

ORIGINAL ARTICLE

Estimating the number of signals using principal component analysis

Joni Virta¹  | Klaus Nordhausen² 

¹Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

²Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria

Correspondence

Klaus Nordhausen, Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstr. 7, A-1040 Vienna, Austria.

Email: klaus.nordhausen@tuwien.ac.at

Funding information

CRoNoS European Cooperation in Science and Technology, Grant/Award Number: IC1408

In this work, we develop inferential tools for determining the correct number of principal components under a general noisy latent variable model, which includes as a special case, for example, the noisy independent component model. The problem is approached using hypothesis testing, and we provide both a large-sample test and several resampling-based alternatives. Simulations and an application to sound data reveal that both types of approaches keep the desired levels and have good power.

KEYWORDS

bootstrap, covariance matrix, dimension reduction, independent component analysis, order determination

1 | INTRODUCTION

Many modern high-dimensional data sets are too large to be directly subjected to various methods of multivariate data analysis. This is especially common with independent component analysis (ICA; Comon & Jutten, 2010; Nordhausen & Oja, 2018), where methods often have computational complexities that grow too fast with the number of variables p to be of any practical use. For example, the popular affine equivariant joint approximate diagonalization of eigen-matrices (JADE) algorithm (Cardoso & Souloumiac, 1993; Miettinen, Taskinen, Nordhausen, & Oja, 2015) requires the joint diagonalization of p^2 fourth cumulant matrices of size $p \times p$ and has the computational complexity $\mathcal{O}(np^4 + p^5)$.

The standard workaround is to precede ICA with a preliminary dimension reduction step, commonly done using principal component analysis (PCA), having the relatively low complexity $\mathcal{O}(np^2 + p^3)$. To ensure that no essential information is lost in the PCA step, a large enough number of principal components has to be retained. However, for the preliminary dimension reduction to be of any use, we still want to discard as many components as possible (without losing information). The key question is then how many components should we exactly retain?

Although the use of the preliminary PCA step is very popular in applied data analysis, no inferential methods for the selection of a suitable dimension under a general noisy model can be found in the literature. Instead, the number of retained components is most often picked with heuristic methods such as the scree plot or Kaiser's rule (Jolliffe, 2002).

In this paper, we approach the problem by using hypothesis testing based on both asymptotic distributions and bootstrapping to infer the appropriate number of principal components. Our analysis is based on a general noisy latent variable model presented in Section 2 along with a suitable test statistic to test for the signal dimension. The asymptotic and bootstrapping null distributions of the test statistic are discussed in Sections 3 and 4, the former also introducing a consistent estimator of the latent dimension. As independent component analysis is perhaps the most popular method applied after a preliminary PCA, our examples in Section 5 are formulated in the context of ICA. However, the chosen model works just as well outside of ICA and can easily be adapted to other situations.

A similar analysis under a noiseless elliptical model was conducted in Nordhausen, Oja, and Tyler (2016), and the current work improves upon their results by discarding both of the aforementioned assumptions, noiselessness and ellipticity (although; Nordhausen et al., 2016, experimented

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Stat Published by John Wiley & Sons, Ltd.

also with a noisy model using simulations). Similar studies include also Luo and Li (2016) and Nordhausen, Oja, Tyler, and Virta (2017) which provide tools to test for the number of non-Gaussian components in an ICA model with internal noise. These techniques were extended in Matilainen, Nordhausen, and Virta (2018), Nordhausen and Virta (2018), and Virta and Nordhausen (2019) to the second-order source separation model to test for white noise.

2 | NOISY ICA MODEL

We assume that the observed random vector $\mathbf{x} \in \mathbb{R}^p$ obeys the noisy model,

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\epsilon}, \quad (1)$$

where \mathbf{A} is an unknown full rank $p \times d$ matrix, $\mathbf{z} \in \mathbb{R}^d$ is a latent signal vector with finite fourth moments, $d < p$, and $\boldsymbol{\epsilon} \in \mathbb{R}^p$ is a noise vector that has finite fourth moments and is independent of \mathbf{z} . The covariance matrix of \mathbf{z} is confounded with the matrix \mathbf{A} , and without loss of generality, we fix $\text{Cov}(\mathbf{z}) = \mathbf{I}_d$. To enable the derivation of our null distributions, we make the standard assumption that the noise vector is spherical (Fang, Kotz, & Ng, 1990) around the origin. In the context of ICA, the vector \mathbf{z} could be assumed to have independent components (or subvectors), but as no such assumptions are required in the PCA step, we refrain from making them here. Similarly, to enable the subsequent ICA step, the noise most likely has to be Gaussian to be annihilated by the standard cumulant-based ICA methods; but again, this assumption is not needed in the PCA step, and the following results are valid for any spherical noise. The noisy ICA model is discussed, for example, in Cichocki and Amari (2002) and Comon and Jutten (2010) where also many references are given.

