A Spectral Decomposition Approach To Separating Independent Factors: The Case Of Foreign Exchange Rates

Sorin-Manuel Delureanu Ph. D Student
University of Craiova
Faculty of Economics and Business Administration
Craiova, Romania

Abstract: In this paper, Independent Component Analysis (ICA) is addressed for revealing fundamental factors behind 6 parallel series of foreign exchange rates. ICA is belonging to the blind source separation (BSS) area of research. Each measured signal is considered a mixture of several distinct underlying factors. Separating such fundamental causal factors is very important for multivariate financial time series analysis, in order to explain past co-movements and to predict future evolutions.

JEL classification: C14, C22, C38, C53.

Key words: blind source separation; independent component analysis; financial time series critical

1. Introduction

Almost every signal measured within a real-world system is actually a mixture of statistically independent source signals. Independent component analysis (ICA) is a statistical technique which aims at decomposing the combination of a set of random variables into statistically independent components (ICs). In the literature, most of the research focuses on applying ICA in engineering fields, while applications of ICA in finance are quite scarse. Because ICA is able to model multivariate data with non-normal distributions such as financial time series data, there are great opportunities of applying ICA in finance. Indeed, financial markets are driven by a few latent common factors. Many financial times series evolve in parallel, which is known as co-movement. Currency exchange rates or daily stock returns are good examples.

The concept of ICA is close related to Bind Source Separation (BSS) [15]. BSS is a term used to describe the method of separation of underlying source signals from a set of observed signal mixtures with little or no information as to the nature of those source signals. ICA is just one of many solutions to the BSS problem.

The property of signals to be unrelated can be captured in terms of statistical independence. Fundamental factors are assumed to be not only uncorrelated with each other, but also unconditionally independent. They are supposed to typically follow non-Gaussian distributions. The nongaussianity is a common feature for a large number of financial time series. Identifying independent factors behind financial data might help in minimizing the risk in the investment strategy. Many researches ([4],[6],[7],16]) showed that most financial time series data (e.g., daily returns series of stocks) follow non-Gaussian distributions, such as fat-tails. In other words, ICA models would be very suitable in modeling so-called non-Gaussian distributed financial data.

ICA separates a set of signal mixtures into a corresponding set of statistically independent component (source) signals. One of the first applications of ICA was the cocktail party problem, where the simultaneous sound signals in the room are separated from the mixture sound data. It considers a room full of people speaking simultaneously. Microphones (equal to the number of people in the room) are scattered throughout the room, and each microphone records a mixture of all the voices in the room. The problem, then, is to separate the voices of the individual speakers using only the recorded mixtures of their voices. ICA tries to recover the original independent components by maximizing the statistical independence of the estimated sources.

ICA is related to conventional methods for analyzing large data sets, such as Principal Component Analysis (PCA) and Factor Analysis (FA), which are classical examples of linear transformation methods proposed for finding projections of the data that have interesting structure. PCA (Hotelling (1933)) and FA (Spearman (1904)) can be seen as dimension reduction techniques that transform the original data (highly correlated) in a set of a few underlying components that are maximally uncorrelated. Both methods compute the components of interest by using only the information contained in the data covariance matrix. Although PCA and FA are very related, they are not identical. PCA takes into account all variability in the variables and the principal components are computed by maximizing the amount of total explained variability. The main difference between ICA and PCA is that ICA not only decorrelates the signals up to second-order statistics, but also reduce the dependency up to higher-order statistics, such as kurtosis. So, ICA requires stronger conditions than PCA.

2. DATA PREPROCESSING

In order to simplify the estimation of Independent Components, there are some data preprocessing methods. One is the data centering. Without loss of generality, we can first perform centering in the mixture data, i.e., we subtract the sample mean of the observable variables xi, making the model zero-mean.

Let us consider a random vector $x = (x_1 \ldots x_N)^T$ with N elements. There is available a sample $x(1), \ldots, x(T)$ from this random vector. The most basic and necessary preprocessing is to center the observed vector $x = (x_1 \ldots x_N)^T$, i.e., substract its mean vector $m_x = E\{x\}$ so as to transform x into a zero mean vector *y*.

After centering, another useful preprocessing strategy is to perform a pre-whitening of the data. The most common way to do this is to assume that each independent component has unit variance. This means that the already centered vector *y* is linearly transformed into a new vector z, which is white (spherical), i.e., its covariance matrix equals the identity matrix: $\Sigma_z = E\{z\,z^T\} = I$.

