AUTHOR	Klaus Nordhausen; Sara Taskinen; Joni Virta
TITLE	Signal dimension estimation in BSS models with serial dependence
YEAR	2022
DOI	10.1109/ICECCME55909.2022.9988152
VERSION	Author's accepted manuscript
CITATION	K. Nordhausen, S. Taskinen and J. Virta, "Signal dimension estimation in BSS models with serial dependence," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, 2022, pp. 1-7, doi: 10.1109/ICECCME55909.2022.9988152.

Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)
16-18 November 2022, Maldives

Signal dimension estimation in BSS models with serial dependence

Klaus Nordhausen
Department of Mathematics and Statistics
University of Jyväskylä
Jyväskylä, Finland
ORCID: 0000-0002-3758-8501

Sara Taskinen
Department of Mathematics and Statistics
University of Jyväskylä
Jyväskylä, Finland
ORCID: 0000-0001-9470-7258

Joni Virta
Department of Mathematics and Statistics
University of Turku
Turku, Finland
ORCID: 0000-0002-2150-2769

***Abstract*—Many modern multivariate time series datasets contain a large amount of noise, and the first step of the data analysis is to separate the noise channels from the signals of interest. A crucial part of this dimension reduction is determining the number of signals. In this paper we approach this problem by considering a noisy latent variable time series model which comprises many popular blind source separation models. We propose a general framework for the estimation of the signal dimension that is based on testing for sub-sphericity and give examples of different tests suitable for time series settings. In the inference we rely on bootstrap null distributions. Several simulation studies are used to demonstrate the performances of the tests in different time series settings.**

***Index Terms*—Dimension reduction, nonstationary source separation, second order source separation, sub-sphericity, block bootstrap.**

I. INTRODUCTION

Nowadays numerous fields of applied science (e.g. engineering, economics and medicine) collect multivariate time series data. As data collection techniques and measuring devices have become more and more complex, a large amount of variables are often measured, resulting into very high-dimensional data. This poses problems for the data analysis as many multivariate methods become computationally impractical when the data dimension is large.

When high-dimensional data are collected, it is increasingly common that datasets include a large amount of noise and redundancy. The first step of the analysis is thus to separate the signals of interest from the noise. In case of independent and identically distributed (iid) data, the most common method for dimension reduction is principal component analysis (PCA) where the key question is to decide how many components should be retained for further analyses. Many existing methods for determining the number of important components rely on visual inspection of the magnitudes of the eigenvalues of the covariance matrix (see [1] and references therein), on bootstrap variability of eigenvectors [2] or on the combination of the two [3]. Inferential tools based on eigenvalues as well as information theoretic criteria in models assuming Gaussian signal and noise were considered for example in [4]–[6]. Recently, interest on inferential eigenvalue-based tools resurfaced under the assumption of non-Gaussian signals. In [7], [8] tests for the equality of the smallest eigenvalues of various matrices were proposed. Later, [9] used similar approach when testing for the number of non-Gaussian components in an independent component analysis setting. None of these approaches considered models with external (additive) noise components.

In [10], asymptotic and bootstrap tests for signal dimension in the case of general iid noisy latent variable model were proposed by extending the probabilistic PCA of [11]. It was shown that due to the model structure and properties of the covariance matrix, a test for the signal dimension can again be formulated as testing for the equality of the smallest eigenvalues. An estimate for the signal dimension was obtained by performing successive hypothesis tests. We review the model as well as the signal dimension estimation procedure of [10] in detail in Section II. The aim of this study is to generalize the resampling-based method of [10] to allow also serial dependence in such a way that many popular blind source separation (BSS) models such as second order separation (SOS), nonstationary source separation (NSS) model or stochastic volatility (SV) separation models are included. For an overview of these models see for example [12], [13] and note that usually the assumption of Gaussian signals is avoided in these contexts. Notice that similar approaches have earlier been suggested in the context of internal noise model (where the noise is included as latent variables) and SOS in [14], [15]. To achieve our goal, we propose a general framework for the testing and estimation of the signal dimension. That is, we allow also other metrics besides variance, such as autocovariance, to be used when separating the signals of interest from noise. The noisy latent time series model, justification for the proposed method and an algorithm for performing the bootstrap test for

The work of KN was supported by Austrian Science Fund (FWF) Grant P31881-N32. The work of JV was supported by Academy of Finland, Grant 335077.

978-1-6654-7095-7/22/$31.00

the signal dimension are described in Section III. In Section IV we compare the performances of the suggested tests and their capabilities in estimating the correct signal dimension in several different time series settings. The paper is concluded with some discussion in Section VI.

Thus, to conclude with a summary of the contributions of this paper:

- We consider a signal dimension testing and estimation method in a BSS framework with serial dependence without making strong distributional assumptions on the signals.
- Our method can be based on an arbitrary matrix $\mathbf{S_K}$ where the choice of the kernel $\mathbf{K}$ has an impact on the performance.
- The performance of our methods is evaluated in an extensive simulation study and in an example. Our methods are also compared with competing methods such as Akaike's information criterion (AIC) and minimum description length (MDL) of [5], [6] for which so far only little empirical evidence seems available and which were developed under much stronger assumptions.

