On Independent Component Analysis with Stochastic Volatility Models

Markus Matilainen
University of Turku

Jari Miettinen
University of Jyvaskyla

Klaus Nordhausen
University of Turku

Hannu Oja
University of Turku

Sara Taskinen
University of Jyvaskyla

Abstract

Consider a multivariate time series where each component series is assumed to be a linear mixture of latent mutually independent stationary time series. Classical independent component analysis (ICA) tools, such as fastICA, are often used to extract latent series, but they don't utilize any information on temporal dependence. Also financial time series often have periods of low and high volatility. In such settings second order source separation methods, such as SOBI, fail. We review here some classical methods used for time series with stochastic volatility, and suggest modifications of them by proposing a family of vSOBI estimators. These estimators use different nonlinearity functions to capture nonlinear autocorrelation of the time series and extract the independent components. Simulation study shows that the proposed method outperforms the existing methods when latent components follow GARCH and SV models. This paper is an invited extended version of the paper presented at the CDAM 2016 conference.

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 $\boldsymbol{x} = (\boldsymbol{x}_t)_{t=0,\pm1,\pm2,\dots}$ follows the basic *independent component (IC) model*

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

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

$$(i) \ \ \mathsf{E}(\boldsymbol{z}_t) = \boldsymbol{0}, \ \ \ (ii) \ \ \mathrm{COV}(\boldsymbol{z}_t) = \boldsymbol{I}_p \ \ \text{and}$$

$$(iii) \ \text{the component series of } \boldsymbol{z} \text{ are independent.}$$

Then $\boldsymbol{x}$ is also stationary with $\mathsf{E}(\boldsymbol{x}_t) = \boldsymbol{\mu}$ 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 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_T$, an

estimate of an unmixing matrix $\boldsymbol{W}$ such that $\boldsymbol{W}\boldsymbol{x} = (\boldsymbol{W}\boldsymbol{x}_t)_{t=0,\pm1,\pm2,...}$ has independent component series.

The IC model has recently received 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 while the rest just present noise. For some recent contributions, see Broda and Paolella (2009); Chen, Härdle, and Spokoiny (2007); García-Ferrer, González-Prieto, and Peña (2012); Lu, Wu, and Lee (2009); Oja, Kiviluoto, and Malaroiu (2000).

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 (Belouchrani, Abed Meraim, Cardoso, and Moulines 1997), 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. In our simulations we consider two popular choices, the GARCH model (Bollerslev 1986) and the SV model (Taylor 1982). For further information on stochastic volatility and a recent overview of stochastic volatility models, see for example Matteson and Ruppert (2011).

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 independent time series components follow GARCH and SV models. The paper has the following structure. First, in Section 2 we define the aforementioned univariate stochastic volatility models. In Section 3 we discuss the ICA methods which are considered in this paper and suggest our extension. In Section 4 we show that this extension has the important affine equivariance property. Section 5 consists of the simulation study.

2. Stochastic volatility models for univariate series

There are several different stochastic volatility models. Here we concentrate on two widely used ones. The first one is GARCH (Generalized Autoregressive Conditional Heteroscedasticity) process (Bollerslev 1986) defined as follows. 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 is a conditional variance process

$$\sigma_t^2 = Var(x_t | x_u, u < t) = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

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

Another popular model is the SV (Stochastic Volatility) model (Taylor 1982), defined as

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

where ϵ_t and η_t are two independent white noise innovation processes. The parameter μ is the level, ϕ is the persistence and $\sigma\eta_t$ is the volatility of the 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) (Miettinen, Taskinen, Nordhausen, and Oja 2015). 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* (Hyvärinen and Oja 1997) approach $\boldsymbol{U}$ maximizes

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

and in the symmetric squared fastICA (Miettinen, Nordhausen, Oja, Taskinen, and Virta 2017a) $\boldsymbol{U}$ maximizes

$$\sum_{i=1}^{p} \left(\mathsf{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st}) \right] \right)^2.$$