Whitening uses the eigenvalue decomposition (EVD) of the covariance matrix of *y*, which is computed from the available sample. Performing its EVD, we obtain $\Sigma_y = E\{y\,y^T\} = EDE^T$, where *E* is the orthogonal matrix of eigenvectors, and *D* is the diagonal matrix of its eigenvalues. Whitening can now be done by $z = D^{-1/2}E^T y$.

After these preprocessing steps, the independent components estimated by ICA algorithms will have zero mean and unit variance. Note that the sign of an independent component may be uncertain. However, multiplying an independent component by -1 will not affect the model and thus the problem is solved.

3. APPROACHES TO EXTRACTING INDEPENDENT FACTORS FROM SIGNAL MIXTURES

3.1 Independence versus Uncorrelatedness

Two random variables y1 and y2 are said to be independent if information on the values of y1 does not give any information on the values of y2, and vice versa. Mathematically, statistical independence can be defined by the density functions of the random variables. Assume that p(y1, y2) denotes the joint probability density function (PDF) of y1 and y2. The marginal PDFs are p(y1) and p(y2), respectively. So, y1 and y2 are said to be independent if and only if the joint PDF is the product of the marginal PDFs:

$$p(y1, y2) = p(y1) \cdot p(y2) \qquad (1)$$

A similar concept that has wide applications in finance (e.g., asset allocation) is uncorrelation. Uncorrelation is a weaker form of independence. If the covariance of two random variables y1 and y2 is zero, then y1 and y2 are said to be uncorrelated:

$$E[y1\cdot y2] - E[y1]\cdot E[y2] = 0 \qquad (2)$$

If the variables are independent, it suggests that they are uncorrelated. However, the inverse relationship does not hold. Since independence implies uncorrelatedness, independent components estimated by ICA algorithms are always un-correlated estimates. It also means that the correlations among the ICs should be zero, which is a very desirable property in financial applications. PCA seeks sources that are Gaussian and uncorrelated, rather than the non-Gaussian, independent sources of ICA. As uncorrelatedness is not as strong a property as independence, PCA is not as robust as ICA, but is less computationally challenging. Additionally, PCA can be utilized as a precursor to the ICA algorithm

3.2 The Basic ICA Model

While we focus on linear mixtures, the already centered and whitened vector $z_t \in \Re^N$ can be expressed as:

$$z_t = A s_t, \quad t = 1, 2, \ldots, T \qquad (3)$$

where $s_t \in \Re^M$ is the unknown source vector, with M$\leq$N, and $A \in \Re^{N\times M}$ the unknown mixing matrix. The objective of BSS is to recover the source vector realizations $\{s_t\}$ from the corresponding realizations of the observed vector $\{z_t\}$. This means to find a matrix W that makes the components of z_t statistically independent. We call such a matrix a separating matrix. The process involves, explicitly or implicitly, the inversion of matrix A, which is thus assumed to be full column rank. The separator output is computed as $\hat{s}_t = W z_t$, where $W \in \Re^{M\times N}$ is the separating matrix. A perfect separating matrix is such that the global filter $G = W A$ can be written as the product of a non-singular diagonal matrix D and a permutation matrix P, i.e., $G = PD$. The diagonal elements of matrix D account for the unknown scales of the separated sources, whereas the permutation matrix P reflects the fact that the sources can appear in any order at the separator output. Such source scale and permutation indeterminacies are typically admissible in BSS/ICA.

A solution to the problem of estimating a separating matrix W that will give us the independent components $\hat{s}_i$, can be found based on the optimization of a cost function, also called contrast function.

3.3 Alternative Approaches to Source Separation

In 1995, A.J. Bell and T.J. Sejnowski developed an "information-maximization approach to blind separation and blind deconvolution," which is now referred to as Infomax [1]. Infomax is a method of finding mutually independent signals by maximizing information-flow, or entropy. Infomax yields the same result as another ICA method, Maximum Likelihood Estimation (MLE) ([14], [17]). MLE optimizes parameter values to find the best fit of observed data given some model (MLE optimizes W to find the best fit of the extracted signals y to the source signals s) ([14], [17]). Both Infomax and MLE make several assumptions about the source signals.

