Deflation-based separation of uncorrelated stationary time series

Jari Miettinen[a,*], Klaus Nordhausen[b], Hannu Oja[c], Sara Taskinen[a]

[a]*Department of Mathematics and Statistics, 40014 University of Jyväskylä, Finland*
[b]*School of Information Sciences, 33014 University of Tampere, Finland*
[c]*School of Health Sciences, 33014 University of Tampere, Finland*

Abstract

In this paper we assume that the observed p time series are linear combinations of p latent uncorrelated weakly stationary time series. The problem is then to find an estimate for the unmixing matrix that transforms the observed time series back to uncorrelated time series. The so called SOBI (Second Order Blind Indentification) estimate aims at a joint diagonalization of the covariance matrix and several autocovariance matrices with varying lags. In this paper, we propose a new procedure that extracts the latent time series one-by-one. The limiting distribution of this deflation-based SOBI is found under general conditions, and we show how the results can be used for the comparison of estimates. The exact formula for the limiting covariance matrix of the deflation-based SOBI estimate is given for general multivariate $MA(\infty)$ processes. Also, a whole family of estimates is proposed with the deflation-based SOBI as a special case, and the limiting properties of these estimates are found as well. The theory is widely illustrated by simulation studies.

Keywords:
Asymptotic Normality, Autocovariance, Blind Source Separation, $MA(\infty)$ Processes, Minimum Distance Index, SOBI

[*]Corresponding author
Email address: `jari.p.miettinen@jyu.fi` (Jari Miettinen)

1. Introduction

The blind source separation (BSS) model is a semiparametric model, where the components of the observed p-variate vector $\boldsymbol{x}$ are assumed to be linear combinations of the components of an unobserved p-variate source vector $\boldsymbol{z}$. The BSS model can then simply be written as $\boldsymbol{x} = \Omega\boldsymbol{z}$, where Ω is an unknown full rank $p \times p$ mixing matrix, and the aim is, based on the observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, to find an estimate of the mixing matrix Ω (or its inverse). Notice that, in the independent component analysis (ICA), see for example Hyvärinen et al. (2002), which is perhaps the most popular BSS approach, it is further assumed that the components of $\boldsymbol{z}$ are mutually independent and at most one of them is gaussian.

It is often assumed in BSS applications that the observation vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ are independent and identically distributed (iid) random vectors. In this paper $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ are observations of a time series. We assume that the p-variate observations $\boldsymbol{x}_1, ..., \boldsymbol{x}_T$ obey a BSS model such that

$$\boldsymbol{x}_t = \Omega\boldsymbol{z}_t, \quad t = 0, \pm 1, \pm 2, \ldots \tag{1}$$

where Ω is a full-rank $p \times p$ mixing matrix, and $\boldsymbol{z} = (\boldsymbol{z}_t)_{t=0,\pm 1,\pm 2,\ldots}$ is a p-variate time series that satisfies

(A1) $E(\boldsymbol{z}_t) = 0$,

(A2) $E(\boldsymbol{z}_t\boldsymbol{z}_t') = I_p$, and

(A3) $E(\boldsymbol{z}_t\boldsymbol{z}_{t+\tau}') = E(\boldsymbol{z}_{t+\tau}\boldsymbol{z}_t') = D_\tau$ is diagonal for all $\tau = 1, 2, \ldots$.

The p time series in $\boldsymbol{z}$ are thus weakly stationary and uncorrelated. Given $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$, the aim is to find estimate $\hat{\Gamma}$ of an unmixing matrix Γ such that $\Gamma\boldsymbol{x}$ has uncorrelated components. $\Gamma = \Omega^{-1}$ is naturally one possible unmixing matrix. Notice that Ω and $\boldsymbol{z}$ in the definition are confounded in the sense the signs and order of the components of $\boldsymbol{z}$ (and the signs and order of the columns of Ω, respectively) are not uniquely defined.

In the signal processing community, model (1) is often described as a model with components which are temporally correlated but spatially uncorrelated or as a "colored" data model as opposite to the "white" iid model (Cichocki & Amari, 2002; Comon & Juten, 2010). Contrary to the ICA model, multiple gaussian sources are not excluded in the BSS model assumption. In ICA higher order moments are often used to recover the underlying sources,

2

whereas in model (1) the use of second order statistics is adequate, and the separation is based on the information coming from the serial dependence. Due to this property, this blind source separation approach is also called second order source separation (SOS) approach.

The outline of this paper is as follows. In Section 2.1 we first discuss the so called AMUSE estimator (Tong et al., 1990) which was the first estimator of the unmixing matrix developed for this problem and is based on the simultaneous diagonalization of two autocovariance matrices. The behavior of this estimate however depends heavily on the chosen lags. As a solution to this dilemma the so called SOBI estimator (Belouchrani et al., 1997), which jointly diagonalizes $k > 2$ autocovariance matrices, is considered in Section 2.2. In Section 3 the limiting distribution of a new deflation-based SOBI estimate is derived. The algorithm for the computation of this estimate was suggested in Nordhausen et al. (2012). In Section 4, a large family of unmixing matrix estimates is proposed with the deflation-based SOBI as a special case, and the limiting properties of these estimates are found as well. The limiting behavior and finite-sample behavior of the estimates are illustrated in simulation studies in Section 5. T he paper is concluded with some final remarks.

2. BSS functionals based on autocovariance matrices

2.1. Functionals based on two autocovariance matrices

Let us first recall the statistical functional corresponding to the AMUSE (Algorithm for Multiple Unknown Signals Extraction) estimator introduced by Tong et al. (1990). Assume that $\boldsymbol{x}$ follows a blind source separation (BSS) model such that, for some lag $k > 0$, the diagonal elements of the autocovariance matrix $E(\boldsymbol{z}_t \boldsymbol{z}'_{t+k}) = D_k$ are distinct, and write

$$S_k = E(\boldsymbol{x}_t \boldsymbol{x}'_{t+k}) = \Omega D_k \Omega', \quad k = 0, 1, 2, \ldots$$

for the autocovariance matrices. The unmixing matrix functional Γ_k is then defined as a $p \times p$ matrix that satisfies

$$\Gamma_k S_0 \Gamma'_k = I_p \quad \text{and} \quad \Gamma_k S_k \Gamma'_k = \Lambda_k,$$

where Λ_k is a diagonal matrix with the diagonal elements in a decreasing order. Notice that Γ_k is affine equivariant, that is, if Γ_k and Γ_k^* are the values

3

of the functionals at $\boldsymbol{x}$ and $\boldsymbol{x}^* = A\boldsymbol{x}$ in the BSS model (1) then $\Gamma_k^* = \Gamma_k A^{-1}$ and further $\Gamma_k \boldsymbol{x} = \Gamma_k^* \boldsymbol{x}^*$ (up to sign changes of the components).

