Statistica Sinica Preprint No: SS-2018-0347	
Title	Determining the signal dimension in second order source separation
Manuscript ID	SS-2018-0347
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0347
Complete List of Authors	Joni Virta and Klaus Nordhausen
Corresponding Author	Joni Virta
E-mail	joni.virta@outlook.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Statistica Sinica

DETERMINING THE SIGNAL DIMENSION IN SECOND ORDER SOURCE SEPARATION

Joni Virta[1] and Klaus Nordhausen[2]

[1] *Aalto University School of Science, Finland*

[2] *Vienna University of Technology, Austria*

Abstract: While an important topic in practice, the estimation of the number of non-noise components in blind source separation has received little attention in the literature. Recently, two bootstrap-based techniques for estimating the dimension were proposed, and although very efficient, they suffer from the long computation times caused by the resampling. We approach the problem from a large sample viewpoint and develop an asymptotic test and a corresponding consistent estimate for the true dimension. Our test statistic based on second-order temporal information has a very simple limiting distribution under the null hypothesis and requires no parameters to estimate. Comparisons to the resampling-based estimates show that the asymptotic test provides comparable error rates with significantly faster computation time. An application to sound recording data is used to illustrate the method in practice.

Key words and phrases: Blind source separation, chi-square distribution, second order blind identification, second order stationarity, white noise.

Statistica Sinica

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

1. Introduction

The modelling of multivariate time series is notoriously difficult and an increasingly common option is to use latent variable or factor models (see for example Ensor, 2013; Chang et al., 2018, and the references therein). In this paper we will follow the *blind source separation* (BSS) approach, as an intermediary step prior to modelling. In BSS the observed multivariate time series is bijectively decomposed into several univariate time series that exhibit some form of mutual independence, such as second order uncorrelatedness or even full statistical independence. After such a decomposition, the lack of interaction between the univariate series allows us to model them separately, requiring much smaller numbers of parameters.

A particularly popular choice among BSS models for time series is the second order source separation (SOS) model (Comon and Jutten, 2010) which assumes that the observed zero-mean p-variate time series $\mathbf{x}_1, \ldots, \mathbf{x}_T$ is generated as,

$$\mathbf{x}_t = \mathbf{\Omega} \mathbf{z}_t, \quad t = 1, \ldots, T, \tag{1.1}$$

where the *source* series $\mathbf{z}_1, \ldots, \mathbf{z}_T$ is a latent non-degenerate, zero-mean, second-order stationary p-variate time series with uncorrelated component series and $\mathbf{\Omega} \in \mathbb{R}^{p \times p}$ is an unknown, invertible matrix-valued parameter.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

The assumption of zero mean is without loss of generality as we can always center our series. The objective in model (1.1) is to estimate an inverse $\hat{\mathbf{\Gamma}}$ for $\mathbf{\Omega}$, giving us an estimate $\hat{\mathbf{z}}_t = \hat{\mathbf{\Gamma}}\hat{\mathbf{x}}_t$ for the p sources, which can then further be modelled univariately.

However, noise is often an inevitable part of any real world signal and we incorporate it in the model (1.1) through the sources. That is, we assume that the sources consist of two parts, $\mathbf{z}_t = (\mathbf{z}_{1t}^\top, \mathbf{z}_{2t}^\top)^\top$, where $\mathbf{z}_{1t} \in \mathbb{R}^q$ contains the signals and $\mathbf{z}_{2t} \in \mathbb{R}^{p-q}$ is a white noise vector. To avoid overfitting in the modelling phase, a crucial step in BSS is to identify the noise subvector $\mathbf{z}_{2t}$ among the sources and discard it prior to the modelling. This problem, signal dimension estimation, has only recently been considered in the context of statistical blind source separation and in this paper we propose a novel estimate that relies on asymptotic hypothesis testing. But first, we review two classical SOS-methods that serve as the basis for both our method and the two existing ones.

The standard way of estimating the sources in (1.1) is via second-order temporal moments. In algorithm for multiple signals extraction (AMUSE) (Tong et al., 1990), an estimate $\hat{\mathbf{\Gamma}}$ is obtained from the generalized eigendecomposition,

$$\hat{\mathbf{\Gamma}}\hat{\mathbf{S}}_0\hat{\mathbf{\Gamma}}^\top = \mathbf{I}_p \quad \text{and} \quad \hat{\mathbf{\Gamma}}\hat{\mathbf{R}}_\tau\hat{\mathbf{\Gamma}}^\top = \hat{\mathbf{D}}_\tau,$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

where $\hat{\mathbf{S}}_0 = (1/T)\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^\top$ is the marginal covariance matrix, $\hat{\mathbf{R}}_\tau = (\hat{\mathbf{S}}_\tau + \hat{\mathbf{S}}_\tau^\top)/2$, $\hat{\mathbf{S}}_\tau = [1/(T-\tau)]\sum_{t=1}^{T-\tau}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t+\tau} - \bar{\mathbf{x}})^\top$ is the τ-lag autocovariance matrix and $\hat{\mathbf{D}}_\tau$ is a diagonal matrix. If the lag-τ autocovariances of the latent series are distinct, then $\hat{\mathbf{\Gamma}}$ is consistent for $\mathbf{\Omega}^{-1}$ up to permutation and signs of its rows. For the statistical properties of AMUSE, see Miettinen et al. (2012).

A usually more efficient estimate, which does not depend so much on the selection of a single parameter τ, is given by second order blind identification (SOBI) (Belouchrani et al., 1997), an extension of AMUSE to multiple autocovariance matrices. In SOBI we choose a set of lags $\mathcal{T}$ and estimate the orthogonal matrix $\hat{\mathbf{U}}$ by maximizing,

$$\sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{U}^\top \hat{\mathbf{S}}_0^{-1/2} \hat{\mathbf{R}}_\tau \hat{\mathbf{S}}_0^{-1/2} \mathbf{U} \right) \right\|^2, \tag{1.2}$$

over the set of all orthogonal $\mathbf{U} \in \mathbb{R}^{p \times p}$, where $\mathrm{diag}(\mathbf{A})$ denotes a diagonal matrix with diagonal elements equal to those of $\mathbf{A}$ and $\hat{\mathbf{S}}_0^{-1/2}$ is the unique symmetric inverse square root of the almost surely positive definite matrix $\hat{\mathbf{S}}_0$. This procedure is called orthogonal (approximate) joint diagonalization and provides a natural extension of the generalized eigendecomposition to more than two matrices. Note that this makes AMUSE a special case of SOBI with $|\mathcal{T}| = 1$. An estimate for $\mathbf{\Omega}^{-1}$ is given by $\hat{\mathbf{\Gamma}} = \hat{\mathbf{U}}^\top \hat{\mathbf{S}}_0^{-1/2}$ and it is consistent (up to permutation and signs of its rows) if, for all pairs of

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

sources, there exists a lag $\tau \in \mathcal{T}$, such that the lag-τ autocovariances of the two sources differ, see Belouchrani et al. (1997); Miettinen et al. (2014, 2016). For more details about solving the maximization problem in (1.2) see for example Illner et al. (2015) and the references therein.