Our model can also be seen as a generalization of a factor analysis model or the probabilistic PCA model (Tipping & Bishop, 1999). In both of these models, the random vectors \mathbf{z} and $\boldsymbol{\epsilon}$ are assumed to be Gaussian, and the signal dimension can be estimated, for example, using information criteria-based techniques; see Comon and Golub (1990), Liavas, Regalia, and Delmas (1999), Minka (2001), Cichocki and Amari (2002), and Comon and Jutten (2010) and the references therein. A quite different model is considered in Choi, Taylor, and Tibshirani (2017) where $\mathbf{A}\mathbf{z}$ is considered deterministic and only the Gaussian noise $\boldsymbol{\epsilon}$ is random. Karhunen, Cichocki, Kasprzak, and Pajunen (1997) argued then in the context of noisy blind source separation that due to a lack of more general methods, approaches that assume Gaussian signals and noise could prove useful as the observed mixture \mathbf{x} is heuristically more Gaussian than \mathbf{z} itself. In this paper, however, even though we have the noisy ICA model in mind, the presented methods hold for any general random signal vector \mathbf{z} and spherical noise $\boldsymbol{\epsilon}$. The only further assumptions we need are the moment conditions given after Equation 1.

Using the singular value decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$, $\mathbf{U} \in \mathbb{R}^{p \times d}$, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d \times d}$, $\mathbf{V} \in \mathbb{R}^{d \times d}$, it is easy to see that the singular values of the unknown parameter \mathbf{A} are identifiable as are its left singular vectors, possibly up to an orthogonal transformation. Namely, assume that a random vector has two representations of the form (1), that is, $\mathbf{A}\mathbf{z} + \boldsymbol{\epsilon} = \mathbf{A}_0\mathbf{z}_0 + \boldsymbol{\epsilon}_0$. Taking covariance matrices, we obtain

$$\mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T + \sigma^2\mathbf{I}_p = \mathbf{U}_0\boldsymbol{\Lambda}_0^2\mathbf{U}_0^T + \sigma_0^2\mathbf{I}_p. \quad (2)$$

As eigenvalues of a matrix are unique, we must have $\sigma_0^2 = \sigma^2$ and $\boldsymbol{\Lambda}_0 = \boldsymbol{\Lambda}$. Moreover, as eigenspaces related to distinct eigenvalues are unique, we must have $\mathbf{U}_0 = \mathbf{U}\mathbf{W}$ for some block-diagonal orthogonal matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$.

Our main objective is to estimate the dimension d , given an independent and identically distributed sample from the model. After obtaining the estimate, we can compute the eigenvectors $\mathbf{U}\mathbf{W}$ and project the observations to obtain a reduced d -dimensional vector where no information on the signal \mathbf{z}_i has been lost, $(\mathbf{U}\mathbf{W})^T\mathbf{x}_i = \mathbf{W}^T\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{z}_i + \boldsymbol{\epsilon}_i^*$, where $\boldsymbol{\epsilon}_i^* \in \mathbb{R}^d$ is again spherical (Fang et al., 1990).

In the following, we propose two different methods, asymptotic and bootstrap based, for estimating the dimension. Both are based on testing the set of null hypotheses

$$H_{0k} : d = k, \quad k = 0, \dots, p-1.$$

Denoting in the following the covariance matrix by $\boldsymbol{\Sigma}$, the form (2) reveals that the eigenvalues of $\boldsymbol{\Sigma}(\mathbf{x})$ are $(\lambda_1^2 + \sigma^2, \dots, \lambda_d^2 + \sigma^2, \sigma^2, \dots, \sigma^2)$ and estimating d is thus equivalent to estimating the index after which the eigenvalues stay constant. Thus, a logical choice for the test statistic \hat{t}_k for H_{0k} is the variance $\hat{t}_k = s_k^2(\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i))$ of the smallest $r = p - k$ eigenvalues of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i)$.

Under the model, $\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i)$ can alternatively be written as

$$\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i) = (\mathbf{U}|\mathbf{R})\hat{\boldsymbol{\Sigma}}((\mathbf{I}_d|\mathbf{0})^T\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{z}_i + \boldsymbol{\epsilon}_i^*)(\mathbf{U}|\mathbf{R})^T,$$

where $\mathbf{R} \in \mathbb{R}^{p \times (p-d)}$ is an arbitrary matrix with orthonormal columns such that $(\mathbf{U}|\mathbf{R})$ is a full $p \times p$ orthogonal matrix and $\boldsymbol{\epsilon}_i^* \sim \boldsymbol{\epsilon}_i$. Consequently, the eigenvalues of $\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i)$ and $\hat{\boldsymbol{\Sigma}}((\mathbf{I}_d|\mathbf{0})^T\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{z}_i + \boldsymbol{\epsilon}_i^*)$ are equal, and we may in Section 3 assume without loss of generality that $\mathbf{U} = (\mathbf{I}_d|\mathbf{0})^T$.

