Analysis of Financial Time Series Morphology with AMUSE Algorithm and Its Extensions

R. Szupiluk[a,*], T. Ząbkowski[b] and T. Soboń[a]

[a]Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland
[b]Warsaw University of Life Sciences, Nowoursynowska 166, 02-787 Warsaw, Poland

The proposed article presents a new approach to analyze the relationships between financial instruments. We use blind signal separation methods to decompose time series into the core components. The components common to the various instruments provide broad set of characteristics to describe the internal morphology of the time series. In this research a modified and extended version of AMUSE algorithm is used. The concept is presented based on real financial instruments.

1. Introduction

The statistical assessment of financial instruments and the relationship between them is one of the fundamental problems in the theory and practice of financial markets and arise in such areas like the Markowitz portfolio theory, or CAPM and ATP models [1, 2]. It is also a key component of investment systems, especially for those based on automatic algorithmic strategies. However, in the multidimensional case, the assessment of interdependences between the different signals based on standard correlation analysis is relatively ambiguous and often difficult due to such factors like the outliers, the noises, the delays or miscellaneous random fluctuations.

We propose a new approach in which comparison of mutual dependence is based on the hidden components. They can be determined as a result of times series morphological analysis which is crucial to reveal the internal structure of the time series. The hidden components are obtained from blind signal separation methods [3–5] where we use a novel version of the AMUSE algorithm [5, 6]. Since the standard version of this algorithm is based on time delay second-order statistics and it is very sensitive to additive noise, we propose to use filtered nonlinear covariance matrices in order to increase its effectiveness. Additionally, we propose to explore flexible Perks model distributions [7, 8] as basis for nonlinear function used in our separation algorithm, which enables to explore higher-order statistical dependences.

2. Blind signal separation and AMUSE algorithm

The idea of blind signal separation (BSS) is to reconstruct source signals mixed in a some system. Both, the

system and the source signals are unknown and the identification is only based on mixed data [3, 4]. The solution to this particular problem requires a number of assumptions, including the generating model. The most common one is a linear model

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) \qquad (1)$$

as well as the model with additive noise

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t), \qquad (2)$$

where $\boldsymbol{A} \in R^{m \times n}$ is full column rank mixing matrix, $\boldsymbol{x} = [x_1, x_2, \ldots, x_m]^T$ is the vector of observed mixed signals, and $\boldsymbol{s} = [s_1, s_2, \ldots, s_m]^T$ is the vector of source signals. The difference between the model (1) and (2) is crucial because additive noises are still a challenging task for BSS methods. Beyond the models (1)–(2) there are other families of models including non-linear mixing, dynamic mixing, and non-linear dynamic mixing. However, in most cases, the starting point is the model of the first form (1) where the solution is determined by a separating matrix $\boldsymbol{W}$ which can be fitted into the following formula:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} = \boldsymbol{W}\boldsymbol{A}\boldsymbol{s} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{s}, \qquad (3)$$

where $\boldsymbol{P}$ is permutation matrix which determines the order of the estimated signals, $\boldsymbol{D}$ is diagonal scaling matrix, $\boldsymbol{W} = \boldsymbol{A}^+$ is pseudo-inverse matrix to $\boldsymbol{A}$ [3, 4]. From (3) we see that the estimated source signals y can be rescaled and reordered, as versions of signals $\boldsymbol{s}$, but in our case it is not significant, so we can assume $\boldsymbol{y} \approx \boldsymbol{s}$. The particular solution of the BSS problem is related to signals and mixing system characteristics. The most popular approaches are: independent component analysis (ICA), sparse component analysis (SCA), smooth component analysis (SMCA) or separation methods based on second-order statistics, like SOBI and AMUSE [3, 4, 9].

In further considerations we focus on AMUSE algorithm, which is addressed for signals with time structure like financial time series. For model (1) this algorithm is simple and there is no need to set arbitrary parameters. Unfortunately, in case of even small level of additive noise the quality of separation declines drastically.

———————

*corresponding author; e-mail: rszupi@sgh.waw.pl

To overcome the above problems we propose include to AMUSE ideas underlying to the ICA algorithms, where exploration of the higher order statistics gives additional information for separation process. It means the usage of non-linearities functions with optimal form as (1):

$$f(y) = -\frac{\partial \log(p(y))}{\partial y}, \qquad (4)$$

where $p(y)$ is probability density function of y, which is in general not known a priori and we need to adopt certain assumed probability model.

