In Intelligent Data Engineering and Automated Learning - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents, ed. K.S.Leung, L.W.Chan and H.Meng, Springer, pages 538-544, 2000.

Applying Independent Component Analysis to Factor Model in Finance

Siu-Ming CHA and Lai-Wan CHAN

Computer Science and Engineering Department
The Chinese University of Hong Kong, Shatin, HONG KONG
Email : {smcha,lwchan}@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/{~smcha,~lwchan}

Abstract. Factor model is a very useful and popular model in finance. In this paper, we show the relation between factor model and blind source separation, and we propose to use Independent Component Analysis (ICA) as a data mining tool to construct the underlying factors and hence obtain the corresponding sensitivities for the factor model.

1 Introduction

Factor model is a fundamental model in finance. Many financial theories are established based on it, for examples, Modern Portfolio Theory and Arbitrage Pricing Theory(APT). These theories assume that the returns of securities are represented as linear combinations of some factors. Modern Portfolio Theory aims at analyzing the composition of securities in the portfolio and relates the return and risk of the portfolio with the security returns and risks [20]. Factor model serves as an efficient and common model for the return generating process [21, 24, 17]. Furthermore, factor model is also the foundation of Arbitrage Pricing Theory (APT) [5, 22]. APT plays an important role in modern finance and it analyses the capital asset pricing in finance [9, 10].

Factor model relates the returns of securities to a set of factors. The factors can be system (market) factors or non-system (individual) factors. Finding the factors for the model is a challenge but not an easy task to researchers, as the factors are hidden and not necessary directly related to the fundamental factors, such as GDP, interest rate[12]. In this paper, we apply independent component analysis (ICA), a modern signal processing method, to recover the hidden factors and the corresponding sensitivities. Section 2 and 3 review the backgrounds of factor model and ICA. We apply ICA to factor model in section 4. Section 5 contains the experiment and results.

2 Factor model in finance

Multifactor model is a general form of factor model [2, 9, 21], and is the most popular model for the return generating process. The return r_i on the ith security is represented as,

$$r_i = \alpha_i + \sum_{m=1}^{k} \beta_{im} F_m + u_i \qquad (1)$$

where k is the number of factors and it is a positive integer larger than zero, $F_1, F_2, ..., F_k$ are the factors affecting the returns of ith security and β_{i1}, β_{i2}, ..., β_{ik} are the corresponding sensitivities. α_i is regarded as "zero" factor that is invariant with time; u_i is a zero mean random variable of ith security. It is generally assumed that the covariance between u_i and factors F_i are zero. Also u_i and u_j for security i and j are independent if $i \neq j$.

The simplest factor model is one-factor model, i.e., $k = 1$. One-factor model with market index as the factor variable is called market model. However, factor model does not restrict the factor to be the market index. Investigators use different approaches in factor model [19,6]. The first one assumes some known fundamental factors are the factors that influence the security and β's are evaluated accordingly. The second approach assumes the sensitivities to factors are known, and the factors are estimated from the security returns [12]. The third approach is factor analysis. This one assumes neither factor values nor the security sensitivities is known. Under factor analysis approach, principle component analysis(PCA) was the most successful method [11,23,25]. PCA was used to find the factors and their sensitivities[2,8]. However it was also shown that the separated factors are not able to truly reflect the real case but only one meaningful factor, which corresponds to the market effect, is extracted. This is due to two limitations of PCA. First, the separated principal components must be orthogonal to each other. Second, PCA uses only up to second order statistics, i.e. the covariance and correlation matrix. In this paper, we apply ICA to factor model because ICA does not have those limitations PCA has. More importantly, ICA is able to reflect the underlying structures of securities[1,18].

3 Independent Component Analysis

Blind source separation(BSS), a well-known problem, aims at recovering the sources from a set of observations. Applications include separating individual voices in cocktail party. In BSS problem, it contains two processes. They are the mixing process and demixing process. First, we observe a set of multivariate signals $x_i(t), i = 1, 2, ..., n$, that are assumed to be linearly mixed with a set of source signals. The mixing process is hidden so we can only observe the mixed signals. The task is to recover the original source signals from the observations through a demixing process. Equation 2 and 3 describe the mixing and demixing processes mathematically.

$$\text{Mixing:} \qquad x = As \tag{2}$$

$$\text{Demixing:} \qquad y = Wx \tag{3}$$

Each signal x_i is a t time steps series, *i.e.* $x_i = [x_i(1), x_i(2), ..., x_i(t)]$; x is the $[n \times t]$ observation matrix, *i.e.* $x = [x_1, x_2, ..., x_n]'$. In BSS problem, we assume the number of observations is equal to the number of source signals. Matrix s contains the original source signals driving the observations whereas the separated signals are stored in matrix y. They are both $[n \times t]$ matrices. A and W

are both $[n \times n]$ matrices, called mixing and demixing matrix respectively. If the separated signals are the same as the original sources, the mixing matrix is the inverse of demixing matrix, *i.e.* $A = W^{-1}$.

