Author(s): Matilainen, M.; Miettinen, Jari; Nordhausen, K.; Taskinen, Sara

Title: ICA and stochastic volatility models

Year: 2016

Version:

Please cite the original version:

Matilainen, M., Miettinen, J., Nordhausen, K., & Taskinen, S. (2016). ICA and stochastic volatility models. In S. Aivazian, P. Filzmoser, & Y. Kharin (Eds.), CDAM 2016 : Proceedings of the XI International Conference on Computer Data Analysis and Modeling (pp. 30-37). Belarusian State University Publishing House.

ICA and stochastic volatility models

M. Matilainen, J. Miettinen*, K. Nordhausen, H. Oja, S. Taskinen*
Department of Mathematics and Statistics, University of Turku
FI-20014 Turku, Finland
**Department of Mathematics and Statistics, University of Jyvaskyla*
FI-40014 Jyvaskyla, Finland
e-mail: `markus.matilainen@utu.fi, jari.p.miettinen@jyu.fi, klaus.nordhausen@utu.fi, hannu.oja@utu.fi, sara.l.taskinen@jyu.fi`

Abstract

We consider multivariate time series where each component series is an unknown linear combination of latent mutually independent stationary time series. Multivariate financial time series have often periods of low volatility followed by periods of high volatility. This kind of time series have typically non-Gaussian stationary distributions, and therefore standard independent component analysis (ICA) tools such as fastICA can be used to extract independent component series even though they do not utilize any information on temporal dependence. In this paper we review some ICA methods used in the context of stochastic volatility models. We also suggest their modifications which use nonlinear autocorrelations to extract independent components. Different estimates are then compared in a simulation study.
Keywords: blind source separation, GARCH model, nonlinear autocorrelation, multivariate time series

1 Introduction

In this paper we assume that the observed p-variate time series $\boldsymbol{x} = (\boldsymbol{x}_t)_{t=0,\pm1,\pm2,\ldots}$ follows the basic *independent component (IC) model*

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{z}_t, \quad t = 0, \pm1, \pm2, \ldots,$$

where $\boldsymbol{\mu}$ is a p-variate location vector, $\boldsymbol{\Omega}$ is a full-rank $p \times p$ mixing matrix and $\boldsymbol{z} = (\boldsymbol{z}_t)_{t=0,\pm1,\pm2,\ldots}$ is an unobservable p-variate stationary time series such that

$$(i) \quad \mathrm{E}(\boldsymbol{z}_t) = \boldsymbol{0}, \quad (ii) \quad \mathrm{COV}(\boldsymbol{z}_t) = \boldsymbol{I}_p \text{ and}$$
$$(iii) \text{ the component series of } \boldsymbol{z} \text{ are independent.}$$

Then $\boldsymbol{x}$ is also stationary with $\mathrm{E}(\boldsymbol{x}_t) = \boldsymbol{\mu}$ and $\mathrm{COV}(\boldsymbol{x}_t) = \boldsymbol{\Sigma} = \boldsymbol{\Omega}\boldsymbol{\Omega}'$. In *independent component analysis (ICA)* the goal is to find, using the observed time series $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, an estimate of an unmixing matrix $\boldsymbol{W}$ such that $\boldsymbol{W}\boldsymbol{x} = (\boldsymbol{W}\boldsymbol{x}_t)_{t=0,\pm1,\pm2,\ldots}$ has independent component series.

The IC model has recently achieved a lot of attention in financial time series analysis as complicated p-variate time series models can then be replaced by p simple univariate (e.g. ARMA or GARCH) models in parameter estimation and prediction problems.

The model also serves as a dimension reduction tool as often only few component series in $\boldsymbol{z}$ are relevant and the rest of the components just present noise. For some recent contributions, see [3, 6, 7, 11, 17].

In the literature standard ICA methods, such as fastICA, are often used to estimate an unmixing matrix $\boldsymbol{W}$ in a time series context although such methods only use the marginal distribution of $\boldsymbol{x}_t$ and make no use of the information on temporal dependence. On the other hand, there exist second order source separation methods, like SOBI [1], which are particularly popular for analyzing biomedical data. Such methods use autocovariances and cross-autocovariances for the estimation. They are capable of separating time series with nonzero linear autocorrelations, but they do not utilize nonlinear autocorrelations.