The statistical properties of the AMUSE estimator were studied recently in Miettinen et al. (2012). The exact formula for limiting covariance matrix was derived for $MA(\infty)$ processes, and the asymptotic as well as finite sample behavior was investigated. It was shown that the behavior of the AMUSE estimate depends crucially on the choice of the lag k; the p time series can for example be separated consistently only if the eigenvalues in Λ_k are distinct. Without additional information on the time series, it is not possible to decide which lag k should be used in estimation. To circumvent this problem Belouchrani et al. (1997) proposed the SOBI (Second Order Blind Indentification) algorithm that aims to jointly diagonalize several autocovariance matrices $S_0, S_1, \ldots, S_K$. Belouchrani et al. (1997) introduced an algorithm that uses iterative Givens rotations to (approximately) jointly diagonalize the autocovariance matrices. In this paper, we propose the use of an algorithm that finds the latent uncorrelated time series in a deflation-based manner.

2.2. Functionals based on several autocovariance matrices

Let $S_1, \ldots, S_K$ be autocovariance matrices with lags $1, \ldots, K$. The $p \times p$ unmixing matrix functional $\Gamma = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)'$ is the matrix that minimizes

$$\sum_{k=1}^{K} ||\mathrm{off}(\Gamma S_k \Gamma')||^2$$

under the constraint $\Gamma S_0 \Gamma' = I_p$, or, equivalently, maximizes

$$\sum_{k=1}^{K} ||\mathrm{diag}(\Gamma S_k \Gamma')||^2 = \sum_{j=1}^{p} \sum_{k=1}^{K} (\boldsymbol{\gamma}_j' S_k \boldsymbol{\gamma}_j)^2 \qquad (2)$$

under the same constraint. Here $\mathrm{diag}(S)$ is a $p \times p$ diagonal matrix with the diagonal elements as in S and $\mathrm{off}(S) = S - \mathrm{diag}(S)$. In this paper, the columns of an unmixing matrix functional Γ are found one by one so that $\boldsymbol{\gamma}_j$, $j = 1, \ldots, p-1$ maximizes

$$\sum_{k=1}^{K} (\boldsymbol{\gamma}_j' S_k \boldsymbol{\gamma}_j)^2, \qquad (3)$$

4

under $\gamma_r' S_0 \gamma_j = \delta_{rj}$, $r = 1, \ldots, j$. Notice that, for population $S_0, \ldots, S_K$, the solution found using (3) maximizes also (2). This is not usually true for the sample estimates.

Remark 1. *In the literature, there are several algorithms available for a simultaneous diagonalization of K autocovariance matrices. Perhaps the most widely used algorithm based on the Jacobi rotation technique was introduced in Cardoso & Souloumiac (1996).*

Now using (3) one can see that the vector γ_j optimizes the Lagrangian function

$$L(\gamma_j, \lambda_j) = \sum_{k=1}^{K} (\gamma_j' S_k \gamma_j)^2 - \lambda_{jj}(\gamma_j' S_0 \gamma_j - 1) - \sum_{r=1}^{j-1} \lambda_{jr} \gamma_r' S_0 \gamma_j,$$

where $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jj})'$ are the Lagrangian multipliers for γ_j. By solving above optimization problem one can easily show that the unmixing matrix functional Γ satisfies the following $p - 1$ estimating equations:

Definition 1. *The unmixing matrix functional $\Gamma = (\gamma_1, \ldots, \gamma_p)'$ based on S_0 and $S_1, \ldots, S_K$ solves the estimating equations*

$$\boldsymbol{T}(\gamma_j) = S_0 \left(\sum_{r=1}^{j} \gamma_r \gamma_r' \right) \boldsymbol{T}(\gamma_j), \quad j = 1, \ldots, p - 1,$$

where

$$\boldsymbol{T}(\gamma) = \sum_{k=1}^{K} (\gamma' S_k \gamma) S_k \gamma.$$

Also now the resulting Γ is affine equivariant, that is, $\Gamma \boldsymbol{x} = \Gamma^* \boldsymbol{x}^*$ (up to sign changes and permutation of the components), where Γ and Γ^* are the values of the functionals computed at $\boldsymbol{x}$ and $\boldsymbol{x}^* = A\boldsymbol{x}$.

3. Statistical properties of the SOBI estimator

3.1. Limiting multivariate normality

Assume now that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ follow the BSS model (1). We also assume (wlog) that the source vectors are centered at zero. The (population) autocovariance matrices S_k can naturally be estimated by the sample autocovariance

matrices

$$\hat{S}_k = \frac{1}{T-k} \sum_{t=1}^{T-k} \boldsymbol{x}_t \boldsymbol{x}'_{t+k}, \quad k = 0, 1, 2, \ldots$$

As the population autocovariance matrices are symmetric in our model, we estimate S_k with a symmetrized version

$$\hat{S}_k^S = \frac{1}{2}(\hat{S}_k + \hat{S}'_k).$$

Applying Definition 1 to observed data, a natural unmixing matrix estimate is obtained as follows.

Definition 2. *The unmixing matrix estimate $\hat{\Gamma} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)'$ based on $\hat{S}_0$ and $\hat{S}_1^S, \ldots, \hat{S}_K^S$ solves the estimating equations*

$$\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) = \hat{S}_0 (\sum_{r=1}^{j} \hat{\boldsymbol{\gamma}}_r \hat{\boldsymbol{\gamma}}'_r) \hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j), \quad j = 1, \ldots, p-1, \tag{4}$$

where

$$\hat{\boldsymbol{T}}(\boldsymbol{\gamma}) = \sum_{k=1}^{K} (\boldsymbol{\gamma}' \hat{S}_k^S \boldsymbol{\gamma}) \hat{S}_k^S \boldsymbol{\gamma}.$$

The estimating equations also suggest a simple algorithm for computing the rows of the unmixing matrix estimate $\hat{\Gamma}$ one-by-one, see Appendix A.

Solving the rows of $\hat{\Gamma}$ in a deflation-based manner allows us to derive asymptotic properties of the unmixing matrix estimate $\hat{\Gamma}$. Without loss of generality, we derive the limiting distribution under $\Omega = I_p$ only. The limiting distribution for other values of Ω can then be found using the affine equivariance property of $\hat{\Gamma}$.