To combine time structure exploration which is underlying idea of the AMUSE algorithm with higher order statistics exploration via ICA algorithms we define the non-linear filtered correlation matrix in the following form:

$$\boldsymbol{R}_F = \lambda E(\boldsymbol{y}(t)(\sum_{k \in K} a_k \boldsymbol{y}(t-k))^T$$
$$+ \beta E(\boldsymbol{f}(\boldsymbol{y}(t))(\sum_{j \in J} a_j \boldsymbol{y}(t-j))^T), \qquad (5)$$

where λ, β are parameters defining the ratio between the correlation matrix and non-linear correlation matrix, and K and J are sets of delay values for the ordinary correlation matrix and non-linear correlation matrix. In order to choose the nonlinearity we propose a new approach based on Perks family distributions [7, 8]:

$$p(y) = \frac{a_0 + a_1 e^{-y} + a_2 e^{-2y} + \ldots + a_m e^{-my}}{b_0 + b_1 e^{-y} + b_2 e^{-2y} + \ldots + b_n e^{-ny}}, \qquad (6)$$

where the parameters are selected to meet the conditions of the probability density function. For example setting $m = 1, a = 0, a_1 = 1$ and $n = 2, b_0 = 1, b_1 = 0, b_2 = 1$, reduces $p(y)$ to hyperbolic secant distribution. The nonlinearity (4) for distribution (6) in the general case has the form

$$f(y) = \frac{b_1 e^{-y} + \ldots + b_n e^{-ny}}{b_0 + b_1 e^{-y} + \ldots + b_n e^{-ny}} - \frac{a_1 e^{-y} + \ldots + a_m e^{-my}}{a_0 + a_1 e^{-y} + \ldots + a_m e^{-my}}. \quad (7)$$

As a result, our version of separation algorithm called extended AMUSE (EAMUSE) has the following form:
(1) Let $\boldsymbol{z}(t) = \boldsymbol{x}(t), p = 0, \boldsymbol{W} = \boldsymbol{I}$.
(2) Estimate the generalized time delay correlation matrix R_F defined by (5).
(3) Perform the symmetrization process

$$R_{zz}(p) = \frac{1}{2}[R_F + R_F^T]. \qquad (8)$$

(4) Find matrix orthogonal to matrix $\boldsymbol{Q}_p$ (eg. from SVD) which diagonalizes the $\boldsymbol{R}_{zz}$.
(5) Perform decorrelation for given delay

$$y(t) = Q_p z(t). \qquad (9)$$

(6) Let $\boldsymbol{W} \leftarrow \boldsymbol{Q}_p\boldsymbol{W}$, $\boldsymbol{z}(t) \leftarrow \boldsymbol{y}(t)$, $p \leftarrow p + 1$ and go to step 2 until chosen stop criterion is fulfilled.
(7) The separation matrix is $\boldsymbol{W}$.

3. Financial data separation — time series morphology

Due to the multidimensional nature of BSS we gain direct insight into the structure of dependences existing in financial data. The relationship analysis performed on certain set of internal components will be referred as morphological analysis. To simplify the assessment of the extracted components, they are standardized into a unitary variance and the corresponding standardization coefficients will be included in the modified mixing matrix

$$\boldsymbol{A} = \boldsymbol{W}^{-1}\boldsymbol{V}, \qquad (10)$$

where matrix $\boldsymbol{V}$ is diagonal matrix with elements $v_{ii} = \frac{1}{\text{var}(y_i)}$. For the extracted source signals and for the mixing matrix we define several morphological characteristics which enable the synthetic analysis of the internal structure of financial instruments and the relationships existing between them. First, we define the concept of BSS correlation defined as:

$$G\text{cov}(x_1, x_2) =$$
$$\frac{\min(|a_{11}|, |a_{12}|) + \min(|a_{21}|, |a_{22}|)}{\max(|a_{11}|, |a_{12}|) + \max(|a_{21}|, |a_{22}|)}, \qquad (11)$$

where a_{ij} are the elements of the separating matrix $\boldsymbol{A}$, which are the results of separating the signals x_1 , x_2. The interpretation of this relationship is as follows. With two observed signals and having, as result of the decomposition (separation), two source signals we can analyze the direct impact of individual components on the observed signals. The dominant component in the observed signal will be called characteristic or dominant one. Non-dominant component will be called external one. As the correlation strength we take the ratio of the dominant components to the external components. A value of 1 is achieved if the proportion of the source signal in the observed signals is equal. Signals are fully independent if each of them is associated with only one hidden component. In the case of ICA method and its properties, it is in fact the actual independence (not only uncorrelated).