Volatility clustering is a common feature in economic and financial time series, i.e. there are periods of lower and higher volatility. As the transitions between such periods do not typically have any clear pattern, they are treated as random occurrences. There are a vast amount of different models that have been invented for such situations. Among stochastic volatility models, the GARCH process [2] has been the most popular one. Another popular model is the SV (Stochastic Volatility) model [20]. In our simulations we consider these two models. For further information on stochastic volatility and a recent overview of stochastic volatility models, see for example [13].

In this paper we review various independent component estimators that use nonlinear autocorrelations, and compare their performance to that of fastICA in a simulation study where the independent time series components come from GARCH and SV models. The paper has the following structure. First, in Section 2 we define the univariate stochastic volatility models. In Section 3 we discuss the ICA methods which are considered in this paper. Section 4 consists of the simulation study.

2 Stochastic volatility models for univariate series

Among stochastic volatility models, the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) process [2] has been the most popular one. A univariate GARCH(p, q) process is given by

$$x_t = \sigma_t \epsilon_t,$$

where ϵ_t is an independent white noise process and σ_t^2 a deterministic conditional variance process

$$\sigma_t^2 = Var(x_t|\mathcal{F}_{t-1}) = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

with $\omega > 0$ and $\alpha_i, \beta_j \geq 0 \; \forall i, j$. For (second order) stationarity, $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

Another popular model is the SV (Stochastic Volatility) model [20], defined as

$$\begin{aligned} x_t &= e^{h_t/2} \epsilon_t, \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma \eta_t, \end{aligned}$$

where ϵ_t and η_t are two independent white noise innovation processes. Parameter μ is the level, ϕ is the persistence and $\sigma\eta_t$ is the volatility of log-variance. The process h_t is called the volatility process and it is strongly stationary with $N(0,1)$ innovations and initial state $h_0 \sim N(\mu, \sigma^2/(1-\phi^2))$. For stationarity, we require $|\phi| < 1$ and $\mu \in \mathbb{R}$.

3 Source separation for multivariate time series

Under our model assumption, the standardized multivariate series of $\boldsymbol{x}_t$ is given by $\boldsymbol{x}_t^{st} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu})$. One of the key results in ICA states that there exists an orthogonal matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$ such that $\boldsymbol{z}_t = \boldsymbol{U}\boldsymbol{x}_t^{st}$ (up to signs and order of the components) [16]. Here $\boldsymbol{z}_t$ denotes the vector of independent series. The final unmixing matrix functional is then given by $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1/2}$. The estimate of $\boldsymbol{W}$ is then obtained by replacing $\boldsymbol{\Sigma}$ and $\boldsymbol{U}$ by their sample counterparts. For finding $\boldsymbol{U}$, we next list the criterion functions in different approaches.

In the symmetric *fastICA* [9] approach and symmetric squared fastICA [15], $\boldsymbol{U}$ maximizes

$$\sum_{i=1}^{p} |\mathrm{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})\right]| \quad \text{and} \quad \sum_{i=1}^{p} \left(\mathrm{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})\right]\right)^2,$$

with a choice of a twice continuously differentiable, nonlinear and nonquadratic function G such that $E[G(y)] = 0$ if $y \sim N(0,1)$. Two common options are $G(z) = z^4 - 3$ and $G(z) = \log(\cosh(z)) - E[G(y)]$, where $y \sim N(0,1)$. Notice that both utilize only the stationary (marginal) distribution of $\boldsymbol{x}_t$.

The estimators presented below make use of the joint distributions of $(\boldsymbol{x}_t, \boldsymbol{x}_{t+k})$, $k = 1, 2, \ldots$. The classical *SOBI* uses only second moments and it was originally defined as a method which jointly diagonalizes several autocovariance matrices. However, SOBI can be reformulated as the maximizer of

$$\sum_{i=1}^{p}\sum_{k=1}^{K} \left(\mathrm{E}\left[(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})\right]\right)^2.$$

The solution is unique if, for all pairs $i \neq j$ there exists a k, $1 \leq k \leq K$, such that $\mathrm{E}(z_{t,i}z_{t+k,i}) \neq \mathrm{E}(z_{t,j}z_{t+k,j})$. SOBI fails to separate GARCH and SV time series as all lagged autocovariances are then zero.

