QUANTITATIVE METHODS IN ECONOMICS
Vol. XIV, No. 2, 2013, pp. 240 – 250

MULTIVARIATE DECOMPOSITIONS FOR VALUE AT RISK MODELLING

Ryszard Szupiluk, Piotr Wojewnik
Department of Business Informatics, Warsaw School of Economics
e-mail: rszupi@sgh.waw.pl, piotr.wojewnik@gmail.com
Tomasz Ząbkowski
Department of Informatics
Warsaw University of Life Sciences – SGGW
e-mail: tomasz_zabkowski@sggw.pl

Abstract: This paper presents the application of independent component analysis (ICA) for value at risk modelling (VaR). The probabilistic models fitted to hidden components from the time series help to identify the independent factors influencing the portfolio value. An important issue here is the choice of the ICA algorithm, especially taking into account the characteristics of the instruments with respect to higher-order statistics. The proposed ICA-VaR concept has been tested on transactional data of selected stocks listed on Warsaw Stock Exchange.

Keywords: multivariate decompositions, value at risk modelling, independent components analysis

VALUE AT RISK MODELING

One of the most popular concept of investment risk modelling is the concept of Value at Risk, which consists of estimating the risk for a specified time horizon at a given probability [Jorion 2001, Jajuga 2001, JP Morgan, 1995]. Although the concept itself is simple and intuitive, it is associated with a fundamental problem of estimating the probability that a given financial instrument (or portfolio) will reach specific values in the future. In this area, one of the most commonly used approache~~s~~ are these based on simulations. They aim to find the possibly best mathematical model for the instrument based on historical data, and then performing predictive simulation of the model behaviour. This opens up a discussion whether the model is adequately fitted to empirical data. In practice, the model fit is a compromise between the characteristics of the empirical data, a priori assumptions about the nature of the original phenomenon and properties of the mathematical apparatus [Bollerslev et al. 1992, Shiryaev 1999].

Since the Markowitz publications [Markowitz 1952] the phenomenon of uncertainty and risk was mainly interpreted in terms of financial instruments volatility and correlation. Additionally, the rational expectations hypothesis justified the random nature of the changes in financial instruments, making popular the models which were based on Gaussian distribution. Bearing in mind that the variance/covariance fully identifies the Gaussian distribution, volatility expressed by the variance gives complete statistical information about the phenomenon of uncertainty and risk. Taking into account also the relationship between Gaussian distribution and Central Limit Theorem a comprehensive conceptual system was established and it prevails in the description of uncertainty and risk over the last decades.

The experience of recent years and the current situation on the financial markets, however, indicate some limitations of this conceptual apparatus. It turned out that models based on Gaussianity and correlations cannot recognize some critical features of the financial markets such as rare events. Historical and actual volatility described by variance is often not applicable for the future situation reasoning, which, in fact, can change dramatically due to rare events such as market breakdown. As a result, unexpected change on financial instruments is not so much associated with the current volatility, but it is rather due to the occurrence of rare events such as panic in the stock market or bankruptcy of a large financial institution.

That types of phenomena and behaviour can be observed in the short-term scales. As a result, we need to look for mathematical apparatus that can better deal with the rare phenomenon. Although the number of works devoted to rare events is substantial, the problem is still an open research issue [Embrechts et al. 1997, Harvey 2013].

One reason for the difficulty of modelling specific market behaviour may be the fact that they do not occur as isolated events. In other words, a crisis or crash is frequently preceded by a long process that develops on a seemingly normally functioning market. Similarly, after the collapse, its effect lasts for a long time. In addition, it should be noted that market instruments are generally interrelated, although the nature, timing and scope of this relationship can be difficult to determine and predict. As a result, we can assume that the morphology of such time series is so complex and the only one signal analysis can be very confusing. This is natural motivation for time series decomposition into components associated with its particular characteristics for their individual properties modelling.

There are two main approaches for decompositions: (a) one-dimensional that is based on a decomposition of the time series (e.g. trend, cycles, noise) or (b) multi-dimensional, taking into account and exploring the relationships between few different signals. In the following discussion we will focus on a multi-dimensional approach that recently led to a number of interesting decomposition methods with number of practical application. In particular, we consider the following method of VaR supported by ICA decomposition:

1. Collect the original time series into one multivariate variable;
2. Decompose the multivariate variable into hidden independent components (separation stage);
3. Choose components for further analysis (filtration stage);
4. Estimate the probability distributions for each component;
5. Perform simulation for each component;
6. Do re-mixing of the components using reverse system to decomposition (separation);
7. Calculate VaR for given portfolio.

