ELSEVIER

A novel algorithm for dynamic factor analysis

Jih-Jeng Huang [a], Gwo-Hshiung Tzeng [b,c,*],
Chorng-Shyong Ong [a]

[a] *Department of Information Management, National Taiwan University, Taipei, Taiwan, ROC*
[b] *Institute of Management of Technology, National Chiao Tung University,*
1001 Ta-Hsuch Road, Hsinchu 300, Taiwan, ROC
[c] *Department of Business Administration, Kainan University, Taoyuan, Taiwan, ROC*

Abstract

In this paper, a dynamic factor model is proposed to extract the dynamic factors from time series data. In order to deal with the problem of scaling, the cross-correlation matrices (CCM) are first employed to cluster the time series data. Then, the dynamic factors are extracted using the revised independent component analysis (ICA). In addition, a numerical study is used to demonstrate the proposed method. On the basis of the simulated results, we can conclude that the proposed method can really extract the effective dynamic factors.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Dynamic factor model; Factor analysis; Cross-correlation matrices (CCM); Independent component analysis (ICA); Time series

* Corresponding author. Address: Institute of Management of Technology, National Chiao Tung University, 1001 Ta-Hsuch Road, Hsinchu 300, Taiwan, ROC.
  E-mail address: u5460637@ms16.hinet.net (G.-H. Tzeng).

1. Introduction

Dynamic factor analysis (DFA), which was proposed by Engle and Watson [1,2], is a dimension-reduction approach for extracting the common trends of time series data. The mathematical formulation of DFA can be described as follows. Let the multivariate time series vector at time t be $\mathbf{y}_t$. Then, the dynamic factor model can be formulated as

$$\mathbf{y}_t = \mathbf{\Gamma}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{e}_t, \tag{2}$$

where $\mathbf{\Gamma}$ denotes the factor loading, $\boldsymbol{\alpha}_t \sim N(\mathbf{a}_t, \mathbf{v}_t)$ is the common trends at time t, $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_\varepsilon)$ is the noise component matrix, and $\mathbf{e}_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_e)$ is the diagonal error covariance matrix. In addition, $\boldsymbol{\alpha}_t$, $\boldsymbol{\varepsilon}_t$, and $\mathbf{e}_t$ are independent of each other.

Although DFA has been successfully used in the applications of economics [3–6] and psychology [7,8], two main problems should be considered for adopting DFA in practice. First, the computational cost of estimating parameters in DFA is usually heavy. Several papers have been reported that DFA can only be suitable for small scaling time series data [9,10]. Although several algorithms such as Markov chain Monte Carlo method [11,12], and EM algorithm [9,10] have been proposed to deal with the problem above, these methods cannot truly overcome the problem of scaling. Second, the conventional DFA only extract the linear common trends among time series data using the second-order statistics. However, the information of the high-order statistics should also be considered to response the complex systems in practice.

In this paper, a novel algorithm is proposed to deal with the problems above simultaneously. First, in order to overcome the problem of scaling, the cross-correlation matrices (CCM) [13] are used to cluster time series variables into segments. Next, the revised independent component analysis (ICA) is proposed to extract the dynamic factors by different segments. Sixteen daily indices of stock markets and foreign currency exchange rates from 1995 to 1997 are used to demonstrate the proposed method. In addition, the dynamic factors are used to predict the daily indices and compare with the dynamic regression model. On the basis of the simulated results, we can conclude that the proposed method can really extract the important common trends among time series data and performs the accurate prediction.

The remainder of this paper is organized as follows. The dynamic factor model is proposed in Section 2. A numerical example, which is used to illustrate the proposed method and compare with the dynamic regression model, is presented in Section 3. Discussion and conclusions are in the last section.

2. Dynamic factor model

In order to derive the dynamic factors, the CCM [13] is first employed to calculate the correlation of the multivariate time series so that we can cluster the variables into several segments to reduce the computational cost. Next, the dynamic factors can be derived using the ICA approach.