We now turn back to our problem at hand, the signal dimension estimation. Of the two estimates proposed in literature, both rely on SOBI (with AMUSE as a special case) and the first (Matilainen et al., 2018) bases the approach on the following set of hypotheses for $k = 0, \ldots, p-1$,

$$H_{0k}: \ \mathbf{z}_t \text{ contains a } (p-k)\text{-subvector of white noise.} \tag{1.3}$$

A suitable test statistic for H_{0k} is given, e.g., by the mean of the last $p-k$ squared diagonal elements of $\hat{\mathbf{U}}^\top \hat{\mathbf{S}}_0^{-1/2} \hat{\mathbf{R}}_\tau \hat{\mathbf{S}}_0^{-1/2} \hat{\mathbf{U}}$ over all $\tau \in \mathcal{T}$, where $\hat{\mathbf{U}}$ is the maximizer of (1.2). This is based on the fact that all autocovariances of white noise series vanish, see Section 2 for a more detailed motivation. Matilainen et al. (2018) use bootstrapping to obtain the null distributions of the test statistics and sequence several of the tests together to estimate the signal dimension q, see the end of Section 2 for various strategies. Similar techniques have been used for the dimension estimation of iid data in Nordhausen et al. (2016, 2017).

An alternative approach is proposed by Nordhausen and Virta (2018) who extend the *ladle estimate* of Luo and Li (2016) to the time series

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

framework. The estimate is based on combining the classical scree plot with the bootstrap variability (Ye and Weiss, 2003) of the joint diagonalizer and has the advantage of estimating the dimension directly, without any need for hypothesis testing.

We complement these approaches by devising an asymptotic test for the null hypotheses (1.3), operating under semiparametric assumptions on the source distributions. Sequencing several tests together then allows us to obtain a consistent estimate for the true signal dimension. Using simulations, the test is showed to enjoy the standard properties of asymptotic tests, that is, computational speed and efficiency under time series of moderate and large lengths. The first of these properties is especially important and desirable, considering that the only competitors of the proposed method are based on computationally costly data resampling techniques. Moreover, the mathematical form of the proposed asymptotic test is shown to be particularly simple and elegant.

The paper is structured as follows. In Section 2 we present our main results and discuss the implications and strictness of the assumptions required for them to hold. Section 3 contains the technical derivations that lead to the results in Section 2 and can be safely skipped by a casual reader. The proofs of the results are collected in the supplemental appendix. Section 4

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

sees us comparing the proposed dimension estimate to the bootstrap- and ladle estimates under various settings using simulated data. In Section 5 we apply the proposed method to estimate the dimension of a sound recording data set and in Section 6, we finally conclude with some prospective ideas for future research.

2. Main results

In this section we present our main results and the assumptions required by them. The more technical content is postponed to Section 3 and can be skipped if the reader is not interested in the theory behind the results.

Let the observed time series $\mathbf{x}_t$ come from the SOS-model (1.1) and denote by $\lambda^*_{\tau\ell}$ the τ-lag autocovariance of the ℓth component of $\mathbf{z}_t$. Recall that SOBI jointly diagonalizes the set of standardized and symmetrized autocovariance matrices $\hat{\mathbf{H}}_\tau = \hat{\mathbf{S}}_0^{-1/2}\hat{\mathbf{R}}_\tau\hat{\mathbf{S}}_0^{-1/2}$, $\tau \in \mathcal{T}$, to obtain the orthogonal matrix $\hat{\mathbf{U}}$. Let next $\text{diag}(\mathbf{A})^2$ denote a diagonal matrix with diagonal elements equal to the squares of the diagonal elements of $\mathbf{A}$. Order the columns of $\hat{\mathbf{U}}$ such that the sums of squared pseudo-eigenvalues, $\sum_{\tau\in\mathcal{T}} \text{diag}(\hat{\mathbf{U}}^\top\hat{\mathbf{H}}_\tau\hat{\mathbf{U}})^2$, are in a decreasing order and partition $\hat{\mathbf{U}}$ as $(\hat{\mathbf{V}}_k, \hat{\mathbf{W}}_k)$ where $\hat{\mathbf{V}}_k \in \mathbb{R}^{p\times k}$, $\hat{\mathbf{W}}_k \in \mathbb{R}^{p\times(p-k)}$ for some fixed k. We show in Section 3 that, for large T, this ordering places the noise components after the signals in the estimated sources.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

If the null hypothesis H_{0k} is true, the autocovariance matrices of the last $p-k$ estimated sources,

$$\hat{\mathbf{D}}_{\tau k} = \hat{\mathbf{W}}_k^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_k,$$

are then expected to be close to zero matrices due to the last sources being (at least, asymptotically) white noise. To accumulate information over multiple lags, we use as our test statistic for H_{0k} the mean of the squared elements of the matrices $\hat{\mathbf{D}}_{\tau k}$ over a fixed set of lags $\tau \in \mathcal{T}$,

$$\hat{m}_k = \frac{1}{|\mathcal{T}|(p-k)^2} \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{D}}_{\tau k}\|^2,$$

which likewise measures departure from the null hypothesis H_{0k}. In the special case of AMUSE we have only a single matrix $\hat{\mathbf{D}}_{\tau k}$, which can in practice be obtained using the easier-to-compute generalized eigendecomposition, instead of joint diagonalization. Note that it is possible for the matrices $\hat{\mathbf{D}}_{\tau k}$ to be small in magnitude even if the number of white noise sources in the model is less than $p-k$. This situation can arise if some of the signal series are indistinguishable from white noise based on autocovariances alone and as such we need to restrict the set of signal distributions we can consider. The next assumption guarantees that each signal component exhibits non-zero autocovariance for at least one lag $\tau \in \mathcal{T}$, and is thus distinguishable from white noise.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Assumption 1. For all $\ell = 1, \ldots, d$, there exists $\tau \in \mathcal{T}$ such that $\lambda^*_{\tau\ell} \neq 0$.

Considering that most signals encountered in practice exhibit autocorrelation, Assumption 1 is rather non-restrictive. Moreover, we can always increase the number of feasible signal processes by incorporating more lags in $\mathcal{T}$. However, there exists time series which, while not being white noise, still have zero autocorrelation for all finite lags. For example, stochastic volatility models (see for example Mikosch et al., 2009) belong to this class of processes, and consequently, by Assumption 1, they are excluded from our model (see, however Section 6 for an idea on how to incorporate these distributions in the model).

The second assumption we need is more technical in nature and requires that the source series come from a specific, wide class of stochastic processes. A similar assumption is utilized also in Miettinen et al. (2012, 2014, 2016).

