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: Despite being an important topic in practice, estimating 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; however, although very efficient, they suffer from long computation times as a result of 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 with resampling-based estimates show that the asymptotic test provides comparable error rates, with significantly faster computation times. Lastly, we illustrate the method by applying it to sound recording data.

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

1. Introduction

Given the difficulty of modeling multivariate time series; an increasingly common option is to use latent variable or factor models (e.g., see Ensor, 2013; Chang et al., 2018, and the references therein). In this study we follow the *blind source separation* (BSS) approach as an intermediate step prior to modeling. 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 far fewer parameters.

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

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

where the *source* series $\mathbf{z}_1, \ldots, \mathbf{z}_T$ is a latent nondegenerate, zero-mean, second-order stationary p-variate time series with uncorrelated component series, and $\boldsymbol{\Omega} \in \mathbb{R}^{p \times p}$ is an unknown, invertible matrix-valued parameter. We assume a zero mean without loss of generality, because we can always

center our series. The objective in model (1.1) is to estimate an inverse $\hat{\boldsymbol{\Gamma}}$ for $\boldsymbol{\Omega}$, giving us an estimate $\hat{\mathbf{z}}_t = \hat{\boldsymbol{\Gamma}}\hat{\mathbf{x}}_t$ for the p sources, which can then be modeled univariately.

However, because noise is often an inevitable part of a real-world signal, we incorporate it in model (1.1) through its source vector. That is, we assume that a source vector consists 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 $\mathbf{z}_{2t} \in \mathbb{R}^{p-q}$, which is a white noise vector. To avoid overfitting in the modeling phase, a crucial step in BSS is to identify the noise subvector $\mathbf{z}_{2t}$ within the sources, and then to discard it prior to modeling. This problem, called a signal-dimension estimation, has only recently been considered in the context of a statistical BSS. Thus, we propose a novel estimate that relies on asymptotic hypothesis testing. First, we review two classical SOS methods that serve as the basis for both our method and two existing methods.

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

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

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,

3

$\hat{\mathbf{R}}_\tau = (\hat{\mathbf{S}}_\tau + \hat{\mathbf{S}}_\tau^\top)/2$, $\hat{\mathbf{S}}_\tau = [1/(T - \tau)] \sum_{t=1}^{T-\tau} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t+\tau} - \bar{\mathbf{x}})^\top$ 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 $\hat{\boldsymbol{\Gamma}}$ is consistent for $\boldsymbol{\Omega}^{-1}$ up to a permutation and the signs of its rows. For the statistical properties of AMUSE, see Miettinen et al. (2012).

A usually more efficient estimate, which does not depend as much on the selection of a single parameter τ, is given by the second-order blind identification (SOBI) (Belouchrani et al., 1997), an extension of the 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 $\boldsymbol{\Omega}^{-1}$ is given by $\hat{\boldsymbol{\Gamma}} = \hat{\mathbf{U}}^\top \hat{\mathbf{S}}_0^{-1/2}$, and is consistent (up to permutation and signs of its rows) if, for all pairs of sources, there exists a lag $\tau \in \mathcal{T}$, such that the lag-τ autocovariances of

4

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 return to the problem at hand, namely, the signal dimension estimation. Both estimates that have been proposed in the literature rely on SOBI (with AMUSE as a special case). The first estimate (Matilainen et al., 2018) bases its approach on the following set of hypotheses, for $k = 0, \ldots, p - 1$:

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

A suitable test statistic for H_{0k} is given, for example, 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. Then, they sequence several of the tests together to estimate the signal dimension q; see the end of Section 2 for various strategies. Similar techniques are used for the dimension estimation of independent and identically distributed (i.i.d.) 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 a time series framework. The estimate combines 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, we show that the test enjoys 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 remainder of 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 safely be skipped

by a casual reader. The proofs of the results are collected in the online Supplementary Material. Section 4 compares the proposed dimension estimate with 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. Lastly, Section 6 concludes the paper.