2.1. Cross-correlation matrices

Consider the multivariate time series $\boldsymbol{Z}_t$, and the mean vector $\boldsymbol{\mu}$, then the cross-covariance matrices at the lth lag can be defined as

$$
\boldsymbol{\Sigma}_l = \mathrm{Cov}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t-l}) = E[(\boldsymbol{Z}_t - \boldsymbol{\mu})(\boldsymbol{Z}_{t-l} - \boldsymbol{\mu})']
$$

$$
= E \begin{bmatrix} z_{1t} - \mu_1 \\ z_{2t} - \mu_2 \\ \vdots \\ z_{kt} - \mu_k \end{bmatrix} [z_{1(t-l)} - \mu_1, z_{2(t-l)} - \mu_2, \ldots, z_{k(t-l)} - \mu_k]
$$

$$
= \begin{bmatrix} v_{11}(l) & v_{12}(l) & \cdots & v_{1k}(l) \\ v_{21}(l) & v_{22}(l) & \cdots & v_{2k}(l) \\ \vdots & \vdots & \cdots & \vdots \\ v_{k1}(l) & v_{k2}(l) & \cdots & v_{kk}(l) \end{bmatrix}.
$$

(3)

On the basis of the cross-covariance matrices, we can obtain the CCM as follows:

$$
\boldsymbol{P}_l = \begin{bmatrix} \rho_{11}(l) & \rho_{12}(l) & \cdots & \rho_{1k}(l) \\ \rho_{21}(l) & \rho_{22}(l) & \cdots & \rho_{2k}(l) \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{k1}(l) & \rho_{k2}(l) & \cdots & \rho_{kk}(l) \end{bmatrix},
$$

(4)

where

$$
\rho_{ij}(l) = \frac{v_{ij}(l)}{[v_{ii}(l)v_{jj}(l)]^{1/2}}.
$$

(5)

By detecting the coefficients of the CCM, we can cluster the correlated time series variables into several segments. Next, we can introduce the ICA method and present how the dynamic factors can be obtained using ICA.
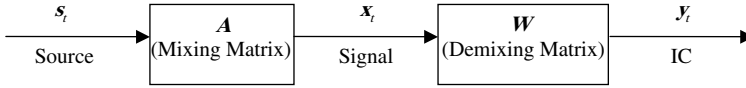
Fig. 1. The concept of ICA.

2.2. Independent component analysis

ICA [14,15] is a statistical tool to extract the independent component (IC) from an observed multivariate time series. ICA has been proposed to deal with many real-world applications such as signal processing [16,17], magnetoencephalography (MEG) [18], and image analysis [19,20]. The concepts of ICA can be described as follows. Let a time signal vector be $x_t = \{x_1, x_2, \ldots, x_n\}$, the ICA model can be formulated as

$$x_t = As_t, \tag{6}$$

where A denotes the unknown mixing matrix and s denotes the sources.

The problem of ICA is to extract the IC vector, y_t, from the signal vector, s_t. We can depict the problem above as shown in Fig. 1.

In order to derive the ICs, we can calculate the demixing matrix, W, such that

$$y_t = Wx_t = WAs_t. \tag{7}$$

Therefore, if we can find $W = A^{-1}$, then $y_t = s_t$, and the perfect separation occurs. It should be highlighted that the conventional ICA only deal with the random variables and cannot handle the time series data. In this paper, a revised ICA, which was proposed in [21,22], is proposed to deal with non-stationary and temporally correlated data.

In addition, although ICA and principal component analysis (PCA) share some common characteristics like building generative model and performing dimension reduction, PCA only process the second-order dependencies in the data. However, ICA is a generation of PCA that separates the higher-order dependencies in the data. In addition, conventional PCA can only deal with the random variable data instead of the time series data. We can depict Fig. 2 to present the proposed algorithm as follows.

In the next section, a numerical study is used to demonstrate the proposed method.

3. Numerical study

In this section, 16 daily indices of stock markets and foreign currency exchange rates from 1995 to 1997, including Amsterdam, Frankfurt, Hong Kong,
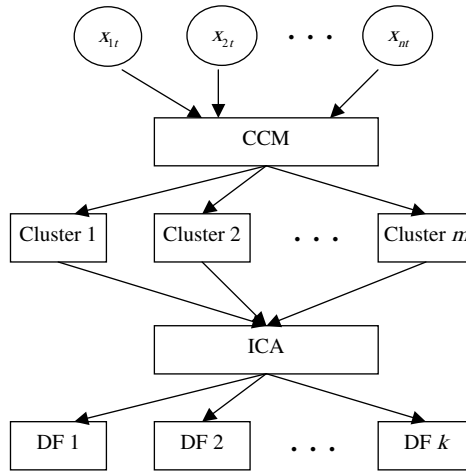
Fig. 2. The procedures of the proposed algorithm.