Here a twice continuously differentiable, nonlinear and nonquadratic function G is chosen so that $E[G(y)] = 0$ if $y \sim N(0,1)$. Two common choices for G are $G(z) = z^4 - 3$ and $G(z) = \log(\cosh(z)) - E[\log(\cosh(y))]$, 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 so that $\boldsymbol{U}$ maximizes

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

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

The *gFOBI* procedure proposed in Matilainen, Nordhausen, and Oja (2015) maximizes a sum of fourth moments

$$\sum_{i=1}^{p} \sum_{k=0}^{K} \left(\mathsf{E}\left[(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})\|\boldsymbol{x}_t^{st}\|^2 \right] \right)^2,$$

where $\|\cdot\|$ is the Frobenius (matrix) norm. For $K = 0$, the regular ICA method FOBI (Cardoso 1989) is obtained.

The *gJADE* procedure (Matilainen *et al.* 2015), in turn, uses a much richer sum of fourth cumulants and maximizes

$$\sum_{i=1}^{p} \sum_{r=1}^{p} \sum_{s=1}^{p} \sum_{k=0}^{K} \left(\kappa(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st}, \boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st}, \boldsymbol{x}_{t,r}^{st}, \boldsymbol{x}_{t,s}^{st}) \right)^2,$$

where

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

Again, for $K = 0$, the regular ICA method JADE (Cardoso and Souloumiac 1993) 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 exists some fixed-point algorithms aimed for a time series context. The *FixNA* (Fixed-point algorithm for maximizing the nonlinear autocorrelation) method was introduced in Shi, Jiang, and Zhou (2009), and its criterion function to be maximized is

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

where G is a twice continuously differentiable function. The G-functions suggested in Shi *et al.* (2009) are $G(z) = \log(\cosh(z))$ and $G(z) = z^2$.

A similar function to be maximized is of the form

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

and we will denote it as *FixNA2*. It was first proposed in Hyvärinen (2001), however only with $G(z) = z^2$ and $K = 1$. We further similarly suggest a natural extension of SOBI with the criterion function

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

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

The term $\mathsf{E} \left[G(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}) \right]^2$ in $D_2(\boldsymbol{U})$ and $D_3(\boldsymbol{U})$ is used to normalize the summands. Notice that in SOBI, $G(z) = z$, and hence the aforementioned term equals to 0. When $G(z) = z^2$, the term equals to 1.

Remark 1. Instead of using lags $k = 1, 2, \ldots, K$, also more flexible lag combinations could be used (see Taskinen, Miettinen, and Nordhausen 2016, for example).

To obtain the estimating equations for matrix $\boldsymbol{U}$, the Lagrangian multiplier technique can be used as in Miettinen, Illner, Nordhausen, Oja, Taskinen, and Theis (2016). 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 the $p(p+1)/2$ 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 $\boldsymbol{T}_r = \boldsymbol{T}_r(\boldsymbol{U}) = (\boldsymbol{T}_{r,1}, \ldots, \boldsymbol{T}_{r,p})'$. Now

$$\frac{\partial}{\partial \boldsymbol{u}_i} L(\boldsymbol{U}, \boldsymbol{\Lambda}) = \boldsymbol{T}_{r,i} - \sum_{j<i} \lambda_{ji} \boldsymbol{u}_j - \sum_{j>i} \lambda_{ij} \boldsymbol{u}_j - 2\lambda_{ii} \boldsymbol{u}_i = 0.$$

Then multiplying this by $\boldsymbol{u}_j'$ from the left side gives

$$\begin{cases} j > i: & \boldsymbol{u}_j' \boldsymbol{T}_{r,i} - \lambda_{ij} = 0 \\ j < i: & \boldsymbol{u}_j' \boldsymbol{T}_{r,i} - \lambda_{ji} = 0 \end{cases}$$

Now notice that $\boldsymbol{u}_j' \boldsymbol{T}_{r,i} - \lambda_{ji} = \boldsymbol{u}_i' \boldsymbol{T}_{r,j} - \lambda_{ij}$. Then

$$\boldsymbol{u}_j' \boldsymbol{T}_{r,i} = \boldsymbol{u}_i' \boldsymbol{T}_{r,j}, \ i \neq j, \ i, j = 1, \ldots, p.$$

From here the estimating equations for an orthogonal $\boldsymbol{U}$ can be obtained.