II. IID METHOD

To motivate our contribution, we begin by reviewing the analogous procedure for iid data sources, as presented in [10]. That is, assume that the rows of $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ form a sequence of p-dimensional random vectors generated as,

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ specifies the location vector, $\mathbf{A} \in \mathbb{R}^{p\times d}$ is an unknown, deterministic matrix such that $\operatorname{rank}(\mathbf{A}) = d \leq p-2$, $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a sequence of zero-mean iid latent d-dimensional random vectors and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ is a random sample from $N_p(\mathbf{0}_p, \sigma^2\mathbf{I})$, for some unknown $\sigma^2 > 0$. Furthermore, assume that the errors $\boldsymbol{\varepsilon}_i$ are independent of the latent random vectors $\mathbf{z}_i$, that the fourth moment of $\mathbf{z}_1$ is finite and that $\mathbf{z}_1$ is not concentrated on a hyperplane in $\mathbb{R}^d$.

In order to estimate the signal dimension d, the authors of [10] suggested basing the inference on the final $r := p - d$ eigenvalues $\rho_{d+1}, \ldots, \rho_p$ of the sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}'\mathbf{H}\mathbf{X}$, where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$ and $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^n$.

To see why this is a sound strategy, consider the population version of the model (1) with the sample covariance matrix replaced by the covariance matrix,

$$\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{x}\mathbf{x}') - \mathbb{E}(\mathbf{x})\mathbb{E}(\mathbf{x}') = \mathbf{A}\mathbb{E}(\mathbf{z}\mathbf{z}')\mathbf{A}' + \sigma^2\mathbf{I}, \tag{2}$$

where the decomposition uses the additivity property of the covariance matrix. Hence, under our assumptions, the ordered eigenvalues of $\boldsymbol{\Sigma}$ are

$$\lambda_1 + \sigma^2 \geq \cdots \geq \lambda_d + \sigma^2 > \sigma^2 = \cdots = \sigma^2, \tag{3}$$

where $\lambda_1 \geq \cdots \geq \lambda_d$ are the ordered d non-zero eigenvalues of the rank-d matrix $\mathbf{A}\mathbb{E}(\mathbf{z}\mathbf{z}')\mathbf{A}'$. Hence, for the model (1) with the signal dimension d and "pure noise dimension" r, we expect the final r eigenvalues of $\mathbf{S}$ to be close to each other (and well-separated from the remaining eigenvalues), for sufficiently large n. In particular, the sample variance of the final r eigenvalues should be small.

Indeed, denoting by m_{1k}, m_{2k} and $s_k^2 = m_{2k} - m_{1k}^2$, respectively, the first and second sample moment and the (biased) sample variance of the final $p-k$ eigenvalues $\rho_{k+1}, \ldots, \rho_p$ of $\mathbf{S}$, [10] showed that the suitably scaled sample variance, $T_k := \frac{n(p-k)s_k^2}{2m_{1k}^2}$, has a chi-squared limiting distribution for $k = d$,

$$T_d \rightsquigarrow \chi^2_{\frac{1}{2}(r+2)(r-1)}. \tag{4}$$

Further, for any $k < d$, the quantity T_k/n converges to a positive constant in probability (this is clear since then at least one of the eigenvalues making up s_k^2 differs from the others in the limit, yielding an asymptotically positive sample variance). Whereas, for any $k > d$, the quantity T_k is bounded in probability (this is heuristically clear from (4) as, if d of the eigenvalues are asymptotically equal then so are $d-1$ of them, etc.).

The above observations lead to an estimator for d as follows. Let $H_{0,k}$ be the null hypothesis that the true signal dimension is $d = k$. Then, we test the sequence of hypotheses $H_{0,0}, \ldots, H_{0,p-2}$ using, respectively, the test statistics $T_0, \ldots, T_{p-2}$ and take the true dimension to be the smallest k for which $H_{0,k}$ is not rejected. In practice, the testing is done by selecting an appropriate significance level and using the chi-squared limiting distribution in (4) as the null distribution. However, an asymptotic power of one can be obtained with a suitable sequence of critical values, see [8]. Note that as the test statistic is based on the variance of the eigenvalues, at least two noise components are assumed which does not differ so much from the other approaches which usually assume at least one noise component. We would like to emphasize that in model (1) the distribution of the signals is not specified and only moment assumptions are made. The model is well investigated in case of an assumption of Gaussianity and then a test for $H_{0,k}$ is discussed for example in [4]. Further, [5], [6] suggested iid normal likelihood-based information theoretic criteria (ITC) of the form

$$\mathrm{ITC}(k) = -a\log\left(\frac{\prod_{i=k+1}^{p}\rho_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p}\rho_i}\right)^{(p-k)n} + b\ k(2p-k), \tag{5}$$

which yields the Akaike information criterion $\mathrm{AIC}(k)$ with $a = b = 2$ and the minimum descriptive length criterion $\mathrm{MDL}(k)$ with $a = 1$ and $b = 0.5\log n$. In both cases the estimate for the signal dimension minimizes the criterion over $k = 0, \ldots, p-1$. [5] argue that AIC will asymptotically overestimate the dimension while MDL yields a consistent estimate. However, we are not aware of any simulation studies verifying this claim. Note that [6] relax the normality assumption of noise and signal by showing that the ITC from above hold also when $\mathbf{x}$ in (1) has an elliptical distribution. In this case it is, however, not at all clear what could be the signal

and noise distributions and how independence between them could hold outside the Gaussian model.