This concept establishes the general research framework, in which components can be identified using different mathematical characteristics. In case of the components identified as noises the algorithm can also realize the filtration [Szupiluk 2004].

The following discussion will focus primarily on classic independent component analysis [Hyvärinen et al. 2001]. This method of decomposition has many practical applications including the blind signal separation problem [Cichocki and Amari, 2002]. However, in contrast to the relatively clear results obtained with a principal component analysis - PCA [Jolliffe 1986], ICA needs the deeper insight into the characteristics of the used algorithms. The motivation for ICA decomposition is due to the fact that blind separation is one of the most general methods of separation, exploring higher-order statistics. This is particularly important in case of rare events observed in financial time series, for which the kurtosis is one of the most important criterion of their assessment.

CHARACTERISTICS OF INDEPENDENT COMPONENT ANALYSIS

In the classical meaning independent component analysis is formulated as a method that allows separation of a multi-dimensional observation vector $\mathbf{x} = [x_1, x_2, ..., x_n]^{\mathrm{T}}$ into statistically independent components $\mathbf{y} = [y_1, y_2, ..., y_m]^{\mathrm{T}}$. It is assumed that the estimation of the independent components is performed using a linear transformation $\mathbf{y} = \mathbf{Wx}$, where $\mathbf{W}$ is separating matrix.

Independent Component Analysis has been widely used for modelling economic and financial phenomena [Back and Weigend 1997]. Also, there are number of publications indicating the effectiveness of ICA for risk analysis [Chen et al. 2007, Wu et al. 2006]. Independent component analysis can be considered twofold. In first case, it can be assumed as a strict statistical decomposition method, which allows to extract independent components from the multidimensional observations. In the second case, ICA can be treated as a method to solve blind separation problem.

ICA as a statistical method

The assessment of signals independence requires the knowledge of their probability distributions, which is a relatively complex task in case of financial time series. In addition, ICA models assume that independent components are mixed and hidden and their distributions are, by definition, unknown. As a result, the practical notion of statistical independence of components obtained in the ICA is not precise and the final effect is verifiable to a limited extent. Unlike the principal components analysis with linear algebra apparatus, separation of independent components requires adoption of certain criteria, concepts, principles and characteristics, which exploration (optimization) may result in certain numerical algorithms. The most popular approaches include the minimization of mutual information, entropy maximization, non-Gaussianity maximization (measured by negentropy or kurtosis) or non-linear decorrelation.

One of the standard algorithms for finding the matrix **W** are Natural Gradient [Amari et al. 1997]

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \mu(t)\left[\mathbf{I} - E\left\{\mathbf{f}(\mathbf{y})\mathbf{y}^{\mathrm{T}}\right\}\right]\mathbf{W}(t)\,. \tag{1}$$

and FASTICA [Hyvärinen et al. 2001]

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \mu(t)\mathbf{D}\left[E\left\{\mathbf{f}(\mathbf{y}(t))\mathbf{y}^{T}(t)\right\} - \mathrm{diag}\left(E\left\{f(y_i)y_i\right\}\right)\right]\mathbf{W}(t) \tag{2}$$

where E is expectation operator, $\mu(t)$ is a learning rate, $\mathbf{f}(\mathbf{y}) = \left[f_1(y_1),..., f_n(y_n)\right]^{\mathrm{T}}$ and $\mathbf{D} = \mathrm{diag}\left(1/E\left\{f(y_i)y_i\right\} - E\left\{f'(y_i)\right\}\right)$ are the vector and the matrix with nonlinearities of the:

$$f_i(y_i) = -\frac{\partial \log(p_i(y_i))}{\partial y_i} \tag{3}$$

where $p_i(y_i)$ is *pdf* of signal y_i .