Result 2. The estimating equations for an orthogonal matrix $\boldsymbol{U}$ are

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

or, equivalently,

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

For vSOBI for example, to maximize $D_3(\boldsymbol{U})$, we need to calculate $\boldsymbol{T}_3 = (\boldsymbol{T}_{3,1}, \ldots, \boldsymbol{T}_{3,p})'$, where

$$\boldsymbol{T}_{3,i} = \sum_{i=1}^{p}\sum_{k=1}^{K}\Big(\mathsf{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})G(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})\right] - \mathsf{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})\right]^2\Big)$$
$$\cdot\Big(\mathsf{E}\left[G'(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})G(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})\boldsymbol{x}_t^{st}\right] + \mathsf{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})G'(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})\boldsymbol{x}_{t+k}^{st}\right]$$
$$-2\mathsf{E}\left[G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})\right]\mathsf{E}\left[G'(\boldsymbol{u}_i'\boldsymbol{x}_t^{st})\boldsymbol{x}_t^{st}\right]\Big),$$

for $i = 1, \ldots, p$.

Therefore in all three cases $r = 1, 2, 3$ the computation of $\boldsymbol{U}$ which maximizes $D_r(\boldsymbol{U})$ can be done iteratively given some initial orthogonal matrix $\boldsymbol{U}_0$ and some tolerance limit ε as described in the algorithm below.

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**
$\quad\Big|\quad \boldsymbol{T}_r = \boldsymbol{T}_r(\boldsymbol{U}_{old})$;
$\quad\Big|\quad \boldsymbol{U}_{new} = (\boldsymbol{T}_r\boldsymbol{T}_r')^{-1/2}\boldsymbol{T}_r$;
$\quad\Big|\quad \Delta = ||\boldsymbol{U}_{new} - \boldsymbol{U}_{old}||$;
$\quad\Big|\quad \boldsymbol{U}_{old} = \boldsymbol{U}_{new}$;
end
$\boldsymbol{U} = \boldsymbol{U}_{new}$;

4. Affine equivariance

In blind source separation it is desirable that an unmixing matrix estimator is affine equivariant, which means that the separation performance does not depend on the actual value of the mixing matrix $\boldsymbol{\Omega}$. Let $\boldsymbol{x}_t^* = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{b}$ be an affine transformation of $\boldsymbol{x}_t$, where $\boldsymbol{A}$ is a non-singular $p \times p$ matrix and $\boldsymbol{b}$ is a p-vector. Then a method is called affine equivariant if $\boldsymbol{W}^* = \boldsymbol{W}\boldsymbol{A}^{-1}$ and $\boldsymbol{W}\boldsymbol{x}_t = \boldsymbol{W}^*\boldsymbol{x}_t^*$, up to location shifts, sign changes and the order of the components.

Result 3. FixNA, FixNA2 and vSOBI algorithms are affine equivariant.

As an example, consider the affine equivariance of vSOBI algorithm: As $\mathrm{COV}(\boldsymbol{x}_t^*) = \boldsymbol{\Sigma}^* = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$, then by Theorem 2.1 in Ilmonen, Oja, and Serfling (2012), $(\boldsymbol{\Sigma}^*)^{-1/2} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}^{-1}$ and therefore $\boldsymbol{x}_t^{*st} = \boldsymbol{V}\boldsymbol{x}_t^{st}$ for some orthogonal $p \times p$ matrix $\boldsymbol{V}$. Let $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$ be the orthogonal matrix which maximizes

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

Next write $\boldsymbol{U}^* = \boldsymbol{U}\boldsymbol{V}' = (\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_p^*)'$, where $\boldsymbol{u}_i^* = \boldsymbol{V}\boldsymbol{u}_i$, for $i = 1, \ldots, p$. Now $\boldsymbol{U}^*$ is an orthogonal matrix and

$$G(\boldsymbol{u}_i^{*\prime}\boldsymbol{x}_t^{*st}) = G((\boldsymbol{V}\boldsymbol{u}_i)'(\boldsymbol{V}\boldsymbol{x}_t^{st})) = G(\boldsymbol{u}_i'\boldsymbol{x}_t^{st}).$$