III. GENERAL METHOD

We next generalize the method described in Section II beyond iid data generating processes. Namely, assume that the rows of $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ are again a sequence of p-dimensional random vectors generated as,

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}_i + \boldsymbol{\varepsilon}_i, \tag{6}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{A} \in \mathbb{R}^{p \times d}$ is an unknown, deterministic matrix such that $\operatorname{rank}(\mathbf{A}) = d \leq p - 2$, $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is an ordered sequence of zero-mean, *not necessarily independent and identically distributed* latent d-dimensional random vectors and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ is a random sample from $N_p(\mathbf{0}_p, \sigma^2\mathbf{I})$ for some unknown $\sigma^2 > 0$. Furthermore, assume again that the errors $\boldsymbol{\varepsilon}_i$ are independent of the latent random vectors $\mathbf{z}_i$.

The procedure in Section II is essentially a test for sphericity applied in the noisy low-rank model (1). There are three critical aspects that make it work. (i) The covariance matrix decomposes as in (2), leading into the eigenvalue structure (3) where the r pure noise components are separated from the signal. (ii) The signal $\mathbf{z}_i$ is truly d-dimensional, making the noise not just separated but well-separated from the signal in (3), thus allowing us to distinguish the two on the basis of variance. (iii) The noise is assumed spherical Gaussian, giving rise to the chi-squared distribution in (4).

Having moved outside of iid data in (6), it is natural to relax also the property (ii) above and allow the separation of the signal and the noise to be based on some other metric besides variance. For example, if the signal $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is an equally spaced multivariate time series with suitable autocovariances, it is natural to differentiate between the signal and the noise on the basis of autocovariance, which the latter does not exhibit at all, being white noise. To facilitate this, it is necessary to change the covariance matrix to an autocovariance matrix, and we are then faced with the question whether the property (i), a combination of additivity and affine equivariance, still holds for autocovariance matrices. Conveniently, it does, a fact that follows from their functional similarity to the covariance matrix. Finally, the tractability of any limiting distributions is unnecessary in the current framework where the estimation is based on bootstrapping techniques and, hence, the property (iii) is strictly not needed. However, we still make the assumption of Gaussian noise both for simplicity and for its ubiquitousness in the signal processing literature and later, in Section VI, discuss the implications of relaxing it.

Based on the above, we propose the following framework for estimating the unknown signal dimension d in the model (6). Let $\mathbf{K} = \mathbf{K}_n = (k_{ij}) \in \mathbb{R}^{n \times n}$ be a fixed sequence of symmetric matrices and define the sample covariance matrix of $\mathbf{X}$ with respect to the *kernel* $\mathbf{K}$ as

$$\mathbf{S}_\mathbf{K} = \frac{1}{n}\mathbf{X}'\mathbf{K}\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}\mathbf{x}_i\mathbf{x}_j'. \tag{7}$$

As a weighted sum of the second-order terms $\mathbf{x}_i\mathbf{x}_j'$, the matrix $\mathbf{S}_\mathbf{K}$ retains the functional form of the covariance matrix (meaning that if a population counterpart for $\mathbf{S}_\mathbf{K}$ exists, we can expect the property (i) to hold), while simultaneously allowing capturing a wide range of different structures. For example, $\mathbf{K} = \mathbf{H}$ yields the covariance matrix as in the original method of Section II and the choice $\mathbf{K} = \mathbf{H}\mathbf{K}_1\mathbf{H}$, where

$$\mathbf{K}_1 = \begin{pmatrix} 0 & 0.5 & 0 & \cdots & \cdots & \cdots & 0 \\ 0.5 & 0 & 0.5 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0.5 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0.5 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0.5 & 0 & 0.5 \\ 0 & \cdots & \cdots & \cdots & 0 & 0.5 & 0 \end{pmatrix} \tag{8}$$

yields the centered symmetrized autocovariance matrix on lag 1. Autocovariance matrices with other lags are obtained by moving the two bands in (8) further from the diagonal.

Inference on the signal dimension d can now be based on the r eigenvalues of $\mathbf{S}_\mathbf{K}$ corresponding to the pure noise components. Note that, for general $\mathbf{K}$, these are not necessarily the final r eigenvalues. In the case of symmetrized autocovariance matrices, for example, the signals might exhibit both positive and negative autocovariances, implying that the eigenvalues of interest (the zeroes corresponding to the white noise) are in neither end of the spectrum. Hence, we select the noise eigenvalues as those r eigenvalues which have the smallest variance out of all subsets of r adjacent eigenvalues. Note that in some special cases we might know what the noise eigenvalues are (for example, they are zeroes when using $\mathbf{S}_{\mathbf{H}\mathbf{K}_1\mathbf{H}}$), and this information could also be exploited when selecting the eigenvalues, see Section IV.