London, New York, Paris, Singapore, Tokyo, and so on, are used to extract the dynamic factors. These daily indices can be represented using Fig. 3.

In order to cluster the indices above to reduce the computational cost, the CCM is used to calculate the correlation among indices as shown in Table 2. On the basis of the CCM, we can cluster these indices into three segments as shown in Table 1 (Table 2).

Next, we can extract the dynamic factors from the segments using ICA. Since the cluster 1 contains many indices, we extract two dynamic factors from cluster 1 as shown in Fig. 4. However, only one dynamic factor is extracted from clusters 2 and 3 as shown in Figs. 5 and 6.

Next, the dynamic regression (DR) model is employed to test the efficiency of the proposed method. First, we select six variables to be the dependent variable and the other 15 variables are used to predict the dependent variable in the six dynamic regression models. Next, we use the same dependent variables but the dynamic factors to be the independent variables in other six dynamic regression models. Finally, we use Akaike information criterion (AIC), Hannan-Quinn criterion (HQC), corrected AIC (AICC), and Schwarz Bayesian criterion (SBC) to compare the proposed method with the dynamic regression model as shown in Table 3.

On the basis of the simulated results, we can conclude that the dynamic factor model performs almost the same accuracy with the dynamic regression model. It indicates the dynamic factors can really be extracted and reflect the common trends of the multivariate time series. Next, we provide the depth discussion according to our implementation.
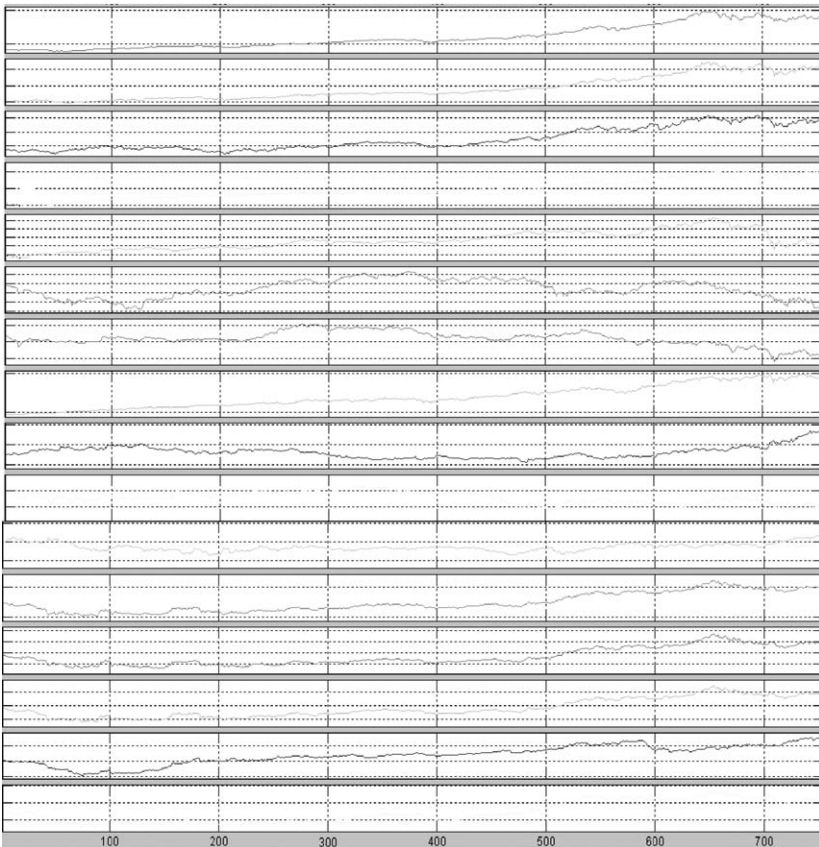
Fig. 3. The trend chart of the 16 daily indices.