Theorem 1. *Assume that $(\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$ is an observed time series from the BSS model with $\Omega = I_p$ and*

$$\sqrt{T} \left[\mathrm{vec}(\hat{S}_0, \hat{S}_1^S, \ldots, \hat{S}_K^S) - \mathrm{vec}(I_p, \Lambda_1, \ldots, \Lambda_K) \right] = O_p(1),$$

where Λ_k, $k = 1, \ldots, K$, satisfy $\sum_k \lambda_{ki} \lambda_{kj} \neq \sum_k \lambda_{ki}^2$ for $i < j$. Assume also

that $\hat{\Gamma} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)$ is such solution for (4) that $\hat{\Gamma} \to_p I_p$. Then

$$\sqrt{T}\hat{\gamma}_{ji} = -\sqrt{T}\hat{\gamma}_{ij} - \sqrt{T}(\hat{S}_0)_{ij} + o_p(1), \qquad \text{for } i < j,$$

$$\sqrt{T}\hat{\gamma}_{ji} = \frac{\sum_k \lambda_{kj}(\sqrt{T}\hat{S}_k^S)_{ji} - \sum_k \lambda_{kj}^2(\sqrt{T}\hat{S}_0)_{ji}}{\sum_k \lambda_{kj}(\lambda_{kj} - \lambda_{ki})} + o_p(1), \qquad \text{for } i > j,$$

$$\sqrt{T}\hat{\gamma}_{jj} = -\frac{1}{2}\sqrt{T}((\hat{S}_0)_{jj} - 1) + o_p(1).$$

For the proof, see the proof of more general Theorem 2 in Appendix B. It is interesting that the joint limiting distribution of $\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p$ depends on the order in which they are found. Theorem 1 also implies the following result.

Corollary 1. *Assume that $(\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$ is an observed time series obeying the BSS model. Assume also that the joint limiting distribution of*

$$\sqrt{T}\left[\text{vec}(\hat{S}_0, \hat{S}_1^S, \ldots, \hat{S}_K^S) - \text{vec}(S_0, S_1, \ldots, S_K)\right]$$

is a (singular) $(K+1)p^2$-variate normal distribution with mean value zero. Then, if $\hat{\Gamma} \to_p \Gamma$, the limiting distribution of $\sqrt{T}\text{vec}(\hat{\Gamma} - \Gamma)$ is a singular p^2-variate normal distribution.

3.2. MA(∞) processes

To examine the limiting properties of unmixing matrices under some specific time series model, we now assume that $\boldsymbol{z}_t$ are uncorrelated multivariate MA(∞) processes, that is,

$$\boldsymbol{z}_t = \sum_{j=-\infty}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j}, \tag{5}$$

where Ψ_j, $j = 0, \pm 1, \pm 2, \ldots$, are diagonal matrices satisfying $\sum_{j=-\infty}^{\infty} \Psi_j^2 = I_p$, and $\boldsymbol{\epsilon}_t$ are p-variate iid random vectors with $E(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$ and $Cov(\boldsymbol{\epsilon}_t) = I_p$. Hence

$$\boldsymbol{x}_t = \Omega \boldsymbol{z}_t = \sum_{j=-\infty}^{\infty} (\Omega\Psi_j)\boldsymbol{\epsilon}_{t-j}$$

is also multivariate MA(∞) process. Notice that every second-order stationary process is either a linear process (MA(∞)) or can be transformed to one using the Wold's decomposition. Notice also that causal ARMA(p, q) processes are MA(∞) processes. See Chapter 3 in Brockwell & Davis (1991). We further assume that

7

(A4) the components of $\boldsymbol{\epsilon}_t$ have finite fourth order moments, and

(A5) the components of $\boldsymbol{\epsilon}_t$ are exchangeable and marginally symmetric, that is,
$$JP\boldsymbol{\epsilon}_t \sim \boldsymbol{\epsilon}_t$$
for all sign-change matrices J and for all permutation matrices P. (J is a sign-change matrix if it is a diagonal matrix with diagonal elements ± 1, and P is a permutation matrix if it is obtained from an identity matrix by permuting its rows and/or columns.)

Assumptions (A4) and (A5) imply that $E(\epsilon_{ti}^3 \epsilon_{tj}) = 0$ and justify notations $E(\epsilon_{ti}^4) = \beta_{ii} < \infty$ and $E(\epsilon_{ti}^2 \epsilon_{tj}^2) = \beta_{ij} < \infty$.

The limiting distributions of the sample autocovariance matrices $\hat{S}_k$ depend on the Ψ_t, $t = 0, \pm 1, \pm 2, \ldots$ only through autocovariances of the parameter vector series,

$$F_k = \sum_{t=-\infty}^{\infty} \boldsymbol{\psi}_t \boldsymbol{\psi}_{t+k}', \quad k = 0, 1, 2, \ldots$$

where $\boldsymbol{\psi}_t = (\psi_{t1}, \ldots, \psi_{tp})'$, and $\psi_{t1}, \ldots, \psi_{tp}$ are the diagonal elements of Ψ_t, $t = 0, \pm 1, \pm 2, \ldots$. See Miettinen et al. (2012).

Under the above assumptions the asymptotic normality of $\hat{\Gamma}$ is obtained, and the asymptotic variances are as follows.

Corollary 2. *Assume that $(\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$ is an observed time series from the BSS model $\boldsymbol{x}_t = \Omega \boldsymbol{z}_t$, where $(\boldsymbol{z}_t)$ is a multivariate $MA(\infty)$ process defined in (5) that satisfies also (A4) and (A5). Assume (wlog) that $\Omega = I_p$, $S_0 = I_p$ and $S_k = \Lambda_k$, for $k = 1, 2, \ldots$. Then*

(i) the joint limiting distribution of
$$\sqrt{T}(\text{vec}(\hat{S}_0, \hat{S}_1^S, \ldots, \hat{S}_K^S) - \text{vec}(I_p, \Lambda_1, \ldots, \Lambda_K))$$

is a singular $(K+1)p^2$-variate normal distribution with mean value zero and covariance matrix

$$V = \begin{pmatrix} V_{00} & \cdots & V_{0K} \\ \vdots & \ddots & \vdots \\ V_{K0} & \cdots & V_{KK} \end{pmatrix}. \tag{6}$$

8

with submatrices of the form

$$V_{lm} = \text{diag}(\text{vec}(D_{lm}))(K_{p,p} - D_{p,p} + I_{p^2}).$$

Here $K_{p,p} = \sum_i \sum_j (e_i e_j^T) \otimes (e_j e_i^T)$, $D_{p,p} = \sum_i (e_i e_i^T) \otimes (e_i e_i^T)$ and D_{lm} is a $p \times p$ matrix with elements

$$(D_{lm})_{ii} = (\beta_{ii} - 3)(F_l)_{ii}(F_m)_{ii} + \sum_{k=-\infty}^{\infty} ((F_{k+l})_{ii}(F_{k+m})_{ii} + (F_{k+l})_{ii}(F_{k-m})_{ii}),$$

$$(D_{lm})_{ij} = \frac{1}{2} \sum_{k=-\infty}^{\infty} ((F_{k+l-m})_{ii}(F_k)_{jj} + (F_k)_{ii}(F_{k+l-m})_{jj})$$

$$+ \frac{1}{4}(\beta_{ij} - 1)(F_l + F_l')_{ij}(F_m + F_m')_{ij}, \quad i \neq j.$$