The above proposal makes several implicit assumptions. First, in order for it to be able to separate the signal from the noise, all signal components must be distinguishable from white noise with respect to the "information" measured by $\mathbf{S}_\mathbf{K}$ (for example, all signals must exhibit non-zero autocovariances in case of $\mathbf{K} = \mathbf{H}\mathbf{K}_1\mathbf{H}$, or, the signals must not be concentrated on a hyperplane in case of $\mathbf{K} = \mathbf{H}$). Second, for the selection of the eigenvalues through the minimal variance subset to work, the noise must form the largest eigenspace of $\mathbf{S}_\mathbf{K}$ (for example, in case of $\mathbf{K} = \mathbf{H}\mathbf{K}_1\mathbf{H}$, having five signals with equal non-zero autocovariance on lag 1 along with only four pure noise components would not be allowed as then the former would be thought of as the noise). Thirdly, the kernel matrix $\mathbf{K}$ must be able to capture enough information on the dimensions in the first place to allow the whole procedure. See Section VI for more discussion on this point.

Assuming that the above hold, we compute, for a given candidate dimension k, the test statistic

$$T_k := s_k^2, \tag{9}$$

where s_k^2 denotes the (biased) sample variance of the $p - k$ eigenvalues of $\mathbf{S}_\mathbf{K}$ with the smallest sample variance out of all

subsets of $p-k$ adjacent eigenvalues. As in Section II, we expect T_k to be large when $k < d$ and small when $k \geq d$, again allowing the pin-pointing of the true signal dimension d. To quantify "small" and "large", we use the bootstrapping Algorithm 1 to obtain the null distribution of T_k under $H_{0,k}$.

Algorithm 1 Algorithm for testing $H_{0,k} : d = k$.

Set the proposed dimension k and $r = p - k$;
Set the number of bootstrap samples M and the block size l;
Choose a kernel $\mathbf{K}$;
Starting with the sample $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, compute $\mathbf{S_K}$ together with its eigendecomposition $\mathbf{S_K} = \mathbf{UDU}'$;
Order the eigenvalues in $\mathbf{D}$ such that the variance of the last $p-k$ eigenvalues in $\mathbf{D}$ is minimal and derive the corresponding partitioning of $\mathbf{U} = (\mathbf{U}_1\ \mathbf{U}_2)$;
Compute the test statistic
$T_k = \frac{1}{r}\sum_{i=k+1}^{p}\left(\rho_i - \frac{1}{r}\sum_{j=k+1}^{p}\rho_j\right)^2$ as the sample variance of the final $p-k$ eigenvalues in $\mathbf{D}$;
Calculate $\mathbf{Y}_1 = \mathbf{XU}_1$, $\mathbf{Y}_2 = \mathbf{XU}_2$ and $\mathbf{S}_2$, the sample covariance matrix of $\mathbf{Y}_2$; and the mean of $\mathbf{Y}_2$, $\bar{\mathbf{Y}}_2$;
Estimate the noise variance as $\hat{\sigma}^2 = \frac{1}{r}\mathrm{tr}(\mathbf{S}_2)$;
for $j \in \{1, \ldots, M\}$ **do**
  Block bootstrap $\mathbf{Y}_1$ with block length l to obtain $\mathbf{Y}_1^*$.
  Simulate n observations from $N_r(\bar{\mathbf{Y}}_2, \hat{\sigma}^2\mathbf{I})$ to obtain $\mathbf{Y}_2^*$.
  Compute $\mathbf{X}^* = (\mathbf{Y}_1^*\ \mathbf{Y}_2^*)\mathbf{U}'$;
  Compute $T_{j,k}^*$ as T_k but based on $\mathbf{X}^*$;
Return bootstrap p-value: $p_k = [\#(T_{j,k}^* \geq T_k) + 1]/(M+1)$

The algorithm divides, under the assumed dimension k, the time series into a part containing a mixture of signal and noise and a part containing only noise. As the signal is assumed to be ordered and not necessarily iid, overlapping block bootstrap with fixed block length l, as described, for example, in [16], is used. For the pure noise part standard parametric bootstrap is used where the noise variance is estimated from the hypothetical noise part as its average sample variance. The two bootstrap samples are then combined and backtransformed in order to recompute the test statistic. Having obtained M bootstrap test statistic values, the p-value can be computed by comparing how extreme the test statistic of the observed sample is relative to the M bootstrap statistics which were generated under the null hypothesis. To conclude this section we would like to point out that [5], [6] argue that in a framework of Gaussian signals and noise the information theoretic criteria introduced in Section II can be used also when the signals are stationary and ergodic.

IV. SIMULATIONS

To evaluate the performance of our suggested method, we performed a simulation study using R [17] and the package rugarch [18]. To be as general as possible, we first considered in model (6) four different BSS settings for the latent stochastic processes $\mathbf{z}_i$, $i = 1, \ldots, n$, always assuming that $d = 3$ and that z_1, z_2 and z_3 are independent:

- SOS: Three different autoregressive AR(1) processes where z_1 has Exp(1) distributed innovations, z_2 has t_5 distributed innovations and z_3 has χ^2_3 distributed innovations.
- BS: Three block stationary processes where blocks are independent and have moving average (MA) or AR processes of different order with different variances where the innovation distributions are as in the previous setting.
- NSS: Three different MA processes with Exp(1) innovations whose variances change smoothly over time.
- SV: Three different autoregressive conditional heteroskedastic ARCH(1) processes with Gaussian innovations.