Clearly also $G(\boldsymbol{u}_i^{*\prime}\boldsymbol{x}_{t+k}^{*st}) = G(\boldsymbol{u}_i'\boldsymbol{x}_{t+k}^{st})$, where $k = 1, \ldots, K$. Then $D_3(\boldsymbol{U}^*)$ is the maximum of the criterion function for $\boldsymbol{x}_t^*$, and

$$\boldsymbol{W}^* = \boldsymbol{U}^*(\boldsymbol{\Sigma}^*)^{-1/2} = \boldsymbol{U}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}^{-1} = \boldsymbol{W}\boldsymbol{A}^{-1}.$$

Using the same arguments it can be shown that also FixNA and FixNA2 algorithms are affine equivariant. Notice also that the fastICA versions discussed here, SOBI, gFOBI and gJADE are all affine equivariant.

5. Simulation study

The following simulations are conducted using R 3.2.2 (R Core Team 2016) with the packages fGarch (Wuertz and Rmetrics Core Team 2013), fICA (Miettinen, Nordhausen, Oja, and Taskinen 2014), JADE (Miettinen, Nordhausen, and Taskinen 2017b), stochvol (Kastner 2016) and tsBSS (Matilainen, Miettinen, Nordhausen, Oja, and Taskinen 2016). In our simulation studies we compare the following methods:

- FixNA, FixNA2 and vSOBI with both $G(z) = z^2$ and $G(z) = \log(\cosh(z))$ and lags $k = 1, \ldots, 12$

- symmetric fastICA and symmetric squared fastICA with both $G(z) = z^4 - 3$ and $G(z) = \log(\cosh(z)) - E[\log(\cosh(y))]$, where $y \sim N(0,1)$

- gFOBI, gJADE with lags $k = 0, 1, \ldots, 12$ and SOBI with lags $k = 1, \ldots, 12$

In the following we refer to G-functions as pow and lcosh. Hence the methods are denoted as vSOBI(pow) or FixNA2(lcosh), for example.

As a performance measure we use the Minimum Distance Index (Ilmonen, Nordhausen, Oja, and Ollila 2010), which is defined as

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

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 $T = 100, 200, 400, \ldots, 25600$ 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 Ilmonen *et al.* (2010) for details. As all the methods are affine equivariant, we can choose $\boldsymbol{\Omega} = \boldsymbol{I}_p$ without loss of generality 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.

Performance Figure 1 summarizes the results for both settings. Notice that for clarity we omitted the results for SOBI as it, as expected, failed to find any latent sources. From the four different fastICA versions only the best one is presented, and for FixNA only the pow version is shown as FixNA(lcosh) was much worse than any other FixNA method. FixNA(lcosh) also exhibited a strange convergence pattern as it was deteriorating when the sample sizes increased.
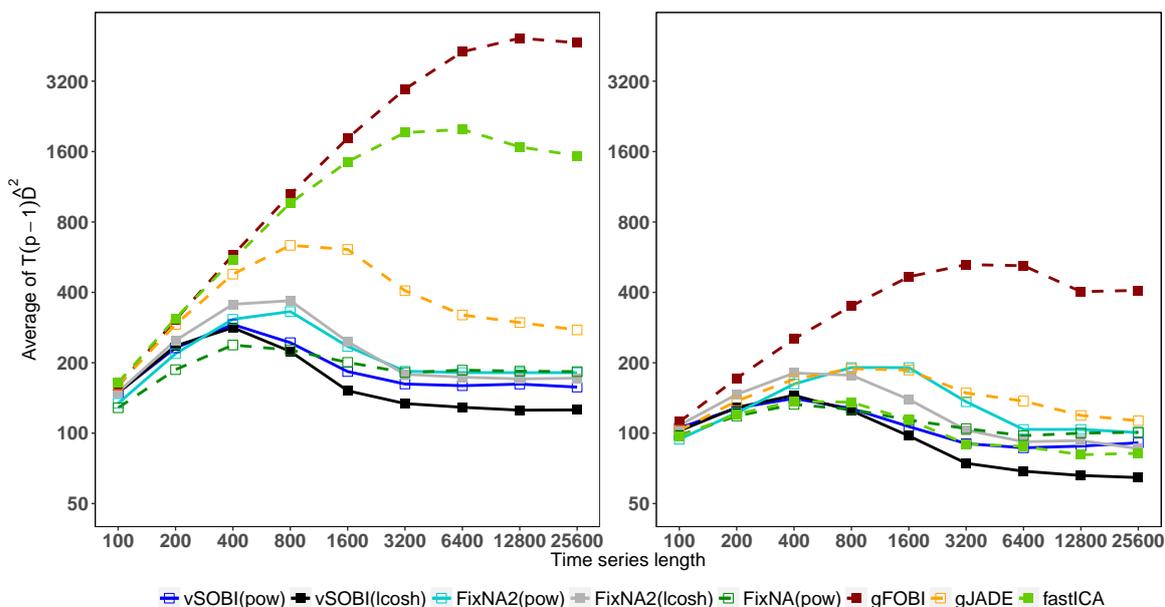
Figure 1: Comparison of performance of algorithms in the GARCH setting (left panel) and in the SV setting (right panel).