In Infomax and MLE, as well as PCA and projection pursuit, optimization is achieved by gradient ascent. Gradient ascent is an optimization method that maximizes a function of multiple parameters by iteratively improving an initial guess using the gradient, which points in the direction of maximum slope [17]. A disadvantage of this method is that if the step size, or learning rate, is not chosen carefully, the function does not converge properly. Additionally, the Gradient Ascent method only finds a local maximum, so if the initial starting value is closer to a local maximum than the global maximum, the algorithm does not converge to find the maximum function value. Either of these situations can produce erroneous results [14].

FastICA is a more advanced approach to ICA and describes a fixed point algorithm that can be applied (in lieu of gradient ascent) to perform more efficient calculations [14].

3.4 The FastICA Algorithm

FastICA is a computationally efficient algorithm for performing ICA estimations by optimizing orthogonal contrast functions of the form

$$\Upsilon_G(W) = \sum_{i=1}^{N} E\left\{G_i\left(w_i^T z\right)\right\} \qquad (4)$$

where the data z are prewhitened and matrix W is constrained to be orthogonal. The functions $G_i(\cdot)$ can fundamentally be chosen in two ways:

- In an maximum likelihood framework, $G_i(\hat{s}_i) = \log p_{s_i}(\hat{s}_i)$ where p_{s_i} is the pdf of the i-th independent component. The functions $G_i(\cdot)$ can be fixed based on prior information of the independent components, or they can be estimated from the current estimates $\hat{s}_i$. In many applications, the independent components are known to be super-Gaussian (sparse), in which case all the functions $G_i(\cdot)$ are often set a priori as

$$G_i(\hat{s}) = -\log\cosh\hat{s} \qquad (5)$$

This corresponds to a symmetric super-Gaussian pdf and has the great benefit of being very smooth; a non-smooth $G_i(\cdot)$ can pose problems for any optimization method. Maximum likelihood estimation is equivalent, under some assumptions, to information-theoretic approaches such as minimum mutual information and minimum marginal entropy.

• A second framework consists in using cumulants. Consider the simple case of maximizing the kurtosis $kurt(\hat{s}) = E\{\hat{x}^4\} - 3\left(E\{\hat{s}^2\}\right)^2$ of the estimated signals $\hat{s}$. Because we assume that the estimated signals have unit variance, due to prewhitening and the orthogonality of W, this reduces to maximizing the fourth moment $E\{\hat{s}^4\}$, thus $G_i(\hat{s}) = \hat{s}^4$. This means that the contrast is essentially the sum of kurtosis. If we know a priori that the independent components are super-Gaussian, the maximization of this contrast function is a valid estimation method. If the sources are all sub-Gaussian, we switch the sign and choose $G_i(\hat{s}) = -\hat{s}^4$, which becomes again a valid contrast.

Function $G_i(\cdot)$ is implicitly expressed in the algorithm via its derivative, a nonlinearity $g_i(\cdot)$; that corresponding to $G_i(\hat{s}) = -\log\cosh\hat{s}$ is the hyperbolic tangent function ($\tanh$).

FastICA is an efficient fixed point based algorithm. For a given sample of the prewhitened vectors z, an iteration is defined as follows:

1. Take a random initial vector w_0 of norm 1. Let $k = 1$.

2. Let $w_k = E\left\{z\left(w_{k-1}^T z\right)^3\right\} - 3w_{k-1}$. The expectation can be calculated using a sample of z vectors.

3. Divide w_k by its norm, i.e. $w_k = w_k / \|w_k\|$.

4. If $\|w_k - w_{k-1}\|$ is not small enough, let $k = k+1$ and go back to step 2. Otherwise, output the vector w_k.

The final vector w_k given by the algorithm separates one of the underlying factors in the sense that $w_k^T x'(t)$, $t = 1, 2, \ldots$ equals one of the factors. To separate m independent factors, we run this algorithm m times. To ensure that we separate each time a different factor, we use the deflation algorithm that adds a simple projection inside the loop. Recalling that the rows of the separating matrix W are orthonormal because of the prewhitening (sphering), we can separate the factors one-by-one by projecting the current estimate w_k on the space orthogonal to the rows of the separating matrix W previously found.

4. An Application to Separating and Remixing Independent Factors: the Case of Foreign Exchange Rates

The proposed approach starts by considering 6 parallel series of foreign exchange rates and is based on the assumption that there are some independent components, such as macroeconomic fundamentals, extraneous influences, interest rates, psychological factors, and so on, which are closely tied to the evolution of the currencies and affect their co-movement. The ICA transformation tends to produce component signals that are more structured and regular. We actually use the FastICA algorithm for performing ICA estimations.