The *gFOBI* procedure proposed in [12] maximizes a sum of fourth moments

$$\sum_{i=1}^{p}\sum_{k=1}^{K} \left(\mathrm{E}\left[(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})||\boldsymbol{x}_t^{st}||^2\right]\right)^2.$$

For $K = 0$, the regular ICA method FOBI [4] is obtained.

The *gJADE* procedure [12], in turn, uses a much richer sum of fourth cumulants and maximizes

$$\sum_{i=1}^{p}\sum_{r=1}^{p}\sum_{s=1}^{p}\sum_{k=1}^{K} \left(\kappa(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st}, \boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st}, \boldsymbol{x}_{t,r}^{st}, \boldsymbol{x}_{t,s}^{st})\right)^2.$$

where

$$\kappa(z_1, z_2, z_3, z_4) = \mathrm{E}(z_1 z_2 z_3 z_4) - \mathrm{E}(z_1 z_2)\mathrm{E}(z_3 z_4) - \mathrm{E}(z_1 z_3)\mathrm{E}(z_2 z_4) - \mathrm{E}(z_1 z_4)\mathrm{E}(z_2 z_3).$$

Again, for $K = 0$, the regular ICA method JADE [5] is obtained. Both, gFOBI and gJADE, were created having stochastic volatility models in mind.

FastICA does not use any knowledge of temporal dependence, but there exist some fixed-point algorithms aimed for time series context. The *FixNA* (Fixed-point algorithm for maximizing the nonlinear autocorrelation) method was introduced in [19], and its criterion function to be maximized is

$$D_1(\boldsymbol{U}) = \sum_{i=1}^{p} \sum_{k=1}^{K} E\left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}) G(\boldsymbol{u}_i' \boldsymbol{x}_{t+k}^{st})\right],$$

where G is a twice continuously differentiable function. The G-functions suggested in [19] are $G(z) = \log(\cosh(z))$ and $G(z) = z^2$.

A similar function to be maximized is of the form

$$D_2(\boldsymbol{U}) = \sum_{i=1}^{p} \sum_{k=1}^{K} \left| E\left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}) G(\boldsymbol{u}_i' \boldsymbol{x}_{t+k}^{st})\right] - E\left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st})\right]^2 \right|,$$

and we will denote it as *FixNA2*. It was first proposed in [8], however only with $G(z) = z^2$, and $K = 1$. We further similarly suggest a natural extension of SOBI with the criterion function

$$D_3(\boldsymbol{U}) = \sum_{i=1}^{p} \sum_{k=1}^{K} \left(E\left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}) G(\boldsymbol{u}_i' \boldsymbol{x}_{t+k}^{st})\right] - E\left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st})\right]^2 \right)^2.$$

As a variant of SOBI, we call this estimator *vSOBI*.

To obtain the estimating equations for matrix $\boldsymbol{U}$, the Lagrangian multiplier technique can be used as in [14]. The Lagrangian function to be optimized is

$$L(\boldsymbol{U}, \boldsymbol{\Lambda}) = D_r(\boldsymbol{U}) - \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \lambda_{ij} \boldsymbol{u}_i' \boldsymbol{u}_j - \sum_{i=1}^{p} \lambda_{ii} (\boldsymbol{u}_i' \boldsymbol{u}_i - 1), \text{ for } r = 1, 2, 3,$$

where $\boldsymbol{\Lambda} = (\lambda_{ij})$ is a symmetric matrix that contains $p(p+1)/2$ Lagrangian multipliers. Write next

$$\boldsymbol{T}_{r,i} = \boldsymbol{T}_{r,i}(\boldsymbol{U}) = \frac{\partial}{\partial \boldsymbol{u}_i} D_r(\boldsymbol{U}), \ i = 1, \ldots, p, \ r = 1, 2, 3,$$

and $\boldsymbol{T}_r = \boldsymbol{T}_r(\boldsymbol{U}) = (\boldsymbol{T}_{r,1}, \ldots, \boldsymbol{T}_{r,p})'$. Solving the optimizing problem then gives the estimating equations for $\boldsymbol{U}$, namely,

$$\boldsymbol{U}\boldsymbol{T}_r' = \boldsymbol{T}_r \boldsymbol{U}' \ \text{ and } \ \boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_p,$$

or, equivalently,

$$\boldsymbol{U} = (\boldsymbol{T}_r \boldsymbol{T}_r')^{-1/2} \boldsymbol{T}_r.$$

For some tolerance limit ε and initial value $\boldsymbol{U}_0$, this leads to Algorithm 1.