The next two sections propose different strategies for approximating the null distribution of \hat{t}_k .

Remark 1. A standard way of ‘‘robustifying’’ methods to be more tolerant to outliers is to replace $\boldsymbol{\Sigma}$ with a robust scatter matrix (Nordhausen & Tyler, 2015). However, our proposed method relies on the variance decomposition (2), which is only satisfied by scatter matrices that are *additive*. Virta (2018) showed that $\boldsymbol{\Sigma}$ is, under natural regularity conditions, the only additive scatter matrix, meaning that the use of other scatter matrices cannot be justified using our current approach.

3 | ASYMPTOTIC NULL DISTRIBUTION

The development very closely mimics that of Nordhausen et al. (2016). By Eaton and Tyler (1991), we have under H_{0k} that $n \cdot s_k^2(\hat{\Sigma}(\mathbf{x}_i)) = n \cdot s^2(\hat{\Sigma}_k(\mathbf{x}_i)) + \mathcal{O}_p(n^{-1/2})$, where $s^2(\hat{\Sigma}_k(\mathbf{x}_i))$ is the variance of the eigenvalues of $\hat{\Sigma}_k(\mathbf{x}_i)$, the lower right $r \times r$ block of $\hat{\Sigma}(\mathbf{x}_i)$. This block equals $\hat{\Sigma}(\boldsymbol{\epsilon}_{ki})$, where $\boldsymbol{\epsilon}_{ki}$ contains the last r components of the noise vectors $\boldsymbol{\epsilon}_i^*$. By the central limit theorem, $\sqrt{n} \cdot \text{vec}(\hat{\Sigma}(\boldsymbol{\epsilon}_{ki}) - \sigma^2 \mathbf{I}_r)$ has a limiting multivariate normal distribution, and following the proof of Nordhausen et al., 2016 (Theorem 1), we have under H_{0k} that

$$\frac{nr \cdot s_k^2(\hat{\Sigma}(\mathbf{x}_i))}{2\sigma^4 \beta} \rightsquigarrow \chi_{\frac{1}{2}(r-1)(r+2)}^2$$

for a specific constant β . The limiting distribution stays the same if σ^2 and β are replaced by their consistent estimates. For σ^2 , such an estimate is given by the mean of the r smallest eigenvalues of $\hat{\Sigma}(\mathbf{x}_i)$. As in Nordhausen et al. (2016), an estimate for β is given by

$$\hat{\beta} = \frac{1}{r(r+2)} \frac{1}{n} \sum_{i=1}^n [(\hat{\boldsymbol{\epsilon}}_{ki} - \bar{\boldsymbol{\epsilon}}_k)^\top \hat{\Sigma}(\hat{\boldsymbol{\epsilon}}_{ki})^{-1} (\hat{\boldsymbol{\epsilon}}_{ki} - \bar{\boldsymbol{\epsilon}}_k)]^2,$$

where $\hat{\boldsymbol{\epsilon}}_{ki} = \hat{\mathbf{R}}^\top \mathbf{x}_i$ and $\hat{\mathbf{R}}$ contains the last r eigenvectors of $\hat{\Sigma}(\mathbf{x}_i)$. If the assumption of normally distributed noise $\boldsymbol{\epsilon}$ is made, then the population value $\beta = 1$ can be substituted instead.

As in Nordhausen et al. (2016) and Virta and Nordhausen (2019), asymptotic tests for the different null hypotheses can be combined to yield an estimator for the signal dimension d . That is, for all $k = 0, \dots, p-1$, let $c_{k,n}$ be a deterministic sequence such that $c_{k,n} \rightarrow \infty$ and $c_{k,n} = o(n)$. Then, the least value of k for which the test statistic is asymptotically bounded is a consistent estimate of d ,

$$\hat{d} = \min \left\{ k \left| \frac{nr \cdot s_k^2(\hat{\Sigma}(\mathbf{x}_i))}{2\sigma^4 \beta} < c_{k,n} \right. \right\} \rightarrow_p d.$$

Despite the guarantee of consistency, the previous construction is difficult to use in practice as the sequences $c_{k,n}$ would need to be chosen very carefully to obtain good results also for finite samples. As such, we advocate carrying out the hypothesis testing (whether asymptotic or the bootstrap-based introduced in the next section) in practice by chaining together tests for different values of k and taking as the estimate, for example, the smallest value of k for which H_{0k} is not rejected on some particular level of significance.