(ii) *The limiting distribution of $\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - e_j)$ is a p-variate normal distribution with mean zero and covariance matrix*

$$ASV(\hat{\boldsymbol{\gamma}}_j) = \sum_{r=1}^{j-1} ASV(\hat{\gamma}_{jr}) e_r e_r' + ASV(\hat{\gamma}_{jj}) e_j e_j' + \sum_{t=j+1}^{p} ASV(\hat{\gamma}_{jt}) e_t e_t',$$

where

$$ASV(\hat{\gamma}_{ji}) = \frac{\sum_{l,m} \lambda_{li} \lambda_{mi}(D_{lm})_{ji} - 2\mu_{ij} \sum_k \lambda_{ki}(D_{k0})_{ji} + \mu_{ij}^2 (D_{00})_{ji}}{(\mu_{ij} - \mu_{ii})^2}, \quad \text{for } i < j$$

$$ASV(\hat{\gamma}_{ji}) = \frac{\sum_{l,m} \lambda_{lj} \lambda_{mj}(D_{lm})_{ji} - 2\mu_{jj} \sum_k \lambda_{kj}(D_{k0})_{ji} + \mu_{jj}^2 (D_{00})_{ji}}{(\mu_{jj} - \mu_{ji})^2}, \quad \text{for } i > j$$

$$ASV(\hat{\gamma}_{jj}) = 4^{-1}(D_{00})_{jj}$$

with $\mu_{ij} = \sum_k \lambda_{ki} \lambda_{kj}$.

Remark 2. *Notice that if we assume (A4) but replace (A5) by*

(A6) The components of ϵ_t are mutually independent,

then, in this independent component model case, the joint limiting distribution of $(\hat{S}_0, \hat{S}_1^S, \ldots, \hat{S}_K^S)$ is again as given in Corollary 2 but with $\beta_{ij} = 1$ for $i \neq j$. If we further assume that innovations ϵ_t are iid from $N_p(\mathbf{0}, I_p)$, then $\beta_{ii} = 3$ and $\beta_{ij} = 1$ for all $i \neq j$, and the variances and covariances in Corollary 2 become even more simplified.

9

4. A new family of deflation-based estimates

The criterion function for the deflation-based SOBI estimator considered so far is based on fourth moments. We next consider a modification of the criterion in (3) that is obtained if the second power in $\sum_{k=1}^{K}(\boldsymbol{\gamma}_j' S_k \boldsymbol{\gamma}_j)^2$ is replaced by $\sum_{k=1}^{K} G(\boldsymbol{\gamma}_j' S_k \boldsymbol{\gamma}_j)$ with other choices of G. Different choices of G then yield estimates with different robustness and efficiency properties. Notice however that the estimate still depends on the observations through the second moments and to find an estimate with a bounded influence function, for example, does not seem to be possible for any choice of G. To obtain an estimate with a high break-down point and a bounded influence function, either (i) the covariance and the autocovariance matrices should be replaced by a robust scatter matrix and by robust autocovariance matrices, respectively, or (ii) the covariance matrix should be replaced by a robust scatter matrix and G should be bounded. However, further research is needed here.

The columns of Γ are again found one by one so that $\boldsymbol{\gamma}_j$, $j = 1, \ldots, p-1$ maximizes

$$\sum_{k=1}^{K} G(\boldsymbol{\gamma}_j' S_k \boldsymbol{\gamma}_j), \tag{7}$$

under $\boldsymbol{\gamma}_r' S_0 \boldsymbol{\gamma}_j = \delta_{rj}$, $r = 1, \ldots, j-1$. In Section 2.2 we used $G(x) = x^2$, but now we let G be any increasing twice continuously differentiable function.

As before, the functional $\Gamma = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)'$ based on functionals S_0 and $S_1, \ldots, S_K$ can be shown to solve the estimating equations

$$\boldsymbol{T}(\boldsymbol{\gamma}_j) = S_0 \left(\sum_{r=1}^{j} \boldsymbol{\gamma}_r \boldsymbol{\gamma}_r' \right) \boldsymbol{T}(\boldsymbol{\gamma}_j), \quad j = 1, \ldots, p-1, \tag{8}$$

where

$$\boldsymbol{T}(\boldsymbol{\gamma}) = \sum_{k=1}^{K} g(\boldsymbol{\gamma}' S_k \boldsymbol{\gamma}) S_k \boldsymbol{\gamma},$$

and $g = G'$. Corresponding unmixing matrix estimate $\hat{\Gamma} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)'$ is obtained by replacing autocovariance matrices with symmetrized sample autocovariance matrices in estimating equations. Further, the limiting distribution of $\hat{\Gamma}$ can be derived using the following result:

Theorem 2. *Assume that $(\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$ is an observed time series obeying the BSS model with $\Omega = I_p$. Assume also that*

$$\sqrt{T}\left[\text{vec}(\hat{S}_0, \hat{S}_1^S, \ldots, \hat{S}_K^S) - \text{vec}(I_p, \Lambda_1, \ldots, \Lambda_K)\right] = O_p(1),$$

where Λ_k, $k = 1, \ldots, K$, satisfy $\sum_k g(\lambda_{ki})\lambda_{kj} \neq \sum_k g(\lambda_{ki})\lambda_{ki}$ for $i < j$, and $\hat{\Gamma} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)$ is such solution for (8) that $\hat{\Gamma} \to_p I_p$. Then

$$\sqrt{T}\hat{\gamma}_{ji} = -\sqrt{T}\hat{\gamma}_{ij} - \sqrt{T}(\hat{S}_0)_{ij} + o_p(1), \qquad\qquad \text{for } i < j,$$

$$\sqrt{T}\hat{\gamma}_{ji} = \frac{\sum_k g(\lambda_{kj})\sqrt{T}(\hat{S}_k^S)_{ji} - \sum_k g(\lambda_{kj})\lambda_{kj}\sqrt{T}(\hat{S}_0)_{ji}}{\sum_k g(\lambda_{kj})(\lambda_{kj} - \lambda_{ki})} + o_p(1), \qquad \text{for } i > j,$$

$$\sqrt{T}\hat{\gamma}_{jj} = -\frac{1}{2}\sqrt{T}((\hat{S}_0)_{jj} - 1) + o_p(1).$$

The results in Corollary 1 and Corollary 2 also hold for any twice continuously differentiable function G. For MA(∞) processes the asymptotic variances are then given by

$$ASV(\hat{\gamma}_{ji}) = \frac{\sum_{l,m} g(\lambda_{li})g(\lambda_{mi})(D_{lm})_{ji} - 2\mu_{ij}\sum_k g(\lambda_{ki})(D_{k0})_{ji} + \mu_{is}^2(D_{00})_{ji}}{(\mu_{ij} - \mu_{ii})^2},$$

for $i < j$,

$$ASV(\hat{\gamma}_{ji}) = \frac{\sum_{l,m} g(\lambda_{lj})g(\lambda_{mj})(D_{lm})_{ji} - 2\mu_{jj}\sum_k g(\lambda_{kj})(D_{k0})_{ji} + \mu_{jj}^2(D_{00})_{ji}}{(\mu_{jj} - \mu_{ji})^2},$$

for $i > j$,

$$ASV(\hat{\gamma}_{jj}) = 4^{-1}(D_{00})_{jj},$$

where $\mu_{ij} = \sum_k g(\lambda_{ki})\lambda_{kj}$, and matrices D_{lm} are as defined in Corollary 2.