Data: Standardized time series $\boldsymbol{x}_t^{st} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu})$
Result: $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1/2}$
$\boldsymbol{U}_{old} = \boldsymbol{U}_0$;
$\Delta = \infty$;
while $\Delta > \varepsilon$ **do**
  $\boldsymbol{T}_r = \boldsymbol{T}_r(\boldsymbol{U}_{old})$;
  $\boldsymbol{U}_{new} = (\boldsymbol{T}_r\boldsymbol{T}_r')^{-1/2}\boldsymbol{T}_r$;
  $\Delta = ||\boldsymbol{U}_{new} - \boldsymbol{U}_{old}||$;
  $\boldsymbol{U}_{old} = \boldsymbol{U}_{new}$;
end
$\boldsymbol{U} = \boldsymbol{U}_{new}$;

Algorithm 1: Algorithm for maximizing the criterion function $D_r,\ r = 1, 2, 3$.

4 Simulation study

The following simulations are conducted using R 3.2.2 [18] with the packages fGarch, fICA, JADE and tsBSS. In the simulation study we compare due to space limitations only the following methods:

- FixNA, FixNA2 and vSOBI with $G(z) = z^2$ and lags $1, \ldots, 12$
- symmetric fastICA and symmetric squared fastICA with $G(z) = z^4 - 3$
- gFOBI, gJADE with lags $0, 1, \ldots, 12$ and SOBI with lags $1, \ldots, 12$.

The comparison is based on the Minimum Distance Index [10], which is defined as

$$\hat{D} = \hat{D}(\hat{\boldsymbol{W}}) = \frac{1}{\sqrt{p-1}} \inf_{\boldsymbol{C} \in \boldsymbol{\mathcal{C}}} ||\boldsymbol{C}\hat{\boldsymbol{W}}\boldsymbol{\Omega} - \boldsymbol{I}_p||,$$

where $\boldsymbol{\mathcal{C}}$ is the set of all matrices with exactly one non-zero element in each row and column, and $||\cdot||$ is the Frobenius (matrix) norm. The index has the range $0 \leq \hat{D} \leq 1$, where zero indicates perfect separation.

For time series of lengths $T = 100, 200, \ldots, 25600$ we report the averages $T(p-1)\hat{D}^2$ based on 2000 repetitions. Such an average represents a global measure of variation of an unmixing matrix, see [10] for details. As all the methods are affine equivariant, we choose wlog $\boldsymbol{\Omega} = \boldsymbol{I}_p$ and consider the following two 4-variate settings:

- **GARCH setting:** The sources are four GARCH$(1, 1)$ processes with normal innovations. The parameters (α_1, β_1) are chosen so that the first eight moments are finite, and are: (i) (0.05, 0.9), (ii) (0.1, 0.7), (iii) (0.1, 0.8) and (iv) (0.2, 0.5).
- **SV setting:** In the second setup the four sources are SV processes with normal innovations and (μ, ϕ, σ)-parameter vectors $(-10, 0.8, 0.1)$, $(-10, 0.9, 0.2)$, $(-10, 0.9, 0.3)$ and $(-10, 0.95, 0.4)$. Again, all the first eight moments exist.