4 | BOOTSTRAP NULL DISTRIBUTION

For the bootstrap approach, samples that mimic the distribution of the data and simultaneously satisfy the null hypothesis, even if it does not hold for the original data, need to be drawn.

If H_{0k} holds, all eigenvectors of $\Sigma(\mathbf{x})$ are given as $\mathbf{H} = (\mathbf{U}\mathbf{W} | \mathbf{R})$, where $\mathbf{W} \in \mathbb{R}^{k \times k}$ is an orthogonal matrix and $\mathbf{R} \in \mathbb{R}^{p \times r}$ has orthonormal columns and satisfies $\mathbf{R}^\top \mathbf{U} = \mathbf{0}$. Consequently, $\mathbf{R}^\top \mathbf{x}$ gives us access to the last r (rotated) noise components. Thus, regardless whether H_{0k} holds or not, replacing the values $\hat{\mathbf{R}}^\top \mathbf{x}_i$ with a sample from a spherical distribution lets us force the data distribution to satisfy H_{0k} .

Algorithm 1 Testing $H_{0k} : d = k$

Set the number of resamples R ;

Set the observed sample $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$;

Estimate $\bar{\mathbf{x}}$, $\hat{\Sigma}$ and the eigenvectors $\hat{\mathbf{H}}$ and the eigenvalues $\hat{\lambda}_1^2 \geq \dots \geq \hat{\lambda}_p^2$ of the latter;

Compute the principal components $\mathbf{S} = (s_1, \dots, s_n)$ with $s_i = \hat{\mathbf{H}}^\top (\mathbf{x}_i - \bar{\mathbf{x}})$, $\hat{\sigma}^2 = \sum_{i=k+1}^p \hat{\lambda}_i^2 / r$ and the test statistic \hat{t}_k ;

Choose a bootstrapping strategy;

for $j \in \{1, \dots, R\}$ do

$\mathbf{S}^* \leftarrow$ bootstrap sample of size n from \mathbf{S} ;

 Partition \mathbf{S}^* into \mathbf{S}_1^* and \mathbf{S}_2^* where \mathbf{S}_1^* contains the first k components and $\mathbf{S}_2^* = (s_{12}^*, \dots, s_{n2}^*)$ the remaining r components.;

 if Strategy = parametric then

$\mathbf{s}_{12}^* \leftarrow N_r(\mathbf{0}, \hat{\sigma}^2 \mathbf{I}_r)$;

 if Strategy = orthogonal then

$\mathbf{s}_{12}^* \leftarrow \mathbf{O}_i \mathbf{s}_{12}^*$ where \mathbf{O}_i is a random $r \times r$ orthogonal matrix;

 if Strategy = permute then

$\mathbf{s}_{12}^* \leftarrow \mathbf{P}_i \mathbf{s}_{12}^*$ where \mathbf{P}_i is a random $r \times r$ permutation matrix;

$\mathbf{x}_j^* = \hat{\mathbf{H}} \mathbf{s}_j^* + \bar{\mathbf{x}}$;

 Compute \hat{t}_k^* based on $\mathbf{X}^* = (\mathbf{x}_1^*, \dots, \mathbf{x}_n^*)$;

Return the p -value: $[\#(\hat{t}_k^* \geq \hat{t}_k) + 1] / (R + 1)$

In Algorithm 1, we propose two different ways of sampling from a spherical distribution: (a) replacing the vectors $\hat{\mathbf{R}}^\top \mathbf{x}_i$ with independent and identically distributed Gaussian vectors, which can be seen as parametric bootstrapping; and (b) spherifying the original $\hat{\mathbf{R}}^\top \mathbf{x}_i$ by random orthogonal transformations (this does not change the distribution of the noise process under the null). As randomly rotating all observations is computationally expensive and as permutations are a subset of rotations, we also consider as third strategy whether it is enough to permute the estimated noise components for each observation. The different bootstrapping approaches are detailed in Algorithm 1.

5 | NUMERICAL EXAMPLES

To evaluate our tests, we use the following noisy ICA design. We have three independent signals having a logistic, t_5 , and uniform distribution, all standardized to have mean zero and variance one. The noise is taken either from a $N_6(\mathbf{0}, \sigma^2 \mathbf{I}_6)$ -distribution (normal setting) or from a t_5 -distribution standardized to have covariance matrix equal to $\sigma^2 \mathbf{I}_6$ (t_5 setting). The sample sizes under consideration are $n = 1,000$ and $n = 2,000$. As the performance of the methods depends on the mixing matrix \mathbf{A} , we randomly choose at each iteration its 18 elements independently from $n(0, 1)$. We then record, on the basis of 1,000 repetitions, the rejection proportions at the level $\alpha = .05$ for the null hypotheses $H_{0k} = \{2, 3, 4\}$, where $k = 3$ corresponds to the correct dimension. We include in our comparisons the asymptotic test both with the constant β fixed to 1, that is, assuming Gaussian noise (Asymp Fix), and with $\hat{\beta}$ estimated from the data (Asymp Est). The three bootstrap variants are all applied using 200 bootstrap samples and respectively denoted Boot Para, Boot Orth, and Boot Perm.