Assumption 2. The latent series $\mathbf{z}_t$ are linear processes having the MA(∞)-representation,

$$\mathbf{z}_t = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j},$$

where $\boldsymbol{\epsilon}_t \in \mathbb{R}^p$ are second-order standardized, iid random vectors with exchangeable, marginally symmetric components having finite fourth order moments and $\boldsymbol{\Psi}_j \in \mathbb{R}^{p\times p}$ are diagonal matrices satisfying $\sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j^2 = \mathbf{I}_p$ and $\| \sum_{j=-\infty}^{\infty} |\boldsymbol{\Psi}_j| \| < \infty$ where $|\boldsymbol{\Psi}_j| \in \mathbb{R}^{p\times p}$ denotes the matrix of

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

component-wise absolute values of $\boldsymbol{\Psi}_j$. Moreover the lower right $(p-q) \times (p-q)$ blocks of $\boldsymbol{\Psi}_j$ (the noise) equal $\boldsymbol{\Psi}_{j00} = \delta_{j0}\mathbf{I}_{p-q}$, where $\delta_{\cdot\cdot}$ is the Kronecker delta.

Note that all second-order stationary multivariate time series can by Wold's decomposition be given a MA(∞)-representation, meaning that the most stringent part of Assumption 2 is that it requires the innovations of the sources to have identical, symmetric marginal distributions. The importance of Assumption 2 to the theory comes from the fact that under it the joint limiting distribution of the sample autocovariance matrices can be derived. As such, it could also be replaced with some other assumption guaranteeing the same thing.

With the previous, we are now able to present our main results.

Proposition 1. *Under Assumptions 1, 2 and the null hypothesis H_{0q},*

$$T|\mathcal{T}|(p-q)^2 \cdot \hat{m}_q \rightsquigarrow \chi^2_{|\mathcal{T}|(p-q)(p-q+1)/2},$$

where χ^2_ν denotes the chi-squared distribution with ν degrees of freedom.

Proposition 2. *For all $k = 0, \ldots, p-1$, let $(c_{k,T})$ be a sequence such that $c_{k,T} \to \infty$ and $c_{k,T} = o(T)$. Then, under Assumptions 1, 2 and the null hypothesis H_{0q},*

$$\hat{q} = \min\{k \mid T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k < c_{k,T}\} \to_p q.$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

The limiting distribution in Proposition 1 is remarkably simple, does not depend on the type of white noise and requires no parameters to estimate, implying that it is also fast to use in practice. Note that the number of degrees of freedom of the limiting distribution is equal to the total number of free elements in the symmetric matrices $\hat{\mathbf{D}}_{\tau q}$, $\tau \in \mathcal{T}$. Thus, each of the elements can be seen to asymptotically contribute a single χ_1^2 random variable to the test statistic.

The impact of the number of used autocovariance matrices is visible in Proposition 1 in that using more lags increases the degrees of freedom of the limiting distribution. This effect is also seen for finite samples where using a higher number of lags, which requires a larger number of parameters to estimate, induces more variability in the results (see, e.g., Table 2 in Section 4 where AMUSE with just a single lag beats SOBI in asymptotic testing). Thus, for smaller sample sizes it might be advisable to restrict to a smaller number of lags. However, on the other hand, using more lags equates to using more information to separate the signals from the noise and if we are not certain of what kind of autocovariances the signals might exhibit, incorporating additional lags could help us identify additional signals. Thus, as a compromise between these conflicting viewpoints, we suggest using SOBI with a small or moderate amount of lags, such as 1-6. This matter

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

will be further investigated in the simulation studies in Section 4.

Proposition 2 introduces a consistent estimate for the true dimension. However, since the result is asymptotical and gives no indication of the required time series length T for which the estimate $\hat{q}$ takes values close enough to q for a given set of sequences $c_{k,T}$, using Proposition 2 in practice would require a very carefully tailored selection of $c_{k,T}$. As such, in the simulation section we approach the estimation in the following, more practical way. That is, to estimate the signal dimension in the p-dimensional BSS-model (1.1), we sequence together a set of asymptotic tests for the null hypotheses $H_{00}, H_{01}, \ldots, H_{0(p-1)}$. Denote the string of p-values produced by these tests by $(p_0, p_1, \ldots, p_{p-1})$ and fix a level of significance α. Different estimates for q are now obtained by considering the p-values via various strategies. The forward estimate of q is the smallest k for which $p_k \geq \alpha$. The backward estimate of q is $k+1$ where k is the largest value for which $p_k < \alpha$. The divide-and-conquer estimate is obtained by iteratively halving the search interval until a change point from $< \alpha$ to $\geq \alpha$ is found.

The proof of Lemma 5 in Section 3 shows that the test statistic is monotone in k in the sense that $T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k$ is a decreasing function of k. However, the associated p-values for the null hypotheses H_{0k} need not be monotone as the null distributions (and their quantiles) are also functions

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

of k. This means that the different estimation strategies, forward, backward and divide-and-conquer, can possible yield different estimates in practice.

3. Theoretical derivations

Throughout this section, we assume the SOS-model (1.1) and a fixed set of lags $\mathcal{T} = \{\tau_1, \ldots, \tau_{|\mathcal{T}|}\}$. Moreover, we work under the assumption of identity mixing, $\mathbf{\Omega} = \mathbf{I}_p$, which is without loss of generality as SOBI is affine equivariant, meaning that the source estimates do not depend on the value of $\mathbf{\Omega}$ (Miettinen et al., 2016). To ensure identifiability of $\mathbf{\Omega}$ we may further set $\mathbf{S}_0 = \mathrm{E}(\mathbf{x}_t \mathbf{x}_t^\top) = \mathbf{I}_p$. We assume a fixed null hypothesis H_{0q} and denote the number of white noise components by $r = p - q$.

The population autocovariance matrices are denoted by $\mathbf{S}_\tau = \mathrm{E}(\mathbf{x}_t \mathbf{x}_{t+\tau}^\top)$ and $\mathbf{R}_\tau = (\mathbf{S}_\tau + \mathbf{S}_\tau^\top)/2$, and by the identity mixing and uncorrelatedness of the latent series we have,

$$\mathbf{S}_\tau = \mathbf{R}_\tau = \mathbf{D}_\tau = \begin{pmatrix} \mathbf{\Lambda}_\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0}, \end{pmatrix},$$

where $\mathbf{\Lambda}_\tau$ is a $q \times q$ diagonal matrix, $\tau \in \mathcal{T}$. The lower right block of the matrix $\mathbf{D}_\tau$ vanishes for all $\tau \in \mathcal{T}$ as autocovariances of a white noise series are zero. Without loss of generality, we assume that the signals are ordered in $\mathbf{z}_{1t}$ such that the diagonal elements of $\sum_{\tau \in \mathcal{T}} \mathbf{\Lambda}_\tau^2$ are in decreasing

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

order. Moreover, if there are ties, we fix the order by ordering the tied components in decreasing order with respect to the diagonal elements of $\boldsymbol{\Lambda}_{\tau_1}^2$. If we still have ties, we order the tied components in decreasing order with respect to the diagonal elements of $\boldsymbol{\Lambda}_{\tau_2}^2$ and so on. If after all this we still have tied components, we set them in arbitrary order and note that such sets of tied components have the same autocovariance structure for all lags $\tau \in \mathcal{T}$, making them indistinguishable by SOBI. However, this just makes the individual signals unestimable and does not affect the estimation of the dimension in any way, as long as Assumption 1 holds.