5. Simulation studies

5.1. Minimum distance index

There are several possible performance indices to measure the accuracy of an unmixing matrix (for an overview see for example Nordhausen et al., 2011a). Most of the indices are functions of the gain matrix $\hat{G} = \hat{\Gamma}\Omega$. For a good estimate $\hat{\Gamma}$, $\hat{G}$ is then close to some C in

$$\mathcal{C} = \{C \ : \ \text{each row and column of } C \text{ has exactly one non-zero element.}\}$$

In this paper, the estimates are compared using the minimum distance index (Ilmonen et al., 2010)

$$\hat{D} = D(\hat{\Gamma}\Omega) = \frac{1}{\sqrt{p-1}} \inf_{C \in \mathcal{C}} \|C\hat{\Gamma}\Omega - I_p\| \tag{9}$$

with the matrix (Frobenius) norm $\|\cdot\|$. It is shown (Ilmonen et al., 2010) that the minimum distance index is affine invariant and $0 \le \hat{D} \le 1$. Moreover, if $\Omega = I_p$ and $\sqrt{T}\,\mathrm{vec}(\hat{\Gamma} - I_p) \to N_{p^2}(\mathbf{0}, \Sigma)$, then the limiting distribution of $T(p-1)\hat{D}^2$ is that of weighted sum of independent chi squared variables with the expected value

$$\mathrm{tr}\left((I_{p^2} - D_{p,p})\Sigma(I_{p^2} - D_{p,p})\right). \tag{10}$$

Notice that $\mathrm{tr}((I_{p^2} - D_{p,p})\Sigma(I_{p^2} - D_{p,p}))$ is the sum of the limiting variances of the off-diagonal elements of $\sqrt{T}\,\mathrm{vec}(\hat{\Gamma} - I_p)$ and therefore provides a global measure of the variation of the estimate $\hat{\Gamma}$.

5.2. Simulation setups

In our simulations, the three-variate time series $\boldsymbol{z}_1, ..., \boldsymbol{z}_T$ were generated from the following five models with different choices of T and with 10 000 repetitions. In all simulation settings, the mixing matrix was $\Omega = I_p$. Notice that, due to affine equivariance of the estimates, the performance comparisons do not depend on this choice.

(A) AR(3), AR(5) and AR(6) processes with coefficient vectors $(0.4, -0.5, 0.2)$, $(-0.1, 0, 0, 0, -0.1)$ and $(0, 0.2, 0, -0.4, 0, 0.3)$, respectively, and spherical three-variate t_5-distributed innovations,

(B) three independent MA(50) processes with normal innovations,

(C) three independent AR(3) processes with coefficient vectors $(0.5, 0.2, -0.4)$, $(-0.3, 0.2, -0.1)$ and $(0.1, 0.1, 0.3)$, respectively, and normal innovations,

(D) AR(2), AR(5) and MA(6) processes with coefficient vectors $(0.1, 0.5)$, $(0.2, 0.5, 0.3, 0.1, -0.5)$ and $(0.4, 0.2, 0.1, -0.6, -0.2, -0.4)$, respectively, and normal innovations,

(E) AR(1), AR(4) and MA(5) processes with coefficient vectors (0.5), $(0.1, -0.3, -0.1, 0.1)$ and $(0.1, -0.3, -0.1, 0.4, 0.1)$, respectively, and normal innovations.

The averages of $T(p-1)\hat{D}^2$ over 10 000 repetitions were used to compare the finite-sample efficiencies of different estimates. We also calculated the expected values of the limiting distributions of $T(p-1)E(\hat{D}^2)$ in each case to compare the asymptotic efficiencies.

5.3. Simulation results

Main results from our simulations are as follows.

(i) The SOBI estimate based on several autocovariance matrices seems to outperform the AMUSE estimates that are based on two autocovariance matrices only: Figure 1 illustrates the finite sample behavior of the AMUSE estimates with lags 1 or 4 as compared to the deflation-based SOBI estimate with ten autocovariance matrices in model (A). It is seen that the deflation-based SOBI estimate is much more efficient than the AMUSE estimates. Notice also that, in practice, the best lag for AMUSE is not known.

(ii) The number of autocovariance matrices used in SOBI affects the performance: In Figure 2, estimates using different numbers of autocovariance matrices are compared in model (B). Since the source components are MA(50) processes, the autocovariances are zero for lags larger than 50. Hence, by Corollary 2, the asymptotic variances of the SOBI estimates are equal when $K \geq 50$. For small sample sizes, the use of too many matrices seems to have a deteriorating effect.

(iii) The order in which the components are found affects the performance: First notice that the order is mainly determined by the initial values given to the algorithm. In our simulations studies, the order is controlled by selecting different permutations as initial values in the algorithm. Figure 3 illustrates the influence of this choice when estimating the unmixing matrix using the deflation-based SOBI with ten autocovariance matrices in case of model (C). The two initial values are chosen to yield estimates with the lowest and highest values of the averages of $T(p-1)\hat{D}^2$. In Table 1, the expected values of the limiting distributions of $T(p-1)\hat{D}^2$ for all six different extraction orders are
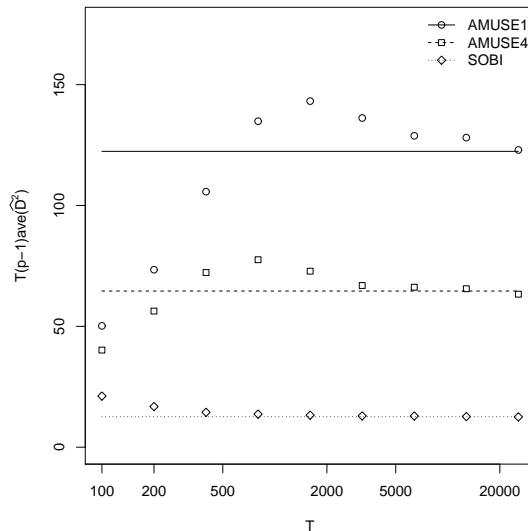
13

Figure 1: The averages of $T(p-1)\hat{D}^2$ for the SOBI estimate ($K = 10$) and the AMUSE estimates with lags 1 and 4 over 10 000 repetitions of the observed time series with length T from model (A). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

listed using models (A)-(C). Finally notice that finding an easy rule for the choice of the initial value to obtain the optimal order (as for example in the case of deflation-based fastICA was done in Nordhausen et al. (2011b)) seems difficult. Luckily, our simulations have shown that in practice the algorithm seems to find the components mainly in a "good" order, even if the initial values are random, and the averages of $T(p-1)\hat{D}^2$ are relatively close to the expected value corresponding to the optimal order.