In each setting, all innovations were standardized so that $E(z_i) = 0$ and $Var(z_i) = 1$, $i = 1, 2, 3$, and the components were further scaled to have unit variance. The setting **SOS** has now three basic second order stationary processes. The setting **BS** differs from **SOS** in that each component consists of independent blocks which are itself stationary, giving a so-called block stationary model. The setting **NSS** is a nonstationary model where the mean stays constant but the variances of the processes change smoothly over time. The last setting is a stochastic volatility model especially popular in financial context. Popular BSS methods under the four settings are, for example, AMUSE and SOBI [19], [20] for **SOS**, NSS-SD and NSS-JD [21] for **NSS**, NSS-JD-TD [22] for **BS**, and gFOBI, gJADE, vSOBI and gSOBI [23]–[25] for **SV**.

Finally, the $p \times 3$ mixing matrix $\mathbf{A}$ was for each simulated data populated with random $N(0, 1)$ elements.

A. Evaluation of test performance

To evaluate the test performance, we set $p = 6$ and tested the null hypotheses $H_{0,2}$ $H_{0,3}$ (true) and $H_{0,4}$ with the α-level 0.05 for samples of sizes $n = 500, 1000, 2000$. As methods to be compared we considered the asymptotic test of [10] with the iid assumption, along with the new bootstrap tests in Algorithm 1 based on the covariance matrix ($\mathbf{K} = \mathbf{H}$) and the symmetrized autocovariance matrix with lag 1 ($\mathbf{K} = \mathbf{HK}_1\mathbf{H}$), each with block sizes $l = 10, 20, 40$. The number of bootstrap samples was $M = 200$ for all bootstrap-based methods and a range of noise levels σ was used. In addition, we also computed the autocovariance tests by exploiting the knowledge that the noise corresponds to zero eigenvalues (ACOV0).

Figure 1 gives the proportion of rejections based on 2000 repetitions for the asymptotic test (COV), the three bootstrap tests based on the covariance matrix (COV (l=10), COV (l=20) and COV (l=40)) and the three bootstrap tests based on the symmetrized autocovariance matrix (ACOV (l=10), ACOV (l=20) and ACOV (l=40)) for $n = 1000$. The results show that when using COV there are no differences between the performances of the asymptotic test and the bootstrap test (irrespective of the block size). Similarly, the number of bootstrap samples has no impact on the ACOV test. Surprisingly, the performance of the asymptotic test (which assumes iid data) is not hampered by the dependency structures present