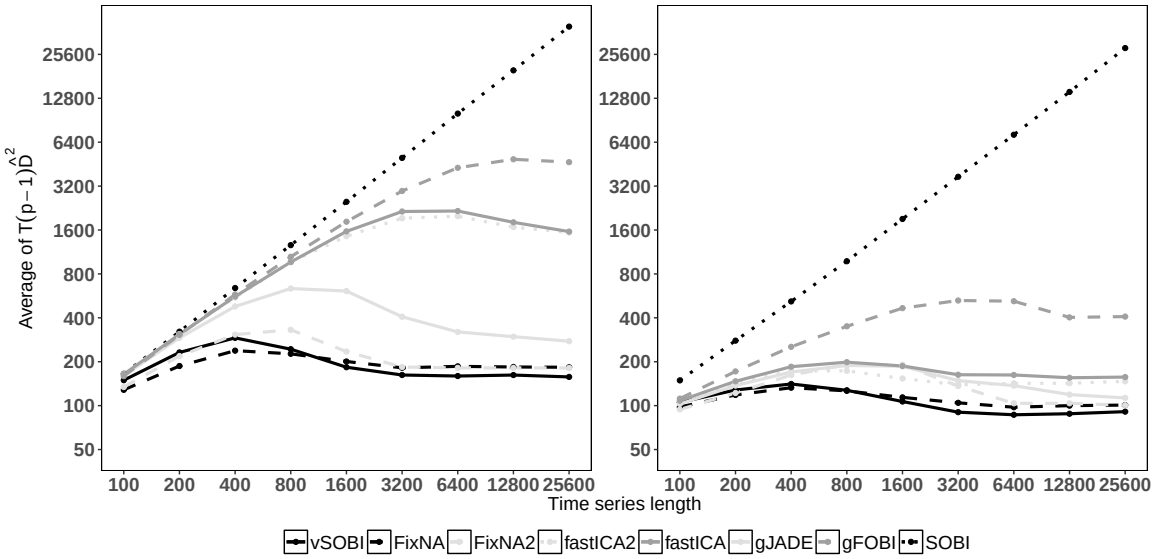

Figure 1: Comparison of performance of algorithms in the GARCH setting (left panel) and SV setting (right panel).

Figure 1 summarizes the results for both settings. As expected, SOBI does not work here. The proposed vSOBI estimator works very well in both cases and outperforms all the other estimators. Interestingly, both fastICA algorithms perform well in the SV example but not in the GARCH example. FastICA2 algorithm produces slightly better results than the fastICA algorithm. While gJADE works quite well in both cases, gFOBI has much poorer performance. FixNA and FixNA2 algorithms are among the best methods.

Convergence of FixNA2 algorithm and both fastICA algorithms is low in short time series (see Figure 2), but gets much better when the time series length increases. Convergence percentage of vSOBI is also good, and in time series of length 800 onwards very close to 100%. SOBI, gFOBI and gJADE have very few convergence issues, if any.

5 Discussion

In this paper we surveyed different blind source separation methods suitable for multivariate time series with stochastic volatility features. Such methods were earlier quite scattered in the literature. We also suggested some small modification yielding the family of vSOBI estimators which showed in our simulations the best performance. We have shown here the simulation results of vSOBI, both FixNA and both FastICA algorithms only based on G functions of the form $G(z) = z^c$. However, in an extended version of this paper we plan to have a larger simulation study, including for example also $\log(\cosh(z))$ as a nonlinearity.