Table 1: The expected values of the limiting distributions of $T(p-1)\hat{D}^2$ for SOBI estimates with $K = 10$. All six different extraction orders were considered.

	order					
model	123	132	213	231	312	321
(A)	47.7	12.6	49.2	50.8	14.2	15.7
(B)	13.7	14.1	11.4	9.8	12.5	10.2
(C)	7.9	8.6	9.5	21.9	21.0	22.7

(iv) The choice of G affects the efficiency: In Figure 4 and Table 2 we
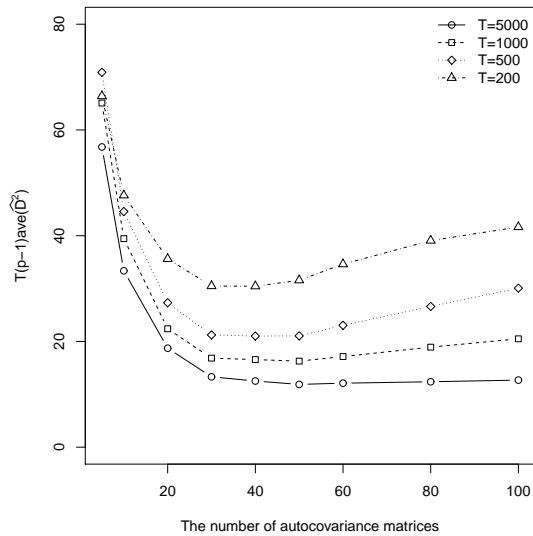
14

Figure 2: The averages of $T(p-1)\hat{D}^2$ for the SOBI estimates using 5, 10, 20, 30, 40, 50, 60, 80 and 100 autocovariance matrices over 10 000 repetitions of observed time series with length T from model (B).

compare the estimates corresponding to different choices of G (and its derivative g). We use g-functions $g_1(x) = 2x$, $g_2(x) = 4x^3$ and $g_3(x) = \arctan(50x)$, respectively. In models (A)-(C) the estimate based on g_1 yields the best results of the three alternatives, but in models (D) and (E) functions g_3 and g_2, respectively, are better.

Table 2: The expected values of the limiting distributions of $T(p-1)\hat{D}^2$ for SOBI estimates with $K = 10$ and selected g-functions. The components are found in the optimal order.

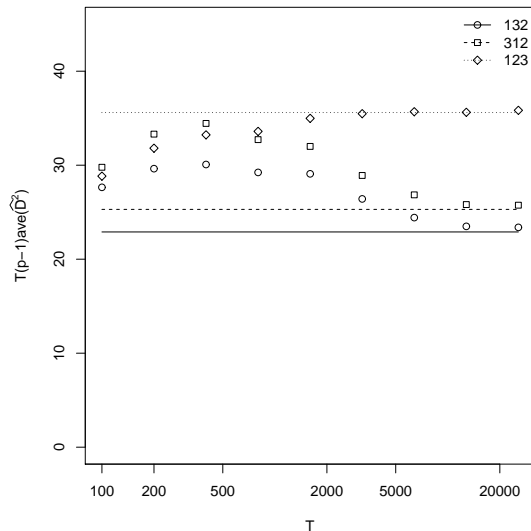| model	$g(x) = 2x$	$g(x) = 4x^3$	$g(x) = \arctan(50x)$
(A)	12.6	14.3	14.8
(B)	9.8	14.1	11.3
(C)	7.9	8.5	12.1
(D)	23.2	40.3	19.1
(E)	96.7	49.9	161.0

Figure 3: The averages of $T(p-1)\hat{D}^2$ for two SOBI estimates ($K = 10$) with different initial values over 10 000 repetitions of observed time series with length T from model (C). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

6. Concluding remarks

In this paper we presented a thorough analysis of the deflation-based SOBI estimate by rigorously developing its asymptotic distribution. For large sample sizes, the asymptotic results then provided a nice approximation of the average criterion values for the estimates. It is not difficult to argue that SOBI is a safer choice in practical applications than AMUSE. Nevertheless the choice of the number of lags and which lags to choose is still an open problem for SOBI. The results in this paper show that for large sample sizes it may be wise to use as many autocovariance matrices as possible but that for small sample sizes one must be more careful. There are no clear guidelines in the literature. Belouchrani et al. (1997) for example considered estimates based on up to 10 lags, whereas Tang et al. (2005) investigate much larger lag sets in the analysis of EEG data. Further research is needed here.

In this paper the autocovariance matrices were jointly diagonalized using a deflation-based algorithm. Other algorithms have been used in practice, however, and our future research will investigate the statistical properties of these estimates which naturally differ from the properties of deflation-based SOBI estimates. It is worth mentioning that Ziehe & Müller (1998) suggested a BSS method which simultaneously diagonalizes the covariance matrix and
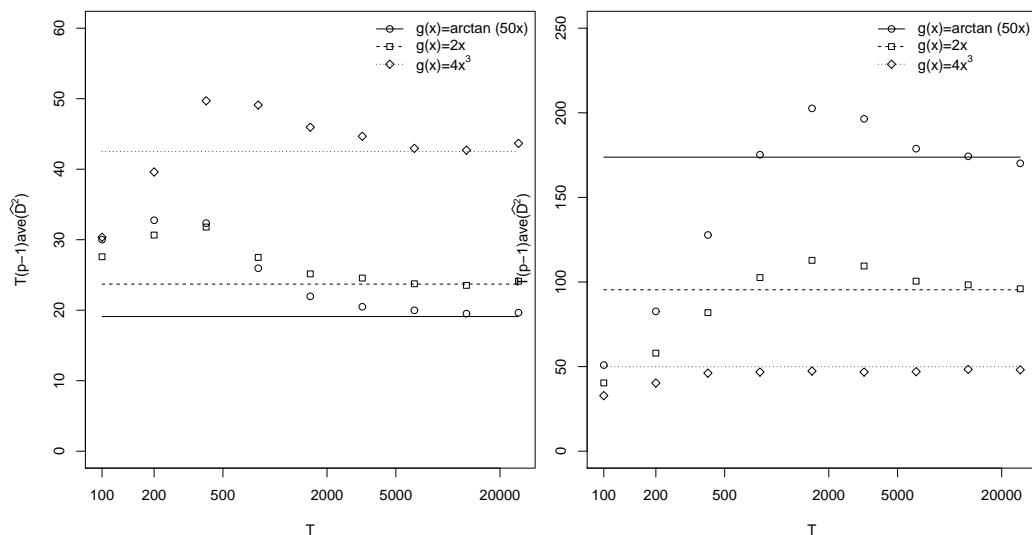
16