2. Main results

In this section, we present our main results and the assumptions they require. 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 $\hat{\mathbf{H}}_\tau = \hat{\mathbf{S}}_0^{-1/2}\hat{\mathbf{R}}_\tau\hat{\mathbf{S}}_0^{-1/2}$, for $\tau \in \mathcal{T}$, to obtain the orthogonal matrix $\hat{\mathbf{U}}$. Let next $\mathrm{diag}(\mathbf{A})^2$ 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 the squared pseudo-eigenvalues, $\sum_{\tau\in\mathcal{T}}\mathrm{diag}(\hat{\mathbf{U}}^\top\hat{\mathbf{H}}_\tau\hat{\mathbf{U}})^2$, are in decreasing order, and partition $\hat{\mathbf{U}}$ as $(\hat{\mathbf{V}}_k, \hat{\mathbf{W}}_k)$, where $\hat{\mathbf{V}}_k \in \mathbb{R}^{p\times k}$ and $\hat{\mathbf{W}}_k \in \mathbb{R}^{p\times(p-k)}$, for some fixed k. We show in Sec-

7

tion 3 that, for large T, this ordering places the noise components after the signals in the estimated sources. 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, owing 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 also measures the 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 rather than using a 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. As such, we need to restrict the set of signal distributions we can consider. The next assumption guarantees that each signal component exhibits nonzero autocovariance for at least one lag $\tau \in \mathcal{T}$, and

8

is thus distinguishable from white noise.

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 nonrestrictive. Moreover, we can always increase the number of feasible signal processes by incorporating additional lags in $\mathcal{T}$. However, time series exist that, while not white noise, still have zero autocorrelation for all finite lags. For example, stochastic volatility models (e.g., see Mikosch et al., 2009) belong to this class of processes and, consequently, by Assumption 1, are excluded from our model (however, Section 6 discusses 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 used in Miettinen et al. (2012, 2014, 2016).

Assumption 2. The latent series $\mathbf{z}_t$ is a linear process, with the MA(∞)-representation

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

where $\boldsymbol{\epsilon}_t \in \mathbb{R}^p$ is a second-order standardized series of i.i.d. random vectors with exchangeable, marginally symmetric components that have finite

fourth-order moments, and $\mathbf{\Psi}_j \in \mathbb{R}^{p \times p}$ are diagonal matrices satisfying $\sum_{j=-\infty}^{\infty} \mathbf{\Psi}_j^2 = \mathbf{I}_p$ and $\| \sum_{j=-\infty}^{\infty} |\mathbf{\Psi}_j| \| < \infty$, where $|\mathbf{\Psi}_j| \in \mathbb{R}^{p \times p}$ denotes the matrix of component-wise absolute values of $\mathbf{\Psi}_j$. Moreover, the lower-right $(p-q) \times (p-q)$ blocks of $\mathbf{\Psi}_j$ (the noise) are equal to $\mathbf{\Psi}_{j00} = \delta_{j0} \mathbf{I}_{p-q}$, where $\delta_{..}$ is the Kronecker delta.

Note that Wold's decomposition states that all second-order stationary multivariate time series can be given an MA(∞)-representation. Thus the most stringent part of Assumption 2 is that it requires that the innovations of the sources have identical, symmetric marginal distributions. The importance of Assumption 2 is that it enables us to derive the joint limiting distribution of the sample autocovariance matrices. As such, it can be replaced with any other assumption that guarantees the same property.

We can now present our main results.

Proposition 1. *Under Assumptions 1 and 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 a 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 and 2 and the null*

10

hypothesis H_{0q},

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

The limiting distribution in Proposition 1 is remarkably simple, does not depend on the type of white noise, and requires no parameters to estimate; thus, it can be implemented quickly and easily 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 $\hat{\mathbf{D}}_{\tau q}$, for $\tau \in \mathcal{T}$. Thus, each element asymptotically contributes a single χ_1^2 random variable to the test statistic.

The impact of the number of autocovariance matrices used is visible in Proposition 1, in that using more lags increases the number of degrees of freedom of the limiting distribution. This effect is also evident in finite samples, where using a higher number of lags, which requires a larger number of parameters to estimate, induces greater variability in the results (e.g., see 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 the number of lags. However, on the other hand, using more lags equates to using more information to separate the signals from the noise. Furthermore, if we are not certain of the kinds of autocovariances the signals might exhibit, incorporating extra lags could help us identify additional

11

signals. Thus, as a compromise between these conflicting viewpoints, we suggest using SOBI with a small or moderate number of lags (e.g., between one and six). This matter is investigated further in the simulation studies in Section 4.