In general, for $x_1, x_2, \ldots, x_n$ signals we can define correlation matrix in the form $\boldsymbol{GCOV} = G\text{cov}(x_k, x_l)_{kl}$, in which the particular elements are calculated in the following way:

$$G\text{cov}(x_k, x_l) = \frac{\sum_{i=1}^{2} \min\left[|a_{i1}^{(k,l)}|, |a_{i2}^{(k,l)}|\right]}{\sum_{i=1}^{2} \max\left[|a_{i1}^{(k,l)}|, |a_{i2}^{(k,l)}|\right]}, \qquad (12)$$

where $a_{i1}^{(k,l)}$ are the elements of separating matrix $\boldsymbol{A}$ which are the results of separating the pair of the signals x_1, x_2. This means that each pair of signals is separated individually with the same algorithm.

Another characteristic which needs to be defined is the morphological similarity determined in the system that consists of n selected signals. The similarity is also defined for the pairs of signals, but it assumes a uniform and simultaneous separation of all signals where common separation matrix is analyzed. Therefore, the common morphological similarity (morphological variance/covariance — mc) is defined as:

$$M_{ij} = \sum_{k=1}^{n} |a_{ik}||a_{jk}|. \tag{13}$$

The last characteristic will refer to the individual financial instrument and will be expressed in terms of separated signals. For this reason we will use the measure of morphological diversification (morphological entropy), expressed as

$$Hm = - \sum_{i=1}^{n} \left(\frac{a_i}{\sum_1^n a_i} \log \frac{a_i}{\sum_1^n a_i} \right). \tag{14}$$

4. Practical applications

In this section we present an example of time series morphology analysis based on the above concept. The following stocks and indices from Warsaw Stock Exchange are selected for the analysis: Mostostal Zabrze (MSZ), Mostostal Warszawa (MSW), BZWBK (BZW), Kety (KTY), Budimex (BDX), WIG20, mWIG40, sWIG80. Firstly, we are interested if the morphological characteristics are stable over the time and whether they have an interesting economic interpretation, which is presented for the two companies. In case of BDX and BZW we observed quite stable behavior: the GC coefficient for 2002.06.26–2007.03.30 period was 0.093, while for 2007.03.30–2008.06.17 it was 0.106. The separating matrix for the first period (2002.06.26–2007.03.30) was

$$A = \begin{bmatrix} 0.84 & 0.16 \\ 0.01 & 0.99 \end{bmatrix}$$

and for the second period (2007.03.30–2008.06.17) it was

$$A = \begin{bmatrix} 3.19 & -0.25 \\ 0.34 & -2.50 \end{bmatrix}.$$

Figure 1 shows the percentage of absolute values of the individual components. The dominant component in BDX is y_1 and for BZW it is y_2. We can observe that for the long-term bull market (the first period) and the declines (the second period) these relationships are relatively stable. It can be also noted that in the first period the influence of the dominant BZW component on BDX is greater than the dominant BDX component on BZW. This seems consistent with the market experience and the theory because BZW is included in the leading WIG20 index and its share is approximately 5 percent. The similar impact can be observed on the morphological analysis. At the same time we can observe that the impact of BDX (the dominant component) on BZWBK is not important, which can be consistent with intuition and market practice.

The second study was aimed to compare the signals generated by the GC coefficient with the signals generated by a standard correlation coefficient (corrcoef). This was done for two indices: WIG20 and mWIG40, taking into account the closings of 3500 sessions. The values of GC coefficient were compared with the standard corre-
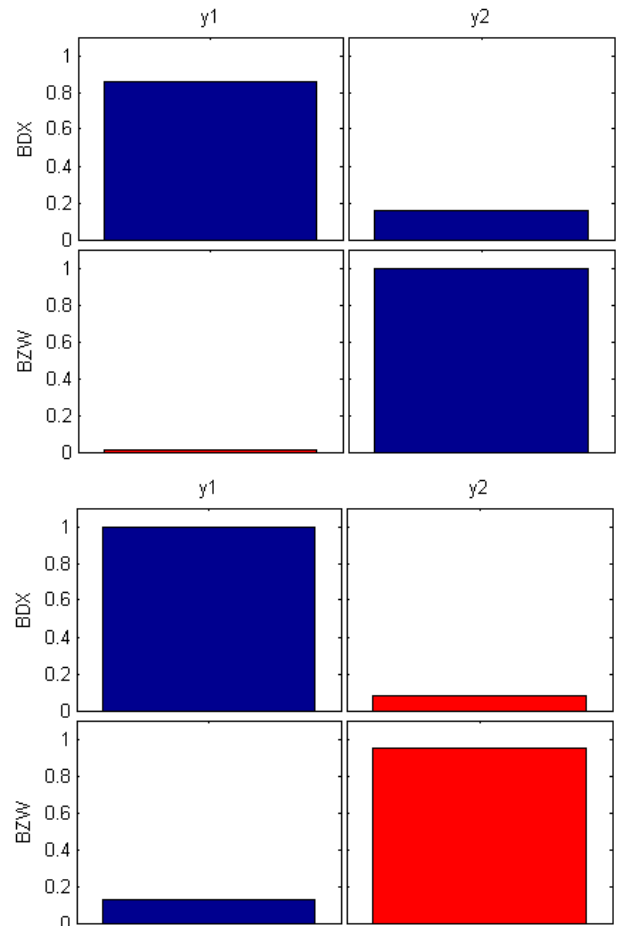