BSS is a difficult task because we do not have any information about the sources and the mixing process. ICA is a method tackling this problem by assuming that the sources are independent to each other[16], and finds the demixing matrix W and corresponding independent signals y from the observations x with some criteria making the separated signals as independent as possible. Various ICA algorithms have been proposed. Most of them use higher order statistics to obtain the independent components, e.g. [13, 7, 15, 14, 3] and [4] etc.

4 ICA and Factor model

4.1 Relationships between BSS and Factor Model

Previous works have been done on using ICA to extract components for stocks [1]. However, the independent components have never been related to the factor models. By relating the independent components to the factor model, we hope that this technique can be used in future applications of the factor model. In this section, we illustrate the application of ICA in factor model. Both of them assume the observations are under driven by a set of factors (or sources). We firstly zero mean the return as

$$r_i - E[r_i] = \sum_{m=1}^{k} \beta_{im}\{F_m - E[F_m]\} + u_i \tag{4}$$

We put $R_i = r_i - E[r_i]$ and $F'_m = F_m - E[F_m]$. Without loss of generality, we treat the noise term, u_i, as an extra factor, i.e. $u_i = \beta_{ik+1}F'_{k+1}$

$$R_i = \sum_{m=1}^{k+1} \beta_{im} F'_m \tag{5}$$

The above is a typical mixing process of observations in blind source separation problem. The factor models are under transformed to mixing matrix and factor series. After the transformation, we can apply ICA to separate the sources (or factors).

4.2 Procedures of finding factors by ICA

Here we show the procedures of finding the factors for factor model using ICA.

1. Select securities' price series as observations. We transform the security prices to returns *i.e.* $r_i(t) = (p_i(t) - p_i(t-1))/p_i(t-1)$ and making the return series zero mean *i.e.* $R_i = r_i - E[r_i]$.

2. Perform independent component separation on the return series R_i.

3. Sort the independent signals with their importance. Importance of a signal can be measured by its L_∞ [1].

4. Select the number of independent signals according to the requirements of factor model. The rest of separated signals are regarded as residuals.

5. Evaluate the sensitivities to the factors using the mixing matrix.

The separated independent signals and the corresponding sensitivities are obtained from the above procedures. Hence the factor model is constructed using the observable security movements. We will demonstrate this in the experiment.

4.3 Remarks of applying ICA to find the factors

From above, the expected return of each security, $E[r_i]$, is equal to the sum of factor means and zero factor. There is no information about zero factor given to the ICA algorithm during decomposition, because we cancelled the zero factor while subtracting the mean of each observation signal as in equation 4. As a result, we cannot separate the zero factor from the observations.[1] However we can retain the original pricing level of each security by adding its expect value $E[r_i]$ to the factor model.

5 Experiments and Results

In the experiment, we used 7 stocks, selected from the Hang Seng Index constitutes. Daily closing prices started form 2/1/1992 to 23/8/2000 were used(Figure 1).

In the experiment, we reconstruct the multifactor model of each stock using the procedures in section 4.2. Figure 2 shows the separated signals. Starting from top to bottom, the top most signals is the most important hidden factor, F_1', and so on, the last signal is named as F_7'.

We reconstruct the factor models with six hidden factors, $F_1', F_2', ..., F_6'$ where the least important factor F_7' is regarded as residual. The mixing matrix found is shown as below

$$\begin{bmatrix} 0.0145 & -0.0119 & -0.0034 & 0.0055 & 0.0027 & 0.0138 & -0.0059 \\ 0.0071 & -0.0169 & -0.0009 & 0.0067 & 0.0019 & -0.0016 & -0.0018 \\ 0.0072 & -0.0137 & -0.0014 & 0.0001 & 0.0154 & 0.0031 & -0.0048 \\ 0.0095 & -0.0137 & -0.0195 & 0.0016 & 0.0041 & 0.0053 & -0.0051 \\ 0.0056 & -0.0180 & -0.0014 & -0.0022 & -0.0002 & 0.0122 & -0.0117 \\ 0.0166 & -0.0105 & -0.0070 & 0.0035 & 0.0038 & 0.0020 & -0.0154 \\ 0.0222 & -0.0158 & -0.0058 & -0.0085 & 0.0014 & 0.0037 & -0.0016 \end{bmatrix}$$

The rows in the mixing matrix are the corresponding sensitivities to the hidden factors for the stock. To reconstruct the factor model, we take stock 1 as an example. Equations 6 and 7 show its return expressed as a 6-factor model and 3 factor model respectively.

$$R_1(t) = 0.0145 \times F_1'(t) - 0.0119 \times F_2'(t) - 0.0034 \times F_3'(t) + 0.0055 \times F_4'(t)$$

[1] It is also a common practice to assume the expected values of the factors are zero. In that case, the zero factor can be obtained.