Table 1
Cluster for multivariate time series

Cluster for multivariate time series	
Cluster 1	AMSTEOE, DAXINDX, FRCAC40, FTSE100, HNGKNGI, SPCOMP, DTCHGUS, FRNFRUS, GERMDUS, JAPYNUS, SWISFUS
Cluster 2	JAPDOWA, AUSTRUS, CDNDLUS
Cluster 3	SNGALLS, BRITPUS

4. Discussion and conclusions

Dynamic factor analysis is a useful tool for extracting the common trends among time series data. The dynamic factors are useful for the decision-maker.

Table 2
The cross-correlation matrix at the first lag

| Variable	AMSTEOE	DAXINDX	FRCAC40	FTSEI00	HNGKNGI	JAPDOWA	SNGALLS	SPCOMP	AUSTRUS	BRITPUS	CDNDLUS	DTCHGUS	FRNFRUS	GERMDUS	JAPYNUS	SWISFUS
Cross-correlation matrix																
AMSTEOE	1.000	0.997	0.984	0.980	0.820	0.048	-0.497	0.982	0.219	-0.582	0.326	0.964	0.946	0.963	0.836	0.918
DAXINDX	0.997	1.000	0.984	0.977	0.803	0.017	-0.521	0.975	0.256	-0.581	0.359	0.966	0.953	0.965	0.827	0.918
FRCAC40	0.984	0.984	1.000	0.957	0.799	-0.027	-0.508	0.956	0.243	-0.619	0.340	0.963	0.953	0.962	0.798	0.938
FTSEI00	0.980	0.977	0.957	1.000	0.815	0.060	-0.475	0.991	0.226	-0.531	0.253	0.918	0.889	0.917	0.851	0.860
HNGKNGI	0.820	0.803	0.799	0.815	1.000	0.365	-0.081	0.826	-0.215	-0.413	-0.093	0.759	0.730	0.761	0.714	0.765
JAPDOWA	0.048	0.017	-0.027	0.060	0.365	1.000	0.599	0.091	-0.739	0.400	-0.336	0.041	-0.022	0.048	0.269	0.021
SNGALLS	-0.497	-0.521	-0.508	-0.475	-0.081	0.599	1.000	-0.449	-0.650	0.648	-0.525	-0.490	-0.520	-0.482	-0.248	-0.435
SPCOMP	0.982	0.975	0.956	0.991	0.826	0.091	-0.449	1.000	0.197	-0.564	0.248	0.919	0.887	0.918	0.865	0.867
AUSTRUS	0.219	0.256	0.243	0.226	-0.215	-0.739	-0.650	0.197	1.000	-0.295	0.531	0.190	0.238	0.184	-0.062	0.094
BRITPUS	-0.582	-0.581	-0.619	-0.531	-0.413	0.400	0.648	-0.564	-0.295	1.000	-0.264	-0.541	-0.548	-0.534	-0.412	-0.597
CDNDLUS	0.326	0.359	0.340	0.253	-0.093	-0.336	-0.525	0.248	0.531	-0.264	1.000	0.423	0.463	0.419	0.223	0.374
DTCHGUS	0.964	0.966	0.963	0.918	0.759	0.041	-0.490	0.919	0.190	-0.541	0.423	1.000	0.993	1.000	0.849	0.976
FRNFRUS	0.946	0.953	0.953	0.889	0.730	-0.022	-0.520	0.887	0.238	-0.548	0.463	0.993	1.000	0.993	0.797	0.971
GERMDUS	0.963	0.965	0.962	0.917	0.761	0.048	-0.482	0.918	0.184	-0.534	0.419	1.000	0.993	1.000	0.850	0.976
JAPYNUS	0.836	0.827	0.798	0.851	0.714	0.269	-0.248	0.865	-0.062	-0.412	0.223	0.849	0.797	0.850	1.000	0.834
SWISFUS	0.918	0.918	0.938	0.860	0.765	0.021	-0.435	0.867	0.094	-0.597	0.374	0.976	0.971	0.976	0.834	1.000
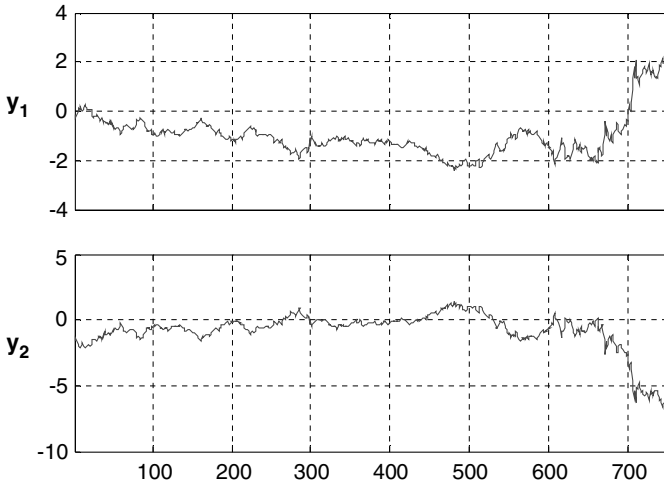
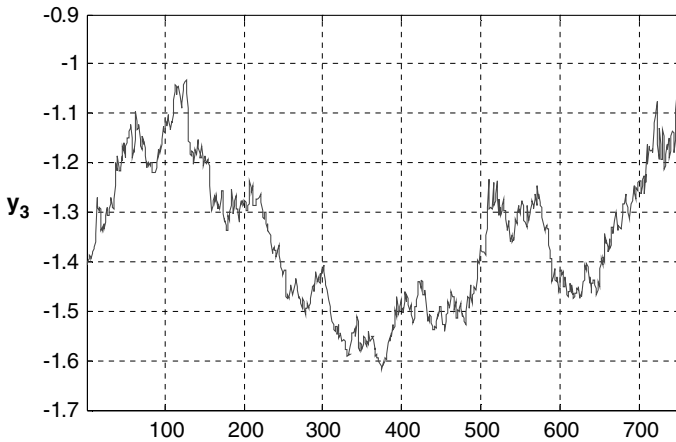Fig. 4. The first and second dynamic factors derived form cluster 1.