Proposition 2 introduces a consistent estimate for the true dimension. However, 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 $c_{k,T}$. Thus, using Proposition 2 in practice requires a careful choice 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 using 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

12

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, because the null distributions (and their quantiles) are also functions of k. As a result, the various estimation strategies can 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 assume identity mixing, $\boldsymbol{\Omega} = \mathbf{I}_p$, which is without loss of generality as SOBI is affine equivariant. Thus, the source estimates do not depend on the value of $\boldsymbol{\Omega}$ (Miettinen et al., 2016). To ensure the identifiability of $\boldsymbol{\Omega}$, 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 $\mathbf{R}_\tau = (\mathbf{S}_\tau + \mathbf{S}_\tau^\top)/2$. Then, from the identity mixing and uncorrelatedness of the latent series, we have

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

where $\boldsymbol{\Lambda}_\tau$ is a $q \times q$ diagonal matrix, for $\tau \in \mathcal{T}$. The lower-right block

13

of the matrix $\mathbf{D}_\tau$ vanishes for all $\tau \in \mathcal{T}$, because the 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 order. Moreover, we settle ties by ordering the tied components in decreasing order with respect to the diagonal elements of $\mathbf{\Lambda}_{\tau_1}^2$. Further ties are resolved by arranging the tied components in decreasing order with respect to the diagonal elements of $\mathbf{\Lambda}_{\tau_2}^2$, and so on. If, after all this, we still have tied components, we set them in arbitrary order and note that such components have the same autocovariance structure for all lags $\tau \in \mathcal{T}$, making them indistinguishable to 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.

Next, we partition 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. In the following, we refer to the noise group using the index 0, as in $p_0 = r$. If $v = 1$, all signal components are indistinguishable by their autocovariances;

at the other extreme, $v = q$, no ties occurred when 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 Assumption 2 and, as such, is 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 $\hat{\mathbf{H}}_{\tau00}, \hat{\mathbf{R}}_{\tau00}$ 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*

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

Furthermore, if H_{0q} holds, then

$$\hat{\boldsymbol{H}}_{\tau00} = \hat{\boldsymbol{R}}_{\tau00} + \mathcal{O}_p(1/T), \quad \text{for all } \tau \in \mathcal{T}.$$

15

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

Lemma 2. *Under Assumptions 1 and 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, for each $j = 0, 1, \ldots, v$, we have that*

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

The first v diagonal blocks in the block matrix of Lemma 2 correspond to the groups of signals that are mutually indistinguishable. The final diagonal block corresponds to the r noise components (which are also indistinguishable from each other). The main implication of Lemma 2 is that

SOBI cannot separate sources within a single group, but it can separate the signals coming from two different groups, with the mixing vanishing at the rate of root-T. 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 that does not depend on the estimation of the SOBI-solution $\hat{\mathbf{U}}$.

Lemma 3. *Under Assumptions 1 and 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 with the (i, j)th element equal to one, and all others equal to

zero.

Lemma 4. *Under Assumption 2 and the null hypothesis H_{0q}, the blocks $\hat{\boldsymbol{R}}_{\tau_1 00}, \ldots, \hat{\boldsymbol{R}}_{\tau_{|\mathcal{T}|} 00}$ have a joint limiting normal distribution,*

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

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 $\boldsymbol{V}_0 = \mathrm{diag}(\mathrm{vec}(\boldsymbol{J}_r + \boldsymbol{I}_r)/2)(\boldsymbol{K}_{rr} - \boldsymbol{D}_{rr} + \boldsymbol{I}_{r^2})$, 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 as the remarkably simple chi-squared distribution; see Proposition 1 in Section 2.

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

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

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

2. *for $k \geq q$, $T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k = \mathcal{O}_p(1)$.*

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 of Matilainen et al. (2018), which takes bootstrap samples from the hypothetical multivariate noise part. The number of bootstrap samples used was 200. We also computed the three other bootstrapping strategies described in Matilainen et al. (2018). However, the results remain essentially the same; thus, we omit them here for brevity.

We considered three settings for the latent sources:

Setting H1: MA(3), AR(2), and ARMA(1,1) processes, with Gaussian innovations and two Gaussian white noise components.

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

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

Hence, in all three settings, the signal dimension is $q = 3$, and the total dimension is $p = 5$. Owing to the affine equivariance of the aforementioned methods, without loss of generality, we take $\mathbf{\Omega} = \mathbf{I}_5$. In general, setting H1 can be viewed as a short-range dependence model, and H2 can be viewed as a long range dependence model. H3 is special, in that the methods should not be able to separate its signals, but should still be able to separate the noise space from the signal space. We also considered several additional settings H1t, H2t, and H3t. These are otherwise identical to H1, H2, and H3, respectively, except that we replace the Gaussian innovations and white noise series with 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. This im-

plies that the convergence of the test statistic to its asymptotic distribution depends very little on the heavy-tailedness of the underlying innovations, if the signals are identifiable. Owing to this similarity, the results for settings H1t, H2t, and H3t are presented in the Supplementary Material only.