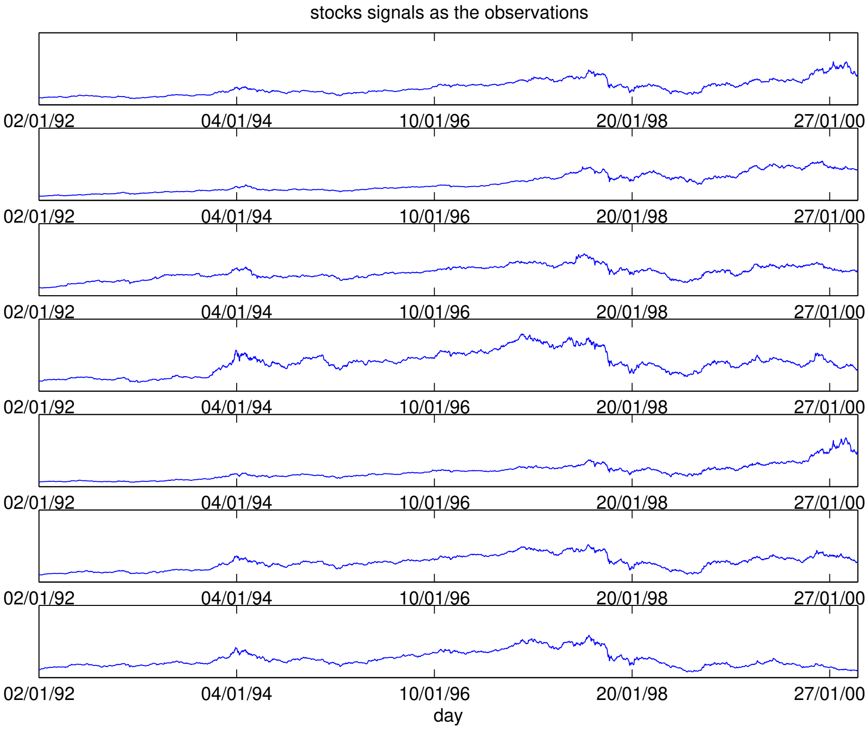

Fig. 1. Seven stocks' series in the experiment

$$+0.0027 \times F_5'(t) + 0.0138 \times F_6'(t) + u_1 \quad (6)$$

$$where \quad u_1 = -0.0059 \times F_7'(t)$$

$$R_1(t) = 0.0145 \times F_1'(t) - 0.0119 \times F_2'(t) - 0.0034 \times F_3'(t) + v_1 \quad (7)$$

$$where \quad v_1 = 0.0055 \times F_4'(t) + 0.0027 \times F_5'(t) + 0.0138 \times F_6'(t) - 0.0059 \times F_7'(t)$$

To express the return as in the factor model, we simply add the expected returns to R_i as $r_i = R_i + E[r_i]$.

6 Discussions and Conclusion

In this paper, we propose to apply independent component analysis (ICA) to extract the factors and the sensitivities of securities in the factor model. In some traditional applications of factor models, the returns are related to some systematic factors or macro-economic variables; for examples, unexpected changes in the rate of inflation and the rate of return on a treasury bill. On one hand, it is useful to know what the exact underlying factors are. On the other hand, the financial market nowaday is extremely complex and dynamic, especially due