The results are given in Figures 1 and 2 for different values of σ . In the normal case, we see that knowing the constant β for the asymptotic test is not a real benefit, but it has a huge negative effect when it is misspecified in the t_5 noise setting. Similarly, parametric bootstrap works well if the noise distribution is correctly specified but fails otherwise. In general, it seems that the asymptotic test that estimates the constant and the bootstrap test based on orthogonal transformations are the most preferable. They hold the alpha level and have good power, which, however, deteriorates with an increasing noise level. Permutation-based bootstrapping, on the other hand, has some trouble keeping the alpha level.

Restricting to the asymptotic test with estimated β , we consider three sound samples ($n = 50,000$), freely available in the R package JADE and used for example in Miettinen, Nordhausen, and Taskinen (2017). The three sound signals are mixed with a 20×3 matrix \mathbf{A} where the elements come from a uniform distribution on $[0, 1]$. The three nonzero eigenvalues of $\Sigma(\mathbf{Az})$ are 14.051, 2.391, and 0.639. We add then either 20-variate Gaussian or t_5 noise to the mixed signals and study at which noise level the number of signals gets undetectable by a divide-and-conquer strategy

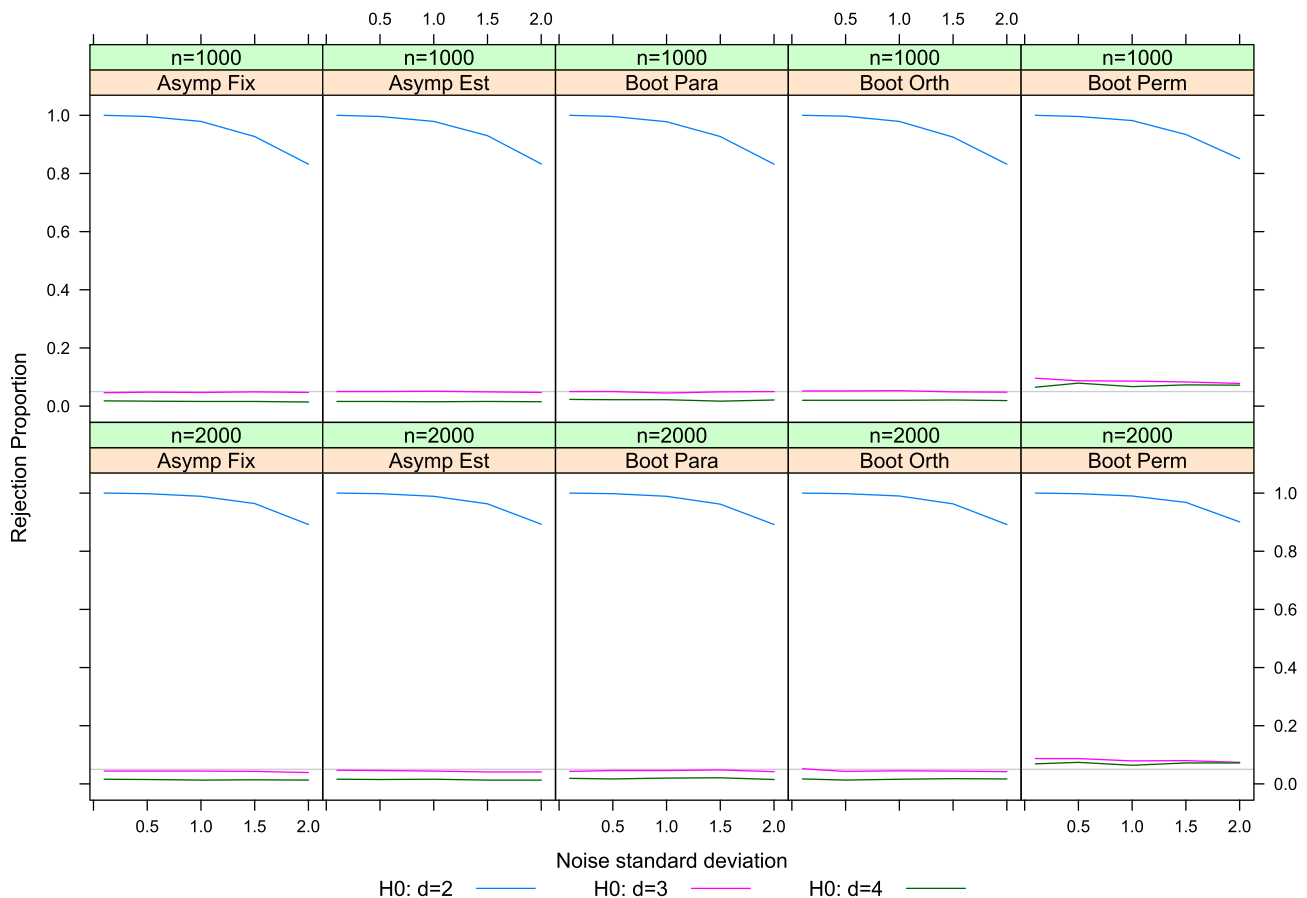


FIGURE 1 Rejection proportions for Gaussian noise based on 1,000 repetitions

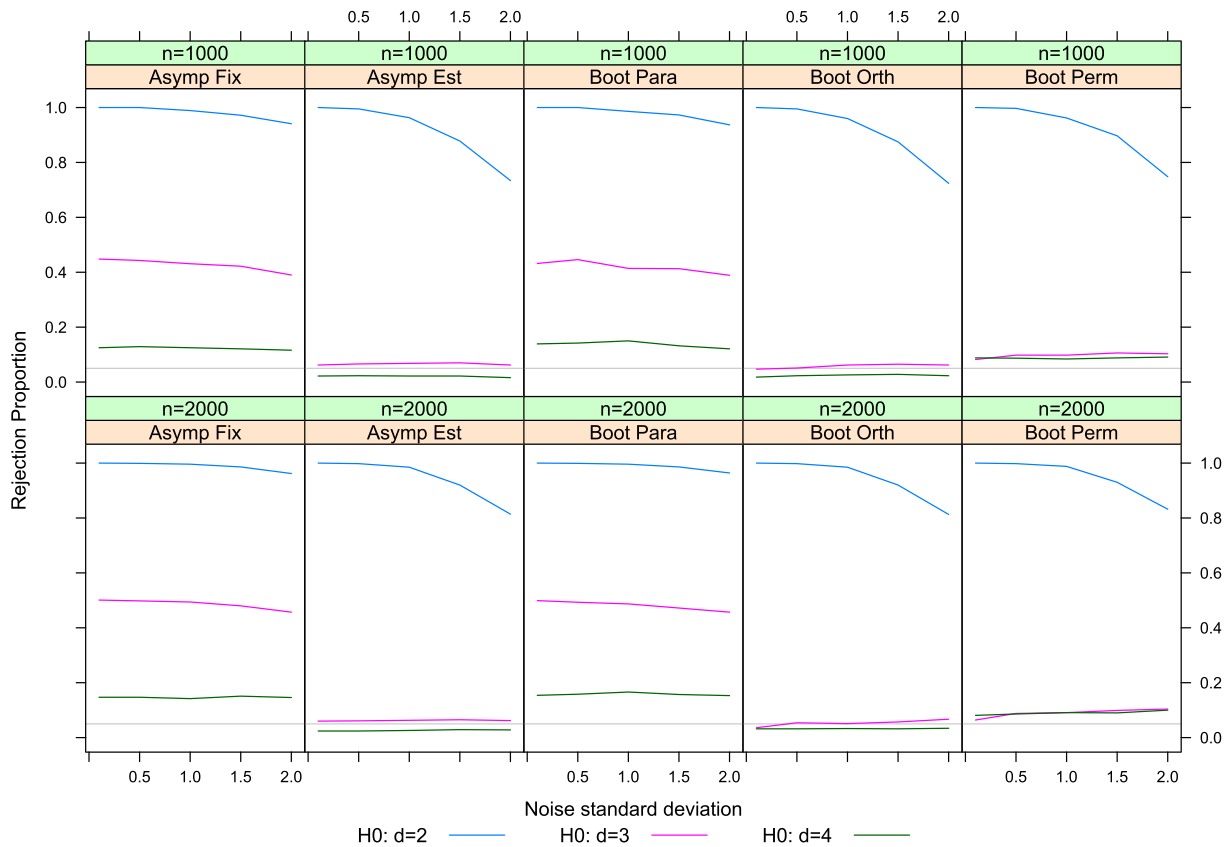


FIGURE 2 Rejection proportions for t_5 spherical noise based on 1,000 repetitions