The key issue in these algorithms is the selection of non-linearity (3). One of the simplest but well working rules are these using higher-order statistics and based on the observation that the non-linearity (3): (i) takes the linear form for the Gaussian distribution; (ii) for the distributions with the slope higher than Gaussian it is growing faster than linear; (iii) for the distributions with the slope lower than Gaussian it is growing slower than linear. Taking the kurtosis $\kappa_4(y_i)$ as distribution slope parameter we obtain a rule of non-linearity selection [Cichocki et al. 2007]:

$$f_i(y_i) = \begin{cases} y_i^3 & \text{for} \quad \kappa_4(y_i) > 0, \\ \tanh(y_i) & \text{for} \quad \kappa_4(y_i) < 0, \\ y_i & \text{for} \quad \kappa_4(y_i) = 0. \end{cases} \tag{4}$$

The above rule can be effective for simple, unimodal and symmetric cases, taking into account a small number of signals with relatively different distributions. In more complex cases it requires more accurate non-linearity determination what makes it difficult since the distribution of independent component is a priori unknown. As a possible solution we can propose either, heuristic methods, parametric models or adaptive techniques. Nevertheless, one of the most versatile approaches is a method based on the Extendend Generalized Lambda Distribution (EGLD) in which it is possible to model a wide range of distributions with different kurtosis and skewness parameters [Karian et al. 1996, Karvanen et al. 2002].

It should be noted that at the theoretical level, with certain assumptions and simplifications, the criteria that establish the ICA algorithms lead to a situation in which the mutual independence is approximated or reduced to a fourth-order statistical relationships. In this case, for decorrelated and symmetric data ICA can be expressed as [Comon 1994]:

$$\max_{\mathbf{W}} \sum_{i=1}^{n} J(y_i) \approx \min_{\mathbf{W}} I(\mathbf{y}) \approx \max_{\mathbf{W}} \left(\sum_{i=1}^{n} \kappa_4^2(y_i) \right) \tag{5}$$

where $J(y_i)$ is negentropy of the signals y_i; $I(\mathbf{y})$ is join mutual entropy.

Therefore, it can be assumed that, although the ICA algorithms derived from different criteria have their particular numerical specificity, in practice the effect is linear and non-linear decorrelation, resulting in removal of the second and fourth order statistics. These observations led to the development of ICA methods based on fourth-order statistics, including also tensor approach, e.g. JADE algorithm [Cardoso 1999].

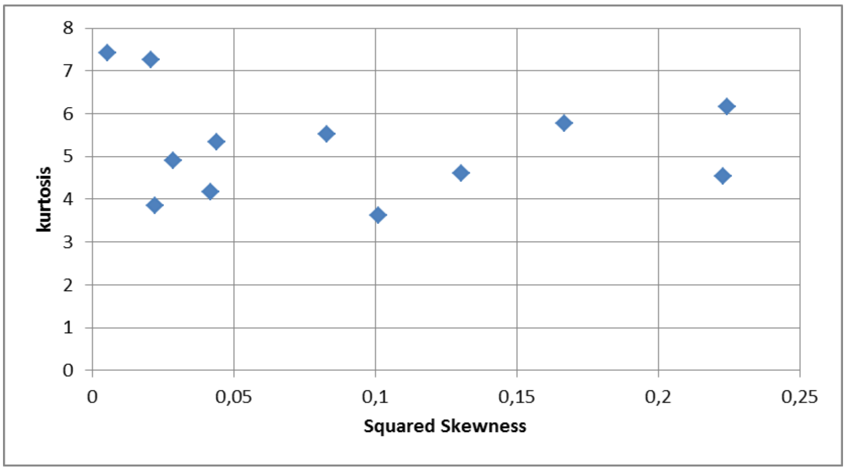
ICA as a blind signal separation method

The problem of blind signal separation is defined as the extraction of unknown signals mixed in the unknown system, with limited a priori knowledge. If we additionally assume that the sought source signals are statistically independent and the mixing is a linear combination of the signals, then the problem of blind signal separation and independent component analysis are identical. However, to separate the real signal, then we need a criterion to measure the independence. This is conditioned by the approximation accuracy using fourth order statistics of the estimated signal probability distribution function. Unfortunately, the form of this distribution is not known a priori in separation problem. Therefore, there is also a question of accuracy assessment in the separation process. While in some practical applications e.g. physical separation of speech signals, the evaluation can be quite simple; then in case of financial time series (signals), the situation is quite complex. First of all, a separation using different algorithms within ICA is possible. If their results comply then we can state that satisfactory solution was achieved, otherwise it will be difficult to assess which solution is better.

