the largest IC shocks retains the overall shape, this is not the case for the PCs.

4 Conclusions

This paper applied independent component analysis (ICA) to decompose a portfolio of 28 instantaneous stock returns into statistically independent components (ICs). The components of the instantaneous vectors of observed daily stock are statistically dependent; stocks on average move together. In contrast, the components of the instantaneous daily vector of ICs are constructed to be statistically independent. This can be viewed as decomposing the returns into statistically independent sources. On three years of daily data from the Tokyo stock exchange, we showed that the estimated ICs fall into two categories, (i) infrequent but large shocks (responsible for the major changes in the stock prices), and (ii) frequent but rather small fluctuations (contributing only little to the overall level of the stocks). The October 1987 crash, for example, is given by only a few ICs in the first group and does not appear in the other group.

We have shown that by using a portfolio of stocks, ICA can reveal some underlying structure in the data. Interestingly, the 'noise' we observe may be attributed to signals within a certain *amplitude* range and not to signals in a certain (usually high) *frequency* range. Thus, ICA gives a fresh perspective to the problem of understanding the mechanisms that influence the stock market data.

In comparison to PCA, ICA is a complimentary tool which allows the underlying structure of the data to be more readily observed. There are clearly many other avenues in which ICA techniques can be applied to finance. Implications to risk management and asset allocation using ICA are explored in Chin and Weigend (1998).

Acknowledgements

We are grateful to Morio Yoda, Nikko Securities, Tokyo for kindly providing the data used in this paper, and to Jean-François Cardoso for making the source code for the JADE algorithm available. Andrew Back acknowledges support of the Frontier Research Program and RIKEN and would like to thank Seungjin Choi and Zhang Liqing for helpful discussions. Andreas Weigend acknowledges support from the National Science Foundation (ECS-9309786) and would like to thank Fei Chen, Elion Chin and Juan Lin for stimulating discussions.

References

Amari, S. (1998). Natural gradient works efficiently in learning, *Neural Computation* **10**(2): 251–276.

Amari, S., Cichocki, A. and Yang, H. (1996). A new learning algorithm for blind signal separation, *in* G. Tesauro, D. Touretzky and T. Leen (eds), *Advances in Neural Information Processing Systems 8 (NIPS*95)*, The MIT Press, Cambridge, MA, pp. 757–763.

Baram, Y. and Roth, Z. (1995). Forecasting by density shaping using neural networks, *Proceedings of the 1995 Conference on Computational Intelligence for Financial Engineering (CIFEr)*, IEEE Service Center, Piscataway, NJ, pp. 57–71.

Bell, A. and Sejnowski, T. (1995). An information maximization approach to blind separation and blind deconvolution, *Neural Computation* **7**: 1129–1159.

Belouchrani, A., Abed Meraim, K., Cardoso, J. and Moulines, E. (1997). A blind source separation technique based on second order statistics, *IEEE Trans. on S.P.* **45**(2): 434–44.

Bogner, R. E. (1992). Blind separation of sources, *Technical Report 4559*, Defence Research Agency, Malvern.

Burel, G. (1992). Blind separation of sources: a nonlinear neural algorithm, *Neural Networks* **5**: 937–947.

Cardoso, J. (1989). Source separation using higher order moments, *International Conference on Acoustics, Speech and Signal Processing*, pp. 2109–2112.

Cardoso, J. and Laheld, B. (1996). Equivariant adaptive source separation, *IEEE Trans. Signal Processing* **44**(12): 3017–3030.

Cardoso, J. and Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals, *IEE Proc. F.* **140**(6): 771–774.

Chin, E. I. A. and Weigend, A. S. (1998). Independent component analysis and stock returns, *Technical report*, Information Systems Department, Leonard N. Stern School of Business, New York University.

Choi, S., Liu, R. and Cichocki, A. (1998). A spurious equilibria-free learning algorithm for the blind separation of non-zero skewness signals, *Neural Processing Letters* **7**: 1–8.

Cichocki, A. and Moszczyński, L. (1992). New learning algorithm for blind separation of sources, *Electronics Letters* **28**(21): 1986–1987.

Cichocki, A., Unbehauen, R. and Rummert, E. (1994). Robust learning algorithm for blind separation of signals, *Electronics Letters* **30**(17): 1386–1387.

Comon, P. (1994). Independent component analysis – a new concept?, *Signal Processing* **36**(3): 287–314.

Girolami, M. and Fyfe, C. (1997). An extended exploratory projection pursuit network with linear and nonlinear anti-hebbian connections applied to the cocktail party problem, *Neural Networks* **10**(9): 1607–1618.

Herault, J. and Jutten, C. (1986). Space or time adaptive signal processing by neural network models, *in* J. S. Denker (ed.), *Neural Networks for Computing. Proceedings of AIP Conference*, American Institute of Physics, New York, pp. 206–211.

Hyvärinen, A. (1996). Simple one-unit algorithms for blind source separation and blind deconvolution, *Progress in Neural Information Processing ICONIP'96*, Vol. 2, Springer, pp. 1201–1206.