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 BSS estimators: AMUSE, with $\tau = 1$; SOBI, with $\mathcal{T} = \{1,\ldots,6\}$ (denoted by SOBI6); and SOBI, with $\mathcal{T} = \{1,\ldots,12\}$ (denoted by 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

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

n	AMUSE		SOBI6		SOBI12	
	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 $\alpha = 0.05$.

n	AMUSE		SOBI6		SOBI12	
	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
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

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

| 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

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

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. First, there is no significant difference between the limiting theory and the bootstrap tests. This is a clearly an advantage of the asymptotic test, because we do not need to select a bootstrapping strategy, and the test is not computationally demanding. The rejection rates of the two types of tests are, in most cases, within 0.01 of each other already for time series of length $T = 1000$, which is a relatively small number of observations in signal processing applications. Second, the number of matrices to be diagonalized appears to be important. If the dependence structure is of a short range, AMUSE works well; however, it seems to struggle in the case of long-range dependence. In the considered settings, SOBI with six matrices seems to be a good compromise. Third, 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 Supplementary Material.

In general, having good power under the alternative hypotheses for

27

noise subspaces that are too large is desirable when using successive testing strategies to estimate the dimension. This was not evaluated in Matilainen et al. (2018); thus, we do so in the next section.

4.2   Signal-dimension test performance

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

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

The settings considered in this simulation are as follows.

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

Setting D2: The same processes as in D1, but 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) with Gaussian innovations and five Gaussian white-noise processes.

Hence, in all settings, $p = 10$ and $q = 5$. Setting D1 is the basic setting, and Setting D2 includes 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, and D3t, where the Gaussian innovations and white-noise series in D1, D2, and D3 were replaced with independent univariate standardized t_5-distributed random variables. The results with the t_5-distribution are again similar to the Gaussian results, and so are relegated to the Supplementary Material.

As in the previous simulation, the mixing matrix used is $\mathbf{\Omega} = \mathbf{I}_{10}$. Figures 1–3 show, based on 2000 repetitions, the frequencies of the estimated signal dimensions.
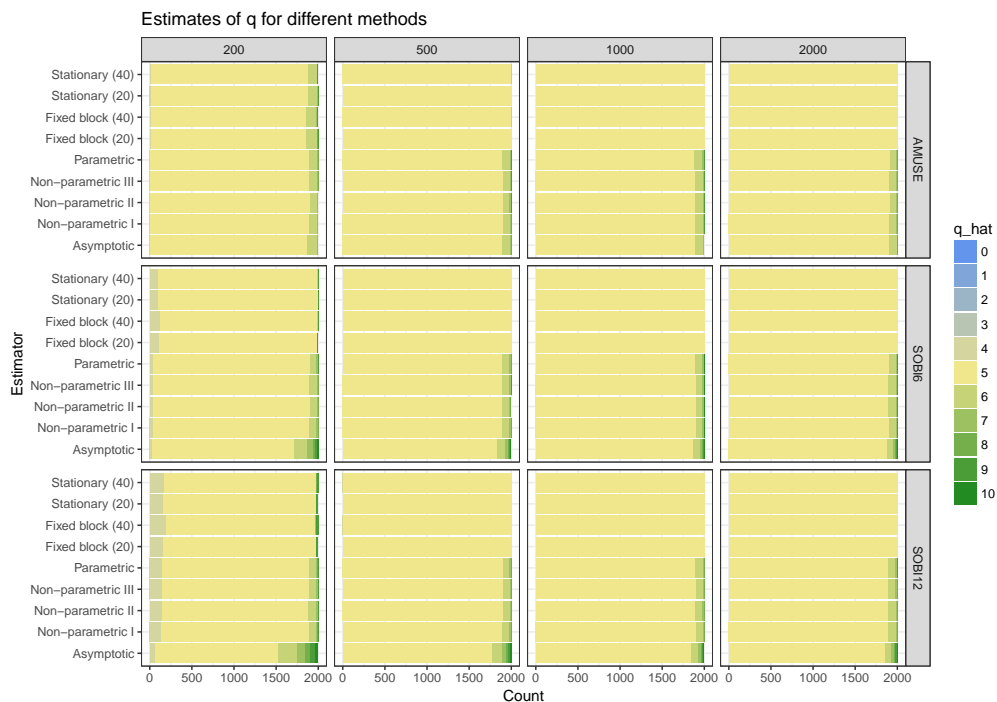
29