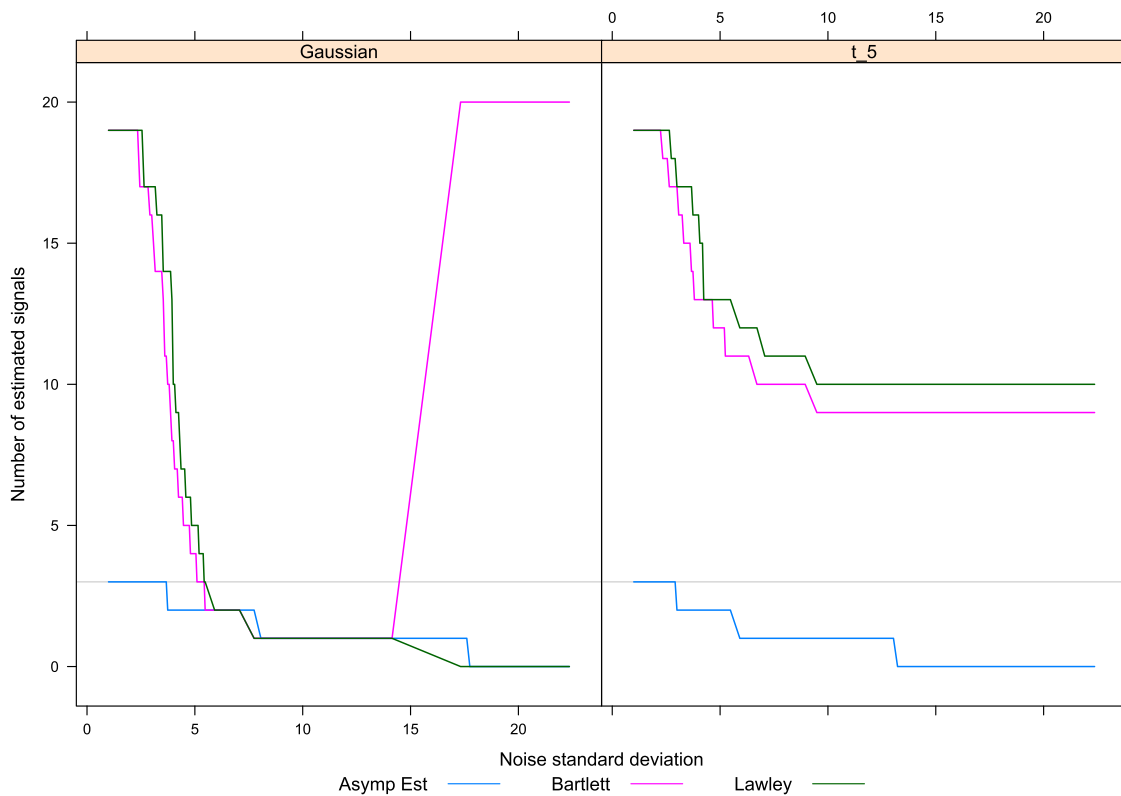


FIGURE 3 Estimated signal dimensions for different noise distributions and levels in the sound example. The true dimension is indicated by the grey horizontal line

(all individual tests are performed at level $\alpha = .05$). As competitors, we use the Bartlett test (Bentler & Yuan, 1996) and Lawley (1956) test for equality of eigenvalues in pure Gaussian model as implemented in the R package nFactors (Raiche, 2010).

The estimated signal dimensions are plotted against the noise standard deviations in Figure 3, with the grey horizontal line indicating the correct value. With Gaussian noise, our test can still detect all three signals if the noise standard deviation is at most 3.5 (meaning that its variance is roughly equal to the largest signal eigenvalue), and the corresponding number for t_5 noise is 2.5. And it is very clear that the two tests assuming a normal model in this context are both not very appropriate here. They hardly ever estimate the correct signal dimension but instead overestimate it quite dramatically for modest noise level and are in general too optimistic for the signal dimension.

6 | CONCLUSION

In this work, we developed inference tools to detect the true latent signal dimensionality using PCA. The problem has been previously considered in the context of noiseless elliptical data (Nordhausen et al., 2016), known as the subsphericity problem, and in a fully Gaussian case using different information criteria (see Chapter 3.2.3 of Cichocki & Amari, 2002, and the references therein). Our model, however, allows for any signal distribution and any spherical noise distribution, making it therefore of special interest in a noisy ICA context.

One drawback of the method is that it requires both the signal and the noise to have finite fourth moments. Consequently, the next step is to develop robust alternatives to the proposed method. Interestingly, additional simulation studies (not shown here) revealed that bootstrap testing actually succeeded in estimating the dimension also when the covariance matrix was replaced with certain robust alternatives, even though the alternatives lack the additivity property. The theoretical justification of this will provide an interesting future prospect, as will investigating whether the theoretical properties of various ICA estimators are retained in the noisy ICA model when the dimension is first estimated.

DATA AVAILABILITY STATEMENT

The data used in the sound example were obtained from the R package JADE (Nordhausen et al. (2018)), available for download from CRAN (<http://CRAN.R-project.org/package=JADE>).