Consequently, an essential issue of ICA algorithm choosing is to fit properties and characteristics of the algorithm to a given problem. In case of financial time series it is a non-stationarity of time series with respect to the higher order statistics.

Figure 1 shows an example of kurtosis and squared skewness distribution calculated on logarithmic returns of the WIG20 index (for different periods, covering time span from 1994 till 2012). Kurtosis is calculated in a standardized form and for a Gaussian distribution it has zero value.

Figure 1. The points represent WIG20 index characteristics (kurtosis and squared skewness for different length of time windows (minimum 2 years) and covering time span from 1994 till 2012



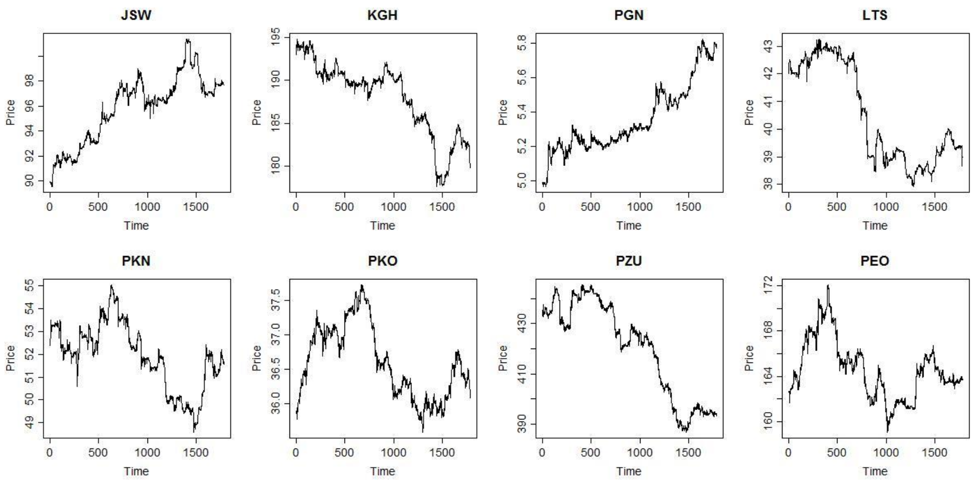
Source: own calculation

This example from Warsaw Stock Exchange shows that even for single time series the statistical characteristic is volatile and highly influenced by chosen time-window.

EXPERIMENT ON EMPIRICAL DATA

In this section an experiment on financial data was conducted. We considered eight stocks from Warsaw Stock Exchange covering the time span 19/12/2012 – 18/01/2013. Data consisted of 1783 observations of 5 minute data. Half of them (19/12/2012 – 07/01/2013) was used for decomposition matrix and probability density functions (PDF's) estimation and the second half (08/01/2013–18/01/2013) was used for testing. A portfolio of stocks consisted of JSW, KGH, PGN, LTS, PKN, PKO, PZU, PEO. For their characteristics, please see Fig. 2.

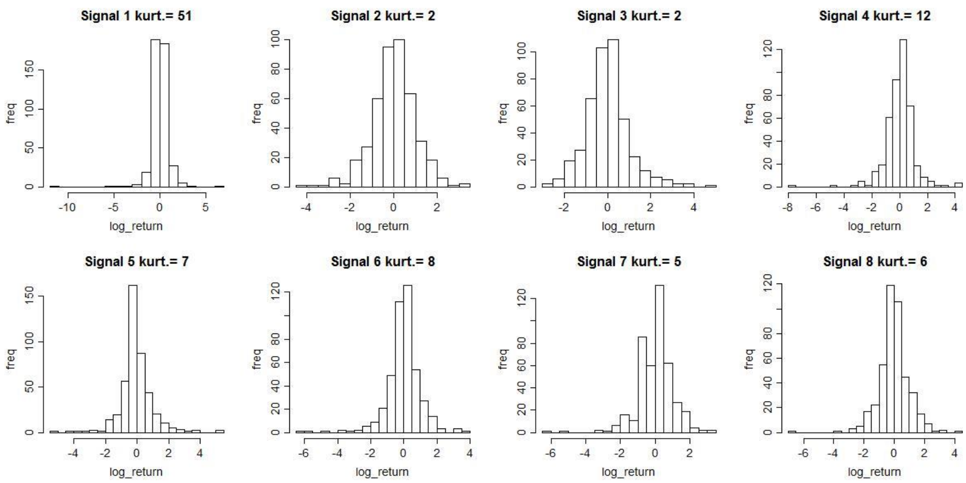
Figure 2. Stocks used in the experiment



Source: own preparation

Histograms of latent signals obtained as a result of JADE decomposition calculated on log returns for all stocks (time span for analysis 19/12/2012–07/01/2013) are presented on Fig. 3. We can observe that kurtosis of the signals ranges from 2 to 51 in this sample.