It is seen in Figure 1 that the proposed vSOBI estimator works very well both in GARCH and in SV settings and, when using $G(z) = \log(\cosh(z))$ as the nonlinearity function, it clearly outperforms all the other estimators. Squared fastICA algorithms produce slightly better results than the original fastICA algorithms. In GARCH setting the symmetric squared fastICA with $G(z) = z^4 - 3$ produces better results that other fastICA algorithms, but compared to other algorithms its performance is still not good, as only gFOBI and SOBI have poorer performance. In the SV setting however, while regular fastICA and squared fastICA algorithms with $G(z) = z^4 - 3$ are not among the best, both versions with $\log(\cosh(z))$ type nonlinearities give good results. The squared version is the better one and, when compared to other algorithms, we can see that only vSOBI with $G(z) = \log(\cosh(z))$ is better. Notice that unlike the other estimators, fastICA algorithms do not utilize any information on temporal dependence.

In general it seems better to use FixNA2 instead of FixNA, perhaps due to FixNA2 having the more natural centering part in the objective function. The performance of gFOBI is poor in both settings, but gJADE seems to work decently in SV setting.

Convergence The convergence percentages of the proposed vSOBI algorithms are good, and in time series of length 800 onwards very close to 100 %. SOBI has no converge issues, but as stated before, it has very poor performance. Also gFOBI and gJADE have very few convergence issues, and only when the time series is very short.

On the other hand, FixNA2 algorithms have lots of convergence problems in short time series, as they converged in less than 25 % of repetitions in time series of length 100. As the time series length increases, the convergence rate approaches 100 %. FastICA algorithms have more issues in the GARCH setting (less than 50 % of repetitions converging in the beginning) than in the SV setting.