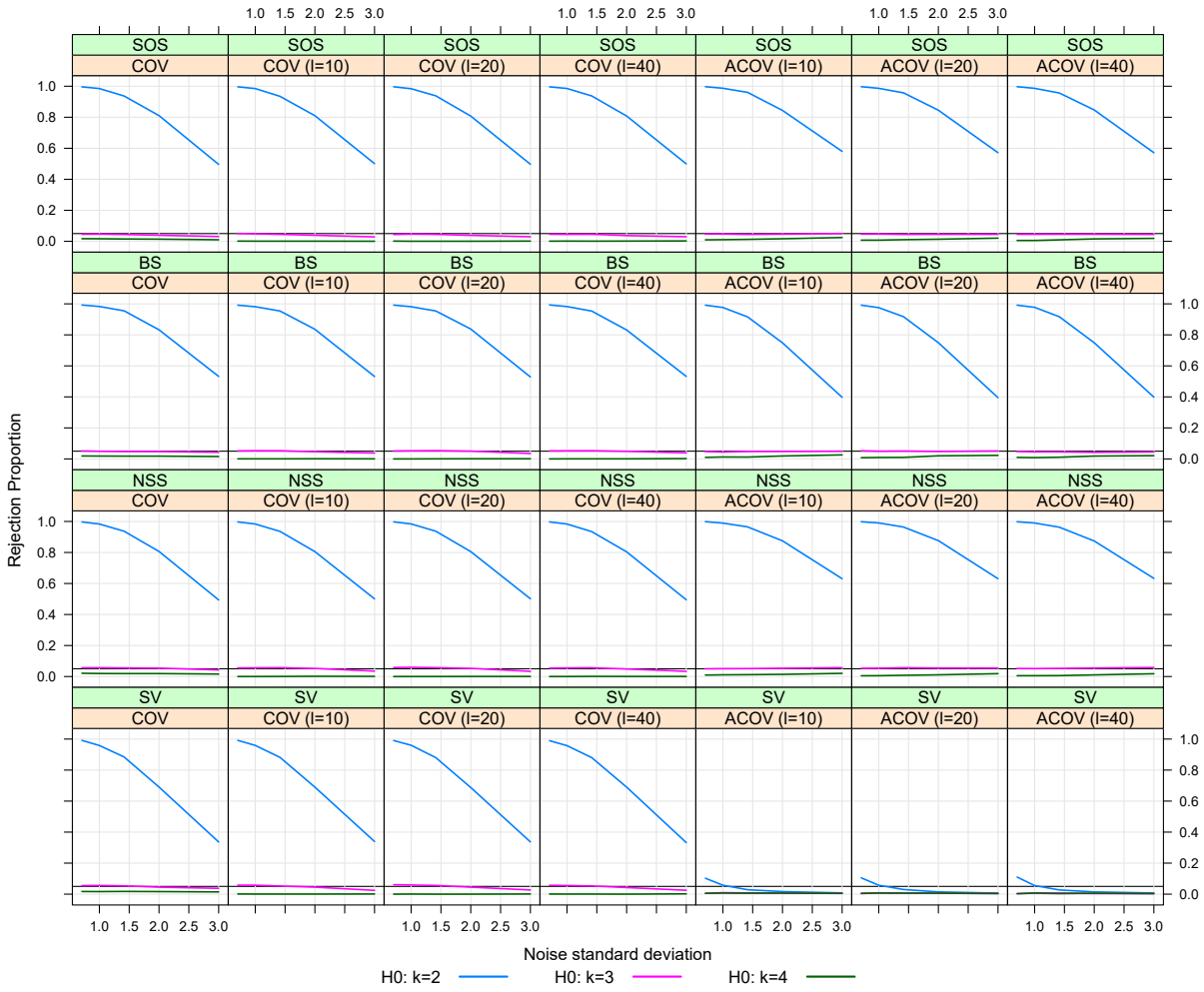

Fig. 1. Proportions of rejections in the four settings when $n = 1000$. The black horizontal line indicates the α-level $= 0.05$.

in the data. In the **SOS**, **BS** and **NSS** settings, all tests keep the null level and the hypothesis $H_{0,4}$ is rejected less often than $H_{0,3}$ as the corresponding test statistic uses a "tighter" subset of eigenvalues, see Section II. The power of the tests is also quite good and naturally decreases with increasing noise variance. As expected, in the setting **SV**, the covariance matrix based tests work nicely, whereas the autocovariance matrix based tests fail completely as there are no second order correlations present in the data. The power of the tests based on the autocovariance matrix is however better in the SOS and NSS setting as compared to the COV based tests. The results for $n = 500$ and $n = 2000$ are omitted as the only visible difference to $n = 1000$ is the power of the test which increases with sample size. Similarly, the results for the bootstrap test which exploits the knowledge of the value of the noise eigenvalue are not shown as the difference to the version shown in Figure 1 is minimal. Finally, we also repeated the experiment on the α-level 0.10. The conclusions were exactly analogous to the main study and, hence, these results are not shown here.

B. *Evaluation of estimation performance*

Of more practical relevance is how well we can actually estimate the signal dimension using successive testing. There are many strategies available and we consider divide-and-conquer to find a pair of consecutive hypotheses $H_{0,k}$ from the range $k = 0, \ldots, p-2$, where a rejection changes to a non-rejection on the level $\alpha = 0.05$. As competing methods we estimate the signal dimension using AIC and MDL as described in Section II. We consider in total 5 signal settings where the BS, NSS and SV settings are as in the previous section, but for the SOS setting we double the number of signals by adding three more signals with the same dependence structure but having Gaussian innovations. In the SOS setting we furthermore consider two different mixing matrix settings. In setting SOS A the mixing matrix is again filled with random elements from N(0,1) while in SOS O the mixing matrix has orthonormal columns meaning that all non-zero eigenvalues of the covariance matrix of $\mathbf{Az}$ are equal. In all 5 settings the data dimension is increased to $p = 10$ and we use only the bootstrap tests with block size $l = 20$ along with the iid asymptotic test.

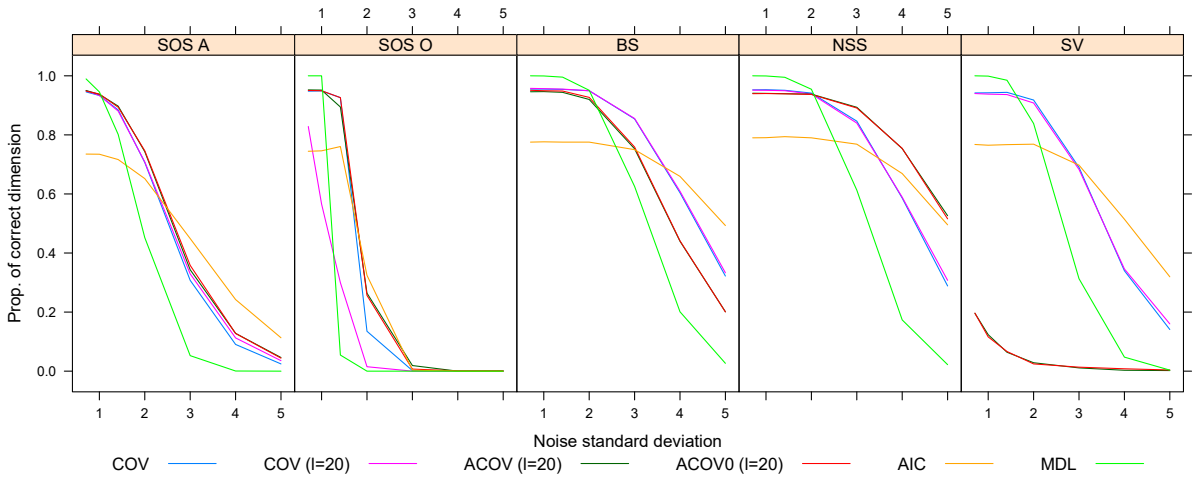

Fig. 2. Proportions of correct estimates of the signal dimension in 5 different settings based on 2000 repetitions.