Figure 3. Histograms of latent components obtained after decomposition



Source: own preparation

The PDFs of latent signals were estimated using t-scale location distribution with shape parameter *df*, location parameter *m* and scale parameter *s*. Degrees of freedom ranges from 1 to 4. Please see Table 1 for details.

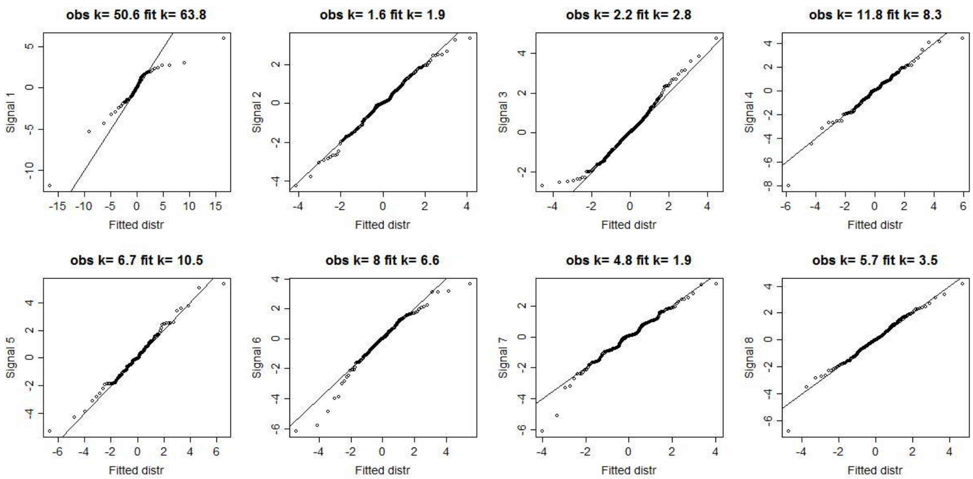
Table 1. PDF characteristics of latent signals

m	s	df
-0.0031	0.2110	1.1382
0.0171	0.7546	4.2239
-0.0445	0.7035	3.5227
0.0360	0.5241	2.3067
-0.0352	0.5056	2.1233
0.0429	0.5553	2.4985
0.0251	0.7277	4.1791
-0.0076	0.6633	3.2060

Source: own calculation

QQ plots of observed PDF's vs. fitted distribution are presented in Figure 4. Fitted kurtosis ranges from 2 to 60 and it is close to expectations. Therefore, we can conclude that used distributions seem to be an adequate choice.

Figure 4. QQ plots observed PDF's vs. fitted distribution



Source: own preparation

Finally, the VaR calculation results on test data (08/01/2013–18/01/2013) are presented in Table 2. It presents the percentage of cases exceeding specified VaR level using 100 000 simulations based on given decompositions and also in comparison to VaR level estimated on historical data (19/12/2012–07/01/2013).

In other words, we show VaR level not taking into account the value of a specific stock in this portfolio but total portfolio value.

Table 2. VaR results of portfolio consisted of eight stocks

VaR level	EGLD ICA VaR	Jade VaR	Historical VaR
0,01	0 %	1,2%	2,1%
0,025	1,4%	2,1%	4,2%
0,05	4,6%	6,0%	7,9%
0,1	9,7%	10,9%	12,7%

Source: own calculation

We showed that in an environment characterized by non-linearities and the occurrence of interactions between the stocks, the ICA methodology can advantageously reveal an underlying structure in financial time series for the purpose of risk measurement. The results obtained on the test dataset indicate the advantage of ICA approach over VaR calculation based on historical data. The best results were achieved for EGLD ICA method, which allows to simulate distributions in a wide range of kurtosis and skewness.