FixNA(pow) has only some convergence issues in short time series. On the other hand, convergence of FixNA algorithm when using $G(z) = \log(\cosh(z))$ as the nonlinearity function is surprising in SV setting, as there are more convergence issues when time series length increases, contrary to what is expected (not shown in figures). For time series length 100 its convergence percentage is 98.5, while with time series length 25600 it is only 76.2. The reason
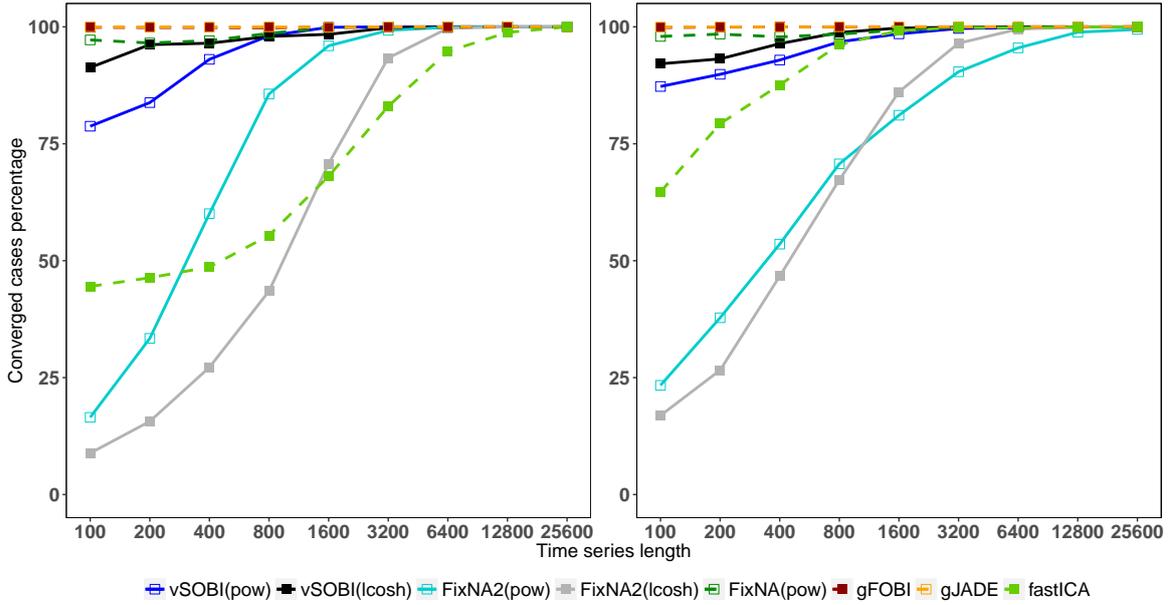
Figure 2: Comparison of convergence percentages of algorithms in the GARCH setting (left panel) and in the SV setting (right panel)

behind this behaviour is not yet known. In GARCH setting this behaviour is not seen after time series length of 800.

Converge results are summarized in Figure 2. Except for FixNA(lcosh) in the SV setting, convergence percentages of all algorithms are larger than 99.4 % in time series of length 25600.

6. 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 suggested a small modification to existing methods yielding the family of vSOBI estimators. The simulations were used to compare the vSOBI estimators with previously proposed methods using stochastic volatility models. The proposed vSOBI estimator with $G(z) = \log(\cosh(z))$ as the nonlinearity function showed the best performance among its competitors.

SOBI as a second order method was designed to function on time series with nonzero linear autocovariances, such as ARMA processes. In stochastic volatility models linear autocovariances are zero, and therefore SOBI is useless. On the other hand, the methods which exploit nonlinear autocorrelations, considered in this paper, are far from optimal in separating ARMA processes. In practical situations one can easily imagine a time series independent component model where some components are ARMA processes and others exhibit stochastic volatility features. Therefore it is desirable to derive methods which work in such cases.

In our future research we will consider a weighted combination of SOBI and vSOBI yielding for example the objective function

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

with some suitable weight $a \in [0, 1]$ and function G. This could still be generalized to case where SOBI and vSOBI parts use different combinations of lags.

Acknowledgements

References

Belouchrani A, Abed Meraim K, Cardoso JF, Moulines E (1997). "A Blind Source Separation Technique Based on Second Order Statistics." *IEEE Transactions on Signal Processing*, **45**, 434–444.

Bollerslev T (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, **31**(3), 307–327.

Broda SA, Paolella MS (2009). "CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation." *Journal of Financial Econometrics*, **7**(4), 412–436.

Cardoso JF (1989). "Source Separation Using Higher Order Moments." In *International Conference on Acoustics, Speech, and Signal Processing*, pp. 2109–2112.

Cardoso JF, Souloumiac A (1993). "Blind Beamforming for Non-Gaussian Signals." In *IEE Proceedings F*, volume 140, pp. 362–370.

Chen Y, Härdle W, Spokoiny V (2007). "Portfolio Value at Risk Based on Independent Component Analysis." *Journal of Computational and Applied Mathematics*, **205**, 594–607.

García-Ferrer A, González-Prieto E, Peña D (2012). "A Conditionally Heteroskedastic Independent Factor Model With an Application to Financial Stock Returns." *International Journal of Forecasting*, **28**(1), 70 – 93.