Partition then the signals into v groups such that each group consists solely of signals with matching autocovariance structures on all lags $\tau \in \mathcal{T}$ and such that each pair of distinct groups has a differing autocovariance for at least one lag $\tau \in \mathcal{T}$. The size of the jth group is denoted by p_j, implying that $p_1 + \cdots p_v = q$. By Assumption 1, the white noise forms its own group not intersecting with any of the signal groups, and in the following we refer to the noise group with the index 0, as in $p_0 = r$. If $v = 1$, all signal components are indistinguishable by their autocovariances and in the other extreme, $v = q$, no ties occurred in ordering the signals and each signal pair has differing autocovariances for at least one lag $\tau \in \mathcal{T}$.

We introduce yet one more assumption which is actually implied by

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Assumption 2 and is as such not strictly necessary. However, some of the following auxiliary results are interesting on their own and can be shown to hold under Assumption 3, without the need for Assumption 2.

Assumption 3. The sample covariance matrix and the sample autocovariance matrices are root-T consistent, $\sqrt{T}(\hat{\mathbf{S}}_\tau - \mathbf{D}_\tau) = \mathcal{O}_p(1)$, for $\tau \in \mathcal{T} \cup \{0\}$, where $\mathbf{D}_0 = \mathbf{I}_p$.

We begin with a simple linearization result for the standardized autocovariance matrices. The notation $\hat{\mathbf{H}}_{\tau 00}, \hat{\mathbf{R}}_{\tau 00}$ in Lemma 1 refers to the lower right $r \times r$ diagonal blocks of the matrices $\hat{\mathbf{H}}_\tau = \hat{\mathbf{S}}_0^{-1/2} \hat{\mathbf{R}}_\tau \hat{\mathbf{S}}_0^{-1/2}$ and $\hat{\mathbf{R}}_\tau$. Under H_{0q} these sub-matrices gather the autocovariances of the noise components.

Lemma 1. *Under Assumption 3 we have*

$$\hat{\boldsymbol{H}}_\tau = \hat{\boldsymbol{R}}_\tau + \mathcal{O}_p(1/\sqrt{T}), \quad \textit{for all } \tau \in \mathcal{T}.$$

If H_{0q} further holds, then,

$$\hat{\boldsymbol{H}}_{\tau 00} = \hat{\boldsymbol{R}}_{\tau 00} + \mathcal{O}_p(1/T), \quad \textit{for all } \tau \in \mathcal{T}.$$

Our second auxiliary result shows that, under Assumptions 1 and 3, the SOBI solution is, while not identifiable, of a very specific asymptotic form (up to permutation). The block division and indexing in Lemma 2

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

are based on the division of the sources into the $v+1$ groups of equal autocovariances.

Lemma 2. *Under Assumptions 1, 3 and the null hypothesis H_{0q}, there exists a sequence of permutation matrices $\hat{\boldsymbol{P}}$ such that,*

$$\hat{\boldsymbol{U}}\hat{\boldsymbol{P}} = \begin{bmatrix} \hat{\boldsymbol{U}}_{11} & \cdots & \hat{\boldsymbol{U}}_{1v} & \hat{\boldsymbol{U}}_{10} \\ \vdots & \ddots & \vdots & \vdots \\ \hat{\boldsymbol{U}}_{v1} & \cdots & \hat{\boldsymbol{U}}_{vv} & \hat{\boldsymbol{U}}_{v0} \\ \hat{\boldsymbol{U}}_{01} & \cdots & \hat{\boldsymbol{U}}_{0v} & \hat{\boldsymbol{U}}_{00} \end{bmatrix},$$

where the diagonal blocks (shaded) satisfy $\hat{\boldsymbol{U}}_{ii} = \mathcal{O}_p(1)$ and the off-diagonal blocks satisfy $\hat{\boldsymbol{U}}_{ij} = \mathcal{O}_p(1/\sqrt{T})$.

Corollary 1. *Under the assumptions of Lemma 2, we have for each $j = 0, 1, \ldots, v$ that,*

$$\hat{\boldsymbol{U}}_{jj}^\top \hat{\boldsymbol{U}}_{jj} - \boldsymbol{I}_{p_j} = \mathcal{O}_p(1/T) \quad \text{and} \quad \hat{\boldsymbol{U}}_{jj} \hat{\boldsymbol{U}}_{jj}^\top - \boldsymbol{I}_{p_j} = \mathcal{O}_p(1/T).$$

The first v diagonal blocks in the block matrix of Lemma 2 correspond to the groups of signals that are mutually indistinguishable and the final diagonal block to the r noise components (which are also indistinguishable from each other). The main implication of Lemma 2 is that the sources within a single group can not be separated by SOBI but the signals coming from two different groups can be, the mixing vanishing at the rate of root-T.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

In the special case of $p_j = 1$, for all $j = 0, 1, \ldots, v$, Lemma 2 is an instant consequence of (Miettinen et al., 2016, Theorem 1(ii)).

The next lemma states that our test statistic is under the null asymptotically equivalent to a much simpler quantity, not depending on the estimation of the SOBI-solution $\hat{\mathbf{U}}$.

Lemma 3. *Under Assumptions 1, 3 and the null hypothesis H_{0q}, we have,*

$$T \cdot \hat{m}_q = T \cdot \hat{m}_q^* + o_p(1),$$

where

$$\hat{m}_q^* = \frac{1}{|\mathcal{T}| r^2} \sum_{\tau \in \mathcal{T}} \|\hat{\boldsymbol{R}}_{\tau 00}\|^2,$$

and $\hat{\boldsymbol{R}}_{\tau 00}$ is the lower right $r \times r$ block of $\hat{\boldsymbol{R}}_{\tau}$.

To compute the limiting distribution of the proxy $\hat{m}_q^*$ we next show that the joint limiting distribution of the blocks $\hat{\mathbf{R}}_{\tau 00}$ is under Assumption 2 and H_{0q} conveniently a multivariate normal distribution. The result is a slight modification of (Miettinen et al., 2016, Lemma 1). In the statement of Lemma 4, $\mathbf{J}_r$ denotes the $r \times r$ matrix filled with ones and $\mathbf{E}_r^{ij}$ denotes the $r \times r$ matrix filled otherwise with zeroes but with a single one as the (i, j)th element.

