“ICA-based High Frequency VaR for Risk Management

Patrick Kouontchou[1] and Bertrand Maillet[2] *

1- Variances and Paris 1 University - CES/CNRS
patrick.kouontchou@variances-fr.com - France

2- A.A.Advisors-QCG (ABN Amro Group), Variances,
Paris 1 University - CES/CNRS and EIF
bmaillet@univ-paris1.fr - France

Abstract. Independent Component Analysis (ICA, see Comon, 1994 and Hyvärinen *et al.*, 2001) is more appropriate when non-linearity and non-normality are at stake, as mentioned by Back and Weigend (1997) in a financial context. Using high-frequency data on the French Stock Market, we evaluate this technique when generating *scenarii* for accurate Value-at-Risk computations, reducing by this mean the effective dimensionality of the *scenario* specification problem in several cases, without leaving aside some main characteristics of the dataset. Various methods for specifying stress *scenarii* are discussed, compared to other published ones and classical tests of rejection are presented (Christoffersen and Pelletier, 2003).

1 Introduction

Value-at-Risk (VaR) shows the loss over a preset time horizon for a given probability, and has been popularly used to control and manage risks, including both credit risks and operational risks. Recent literature on risk has focused in the last years on the use of intra-day data for better measurement of financial risk (see Dacorogna *et al.*, 2001). There are many factors that drive the movements of asset returns. But is not unusual to assume that a set of different asset returns are driven by some common factors. By extracting the common factors from the high-frequency variance-covariance matrix, it is also possible to consider other applications. Specifically, some statistical issues that arise in the formulation of stress *scenarii* for market risk are studied in this text. The possibility of reducing the number of *scenarii* through the use of data-based statistical dimension reduction methods is proposed in Loretan (1997), when uses a Principal Component Analysis. Nevertheless, a loss in information may be entailed by the linear filter of the PCA when high-frequency data is considered. Indeed, non-linearities and non-gaussianities are specific features of this type of database. Independent Component Analysis (ICA, see Comon, 1994; Hyvärinen *et al.*, 2001), as mentioned by Back and Weigend (1997) in a financial context, is more appropriate when non-linearity and non-normality are at stake. Independent Component Analysis, a well-known method for finding latent structure

*We would like to thanks Professor Thierry Chauveau, and Michel Verleysen, Emmanuel Jurczenko, Thierry Michel for help and encouragement in preparing this work. Preliminary version; do not quote or diffuse without permission. The usual disclaimers apply.

in data, is a statistical method that expresses a set of multidimensional observations as a combination of unknown latent variables. These underlying latent variables are called sources or Independent Components which are assumed to be statistically independent one from another. Using high-frequency data on the French Stock Market, we evaluate this technique when generating *scenarii* for accurate Value-at-Risk (VaR) computations, reducing by this means the effective dimensionality of the scenario specification problem in several cases, without leaving aside some main characteristics of the dataset. Section 2 and 3 review the backgrounds of ICA model and VaR computation. Section 4 contains the experiment and results.

2 Independent Component Analysis

The method of Independent Component Analysis developed in this section is well-known in this Signal Processing field: biomedical, speech and telecommunications signals to mention a few. These signals can be seen as mixtures of different physical activities and external noise sources, and the task is to find this independent sources. We firstly present, the formal definition of the method and, we show after the reason for exploring this technique in a Times series context.

2.1 Definition

Independent Component Analysis (ICA, Comon, 1994; Hyvärinen *et al.*, 2001) is a well-known method of finding latent structure in data. ICA is a statistical method that expresses a set of multidimensional observations as a combination of unknown latent variables. The ICA model is:

$$X = f(\beta, S) \tag{1}$$

where $X(m \,\mathrm{x}\, 1)$ is an observed vector and $f(.)$ is a general unknown function with parameters β that operates on statistically independent latent variables listed in the vector $s(n \,\mathrm{x}\, 1)$. The task is to recover the original source signals from the observations through a demixing process. Various ICA algorithms have been proposed in the literature (see for example, Comon, 1994; Cardoso, 1999; Hyvärinen *et al.*, 2001). The approach is maximization of non-Gaussianity of the components; non-Gaussianity is often measured by higher order cumulants such as kurtosis or skewness, although they are not robust against outliers. Robust measures have been presented in Hyvärinen (1999) and Hyvärinen *et al.* (2001). They propose the FastICA algorithm which is an iterative fixed-point algorithm with the following update for unmixing matrix W:

$$W \leftarrow E[Xg(WX)] - E[g'(WX)]W \tag{2}$$

The nonlinear function g is the derivative of the non-quadratic contrast function G that measures negentropy or non-Gaussianity. A remarkable property of the FastICA algorithm is the high speed of convergence in the iterations.

2.2 Reasons for Exploring ICA in Financial Times-series

The ICA technique assumes that each particular signal of an ensemble of signals is a superposition of elementary components that occur independently of each other. The technique is different from the Principal Component Analysis (PCA), because it also imposes higher-order independence, and not just up to the second order (decorrelation) as in PCA (see for example Cardoso, 1989 and 1999; Comon, 1994; Hyvärinen, 1999; Hyvärinen *et al.*, 2001 and Hyvärinen and Oja, 1997). However, recently, it has been realized that Independent Component Analysis (ICA), rather than PCA, is more appropriate for time series analysis (Back and Weigend, 1997). By analyzing time series with ICA, we often need to select several dominant components that are robust in series forecasting. We can also select some factors which have an economic interpretation. Such factors could include news (government intervention, natural or man-made disasters, political upheaval), response to very large trades and of course, unexplained noise (Cheung and Xu, 2001).