Fig. 1 shows the 6 original foreign exchange series (mixtures, before preprocessing). As we easily can see, the time series are clearly non-stationary and present some trends.

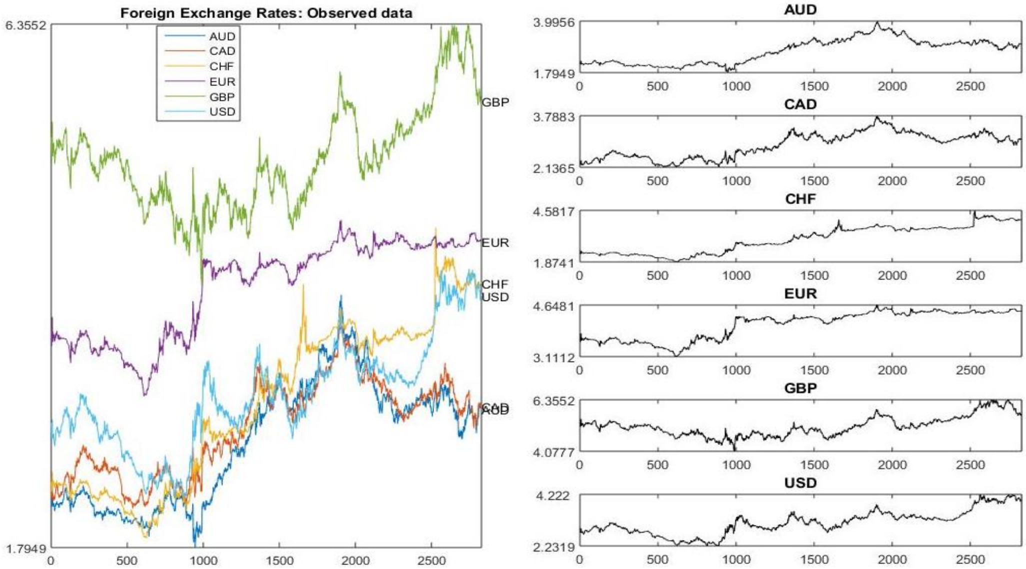

Figure no. 1. Original foreign exchange series (mixtures)

Whitening uses the eigenvalue decomposition (EVD) of the covariance matrix of centered data. Fig. 2 shows the 6 eigenvalues. The contribution in percents of the first 3 eigenvalues is: $\lambda_1 : 81.233\%$; $\lambda_2 : 12.822\%$; $\lambda_3 : 3.6637\%$. Their cumulate contribution is 97.718%.

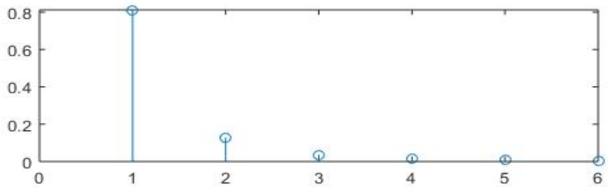

Figure no. 2 . Eigenvalues

Fig. 3 (left) shows the 6 series, after centering and whitening. Clearly, after preprocessing, the time series become stationary. The independent components are shown in Fig. 3 (right). As one can see, they have very distinct shapes. By contrast with the original time series, which display similar shapes, due to common influences affecting their co-movement, one can see that the independent components have very distinctive shapes, corresponding to mutually different factors.