Lemma 4. *Under Assumption 2 and the null hypothesis H_{0q}, the blocks*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$\hat{\boldsymbol{R}}_{\tau_1 00}, \ldots, \hat{\boldsymbol{R}}_{\tau_{|\mathcal{T}|} 00}$ *have a joint limiting normal distribution,*

$$\sqrt{T}\text{vec}\left(\hat{\boldsymbol{R}}_{\tau_1 00}, \ldots, \hat{\boldsymbol{R}}_{\tau_{|\mathcal{T}|} 00}\right) \rightsquigarrow \mathcal{N}_{|\mathcal{T}|r^2}(\boldsymbol{0}, \boldsymbol{V}),$$

where vec *is the column-vectorization operator,*

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_0 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{V}_0 \end{pmatrix} \in \mathbb{R}^{|\mathcal{T}|r^2 \times |\mathcal{T}|r^2},$$

and $\boldsymbol{V}_0 = \text{diag}(\text{vec}(\boldsymbol{J}_r + \boldsymbol{I}_r)/2)(\boldsymbol{K}_{rr} - \boldsymbol{D}_{rr} + \boldsymbol{I}_{r^2})$ *where* $\boldsymbol{K}_{rr} = \sum_{i=1}^{r} \sum_{j=1}^{r} \boldsymbol{E}_r^{ij} \otimes \boldsymbol{E}_r^{ji}$ *and* $\boldsymbol{D}_{rr} = \sum_{i=1}^{r} \boldsymbol{E}_r^{ii} \otimes \boldsymbol{E}_r^{ii}$.

Lemmas 3 and 4 now combine to establish the limiting null distribution of the test statistic to be the remarkably simple chi-squared distribution, see Proposition 1 in Section 2.

Finally, to prove Proposition 2, we establish the following result detailing the asymptotic behavior of the test statistics for different k under a fixed null hypothesis.

Lemma 5. *Under Assumptions 1, 2 and the null hypothesis* H_{0q}*,*

1. *for* $k < q$, $\quad T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k \geq T(b + o_p(1))$, *for some* $b > 0$,
2. *for* $k \geq q$, $\quad T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k = \mathcal{O}_p(1)$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Lemma 5 shows that the true dimension q is the smallest value of k for which the test statistic $T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k$ is bounded in probability. This idea is formalized in Proposition 2 in Section 2.

4. Simulations

The following results are all obtained in R (R Core Team, 2017) using the packages JADE (Nordhausen et al., 2017) and tsBSS (Matilainen et al., 2018).

4.1 Evaluation of the hypothesis testing

In the first set of simulations we consider the performance of the hypothesis tests. As our competitor we use the recommended and most general non-parametric bootstrapping strategy from Matilainen et al. (2018), which takes bootstrap samples from the hypothetical multivariate noise part. The number of bootstrap samples used was 200. We computed also the three other bootstrapping strategies as listed in Matilainen et al. (2018), but as the results were basically the same, we report for simplicity only the strategy mentioned above.

We considered three different settings for the latent sources:

Setting H1: MA(3), AR(2) and ARMA(1,1) having Gaussian innovations

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

together with two Gaussian white noise components.

Setting H2: MA(10), MA(15) and M(20) processes having Gaussian innovations together with two Gaussian white noise components.

Setting H3: Three MA(3) processes having Gaussian innovations and identical autocovariance functions together with two Gaussian white noise processes.

Hence, in all three settings the signal dimension is $q = 3$ and the total dimension is $p = 5$. Due to affine equivariance of the used methods, we take without loss of generality $\boldsymbol{\Omega} = \mathbf{I}_5$. In general, setting H1 can be considered a short range dependence model and H2 a long range dependence model. Setting H3 is special in that the methods should not be able to separate its signals, but they should still be able to separate the noise space from the signal space. We also considered additional settings H1t, H2t, H3t which were otherwise identical to H1, H2, H3 but had the Gaussian innovations and white noise series replaced by independent univariate standardized t_5-distributed random variables. Interestingly, the results with the t_5-distribution were almost identical to the Gaussian results for the first two settings, showing especially that the convergence of the test statistic to its asymptotic distribution seems to depend very little on the heavy-tailedness

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

of the underlying innovations if the signals are identifiable. Due to this similarity of the results, the results for settings H1t, H2t, H3t are presented only in the supplemental appendix.

Based on 2000 repetitions, we give the rejection frequencies of the null hypotheses H_{02}, H_{03} and H_{04} at level $\alpha = 0.05$ in Tables 1-9. We considered three different BSS-estimators, AMUSE with $\tau = 1$, SOBI with $\mathcal{T} = \{1, \ldots, 6\}$ (denoted SOBI6) and SOBI with $\mathcal{T} = \{1, \ldots, 12\}$ (denoted SOBI12). The optimal rejection rates at level $\alpha = 0.05$ are 1.00 for H_{02}, 0.05 for H_{03} and < 0.05 for H_{04}.

Table 1: Rejection frequencies of H_{02} in Setting H1 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	1.000	1.000	1.000	0.999	0.998	0.998
500	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000
5000	1.000	1.000	1.000	1.000	1.000	1.000

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 2: Rejection frequencies when testing H_{03} in Setting H1 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.059	0.050	0.078	0.050	0.102	0.050
500	0.053	0.048	0.064	0.049	0.071	0.052
1000	0.048	0.047	0.059	0.053	0.054	0.050
2000	0.050	0.054	0.048	0.049	0.054	0.046
5000	0.048	0.052	0.052	0.047	0.056	0.053

Table 3: Rejection frequencies when testing H_{04} in Setting H1 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.006	0.008	0.015	0.006	0.024	0.004
500	0.006	0.007	0.009	0.004	0.016	0.006
1000	0.007	0.010	0.012	0.005	0.012	0.006

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

2000	0.003	0.006	0.008	0.003	0.009	0.002
5000	0.006	0.006	0.006	0.002	0.008	0.004

Table 4: Rejection frequencies when testing H_{02} in Setting H2 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.038	0.043	0.608	0.484	0.911	0.848
500	0.090	0.094	0.988	0.987	1.000	1.000
1000	0.190	0.189	1.000	1.000	1.000	1.000
2000	0.252	0.256	1.000	1.000	1.000	1.000
5000	0.558	0.550	1.000	1.000	1.000	1.000

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 5: Rejection frequencies when testing H_{03} in Setting H2 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.002	0.006	0.125	0.050	0.148	0.063
500	0.008	0.014	0.075	0.041	0.074	0.050
1000	0.010	0.014	0.067	0.046	0.068	0.047
2000	0.020	0.024	0.056	0.052	0.066	0.061
5000	0.031	0.039	0.051	0.048	0.054	0.047