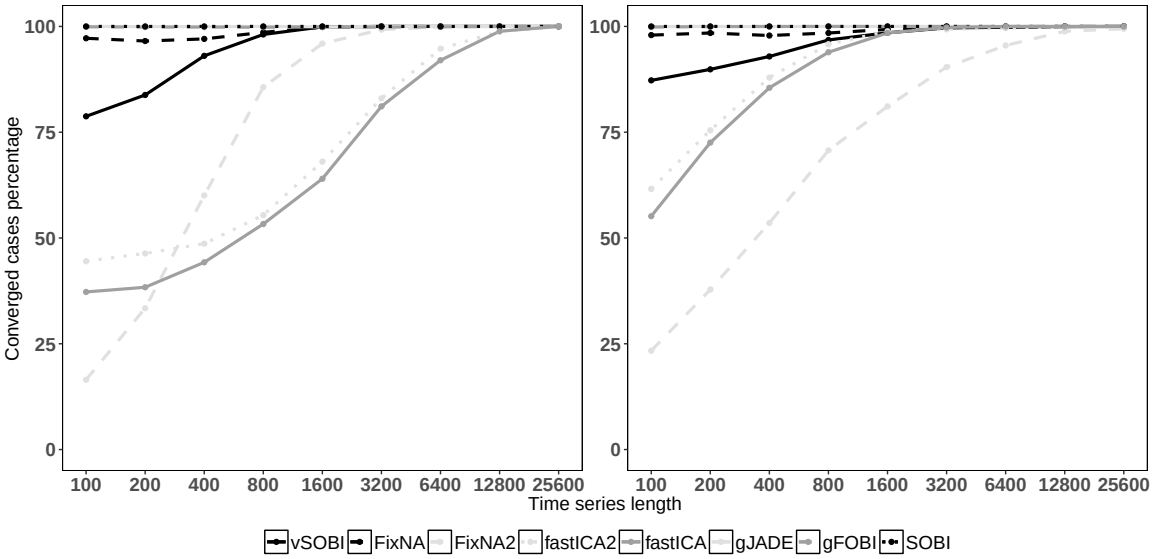

Figure 2: Comparison of convergence percentages of algorithms in the GARCH setting (left panel) and SV setting (right panel).

Acknowledgements

This work was supported by the Academy of Finland (grants 251965, 256291 and 268703).

References

[1] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE T. Signal Proc.*, 45:434–444, 1997.

[2] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31(3):307–327, 1986.

[3] S. A. Broda and M. S. Paolella. CHICAGO: A fast and accurate method for portfolio risk calculation. *J. Financ. Economet.*, 7(4):412–436, 2009.

[4] J.-F. Cardoso. Source separation using higher order moments. In *Int. Conf. Acoust. Spee.*, pages 2109–2112, 1989.

[5] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. In *IEE-Proc.-F*, volume 140, pages 362–370, 1993.

[6] Y. Chen, W. Härdle, and V. Spokoiny. Portfolio value at risk based on independent component analysis. *J. Comput. Appl. Math.*, 205:594–607, 2007.

[7] A. García-Ferrer, E. González-Prieto, and D. Peña. A conditionally heteroskedastic independent factor model with an application to financial stock returns. *Int. J. Forecasting*, 28(1):70 – 93, 2012.

[8] A. Hyvärinen. Blind source separation by nonstationarity of variance: A cumulant-based approach. *IEEE T. Neural Networ.*, 12(6):1471–1474, 2001.

[9] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analyis. *Neural Comput.*, 9:1483–1492, 1997.

[10] P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila. A new performance index for ICA: Properties computation and asymptotic analysis. In V. Vigneron et al., editor, *LVA/ICA 2010. LNCS*, volume 6365, pages 229–236, Heidelberg, 2010. Springer.

[11] C.-J. Lu, J.-Y. Wu, and T.-S. Lee. Application of independent component analysis preprocessing and support vector regression in time series prediction. In *International Joint Conference on Computational Sciences and Optimization*, volume 1, pages 468–471, 2009.

[12] M. Matilainen, K. Nordhausen, and H. Oja. New independent component analysis tools for time series. *Stat. Probabil. Lett.*, 105:80–87, 2015.

[13] D.S. Matteson and D. Ruppert. Time-series models of dynamic volatility and correlation. *IEEE Signal Proc. Mag.*, 28(5):72–82, 2011.

[14] J. Miettinen, K. Illner, K. Nordhausen, H. Oja, S. Taskinen, and F. Theis. Separation of uncorrelated stationary time series using autocovariance matrices. *J. Time Ser. Anal.*, 37(3): 337–354, 2016.

[15] J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, and J. Virta. The squared symmetric FastICA estimator, 2015. http://arxiv.org/abs/1512.05534.

[16] J. Miettinen, S. Taskinen, K. Nordhausen, and H. Oja. Fourth moments and independent component analysis. *Stat. Sci.*, 30:372–390, 2015.

[17] E. Oja, K. Kiviluoto, and S. Malaroiu. Independent component analysis for financial time series. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 111–116, 2000.

[18] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015. R version 3.2.2.

[19] Z. Shi, Z. Jiang, and F. Zhou. Blind source separation with nonlinear autocorrelation and non-gaussianity. *J. Comput. Appl. Math.*, 223(1):908–915, 2009.

[20] S. J. Taylor. Financial returns modelled by the product of two stochastic processes – a study of daily sugar prices 1961–79. In O. D. Anderson, editor, *Time Series Analysis: Theory and Practice 1*, pages 203–216. Springer, North-Holland, Amsterdam, 1982.