3 Value-at-Risk Computation

Value-at-Risk (VaR) shows the loss over a preset time horizon for a given probability. In this section, we present the framework of VaR estimation in the case of multi-asset portfolio and, we after show the link between ICA and computation of VaR.

3.1 General Framework of VaR Estimation

Let R_i be the return of the ith asset class of the entire portfolio during a specific time interval. If a portfolio consists of n asset classes, then the return of the portfolio can be written as:

$$R_P = \sum_{i=1}^{n} w_i R_i = \mathbf{W}'\mathbf{R} \tag{3}$$

An estimate of the VaR at the $\alpha\%$ level for a global portfolio can be defined as:

$$VaR_{P,\,\alpha} = G^{-1}_{P,\,\alpha}(R_P) \times V_P \tag{4}$$

where $G^{-1}_{P,\,\alpha}(.)$ is the quantile function of a portfolio and V_P is the market value. The VaR measure is not a coherent measure in the sense that it is not additive; meaning here that the Quantile Function is unknown and thus should be also estimated. Investors are in general not able to aggregate their portfolio because of liquidity problems and/or transaction costs between the different security markets.

3.2 Computing VaR with ICA

Since the return of an asset class is often caused by the change of the underlying risk factors, one can write the return of the ith asset class as:

$$R_i - \alpha_i = B_i'\mathbf{F} + \epsilon_i$$

where $B_i(1 \text{ x } k)$ is a vector of constant coefficients for the ith asset class and $\mathbf{F}(k \text{ x } 1)$ is a random vector of risk factors. The random error ϵ_i with mean 0 accounts for the information not captured by the risk factors.

In traditional factor models, we require the factors are uncorrelated to each other. It is possible that the factors are uncorrelated, but not independent. Therefore, the typical assumption on uncorrelated factors in the factor models cannot guarantee the return of the portfolio be free from the influence of other factors. On the contrary, if the factors are independent to each other, it is possible to construct a portfolio which is free from the influence of neither factors. Loretan (1997) propose PCA to extract common factors. ICA is an ideal candidate for the extraction of independent factors.

We then fit a parametric models for the independent signals of ICA or uncorrelated factor of PCA. Since we need to capture the extreme events and focus on the tails of the portfolio return distribution, a leptokurtic distribution, like t Location Scale distribution is used (This choose of leptokurtic models is motivate by the fact that ICA determines risk factors by maximizing their non-gaussianity). The Gaussian distribution is also fitted in the ICA and PCA factors.

4 Experiments and Results

To investigate the effectiveness of ICA techniques for Value-at-Risk computation, we apply ICA to data from the French Stock Market. We use intraday prices from January 2000 until March 2000 of 7 firms. The preprocessing consists of two steps: we obtain the intraday stock returns (in this paper, 10 minutes subsampling is used) and normalize the resulting values to lie within the range [-1,1]. We performed ICA on the stock returns using the FastICA algorithm (Hyvärinen *et al.*, 2001) described in preview section. In all the experiments, we assume that the number of stocks equals the number of sources supplied to the mixing model. Figure 1 (a) shows a independent components (ICs) obtained from the algorithm and figure 1 (b) shows the PCA factors.

With the rebuilt stock price, we compute a ICA-based VaR of the returns of portfolio with equal weight and the corresponding VaR computation (Normal VaR and Historical VaR). A backtest is performed using Loss Function-based Backtests (see Christoffersen and Pelletier, 2003). The failure rates are given in Table 1. Generally speaking, the results are quite good for all VaR computation methods. But we clearly observe that ICA-VaR is the best method for many level. This result confirms the fact that, the classical evaluation of VaR model cannot be use in the case of High Frequency prices.

5 Conclusion

In this paper, we have applied the Independent Component Analysis (ICA) to compute the Value-at-Risk of a portfolio. We show in an environment characterised by non-linearities and non-Gaussianity, the ICA methodology can advantageously reveal an underlying structure in intraday data that can be exploited for risk measurement. Interestingly, the "noise" we observe in the intraday data may be attributed to signals with a small relative range and therefore the VaR computation with the denoised data (ICA-VaR) should be more robust than VaR computation done directly with the original data. Finally, based on our first results, the ICA technique, directly dealing with the factors that drive the market, could have some interest in some further financial applications, such as stress test assessment or asset pricing.

References

[1] A. Back A. and A. Weigend, A First Application of Independent Component Analysis to Extracting Structure from Stock Returns, *International Journal of Neural Systems 8*, 1-16, 1997.

[2] J. Cardoso, High-order Contrasts for Independent Component Analysis, *Neural Computation 11(1)*, 157-192, 1999.

[3] Y. Cheung and L. Xu, Independent Component Ordering in ICA Time Series Analysis, *Neurocomputing 41*, 145-152, 2001.

[4] P. Christoffersen and D. Pelletier, Back Testing Value-at-Risk: A Duration-Based Approach, *Journal of Financial Econometrics*, 2:84-108, 2003.

[5] P. Comon, Independent Component Analysis - a New Concept?, *Signal Processing 36*, 287-314, 1994.

[6] M. Dacorogna, U. Müller, R. Gencay, R. Olsen and O. Pictet, *An Introduction to High-Frequency Finance*, Academic Press, San Diego, CA, 2001.

[7] P. Giot and S. Laurent, Modelling Daily Value-at-Risk using Realized Volatility and ARCH type Models, *Journal of Empirical Finance*, 11:379-398, 2004.

[8] A. Hyvärinen, Fast and Robust Fixed-point Algorithms for Independent Component Analysis, *IEEE Transactions on Neural Networks*, 10:626-634, 1999.

[9] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley&Sons, New York, 2001.

[10] M. Loretan, Generating Market Risk Scenarios using Principal Component Analysis: Methodological and Practical Considerations, *manuscript*, Federal Reserve Board, 60 pages, 1997.