CONCLUSIONS

In this paper the application of independent component analysis in the multidimensional decomposition framework for VaR was proposed. In contrast to the relatively explicit algebraic decomposition methods, different ICA algorithms have their own characteristics, what can significantly influence the results. In particular, it refers to the case of ICA as a blind source separation method. Therefore, the choice of appropriate algorithm with respect to the third and fourth order statistics is the key issue here. These higher order statistics play an important role in ICA algorithms both, for the optimization task and for the numerical implementation of the algorithm. In the case of financial instruments we can observe both the instability of these statistics and the volumes which are significantly different from the Gaussian distribution. In such circumstances, it seems appropriate to select an adaptive algorithm for non-linearity selection taking into account a wide range of kurtosis and skewness. One of them is a system based on EGLD distribution. For the VaR modelling EGLD adoption led to better results than JADE, although the last one is recognized as one of the most popular and effective algorithms.

In this study, we focused mainly on distribution fitting and forecasting. Nevertheless, it should be noted that under the proposed approach a filtration and elimination of the specific components, can be also considered.

Acknowledgments: This work was funded by the National Science Center in Poland based on decision number DEC-2011/03/B/HS4/05092.

REFFERENCES

Amari S., Cichocki A., Yang H.H. (1996) A new learning algorithm for blind signal separation, Advances in Neural Information Processing Systems NIPS-1995, MIT Press, Cambridge MA, pp. 757–763.

Back A.D., Weigend A.S. (1997) A first application of independent component analysis to extracting structure from stock returns, Int. Journal of Neural Systems 8(4), pp. 473–484.

Bollerslev, T., Chou, R.Y., Kroner, K.F. (1992) ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence, Journal of Econometrics 52, pp. 5–59.

Cardoso J.F. (1999) High-order contrasts for independent component analysis, Neural Computation 11, pp. 157–192.

Chen Y., Hardle W., Spokoiny V. (2007) Portfolio value at risk based on independent component analysis Journal of Computational and Applied Mathematics 205(1), pp. 594–607.

Cichocki A., Amari S. (2002) Adaptive Blind Signal and Image Processing, John Wiley, Chichester.

Cichocki A., Sabala I., Choi S., Orsier B., Szupiluk R. (1997) Self adaptive independent component analysis for sub-Gaussian and super-Gaussian mixtures with unknown number of sources and additive noise, Proc. of NOLTA-97, vol. 2, Hawaii USA, pp. 731–734.

Comon P. (1994) Independent component analysis, a new concept?, Signal Processing 36, pp. 287-314

Embrechts P., Klüppelberg C., Mikosch T. (1997) Modelling Extremal Events for Insurance and Finance. Berlin, Springer.

Harvey A.C. (2013) Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, Cambridge University Press.

Hyvärinen A., Karhunen J., Oja E. (2001) Independent Component Analysis, John Wiley.

Jajuga K. (2001) Value at Risk, Rynek Terminowy13, pp. 18-20.

Jolliffe T. (1986) Principal Component Analysis, Springer-Verlag.

Jorion P. (2001) Value at Risk, McGraw-Hill.

JP Morgan (1995) Riskmetrics Technical Document, 3rd ed., New York

Karian Z.A., Dudewicz E.J., McDonald P. (1996) The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the Final Word on Moment Fits, Communications in Statistics - Simulation and Computation 25, pp. 611–642.

Karvanen J., Eriksson J., Koivunen V. (2002) Adaptive Score Functions for Maximum Likelihood ICA. VLSI Signal Processing 32, pp. 83–92.

Kouontchou P., Maillet B. (2007) ICA-based High Frequency VaR for Risk Management ESANN'2007 proceedings - European Symposium on Artificial Neural Networks, pp. 385-390.

Markowitz H.M. (1952) Portfolio Selection, The Journal of Finance 7, pp. 77–91.

Shiryaev A.N. (1999) Essentials of stochastic finance: facts, models, theory, World Scientific, Singapore.

Szupiluk R., Wojewnik P., Zabkowski T. (2004) Model Improvement by the Statistical Decomposition, Lecture Notes in Computer Science 3070, pp. 1199–1204.

Wu E.H., Yu P.L., Li W.K. (2006) Value at risk estimation using independent component analysis-generalized autoregressive conditional heteroscedasticity (ICA-GARCH) models, International Journal of Neural Systems 16(5), pp. 371–382.