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

In Setting D1, the asymptotic test seems not to work as well as the other methods for small samples do. However, in general, the difference between this and the bootstrap-based testing procedures is negligible. Furthermore, the ladle is the most preferable option. In setting D2, on the other hand, the ladle consistently underestimates the signal dimension, and is not able to find the weak signal. When using the hypothesis testing-based methods, the weak signal is identified with increasing sample size. However, the more scatter matrices we estimate, the more difficult the estimation becomes and,
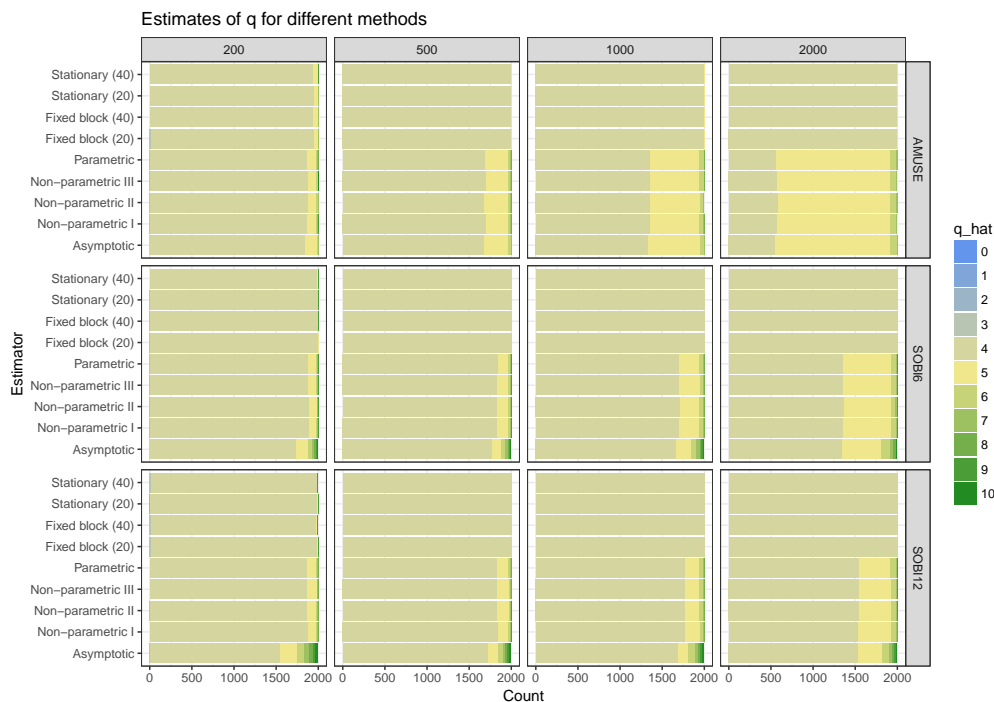
30

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

thus, AMUSE works best.

In Setting D3, the ladle fails completely, and keeps getting worse with increasing sample size. The difference between the bootstrapping and asymptotic tests is at its largest in this setting, and the asymptotic test seems to be the most preferable option. Because two lags are needed to capture all temporal information, AMUSE is at a disadvantage in this setting, as is clearly visible in the plots. In addition, SOBI6 seems to exploit the lag information better than SOBI12 does, possibly because it avoids the inclusion

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

of several unnecessary autocovariance matrices in the estimation.

5. Data example

For our real-data example, we use the recordings of three sounds signals available in the R-package JADE; these data have been analyzed by, among others, Miettinen et al. (2017). To the three signal components, we added 17 white-noise components, all with t_5-distributions, to determine whether the methods work for non-Gaussian white noise as well. After standard-

izing the 20 components to have unit variances, we used a random square matrix in which each element comes from the uniform distribution on $[0, 1]$. The original signals have a length of 50000; for convenience, we selected the first 10000 instances. The 20 mixed components are visualized in the Supplementary Material, but reveal no clear structure.

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

However, the computation times of the nine methods varied significantly; see Table 10.