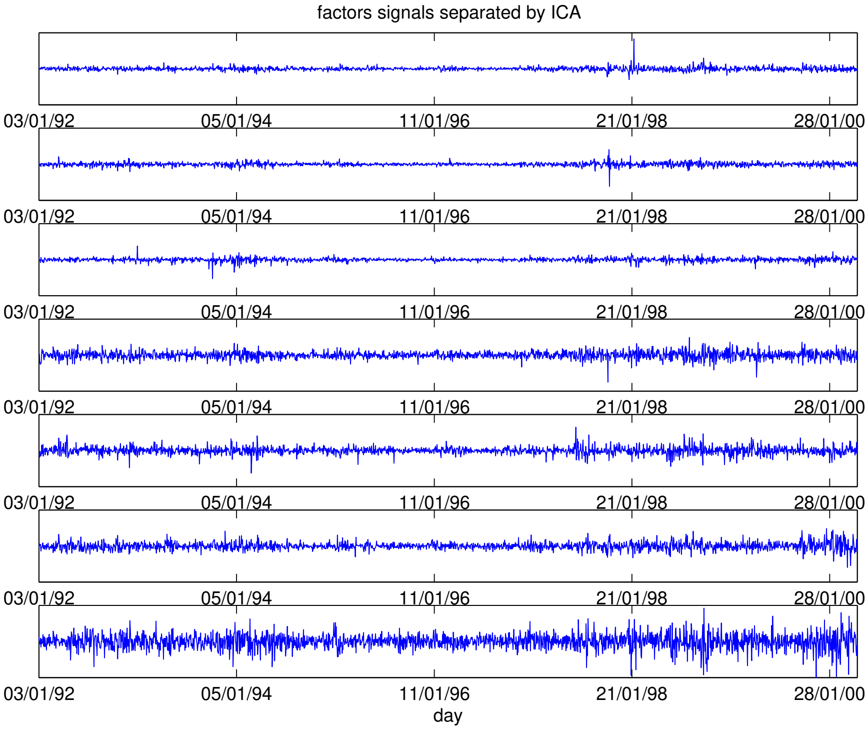

Fig. 2. The separated signals are sorted with their importance. (The y-axes of the sub-figures do not have equal scales.) The uppermost signal is regarded as the most important signal F_1' and so on.

to globalization and many newly introduced indices, such as IT index, it is not an easy task to decide which variables, among so many systematic factors and macro-economic variables, should be included in the model as factors. Our method serves as a data mining technique to automatically identify the hidden factors from historical data. Though attempts can be made to correlate the factors extracted to some known variables, it is still possible to apply these factor models in many aspects in finance. For example, we can perform risk analysis and construct portfolios which are less sensitive to the hidden factors.

Acknowledgement The authors would like to thank The Research Grants Council, HK for support.

References

1. A. Back and A. Weigend. A first application of independent component analysis to extracting structure from stock returns. *Journal of Neural Systems*, 8:473–484, 1997.

2. S. Brown. The number of factors in security returns. *The Journal of Finance*, 44(5):1247–1262, December 1989.
3. J.F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.
4. J.F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. In *IEE Proc-F.140(6)*, pages 771–774, 1993.
5. G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, September 1983.
6. R. Chen, N.F. Roll and S. Ross. Economic forces and the stock market. *Journal of Business*, 59(3):383–403, July 1986.
7. P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36:287–314, April 1994.
8. G. Connor and R. Korajczyk. Performance measurement with the arbitrage pricing theory a new framework for analysis. *Journal of financial economics*, 15:373–394, 1986.
9. G. Connor and R. Korajczyk. A test for the number of factors in an approximate factor model. *The Journal of Finance*, 48(4):1263–1291, September 1993.
10. F. Fabozzi. *Investment Management.* Prentice Hall International, Inc, 1995.
11. G. Feeney and D. Hester. Stock market indices: A principal component analysis. *Cowles Foundation, Monograph 19*, volume 39:110–138, 1967.
12. A. Gordon, W. Sharp, and B. Jeffery. *Fundamentals of investments.* Englewood Cliffs, N.J. : Prentice Hall, second edition, 1993.
13. J. Heradult and C. Jutten. Space or time adaptive signal processing by neural network models. In *Neural Networks for Computing, Proceeding of AIP Conference*, pages 211–206, New York, 1986. American Institute of Physics.
14. A. Hyvärinen. Independent component analysis by minimization of mutual information. Technical report, Helsinki University of Technology, Laboratory of Computer and Information Science, August 1997.
15. A. Hyvärinen and E. Oja. Independent component analysis by general nonlinear hebbian-like learning rules. *Signal Processing*, 64(3):301–313, 1998.
16. A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.
17. B. King. Market and industry factors in stock price behavior. *Journal of Business*, 39:139–190, 1966.
18. R. Lesch, Y. Caille, and D. Lowe. Component analysis in financial time series. In *Proceedings of the IEEE/IAFE 1999*, pages 183–190, 1999.
19. B. Manly. *Multivariate statistical methods: A primer.* Chapman and Hall, 1994.
20. H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, March 1952.
21. H. Markowitz. *Portfolio selection, efficient diversification of investment.* Blackwell Publishers Ltd, 108 Cowley Road Oxford OX4 1JF, UK, second edition, 1991.
22. S. Ross. A arbitrage theory of the capital asset pricing. *Journal of Economic Theory*, 3:343–362, 1976.
23. H. Schneeweiss and H. Mathes. Factor analysis and principal components. *Journal of multivariate analysis*, 55:105–124, 1995.
24. W. Sharp. A simplified model for portfolio selection. In *Management Science*, volume 9, pages 277–293, 1963.
25. J. Utans, W.T. Holt, and A.N. Refenes. Principal components analysis for modeling multi-currency portfolios. In *Proceedings of the Fourth International Conference on Neurals Networks in the Capital Markets, NNCM-96*, pages 359–368. World Scientific, 1997.