Fig. 1. The visualization of absolute values in A matrices for Budimex (BDX) and BZWBK (BZW): (top) in 2002.06.26–2007.03.30 period; (bottom) in 2007.03.30–2008.06.17 period.

lation coefficient. The calculations were prepared on the rolling windows with 400 and 600 observations each.

Based on Fig. 2, in both cases, with increasing number of observations the smoothing of the characteristics is observed. However, in case of Gcov coefficient a significant increase in the volatility is clearly visible for market disaster in 2007 and 2008 (the observations between 1500 and 2000). Similarly, in the 2011 with the markets collapsed due to the crisis in Greece, the characteristics of the Gcov shows the fluctuations (around the observation no. 2500). For the correlation coefficient we do not see those market events so clearly. Therefore, the Gcov characteristics can act as an indicator of significant and/or long term changes in the market.

The third study was performed for the common morphological similarity assessment for all chosen instruments. Figure 3 shows a visualization of the absolute values of the A common matrices normalized to the range $(0,1)$ in the following periods: (a) 2002.06.26–2007.03.30, (b) 2007.03.30–2008.06.17. We can see that each instrument can be associated with the dominant components which can be associated with main trend of each instru-
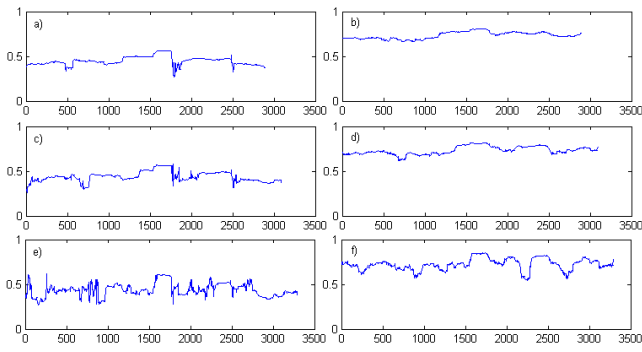
Fig. 2. *G*cov and corrcoef over the time, where: a) GC with 600 observations window, b) corrcoef with 600 observations window, c) GC with 600 observations window, d) corrcoef with 400 observations window.

ment. For instance, for WIG20 it is $y7$ component for the first period and $y4$ for the second period. The components other than the dominant can be interpreted as the impact of the other instruments on the particular stock. For example, $y3$ component in the first period significantly influences several building and construction companies (MSZ, MSW, BDX). Due to that, it can be treated as a component of broad market trends related to the boom in the construction and property development market which was an engine for the growth in the overall market in 2003–2007.

The dependence structure shown in the diagrams in Fig. 1 may be used for the noise elimination and trends estimation. The elimination of the component interpretable as interfering or external is possible by e.g. assuming $y1 = 0$ and applying inverse transformation to (3). As a result the filtered values of each individual instruments can be obtained. The selection of the noise or random component may be based on morphological entropy defined by (14). Such morphological noise detection can be either applied for prediction improvement [10].

TABLE I

The values of variance and morphological covariance mc.