Figure 4: The averages of $T(p-1)\hat{D}^2$ for three SOBI estimates (K=10) with different choices of function G from 10 000 repetitions of observed time series with length T from models (D) (the right panel) and (E) (the left panel). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

a weighted sum of several autocovariance matrices which can then be seen as some kind of compromise between AMUSE and SOBI. They furthermore also suggested a joint diagonalization of several autocovariance matrices together with higher-order moment matrices in the case of possibly identical marginal autocovariances.

In our approach the covariance matrix plays a special role while all other autocovariance matrices are treated symmetrically. Some procedures like WASOBI (Yeredor, 2000) try to optimize the performance of SOBI by weighting the autocovariance matrices and also by removing the special role of the covariance matrix.

References

BELOUCHRANI, A., ABED-MERAIM, K., CARDOSO, J.-F. & MOULINES, E. (1997). A blind source separation technique using second-order statistics. IEEE Transactions on *Signal Processing*, 45, 434–444.

BROCKWELL, P.J. & DAVIS, R.A. (1991). *Time Series: Theory and Methods*. Second edition, Springer-Verlag, New York.

CARDOSO, J.F. AND SOULOUMIAC, A. (1996). Jacobi angles for simultaneous diagonalization. SIAM J. Mat. Anal. Appl., 17, 161–164.

CICHOCKI, A. & AMARI, S. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* Wiley , Chichester.

COMON P. & JUTTEN C. (2010). *Handbook of Blind Source Separation. Independent Component Analysis and Applications.* Academic Press, Amsterdam

HYVÄRINEN, A. KARHUNEN, J. & OJA, E. (2001). *Independent Component Analysis.* John Wiley & Sons, New York.

ILMONEN, P., NORDHAUSEN, K., OJA, H. & OLLILA, E. (2010). A new performance index for ICA: properties computation and asymptotic analysis. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R. and Vincent, E. (editors) *Latent Variable Analysis and Signal Separation.* 229–236, Springer, Heidelberg.

MIETTINEN, J., NORDHAUSEN, K., OJA, H. AND TASKINEN, S. (2012). Statistical properties of a blind source separation estimator for stationary time series. *Statistics and Probability Letters* 82, 1865–1873.

NORDHAUSEN K., OLLILA E., OJA H. (2011). On the performance indices of ICA and blind source separation. In the Proceedings of *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2011)*, 486–490.

NORDHAUSEN, K., ILMONEN, P., MANDAL, A., OJA, H. AND OLLILA, E. (2011). Deflation-based FastICA Reloaded. In the Proceedings of *19th European Signal Processing Conference 2011 (EUSIPCO 2011)*, 1854–1858.

NORDHAUSEN, K., GUTCH, H. W., OJA, H. AND THEIS, F.J. (2012). Joint Diagonalization of Several Scatter Matrices for ICA. In Theis, F., Cichocki, A., Yeredor, A. and Zibulevsky, M. (editors) *Latent Variable Analysis and Signal Separation, LNCS 7191.* 172–179, Springer, Heidelberg.

TANG, A.C., LIU, J.-Y. & SUTHERLAND M.T. (2005). Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI. *NeuroImage*, 7, 507–519.

18

TONG, L., SOON, V.C., HUANG, Y.F. & LIU, R. (1990). AMUSE: a new blind identification algorithm. In Proceedings of IEEE International Symposium on *Circuits and Systems 1990*, pp.1784–1787.

YEREDOR, A. (2000). Blind Separation of Gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, 7, 197–200.

ZIEHE, A. & MÜLLER, K.-R. (1998). TDSEP - an efficient algorithm for blind separation using time structure. In Proceedings of ICANN, pp.675–680.

Appendix A: Joint diagonalization algorithm

The algorithm for computing unmixing matrix estimate $\hat{\Gamma}$ can be easily derived using the estimating equation (4) (Nordhausen et al., 2012). We first use $\hat{S}_0$ to whiten the data and then find the rows of an orthogonal matrix $\hat{W} = (\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_p)'$ one-by-one so that K sample autocovariance matrices of the whitened data become jointly diagonalized.

To be more precise, we write the estimating equation for the whitened data as

$$\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j) = \Big(\sum_{r=1}^{j-1} \hat{\boldsymbol{w}}_r \hat{\boldsymbol{w}}_r' + \hat{\boldsymbol{w}}_j \hat{\boldsymbol{w}}_j'\Big)\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j). \tag{11}$$

Then

$$(I_p - \sum_{r=1}^{j-1} \hat{\boldsymbol{w}}_r \hat{\boldsymbol{w}}_r')\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j) = (\hat{\boldsymbol{w}}_j'\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j))\hat{\boldsymbol{w}}_j,$$

that is, $\hat{\boldsymbol{w}}_j$ that solves the estimating equation (11) must satisfy

$$\hat{\boldsymbol{w}}_j \propto (I_p - \sum_{r=1}^{j-1} \hat{\boldsymbol{w}}_r \hat{\boldsymbol{w}}_r')\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j).$$

This yields to following simple iteration steps for the jth row of $\hat{W}$:

step 1. $\hat{\boldsymbol{w}}_j \leftarrow \hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j)$

step 2. $\hat{\boldsymbol{w}}_j \leftarrow (I_p - \sum_{r=1}^{j-1} \hat{\boldsymbol{w}}_r \hat{\boldsymbol{w}}_r')\hat{\boldsymbol{w}}_j$ (orthogonalization)

step 3. $\hat{\boldsymbol{w}}_j \leftarrow \hat{\boldsymbol{w}}_j/||\hat{\boldsymbol{w}}_j||$ (standardization),

and the final unmixing matrix estimate is then $\hat{\Gamma} = \hat{W}\hat{S}_0^{-1/2}$.

Notice that an initial value for $\hat{W}$ is needed. The choice of the initial value determines the order of the rows of $\hat{W}$, and hence it affects the performance of the unmixing matrix estimate $\hat{\Gamma}$.