Table 6: Rejection frequencies when testing H_{04} in Setting H2 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.000	0.001	0.034	0.004	0.039	0.006
500	0.002	0.004	0.010	0.004	0.016	0.007
1000	0.000	0.004	0.012	0.004	0.007	0.001

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

2000	0.002	0.004	0.010	0.004	0.010	0.003
5000	0.004	0.008	0.010	0.005	0.007	0.003

Table 7: Rejection frequencies when testing H_{02} in Setting H3 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.036	0.042	0.600	0.479	0.906	0.84
500	0.084	0.092	0.986	0.987	1.000	1.00
1000	0.168	0.175	1.000	1.000	1.000	1.00
2000	0.279	0.272	1.000	1.000	1.000	1.00
5000	0.576	0.568	1.000	1.000	1.000	1.00

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 8: Rejection frequencies when testing H_{03} in Setting H3 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.004	0.005	0.122	0.049	0.146	0.047
500	0.006	0.008	0.075	0.043	0.074	0.057
1000	0.010	0.018	0.062	0.050	0.062	0.055
2000	0.016	0.023	0.058	0.044	0.046	0.054
5000	0.034	0.042	0.051	0.050	0.048	0.045

Table 9: Rejection frequencies when testing H_{04} in Setting H3 at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.000	0.002	0.026	0.005	0.034	0.006
500	0.000	0.002	0.012	0.003	0.012	0.003
1000	0.002	0.002	0.010	0.005	0.010	0.005

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

2000	0.004	0.006	0.008	0.007	0.006	0.004
5000	0.005	0.010	0.008	0.004	0.006	0.003

The results of the simulations can be summarized as follows. (i) There is no big difference between the limiting theory and the bootstrap tests, which is a clear advantage for the asymptotic test as neither a bootstrapping strategy has to be selected nor is the asymptotic test computationally demanding. The rejection rates of the two types of tests are in most of the cases within 0.01 of each other already for time series of length $T = 1000$, a relatively small amount of observations in the usual signal processing applications. (ii) The number of matrices to be diagonalized seems to matter. If the dependence structure is of a short range AMUSE works well, but it seems to struggle in the case of long range dependence. In the considered settings SOBI with 6 matrices seems to be a good compromise. (iii) Even when the signals cannot be individually separated, the noise and signal subspaces can be separated accurately. Additionally, it seems that having heavy-tailed innovations benefits the dimension estimation in this setting, especially for AMUSE, see the supplemental appendix.

In general, having a very good power under the alternative hypotheses of too large noise subspaces is desirable when using successive testing strate-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

gies to estimate the dimension. This was not yet evaluated in Matilainen et al. (2018) and will be done in the next section.

4.2 Evaluation of determining the dimension of the signal

In this section we evaluate in a simulation study the performance of our test when the goal is to estimate the signal dimension q. Several different testing strategies are possible, as described in the end of Section 2. We will use in the following the divide-and-conquer strategy as it seems the most practical. For simplicity, all tests will be performed at the level $\alpha = 0.05$.

As competitors we use again the bootstrap tests, this time including all three nonparametric bootstraps and the parametric bootstrap. For details we refer to Matilainen et al. (2018). As an additional contender we use the ladle estimator as described in Nordhausen and Virta (2018). Also for the ladle different bootstrap strategies are possible and we consider the fixed block bootstrap with the block lengths 20 and 40 and the stationary block bootstrap with the expected block lengths 20 and 40, see Nordhausen and Virta (2018) for details. For all bootstrap-based methods the number of bootstrapping samples is again 200 and, as in the previous section, we consider the three estimators, AMUSE, SOBI6 and SOBI12.

The settings considered in this simulation are:

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Setting D1: AR(2), AR(3), ARMA(1,1), ARMA(3,2) and MA(3) processes having Gaussian innovations together with five Gaussian white noise components.

Setting D2: Same processes as in D1 but the MA(3) is changed to an MA(1) process with the parameter equal to 0.1.

Setting D3: Five MA(2) processes with parameters (0.1, 0.1) having Gaussian innovations together with five Gaussian white noise processes.

Hence, in all settings $p = 10$ and $q = 5$. Setting D1 is the basic setting whereas in Setting D2 there is one very weak signal. In Setting D3 all five signals come from identical processes and exhibit weak dependence. We also considered the additional settings D1t, D2t, D3t where the Gaussian innovations and white noise series in D1, D2, D3 were replaced by independent univariate standardized t_5-distributed random variables. Since the results with the t_5-distribution were again very similar to the Gaussian results, we have postponed them to the supplemental appendix.

As in the previous simulation, the mixing matrix used is $\mathbf{\Omega} = \mathbf{I}_{10}$ and Figures 1-3 show, based on 2000 repetitions, the frequencies of the estimated signal dimensions.

In Setting D1, the asymptotic test seems not to work as well as the other

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Figure 1: Estimating q by divide-and-conquer in Setting D1.

methods for small samples but in general the difference to the bootstrap-based testing procedures is negligible. In general, the ladle is the most preferable option. In setting D2, on the other hand, ladle consistently underestimates the signal dimension and is not able to find the weak signal. When using the hypothesis testing-based methods also the weak signal is identified with increasing sample size. However, the more scatter matrices we estimate, the more difficult the estimation gets and thus AMUSE works the best.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

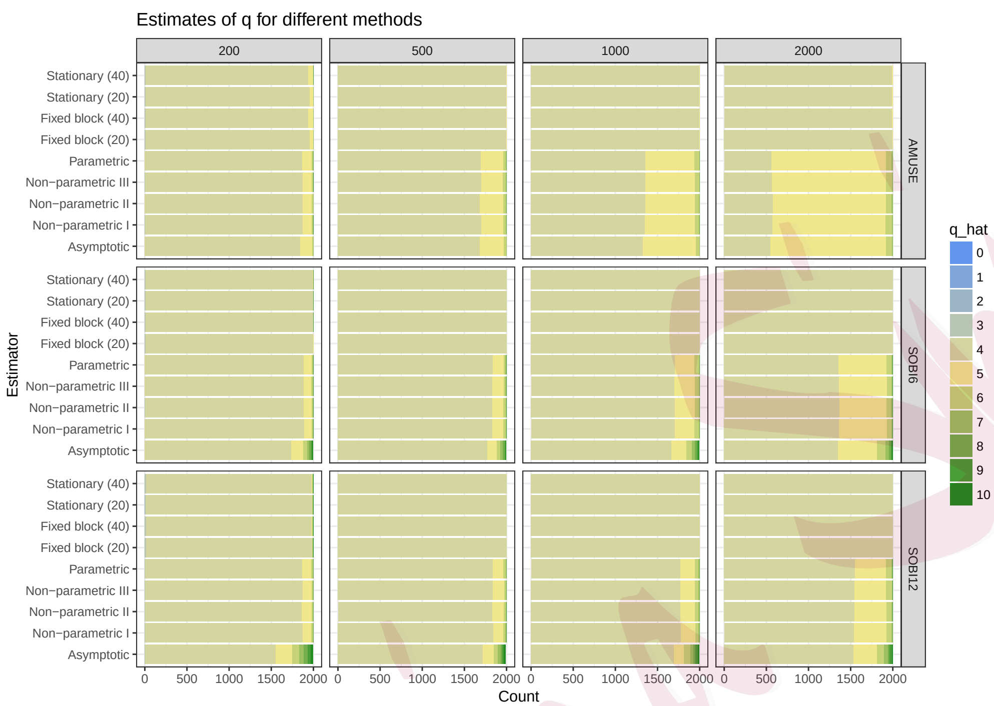

Figure 2: Estimating q by divide-and-conquer in Setting D2.