Figure 2 shows then the proportions of correctly estimated dimension $d = 3$ based on 2000 repetitions with $M = 200$.

The results show that the testing based methods have a good signal-to-noise tolerance, if the mixing matrix is not orthogonal and therefore also provides information, in all settings under which they are expected to work. Whether the covariance or autocovariance based tests are better depends again on the setting. Finally, Figure 2 also shows the results for the autocovariance matrix based bootstrap test which uses the knowledge of the true noise eigenvalue to select the eigenvalues (ACOV0). When comparing the results to those of ACOV, where the eigenvalues are chosen through minimal variance, we see that the knowledge brings no benefit in the estimation. When comparing the test-based methods to the information theoretic criteria it seems that the test-based methods are more reliable. MDL is the best method for small noise but then starts rapidly to fail with increasing noise. AIC performs poorly in case of small noise but outperforms many of its competitors when the noise is large. Based on this simulation, however, it seems that the method which estimates the signal dimension based on repeated testing performs best, but which kernel to choose for $\mathbf{S_K}$ seems data dependent.

V. Example

For the final comparison of the methods we use three sound signals ($n = 50000$) which are freely available in the R package JADE and are for example used in [26]. For the mixing matrix $\mathbf{A}$ we fill a 20×3 matrix with random elements from $U(0, 1)$ and add then to the mixed signals Gaussian noise with different standard deviations. The non-zero signal eigenvalues for $\mathbf{S}(\mathbf{Az})$ are then 14.051, 2.391 and 0.639 and for $\mathbf{S_{HK_1H}}(\mathbf{AZ})$ 11.199, 1.920 and 0.508, respectively. The estimated signal dimensions for the different noise levels are shown in Figure 3 where the curves have been jittered so that the different methods are distinguishable.

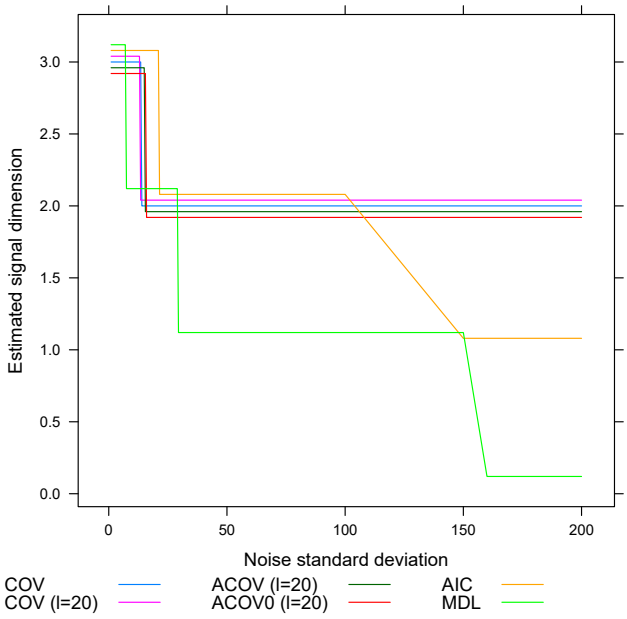

Fig. 3. Detected number signals in the sound example for different noise levels. Estimates are slightly jittered to avoid overlying lines.

For this example it seems again that MDL is the worst method and maybe ACOV the best method which is in line with the simulation results as speech signals are often analyzed using a NSS framework. For a noise standard deviation larger than 30 it seems that all methods lose one signal.

VI. Discussion and Conclusion

We proposed a flexible framework for estimating the signal dimension in a noisy BSS model, with the inference relying on bootstrapped null distributions. Simulations revealed that

the method provides accurate dimension estimates in a wide range of models and for different choices of the kernel $\mathbf{K}$.

The next step is to extend the asymptotic test of [10] to non-iid data by deriving the limiting distribution of the test statistic (9) under suitable conditions on the signal $\mathbf{z}_i$ and the kernel $\mathbf{K}$. Assuming that the limiting distribution is tractable (which we expect under the Gaussian noise), an instant benefit of such a result would be freedom from the computationally intensive bootstrapping in computing the null distributions. Moreover, from the performance of the asymptotic test based on the iid assumption in the simulations, it seems reasonable to expect the asymptotic null distribution of (9) to be accurate already for moderate sample sizes.

A related objective involves establishing suitable conditions for the kernel matrix $\mathbf{K}$ under which dimension estimation based on $\mathbf{S}_\mathbf{K}$ is feasible. Namely, it is simple to come up with non-trivial examples of matrices $\mathbf{K}$ which cannot be used to estimate the signal dimension. For example, the matrix $\mathbf{K} = \frac{1}{n}\mathbf{1}\mathbf{1}'$ yields $\mathbf{S}_\mathbf{K} = \bar{\mathbf{x}}\bar{\mathbf{x}}'$ which, while measuring a meaningful quantity, has rank one and carries no information on the separation of the signal from the noise. Conversely, having a full rank for $\mathbf{K}$ is not necessary either as can be seen by choosing $\mathbf{K}$ equal to $\mathbf{I}$ with the first diagonal element set to zero. This yields the (non-centered) covariance matrix of the data with the first observation removed and, as such, carries asymptotically the same information as the covariance matrix. From the simulations it is clear that different kernels $\mathbf{K}$ might be more suitable for some models than for the others. Therefore, in future research, the plan is to develop data-driven guidelines for choosing a correct kernel for data at hand. We also plan to investigate if the information from several kernels can be combined.