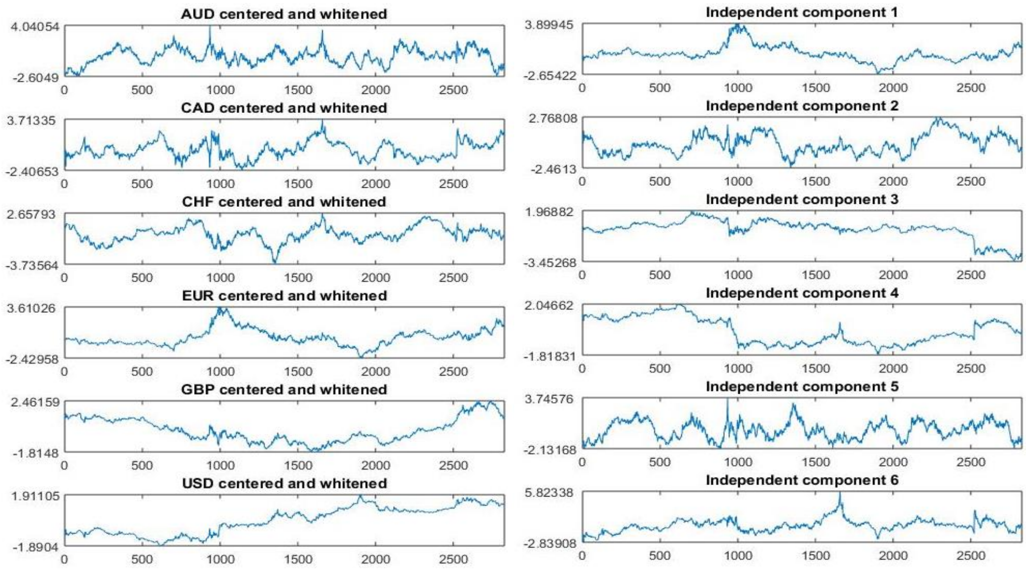

Figure no. 3. Left: centered and whitened series; Right: independent components

The mixing and demixing matrices, A and W, respectively, are:

```
A =
  -0.54094 -0.055219   -0.3443  -0.73965  0.019696   0.19486
  -0.42744 -0.074556  -0.36568  -0.76257   0.25508   0.17632
  -0.31087   0.20377  -0.67217  -0.58318  0.087596    0.2488
  -0.17283   0.27728  -0.46228  -0.81058  0.093636   0.11571
  -0.49572   0.19465  -0.76299  0.022738   0.31023  -0.19265
    -0.154  0.064991  -0.83028  -0.45774   0.26248 -0.062598
W =
   -1.3454   0.08516   0.30186   0.10641   -1.2825    1.3953
   -1.5103  -0.16708   0.56489    2.4816    1.2969   -2.3308
  -0.60363    1.2071  -0.98589    1.0452   0.37247   -1.6118
  -0.66725  0.042184    1.3388   -1.3148    0.4754  -0.53058
   -4.0188    4.4673   0.28116  -0.04916   0.74165   -1.1828
   -2.2239    2.0304    4.3095   -2.1403  -0.80511   -1.5293
```

Now, let us reconstruct the original mixtures starting from the independent component estimates $\hat{s}$ and using the mixing matrix A, i.e., $\hat{z} = A\hat{s}$. Aditionally, we can apply to $\hat{z}$ the de-whitening matrix $E\,D^{1/2}$ in order to reconstruct the centered data, i.e., $\hat{y} = E\,D^{1/2}\,\hat{z}$, and finally to add the mean, i.e. $\hat{x} = \hat{y} + m_x$.

Fig. 4 shows the remixed times series. Our experiments show that the reconstruction is absolutely exact for all these time series, i.e., $\hat{z} = z$, $\hat{y} = y$ and finally $\hat{x} = x$.

Another goal could be that of obtaining a smooth reconstruction of the original time series, via denoising the independent components. Indeed, the independent components we have separated are really more structured and regular, representing distinctive influence sources. However, they are somehow noisy. By denoising them, we expect to obtain a smooth reconstruction of the original time series.