In Setting D3 the ladle fails completely and keeps getting worse with increasing sample size. The difference between bootstrapping and asymptotic testing is at its largest in this setting, the asymptotic test seems to be the most preferable option. As two lags are needed to capture all the temporal information, AMUSE is at an disadvantage in this setting, and this is clearly visible in the plots. Also, SOBI6 seems to exploit the lag information better than SOBI12, possibly because it avoids the inclusion of several unnecessary autocovariance matrices in the estimation.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

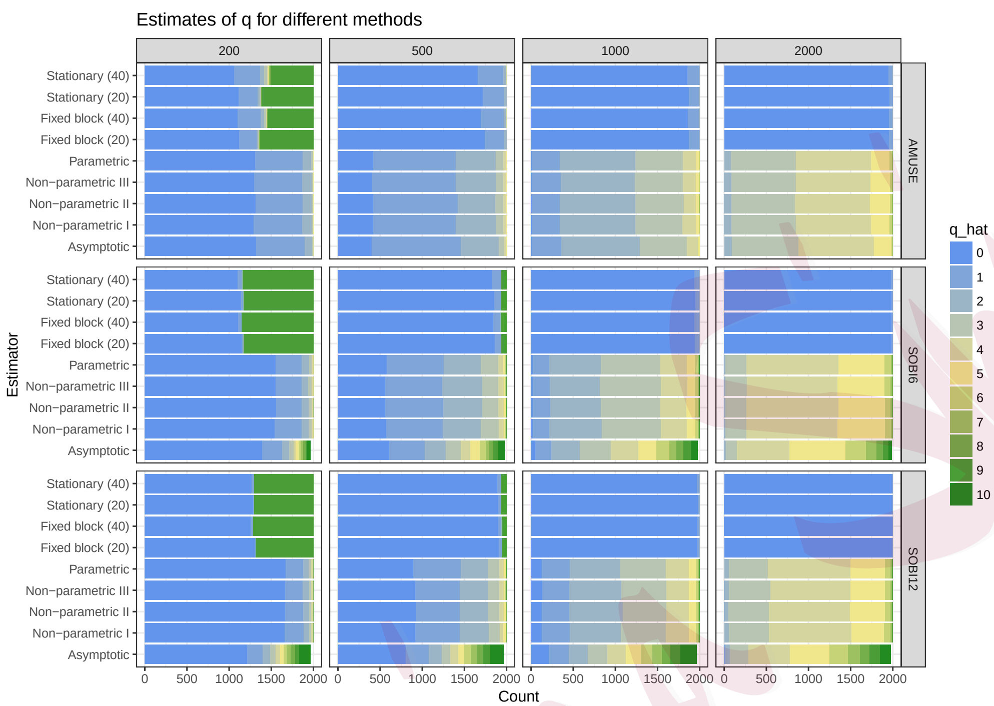

Figure 3: Estimating q by divide-and-conquer in Setting D3.

5. Data example

For our real data example we use the recordings of three sounds signals available in the R-package JADE and analysed, for example, in Miettinen et al. (2017). To the three signal components we added 17 white noise components which all had t_5-distributions to study whether the methods also work in case of non-Gaussian white noise. After standardizing the 20 components to have unit variances, we used a random square matrix where each element came from the uniform distribution on $[0, 1]$. The original

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

signals had a length of 50000 and for convenience we selected the first 10000 instances. The 20 mixed components are visualized in the supplemental appendix and reveal no clear structure.

We used the divide-and-conquer approach to estimate the signal dimension with our asymptotic test and the bootstrapping strategy of Matilainen et al. (2018) used in Section 4.1. Additionally, we considered also the ladle estimator using stationary bootstrapping with the expected block length 40. Of each estimator, three versions, AMUSE, SOBI6 and SOBI12, were computed. All nine estimators estimated correctly 3 as the signal dimension and the estimated signals based on SOBI6 are shown in Figure 4.

The computation times of the nine methods were, however, quite different and are given in Table 10.

Table 10: Computation times (in seconds) of the nine estimators for the sound example data.

Asymptotic tests			Bootstrap tests			Ladle estimator		
AMUSE	SOBI6	SOBI12	AMUSE	SOBI6	SOBI12	AMUSE	SOBI6	SOBI12
0.07	0.19	0.49	15.08	47.24	88.08	2.75	9.85	18.17

As all the approaches estimated the dimension correctly, the ones based on the asymptotic test are clearly favourable due to their computational

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

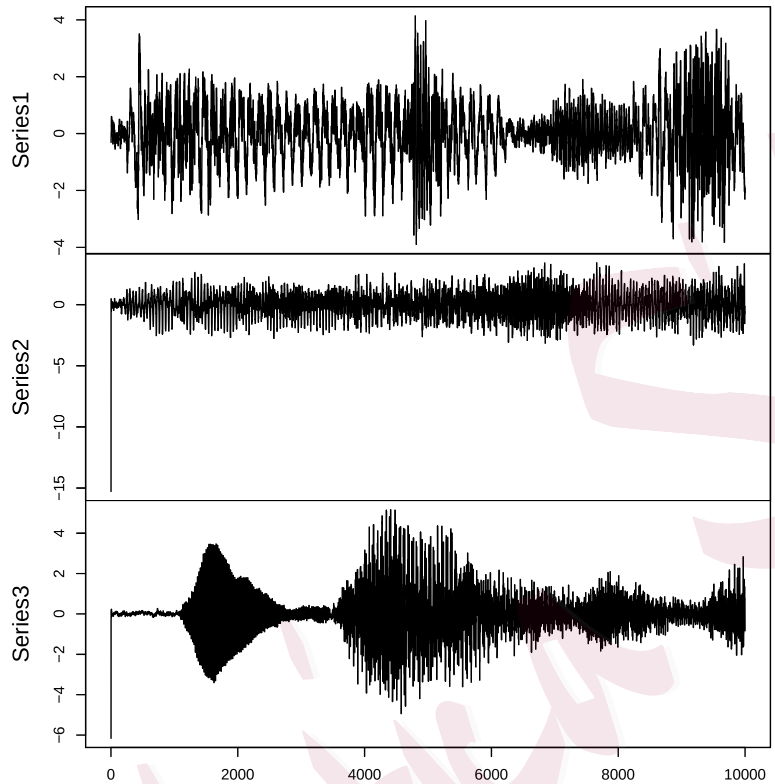

Figure 4: The three estimated sound signals based on SOBI6.