Fig. 5. The third dynamic factor derived form cluster 2.

For example, by extracting the important factors, the decision-maker can understand the changing trends of the future and effectively manage the strategic planning.

In this paper, the 16 daily indices of stock markets and foreign currency exchange rates are used to extract the dynamic factors. Since the computational cost of dynamic factor analysis is heavy, the CCM is first used to cluster the 16 indices into three segments. Next, the dynamic factors are extracted by each cluster. In the first cluster, two dynamic factors are extracted. From the shape
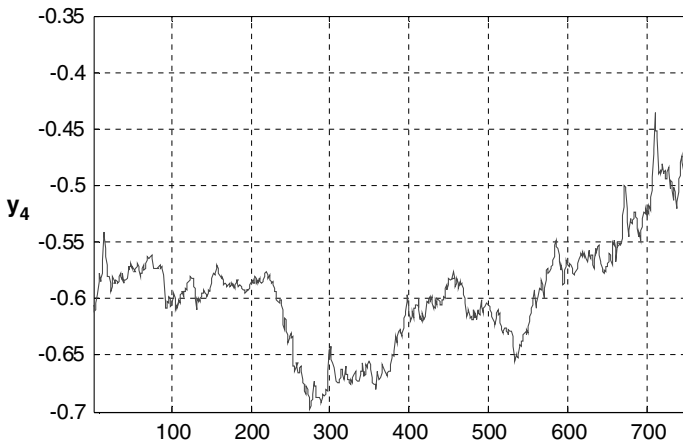
Fig. 6. The fourth dynamic factor derived form cluster 3.

Table 3
The comparison of the dynamic regression model and the proposed method

	Dependent	Independent	AIC	HQC	AICC	SBC
DFA	AMSTEOE	F1 and F2	4.2023	4.2094	4.2023	4.2207
DR	AMSTEOE	Others	4.1956	4.2335	4.1966	4.2939
DFA	JAPDOWA	F3	15.1852	15.1899	15.1852	15.1975
DR	JAPDOWA	Others	13.2089	13.2467	13.2098	13.3071
DFA	SNGALLS	F4	−8.6634	−8.6587	−8.6634	−8.6511
DR	SNGALLS	Others	3.0958	3.1336	3.0967	3.1940
DFA	DAXINDX	F1 and F2	7.2486	7.2557	7.2487	7.2671
DR	DAXINDX	Others	7.2401	7.2779	7.2410	7.3383
DFA	AUSTRUS	F3	−5.7006	−5.6959	−5.7006	−5.6884
DR	AUSTRUS	Others	−10.0310	−9.9955	−10.0302	−9.9389
DFA	BRITPUS	F4	−11.5517	−11.5469	−11.5516	−11.5394
DR	BRITPUS	Others	−11.5466	−11.5087	−11.5456	−11.4483

of the first-two dynamic factors, it can be seen that the direction of the two factors are opposite. It can be interpreted that the two opposite forces control the indices of cluster 1. On the other hand, the second and the third dynamic factors which are extracted from cluster 2 and cluster 3 seem reflect the short-term and the long-term cycle trends.