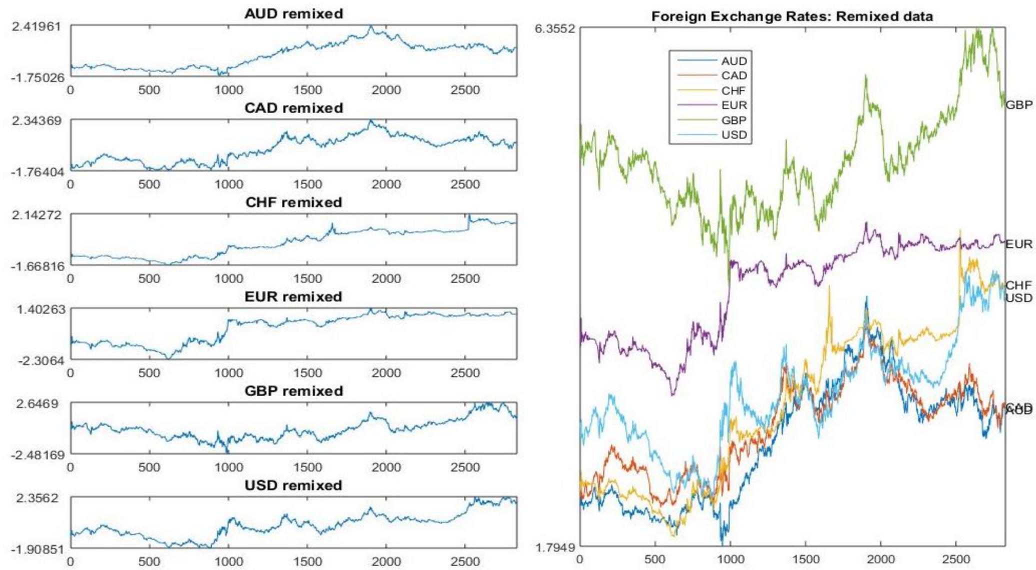

Figure no. 4. Remixed foreign exchange series

Fig. 5 shows the smooth representation of the independent components and remixed foreign exchange series.

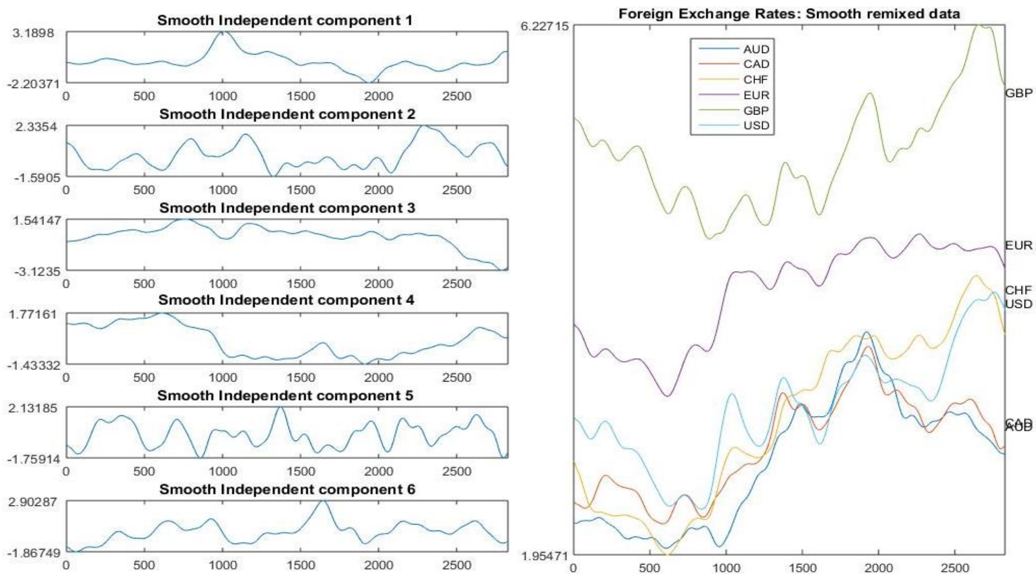

Figure no. 5. Smooth independent components and foreign exchange series

There are many alternative techniques for denoising time series. Singular Spectrum Analysis (SSA) is one of them. SSA is a nonparametric method for time series structure recognition and identification. The SSA algorithm has two basic stages: decomposition and reconstruction. Basically, it first builds the trajectory matrix associated to a time series, whose columns are formed by a sequence o lagged vectors extracted from the time series with a sliding window. Afterward, the singular value decomposition is applied to the trajectory matrix. For the reconstruction of the denoised part of the time series the most dominant eigenvalues are considered, whereas the rest of eigenvalues are used to compute the residuals.

5. FUTURE WORK

The final goal in using ICA is to predict financial time series observed as some mixture signals by first going to the ICA space (spanned by the independent components), doing the prediction there, and then transforming back to the original time series. The prediction can be done separately and with a different method for each independent component, depending on its time structure.

There are several possible ways to forecast with ICA.

One approach is to compute the h-step ahead forecast of x_t regression on s_t and some lagged factors.

$$x_{t+h} = \alpha_h + \sum_{j=1}^{N} \beta_{j,h} \hat{s}_{t+h-j} + e_{t+h}, \quad t = 1, \ldots, T-h$$

The factors are estimated by ICA, using the whole sample. Then, the factors and their lags are used to forecast x_t by the following linear regression model:

$$\hat{x}_{t+h|T} = \hat{\alpha}_h + \sum_{j=1}^{N} \hat{\beta}_{j,h} \hat{s}_{t+h-j|T} + e_{t+h}$$

where h is the forecast time horizon, and the regression coefficients, $\{\alpha_h, \beta_{j,h}\}_{j,h}$, are estimated by OLS.