In [10], the Gaussian noise $\varepsilon_1, \ldots, \varepsilon_n$ was replaced with a random sample from an arbitrary spherical distribution [27] (with suitable moments), with minimal changes to the limiting result (4). However, the assumption of sphericity was made precisely for analytical tractability and, since we base our inference on bootstrapped null distributions, in the current context it is possible to venture still outside of spherical noise distributions. The only requirement is that the noise vectors have covariance matrix proportional to identity, but otherwise the bootstrapping-based method works for any noise distribution when in Algorithm 1 the parametric bootstrap for $\mathbf{Y}_2^*$ is replaced with an appropriate nonparametric alternative.

Finally, while the family of matrices $\mathbf{S}_\mathbf{K}$ is rather general, natural candidate matrices for dimension estimation not belonging to this set obviously exist. One example of such a matrix is the covariance matrix of fourth moments whose population version reads $\mathbf{\Sigma}_4 = \mathbb{E}\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\}$, where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are, respectively, the mean and the covariance matrix of $\mathbf{x}$. While $\mathbf{\Sigma}_4$ does not fit in our framework, some alternative techniques could possibly be used to show that it (along with various other matrices) can also be used for dimension estimation. For an example of using $\mathbf{\Sigma}_4$ for dimension estimation in the context of a specific BSS model, non-Gaussian subspace analysis, see [9].

REFERENCES

[1] I. Jolliffe, *Principal component analysis*, Springer, 2002.

[2] Z. Ye and R.E. Weiss, "Using the bootstrap to select one of a new class of dimension reduction methods," *Journal of the American Statistical Association*, vol. 98, pp. 968–979, 2003.

[3] W. Luo and B. Li, "Combining eigenvalues and variation of eigenvectors for order determination," *Biometrika*, vol. 103, pp. 875–887, 12 2016.

[4] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, pp. 128–136, 1956.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 387–392, 1985.

[6] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, pp. 1–25, 1986.

[7] J.R. Schott, "A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix," *Journal of Multivariate Analysis*, vol. 97, pp. 827–843, 2006.

[8] K. Nordhausen, H. Oja, and D.E. Tyler, "Asymptotic and bootstrap tests for subspace dimension," *Journal of Multivariate Analysis*, vol. 188, pp. 104830, 2022.

[9] K. Nordhausen, H. Oja, D.E. Tyler, and J. Virta, "Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace," *IEEE Signal Processing Letters*, vol. 24, pp. 887–891, 2017.

[10] J. Virta and K. Nordhausen, "Estimating the number of signals using principal component analysis," *Stat*, vol. 8, pp. e231, 2019.

[11] M.E. Tipping and C.M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society B*, vol. 61, pp. 611–622, 1999.

[12] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, Wiley & Sons, USA, 2002.

[13] Y. Pan, M. Matilainen, S. Taskinen, and K. Nordhausen, "A review of second-order blind identification methods," *Wires Computational Statistics*, p. e1550, 2021.

[14] M. Matilainen, K. Nordhausen, and J. Virta, "On the number of signals in multivariate time series," in *Latent Variable Analysis and Signal Separation*, Y. Deville, S. Gannot, R. Mason, M.D. Plumbley, and D. Ward, Eds. 2018, pp. 248–258, Springer.

[15] J. Virta and K. Nordhausen, "Determining the signal dimension in second order source separation," *Statistica Sinica*, vol. 31, pp. 135–156, 2021.

[16] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, 1997.

[17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[18] Ghalanos A., *rugarch: Univariate GARCH models*, 2020, R package version 1.4-2.

[19] L. Tong, V.C. Soon, Y.F. Huang, and R. Liu, "AMUSE: A new blind identification algorithm," in *IEEE International Symposium on Circuits and Systems*, 1990, pp. 1784–1787.

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

[21] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electronics Letters*, vol. 36, pp. 848–849, 2000.

[22] S. Choi and A. Cichocki, "Blind separation of nonstationary and temporally correlated sources from noisy mixtures," in *Neural Networks for Signal Processing X.*, 2000, pp. 405–414.

[23] M. Matilainen, K. Nordhausen, and H. Oja, "New independent component analysis tools for time series," *Statistics & Probability Letters*, vol. 105, pp. 80–87, 2015.

[24] M. Matilainen, J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen, "On independent component analysis and stochastic volatility models," *Austrian Journal of Statistics*, vol. 46, pp. 57–66, 2017.

[25] Jari Miettinen, Markus Matilainen, Klaus Nordhausen, and Sara Taskinen, "Extracting conditionally heteroskedastic components using independent component analysis," *Journal of Time Series Analysis*, vol. 41, pp. 293–311, 2020.

[26] Jari Miettinen, Klaus Nordhausen, and Sara Taskinen, "Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp," *Journal of Statistical Software*, vol. 76, pp. 1–31, 2017.

[27] K.-T. Fang, S. Kotz, and K.W. Ng, *Symmetric multivariate and related distributions*, Chapman & Hall, 1990.