In addition, we use the dynamic factors to predict the daily indices and compare with the dynamic regression model. On the basis of Table 3, it can be seen that the dynamic factors can perform the accurate prediction. That is, the proposed method can really extract the important common trends among time series data. Finally, the problem of scaling can be overcome using the proposed method.

References

[1] R.F. Engle, M.W. Watson, A one-factor multivariate time series model of metropolitan wage rates, Journal of the American Statistical Association 76 (376) (1981) 774–781.

[2] M.W. Watson, R.F. Engle, Alternative algorithms for estimation of dynamic MIMIC, factor, and time varying coefficient regression models, Journal of Econometrics 23 (3) (1983) 385–400.

[3] A.C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge, Cambridge University Press, 1989.

[4] A.W. Gregory, A.C. Head, Common and country-specific fluctuations in productivity, investment, and the current account, Journal of Monetary Economics 44 (3) (1999) 423–451.

[5] A. Gregory, A. Head, J. Raynauld, Measuring world business cycles, International-Economic-Review 38 (3) (1997) 677–701.

[6] S.C. Norrbin, D.E. Schlagenhauf, The role of international factors in the business cycle: a multi-country study, Journal of International Economics 40 (1–2) (1996) 85–104.

[7] P.C.M. Molenaar, Dynamic factor analysis for the analysis of multivariate time series, Psychometrika 50 (1) (1985) 181–202.

[8] P.C.M. Molenaar, J.G. de Gooijer, B. Schmitz, Dynamic factor analysis of nonstationary multivariate time series, Psychometrika 57 (3) (1992) 333–349.

[9] A.F. Zuur, R.J. Fryer, I.T. Jolliffe, R. Dekker, J.J. Beukema, Estimating common trends in multivariate time series using dynamic factor analysis, Environmetrics 14 (7) (2003) 665–685.

[10] A.F. Zuur, I.D. Tuck, N. Bailey, Dynamic factor analysis to estimate common trends in fisheries time series, Canadian Journal of Fisheries and Aquatic Sciences 60 (5) (2003) 542–552.

[11] O. Aguilar, G. Huerta, R. Prado, M. West, Bayesian inference on latent structure in time series, Bayesian Statistics 6 (1) (1998) 1–16.

[12] M. West, P.J. Harrison, Bayesian Forecasting and Dynamic Models, New York, Springer-Verlag, 1997.

[13] G.C. Tiao, R.S. Tsay, Multiple time series modeling and extended sample cross correlations, Journal of Business and Economic Statistics 1 (1) (1983) 43–56.

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

[15] P. Common, Independent component analysis—a new concept? Signal Processing 36 (3) (1994) 287–314.

[16] A. Bell, T. Sejnowski, An information—maximization approach to blind separation and blind deconvolution, Neural Computation 7 (6) (1995) 1129–1159.

[17] S. Ikeda, N. Murata, A method of ICA in time frequency domain, in: Proceedings of International workshop on Independent Component Analysis and Blind Signal Separation, Aussois, France, 1999, pp. 365–370.

[18] R. Vigario, Extraction of ocular artifacts from EEG using independent component analysis, Electroencephalography and Clinical Neurophysiology 103 (3) (1997) 395–404.

[19] A. Bell, T. Sejnowski, The 'independent components' of natural scenes are edge filters, Vision Research 37 (23) (1997) 3327–3338.

[20] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607–609.

[21] S. Choi, A. Cichocki, Blind separation of nonstationary and temporally correlated sources from noisy mixtures, in: Proceeding of NEEE NNSP, Sydney, Australia, 2000, pp. 405–414.

[22] J.V. Stone, Blind source separation using temporal predictability, Neural Computation 13 (7) (2001) 1559–1574.