Another approach is to forecast in the space of the latent factors (using an auto-projective technique) and then transform back to the original data. The three steps of the procedure are the following:

The common factors of x_t are estimated by ICA and the elements of the mixing matrix, A, are simultaneously found.Then, an ARIMA model is fit to each component and the h-step-ahead forecast of each factor is separately computed:

$$\{\hat{s}_{j,T+h|T}\}_{j=1}^{T}$$

The forecasts for the observed data are obtained by multiplying the forecasts for each independent component by the corresponding weights:

$$\hat{x}_{T+h|T} = \hat{A}\hat{s}_{T+h|T}$$

The problem of forecasting financial time series via ICA will be addressed in a future paper.

6. CONCLUSIONS

The approach in this paper is based on interpreting financial time series as mixtures of several distinct underlying factors. In an attempt to produce more structured and regular representations and more accurate predictions, a spectral decomposition approach to separating, smoothing and remixing fundamental factors is used. It consists in exploiting the ICA's capability of revealing independent components behind several parallel series of foreign exchange rates. These latent independent factors can be further used for producing forecasts and the later can be remixed in order to predict the observable time series. More accurate predictions can be performed via these independent components, after their separation, due to their higher regularity, using different forecasting strategies.

REFERENCES

1.	Bell, A.J., Sejnowski, T.J.	An information-maximization approach to blind separation and blind deconvolution. Neural computation. 7, 1129–1159, 1995.
2.	Belouchrani A. et all.	A blind source separation technique based on second order statistics. IEEE Trans. on Singal Processing, 45(2):434-44, 1997.
3.	Cardoso, J.F., Souloumiac, A.	Independent component analysis. John Wiley and Sons, New York, 2001
4.	Cardoso, J.F.	Source separation using higher order moments. Proc. ICASSP'89, pages 2109-2112, 1989.
5.	Comon, P.	Independent component analysis: a new concept? Signal Processing 36, 287-314, 1994.
6.	Georgescu,V.	Robustly Forecasting the Bucharest Stock Exchange BET Index through a Novel Computational Intelligence Approach. Economic Computation and Economic Cybernetics Studies and Research, 3, 23-42, 2010.
7.	Georgescu,V.	An Econometric Insight into Predicting Bucharest Stock Exchange Mean-Return and Volatility–Return Processes. Economic Computation and Economic Cybernetics Studies and Research, 3, 25-42, 2011.
8.	Georgescu,V.	Online Change-Point Detection in Financial Time Series: Challenges and Experimental Evidence with Frequentist and Bayesian Setups. Proceedings of the XVII SIGEF Congress, Word Scientific, 131-145, 2012.
9.	Georgescu,V.	Indications of Chaotic Behavior in Bucharest Stock Exchenge BET Index. Proceedings of the MS'12 International Conference, Word Scientific, 343-352, 2012.
10.	Georgescu,V.	Joint Propagation of Ontological and Epistemic Uncertainty across Risk Assessment and Fuzzy Time Series Models, Computer Science and Information Systems, 2, 11, 881–904, 2014.
11.	Georgescu,V., Delureanu, S.M.	Fuzzy-Valued and Complex-Valued Time Series Analysis using Multivariate and Complex Extensions to Singular Spectrum Analysis. Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015), 1-8, 2015.
12.	Georgescu,V., Delureanu, S.M.	Advanced Spectral Methods and Their Potential in Forecasting Fuzzy-Valued and Multivariate Financial Time Series. Advances in Intelligent Systems and Computing, Springer-Verlag, 129-140, 2015.
13.	Hyvärinen, A.	Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 10(3), 626-634, 1999.
14.	Hyvärinen, A., Karhunen, J., Oja, E	Independent component analysis. John Wiley and Sons, New York, 2001
15.	Jutten, C., Herault, J.	Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Processing 24, 1-10, 1991.
16.	Molgedey, L. Schuster, H.G.	Separation of a mixture of independent signals using time delayed correlations. Phys. Rev. Lett., 72:3634-3636, 1994.
17.	Stone, J.V.	Independent Component Analysis: A Tutorial Introduction. Massachusetts Institute of Technology, Massachusetts, 2004.
18.	Tong, L. et all	Indeterminacy and identifiability of blind identification. IEEE Trans. on Circuits and Systems, 38, 1991.