	BDX	BZW	MSW	MSZ	KTY	sWIG80	mWIG40	WIG20
BDX	0.30							
BZW	0.08	0.35						
MSW	0.11	0.14	0.44					
MSZ	0.10	0.18	0.29	0.26				
KTY	0.16	0.20	0.18	0.19	0.26			
sWIG80	0.05	0.05	0.13	0.07	0.05	0.54		
mWIG40	0.09	0.07	0.15	0.10	0.09	0.07	0.42	
WIG20	0.05	0.09	0.15	0.10	0.08	0.03	0.04	0.51

Based on the value of the common separation matrix it is possible to determine the values of the morpholog-
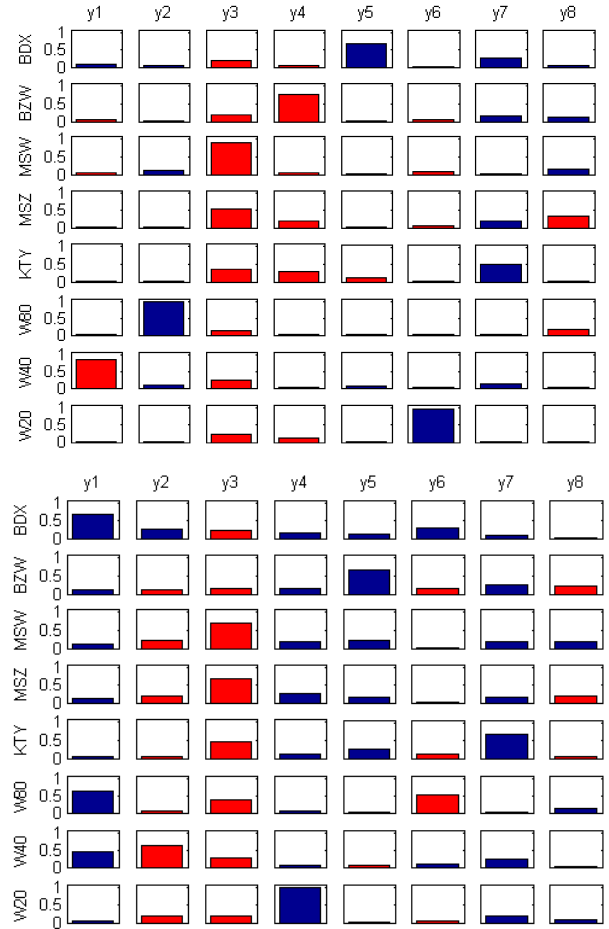

Fig. 3. The visualization of absolute values in A matrices for the analyzed set of instruments in: (top) 2002.06.26–2007.03.30 period, (bottom) 2007.03.30–2008.06.17 period.

ical variance/covariance (13), as shown in Table I. The obtained values are directly related to the size/strength of the components explanatory power. From a practical point of view, the value of the morphological correlation, combined with the knowledge on the structure of the base components can contribute to variables selection, e.g. in arbitrage pricing theory (APT) model which can be an alternative to the models built based on principal component analysis.

5. Conclusions

In this paper, we proposed the use of blind source separation methods to gain an insight into the inner structure of financial instruments, together with the assessment of existing relationships in data which we refer to as morphological analysis. For the separation stage we have extended and modified the AMUSE algorithm to combine the properties of the SOS BSS algorithms with the ICA algorithms. The family of Perks distributions is proposed to choose the non-linear function. Due to the methodological character of our studies, the under-

taken experiments had, to some extent, only exemplary and introductory contribution. Nevertheless, the proposed experiments confirmed the validity of theoretical concepts discussed, giving interesting novel financial data characteristics. Of course, there are many aspects open to further research like: the analysis of the components depending on the separation choice, development of a consistent interpretation of the morphological information derived from different sets of analyzed instruments or/and linking the conclusions of the morphological study with the existing financial and economic theories. We expect that the future research will address these issues and contribute to the methodology on financial markets.

References

[1] J.C. Hull, *Options, Futures and Other Derivatives*, Prentice Hall, Upper Saddle River 2014.

[2] A.N. Shiryaev, *Essentials of Stochastic Finance: Facts, Models, Theory*, World Sci., Singapore 1999.

[3] A. Cichocki, S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, Wiley, Chichester 2003.

[4] P. Comon, Ch. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, Boston 2010.

[5] R. Szupiluk, A. Cichocki, in: *Proc. SPETO*, Wydawnictwo Politechniki Śląskiej, Gliwice/Ustroń 2001, p. 485.

[6] L. Tong, V. Soon, Y.F. Huang, R. Liu, *IEEE Trans. Circ. Syst.* **38**, 499 (1991).

[7] M.J. Fischer, *Generalized Hyperbolic Secant Distributions with Applications to Finance*, Springer, Heidelberg 2014.

[8] W. Perks, *J. Inst. Actuar.* **63**, 1257 (1932).

[9] R. Szupiluk, *Multivariate decompositions for predictive Data Mining models aggregation*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa 2013 (in Polish).

[10] R. Szupiluk, P. Wojewnik, T. Zabkowski, *Lect. Notes Comp. Sci.* **6594**, 133 (2006).