Hyvärinen A (2001). "Blind Source Separation by Nonstationarity of Variance: A Cumulant-based Approach." *IEEE Transactions on Neural Networks*, **12**(6), 1471–1474.

Hyvärinen A, Oja E (1997). "A Fast Fixed-Point Algorithm for Independent Component Analysis." *Neural Computation*, **9**, 1483–1492.

Ilmonen P, Nordhausen K, Oja H, Ollila E (2010). "A New Performance Index for ICA: Properties Computation and Asymptotic Analysis." In V Vigneron, V Zarzoso, E Moreau, R Gribonval, E Vincent (eds.), *"Latent Variable Analysis and Signal Separation", LNCS*, volume 6365, pp. 229–236. Springer, Heidelberg.

Ilmonen P, Oja H, Serfling R (2012). "On Invariant Coordinate System (ICS) Functionals." *International Statistical Review*, **80**, 93–110.

Kastner G (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol." *Journal of Statistical software*, **69**(5), 1–30.

Lu CJ, Wu JY, Lee TS (2009). "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, pp. 468–471.

Matilainen M, Miettinen J, Nordhausen K, Oja H, Taskinen S (2016). *tsBSS: Tools for Blind Source Separation for Time Series*. R package version 0.2, URL https://CRAN.R-project.org/package=tsBSS.

Matilainen M, Nordhausen K, Oja H (2015). "New Independent Component Analysis Tools for Time Series." *Statistics & Probability Letters*, **105**, 80–87.

Matteson D, Ruppert D (2011). "Time-Series Models of Dynamic Volatility and Correlation." *IEEE Signal Processing Magazine*, **28**(5), 72–82.

Miettinen J, Illner K, Nordhausen K, Oja H, Taskinen S, Theis F (2016). "Separation of Uncorrelated Stationary Time Series Using Autocovariance Matrices." *Journal of Time Series Analysis*, **37**(3), 337–354.

Miettinen J, Nordhausen K, Oja H, Taskinen S (2014). *fICA: Classical, Reloaded and Adaptive FastICA Algorithms.* R package version 1.0-2, URL http://CRAN.R-project.org/package=fICA.

Miettinen J, Nordhausen K, Oja H, Taskinen S, Virta J (2017a). "The Squared Symmetric FastICA Estimator." *Signal Processing*, **131**, 402–411.

Miettinen J, Nordhausen K, Taskinen S (2017b). "Blind Source Separation Based on Joint Diagonalization in R: The Packages JADE and BSSasymp." *Journal of Statistical Software*, **76**(2). URL http://dx.doi.org/10.18637/jss.v076.i02.

Miettinen J, Taskinen S, Nordhausen K, Oja H (2015). "Fourth Moments and Independent Component Analysis." *Statistical Science*, **30**, 372–390.

Oja E, Kiviluoto K, Malaroiu S (2000). "Independent Component Analysis for Financial Time Series." In *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pp. 111–116.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. R version 3.2.4, URL http://www.R-project.org/.

Shi Z, Jiang Z, Zhou F (2009). "Blind Source Separation with Nonlinear Autocorrelation and Non-Gaussianity." *Journal of Computational and Applied Mathematics*, **223**(1), 908–915.

Taskinen S, Miettinen J, Nordhausen K (2016). "A More Efficient Second Order Blind Identification Method for Separation of Uncorrelated Stationary Time Series." *Statistics & Probability Letters*, **116**, 21–26.

Taylor SJ (1982). "Financial Returns Modelled by the Product of Two Stochastic Processes – A Study of Daily Sugar Prices 1961–79." In OD Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, pp. 203–216. Springer, North-Holland, Amsterdam.

Wuertz W, Rmetrics Core Team (2013). *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling.* R package version 3010.82, URL http://CRAN.R-project.org/package=fGarch.

Affiliation:

Markus Matilainen
Department of Mathematics and Statistics
FI-20014 University of Turku, Finland
E-mail: markus.matilainen@utu.fi
URL: http://users.utu.fi/manmat