Since $2\hat{T}(\hat{\boldsymbol{w}}_j)$ is the gradient of $\sum_{k=1}^{K}(\hat{\boldsymbol{w}}_j'\hat{S}_k\hat{\boldsymbol{w}}_j)^2$, the described algorithm is a fixed-point algorithm. It is based on the same idea as the gradient algorithm, which is derived by replacing step 1 with

$$\text{step } 1^*. \quad \hat{\boldsymbol{w}}_j \leftarrow \hat{\boldsymbol{w}}_j + 2\hat{\boldsymbol{T}}(\hat{\boldsymbol{w}}_j).$$

The fixed-point algorithm takes longer steps and is usually faster. Unfortunately, it does not always convergence. Therefore we recommend the use of a mixture of the two algorithms, where step 1 is applied, for example, four times and step 1* is then applied once.

Appendix B: Proofs of the results

Proof of Corollary 2 Notice first that

$$\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j) = \sum_{r=1}^{j-1}\hat{\gamma}_{jr}\boldsymbol{e}_r + \hat{\gamma}_{jj}\boldsymbol{e}_j + \sum_{t=j+1}^{p}\hat{\gamma}_{jt}\boldsymbol{e}_t.$$

Then

$$\begin{aligned}
ASV(\hat{\boldsymbol{\gamma}}_j) &= T\,E[(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j)(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j)'] \\
&= \sum_{r=1}^{j-1}E[(\sqrt{T}\hat{\gamma}_{jr})^2]\boldsymbol{e}_r\boldsymbol{e}_r' + E[(\sqrt{T}(\hat{\gamma}_{jj} - 1)^2]\boldsymbol{e}_j\boldsymbol{e}_j' + \sum_{t=j+1}^{p}E[(\sqrt{T}\hat{\gamma}_{jt})^2]\boldsymbol{e}_t\boldsymbol{e}_t',
\end{aligned}$$

where the asymptotic variances can be derived using Theorem 1 and Corollary 1.

Proof of Theorem 2. The limiting distributions of the symmetrized sample autocovariance matrices are given in Miettinen et al. (2012). Notice then that as $\hat{\boldsymbol{\gamma}}_j \rightarrow_p \boldsymbol{e}_j$, then $\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) \rightarrow_p \mu_{jj}\boldsymbol{e}_j$, where $\mu_{jj} = \sum_k g(\lambda_{kj})\lambda_{kj}$. The estimating equation (4) then yields to

$$\sqrt{T}(\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j) = \sqrt{T}(\hat{S}_0\,\hat{U}_j\,\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j), \quad (12)$$

where $\hat{U}_j = \sum_{r=1}^{j} \hat{\boldsymbol{\gamma}}_r \hat{\boldsymbol{\gamma}}_r'$. Now as $\hat{U}_j \to_p U_j = \sum_{r=1}^{j} \boldsymbol{e}_r \boldsymbol{e}_r'$, the right-hand side of the equation can be written as

$$\sqrt{T}(\hat{S}_0 \hat{U}_j \hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j) = \sqrt{T}(\hat{S}_0 - I_p)\hat{U}_j\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) + \sqrt{T}(\hat{U}_j - U_j)\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j)$$
$$+ \sqrt{T}(U_j(\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j)).$$

Thus (12) yields to

$$(I_p - U_j)\sqrt{T}(\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j) = \sqrt{T}(\hat{S}_0 - I_p)\hat{U}_j\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) + \sqrt{T}(\hat{U}_j - U_j)\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j)$$
$$= \sqrt{T}(\hat{S}_0 - I_p)\mu_{jj}\boldsymbol{e}_j + \sqrt{T}(\hat{U}_j - U_j)\mu_{jj}\boldsymbol{e}_j + o_p(1).$$
$$(13)$$

Now

$$\sqrt{T}(\hat{U}_j - U_j)\mu_{jj}\boldsymbol{e}_j = \sum_{r=1}^{j} \sqrt{T}(\hat{\boldsymbol{\gamma}}_r\hat{\boldsymbol{\gamma}}_r' - \boldsymbol{e}_r\boldsymbol{e}_r')\mu_{jj}\boldsymbol{e}_j$$

$$= \sum_{r=1}^{j} \sqrt{T}((\hat{\boldsymbol{\gamma}}_r - \boldsymbol{e}_r)\hat{\boldsymbol{\gamma}}_r' + \boldsymbol{e}_r(\hat{\boldsymbol{\gamma}}_r - \boldsymbol{e}_r)')\mu_{jj}\boldsymbol{e}_j \qquad (14)$$

$$= \mu_{jj}(\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j) + \sum_{r=1}^{j} \sqrt{T}(\hat{\boldsymbol{\gamma}}_r - \boldsymbol{e}_r)'\boldsymbol{e}_j\boldsymbol{e}_r) + o_p(1).$$

And by Taylor???s expansion, we get

$$\sqrt{T}(\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j) = \sum_{k}(g(\lambda_{kj}) - g'(\lambda_{kj})\lambda_{kj})\sqrt{T}\hat{S}_k^S\hat{\boldsymbol{\gamma}}_j$$

$$+ \sum_{k} g'(\lambda_{kj})\lambda_{kj}\Lambda_k\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j) + \sum_{k} g'(\lambda_{kj})\lambda_{kj}\sqrt{T}(\hat{S}_k^S - \Lambda_k)\boldsymbol{e}_j$$

$$+ 2\sum_{k} g'(\lambda_{kj})\lambda_{kj}^2\boldsymbol{e}_j'\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j)\boldsymbol{e}_j + \sum_{k} g'(\lambda_{kj})\lambda_{kj}(\boldsymbol{e}_j'\sqrt{T}(\hat{S}_k^S - \Lambda_k)\boldsymbol{e}_j)\boldsymbol{e}_j,$$

where

$$\sqrt{T}\hat{S}_k^S\hat{\boldsymbol{\gamma}}_j = \sqrt{T}(\hat{S}_k^S - \Lambda_k)\hat{\boldsymbol{\gamma}}_j + \sqrt{T}\Lambda_k(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j) + \sqrt{T}\lambda_{kj}\boldsymbol{e}_j$$

After some tedious computations this reduces to

$$\sqrt{T}(\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \mu_{jj}\boldsymbol{e}_j) = A_j\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j) + \sum_{k} B_{kj}\sqrt{T}(\hat{S}_k^S - \Lambda_k)\boldsymbol{e}_j$$
$$- \sum_{k} \sqrt{T}g'(\lambda_{kj})\lambda_{kj}^2\boldsymbol{e}_j, \qquad (15)$$

21

where

$$A_j = \sum_k (g(\lambda_{kj})\Lambda_k + 2g'(\lambda_{kj})\lambda_{kj}^2 \boldsymbol{e}_j \boldsymbol{e}_j')$$

$$B_{kj} = g(\lambda_{kj})I_p + g'(\lambda_{kj})\lambda_{kj}\boldsymbol{e}_j\boldsymbol{e}_j'$$

The result is then obtained by plugging (14) and (15) into (13).