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

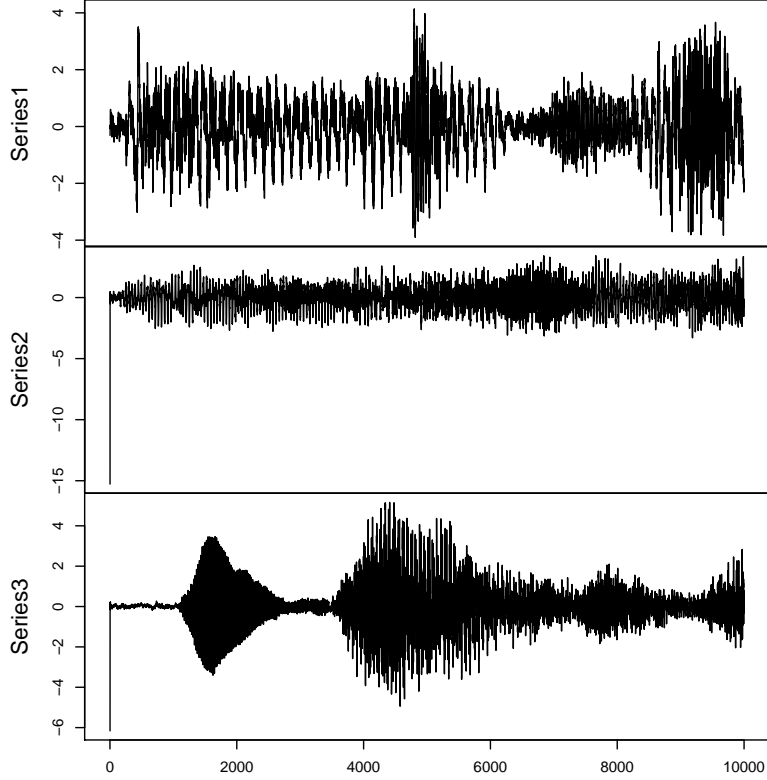| Asymptotic tests			Bootstrap tests			Ladle estimator		
AMUSE	SOBI6	SOBI12	AMUSE	SOBI6	SOBI12	AMUSE	SOBI6	SOBI12
0.07	0.19	0.49	15.08	47.24	88.08	2.75	9.85	18.17

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

Because the approaches all estimated the dimension correctly, those based on the asymptotic test are clearly favourable, owing to their computational speed. However, note that the ladle and bootstrap tests can be parallelized; in the current comparison, we used only a single-core computer. The ladle estimator is naturally faster to compute than the divide-and-conquer hypothesis-testing strategy is. This is because the latter needs to create separate bootstrap samples for each hypothesis to be tested, whereas

the ladle estimator conducts bootstrap sampling only once.

6. Discussion

We have 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 that we estimate any parameters, and makes relatively weak assumptions. Furthermore, its existing competitors use computer-intensive resampling techniques. For all these reasons, the asymptotic test is a very attractive option in practice. This conclusion is supported by the results of our simulations and the real-data example, in which we estimate the dimension of sound recording data.

A drawback of the proposed method is its inability to distinguish between non-autocorrelated signals (e.g., stochastic volatility models) and white noise. One way to get around this limitation is to replace $\mathbf{z}_{2t}$ in (1.1) with a vector of stochastic volatility series, and then reverse the roles of the signal and the noise. That is, we use hypothesis testing to estimate the dimension of the "noise" subspace (containing the stochastic volatility components), and then separate them from the uninteresting "signal" series that exhibit autocorrelation. For this to work, a limiting result equivalent to Lemma 4 is needed for the above combination model. A similar idea was

suggested in the context of the ladle estimator by Nordhausen and Virta (2018).

In addition to (1.1), noise can be incorporated in a time series model by using an additive noise model for $\mathbf{x}_t \in \mathbb{R}^p$, as follows:

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

where $\mathbf{\Omega} \in \mathbb{R}^{p \times q}$ is a nonsquare 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 (2019) who use the PCA-transformation to estimate the $p - k$ principal components with smallest variances. Hypothesis testing can then be used to pinpoint the correct dimension $k = q$ for which these components consist of pure noise. It seems likely to us that a version of this idea could be used in the context of time series data and model (6.4), and 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, not the signals themselves, in contrast to the model proposed here.

Supplementary Material

The online Supplementary Material contains the proofs of all theoretical results, additional simulation plots, and additional simulations.

Acknowledgments

The authors would like to express their gratitude to the anonymous referees for their helpful comments and suggestions.

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.

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-

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-*

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 (2019). Estimating the number of signals using principal component analysis. *Stat 8*(1), e231.

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