ACKNOWLEDGEMENT

The work of K. N. was supported by CRoNoS European Cooperation in Science and Technology Action IC1408.

ORCID

Joni Virta  <https://orcid.org/0000-0002-2150-2769>

Klaus Nordhausen  <https://orcid.org/0000-0002-3758-8501>

REFERENCES

- Bentler, P. M., & Yuan, K. H. (1996). Test of linear trend in eigenvalues of a covariance matrix with application to data analysis. *British Journal of Mathematical and Statistical Psychology*, 49, 299–312.
- Cardoso, J. F., & Souselias, A. (1993). Blind beamforming for non-Gaussian signals. *IEE-Proceedings-F*, 140(6), 362–370.
- Choi, Y., Taylor, J., & Tibshirani, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *The Annals of Statistics*, 45, 2590–2617.
- Cichocki, A., & Amari, S. (2002). *Adaptive blind signal and image processing: Learning algorithms and applications*. Chichester, UK: Wiley.
- Comon, P., & Golub, G. H. (1990). Tracking a few extreme singular values and vectors in signal processing. *Proceedings of the IEEE*, 78, 1327–1343.
- Comon, P., & Jutten, C. (2010). *Handbook of blind source separation. Independent component analysis and applications*. Oxford, UK: Academic Press.
- Eaton, M., & Tyler, D. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *The Annals of Statistics*, 19, 260–271.
- Fang, K.-T., Kotz, S., & Ng, K.-W. (1990). *Symmetric multivariate and related distributions*. London, UK: Chapman and Hall.
- Jolliffe, I. (2002). *Principal component analysis*. Berlin Heidelberg: Springer.
- Karhunen, J., Cichocki, A., Kasprzak, W., & Pajunen, P. (1997). On neural blind separation with noise suppression and redundancy reduction. *International Journal of Neural Systems*, 08, 219–237.
- Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43, 128–136.
- Liavas, A. P., Regalia, P. A., & Delmas, J. P. (1999). Blind channel approximation: Effective channel order determination. *IEEE Transactions on Signal Processing*, 47, 3336–3344.
- Luo, W., & Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, 103, 875–887.
- Matilainen, M., Nordhausen, K., & Virta, J. (2018). On the number of signals in multivariate time series. In *International Conference on Latent Variable Analysis and Signal Separation* (Deville, Y., Gannot, S., Mason, R., Plumbley, M. D., & Ward, D., Eds.), Springer International Publishing, Cham, pp. 248–258.
- Miettinen, J., Nordhausen, K., & Taskinen, S. (2017). Blind source separation based on joint diagonalization in R: The packages JADE and BSSasympt. *Journal of Statistical Software*, 76, 1–31.

- Miettinen, J., Taskinen, S., Nordhausen, K., & Oja, H. (2015). Fourth moments and independent component analysis. *Statistical Science*, 30, 372–390.
- Minka, T. (2001). Automatic choice of dimensionality for PCA, In V. Tresp, T. Leen, & T. Dietterich (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 598–604). MIT Press, Cambridge, MA, USA.
- Nordhausen, K., Cardoso, J.-F., Miettinen, J., Oja, H., Ollila, E., & Taskinen, S. (2018). JADE: Blind source separation methods based on joint diagonalization and some BSS performance criteria. Retrieved from <http://CRAN.R-project.org/package=JADE>. R package version 2.0-1.
- Nordhausen, K., & Oja, H. (2018). Independent component analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10, e1440.
- Nordhausen, K., Oja, H., & Tyler, D. (2016). Asymptotic and bootstrap tests for subspace dimension. Preprint available as arXiv:1611.04908.
- Nordhausen, K., Oja, H., Tyler, D., & Virta, J. (2017). Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. *IEEE Signal Processing Letters*, 24, 887–891.
- Nordhausen, K., & Tyler, D. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika*, 102, 573–588.
- Nordhausen, K., & Virta, J. (2018). Ladle estimator for time series signal dimension. In *Proceedings of IEEE Statistical Signal Processing Workshop 2018, IEEE SSP 2018*, Freiburg, Germany, pp. 428–432.
- Raiche, G. (2010). nFactors: Parallel analysis and non graphical solutions to the Cattell scree test. <http://CRAN.R-project.org/package=nFactors>. R package version 2.3.3.
- Tipping, M., & Bishop, C. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 611–622.
- Virta, J. (2018). On characterizations of the covariance matrix. Preprint available as arXiv:1810.01147.
- Virta, J., & Nordhausen, K. (2019). Determining the signal dimension in second order source separation. To appear in *Statistica Sinica*. <https://doi.org/10.5705/ss.202018.0347>

How to cite this article: Virta J, Nordhausen K. Estimating the number of signals using principal component analysis. *Stat.* 2019;8:e231. <https://doi.org/10.1002/sta4.231>