speed. Although, we note that the ladle and the bootstrap tests can also be run parallelized while in the current comparison we used only a single core. The ladle estimator is naturally faster to compute than the divide-and-conquer hypothesis testing strategy as the latter needs to create separate bootstrap samples for each hypothesis to be tested whereas the ladle estimator conducts the bootstrap sampling only once.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

6. Discussion

We proposed an asymptotic test for estimating the signal dimension in an SOS-model where the sources include both signal series and white noise. The test does not require the estimation of any parameters and makes quite weak assumptions. This combined with the fact that both of its existing competitors are based on the use of computer-intensive resampling techniques makes the asymptotic test a very attractive option in practice. This conclusion was supported by our simulations studies and real data example exploring dimension estimation for sound recording data.

A drawback of the proposed method is its inability to recognize non-autocorrelated signals, such as those coming from stochastic volatility models, from white noise. One way to get around this limitation is to replace $\mathbf{z}_{2t}$ in (1.1) by a vector of stochastic volatility series and to revert the roles of signal and noise. That is, we use hypothesis testing to estimate the dimension of the “noise” subspace (containing the stochastic volatility components) and separate them from the uninteresting “signal” series exhibiting autocorrelation. For this to work, a limiting result equivalent to Lemma 4 is needed for the above combination model. Similar idea was suggested in the context of the ladle estimator already in Nordhausen and Virta (2018).

Besides (1.1), another way to incorporate noise in a time series model

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

for $\mathbf{x}_t \in \mathbb{R}^p$ is through the additive noise model,

$$\mathbf{x}_t = \boldsymbol{\Omega}\mathbf{z}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{6.4}$$

where $\boldsymbol{\Omega} \in \mathbb{R}^{p \times q}$ is a non-square mixing matrix, $\mathbf{z}_t \in \mathbb{R}^q$ consists of $q < p$ signal components and $\boldsymbol{\epsilon}_t \in \mathbb{R}^p$ is a noise process. This problem is considered in the case of i.i.d. data in Virta and Nordhausen (2018) where the authors use the PCA-transformation to estimate the $p - k$ principal components with smallest variances. Hypothesis testing can then be used to pin-point the correct dimension $k = q$ for which these components consist of pure noise. It seems likely to us that a modification of this idea could be utilized also in the context of time series data and model (6.4), and this will be considered in future work. Note, however, that in such a model only the signal dimension and the mixing matrix can be consistently estimated but not the signals itself, contrary to the model proposed in the current work.

Acknowledgements

The authors would like to express their gratitude to the anonymous referees whose comments and suggestions helped greatly to improve the manuscript quality.

Klaus Nordhausen acknowledges support from the CRoNoS COST Action IC1408.

Statistica Sinica

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

References

Belouchrani, A., K. Abed Meraim, J.-F. Cardoso, and E. Moulines (1997). A blind source separation technique based on second order statistics. *IEEE Transactions on Signal Processing 45*, 434–444.

Chang, J., B. Guo, and Q. Yao (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics 46*, 2094–2124.

Comon, P. and C. Jutten (2010). *Handbook of Blind Source Separation. Independent Component Analysis and Applications*. Academic Press.

Ensor, K. B. (2013). Time series factor models. *Wiley Interdisciplinary Reviews: Computational Statistics 5*(2), 97–104.

Illner, K., J. Miettinen, C. Fuchs, S. Taskinen, K. Nordhausen, H. Oja, and F. J. Theis (2015). Model selection using limiting distributions of second-order blind source separation algorithms. *Signal Processing 113*, 95–103.

Luo, W. and B. Li (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika 103*, 875–887.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Matilainen, M., C. Croux, J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, and J. Virta (2018). *tsBSS: Blind Source Separation and Supervised Dimension Reduction for Time Series*. R package version 0.5.2.

Matilainen, M., K. Nordhausen, and J. Virta (2018). On the number of signals in multivariate time series. In Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward (Eds.), *International Conference on Latent Variable Analysis and Signal Separation*, Cham, pp. 248–258. Springer International Publishing.

Miettinen, J., K. Illner, K. Nordhausen, H. Oja, S. Taskinen, and F. Theis (2016). Separation of uncorrelated stationary time series using autocovariance matrices. *Journal of Time Series Analysis 37*, 337–354.

Miettinen, J., K. Nordhausen, H. Oja, and S. Taskinen (2012). Statistical properties of a blind source separation estimator for stationary time series. *Statistics & Probability Letters 82*, 1865–1873.

Miettinen, J., K. Nordhausen, H. Oja, and S. Taskinen (2014). Deflation-based separation of uncorrelated stationary time series. *Journal of Multivariate Analysis 123*, 214–227.

Miettinen, J., K. Nordhausen, and S. Taskinen (2017). Blind source sep-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

aration based on joint diagonalization in R: The packages JADE and BSSasymp. *Journal of Statistical Software 76*, 1–31.

Mikosch, T., Kreiß, J.-P., R. Davis, and T. Andersen (2009). *Handbook of Financial Time Series*. Springer.

Nordhausen, K., J.-F. Cardoso, J. Miettinen, H. Oja, E. Ollila, and S. Taskinen (2017). *JADE: Blind Source Separation Methods Based on Joint Diagonalization and Some BSS Performance Criteria*. R package version 2.0-0.

Nordhausen, K., H. Oja, and D. Tyler (2016). Asymptotic and bootstrap tests for subspace dimension. *Preprint available as arXiv:1611.04908*.

Nordhausen, K., H. Oja, D. Tyler, and J. Virta (2017). Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. *IEEE Signal Processing Letters 24*, 887–891.

Nordhausen, K. and J. Virta (2018). Ladle estimator for time series signal dimension. In *Proceedings of IEEE Statistical Signal Processing Workshop 2018, IEEE SSP 2018.*

R Core Team (2017). *R: A Language and Environment for Statistical Com-*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

puting. Vienna, Austria: R Foundation for Statistical Computing. R version 3.4.1.

Tong, L., V. Soon, Y. Huang, and R. Liu (1990). AMUSE: A new blind identification algorithm. In *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 1784–1787. IEEE.

Virta, J. and K. Nordhausen (2018). Estimating the number of signals using PCA. *Submitted*.

Ye, Z. and R. E. Weiss (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association 98*, 968–979.

Aalto University, Finland

E-mail: joni.virta@aalto.fi

Vienna University of Technology

E-mail: klaus.nordhausen@tuwien.ac.at

Statistica Sinica