Jutten, C. and Herault, J. (1991). Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing* **24**: 1–10.

Jutten, C., Nguyen Thi, H., Dijkstra, E., Vittoz, E. and Caelen, J. (1991). Blind separation of sources, an algorithm for separation of convolutive mixtures, *Proceedings of Int. Workshop on High Order Statistics*, Chamrousse (France), pp. 273–276.

Karhunen, J. (1996). Neural approaches to independent component analysis and source separation, *Proceedings of 4th European Symp. on Artificial Neural Networks (ESANN'96)*, Bruges, Belgium, pp. 249–266.

Lin, J. K., Grier, D. G. and Cowan, J. D. (1997). Faithful representation of separable distributions, *Neural Computation* **9**: 1305–1320.

Moody, J. E. and Wu, L. (1996). What is the "true price"? – State space models for high frequency financial data, *Progress in Neural Information Processing (ICONIP'96)*, Springer, Berlin, pp. 697–704.

Moody, J. E. and Wu, L. (1997a). What is the "true price"? – State space models for high frequency FX data, *in* A. S. Weigend, Y. S. Abu-Mostafa and A.-P. N. Refenes (eds), *Decision Technologies for Financial Engineering (Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets, NNCM-96)*, World Scientific, Singapore, pp. 346–358.

Moody, J. E. and Wu, L. (1997b). What is the "true price"? – State space models for high frequency FX data, *Proceedings of the IEEE/IAFE 1997 Conference on Computational Intelligence for Financial Engineering (CIFEr)*, IEEE Service Center, Piscataway, NJ, pp. 150–156.

Nguyen Thi, H.-L. and Jutten, C. (1995). Blind source separation for convolutive mixtures, *Signal Processing* **45**(2): 209–229.

Oja, E. and Karhunen, J. (1995). Signal separation by nonlinear hebbian learning, *in* M. Palaniswami, Y. Attikiouzel, R. Marks II, D. Fogel and T. Fukuda (eds), *Computational Intelligence - A Dynamic System Perspective*, IEEE Press, New York, NY, pp. 83–97.

Parra, L., Spence, C. and de Vries, B. (1997). Convolutive source separation and signal modeling with ML, *International Symposium on Intelligent Systems (ISIS'97)*, University of Reggio Calabria, Italy.

Pearlmutter, B. A. and Parra, L. C. (1997). Maximum likelihood blind source separation: A context-sensitive generalization of ICA, *in* M. C. Mozer, M. I. Jordan and T. Petsche (eds), *Advances in Neural Information Processing Systems 9 (NIPS*96)*, MIT Press, Cambridge, MA, pp. 613–619.

Pope, K. and Bogner, R. (1996). Blind signal separation. I: Linear, instantaneous combinations, *Digital Signal Processing* **6**: 5–16.

Tong, L., Liu, R., Soon, V. and Huang, Y. (1991). Indeterminacy and identifiability of blind identification, *IEEE Trans. Circuits, Syst.* **38**(5): 499–509.

Tong, L., Soon, V. C., Huang, Y. F. and Liu, R. (1990). AMUSE: A new blind identification algorithm, *International Conference on Acoustics, Speech and Signal Processing*, pp. 1784–1787.

Torkkola, K. (1996). Blind separation of convolved sources based on information maximization, *in* S. Usui, Y. Tohkura, S. Katagiri and E. Wilson (eds), *Proc. of the 1996 IEEE Workshop Neural Networks for Signal Processing 6 (NNSP96)*, IEEE Press, New York, NY, pp. 423–432.

Utans, J., Holt, W. T. and Refenes, A. N. (1997). Principal component analysis for modeling multi-currency portfolios, *in* A. S. Weigend, Y. S. Abu-Mostafa and A.-P. N. Refenes (eds), *Decision Technologies for Financial Engineering (Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets, NNCM-96)*, World Scientific, Singapore, pp. 359–368.

Weinstein, E., Feder, M. and Oppenheim, A. (1993). Multi-channel signal separation by de-correlation, *IEEE Trans. Speech and Audio Processing* **1**(10): 405–413.

Wu, L. and Moody, J. (1997). Multi-effect decompositions for financial data modelling, *in* M. C. Mozer, M. I. Jordan and T. Petsche (eds), *Advances in Neural Information Processing Systems 9 (NIPS*96)*, MIT Press, Cambridge, MA, pp. 995–1001.

Yang, H., Amari, S. and Cichocki, A. (1998). Information-theoretic approach to blind separation of sources in non-linear mixture, *Signal Processing* **64**(3): 291–300.

Yang, H. H., Amari, S. and Cichocki, A. (1997). Information back-propagation for blind separation of sources in non-linear mixtures, *IEEE International Conference on Neural Networks, Houston TX (ICNN'97)*, IEEE-Press, pp. 2141–2146.

Yellin, D. and Weinstein, E. (1996). Multichannel signal separation: Methods and analysis, *IEEE Transactions on Signal Processing* **44**: 106–118.